We identified and characterized 11 polymorphic microsatellite markers suitable for routine testing (three in the MHC class I sub-region, four in MHC class II and four in the MHC class III sub-region) of dromedaries and Bactrian camels. In total, 38 dromedaries and 33 Bactrian camels were genotyped, and interspecific differences were observed in the numbers of alleles and in allelic frequencies, as well as in the observed heterozygosity. These loci may be used as markers to study the adaptive genetic diversity of the MHC region in Old World camels.

**KEYWORDS**

Genetic diversity, camels, microsatellite markers, major histocompatibility complex, *Camelus dromedarius*, *Camelus bactrianus*
types of genes are highly polymorphic, and their polymorphism is of adaptive value (Axtner and Sommer 2007). The multiple roles of the MHC in immunity and disease in humans and various mammalian species have been repeatedly reviewed (Traherne 2008). At the population level, an extensive genetic diversity of MHC genes has been reported in many mammalian species (Yuhki and O’Brien 1990; Kamath and Getz 2012). Often, this diversity is associated with important diseases (Shiina et al., 2009).

The MHC and NKR genomic regions of cameldid species have been characterized only recently (Plasil et al., 2019a; Futas et al., 2019). The genomic structure of the cameldid MHC is similar to other mammalian species (Plasil et al., 2019a). However, the diversity of MHC genes seems to be lower than in most other species (Plasil et al., 2019b). Due to the existence of multiple highly similar loci, individual genotyping of MHC and NKR genes is a challenging task. Therefore, only limited information on the genetic diversity of MHC genes in various camel populations is available.

Genotyping microsatellite markers is an inexpensive and relatively easy method of assessing the genetic diversity of the MHC genomic regions in various animal species. MHC-linked microsatellite markers have been identified and some of them have been used for characterizing MHC polymorphism in various species, including domestic animals such as horses (Sommer 2005; Horecky et al., 2018; Jaworska et al., 2020). Microsatellite loci distributed throughout the genomes of dromedaries have been identified (Khalkhali-Evrigh et al., 2019). However, no detailed information on microsatellite loci mapping to MHC class I, II and III sub-regions is currently available. Recently, we have developed a panel of microsatellite markers for assessing the genetic diversity of NKR genomic regions in camels (Futas et al., 2019). Here, we present a set of MHC-linked microsatellites that can be used as markers of MHC population diversity and for association analyses.

The primary purpose of this work was to study the non-neutral, i.e., adaptive genetic diversity of camel populations. For this purpose, the MHC genomic region is a prime candidate, due to its crucial role in immune responses. The specific objective of this study was therefore to identify polymorphic microsatellite loci located within the MHC region in the two species of domestic camels, the dromedary (Camelus dromedarius) and the Bactrian camel (Camelus bactrianus).

Materials and methods

Camels

DNA samples of 38 dromedaries (Camelus dromedarius) and 33 Bactrian camels (Camelus bactrianus) were analyzed. The dromedary hair samples were obtained from North African and Arabian countries during previous projects (Austrian Science Foundation (FWF) P1084-B17 and P24706-B25), and DNA extractions followed a simple salting-out protocol (De Volo et al., 2008). The Bactrian samples were collected in Mongolia for the purposes of another project and shared by prof. D. Modry, Vetuni Brno, Czechia.

All samples were collected in agreement with all legal and ethical standards by licensed veterinarians for the purposes of other projects. No sample analyzed in this study was collected specifically for the purposes of this particular project.

Genomic resources

Genomic sequences of Camelus dromedarius from the CamDro2 whole genome assembly (https://www.ncbi.nlm.nih.gov/assembly/GCA_000803125.2/) NW_011591952.1 (3.29 Mb), NW_011591249.1 (0.33 Mb) and NW_011591121.1 (1.05 Mb) were screened to detect microsatellites and for primer design.

Microsatellite detection and characterization

The search for microsatellites was conducted using the online microsatellite repeat finder WebSat (http://www.wsmartins.net/websat/) within the sequences NW_011591121.1 (Wu et al., 2014), NW_011591952.1 (Wu et al., 2014) and NW_011591249.1 (Wu et al., 2014).

