BRIEF COMMUNICATION

Complementary DNA sequences encoding the multimammate rat MHC class II DQ α and β chains and cross-species sequence comparison in rodents

J. Goüy de Bellocq¹ & H. Leirs¹,²

¹ Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
² Danish Pest Infestation Laboratory, Department of Integrated Pest Management, University of Aarhus, Skovbrynet 14, DK-2800 Kongens Lyngby, Denmark

Key words
DQ; major histocompatibility complex class II; multimammate rat; Muridae; RACE

Abstract
Sequences of the complete open reading frame (ORF) for rodents major histocompatibility complex (MHC) class II genes are rare. Multimammate rat (Mastomys natalensis) complementary DNA (cDNA) encoding the α and β chains of MHC class II DQ gene was cloned from a rapid amplifications of cDNA Ends (RACE) cDNA library. The ORFs consist of 801 and 771 bp encoding 266 and 256 amino acid residues for DQB and DQA, respectively. The genomic structure of Mana-DQ genes is globally analogous to that described for other rodents except for the insertion of a serine residue in the signal peptide of Mana-DQB, which is unique among known rodents.

The Muridae is the largest family of mammals with >730 recognized species (1), including two classical laboratory models, the mouse (Mus musculus) and rat (Rattus norvegicus), and many reservoirs of human diseases [e.g. reservoirs of old world arenaviruses causing haemorrhagic fevers (2)]. To date, the genetic structure of the major histocompatibility complex (MHC) of Muridae has been intensively investigated in the two laboratory models – labelled as the H2 system in the mouse and the RT1 complex in the rat (3–5).

Mastomys natalensis, the multimammate rat, is a widely distributed African murid (6), a pest causing agricultural damage and/or acting as a natural reservoir of human pathogens (7). In West Africa, M. natalensis is the reservoir of Lassa virus, an arenavirus causing haemorrhagic fever (8). In southern Africa, M. natalensis plays a role in the transmission of the bacteria Yersinia pestis – the aetiologic agent of plague – between wild and commensal rodents (9). M. natalensis is used as laboratory model of human diseases such as papillomavirus infections (10) and for the investigation of bilharziasis, filariasis, relapsing fever and plague (11, 12). It is surprising, then, that the multimammate rat MHC has never been characterized.

The primary aim of our work was to clone transcripts of M. natalensis MHC class II DQ and thereby provide data for further studies on pathogen resistance in natural populations. As sequences encoding the complete open reading frame (ORF) of DQ genes are available for a few other rodent species, we were able to compare the complementary DNA (cDNA) sequences of the multimammate rat with those of other rodents and relate them to those of other mammal species.

The individuals were obtained from a laboratory colony established in 2006 in the Danish Pest Infestation Laboratory. Progenitors of the colony were captured in Morogoro,
Tanzania, in 2006. Four individuals were randomly chosen from the F1 generation. Total RNA was isolated from liver preserved in RNA later using RNase-free DNase set (Qiagen Vertriebs GmbH, Wien, Austria). RNA from the four individuals was pooled two by two into two samples and used as a template for full-length cDNA synthesis using an RNA ligase-mediated RACE with the GeneRacer kit (Invitrogen GmbH, Lofer, Austria). Polymerase chain reaction (PCR) amplifications of both the 5′ and the 3′ ends of the cDNA were performed in two separate reactions using primers provided with the RACE kit and primers designed from mouse and rat full-length cDNA synthesis using an RNA ligase-mediated two by two into two samples and used as a template for Wien, Austria). RNA from the four individuals was pooled column RNase-free DNase set (Qiagen Vertriebs GmbH, preserved in RNAlater using RNeasy mini kit with on-

