ORIGINAL ARTICLE

Characterization and genetic diversity of MHC class II *DRB* genes in the Arabian camel (*Camelus dromedarius*)

Mohamed Habib Yahyaoui[1,2]

Abstract

This study investigated the MHC *DRB* genes in the Arabian camel (*Camelus dromedarius*). The results revealed the presence of — at least — two transcribed DRB-like genes in chromosome 20, designated *MhcCadr-DRB1* and *MhcCadr-DRB2*. These genes are 155 Kb apart, have similar gene structure, and are transcribed in opposite directions. Compared to *DRB1*, the *DRB2* locus contains a deletion of 12 nucleotides in the second exon (270 bp), exhibits lower transcript abundance, and is expressed as two splice variants differing by exon 2 skipping. This gene seems to be of minor functional relevance in the dromedary camel. Conversely, the *DRB1* is thought to be the main gene in this species showing higher transcript abundance and polymorphism levels. A total of seven *DRB1* exon 2 alleles were identified in the Tunisian dromedary camel resulting from 18 amino acid substitutions. Six full length alleles were characterized at the mRNA level. Although there is no clear evidence for balancing selection (i.e., heterozygote advantage), signals of weak historical positive selection acting on the *DRB1* gene were detected, as indicated by the limited number of the sites being positively selected. This trend might be related to the low exposure to pathogens and to the demographic history of the species. Comparative analysis with Bactrian and wild camel genomes suggested occurrence of trans species polymorphism (TSP) in the *Camelus* genus. The results lay the foundation for the MHC *DRB1* genetic diversity analysis in this genus since the developed genotyping protocols are fully applicable in the three *Camelus* species.

Keywords *Camelus dromedarius* · MHC class II · DRB · Genetic diversity · Tunisia

Introduction

Major histocompatibility complex (MHC) plays a central role in adaptive immunity of jawed vertebrates shaping immune defenses against a wide range of pathogens. The MHC consists of a large region of the genome containing several loci, including classical class I and class II genes coding for cell surface glycoproteins (or MHC molecules) involved in the processing and presentation of antigens to T lymphocytes. These antigens are small peptides derived by proteolysis from self- and pathogen-expressed proteins in the host. The first class molecules are present on the surface of all nuclear cells and bind to peptides derived from intracellular parasites (e.g.,

virus), whereas class II are present only on specific immune cells (macrophages, B cells, and dendritic cells) and are involved in the recognition of extra cellular pathogens (e.g., bacteria, protozoa). There are different isotypes of MHC class II molecules, including the DR isotype, found in most mammalian species. These molecules are heterodimers of two non-covalently associated α and β subunits encoded by adjacent genes MHC *DRA* and MHC *DRB*. While the *DRA* gene is usually single copy, the presence of multiple copies of the *DRB* gene appears to be the rule in most mammalian species. For example, nine loci (including five pseudogenes) have been mapped in human MHC (Horton et al. 2004) and three (one pseudogene) in cattle (Burke et al. 1991). However, the functional relevance of these loci is variable due to their varying levels of expression. For instance, *DRB1* is the main *DRB* gene in human and is the most highly expressed (Emery et al. 1993), whereas only the *DRB3* gene is considered to be functionally relevant in cattle (Burke et al. 1991). Structurally, the β-chain, encoded by the *DRB* gene, consists of the extracellular β1- and β2-domains, the transmembrane

✉ Mohamed Habib Yahyaoui
   Habib.Yahyaoui@ira.rnrt.tn

1 Livestock and Wildlife Laboratory LR16IRA04, Arid Lands Institute, Medenine, Tunisia

2 Department of Life Sciences, Faculty of Sciences, University of Gabès, Gabès, Tunisia

domain and the cytoplasmic tail. The β1-domain, which is encoded by the *DRB* exon 2, is part of the structure of the peptide binding groove where foreign peptides are recognized and processed. Residues within this domain that directly bind to antigens define the peptide binding regions or PBR (Brown et al. 1993) and thus determine the antigen binding repertoire. These regions are often found to be under positive diversifying selection resulting in a high number of alleles that allow the host immune system to react to a wide and variable range of foreign pathogens found in the host environment. As a result, the *DRB* genes usually exhibit high levels of polymorphisms mainly clustered in the second exon coding for the β1 domain (Babik 2010). For instance, a total of 384 *DRB3* alleles have been detected in cattle as shown in the IPD-MHC (Immuno Polymorphism Database, https://www.ebi.ac.uk/ipd/mhc/; Maccari et al. 2017) database, and several alleles have been associated with diseases in different breeds (Behl et al. 2012).

The Arabian camel (*Camelus dromedarius*) or dromedary is the one-humped camel belonging to the genus *Camelus* that also includes the domesticated Bactrian (*Camelus bactrianus*) and wild (*Camelus ferus*) two-humped camels. The dromedary is a common livestock in several arid and semi-arid regions of the world with significant socio-economic importance allowing production of meat and milk in these areas where the climate precludes many other animals. The species is notoriously known for its resilience and adaptation to harsh environments characterized by extreme temperatures, drought, low-quality food, and scarcity of the resources. Notwithstanding this impressive adaptation to harsh conditions, the dromedary is indeed susceptible to various infectious and parasitic pathogens (Khalafalla 2017) although it showed also resistance to some infectious diseases such as viral FMD (foot and mouth disease, Wernery and Kinne 2012) which affects ruminant species causing a significant economic impact. The dromedary could also be infested by a number of ectoparasites such as ticks and fleas (Sazmand et al. 2019) and has been identified as potential reservoir of the zoonotic Middle East respiratory syndrome (MERS) virus (Dawson et al. 2019).

The humoral immune system of the dromedary camel and other camelids is characterized by the presence of antibodies (IgG2 and IgG3) devoid of lights chains (Hamers-Casterman et al. 1993) in addition to the conventional antibodies (IgG1) found in other species. As regard adaptive immunity, the MHC region displays an overall organization similar to the structure of the mammalian MHC with genes ordered into class I-class III-class II and has been mapped to the long arm of chromosome 20 (Avila et al. 2014; Plasil et al. 2016). Currently, studies regarding the MHC II genes and their diversity in dromedary are very scarce despite the importance of the species as a livestock animal and its capacity of adaptation and resilience to harsh environment challenges. For instance,

these genes have not yet been characterized at the mRNA level and the available data were derived solely from genomic DNA with no knowledge of the expression status or the number of loci being analyzed. However, this knowledge is primordial for interpreting the significance and the functional relevance of the detected MHC diversity. In dromedary — and in other *Camelus* species — the MHC II genes appeared to display higher mean nucleotide diversity than other immune and rest of genome genes as estimated from genome-wide analysis in nine dromedaries (Lado et al. 2020). On the other hand, an unexpectedly low level of polymorphism has been previously reported in MHC class II *DR* and *DQ* genes in dromedary as well as in Bactrian and wild camels (Plasil et al. 2016). As far as the *DRB* gene is concerned, only one non synonymous mutation has been observed from a total of four variable sites detected in the analyzed exon 2 spanning 270 nucleotides (54 animals were analyzed).

This work sets out with the aim to study the main MHC *DRB* gene in dromedary camel at both mRNA and DNA levels. Specifically, the objectives of the study were to (i) identify and characterize the full length MHC DRB cDNA in dromedary using Sanger sequencing and determine gene structure, (ii) assess the extent of the exon 2 polymorphism and genotype the alleles in the Tunisian dromedary population, and (iii) investigate genetic diversity (segregating sites, nucleotide diversity) and the evolutionary process that shape this diversity. The results showed the presence of at least two transcribed *DRB* genes in the dromedary genome differing in the size of exon 2 (270 and 282 bp), transcript abundance, and the level of polymorphism. Furthermore, occurrence of alternative splicing resulting from exon 2 skipping was observed in one locus. The findings highlight the importance of including cDNA analysis in the investigation of the MHC diversity.

Materials and methods

Sample collection and DNA extraction

A total of 132 blood samples (about 5 ml) were collected in EDTA tubes from unrelated animals in different herds representing all dromedary camel ecotypes found in Tunisia. The panel includes two trios (sire, dam, and offspring) allowing gene segregation analysis. Samples used for RNA extraction were obtained from six healthy adult animals (three males and three females) belonging to the experimental herd of the Arid Lands Institute of Medenine (Tunisia). An additional blood sample for allele validation at the mRNA level was taken from private herd, preserved in RNA later, and stored at –20 °C. Genomic DNA was extracted from blood samples according to the standard phenol–chloroform protocol (Sambrook et al. 1989) and DNA quantity and quality were evaluated by absorbance using UV5Nano spectrophotometer

(Mettler Toledo) and by agarose gel electrophoresis. All sampling, animal handling, and study procedures were in compliance with local regulations.