Specific primers for amplifying the selected 32 microsatellites were designed de novo using the OLIGO software v4 and verified in the Primer Express v2.0 software (Applied Biosystems, Inc, Foster City, CA, United States). Aiming to optimize multiplex PCR protocols, all primers were designed using the same parameters (Ta, Tm, number of nucleotides). The forward primers were labelled by fluorescent dyes (6-FAM, HEX, NED and PET; Table 1). In addition, information on primers amplifying another six loci was obtained from the Baker Institute for Animal Health, Cornell University, Ithaca, NY (D. Antczak, D. Miller). They were also located within the sequences NW_011591121.1 (Wu et al., 2014), NW_011591952.1 (Wu et al., 2014) and NW_011591249.1 (Wu et al., 2014).

Based on the selection of primer pairs, protocols allowing three multiplex amplifications and parallel detection of alleles were optimized. Each PCR was performed in a final volume of 10 µL. The PCR reaction mix consisted of 1.0 µL 10 × Taq Buffer (Top-Bio, Ltd, Prague, Czech Republic), 0.5 U CombiTaq DNA polymerase (Top-Bio, Ltd, Prague, Czech Republic), 200 µM each dNTP (Thermo Fisher Scientific Inc, Waltham, United States), 0.1 µL of each primer of concentration 10 µM, and 50 ng of genomic DNA. The volume of the reaction mix was adjusted to 10.0 µL with PCR H2O (Top-Bio, Ltd, Prague, Czech Republic). Fragments were amplified in three separate PCR mixes: A, B and C. Mix A contained primers for markers CAM001, CAM03 and CAM08. Mix B contained primers for...
amplification of microsatellites \textit{CAM04}, \textit{CAM05}, \textit{CAM09} and \textit{CAM10}. Markers \textit{CAM02}, \textit{CAM06}, \textit{CAM07} and \textit{CAM11} were amplified by mix C (Table 1).

All PCRs were performed in the thermal cycler ABI Verity 96 Well (Applied Biosystems Inc, Foster City, CA, United States) with the following protocol: initial denaturation at 95°C for 3 min; 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C (mix A) and 58°C (mix B,C) for 30 s and elongation at 72°C for 30 s; final elongation at 72°C for 60 min and holding at 7°C.

All three sets of markers were then tested by fluorescent fragment analysis using an ABI PRISM 3500 Genetic Analyzer and sized with LIZ500 dye size standard (Life Technologies, Corp, Carlsbad, United States). The data obtained were analyzed using the GeneMapper software v4.1 (Life Technologies, Corp, Carlsbad, United States).

The dromedary primers were then used for PCR amplifications of Bactrian camel DNA samples.

**Validation of the microsatellite markers**

The specificity of all selected markers (including flanking regions) was verified based on at least two independent PCRs using Sanger sequencing. The Mendelian inheritance of all markers was assessed in seven trios (father, mother, offspring) of \textit{Camelus bactrianus} (Table 2).

**Characterization of the markers on a panel of camels**

The microsatellite Toolkit for Excel (Park 2001) was used to calculate the major characteristics of molecular diversity for all selected markers genotyped in the panel of camels. Haplotype analysis was performed using Arlequin ver 3.5 (Excoffier and Lischer, 2010).

**Results**

One hundred and twenty-two dinucleotide microsatellites were detected in all three MHC subregions (58 within the sequence NW\_011591121.1, 31 in NW\_011591952.1, and 33 in NW\_011591249.1). Based on our previous experience, dinucleotide microsatellites with CA/TG motifs with perfect repetitions and at least seven motif repeats were preferred.

