Turkish Journal of Veterinary and Animal Sciences

http://journals.tubitak.gov.tr/veterinary/

Research Article

Turk J Vet Anim Sci
(2015) 39: 218-228
© TÜBİTAK
doi:10.3906/vet-1402-72

Identification of simple sequence repeat markers in the dromedary (*Camelus dromedarius*) genome by next-generation sequencing

Monther SADDER[1,2,*], Hussein MIGDADI[1], Ahmed AL-HAIDARY[3], Aly OKAB[3,4]

[1]Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia
[2]Department of Plant Production, Faculty of Agriculture, University of Jordan, Amman, Jordan
[3]Department of Animal Production, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia
[4]Department of Environmental Studies, Institute of Graduate Studies and Research, Alexandria University, Alexandria, Egypt

Received: 19.02.2014 • **Accepted:** 04.07.2014 • **Published Online:** 01.04.2015 • **Printed:** 30.04.2015

Abstract: The availability of molecular markers in camels is limited. The aim of this study was to develop new simple sequence repeat (SSR) markers. Four breeds of pooled dromedary genome were sequenced at low coverage utilizing Roche and Illumina platforms. A total of 65,746 contigs, covering approximately 52 Mb (2316 contigs > 2 kb), were assembled. The partial genome revealed 613 SSR loci with a minimum number of 5 repeat units. Comparative chromosomal location for 60 camel loci was predicted against bovine genome assembly Baylor Btau_4.6.1/bosTau7. Ten markers (16.7%) returned matches with a >100 score and >80% identity. SSR abundance was 1 in every 84.3 kb of contigs. The SSR loci mainly comprised di- (80.8%), tri- (10.8%), tetra- (7.6%), and pentamer (0.8%) motifs. (TA)n and (AC)n were the most abundant (58.6%) dimers. Thirty SSR loci were experimentally characterized for both dromedary (16 animals) and Bactrian camels. The number of alleles ranged from 1 to 3, and the average number of fragments scored per animal ranged from 0.81 to 2. Polymorphic information content ranged from 0 to 0.66 with a mean value of 0.38. These SSR markers will be a valuable resource for further genetic studies of camels and related species.

Key words: Camel, dromedary, genome, microsatellites, next-generation sequencing

1. Introduction

The family Camelidae comprises 4 domesticated species belonging to 3 genera (1). These species are the Bactrian (*Camelus bactrianus*), the dromedary (*Camelus dromedarius*), the llama (*Lama glama*), and the alpaca (*Vicugna pacos*). In desert countries, camels provide resources that are integral for society such as milk, meat, and other products. Camels are heat stress-resistant animals (2), possessing the ability to apply remarkable adaptive thermoregulatory mechanisms to survive in arid and semiarid environments. Acquiring thermotolerance is a worldwide goal for animal producers (3,4).

An evaluation of genetic diversity based on morphological traits does not usually provide accurate estimates of genetic differences, as they are highly influenced by environmental factors. Several molecular markers have been developed and utilized in genotyping, breeding, and conservation of animals (5). Among the large variety of marker systems available, microsatellites or simple sequence repeats (SSRs) are the most abundant codominant and multiallelic markers (6,7). They are invaluable genetic tools for animal breeding and quantitative trait locus (QTL) analysis (7,8). The SSR marker system has been widely used for camel genetic diversity (9–15).

Several studies developed SSR markers for different camelids, and each publication reported from 8 to 23 new loci (16–18). However, they were limited in number and not adequate for genetic mapping or QTL analysis. This is because the development of SSR markers is labor-intensive and requires library construction and screening (17). Most recently, high throughput of next-generation sequencing (NGS) enabled the development of genome-wide SSR markers such as alpaca transcriptome (19) and bovine genome (20). The goal of the present study was to identify SSR markers from the dromedary (*Camelus dromedarius*) genome and investigate their polymorphic nature for genetic applications by using camel breeds bred in Saudi Arabia.