RNA extraction and cDNA synthesis

Peripheral blood mononuclear cells (PBMCs) were isolated from fresh EDTA blood samples using Histopaque 1077 (Sigma), washed twice with PBS pH 7.2 (Gibco, Life Technologies), and homogenized in TRIzol reagent (Sigma) using rotor–stator homogenizer (Kinematica). For RNA later-preserved sample, blood cells were collected from 2.4 ml after thawing and centrifugation for 1 min and processed as for PBMCs using TRIzol. Total RNA was isolated according to the manufacturer protocol and purified with the PureLink RNA Mini Kit (Ambion, Life Technologies). RNA quality and concentration were assessed on an UV5Nano spectrophotometer (Mettler Toledo) at 260/280 nm wavelength. All RNA samples were treated with DNase I (Thermo Scientific) prior to cDNA synthesis in order to remove genomic DNA. Reverse transcription was performed using the MMLV reverse transcriptase (Invitrogen) following manufacturer protocol in a final volume of 20 μl containing 1 μg of RNA and 2.5 μM of the oligo d(T)15 primer (Promega). Single strand cDNA was diluted ten-fold and stored at −20 °C.

In silico analysis and primer development

Human DRB1 and bovine DRB3 mRNA sequences (GenBank accession numbers NM_001243965 and NM_001012680, respectively) were used as queries to search dromedary camel ESTs (Al-Swailem et al. 2010) using standalone BLAST (version 2.9.0+, Altschul et al. 1990). These ESTs were generated from several tissues including blood from nine middle-eastern dromedary camels. The retrieved most similar hits ($n = 5$) were blasted against the NCBI non-redundant nucleotide (nr) database using online BLAST in order to confirm they were DRB-like in other mammalian species. Besides, they were further searched in the three available dromedary camel genome assemblies to infer gene structure (exon–intron boundaries following canonical AG/GT rule), and in the dromedary RNAseq available in the NCBI database as transcriptome shotgun assembly (TSA, bioproject

PRJNA82161). The top hit ESTs were aligned to DRBs from other species, including the human DRB1, the bovine DRB3, and the porcine DRB1 (NM_001113695) mRNAs using Clustal omega available in the EBI web server (https://www.ebi.ac.uk/Tools/msa/clustalo/) to compare gene structure and features. Primers used for amplification of the full-coding region of the DRB1 gene were designed based on these detected ESTs. The remaining primers (DRB2 cDNAs and DRB1 exon 2) were constructed over genomic DNA after aligning transcript (exons) to a genomic sequences using BLAST and Splign (www.ncbi.nlm.nih.gov/sutils/splign) tools. All primers (Table 1) were designed using Primer3 version 4.1 (Untergasser et al. 2012), checked with Primer-Blast (NCBI) and SMS (sequence manipulation suite) version 2 (Stothard 2000).

Amplification of the DRB cDNA and gene

The full coding region of the DRB1 cDNA was amplified using primers Cd.DRB1.F-R (Table 1) constructed on exons 1 and 6, respectively. The PCR reactions were performed in a final volume of 25 μl containing 1X PCR buffer, 1.5 mM MgCl$_2$, 0.2 mM dNTPS, 0.4 μM of each primer, 0.5 U of Go Taq DNA Polymerase (Promega), and about 20 ng of cDNA. Thermal profile consisted of an initial denaturation step at 94 °C for 2 min, 35 cycles of denaturation (94 °C–30 s), annealing at 62 °C for 45 s, extension (72 °C–1 min), and a final extension step at 72 °C for 10 min. The same PCR conditions were used for the DRB2 cDNA except for MgCl$_2$ (2 mM) and annealing temperature (63 °C).

Two overlapping fragments spanning whole DRB1 gene were amplified with primers Cd.DRB1.F-I2R (exon1-intron2,~5.3 Kb) and Cd.DRB1.E2F-R (exon2-exon6,~4.5 Kb) using LongAmp Taq DNA Polymerase (New England Biolabs). Amplifications were performed in a total volume of 25 μl and contained 1X LongAmp buffer, 300 μM dNTPS, 1.5 mM MgCl$_2$, 0.4 μM of each primer, 2.5 units of LongAmp Taq polymerase, and about 50 ng of DNA. Thermal cycling started with one cycle at 94 °C for 2 min, followed by 35 cycles of denaturation (94 °C–30 s), annealing (62 °C for 30 s), and extension (65 °C–5 min), with a final extension step at 65 °C for 10 min.

Table 1 Primers used for the analysis of the dromedary camel DRB1 and DRB2 genes and cDNAs. Underlined nucleotides in the Cd.DRB1.E2F primer correspond to intron 1 sequences and the remaining bases from exon 2. The two first nucleotides in the 5′-end of the Cd.DRB2.F primer were added in order to increase melting temperature (GA in gDNA)

Locus	Primer	Sequence (5′-3′)	Location	Ta (used in PCR)
DRB1	Cd.DRB1.F	TGCTCCCATCACACGCTATC	Exon 1	62 °C
	Cd.DRB1.R	CGAACCAGGAGCAAAGCAGA	Exon 6	
	Cd.DRB1.I1F	GGTGACTGACTCTTTCCTGTCC	Intron 1	66 °C
	Cd.DRB1.I2R	CTCTCTCTCCCTCTCTCACACA	Intron 2	
	Cd.DRB1.E2F	GTCCACACAGCACGTTTCATG	Junction intron1/exon 2	65 °C
DRB2	Cd.DRB2.F	CCGCTGTGGGTCTTAATCCTCAC	Exon 1	63 °C
	Cd.DRB2.R	AGGCGCTGCAATCAGTGCT	Exon 6	

Amplification of the *DRB1* exon 2

The entire sequence of the *DRB1* exon 2 (282 bp) was amplified using primers Cd.DRB1.I1F—I2R (Table 1) designed in introns 1 and 2, respectively, resulting in a fragment of 382 nucleotides. All samples were genotyped for exon 2 using this set of primers, including samples used for cDNA analysis in order to compare gDNA and cDNA sequences. PCR reactions were carried out as above (for DRB1 cDNA) in either 10 or 25 µl using Go Taq polymerase (Promega) and 10–30 ng of genomic DNA. Cycling conditions were as follows: initial denaturation at 94 °C for 2 min, then 35 cycles of 94 °C for 30 s, 66 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 10 min. Exon 2 (partial sequences) was also amplified using the forward primer Cd.DRB1. E2F used for gene amplification. This primer was designed at the intron1-exon2 boundary allowing the amplification of a 364 bp-fragment including 271 bp from exon 2 (11 nucleotides missed) when used with Cd.DRB1.I2R reverse primer. A subset of a randomly selected samples ($n = 79$) with known genotype and representing all established alleles were amplified using this set of primers (at annealing temperature of 65 °C) and directly sequenced in order to further validate the genotyping procedure.

Sequencing of MHC *DRB* genes

All PCR reactions included a negative control (no DNA) and products were visualized by electrophoresis in 1.5% agarose gel in 1×TAE buffer containing ethidium bromide (0.5 µg/ml). Prior to sequencing, the PCR products were either gel-extracted using Purelink gel extract kit (Invitrogen) or purified with exonuclease I and shrimp alkaline phosphatase (Thermo Fischer). Purified fragments were directly sequenced in both directions using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI PRISM 3500 Genetic Analyzer (Applied Biosystems). All samples were sequenced at least twice from two independent PCRs (ranging from 2 to 5×). In order to phase alleles from heterozygote samples, DNA fragments (DRB1 cDNA) were cloned using the TOPO TA cloning kit (Invitrogen) and transformed into competent JM109 cells. Positive clones were confirmed for the presence of correct insert by colony PCR and by enzymatic digestion with *Eco*R I endonuclease (Thermo Fischer). Plasmid DNA from five clones was purified with EZ-10 Spin Column Plasmid DNA minipreps kit (Bio Basic, Canada) and sequenced using M13 forward and reverse primers.

Sequence analysis, allele detection, and nomenclature

DNA sequences were analyzed with Sequencing Analysis software v6.0 (Thermo Fischer) and all chromatograms were manually checked. The obtained sequences were aligned using Clustal omega and compared to the NCBI non-redundant nucleotide (nr) database and to the dromedary camel genomic and transcriptome sequences using BLAST. Coding sequences were translated with SMS v2. For *DRB1* genotyping, the obtained homozygote exon 2 sequences were compared to the alleles initially established from full coding cDNAs, and heterozygote samples were resolved using Haplofinder script (Miltiadou et al. 2003). Individuals showing alleles with new mutations were re-sampled — when possible — in order to extract RNA and obtain the full coding cDNA sequences. Alleles were defined on the basis of the standard criterion in MHC genotyping (Babik 2010): sequences are considered to be valid alleles only when they are detected at least twice in two different PCRs from the same or different individual. Alleles were designated following the proposed nomenclature for the MHC in different species (Klein et al. 1990) with serial numbers according to the observed frequencies.

Data analysis

Various indices of nucleotide and amino acid polymorphisms of the *DRB1* exon 2 were calculated using DnaSP v6.12 (Rozas et al. 2017) and MEGAX (Kumar et al. 2018). These include nucleotide diversity (π), haplotype diversity (Hd), segregating and parsimony informative sites, and overall mean nucleotide and amino acid genetic distances based on Tamura 3-parameter (T92) evolutionary distances and Poisson corrected amino acid distances, respectively. Genepop v4.7 (Rousset 2008) was used to calculate allele and genotype frequencies and to test for Hardy–Weinberg (HW) equilibrium.