### TABLE 1 Panel of markers selected for assessing the MHC diversity in camels.

| Region | Marker name | Location in CamDro2 | Primer sequence 5’—3’ | Fluorescent dye (5’) | PCR mix | Locus database Id | Repeated sequence motif |
|--------|-------------|---------------------|-----------------------|---------------------|---------|------------------|-------------------------|
| MHC I  | CAM01 CAMf_I_01 | 317 417 | F: GCA GCA GAG ATA TGC ACA GAA T R: CTG GCT TAC TTC TCT AGT GAA T 6-FAM A | NW\_011591211.1 (TG) | A | | |
| MHC I  | CAM02 CAMf_I_11 | 595 103 | F: ACT GGG TCA GAA ACA CCA AG R: CTG TGA GGG AGC TGT CTT GC 6-FAM C | NW\_011591211.1 (AC) | C | | |
| MHC I  | CAM03 CAMf_I_45 | 333 420 | F: AAC CCT TCC TTC TGA GTG CTG GC R: GGG AAT GAG AGA GAC ACA GGG AAA | NED A | A | NW\_011591211.1 (TC) | |
| MHC II | CAM04 CAMf_II_35 | 3 024 151 | F: GCT GCT CTC TTT CCT TCT GCT CTA C R: GGT TAA GTG GGT GAG GCT ACA G 6-FAM B | NW\_011591952.1 (AC) | B | | |
| MHC II | CAM05 CAMf_II_47 | 3 054 949 | F: CCA GCT GCC CTT TTA GTC GTG AAC A | R: CCC ACC CCG CAA TGT TAA AAT HEX B | B | NW\_011591952.1 (AC) | |
| MHC II | CAM06 1592_II_07 | 3 219 385 | F: TTG TTC CAG TGA TGG GTG TGG TGA TG R: ATG CTG AGG AGA GAC CAG AG | NED C | C | NW\_011591952.1 (GT) | |
| MHC II | CAM07 1592_II_08 | 3 257 334 | F: TTT CAA AAT CCT GCC AAG AC R: TTG TAC CCT GAG GTC TGC TC | NED C | C | NW\_011591952.1 (TG) | |
| MHC III | CAM08 CAMf_III_26 | 2 742 | F: AGA AGC AGT GGT GTG AGA TCC T R: TCA ATT TGA ACC ACA TGG ATG T | PET A | A | NW\_011591249.1 (GT) | |
| MHC III | CAM09 CAMf_III_41 | 190 586 | F: TTC AGA CAA CGT CAG CAT CTC T R: AGC CAG TGA GTG TAG GGA GTG T | NED B | B | NW\_011591249.1 (AC) | |
| MHC III | CAM10 CAMf_III_42 | 232 193 | F: CAG TCC CCC GCT GTG TAA TCT C R: CAG TGT CTA ACG TGG GTT TGT A | NED B | B | NW\_011591249.1 (GT) | |
| MHC III | CAM11 1121_III_02 | 140 350 | F: TGG CTT TAC ACA CAC ACA CG R: CCT GAA GTT TCA AGA TGG CG | NED C | C | NW\_011591249.1 (TG) | |
Thirty-eight microsatellites (including six microsatellites from the Baker Institute for Animal Health) complied with the previously defined criteria. These microsatellites were tested in our set of animals. Even in the initial selection of microsatellites, no markers polymorphic in only one species were found. Some microsatellites were non-polymorphic (in both species) and were immediately removed from further analysis. Only 11 microsatellites were polymorphic and suitable for further characterization and routine testing (three in the MHC class I sub-region, four in MHC class II and four in the MHC class III sub-region).

The physical location of these 11 markers within the MHC region of dromedaries is shown in Figure 1. Their molecular characteristics are summarized in Table 1. A comparison between dromedaries and Bactrian camels is available in Table 3 and the mean number of alleles per loci and observed heterozygosity by species and classes is presented in Supplementary Table S3.

