

Sequence of the *Mrjp3* Microsatellite Locus in Honeybees of Different Origin

N. V. Ostroverkhova^{a, *}, A. N. Kucher^{a, b}, N. P. Babushkina^b, and O. L. Konusova^a

^aNational Research Tomsk State University, Tomsk, 634050 Russia

^bResearch Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, 634050 Russia

*e-mail: nvostrov@mail.ru

Received April 12, 2017; in final form, June 5, 2017

Abstract—The sequencing of the nucleotide sequences of the *mrjp3* repetitive region (*mrjp3* microsatellite locus) in Siberian honeybees was carried out. A high similarity of the studied nucleotide sequences ($\geq 99\%$ identity) with reference sequences was observed, which indicates a high conservation of the *mrjp3* repetitive region in different *Apis mellifera* subspecies.

Keywords: honeybee, *mrjp3* microsatellite locus, evolutionary branches M and C, *Apis mellifera mellifera* L., *Apis mellifera carpatica* Avet., Siberia

DOI: 10.1134/S1022795418030109

The *mrjp3* gene belongs to the *mrjp* subfamily, encoding the Major Royal Jelly Proteins (MRJP) and described in the genome of bees of the genus *Apis* and a number of other hymenoptera insects (the solitary, parasitoid jewel wasp *Nasonia vitripennis*, the alfalfa leafcutter bee *Megachile rotundata*, some ants, and bumblebees) [1–4]. In the process of evolution, the *mrjp* subfamily of genes arose owing to the multiple duplication of the ancestral *yellow* gene and can serve as an example when genes acquire new functions in evolution [4–9]. It is believed that the MRJP family of proteins (the *mrjp* family of genes) most likely evolved in the course of evolution precisely in connection with the function of producing royal jelly proteins—a special feed for the larvae of worker bees and drones up to three days old and queen bee larvae and queens throughout life (up to 4–5 years) [7, 8]. In addition, it is assumed that the genes of the *mrjp* family also have other functions, such as caste and sex differentiation, that is, they are associated with the development of a social way of life (eusociality) in the honeybee [4, 6].

The *mrjp3* gene includes a highly repetitive region, for which association with the production of royal jelly is shown [10–12]. The highly repetitive region is located in the 3'-coding region of the *mrjp3* gene; it is the VNTR region and is designated as *mrjp3* microsatellite locus. Analysis of nucleotide sequence of five allelic variants of the repetitive region identified in *Apis mellifera carnica* showed the presence in the structure of *mrjp3* locus of two different segments represented by 15-nucleotide repeats [1]. The *mrjp3* locus, correspondingly, encodes the C-terminal region of the MRJP3 protein, represented by 20–22

fold regularly repeating motif of five amino acid residues—(N/K/R)QN(A/G/D)(G/D/N) [1, 6].

At present, the *mrjp3* gene and the *mrjp* subfamily of genes as a whole are being rather intensively studied. For example, questions related to the evolution of the family, the expression of family genes in bees of different castes and at different stages of development, and the functional significance of individual genes and the whole family are examined [1–13]. At the same time, there are scarce data on the variability of the structure of genes of the *mrjp* subfamily, including the *mrjp3* microsatellite locus, in bees of different origin and different populations [1, 11, 12].

This aim of this work was to investigate the nucleotide sequence structure of the repetitive region of *mrjp3* gene (*mrjp3* locus) in Siberian honeybees.

MATERIALS AND METHODS

Samples of the DNA of worker bees obtained from bee colonies living in the apiaries of the Siberian region were investigated. DNA extraction and polymerase chain reaction (PCR) were performed according to standard protocols with some modifications [14] using the following primers for amplification of the *mrjp3* microsatellite locus: forward-5'–ATG TAA TTT TGA AGA ATG AAC TTG; reverse-5'–TGT AGA TGA CTT AAT GAG AAA CAC [10]. Genotyping was carried out in the collective center Medical Genomics of the Research Institute of Medical Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences (Tomsk), using an ABI Prism 3730 Genetic Analyzer and GeneScan500-ROX

DNA size standard under conditions recommended by the manufacturer. The size of fragments was analyzed using GeneMapper Software.