The occurrence of recombination events was examined with GARD method (Kosakovsky Pond et al. 2006) available in the datamonkey server (http://datamonkey.org/, Weaver et al. 2018) and using methods implemented in RDP5 (Recombination Detection Program, Martin et al. 2021) software. These methods include RDP, Geneconv, Maxchi, Chimaera, Bootscan, SiScan, and 3Seq, all used with default settings and at alpha value of 0.05 for significance.

Footprint of positive selection acting on individual sites of the *DRB1* exon 2 was evaluated by analyzing the ratio of non-synonymous (dN) to synonymous (dS) substitutions (dN/dS or ω) using maximum likelihood (ML) and Bayesian inference approaches. In the ML approach, sites subject to episodic positive or diversifying selection were detected using the MEME (mixed effects model of evolution, Murrell et al. 2012) method in the HyPhy package implemented in the datamonkey server. Bayesian inference analysis was carried out within MrBayes version 3.2.7 (Ronquist et al. 2012) using the Ny98 model with default priors and two independent runs of four Markov chain Monte Carlo (MCMC) chains for 150,000 generations with a burn-in of 25%. Convergence was assessed following default settings

using the average standard deviation of split frequencies between runs (< 0.01), ESS (estimated sample size, > 100), and PSRF (potential scale reduction factor, ~ 1) indices. Codons identified by these two approaches are considered to be under positive selection, whereas codons are considered only potentially positively selected when detected by MEME method or when ω (dN/dS) > 3 and the probability of being positively selected is ≥ 0.9 using Bayesian inference. Since the dromedary *DRB1* exon 2 contains an insertion of 12 nucleotides compared to human *DRB1*, it is difficult to infer residues of the peptide binding region (PBR) and thus to partition the exon 2 into PBR and non PBR codons. Therefore, only overall dN/dS ratio on the entire exon 2 was calculated using DNaSP v6.12 following the modified Nei–Gojobori method with Jukes–Cantor correction.

Analyses of the phylogenetic relationships among the *DRB1* exon 2 alleles were performed using ML method in MEGAX and the JC + I model (Jukes and Cantor 1969) which is considered the best fitting model on the basis of the Bayesian information criterion (BIC). The tree robustness was assessed with 2000 bootstrap replications, and outgroup included exons 2 from *Bos taurus DRB3* (NM_001012680) and *Sus scrofa DRB1* (NM_001113695) genes.

Results

Evidence of two transcribed DRB loci in the Arabian camel genome

In order to identify functional *DRB* genes in dromedary camel, we firstly focused on transcriptome data through BLAST analysis of the available EST database. We retrieved several contigs including a transcript of 1.3 Kb (CL83contig1) showing 85 and 83% similarity to bovine DRB3 and human DRB1, respectively. Searching for this contig in the dromedary camel genomes revealed the presence of six coding exons exhibiting complete identity to the sequences from the ASM164081v1 genome assembly. Furthermore, several highly similar transcripts were also detected in the dromedary TSA databases.

In addition to these identical matches (to CL83contig1) found in the genomic sequences, further hits covering nearly the entire sequence but with gaps and less identity were observed in the three dromedary camel genomes assemblies available in the NCBI databases. These hits, corresponding to six putative exons, were detected in the minus strand of the dromedary reference genome (assembly accession: GCF_000803125.2) and sequence discrepancies and gaps were mainly located in the 5′ region of the transcript. Identical sequences to these putative exons were detected in raw RNAseq SRA (Sequence Read Archive) though none were found in ESTs and TSA. Close sequence analysis showed that this locus has been annotated in the reference genome

assembly (LOC105101410) as DRB-like albeit the locus product was reported as low quality protein due to the discrepancy between genome and mRNA sequences.

Comparative analysis revealed that these two putative *DRB* loci displayed similarity to the *DRB* genes from mammalian species as shown by BLAST search against the nonredundant (nr) and the reference sequence RNA (refseq rna) NCBI databases. These observations clearly indicate that the Arabian camel genome contains at least two transcribed *DRB* loci, designated *MhcCadr-DRB1* (initially detected in ESTs) and *MhcCadr-DRB2* (observed in the minus strand) in accordance to the proposed nomenclature for the MHC in different species (Klein et al. 1990). These genes are located approximately 155 Kb apart in the chromosome 20 of the dromedary reference genome and are transcribed in opposite directions (Fig. 1A).

Low abundance and alternative splicing of the DRB2 mRNA

To determine whether the *DRB2* locus is transcribed in our panel, a primer pair flanking the coding region (Table 1) was designed. PCR amplification of cDNA derived from seven dromedaries produced a fragment of expected size (~ 1.1 Kb) in addition to a low-size fragment of about 0.8 Kb (Fig. 1B). Sequence analysis from both fragments showed that they correspond to two DRB2 transcript variants (named X1 and X2) resulting from an RNA alternative splicing event. Alignment of the X1 isoform (1032 bp) to the dromedary genomes revealed that the *DRB2* gene is spanned over six exons with an exon 2 size of 270 nucleotides, similar to the *DRB* genes from other species such as human, bovine, and porcine. The open reading frame contains 801 bp coding for 266 amino acids. The locus is mapped to 20,873,286–20,863,270 positions (~ 10 Kb) within the minus strand of the dromedary chromosome 20 (GenBank accession: CM016646.2). Compared to X1, the X2 isoform (764 bp) lacks 270 nucleotides as a result of exon 2 skipping during RNA splicing (Fig. 1B) with no evidence of frameshift or premature codon stop, and translating to a predicted protein of 176 amino acids. The *DRB2* exon 2 sequences are of low complexity, showing a high GC content (about 67.4%). A total of nine SNPs were detected, all in the X2 isoform from two individuals, including five non-synonymous (three in exon 1 and two in exon 3), two synonymous in exons 1 and 3, and two in the 3′-untranslated region. The sequences of the DRB2 X1 and X2 isoforms have been submitted to the GenBank under accession numbers OQ106969 and OQ106970, respectively.

The dromedary DRB2 mRNA (X1) is virtually identical to the genomic sequences from domestic Bactrian (*Camelus bactrianus*) and wild (*Camelus ferus*) camels, with only four SNPs observed in exon 1 and one SNP in the 3′-UTR (exon 6). The *DRB2* locus was found in the plus strand of chromosome
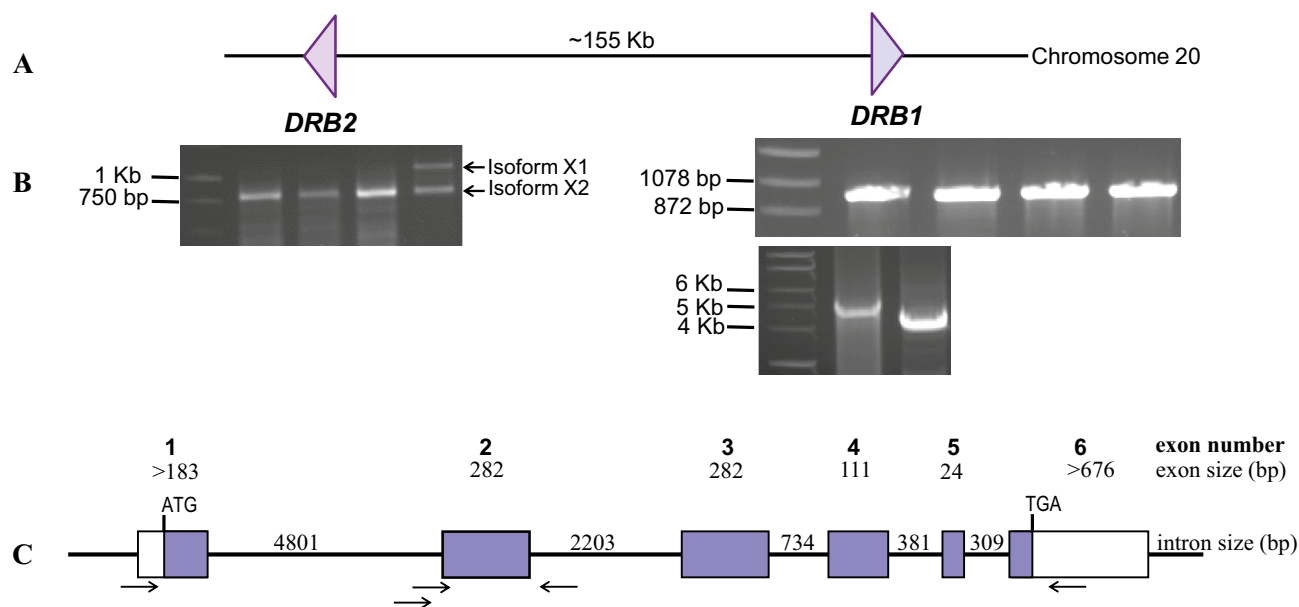
Fig. 1 Dromedary camel MHC *DRB1* and *DRB2* genes. **A** Relative position of the *DRB1* and *DRB2* genes in chromosome 20 of the reference genome. Genes are represented as arrows indicating the direction of transcription. **B** Amplification of the DRB2 cDNA (left) showing isoforms X1 and X2, and the DRB1 cDNA (upper right); lower right, amplification of the whole *DRB1* gene (~9.5 Kb) in two overlapping fragments. **C** Schematic representation of the dromedary camel MHC *DRB1* gene. Introns are shown as lines and exons as rectangles with coding region in purple and UTR in white (sizes are not to scale). ATG indicates the start codon and TGA depicts the stop codon. Arrows represented approximate position of primers used for gene amplification