**TABLE 2 Example of the Mendelian inheritance of all markers in Camelus bactrianus in two trios.**

|       | CAM01 | CAM02 | CAM03 | CAM04 | CAM05 | CAM06 | CAM07 | CAM08 | CAM09 | CAM10 | CAM11 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| father1 | 370   | 374   | 302   | 304   | 194   | 196   | 340   | 340   | 284   | 294   | 147   | 149   | 296   | 302   | 374   | 374   | 288   | 290   | 376   | 360   | 250   | 258   |
| mother1 | 366   | 382   | 304   | 306   | 194   | 196   | 340   | 340   | 274   | 284   | 143   | 149   | 296   | 296   | 374   | 374   | 288   | 292   | 376   | 360   | 256   | 252   |
| offspring | 370   | 382   | 304   | 304   | 194   | 196   | 340   | 340   | 284   | 294   | 143   | 149   | 296   | 296   | 374   | 374   | 288   | 292   | 376   | 360   | 256   | 252   |

|       | CAM01 | CAM02 | CAM03 | CAM04 | CAM05 | CAM06 | CAM07 | CAM08 | CAM09 | CAM10 | CAM11 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| father2 | 366   | 376   | 306   | 320   | 194   | 196   | 340   | 346   | 274   | 288   | 145   | 149   | 296   | 302   | 374   | 374   | 288   | 304   | 376   | 360   | 250   | 246   |
| mother2 | 364   | 372   | 314   | 324   | 194   | 196   | 346   | 346   | 284   | 284   | 141   | 147   | 296   | 302   | 374   | 374   | 288   | 288   | 376   | 360   | 242   | 246   |
| offspring | 364   | 366   | 306   | 314   | 196   | 196   | 340   | 346   | 274   | 284   | 147   | 149   | 296   | 302   | 374   | 374   | 288   | 288   | 376   | 376   | 242   | 246   |

FIGURE 1
Relative positions of MHC-linked microsatellite markers (in bp) dromedary camel genome (A) Class I - in the sequence NW_011591121.1, https://www.ncbi.nlm.nih.gov/nuccore/NW_011591121.1?report=graph. (B) Class II - in the sequence NW_011591952.1, https://www.ncbi.nlm.nih.gov/nuccore/NW_011591952.1?report=graph (C) Class III - in the sequence NW_011591249.1, https://www.ncbi.nlm.nih.gov/nuccore/NW_011591249.1?report=graph.
Allele frequencies in *Camelus bactrianus* and *C. dromedarius* are shown in Supplementary Table S1. Out of the total number of alleles found in our two groups (79), 63 alleles were observed in the Bactrian camel and 59 in the dromedary. Fifteen alleles unique to the Bactrian camel and 16 unique to the dromedary were observed.

Based on the assumption of independent segregation, it is possible to obtain altogether 9480 haplotypes in the Bactrian camel. Out of them, we were able to identify 27 haplotypes composed of all loci analyzed; three of them occurred with frequencies higher than 0.05: 0.1875, 0.1250 a 0.0833, respectively. For the dromedary, 45 haplotypes out of 8486 were found. No haplotype reached the 0.05 frequency. A complete dataset is in Supplementary Table S2.

**Discussion**

Microsatellite markers have been widely used for characterizing the genetic diversity of various human and animal populations (Field and Wills 1996). As unexpressed loci, they are considered to be mostly markers of neutral diversity of the genome (Holderegger et al., 2006). However, microsatellite loci located within complex immunity-related regions may serve for assessing the genetic diversity of these regions, which is assumed to be of adaptive value. As such, especially in combination with analyses of the genetic diversity of expressed loci, they may provide valuable information on the region of interest and may also be used for disease-association studies. For these purposes, we previously characterized the genomic organization of the NKC and LRC regions of camelids, where both expressed loci and microsatellites were identified (Futas et al., 2019).

Identification of polymorphic microsatellite markers in the MHC of the two camel species posed some methodological challenges. Currently, two whole genome assemblies of the dromedary are available, CamDro2 and CamDro3. In this study, CamDro2 was used as the primary resource and all markers were identified (Futas et al., 2019).

Another challenge was the existence of duplicated regions and consequently of duplicated loci reflected by repeated findings of three alleles of some markers, which eventually were eliminated from the definitive panel. Nevertheless, three, four and four markers of the class I, II, and III subregions, respectively, have the potential to be informative for various types of diversity, association and/or evolutionary studies.