2. Materials and methods

2.1. NGS and sequence analysis

Whole-genomic DNA was isolated from 4 female Arabian camels (dromedary) using the Wizard Genomic Kit

* Correspondence: msadder@ksu.edu.sa

(Promega, USA). DNA samples were pooled and used for NGS utilizing 2 sequencing platforms. The first run required the generation of a sequencing library followed by emulsion PCR. The data were generated from a half-plate 454 pyrosequencing reaction using a GS FLX titanium platform (Roche, USA). The second run was performed utilizing the Genome Analyzer (Illumina, USA). The data were generated from 1 lane with 101 paired-end cycles with a gap of approximately 450 bp. Combined reads were assembled in SeqMan NGen (DNAstar, USA). SSRs were retrieved from assembled contigs using the Simple Sequence Repeat Identification Tool (SSRIT) (21) as a web interface. There was no sequence masking for any repetitive element or those with a minimum number of 5 repeat units. A total of 60 SSRs representing di-, tri-, tetra-, and pentamers were randomly selected, and their original contig sequences were retrieved from the assembly. Forward and reverse primers flanking each SSR locus were designed in Vector NTI (Invitrogen, USA). The marker sequences were compared to the bovine whole genome sequence (Baylor Btau_4.6.1/bosTau7) to identify potentially homologous sequences utilizing BLAT genome search. Default search parameters were used for this comparison (https://genome.ucsc.edu/cgi-bin/hgBlat).

2.2. SSR characterization and data analysis

A total of 16 Saudi camels (*C. dromedarius*), representing 4 breeds (ZU: Zurg, MJ: Majaheem, MG: Maghateer, SO: Sofr), were investigated to assess the applicability of the developed SSR markers. In addition, SSR markers were screened for one Bactrian camel (*C. bactrianus*). DNA was isolated using Wizard Genomic DNA purification kit (Promega, USA) from blood samples (dromedary) or hair samples (Bactrian). DNA samples were resuspended in TE buffer overnight at 4 °C and stored at –20 °C. The quality and quantity of genomic DNA were determined with a NanoDrop spectrophotometer.

Isolated DNA samples were first assessed for PCR by amplifying a repetitive sequence, which partially covered the 12S ribosomal gene developed in this study from the GenBank database using forward (5'-ACTCAAAGGACTTGGCGGTGC-3') and reverse (5'-GTGTGCGTGCTCCATGGC-3') primers. If the 12S is successfully amplified, then the DNA sample is ready for SSR analysis; otherwise, it may contain PCR inhibitors that preclude SSR amplification. PCR amplifications (both for 12S and SSR markers) were performed in 20-µL reactions containing 20 ng of genomic DNA template (pooled from all 16 animals), 1X GoTaq Green Master Mix (Promega, USA), 0.1 µM each forward and reverse primer, and nuclease-free water. Thermal cycling profile consisted of an initial denaturation at 94 °C for 5 min, followed by 35 cycles (94 °C for 45 s, 50 °C for 45 s, and 72 °C for 1 min) and a final extension at 72 °C for 20 min. PCR products were separated in 3% MetaPhore agarose (Lonza, USA) in 0.5X TBE buffer. HyperLadder IV (Bioline, UK) was used as the DNA marker. Gels were run under 60 V for 2 h. DNA was visualized with acridine orange (Sigma, USA) under UV light.

The expected heterozygosity (H_e) was calculated according to the Nei equation (22), and the observed heterozygosity (H_o) was calculated by dividing the number of heterozygotes at the locus by the number of individuals typed. Polymorphic information content (PIC) values were calculated for each SSR to estimate its allelic variation according to the formula described by Anderson et al. (23).