For the sequencing of the *mrjp3* locus, 12 samples of DNA of bees were selected: seven samples of the DNA of the Middle Russian bee *Apis mellifera mellifera* (from the apiaries from the villages of Ostyatskoye and Kolmogorovo in Krasnoyarsk krai and Zarechny in Tomsk oblast) and five DNA samples of the Carpathian bee *Apis mellifera carpatica* (from the Tomsk apiaries). The belonging to the subspecies was determined on the basis of the analysis of the mtDNA locus COI-COII, as well as the complex analysis of morphological and ethological characteristics [15–18]. Samples of bees homozygous for alleles of the *mrjp3* locus of the following sizes were used for the study: 406 bp (3 samples), 437 bp (3 samples), 518 bp (2 samples), and 529 bp (4 samples). Bees homozygous for alleles 406 and 518 bp in size originated from families of Carpathian bee (*A. m. carpatica*), and bees with alleles 437 and 529 bp in size originated from families of Middle Russian bee (*A. m. mellifera*).

Sequencing of the samples was carried out using an Applied Biosystems 3730 automatic analyzer with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's protocol. Data analysis was performed using Sequencing Analysis v5.4 software. The resulting nucleotide sequences were aligned with the BioEdit program [19]. The structure of the nucleotide sequences obtained during the present study was compared with the sequences presented in the genetic bank (NCBI Reference Sequence, GenBank) and scientific publication [1].

RESULTS AND DISCUSSION

The choice of alleles in this work to study the structure of the nucleotide sequences of the *mrjp3* locus was due to the wide spreading of the 406, 437, 518, and 529 alleles in honeybees of different origins (M and C evolutionary branches) inhabiting the territory of Siberia (on the basis of the results of study of 697 samples, unpublished data). The 529 allele was predominant in *A. m. mellifera* (the frequency of the allele in different samples varies from 0.600 to 0.929); for bees of southern origin (*A. m. carpatica* and *A. m. carnica*), this allele was recorded at a low frequency (less than 0.01). The 406 allele was typical of *A. m. carpatica* (the frequency of the allele exceeded 0.5) and it was not registered in *A. m. mellifera*. The 518 allele was found only in bees of southern origin (*A. m. carpatica* and *A. m. carnica*). The 437 allele was registered in both the bees of *A. m. mellifera* and the bees of southern origin (*A. m. carpatica*).

In the structure of the repetitive region of all compared sequences (406, 437, 518, and 529 alleles) (Table 1), two distinct segments are clearly distinguished,

described in the honeybees of *A. m. carnica* (Germany) [1]. The first segment is located closer to the 5'-end of the gene and is represented by 6–8 copies of the 15-nucleotide motif—AAT CAG AAT GCT (A/G)A(C/T). The repeat is highly conservative; 13 nucleotides of the motif are constant; nucleotide substitutions are noted only in the last triplet of the sequence (accordingly, the sequence of the protein is represented by five amino acid residues—NQNA(D/N/G)). The second type of repeat is less conservative; it can include from 5 to 20 copies of the 15-nucleotide motif—A(A/G)(A/G) CA(A/G) AAT G(A/G)T AA(C/T); correspondingly, the protein sequence—(K/R)QN(D/G)N [1].

In the studied alleles of the *mrjp3* locus in Siberian bees (Table 1), the first segment also includes a 15-nucleotide motif repeated 6–8 times, but is characterized by an additional single nucleotide substitution at the 11th position described in three variants of the sequence (437, 529, and 518 alleles)—C→T, which leads to the replacement of the amino acid alanine with valine in the protein sequence (A→V). The second segment includes 14–20 copies of the 15-nucleotide motif; in the sequence of the variant 2 (529 allele), a single nucleotide replacement G→A was found, leading to a change in the protein sequence (glycine to serine)—(K/R)QN(D/G/S)N.