122 dinucleotide microsatellites were found in the studied MHC regions. However, the final panel we compiled contains only 11 markers. Based on our previous experience, we eliminated *a priori* mononucleotide microsatellites (uncertain...
identification of individual alleles), three- or more-nucleotide repetitions (low polymorphism) and irregular microsatellites (not suitable as stable markers). Moreover, other common reasons for discarding a given microsatellite were monomorphism and technical problems during amplification (non-specific fragments, low amplification efficiency, etc.), since the aim of the study was to identify potentially good quality markers.

The marker diversity characteristics summarized in Table 3; Supplementary Table S1 were primarily calculated to characterize the marker sets rather than the two species. Although the two cohorts genotyped were composed of randomly selected representatives of their species due to the relatively small numbers of individuals and limited information on their relatedness that could be retrieved from the owners. However, clear interspecific differences in the numbers of alleles and observed heterozygosities (Table 3), as well as in allelic frequencies (Supplementary Table S1), were observed. The lowest observed heterozygosity and at the same time the PIC value were found for the CAM01 and CAM10 loci in the Bactrian camel, while in dromedaries, it was in loci CAM04 and CAM02. We are not sure to what extent this difference is due to real interspecific differences and how much it may be affected by the low number of camels analyzed. Taking into account this limitation, “species-specific” alleles were identified in all loci except CAM06 and CAM08. No differences in diversity were found between markers located in different MHC sub-regions; their characteristics seem to be primarily determined by the particular locus regardless of its linkage to expressed genes in the respective sub-region (see Supplementary Table S3).

The in-silico haplotype analysis revealed high numbers of inferred haplotypes probably reflecting the relatively long distances between the loci analyzed and weak linkage among them (Supplementary Table S2). This assumption is supported by the lower numbers and higher frequencies of haplotypes specific for the three MHC sub-regions. However, no clearly prevailing haplotypes were observed, and at the species level it was not possible to assign potentially species-characteristic haplotypes.

Conclusions

Eleven polymorphic microsatellite loci were identified in the three MHC subregions I, II and III of dromedaries and Bactrian camels. The major characteristics of these markers and their polymorphism were assessed. Interspecific differences were observed in the numbers of alleles and in allelic frequencies, as well as in the observed heterozygosity.

These loci may be used as markers to study the adaptive genetic diversity of the MHC region in Old World camels. Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the animal study because Ethical review and approval were waived for this study because all peripheral blood samples used for DNA extraction were originally collected for previous research projects (GACR 523/09/1972; PI: P. Horin; FWF P1084-B17 and P24706-B25; PI: P. Burger) during standard diagnostic and/or therapeutic procedures. The work was conducted in compliance with all national and international ethical and legal standards for animal welfare by licensed veterinarians, based on owners’ consent. Therefore, in agreement with its rules, no special approval of the Committee for Animal Welfare and Ethics of the University of Veterinary Sciences Brno was needed for working on archived and re-used samples. Written informed consent for participation was not obtained from the owners because archived samples were used. The original sampling was performed during standard diagnostic and/or therapeutic procedures with the oral consent of the owners, in agreement with all legal and ethical rules valid at the time of sampling.

Author contributions

AK performed the annotation of the camel MHC region, designed the microsatellite screening, analyzed molecular and population data, and participated to the project design. JW designed primers, carried out the PCR laboratory analysis, analyzed genome sequences. MP participated to the MHC region annotation and analyzed genome sequences. PH and PB designed the project, AK and PH drafted the manuscript. All authors read, commented on, edited, and approved the final version of the manuscript.

Funding

This study was supported by the Austrian Science Foundation (FWF) project P29623-B25, by the Central European Institute of Technology (CEITEC) CZ.1.05/1.1.00/02.0068 and the Czech National Sustainability Programme NPU LQ1601, which provided equipment and contributed to personal costs, respectively, and by the Internal Grant Agency of the Faculty of Agronomy, Mendel University in Brno (IGA FA MENDELU No. IP 057/2017).
Acknowledgments