3. Results

The NGS with 454 GS FLX System yielded more than 700,000 reads with an average length of 375 bp, while the NGS with Genome Analyzer platform yielded more than 30×10^6 paired reads with approximately 100 bp. The reads were trimmed, and a draft dromedary genome was assembled into 65,746 contigs (2316 contigs longer than 2 kb) with N50 of 973 bp and an average of 786 bp, where N50 is the length of the longest contig of the lower half of all contigs (with a descending order from the longest to the shortest contig).

In total, 613 SSR loci with perfect repeats were detected in the assembly (Table 1). Singletons were not used to extract SSR motifs. The search was limited to motifs with 5 or more repeats. All 4 possible combinations of dimer motif groupings were found in 495 loci, of which 156 were AT/TA motifs. The trimer, tetramer, and pentamer combinations were detected in 66, 47, and 5 loci, respectively.

One-tenth of detected loci were randomly selected to be tested for SSR characterization utilizing local camel breeds. The designed PCR primers are listed in Table 2. The repeat number ranged from 5 to 22. These loci were numbered consecutively (*Cd00801* to *Cd00860*), and their sequences were deposited in GenBank (http://www.ncbi.nlm.nih.gov) with sequential accession numbers (JX093499–JX092558).

Comparative chromosomal location for the selected camel markers was predicted in the bovine genome by BLAT searches against bovine genome. All sequences returned a BLAT match (Table 3). Some markers returned multiple matches; however, 16.7% (10 markers) returned BLAT matches with >100 score and >80% identity. Putative camel homologs were found on each chromosome of bovine genome, except for BTA 25, 28, and Y. One camel SSR locus was placed on BTA 6, 7, 8, 12, 16, 19, 24, and 26, while BTA 11 and 14 reached 5 SSRs each with an average of 2 loci per chromosome. Conversely, 3 markers showed matches to unassigned contigs (UN).

The selected SSR primers were evaluated for their ability to prime PCR amplification of one pooled DNA

Table 1. SSR repeats detected in dromedary camel genome.

Repeat motif grouping	Times repeated	Occurrence
Dimers		
AC/CA/TG/GT	5–149	83
AG/GA/CT/TC	5–61	248
AT/TA	5–19	156
GC/CG	5–9	8
Trimers		
AAC/ACA/CAA/GTT/TTG/TGT	5–17	16
AAG/AGA/GAA/CTT/TTC/TCT	0	0
AAT/ATA/TAA/ATT/TTA/TAT	5–18	8
ACC/CCA/CAC/GGT/GTG/TGG	5–8	5
ACG/CGA/GAC/CGT/GTC/TCG	5	1
AGC/GCA/CAG/GTC/TCG/CGT	5	3
AGG/GGA/GAG/CCT/CTC/TCC	5–14	10
AGT/GTA/TAG/ACT/CTA/TAC	5	2
ATG/TGA/GAT/CAT/ATC/TCA	5–17	4
GGC/GCG/CGG/GCC/CCG/CGC	5–17	17
Tetramers*		
AATT	5	2
ACCC	5	1
ACGC	7	1
AGAC	5–11	6
AGAT	6–17	5
ATGT	6–12	3
CAGG	13–10	2
CCCT	7–19	2
CCGC	5–6	3
CCTT	5	1
GGCT	5	1
GTAA	8	1
TAGT	6	1
TGAA	6	2
TTTA	5–15	3
TTTC	6–20	6
TTTG	5–10	7
Pentamers*		
AAACA	8	1
AATAA	7	1
ACCAC	8	1
CCGCT	5	1
CGTGC	6	1

*: Equivalent motifs in different reading frames or on a complementary strand were not listed to save space. Tetramers have equivalent motifs ACGC, CGCA, GCAC, CACG, GCGT, CGTG, and GTGC, while pentamers have 10 equivalent motifs.

sample (Figure 1). Among the 60 primer pairs, 56 (93%) primers showed clear amplified fragments and 4 (7%) did not amplify detectable products. After 3 independent PCRs, 30 primers showing consistent and reproducible amplification were selected to analyze 16 camels. In addition, they were all positive when tested for the Bactrian camel genome with similar allele amplifications (data not shown).