Characterization of Mastomys MHC class II DQ sequences

For the DQB gene, the entire coding sequence with some 5′ and 3′ UTRs was also successfully amplified by RACE. The 5′ cDNA ends amplified using GeneRacer 5′ and MasDQA-Ex3R-Race primers were ~400 bp long. Three different sequences were obtained from the eight sequenced plasmid inserts. The sequences were identified by BLAST as encoding the DQA gene. The 3′ cDNA ends amplified using the GeneRacer 3′ and MasDQA-Ex2F-Race primers were ~850 bp long. Three different sequences were obtained from the five sequenced plasmid inserts and were identified as coding for DQA gene. Comparing the 168 shared nucleotides between the 5′ and the 3′ cDNA ends, one of the 3′ end sequence matched one of the 5′ end sequence. DQA sequences were named Mana-DQA*01 for the complete coding sequence, Mana-DQA*02 and Mana-DQA*03 for the incomplete 5′ end coding sequences and Mana-DQA*04 and Mana-DQA*05 for the incomplete 3′ end coding sequences (AN: FJ968787–FJ968791). The nucleotide sequences showed high homology with rodent sequences (BLAST: identity with M. musculus: 91%, R. norvegicus: 90% and Peromyscus eremicus: 87%). The ~1100 bp contain a 771 bp ORF coding 256 residues (Figure 2).

The deduced amino acid sequences were aligned and numbered according to the structure of RT1-B of the rat (14) and H2-IA of the mouse (17) (Figures 1 and 2). SIGNAL.P predicted the signal peptide cleavage site between Gly and Arg at −1 and 1 and between Gly and Gln/Glu at −1 and 1a for the mature Mana-DQB and Mana-DQA, respectively. The predicted signal peptide cleavage sites were similar to those for other rodents and mammals, although the human, cattle and dog signal peptides of DQB are four amino acids longer than that of Mastomys (14). The β65–66 deletion, very common in rat and mouse haplotypes (14, 17), was not found in the three Mastomys sequences (Figure 1). In rodents, this deletion has only been described in mouse and rat haplotypes (14, 17), very common in rat and mouse haplotypes (14). The 3′ end coding sequences and

For the DQB gene, the entire coding sequence and some 5′ and 3′ untranslated regions (UTR) were amplified by RACE. The 5′ cDNA ends amplified using GeneRacer 5′ and MusDQB-R2 primers were ~400 bp long. Three different sequences were obtained from the 12 sequenced plasmid inserts: one sequence was identified by BLAST as a DRB allele and the remaining sequences as two different DQB alleles. The 3′ cDNA ends amplified using the GeneRacer 3′ and BetaUniv-F primers were ~900 bp long. Four different sequences were obtained from the 12 sequenced plasmid inserts: 3 were identified as coding for the DQB gene and the last one for the haemopexin gene (BLAST 95% identity with haemopexin of M. musculus).

For the DQA gene, the entire coding sequence with some 5′ and 3′ UTRs was also successfully amplified by RACE. The 5′ cDNA ends amplified using GeneRacer 5′ and MasDQA-Ex3R-Race primers were ~400 bp long. Three different sequences were obtained from the eight sequenced plasmid inserts. The sequences were identified by BLAST as encoding the DQA gene. The 3′ cDNA ends amplified using the GeneRacer 3′ and MasDQA-Ex2F-Race primers were ~850 bp long. Three different sequences were obtained from the five sequenced plasmid inserts and were identified as coding for DQA gene. Comparing the 168 shared nucleotides between the 5′ and the 3′ cDNA ends, one of the 3′ end sequence matched one of the 5′ end sequence. DQA sequences were named Mana-DQA*01 for the complete coding sequence, Mana-DQA*02 and Mana-DQA*03 for the incomplete 5′ end coding sequences and Mana-DQA*04 and Mana-DQA*05 for the incomplete 3′ end coding sequences (AN: FJ968787–FJ968791). The nucleotide sequences showed high homology with rodent sequences (BLAST: identity with M. musculus: 91%, R. norvegicus: 90% and Sigmodon hispidus: 88%). The ~1100 bp contain a 771 bp ORF coding 256 residues (Figure 2).
The amino acid sequence of the signal peptide in the Mana-DQB protein differs from those of other rodents because of insertion of a Ser residue at position 217. This residue is an Ala in HLA-DQB1, DLA-DQB1, and BoLA-DQB. The structural characteristics of Mana-DQA did not differ from those of other rodents. The numbering of Mana-DQB and Mana-DQA was checked by identifying the position of evolutionary conserved amino acids: the Asn and Thr that provide a site for carbohydrate attachment were identified at amino acids 19 and 21; the Cys that form an intradomain disulfide bond was identified at amino acids 82 and 83, respectively. In summary, we have characterized cDNA encoding the entire mature proteins of the Mana-DQA and DQB class II molecules. Comparison with other rodent DQ proteins identifies a major discrepancy in the leader peptide of the Mana-DQB protein, which is insertion of a Ser residue. There has been no study investigating the functional role of polymorphism in the signal peptide in MHC class II molecules. However, this region shows some polymorphism in humans, and the signal peptide of the Mana-DQB shows one polymorphic site.