Analysis of the nucleotide sequences of the second segment made it possible to distinguish a 30-nucleotide motif as a structural unit rather than a 15-nucleotide one, which can be represented as follows: A—(B)_{3–5}—C—(B')_{1–2}—D. Moreover, the B and B' motifs, repeating the most times, while preserving the nucleotide sequence of 15-nucleotide structural fragments differ from each other in the order of their position: the sequence of the first 15 nucleotides of the B motif corresponds to a sequence of 15 nucleotides at the end of the B' motif; and vice versa, the sequence of the last 15 nucleotides of the B motif corresponds to that of the 15-nucleotide fragment at the beginning of the B' motif. The B and B' motifs are flanked on both sides by different 30-nucleotide sequences (A, C, and D), which are identical in all alleles (Table 1). The D motif differs between alleles with the last three nucleotides (CAG, AAT, or AAC). With great probability, it can be assumed that the sequence of the second segment evolved by duplicating the 30-nucleotide motif, despite the fact that it is assumed that the repeats of the first and second segments originated from a common ancestral sequence 15 bp in size.

The repetitive region ends with a sequence of 33 nucleotides, identical in all variants. In some cases, the repetitive region and the 33-nucleotide sequence are separated by short fragments of different lengths (AATCAG—518 allele and CAA—406 allele), which disturb the tandem-repeating structures. However, it should be noted that these changes are triplet sequences, which supports the selection against the

Sequence		Variant 1: 437 allele	Variant 2: 529 allele
Sequence size		412 bp	472 bp
Sequence before the repetitive region			
Segment 1, 15-nucleotide motifs	motif 1	TCGTTGCGGAAAGATATCAC	TCGTTGCGGAAAGATATCAC
	motif 2	(AATCAGAATGCTGGC) ₂	(AATCAGAATGCTGGC) ₃
	motif 3	(AATCAGAATG(C/D)TGAC) ₂	(AATCAGAATG(C/D)TGAC) ₂
	motif 4	AATCAGAATGCTAAC	AATCAGAATGCTAAC
	motif 5	AATCAGAATGCTGAT	AATCAGAATGCTGAT
	motif 6	AATCAGAATGCTAA(T/C)	AATCAGAATGCTAAC
number of motifs	8	8	8
Segment 2, 30-nucleotide motifs	motif 1 (A)	AAACAAAATGGTAATAGACAAAATGATAAC	AAACAAAATGGTAATAGACAAAATGGTAAC
	motif 2 (B)	(AGACAGAATGATAACAAGCAAAATGGTAAAC) ₄	(AGACAGAATGATAACAAGCAAAAT(G/A)GTAA(C/T)) ₅
	motif 3 (C)	AGACAAAATGGTAACAACACAGAATGATAAC	AGACAAAATGGTAACAACACAGAATGATAAC
	motif 4 (B')	AAGCAAAAATGGTAAACAGACAGAATGATAAC	(AAGCAAAAATGGTAA(T/C)AGACA(A/G)AATGATAAC) ₂
	motif 5 (D)	AAGAGGAATGGTAAACAGGCAAAAATGATCAG	AAGAGGAATGGTAAACAGGCAAAAATGATCAG
number of motifs	8	10	10
Additional insertion			
		No	No
Sequence after the repetitive region			
		AATAATCAGAATGATAATAATCGAAAATGATAAT	AATAATCAGAATGATAATAATCGAAAATGATAAT
Sequences for comparison			
	Mrip3c [#]		XM_016917723.1 [#]

Table 1. (Contd.)