The 30 SSR primers revealed 61 amplified DNA fragments (alleles) that ranged from 1 to 3 alleles with an average of 2.03 alleles per primer combination across all 16 animals (Table 4). All primers showed an average of 62.8% polymorphism ranging from 0% (no polymorphism) to 100%. Results showed that more than 76% of primers produced more than 1 allele across all 16 animals. The number of SSR alleles scored per animal ranged from 1 to 3, and the average number of fragments ranged from 0.81 to 2. In total, applied markers generated 592 fragments across the tested animals; 14–32 fragments were generated per SSR marker with an average of 19.7. The PIC for all primers ranged from 0.0 to 0.66 with an average value of 0.38. The H_o and H_e values of each locus are presented in Table 4. The H_o ranged from 0 to 1 with an average of 0.26, whereas the H_e ranged from 0 to 0.69 with an average of 0.38.

4. Discussion

The present investigation was carried out to enrich the content of available camel molecular markers. The generated trace genome sequence served as the basis to achieve this goal. We assembled the reads into genomic contigs to extract SSR sequences. The utilization of NGS technology delivers more coverage than the conventional whole-genome sequencing approach (24). This coverage includes more SSR markers, as recorded in this study. The Illumina platform is very important for delivering good sequence depth and confidence, as shown in the SSR markers identified in alpaca (19). However, the Roche GS FLX platform is equally important in extending contig length, thus capturing long repeats flanked by unique signature sequences. Therefore, a mixed sequence would cover both good sequence depth and contig length.

The assembly generated contigs that were useful for primer design. The total SSR genome coverage varies between mammals. It can extend to 4.16% in mice, but decreases to a mere 0.78% in humans (20). The calculated SSR coverage in the analyzed partial camel genome was 0.021%, which represents a minor portion. However, this does not include motifs repeated twice, thrice, or 4 times. In fact, we observed many mononucleotide repeats within camel contigs. Mononucleotides are highly abundant in humans with an average appearance of 2.9 kb, thus exceeding all other nucleotide SSRs (25).

Table 2. Developed dromedary camel SSR markers with their repeats and PCR primers.