Figure 1 Comparison of the translated amino acid sequences of Mastomys natalensis major histocompatibility complex class II DQB gene (Mana-DQB) and other representative of murids, rodents and mammals. Sequences are split according to their domain structure. The symbols used indicate the following: -, indel; ., same amino acids as those in Mana-DQB*01; /, alignment filler. Numbering of Mana-DQB includes two insertions of amino acids designated −17 and 84a (in bold). The sources of the sequences are as follows: Mus musculus H2-Iab (GenBank accession number: M13537–M13541 and AH002012), Rattus norvegicus RT1-Bb (AY626180–AY626183, AY626186, AY626187 and AY626189), Ctenomys haigi Ctha-DQB (AF312534), Ctenomys sociabilis Ctso-DQB (AF312530), Peromyscus leucopus Pele-DQB (AF300846), Peromyscus eremicus Peer-DQB (AF300853), Peromyscus maniculatus Pema-DQB (AF300855), Bos taurus BoLA-DQB (Y18201), Canis lupus familiaris DLA-DQB (NM_001014381) and Homo sapiens sapiens HLA-DQB1*030201 (IMGT/HLA AN: HLA00627). Positions of primers used for the RACE, BetaUniv-F (5′-acctcacaagggacgcac-3′) and MusDQB-R2 (5′-cccctcgtatggttgtctgc-3′) are underlined. HLA, human leucocyte antigen.
humans, signal sequences of MHC class I molecules have been shown to play a role in NK-cell-mediated lysis through their presentation by the class I human leucocyte antigen-E molecule (22). The interspecific and intraspecific polymorphisms of signal peptide shown here suggest that the functional role of leader peptide sequences in MHC class II molecules deserves closer inspection.

Acknowledgments

The authors thank F. Suchentrunk and the Research Institute of Wildlife Ecology in Vienna for access to laboratory facilities. JGB is a postdoctoral fellow with the Fund for Scientific Research – Flanders (FWO). Financial support was provided by the University of Antwerp (grant number 2964) and the FWO (grant number 1.5.180.09).

References


Figure 2 Comparison of the translated amino acid sequences of Mastomys natalensis MHC class II DQA gene (Mana-DQA) and other representative of murids, rodents and mammals. Sequences are split according to their domain structure. Numbering of Mana-DQA includes an insertion in exon 2, designated amino acid 9a and three extra amino acids at the start of the mature protein designated 1a, 1b and 1c (in bold). These numbers are inserted for structural consistency across MHC class II molecules (15). The symbols used indicate the following: -, indel; ., same amino acids as those in Mana-DQA*01; ~, alignment filler. The sources of the sequences are as follows: Mus musculus H2-IAa (K01922–K01926 and M11356–M11358), Rattus norvegicus RT1-Ba (X14879, AY626190, AY626191, AY626193, AY626194, AY626196, AY626197 and AY626199), Peromyscus leucopus Pele-DQA (U34805), Sigmodon hispidus Sihi-DQA (NM_001011726) and Homo sapiens sapiens HLA-DQA*0101 (IMGT/HLA AN: HLA00601). Positions of the primers used for the RACE, MasDQA-Ex2F-Race (5'-acctggctgaccagcaagc-3') and MasDQA-Ex3R-Race (5'-tgttgggctgacccagcagcac-3') are underlined. HLA, human leucocyte antigen; MHC, major histocompatibility complex.