20 (and *DRB1* in minus strand) in the genomes of both species and the same sequence orientation has been observed in the previous assembly of the dromedary reference genome (Camdro2, GCA_000803125.2). This sequence orientation is likely the correct one, and the sequences were probably reversed during the assembly of the current version of the dromedary reference genome Camdro3. The inter genic region between the two *DRB* loci spanned about 155 and 74 Kb in the dromedary and wild camel genomes, respectively. Compared to other species, the DRB2 mRNA (X1) displayed a nucleotide similarity ranging from ~86% (bovine, caprine) to 81.4% (porcine) with intermediate levels for other common species (human, horse, dog, cat, sheep, chimpanzee). Contrary to DRB1, the DRB2 transcript was not detected neither in the ESTs (derived from blood among other tissues) nor in the TSA databases, though identical sequences were found in the raw RNAseq reads generated from blood of three dromedaries (bioproject accession: PRJNA701734). Although the expression level was not quantitatively assessed, this finding suggests that the *DRB2* is rather a low abundant transcript, which is also indicated by the observed faint cDNA amplification (Fig. 1B) despite extensive attempts of PCR optimization.

From the seven individuals sequenced, four were homozygous for the X2 variant and thus lacking exon 2, three were heterozygous X1/X2, and none homozygous for the full X1 variant. The functionality of the DRβ proteins in mammals is primarily determined by the β1 domain, which is a structural part of the peptide-binding cleft. Residues within this domain are encoded by exon 2, including the amino acids in the PBR that interact directly with the foreign antigens. Consequently, although the *DRB2* gene is transcribed as shown by mRNA analysis, it is unlikely that it would be of functional relevance due to the exon 2 skipping during transcription. These observations, together with the noticed low abundance of the transcripts, clearly indicate that the *DRB2* is not the main *DRB* gene in the dromedary camel and thus, it is of minor importance as regard the adaptive immunity in this species.

Characterization of the DRB1 cDNA and gene

As previously mentioned, the *DRB1* locus corresponds to the transcript (CL83contig1) detected from the EST database displaying complete identity to the dromedary genomic sequences and high similarity to the bovine DRB3 and human DRB1 mRNAs. A couple of primers (Table 1) was designed based on this contig to PCR amplify the full coding region using cDNA derived from four animals. Subsequently, three animals were included in order to validate alleles detected during exon 2 diversity screening. Contrary to the DRB2, a sharp, high intensity expected fragment (980 bp) without extra bands was observed (Fig. 1B). Alignment of the obtained sequences to dromedary genomes revealed that the *DRB1* gene consists of six exons (Fig. 1C) with an

open reading frame of 813 bp encoding for 270 amino acids. All exons were coding and the intron sizes varied from 309 (intron 5) to 4801 (intron 1) nucleotides. The gene size is about 9.5 Kb. The dromedary DRβ1 protein contains an additional four residues (GEQG) at positions 74–77 when compared to DRβ from other mammalian species (266 aa) as a result of an insertion of 12 nucleotides in exon 2 (282 bp) and this feature seems to be specific to the *Camelidae* family. The four amino acids include an acidic residue (glutamic acid) along with one polar (glutamine) and to two non-polar (glycine) residues. The insertion is located upstream of the major part of the PBR sites (16 from a total of 24 codons) making it difficult to infer dromedary PBR by comparison to human DRβ1. The DRB1 had 88% nucleotide similarity and 79.6% amino acid similarity with the DRB2 (isoform X1). Alignment of the dromedary DRB1 and DRB2 nucleotide and protein sequences are shown in Fig. S1.

Several transcripts showing complete identity to the obtained DRB1 cDNA sequences were found in the dromedary transcriptome data (ESTs, TSA, and SRA) indicating a wide expression and transcript abundance and corroborating the observations from cDNA amplification (Fig. 1B). Two cDNA sequences (911 bp) obtained from two different animals were identical to CL83contig1 and to the genomic sequences from one contig detected in the ASM164081v1 genome assembly containing the whole *DRB1* gene (LSZX01099695, positions: 3231–12,569). On the contrary, we found several discrepancies when these sequences were compared to the current assembly of the dromedary reference genome (GCF_000803125.2) likely due sequence or assembly errors of this region of the genome. Specifically, although almost identical sequences of the 5′-region of the gene (from exon 1 to exon 3 with two missed nucleotides in exon 2) were observed, the sequences completely diverge from exon 3 (starting at position 54) onwards. Based on these partial sequences, a DRB-like locus was predicted (LOC105101413) in the NCBI reference genome annotation (the locus product was annotated as low quality protein). In this regard, similar sequence errors and equivocal annotation have been previously reported in the MHC region of the dromedary reference genome (Lado et al. 2021). The *DRB1* gene has been correctly mapped to the previous assembly version Camdro2 (GCA_000803125.2) though with minor sequence errors (281 bp of exon 2).

Given these observed discrepancies and in order to assess the size and the structure of the *DRB1* gene, two overlapping fragments covering the whole gene were amplified by PCR using genomic DNA and their ends were directly sequenced. The size of amplified fragments is ~9.5 Kb as expected (Fig. 1B) and the obtained sequences (exons one, two, and from the end of intron three to exon six as well as partial introns 1 and 2) were identical to the sequences of the contig LSZX01099695. This finding thus confirms that this contig indeed contains the *DRB1* gene. The retrieved genomic sequences have been submitted to the GenBank under accession numbers OQ106978-80. The *DRB1* gene is highly conserved in the genus *Camelus*; the sequences of the whole gene in the dromedary (LSZX01099695:3231–12,569) and in the wild camel (VSZR01000013:c21690641-21,681,287) were practically identical even in introns. Only an insertion of 16 nucleotides in the repetitive sequences at the start of the intron 2 and one SNP in intron 3 were observed in the *Camelus ferus* gene compared to the dromedary camel. Relatively high similarities were also observed between the dromedary DRB1 and the DRBs from other mammalian species at both nucleotide (83–87%) and protein (75–80%) levels.

Polymorphism screening and genetic diversity of the DRB1 exon 2

A total of 132 animals were successfully genotyped by direct sequencing of the entire exon 2 (282 bp) amplified using primers Cd-DRB1.I1F-I2R (Fig. S2). The panel includes two families (sire, dam, and offspring) as well as the seven animals genotyped at the mRNA level. Patterns of chromatograms obtained after direct sequencing are consistent with the presence of a single *DRB1* locus since no more than two peaks were observed in all obtained sequences. Complete identity was also observed between the cDNA and the genomic DNA sequences derived from the same animals. Thirty three (33) variable nucleotide sites in exon 2 were detected in the analyzed panel translating in 18 amino acid changes, corresponding to 19% of the 95 encoded residues (Table 2). Only one site showed three possible amino acid residues at position 104 in the protein, all remaining nucleotide and protein sites were bi-allelic. Overall, an intermediate level of nucleotide diversity (π) was observed (~0.049) in the analyzed population (Table 2).

A total of seven exon 2 alleles (haplotypes) resulting from the combination of these polymorphisms were identified in the Tunisian dromedary population and designated Cadr-DRB1*01 through Cadr-DRB1*07. Alignment of the nucleotide and predicted amino acid sequences of these alleles are shown in Figs. 2 and 3, respectively. Allele-specific sites were found in three alleles (Cadr-DRB1*01, 04, and 06). All alleles differed by at least two amino acids (Cadr-DRB1*02-Cadr-DRB1*03, Cadr-DRB1*05-Cadr-DRB1*07) with a maximum of 15 differences observed between Cadr-DRB1*01 and Cadr-DRB1*02. The average number of nucleotide differences (k) was 13.24 and the mean nucleotide and amino acid distances were 5 and 8%, respectively (Table 2). A low mean ratio of transitions to transversions (Ti/Tv) over all sequence pairs was observed (0.616), close to the theoretical expected value of 0.5, and ranging from 0 (Cadr-DRB1*02-Cadr-DRB1*05) to 2.5 (Cadr-DRB1*02-Cadr-DRB1*03). Analysis of trio

Table 2 Measures of the *DRB1* exon 2 genetic diversity. *S*n, segregating nucleotide sites; *S*i, parsimony informative sites; k, average number of nucleotide differences; π, average nucleotide diversity; *H*d, haplotype diversity; *S*aa, segregating amino acid sites; *d*nt and *d*aa, overall mean nucleotide and amino acid genetic distances (±standard errors); and n, number of animals

Size (bp)	n	Sn (%)	Si (%)	k	π	Saa (%)	Hd ± SE	dnt ± SE	daa ± SE
282	132	33 (11.7)	17 (5.7)	13.24	0.04895	18 (19)	1.00 ± 0.076	0.05 ± 0.01	0.08 ± 0.02

genotypes showed Mendelian inheritance with sharing of identical alleles within each trio without evidence of recombination. The first family was homozygous for the second allele while the second trio was heterozygous displaying Cadr-DRB1*01–05, Cadr-DRB1*01–02, and Cadr-DRB1*01–02 genotypes for sire, dam, and offspring, respectively.