Locus	Accession number	Repeat motif	Primer sequence (5'–3')	T_m (°C)	Size (bp)
Cd00801	JX093499	$(AAAT)_{15}$	F: GATGCAACGGAGAAACGATC R: CCAAGATCATAAAGCTTAAGCC	52.0 52.0	254
Cd00802	JX093500	$(TA)_{12}$	F: GTCTGAATTCCCAATGTAACCC R: CAGGATGCTCTGCAATGTCAC	51.7 53.0	203
Cd00803	JX093501	$(TTG)_6$	F: TGTTCCTTGGGCTTACTTCC R: TGAGTCTTGCTACATACCAGGC	51.0 51.3	204
Cd00804	JX093502	$(CA)_8$	F: ATTCAAACCCAGGTCTCTGG R: GCAGAAGATCCATATGGAGCC	50.4 52.8	239
Cd00805	JX093503	$(GTAA)_8$	F: GTTCGATCTTCAGGACTTCCG R: CTTGCTGTCGTGATTCCAGG	52.9 53.0	322
Cd00806	JX093504	$(GCG)_{12}$	F: GTTCGTTGCTCGTGTGACG R: GCTGAGACTAAACACTGACGGC	52.2 53.2	331
Cd00807	JX093505	$(GA)_{15}$	F: TCAAGCCGGCTTTACAAGG R: AGCCTGCTTGACCCATGG	53.0 53.1	232
Cd00808	JX093506	$(AT)_9$	F: AGTGCAGGCACTTTATTGGG R: CGAGTTGGATGTTGTGTCTCC	51.9 51.8	238
Cd00809	JX093507	$(AGAT)_{10}$	F: GCACACACGCACACACACAC R: TATCTAACGGAGGAGGAGGCC	53.7 54.0	308
Cd00810	JX093508	$(AAC)_9$	F: TGGACTTGGGGAGTATTATGC R: TCCCTATCCCAGTCTTGCC	51.3 51.3	217
Cd00811	JX093509	$(GA)_8$	F: ACGCCCTAGGCTTCAAGG R: CTAGCCCTGAAAATGGATGG	51.3 51.8	283
Cd00812	JX093510	$(AAC)_{10}$	F: CCATGAGGTTCTCTGAAACCC R: GAGTAATTCCCTGAAATGGCC	52.5 52.0	292
Cd00813	JX093511	$(GTTT)_5$	F: AAAGCGTGCTGAACGATCC R: GACGTCAAAATCCTTAGGATGG	52.7 52.1	261
Cd00814	JX093512	$(TG)_{14}$	F: GCATAATGCCATCCAAGTCC R: GCCAAGGTATGGAAGCAACC	51.9 53.6	236
Cd00815	JX093513	$(AAC)_{11}$	F: CCATGAGGTTCTCTGAAACCC R: TGGCCCATCACTTGAAATACC	52.5 53.8	262
Cd00816	JX093514	$(CA)_{23}$	F: GCAGGGTCATTTTTAGCAGG R: ATGGTGAGCACAAGTGAGGG	51.6 52.2	317
Cd00817	JX093515	$(AT)_9$	F: ATCACCTGTGCTTCCTGCC R: GAAGGAAGGGTGCTGAAGG	52.2 51.1	285
Cd00818	JX093516	$(TG)_{12}$	F: AGTTATCCTTGAGGGCCTGC R: ACAGTGTTTCCCCTGTTCCC	52.5 52.6	320
Cd00819	JX093517	$(AT)_{19}$	F: AATCAGAAGCAGAACCCAAGC R: AAGGAGGTAAAGGAGGTGTGG	52.7 51.5	287
Cd00820	JX093518	$(CA)_{20}$	F: CTGTACACGTCCCACGACATG R: AACCATGCAAGAAGCCAGG	53.6 52.5	207
Cd00821	JX093519	$(CA)_{20}$	F: AGCTCATTCTCCCCAACCC R: AGTCCTCAGCTTGTGAATTGC	52.8 51.1	258
Cd00822	JX093520	$(AATAA)_7$	F: ACTCTCCGTATCTAGGGCCC R: GGTTTAGTGGTTCAAAGCCG	51.5 51.5	277
Cd00823	JX093521	$(GCGG)_6$	F: ATCCCTTTCACGCCAACC R: TCGTAACAAGGTTTCCGTAGG	52.0 51.3	298

Table 2. (Continued).