Sequence	Variant 3: 518 allele	Variant 4: 406 allele
Sequence size	463 bp	355 bp
Sequence before the repetitive region	<i>TCGTTGCGGAAGATATCAC</i>	<i>TCGTTGCGGAAGATATCAC</i>
motif 1	(AATCAGAATGCTGGC) ₂	AATCAGAATGCTGGC
motif 2	(AATCAGAATG(C/T/C)TGAC) ₃	(AATCAGAATGCTGAC) ₂
motif 3		AATCAGAATGCTAAC
motif 4	AATCAGAATGCTGAT	AATCAGAATGCTGAT
motif 5		
motif 6	AATCAGAATGCTAAC	AATCAGAATGCTAAC
number of motifs	7	6
motif 1 (A)	AAACAAAATGGTAATAGACAAAATGATAAC	AAACAAAATGGTAATAGACAAAATGATAAC
motif 2 (B)	(AGACAGAATGATAACAAGCAAATGGTAA(C/T)) ₅	(AGACAGAATGATAACAAGCAAATGGTAA(C/T)) ₃
motif 3 (C)	AGACAAAATGGTAAACAAACAGAATGATAAC	AGACAAAATGGTAAACAAACAGAATGATAAC
motif 4 (B')	(AAGCAAAAATGGTAA(T/C)AGACA(A/G)AATGATAAC) ₂	AAGCAAAAATGGTAAACAGACAGAATGATAAC
motif 5 (D)	AAGAGGAATGGTAACAGCAAAAATGATAAC	AAGAGGAATGGTAAACAGCAAAAATGATAAT
number of motifs	10	7
Additional insertion	<u>AATCAG</u>	<u>CAA</u>
Sequence after the repetitive region	<i>AATAATCAGAATGATAATAATCGAAAATGATAAT</i>	<i>AATAATCAGAATGATAATAATCGAAAATGATAAT</i>
Sequences for comparison	AY663104.1** Mrip3d**	GU434675.1# NM_001011601.1# Z26318.1# Mrip3a#

* The allele sizes are given according to the PCR analysis of the *mrip3* locus. Nucleotide substitutions are in bold; the sequences located at the beginning and end of the repetitive region of the *mrip3* gene and identical in all the samples are in italics. The sequence numbers in the Genetic Bank (NCBI Reference Sequence, GenBank) and the allele names described in *A. m. carnica* [1], with which (**) 99% or (*) 100% identity was observed, are shown in the line "Sequences for comparison". For the 518 allele, 99% identity with the reference sequence AY663104.1 is shown; the differences are shown as single nucleotide substitutions in both the first and second segments.

shift of the reading frame, and, accordingly, indicates the functional significance of conservative amino acid residues beyond the repetitive region of MRJP3. In addition, if the nucleotide sequences behind the repetitive region are highly conserved (100% identity) in different alleles, then the size of the repetitive region can differ by 120 nucleotides [1]. For example, comparison of two nucleotide sequences of the *mrjp3* microsatellite locus (529 and 406 alleles), typical of *A. m. mellifera* and *A. m. carpatica*, respectively, showed 98% compliance with each other; differences were recorded only by the number of repeats of individual motifs that were observed in both the first and second segments.

A comparative analysis of the nucleotide sequences of the *mrjp3* locus of Siberian population bees with reference sequences showed the following results.

Variant 1 (437 allele) is completely identical to the *Mrjp3c* sequence described in *A. m. carnica* [1]. Variant 2 (529 allele) completely coincides with the XM_016917723.1 sequence described in hybrid BeeWeaver bees. Variant 3 (518 allele) showed 99% identity with the sequence presented in GenBank under the number AY663104.1 and 99% coincides with the variant *Mrjp3d* described in *A. m. carnica* (AY663104.1 and *Mrjp3d* are identical) [1]. The revealed differences between the two sequences (518 allele and AY663104.1) are associated with single nucleotide substitutions (A→G, C→T, A→G, and T→C), as well as various additional insertions (518 allele—AATCAG sequence; AY663104.1—CAA) (Table 1). Variant 4 (406 allele) showed 100% identity to the sequences represented in GenBank under the numbers GU434675.1, NM_001011601.1, and Z26318.1, as well as 100% identity to the variant *Mrjp3a* described in *A. m. carnica* [1].

Of the five alleles (*Mrjp3a*, *Mrjp3b*, *Mrjp3c*, *Mrjp3d*, and *Mrjp3e*) identified as a result of sequencing in bees of *A. m. carnica* (Germany) [1], two alleles (406 and 437) were registered in the bees of the Siberian region, which had a 100% identity with *Mrjp3a* and *Mrjp3c*, respectively. The 518 allele had a 99% identity with the *Mrjp3d* sequence. Thus, the 406 and 518 alleles registered in *A. m. carpatica* bees in the Siberian population are identical to the alleles described in *A. m. carnica* of the European population.