Of the 132 animals genotyped, only 38 were heterozygous showing ten different allele combinations (from 21 possible), and this deficit of heterozygotes (Ho = 0.29, He = 0.68) resulted in a significant deviation from HW equilibrium. The most frequent allele was Cadr-DRB1*01 found in 48 homozygous and 30 heterozygous animals, followed by Cadr-DRB1*02 present in 24 homozygous and 27 heterozygous individuals. Twenty animals (from 38) were heterozygous Cadr-DRB1*01/Cadr-DRB1*02. Apart from the Cadr-DRB1*07, all other alleles were observed in heterozygote combinations with either Cadr-DRB1*01 or Cadr-DRB1*02. These two highly divergent alleles, differing by 29 nucleotides and 15 amino acids, had a cumulative frequency of 76% (48 and 28 respectively). Frequency of other alleles varied from 9.1% (Cadr-DRB1*03) to 2.7% (Cadr-DRB1*06) with alleles Cadr-DRB1*04 and

Cadr-DRB1*05 having an intermediate frequency (5.7%). The less abundant allele in our panel was Cadr-DRB1*07, found in only one homozygous animal. However, this allele was detected in TSA (GenBank accession: GAEB01042196) as well as in raw RNAseq SRA (SRX10094694). Three other alleles detected here were found in the EST database (Cadr-DRB1*01, Cadr-DRB1*03, and Cadr-DRB1*05), two in RNAseq SRA (Cadr-DRB1*01, Cadr-DRB1*03), and one in TSA (Cadr-DRB1*02). Thus, only the Cadr-DRB1*04 allele present in this panel with 5.7% frequency and the Cadr-DRB1*06 (2.7%) were not found in the available dromedary transcriptome resources.

Both genomic sequences (LSZX01099695) from the ASM164081v1 assembly and the CL83Contig1 transcript initially detected in the ESTs displayed Cadr-DRB1*01 genotype. This allele was also observed in the *Camelus ferus* reference genome (GCF_009834535.1). In the Bactrian camel reference genome (assembly accession: GCF_000767855.1), the *DRB1* gene was split into two contigs with partial sequences of exon 2 (191 nucleotides) found at the end of the contig (JARL01022239). These sequences (which contain 23 of 33 variable sites) are identical to the
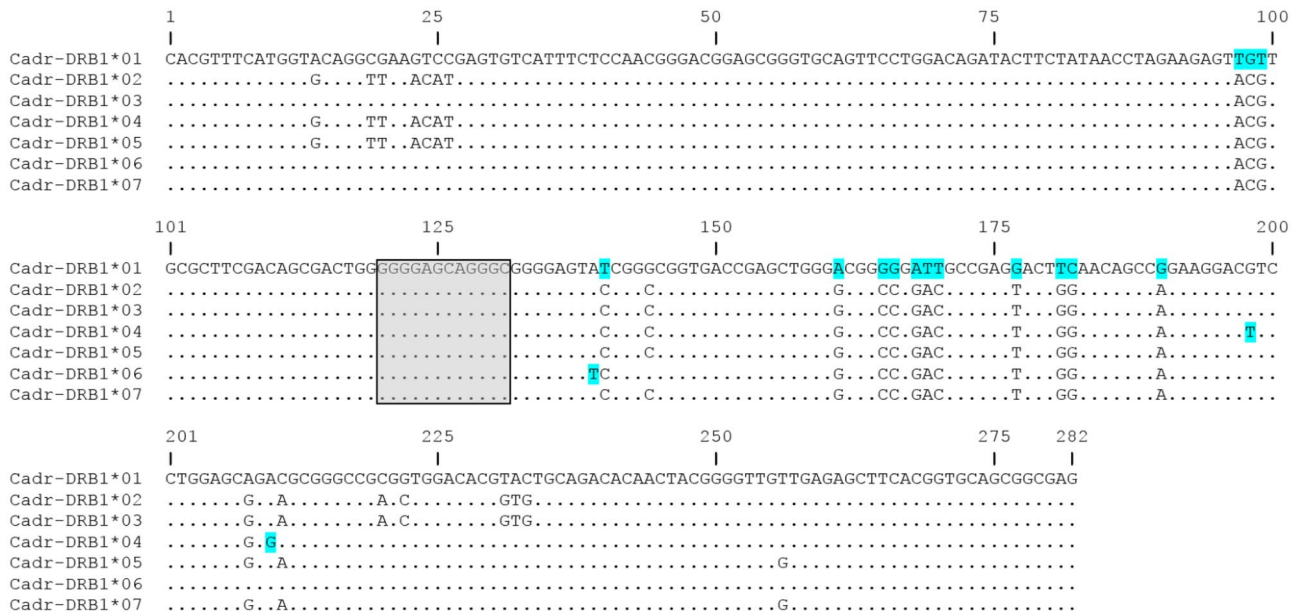

Fig. 2 Alignment of the nucleotide sequences of the dromedary camel MHC *DRB1* exon 2 alleles (282 bp). Allele-specific sites are highlighted in blue and the insertion of the 12 nucleotides at positions 120–131 (corresponding to the positions 220–231 in the coding region) is highlighted by a grey rectangle
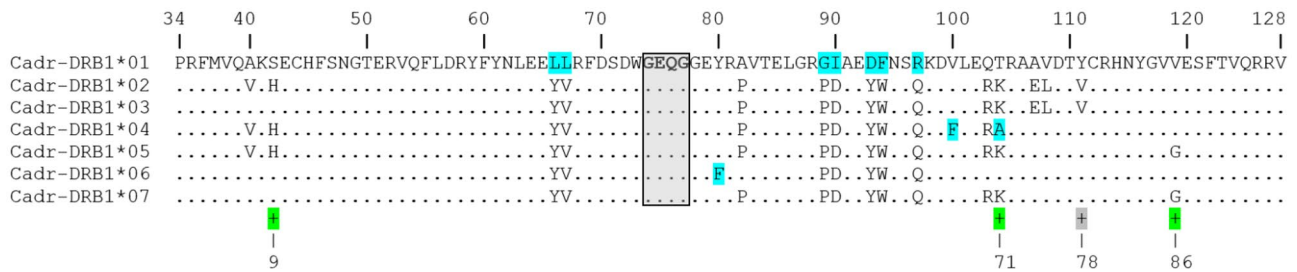
Fig. 3 Alignment of the deduced amino acid sequences of the dromedary camel MHC *DRB1* exon 2 alleles. The first (P) and last (V) residues are also encoded by exons 1 and 3, respectively. Allele-specific sites are highlighted in blue, and the insertion of the four residues at positions 74–77 is indicated in bold and highlighted by a grey rectangle. Numbering of amino acid residues is relative to the DRβ1 protein (270 aa) above the alignment and to the residues encoded by exon 2 below the alignment. Detected sites under positive selection are indicated by + symbol highlighted in green (detected by MEME method and Bayesian inference) and grey (Bayesian inference)

Cadr-DRB1*01 allele. In this species, four transcripts were detected in the TSA (bioproject PRJNA62509), none of them was identical to the alleles detected here. The seven *DRB1* exon 2 alleles have been deposited in GenBank under accession numbers OQ106971-77.

Five alleles (Cadr-DRB1*01–05) were characterized at the mRNA level by sequencing the full coding regions from seven animals. No samples were available for mRNA analysis of the two remaining alleles. Three animals, two homozygous and one heterozygous, displayed the Cadr-DRB1*01 allele. Outside the β1 domain encoded by exon 2, four variations were detected in exon 3, including three non-synonymous (p.V137A, p.N183S, and p.V197A; Cadr-DRB1*01 is the reference). No evidence for mutations was found neither in the remaining exons nor in the untranslated regions. The substitution of valine residue by alanine at position 197 resulted in a new allele since it was observed only in one animal harboring the reference allele Cadr-DRB1*01. Therefore, the reference allele showing valine at this position was renamed Cadr-DRB1*01:01 following MHC nomenclature, whereas alanine at this position in the same allele will corresponds to Cadr-DRB1*01:02. The other four alleles (Cadr-DRB1*02–05) showed A-S-V amino acids at positions 137–183-197, respectively. The synonymous mutation (c.537 T > C) occurred in the alleles Cadr-DRB1*02–05 and was not observed in Cadr-DRB1*01:02. The alignment of the full coding nucleotide and the deduced amino acid sequences of the six alleles are shown in Fig. S3. The sequences of these alleles have been deposited in GenBank under accession numbers OQ106963-68.

As outlined in methods, a subset of animals ($n = 79$) displaying all detected genotypes were also genotyped using different forward primer (Cd.DRB1.E2F) for partial exon 2 amplification (271 bp) (Fig. S2) and the results showed complete concordance of genotypes. This finding validated the genotyping procedure and ruled out the likelihood of null alleles. On the other hand, this method could also be used for *DRB1* genotyping in dromedary camel since the 11 missed nucleotides at the start of exon 2 which are included in the primer sequences turned out to be monomorphic in all genotyped animals.