Cd00824	JX093522	$(TTTG)_5$	F: TCTTGTGATGCCTTTGTCTGG R: CATTCCCACGAGGAAATGC	52.6 52.7	210
Cd00825	JX093523	$(TG)_5$	F: AACACCATGCACTAAGCAAGG R: ATGTCTTGCCTTTCCCTTGC	52.0 53.3	352
Cd00826	JX093524	$(AC)_{11}$	F: TGAATGGTCTTCTAGTGGCCC R: AATGAGCCTGGAGGTAAGTGG	53.2 52.4	269
Cd00827	JX093525	$(TTTG)_5$	F: AATCCCAGTCTATCCCTTCCC R: TGCACCCCAATGTTCATAGC	52.7 53.2	368
Cd00828	JX093526	$(GT)_{20}$	F: AAGTGGTCCTTCTCCTTCAGC R: ACGTCTTGCCTTTCCCTAGC	51.7 52.7	278
Cd00829	JX093527	$(CA)_{10}$	F: CAGTGTTGGCTATGACCAAGC R: GGGGAATACTGACACAGAGGG	52.3 52.4	342
Cd00830	JX093528	$(TTA)_{18}$	F: GCTCAGCAAATACAGCAGCC R: TTCATAGCTGTCTGGCGTGC	52.7 53.8	352
Cd00831	JX093529	$(AATT)_5$	F: TGCTTAGCATGCACAAGGC R: GTGGGGAGGGCTATGTGG	52.3 52.2	215
Cd00832	JX093530	$(CATA)_{10}$	F: TGTGGGTTCATTTCAGGGC R: CTCCCTATAAGCCCACTTTGG	52.9 52.3	326
Cd00833	JX093531	$(AC)_{22}$	F: AATATGGGCTCAATTTGGCC R: CCTCTTGTTCATCTGGACTGG	53.1 51.1	302
Cd00834	JX093532	$(TTG)_{15}$	F: TCTCACTCTGCCTCCAGGG R: CTGAGCTTGACACTGATTGCC	52.3 52.3	237
Cd00835	JX093533	$(AGAC)_6$	F: AGGGAGACAGACAGACACGC R: CGGTGGCAGAAGGACTCC	51.4 52.6	242
Cd00836	JX093534	$(AC)_{10}$	F: ACGTCCCTCTCCCACTGG R: GGGTGGGGCTAGAACTCTACC	51.7 53.4	204
Cd00837	JX093535	$(AC)_{16}$	F: AACTGAGCTGATTCCAGCCC R: GGGAACAGGGAGTAGGTGG	53.2 50.6	236
Cd00838	JX093536	$(TG)_{17}$	F: GAGCCTGGAGGCAAGTGG R: TCTAATGACCCTCCCAGTTGG	52.7 53.0	257
Cd00839	JX093537	$(CA)_{16}$	F: CCAGTTGATTGGGAAATCCC R: TTCCAGATTGTGTGTGTGTGC	53.1 51.4	214
Cd00840	JX093538	$(TG)_{15}$	F: AAAGGTTTGAGCGCCACC R: CTGTCCTTCCAACTGTTCTGC	52.5 51.3	284
Cd00841	JX093539	$(CA)_5$	F: GCGTTCCCAACAAGCTAGG R: TGTGGAGGTGTACCAGCTCC	52.3 52.2	210
Cd00842	JX093540	$(AG)_5$	F: CATACCTCTTTGGCACTGTGG R: TCCTGCTATTGATTAGACACAGG	52.2 50.6	303
Cd00843	JX093541	$(AT)_7$	F: TGCCTGTTTCAAATTCCTGC R: GGAAGGGAAAGTAAATTTTCCG	52.7 53.0	609
Cd00844	JX093542	$(AT)_6$	F: CTTTGTGCTAGATGAACGAACG R: AATGGAACGGGTTGCAGG	52.0 53.0	255
Cd00845	JX093543	$(CA)_5$	F: GACTGGAAAACAGATTTGGAGC R: TCCTGTTTTGCTCGATGTACG	52.2 52.9	127
Cd00846	JX093544	$(TC)_6$	F: TGGTCTTGACAAATCTTACGACC R: TAAGGCATGATCTTTCACTCACC	52.6 52.7	431
Cd00847	JX093545	$(CA)_5$	F: TAAGATGAAAGGAAAAGAGAGCC R: TCTTGCCAATATGAGAAATTGC	51.4 50.9	242

Table 2. (Continued).