The authors acknowledge and wish to thank Don Miller and Doug Antczak (Cornell University, Ithaca, NY) for sharing information on microsatellite primers and for the permission to use it in this work. The authors wish to thank MVDr. Apríl Jelinek for her careful English edit.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Ali, A., Baby, B., and Vijayan, R. (2019). From desert to medicine: A review of camel genomics and therapeutic products. Front. Genet. 10, 17. doi:10.3389/fgene.2019.00017

Astrner, J., and Sommer, S. (2007). Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, Clethrionomys glareolus. Immunogenetics 59 (5), 417–426. doi:10.1007/s00251-007-0205-y

Burger, P. A. (2016). The history of Old World camels in the light of molecular genetics. Trop. Anim. Health Prod. 48 (5), 905–913. doi:10.1007/s11250-016-1032-7

Camillo-Bustamante, P., Keşmir, C., and de Boer, R. J. (2016). The evolution of natural killer cell receptors. Immunogenetics 68 (1), 3–18. doi:10.1007/s00251-015-0869-7

Ciccarese, S., Burger, P. A., Carin, E., Castelli, V., Linguiti, G., Piasil, M., et al. (2019). The camel adaptive immune receptor repertoire as a singular Example of structural and functional genomics. Front. Genet. 10, 997. doi:10.3389/fgene.2019.00997

De Volo, S. B., Reynolds, R. T., Douglas, M. R., and Antolin, M. F. (2008). An improved extraction method to increase DNA yield from molted feathers. Condor 110 (4), 762–766. doi:10.1525/condor.2008.8586

Escotfié, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. Mol. Ecol. Resour. 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Elbers, J. P., Rogers, M. F., Perelman, P. L., Proskuryakova, A. A., Serdyukova, N. A., Johnson, W. E., et al. (2019). Improving illumina assemblies with Hi-C and long reads: An example with the North African dromedary. Mol. Ecol. Resour. 19, 1015–1026. doi:10.1111/1755-0998.13141

Field, D., and Wills, C. (1996). Long, polymorphic microsatellites in simple organisms. Proc. Biol. Sci. 263 (1367), 209–215. doi:10.1098/rspb.1996.0033

Futas, J., Oppell, J., Jelinek, A., Elbers, J. P., Wijkmark, S., Horin, P., et al. (2019). Natural killer cell receptor genes in camels. Another mammalian model. Front. Genet. 10, 620. doi:10.3389/fgene.2019.00620

Holdeggereg, R., Kamn, U., and Gugler, F. (2006). Adaptive vs. neutral genetic diversity: Implications for landscape genetics. Landsc. Ecol. 21, 797–807. doi:10.1007/s10980-005-5245-9

Horecky, C., Horecka, E., Futas, J., Janova, E., Horin, P., and Knoll, A. (2018). The major histocompatibility complex of Old World camels—A synopsis. Genet. Sel. Evol. 50 (1), 39. doi:10.1007/s10704-018-0370-3

Jelinek, A., Johnson, W. E., et al. (2019). Improving illumina assemblies with Hi-C and long reads: An example with the North African dromedary. Mol. Ecol. Resour. 19, 1015–1026. doi:10.1111/1755-0998.13141

Khalilkhani-Evrigh, R., Hedaya-Evrigh, N., Hafezian, S. H., Farhadi, A., and Bakhtiarizadeh, M. R. (2019). Genome-wide identification of microsatellites and transposable elements in the dromedary camel genome using whole-genome sequencing data. Front. Genet. 10, 692. doi:10.3389/fgene.2019.00692

Kado, S., Elbers, J. P., Piasil, M., Loney, T., Weidinger, P., Camp, J. V., et al. (2021). Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. Cells 10 (12), 691. doi:10.3390/cells10012691

Kado, S., Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamasuren, A., Corander, J., 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 Genomics 21, 606. doi:10.1186/s12864-020-05990-4

Kado, S., Futas, J., Piasil, M., Loney, T., Weidinger, P., Camp, J. V., et al. (2022). Crimean-Congo hemorrhagic fever virus past infections are associated with two innate immune response candidate genes in dromedaries. Cells 11 (1), 8. doi:10.3390/cells11010008

Knoll et al. 10.3389/fgene.2022.1015288

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1015288/full#supplementary-material