Recombination and positive selection analysis

No significant evidence of recombination was detected in the *DRB1* exon 2 when using methods implemented in RDP5 software. Although two potential recombination events at sites 61–170 and 100–215 were found, the involved parental sequences were not clearly identified and they turned out to be non-significant for the seven used methods. GARD analysis revealed one recombination event with inferred breakpoints at positions 26 and 189, the same position 26 but in conjunction with position 211 was also detected by DnaSP6 which used the method of Hudson (1987). Even considering the occurrence of this recombination event as highly plausible, it is unlikely that it would have pronounced effects on codon-based analysis of positive selection given the low level of recombination (only one event) and the limited number of the analyzed sequences (Anisimova et al. 2003).

Overall, the mean dN/dS ratio estimated across all codon sites of the second exon was less than 1 ($\omega = 0.89$) indicating a lack of positive selection at the *DRB1* locus. However, when considering individual codons for dN/dS ratio estimation, signals of episodic diversifying selection were detected at three positions (9, 71, and 86) by MEME method (Fig. 3). The same codons were also detected by Bayesian inference, all showing a high estimate of ω value (3.2–3.6) albeit with varying posterior probabilities (0.76, 0.9, and 0.87 for the three codons, respectively). An additional codon at position 78 was detected by the same procedure as being under positive selection ($\omega = 3.84$) with strong evidence (posterior probability of 0.98). Three other codons at positions 33, 57, and 61 were detected as having an omega ratio > to 3 but with reduced posterior probabilities (0.72–0.77). Codon 9 falls within the peptide binding region (PBR) on the basis of the homology to the human DRβ1 molecule (Brown et al.

1993), whereas PBR inference could not be ascertained for codons 71, 78, and 86 since they were located downstream of the four amino acids insertion observed in the dromedary DRβ1 protein.

Phylogenetic analysis of the *DRB1* exon 2 alleles

The ML phylogenetic tree based on the entire *DRB1* exon 2 alleles with bovine and porcine alleles as outgroups is depicted in Fig. 4. The tree divided exon 2 sequences into two separate lineages represented by the two highly divergent Cadr-DRB1*01 (with Cadr-DRB1*03, 06 and 07) and Cadr-DRB1*02 (with Cadr-DRB1*04 and 05) alleles. Within each lineage, two alleles appeared to be closely related and were clustered together with high bootstrap support, Cadr-DRB1*01-Cadr-DRB1*06 (90%) and Cadr-DRB1*04- Cadr-DRB1*05 (87%). The remaining alleles formed each distinct sub-lineage but with weak bootstrap support (less than 70%).

Discussion

To date, few is known as regard the MHC II genes in the dromedary camel, e.g., number of expressed loci, gene structure and polymorphism, despite the importance of the species from the socio-economic point of view (as livestock animal thriving in several arid and semi-arid regions) and from the biological and evolutionary perspectives (genetic basis of adaptation to harsh environments, biological peculiarities). This study thus provided first insights on the expressed MHC class II *DRB* loci and their diversity in the Arabian camel *Camelus dromedarius*. Focusing essentially on the expressed loci rather than solely genomic DNA was motivated by the fact that several *DRB* genes (and pseudogenes) have been observed in several species with varying levels of expression and functionality. Furthermore, alleles detected using genomic DNA are not always observed at the cDNA

level (O'Connor and Westerdahl 2021) and thus not mirroring the real functional diversity.

In this regard, and as in several other species where multiple *DRB* loci were found, the dromedary camel genome contains at least two transcribed loci, designated *MhcCadr-DRB1* and *MhcCadr-DRB2*. These two loci share overall gene structure with six exons but exhibited different exon 2 sizes (282 vs. 270 bp) and transcript abundance patterns. *Cadr-DRB1* is thought to be the main *DRB* locus in the dromedary camel on the basis of transcript abundance and genetic diversity, whereas *Cadr-DRB2* locus seems of secondary relevance due to exon 2 skipping during transcription and to the observed low transcript abundance and polymorphism. The presence of multiple copies of the *DRB* locus with varying levels of expression and polymorphism has been observed in several mammalian species. For example, the main *DRB* gene in human, *HLA-DRB1*, is highly polymorphic and is expressed at much higher level than its paralogs *DRB3*, *DRB4*, and *DRB5* (Emery et al. 1993). Similarly, only *BoLA-DRB3* is functional in cattle; the second gene *BoLA-DRB2* is monomorphic and showing very low levels of expression (Burke et al. 1991). Multiple copies of MHC loci usually arise by gene duplication events following the "birth-and-death" model of evolution (Nei et al. 1997). In this context, and in addition to the two *DRB* genes characterized here, a DRB2-like locus containing a 270 bp exon 2 has been predicted (LOC105101437) in the NCBI annotation of the dromedary camel reference genome. The extent of expression and genetic diversity of this locus as well as its relevance to the dromedary immunity remains to be investigated. Even though no evidence of additional copies of the *DRB1* gene was found in the sequenced dromedary genomes, the presence of duplicated copies could not be ruled out because *DRB* copy number variation between individuals within the same species appears to be not uncommon in mammalian species, including in human (Arvidsson et al. 1995). Thus, and in the light of the limited number of the sequenced dromedary genomes

Fig. 4 Maximum likelihood phylogenetic tree of the seven dromedary MHC *DRB1* exon 2 alleles. The numbers on the nodes represent the bootstrap percentages that support each node after 2000 replicates
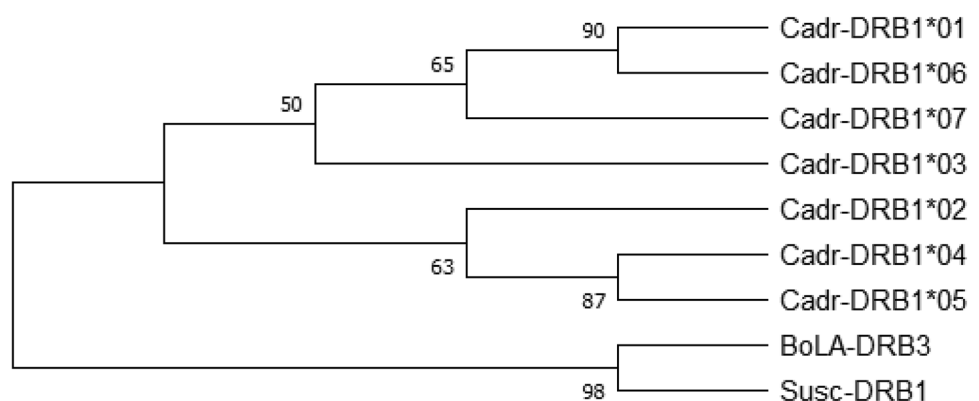
currently available, this possibility should be considered as it remains fairly plausible.

The dromedary *DRB1* gene contains an insertion of twelve nucleotides in exon 2 resulting in the introduction of four amino acids in the antigen presenting domain. This insertion extends the dromedary exon 2 to 282 bp (and the protein to 270 amino acids) and appears to be a specific feature of *Camelidae* family since it was not observed in other mammalian species DRBs (Fig. 5). This feature adds another peculiarity to the camelids' immune system characterized by the presence of the well-known HCAbs (heavy-chain antibodies) devoid of light chains (Hamers-Casterman et al. 1993). It is expected that the four amino acid extensions of the β1 domain in dromedary camel DRβ1 will have an impact on the composition of the PBR sites and the repertoire of bound peptides. Further investigations are needed to evaluate the extent of this impact and to study the characteristics of the dromedary PBR and peptide binding repertoires. Similar insertions of twelve nucleotides in the exon 2 (282 bp) of the dromedary *DQB* gene have been reported (Plasil et al. 2016) and this appears to be a common characteristic of the dromedary MHC *DR* and *DQ* genes.

A straightforward protocol for the dromedary *DRB1* exon 2 amplification was developed using primers in the flanking introns (and thus avoiding polymorphic exon sequences) allowing analysis of the whole exon 2. The reliability of this method was further validated with different forward primer designed at the intron–exon boundary (Fig. S2). High primer annealing temperatures (65–66 °C) ensured specific amplification and facilitated subsequent sequencing steps. Full

concordance between genotypes (from homozygous and heterozygous animals) obtained using these two methods was observed.