Cd00848	JX093546	$(TTG)_5$	F: TGCACATGTTTCCTCAGGG R: AGGTGACTGCTTTCATAAATGC	51.4 50.6	264
Cd00849	JX093547	$(TATT)_5$	F: CCATGCTGTACAGGAGGACC R: GCATTCTGAGTCCCAGAGAGG	51.7 52.8	435
Cd00850	JX093548	$(GT)_7$	F: CCCAAATTTCCCTCTCAACC R: GGTAATTAGCGGAGTTCCCC	52.5 52.0	211
Cd00851	JX093549	$(ATA)_5$	F: TCTTAGGGGTAGGATCAATTCC R: GTCAGTGCATCAGGCATCC	50.9 50.7	310
Cd00852	JX093550	$(TC)_6$	F: TATACGAGGTTCGGTGCTAGC R: CGTGGATGATTGGCTTAAGG	51.5 52.2	224
Cd00853	JX093551	$(CTAT)_{11}$	F: GGCAGCCCAGATCTATCTCC R: GCTCAGTGGTAGAGTGCATGC	52.7 52.3	463
Cd00854	JX093552	$(AC)_{10}$	F: GTGGGAACGAGAGCTCTGC R: TGGAGGACAATTGAGAGATAAGG	52.1 51.8	286
Cd00855	JX093553	$(CA)_{13}$	F: CTAGCCTCTTCCTCCATTTAGC R: CCTACAGGAGGCATACCTGC	51.2 51.3	250
Cd00856	JX093554	$(TC)_7$	F: CAACTGGGTGTTTGCTTGC R: TCCTCAGCCCAAACTCTCC	51.4 51.4	445
Cd00857	JX093555	$(GA)_5$	F: GGGACTATGGTTGCAGATGC R: CCTCCTAGGGTTCTTGAATGC	51.9 52.1	322
Cd00858	JX093556	$(GCC)_7$	F: ATGGGAGCTAATCCTCAAGC R: CGAACTGATGGAATAGCTGC	50.2 50.0	481
Cd00859	JX093557	$(CG)_5$	F: ACAGCCAGACAGACATACTAGCC R: GCTATCTATCTATGTGGGGAGGC	52.0 52.9	288
Cd00860	JX093558	$(TG)_{15}$	F: ACAATGTCAGGAGACCCAGG R: CCTTTGCTTCATTTACCTCTCC	51.0 51.7	513

T_m: Melting temperature.

SSR locus length can be calculated by multiplying the motif length with its repetition frequency (Table 1). Dimer motifs were found to be repeated up to 149 times (298 bp long). Dinucleotide repeat motifs tend to be longer than other repeats in several eukaryotic genomes (26). Long SSR motifs are expected to give a large number of alleles per locus due to greater potential for slippage (27). Few loci with many alleles will give an estimated genetic distance that is equivalent to that of many loci with few alleles (28). On the other hand, many loci with few alleles constitute crucial input for mapping purposes.

The abundance of specific SSR repeat motifs was investigated in several animals such as chicken (29) and alpaca (19). When studying the abundance of certain SSR motifs in any genome, all equivalent motifs in a grouping in different reading frames or on a complementary strand should be considered (26). Dimer SSRs have 4 groupings or classes, while trimers have 10 groupings (Table 1). Camel genome showed high frequency of dimer motif repeats (80.8%). This was likewise observed in several other eukaryotes (26). Camel SSRs with dimer and trimer motifs were compared with those of the related alpaca (19). The most abundant dimer in camel was AG/GA/CT/TC, with 50.1% compared to 30% in alpaca. The lowest dimer occurrence was recorded for GC/CG, and the comparable figures were 1.6% (camel) and 1.4% (alpaca). The motif AT/TA represented 31.5% (camel) and 31.6% (alpaca) of all dimers. As a percentage of all repeats, AT/TA occurrence was 25.4% in camel compared to 13.1% in alpaca (Figure 2). Considering the source of SSR sequences (genomic in camel and ESTs in alpaca) and the presumed synteny between them, it is probable that AT/TA repeats are almost equally dispersed between genic and intergenic sequences in camels. In sheep, the most abundant dimer repeat was found to be AC/CA/TG/GT (67%) (30). However, the SSR sequences were extracted from skin EST sequences and thus do not reflect the whole genome.