The 529 allele, prevailing in the Middle Russian bee *A. m. mellifera* inhabiting Siberia, showed the complete identity of the sequence of the *mrjp3* repetitive region with a sequence of the BeeWeaver bee, which is a hybrid of different subspecies, including the *A. m. mellifera* dark-colored forest bee. This coincidence is of considerable interest: on one hand, the sequence is specific to the bees of Siberian populations (including the isolated population of the Krasnoyarsk krai, existing for more than 60 years without importing bees from other populations), and on the other hand, the sequence has survived in hybrid BeeWeaver bees,

which have undergone a lengthy artificial selection and have in their gene pool genetic material of several European and African subspecies. This variant (529 allele) is the longest repetitive region of all identified alleles.

Thus, a high level of identity between the sequences of the *mrjp3* microsatellite locus studied not only in bees from different populations (Siberia and Europe) but also in different subspecies of honey bees (*A. m. mellifera*, M branch; *A. m. carnica* and *A. m. carpatica*, C branch) was noted, which can be evidence of the conservatism and functional significance of the repetitive region of the *mrjp3* gene.

The evolutionary significance of the repetitive region of the *mrjp3* gene in a honeybee is also indicated by the comparison of the nucleotide sequence of the *mrjp3* gene between four species of honeybees (*A. mellifera*, *A. cerana*, *A. dorsata*, and *A. florea*). Thus, in all studied species of bees, the presence of an extensive repetitive region characterized by a significant similarity in structure and nucleotide composition in different species was determined [3]. The revealed polymorphism between the *Apis* species, primarily, was associated with different lengths of the repetitive region. Thus, the *mrjp3* repetitive region in *A. mellifera* was shorter in comparison with the locus of *A. dorsata* and *A. cerana*. For example, the first segment of *A. mellifera* included a 15-nucleotide motif, repeated 6–8 times, while in *A. dorsata* the motif was repeated 9–12 times; the 15-nucleotide motif of the second segment in *A. mellifera* was repeated 5–21 times, and in *A. dorsata*, 24–26 times [2].

In addition to the *mrjp3* gene, the repetitive regions were also found in other genes of the *mrjp* family—the *mrjp2* and *mrjp5* genes. Similar to the MRJP3 protein, a repeat of five amino acid residues in the C-terminal region of the MRJP2 protein was detected; a repetitive motif of three amino acid residues in the other part is described for the MRJP5 protein. It is assumed that the repetitive regions developed independently in the MRJP family of proteins, with all the repetitive regions of the MRJP family proteins containing a large number of nitrogen-rich amino acids. Their presence significantly increases the nitrogen content in royal jelly. It is assumed that the repetitive regions are the regions that store nitrogen in a biologically accessible form [1]. Owing to the fact that a correlation was found between the content of nitrogen in MRJP proteins and the length of repeat, it is likely that the spreading of repeats is associated with a selection aimed at increasing the nitrogen content for more efficient nutrition of queens and larvae [3]. These assumptions are consistent with the results of this study, which shows a wide distribution of the longest repetitive region in honeybees of different origin and geographical location.