Screening for the *DRB1* exon 2 polymorphism revealed a contrasting pattern of the *DRB1* diversity in the dromedary camel. Although a substantial sequence variation at the *DRB1* exon 2 has been detected with 33 segregating nucleotides (11.7%) and 18 variable amino acids sites (19%), this variation has been "condensed" in a relatively reduced number of haplotypes (alleles), seven in total (and eight when considering exon 3). The reasons behind this observed pattern are not clear. While the number of alleles was relatively small, significant differences between these alleles were observed, with a difference of 2–15 amino acids in the protein sequences. Six from the seven detected exon 2 alleles were found at frequencies more than 1% and five at more than 5%. Low polymorphism levels have been previously reported in the dromedary and Bactrian camel MHC class II genes (Plasil et al. 2016); however, as far as the *DRB* is concerned, this low polymorphism was not surprisingly. Rather, it corroborates the results found in this work since the analyzed fragment was from a DRB2-like locus (exon 2 of 270 bp) characterized by low levels of polymorphism and transcript abundance as shown here for the *DRB2* gene. The *DRB1* alleles appear to be widespread among the dromedary populations since they were observed in animals from diverse geographic and — presumed — different ancestry origins: the ESTs and RNAseq SRA were derived from Middle-Eastern dromedaries, the TSA was derived from Indian animals, and the ASM164081v1 genome (from an Algerian dromedary, breed Targui) and the data from
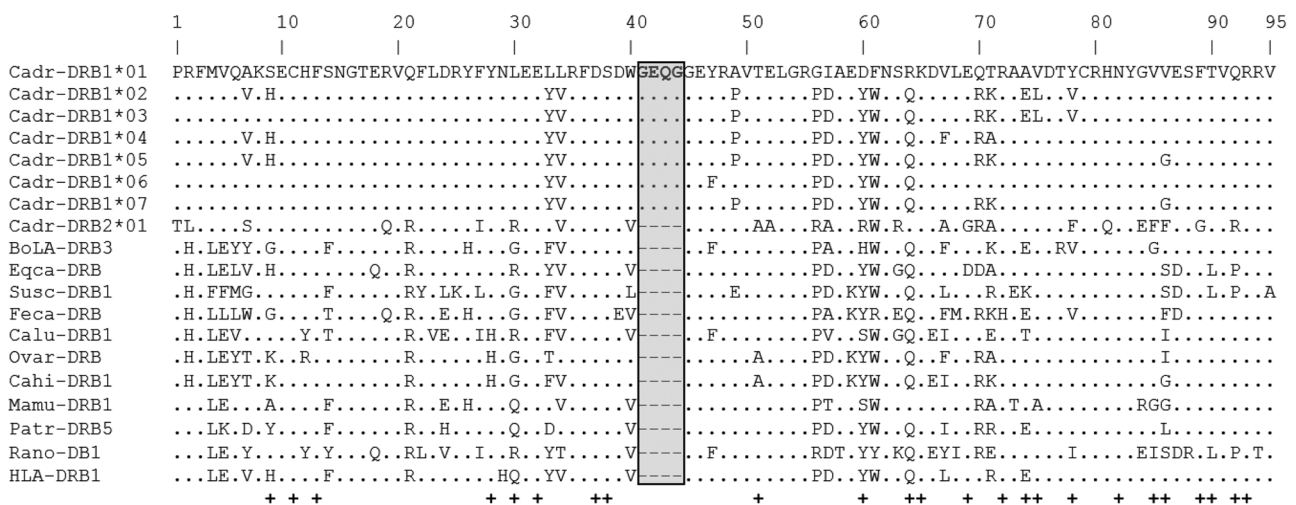

Fig. 5 Alignment of the amino acids encoded by the MHC *DRB* exon 2 in the dromedary camel and in other species. The four amino acids insertion in the dromedary DRβ1 is highlighted by the light grey box and indicated in bold. PBR codons in human DRβ1 chain are indicated by the+symbol. BoLA (*Bos taurus*, NP_001012698), Eqca (*Equus caballus*, NP_001136287), Susc (*Sus scrofa*, NP_001302687), Feca (*Felis catus*,

NP_001121544), Calu (*Canis lupus familiarus*, NP_001014768), Ovar (*Ovis aries*, NP_001116874), Cahi (*Capra hircus*, NP_001301146), Mamu (*Macaca mulatta*, NP_001305273), Patr (*Pan troglodytes*, NP_001266704), Rano (*Rattus norvegicus*, NP_001008884), and HLA (*Homo sapiens*, NP_001346123)

this study (Tunisian dromedary camels) were derived from North-African individuals.

The number of the *DRB* alleles varies considerably among mammalian species from monomorphism or few alleles (Mikko et al. 1999) to several hundred in human or cattle species as shown in the IPD-MHC database. Nevertheless, this high variability of allelic richness should be regarded cautiously considering the number of analyzed samples, population or breed coverage, and gene duplication. In this context, the dromedary camel allelic diversity could be considered within the range of *DRB* polymorphism of other mammalian species, albeit in the lower end. On the other hand, it is anticipated that the *DRB1* allele repertoire will certainly be extended by genotyping more diverse dromedary populations. In this regard, evidence of three additional alleles was found (an allele from one animal from our panel, and two alleles detected in the raw RNAseq SRA) although they remain to be validated. These alleles did not arise from new SNPs but were rather different haplotypes from the variants detected here.

Contrasting with the relatively limited allelic richness, an overall intermediate level of nucleotide diversity ($\pi = 0.049$) was observed at the *DRB1* exon 2. Comparatively, this level is slightly less than that observed in cattle for example, an intensively and widely studied species, where π ranged from 0.068 to 0.09 (Salim et al. 2021). In dromedary, much lower nucleotide diversity has been reported for both SNPs and indels (6.2×10^{-4}) and non-synonymous (1.72×10^{-4}) variants when considering all MHC genes (Lado et al. 2020) estimated from whole genome sequences. Obviously, these are overall estimates of nucleotide diversity trends that could not be compared to the π value shown here as it was obtained from only *DRB1* exon 2. Patterns of nucleotide diversity are not uniform among the MHC region due to the varying levels of selection pressure, and generally showing extreme but interrupted levels (Gaudieri et al. 2000).

The observed heterozygote deficiency (Ho = 0.29, He = 0.68) in the Tunisian dromedary despite the diversity of seven alleles raises the possibility of the presence of missed alleles during PCR amplification, i.e., null alleles. However, this scenario is highly unlikely since fully matched genotypes — in both homozygous and heterozygous animals — were observed using two genotyping protocols. Furthermore, total agreement of the cDNA and gDNA sequences and no Mendelian mismatches in trio analysis were observed. In addition, primer sequences used for amplification of either DRB1 cDNA or exon 2 were fully conserved in both domesticated Bactrian and wild camels. On the other hand, no noticeable genetic structure has been detected in the analyzed population using neutral microsatellites markers (Nouairia et al. 2015) that would explain this heterozygote deficit by the Wahlund effect. Rather, these results could be likely related to the breeding systems practiced (low-input systems with low reproduction rate) or to the population size

fluctuations (pronounced reduction during the second half of last century followed by steady recovery in last decades).

Otherwise, this heterozygote deficiency is indicative of a lack of ongoing balancing selection in the population on the basis of the heterozygote advantage assumption (or overdominant selection). The analysis of non-synonymous and synonymous substitutions revealed no evidence for positive selection acting on the *DRB1* exon 2 in the past when considering the entire sequences (PBR and non-PBR, dN/dS < 1). The distinction between PBR and non PBR codons was not possible due to the insertion of four amino acids in the dromedary protein compared to the human DRβ1. However, when codons were taken individually, four sites were inferred as to be under positive selection, including one codon (at position 9) congruent with predicted antigen-binding sites of the human HLA-DRβ1 (Fig. 5), the three remaining codons could not be inferred since they are located downstream of the insertion. These findings suggest a weak positive selection pressure acting on the dromedary *DRB1* gene that might be related to a reduced likelihood of exposure to pathogens in the harsh and arid environments where dromedary camel usually thrives (at least in Tunisia). In this context, it would be of interest to compare the genetic diversity of the *DRB1* in dromedary populations from regions with different environments and likelihood of pathogen load, for example, from the dry and hyper-arid areas (e.g., Saharan and desert zones) and from the hot and relatively most wetter regions (e.g., sub-Saharan and Sahel). Another possible explanation of the relaxed selection pressure may be related to the breeding systems currently practiced in this species. Compared to other livestock species, the dromedary camel is usually reared under an "extensive" management systems with low-input and low density herds that may contribute to a limited pathogen transmission between animals. Nevertheless, these observations and hypothesis have also to be considered in the light of the evolution and demographic history of the species that might contribute to the observed genetic diversity and selection patterns. Particularly, signals of pronounced and long-term population bottlenecks during Pleistocene and domestication process have been detected in dromedary camel (Fitak et al. 2020) resulting in an overall reduced genome-wide diversity and a relatively lack of substructure in the dromedary populations.

The common allele Cadr-DRB1*01 was the most frequent and distinct one compared to other alleles, differing by 8–15 amino acids, and showing a significant excess of homozygotes in the population. It is therefore possible that this genotype excess was related to pathogen-driven selection where homozygote genotype might confers more protection or survival against specific pathogens (Bernatchez and Landry 2003). However, this remains only speculative as many other confounding factors may be involved and has to be validated by empirical functional studies. On the other

hand, the fact that this allele was present in *Camelus ferus* — and probably in *Camelus bactrianus* — suggests trans-species polymorphism (TSP) across the genus *Camelus*. TSP occurs when identical or similar alleles are detected in related species and this phenomenon has been observed in several animal orders, including primates and rodents (Klein et al. 2007). Allele sharing between related species could result from convergent selection, from introgression, or as unaltered remnant from the speciation process. Although no data is currently available regarding the *DRB1* gene in the two other *Camelus* species, it is likely that the Cadr-DRB1*01 allele may predate the split of dromedary and two-humped wild camel from the common ancestor 4.4 million years ago (Wu et al. 2014) and is still persisting in the two species. Such persistence over a long evolutionary period of the *DRB1* polymorphisms implies that the two species have been — and are still — submitted to a relatively similar parasitic and pathogen challenges (at least those recognized by DRB1). Even though they are found in distinct geographic regions, both species share similar environment and occupy harsh and arid desert habitat characterized by scarcity of resources and extreme temperatures. In this context, analysis of the *DRB1* genetic diversity in the wild *Camelus ferus* and domesticated *Camelus bactrianus* camels would elucidate the extent of allele sharing and relationship within the *Camelus* genus as well as the involved underlying genetic basis (TSP, convergent selection, etc.). This would contribute to better understand the evolutionary history of these species. For this purpose, the genotyping procedures developed in this work could be directly applied in both species since the primers used for the *DRB1* exon 2 screening as well as for cDNA analysis are perfectly conserved in this genus. Particularly, such studies are of high relevance for the wild camel species since it is considered critically endangered on the IUCN (International Union for Conservation of Nature) red list with a census of few hundred individuals.

In summary, the present study lays the groundwork for future investigations into the *DRB1* diversity in the genus *Camelus* by providing the first comprehensive analysis of the expressed MHC class II *DRB* genes in the dromedary camel and by developing optimized protocols for genetic diversity screening. This makes it possible to consider in the future targeted functional studies in these species, for example, by analyzing the association of the *DRB1* genotypes with parasite infestation (i.e., ticks) and infectious diseases (i.e., MERS caused by MERS coronavirus). In this regard, some human *HLA-DRB1* alleles have been associated with the severe acute respiratory syndrome (COVID-19) caused by the SARS-CoV-2 virus (Augusto and Hollenbach 2022). Further investigations are needed to identify other DRB-like genes, analyze their genetic diversity and their expression levels (along with the *DRB1* and *DRB2* loci) using quantitative PCR, and evaluate their relevance to the dromedary camel adaptive immunity.

Declarations

Competing interests The author declares no competing interests.

References

Al-Swailem AM, Shehata MM, Abu-Duhier FM et al (2010) Sequencing, analysis, and annotation of expressed sequence tags for *Camelus dromedarius*. PLoS ONE 5(5):e10720 https://doi.org/10.1371/journal.pone.0010720

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/s0022-2836(05)80360-2

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236. https://doi.org/10.1093/genetics/164.3.1229

Arvidsson AK, Svensson AC, Widmark E, Andersson G, Rask L et al (1995) Characterization of three separated exons in the HLA class II DR region of the human major histocompatibility complex. Hum Immunol 42:254–264. https://doi.org/10.1016/0198-8859(94)00102-V

Augusto DG, Hollenbach JA (2022) HLA variation and antigen presentation in COVID-19 and SARS-CoV-2 infection. Curr Opin immunol 76:102178 https://doi.org/10.1016/j.coi.2022.102178

Avila F, Baily M, P, Perelman P, et al (2014) A comprehensive whole-genome integrated cytogenetic map for the alpaca (*Lama pacos*). Cytogenet Genome Res 144:196–207. https://doi.org/10.1159/000370329

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour 10:237–251. https://doi.org/10.1111/j.1755-0998.2009.02788.x

Behl JD, Verma NK, Tyagi N et al (2012) The major histocompatibility complex in bovines: a review. ISRN Vet Sci 2012:872710 https://doi.org/10.5402/2012/872710

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–377. https://doi.org/10.1046/j.1420-9101.2003.00531.x

Brown JH, Jardetzky TS, Gorga JC et al (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364:33–39. https://doi.org/10.1038/364033a0

Burke MG, Stone RT, Muggli-Cockett NE (1991) Nucleotide sequence and Northern analysis of a bovine major histocompatibility class II DR β-like cDNA. Anim Genet 22:343–352. https://doi.org/10.1111/j.1365-2052.1991.tb00688.x

Dawson P, Malik MR, Parvez F, Morse SS (2019) What have we learned about Middle East respiratory syndrome coronavirus emergence

in humans? A systematic literature review. Vector Borne Zoonotic Dis 19:174–192. https://doi.org/10.1089/vbz.2017.2191

Emery P, Mach B, Reith W (1993) The different level of expression of HLA-DRB1 and -DRB3 genes is controlled by conserved isotypic differences in promoter sequence. Hum Immunol 38:137–147. https://doi.org/10.1016/0198-8859(93)90531-5

Fitak RR, Mohandesan E, Corander J et al (2020) Genomic signatures of domestication in Old World camels. Commun Biol 3:316. https://doi.org/10.1038/s42003-020-1039-5

Gaudieri S, Dawkins RL, Habara K et al (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. Genome Res 10:1579–1586. https://doi.org/10.1101/gr.127200

Hamers-Casterman C, Atarhouch T, Muyldermans S et al (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446–448. https://doi.org/10.1038/363446a0

Horton R, Wilming L, Rand V et al (2004) Gene map of the extended human MHC. Nat Rev Genet 5:889–899. https://doi.org/10.1038/nrg1489

Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. Genet Res 50:245–250. https://doi.org/10.1017/S0016672300023776

Jukes T, Cantor C (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian Protein Metabolism. Academic Press, New York, pp 21–132

Khalafalla AI (2017) Emerging infectious diseases in camelids. In: Bayry J (ed) Emerging and re-emerging infectious diseases of livestock. Springer 425–441 https://doi.org/10.1007/978-3-319-47426-7_20

Klein J, Bontrop RE, Dawkins RL et al (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219. https://doi.org/10.1007/BF00204890

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet 41:281–304. https://doi.org/10.1146/annurev.genet.41.110306.130137

Kosakovsky Pond SL, Posada D et al (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901. https://doi.org/10.1093/molbev/msl051

Kumar S, Stecher G, Li, et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lado S, Elbers JP, Plasil M et al (2021) Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. Cells 10(6):1291. https://doi.org/10.3390/cells10061291

Lado S, Elbers JP, Rogers MF et al (2020) Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. BMC Genom 21:606. https://doi.org/10.1186/s12864-020-06990-4

Maccari G, Robinson J, Ballingall K et al (2017) IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. Nucleic Acids Res 45:D860–D864. https://doi.org/10.1093/nar/gkw1050

Martin DP; Varsani A, Roumagnac P et al (2021) RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol 7:veaa087 https://doi.org/10.1093/ve/veaa087

Mikko S, Røed K, Schmutz S, Andersson L (1999) Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. Immunol Rev 167:169–178. https://doi.org/10.1111/j.1600-065X.1999.tb01390.x

Miltiadou D, Law AS, Russell GC (2003) Establishment of a sequence-based typing system for BoLA-DRB3 exon 2. Tissue Antigens 62:55–65. https://doi.org/10.1034/j.1399-0039.2003.00080.x

Murrell B, Wertheim JO, Moola S et al (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764 https://doi.org/10.1371/journal.pgen.1002764

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799–7806. https://doi.org/10.1073/pnas.94.15.7799

Nouairia G, Kdidi S, Salah R et al (2015) Assessing genetic diversity of three Tunisian dromedary camel (*Camelus dromedarius*) sub populations using microsatellite markers. Emir J Food Agric 27:362–366. https://doi.org/10.9755/ejfa.v27i4.19258

O'Connor EA, Westerdahl H (2021) Trade-offs in expressed major histocompatibility complex diversity seen on a macroevolutionary scale among songbirds. Evolution 75:1061–1069. https://doi.org/10.1111/evo.14207

Plasil M, Mohandesan E, Fitak RR et al (2016) The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. BMC Genom 17:167. https://doi.org/10.1186/s12864-016-2500-1

Ronquist F, Teslenko M, Van Der Mark P et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resources 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Salim B, Takeshima S, Nakao R et al (2021) BoLA-DRB3 gene haplotypes show divergence in native Sudanese cattle from taurine and indicine breeds. Sci Rep 11:17202. https://doi.org/10.1038/s41598-021-96330-7

Sambrook J, Fritsch ER, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Sazmand A, Joachim A, Otranto D (2019) Zoonotic parasites of dromedary camels: so important, so ignored. Parasites Vectors 12:610. https://doi.org/10.1186/s13071-019-3863-3

Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102–1104. https://doi.org/10.2144/00286ir01

Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3-new capabilities and interfaces. Nuc Acids Res 2012 Aug 1;40(15):e115. https://doi.org/10.1093/nar/gks596

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL (2018) Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35:773–777. https://doi.org/10.1093/molbev/msx335

Wernery U, Kinne J (2012) Foot and mouth disease and similar virus infections in camelids: a review. Sci Tech Rev 31:907–918 https://doi.org/10.20506/rst.31.3.2160

Wu H, Guang X, Al-Fageeh M et al (2014) Camelid genomes reveal evolution and adaptation to desert environments. Nat Commun 5:5188. https://doi.org/10.1038/ncomms6188

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.