REFERENCES

1. Albert, S., Klaudiny, J., and Simúth, J., Molecular characterization of MRJP3, highly polymorphic protein of honeybee (*Apis mellifera*) royal jelly, *Insect Biochem. Mol. Biol.*, 1999, vol. 29, no. 5, pp. 427–434.
2. Albert, S., and Simúth, J., Characterization of major royal jelly protein-like DNA sequences in *Apis dorsata*, *J. Apicult. Res.*, 2002, vol. 41, nos. 3–4, pp. 75–82. doi 10.1080/00218839.2002.11101072
3. Albertová, V., Su, S., Brockmann, A., et al., Organization and potential function of the *mrjp3* locus in four honeybee species, *J. Agric. Food Chem.*, 2005, vol. 53, no. 20, pp. 8075–8081. doi 10.1021/jf051417x
4. Buttstedt, A., Moritz, R.F., and Erler, S., More than royal food—Major royal jelly protein genes in sexuals and workers of the honeybee *Apis mellifera*, *Front. Zool.*, 2013, vol. 10, no. 72.
5. Drapeau, M.D., Albert, S., Kucharski, R., et al., Evolution of the Yellow/Major royal jelly protein family and the emergence of social behavior in honey bees, *Genome Res.*, 2006, vol. 16, pp. 1385–1394.
6. Buttstedt, A., Moritz, R.F., and Erler, S., Origin and function of the major royal jelly proteins of the honeybee (*Apis mellifera*) as members of the yellow gene family, *Biol. Rev.*, 2014, vol. 89, no. 2, pp. 255–269. doi 10.1111/brv.12052
7. Albert, S., Bhattacharya, D., Klaudiny, J., et al., The family of major royal jelly proteins and its evolution, *J. Mol. Evol.*, 1999, vol. 49, no. 2, pp. 290–297.
8. Schmitzová, J., Klaudiny, J., Albert, S., et al., A family of major royal jelly proteins of the honeybee *Apis mellifera* L., *Cell. Mol. Life Sci.*, 1998, vol. 54, no. 9, pp. 1020–1030.
9. Albert, S. and Klaudiny, J., The MRJP/YELLOW protein family of *Apis mellifera*: identification of new members in the EST library, *J. Insect Physiol.*, 2004, vol. 50, no. 1, pp. 51–59.
10. Baitala, T.V., Faquinello, P., Toledo, V.A.A., et al., Potential use of major royal jelly proteins (MRJPs) as molecular markers for royal jelly production in Africanized honeybee colonies, *Apidologie*, 2010, vol. 41, pp. 160–168. doi 10.1051/apido/2009069
11. Parpinelli, R.S., Ruvolo-Takasusuki, M.C.C., and Toledo, V.A.A., MRJP microsatellite markers in Africanized *Apis mellifera* colonies selected on the basis of royal jelly production, *Genet. Mol. Res.*, 2014, vol. 13, no. 3, pp. 6724–6733.
12. Ruvolo-Takasusuki M.C.C., Pozza A.P.B.C., Oliveira A.P.N.Z. et al. Improvement and selection of honeybees assisted by molecular markers, *Beekeeping and Bee Conservation—Advances in Research*, Dechechi Chambo, E., Ed., InTech, Croatia, 2016, pp. 63–75. doi 10.5772/62426.10.5772/62426
13. Ji, T., Liu, Z., Shen, J., et al., Proteomics analysis reveals protein expression differences for hypopharyngeal gland activity in the honeybee, *Apis mellifera carnica* Pollmann, *BMC Genomics*, 2014, vol. 15, no. 665. PMID: 25103401. doi 10.1186/1471-2164-15-665
14. Ostroverkhova, N.V., Konusova, O.L., Kucher, A.N., et al., Population genetic structure of honeybee (*Apis mellifera* L.) near the village of Leboter, Chinskii rayon, Tomsk oblast, *Vestn. Tomsk. Gos. Univ., Biol.*, 2013, no. 1 (21), pp. 161–172.
15. Ostroverkhova, N.V., Konusova, O.L., Kucher, A.N., et al., Genetic diversity of the locus COI-COII of mitochondrial DNA in honeybee populations (*Apis mellifera* L.) from the Tomsk region, *Russ. J. Genet.*, 2015, vol. 51, no. 1, pp. 80–90. doi 10.1134/S102279541501010X
16. Konusova, O.L., Ostroverkhova, N.V., Kucher, A.N., et al., Morphometric variability of honeybees *Apis mellifera* L., differing in the variants of the mtDNA COI-COII locus *Vestn. Tomsk. Gos. Univ., Biol.*, 2016, no. 1 (33), pp. 62–81. doi 10.17223/19988591/33/5
17. Ostroverkhova, N.V., Konusova, O.L., Kucher, A.N., and Sharakhov, I.V., A comprehensive characterization of the honeybees in Siberia (Russia). *Beekeeping and Bee Conservation—Advances in Research*, Dechechi Chambo, E., Ed., InTech, Croatia, 2016, pp. 1–37. doi 10.5772/6239510.5772/62395
18. Ostroverkhova, G.P., Konusova, O.L., and Pogorelov, Yu.L., *Biologicheskaya i khozyaistvennaya otsenka pchelinoi sem'i (Apis mellifera L.)* (Biological and Business Assessment of Honey Bee (*Apis mellifera* L.) Family), Tomsk: NTL, 2005.
19. Hall, T.A., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp.*, 1999, vol. 41, pp. 95–98.

Translated by A. Kashevarova

SPELL: I. valine