The camel genome showed 2 abundant trimer motif groupings, namely GGC/GCG/CGG/GCC/CCG/CGC (25.8%) and AAC/ACA/CAA/GTT/TTG/TGT (24.2%).

Table 3. BLAT search results with bovine. Only the top hit is indicated for each locus (the used query-database type was nucleotide–nucleotide).

Locus	BLAT Score	Start	End	Q size	Identity	Chromosome	Start	End	Span
Cd00801	49	18	172	254	71.5%	2	73873046	73873150	105
Cd00802	65	123	199	199	95.9%	4	118771549	118771949	401
Cd00803	61	37	135	204	82.7%	1	150016532	150016618	87
Cd00804	33	17	156	240	55.6%	14	32951266	32951306	41
Cd00805	36	184	219	322	100%	14	14237251	14237286	36
Cd00806	51	159	244	331	78.7%	4	97209810	97209881	72
Cd00807	75	95	204	229	94.4%	7	14999684	14999816	133
Cd00808	23	184	207	238	100%	10	39983556	39983584	29
Cd00809	78	52	173	282	83.5%	17	70719484	70719604	121
Cd00810	96	1	165	216	83.0%	15	13203076	13203224	149
Cd00811	160	2	282	282	89.1%	Un_AAFC02248261	792	1021	230
Cd00812	88	10	281	291	90.9%	Un_JH126266	1826	2255	430
Cd00813	30	99	132	260	97.0%	11	48406748	48406782	35
Cd00814	130	1	234	234	86.3%	X	67270088	67270289	202
Cd00815	47	49	231	260	92.8%	10	5700521	5700865	345
Cd00816	44	90	144	316	83.4%	19	63666336	63666384	49
Cd00817	100	118	284	284	83.1%	3	56808973	56809137	165
Cd00818	85	178	296	320	91.4%	17	321136	321268	133
Cd00819	108	86	225	285	88.6%	3	46283442	46283581	140
Cd00820	56	19	90	207	96.8%	23	42335175	42335346	172
Cd00821	150	19	258	258	85.9%	14	2508875	2509098	224
Cd00822	77	19	205	277	74.0%	11	14487391	14487537	147
Cd00823	124	79	297	297	85.3%	27	7281250	7281432	183
Cd00824	83	1	173	210	78.2%	8	54427216	54427364	149
Cd00825	35	117	302	353	71.8%	13	28557411	28557575	165
Cd00826	48	56	206	270	96.2%	18	28948221	28948421	201
Cd00826	20	216	235	270	100%	18	45252141	45252160	20
Cd00827	208	30	368	368	86.9%	X	80748743	80749108	366
Cd00828	56	26	268	278	98.3%	10	94786630	94787091	462
Cd00829	34	176	244	342	94.6%	29	30233605	30234055	451
Cd00830	159	11	323	351	82.0%	14	2477863	2478107	245
Cd00831	45	130	195	208	94.3%	23	8829852	8829918	67
Cd00832	99	1	228	324	80.8%	15	35395766	35395955	190
Cd00833	33	115	149	302	97.2%	15	32937537	32937571	35
Cd00834	32	36	221	237	58.9%	20	59854501	59854599	99
Cd00835	53	49	122	242	96.7%	12	83018654	83018764	111
Cd00836	23	21	44	203	100%	21	22051621	22051651	31
Cd00837	49	103	171	237	88.9%	5	99220633	99220699	67
Cd00838	32	169	202	256	97.1%	20	60265921	60265954	34
Cd00839	41	161	203	214	97.7%	26	15543536	15543578	43

Table 3. (Continued).

Cd00840	40	67	108	282	100%	Un_JH126349	9255	9462	208
Cd00841	28	105	135	210	86.3%	13	72622131	72622159	29
Cd00842	31	117	157	302	76.5%	14	12022068	12022101	34
Cd00843	28	456	490	608	94.0%	11	75030580	75030616	37
Cd00844	22	165	187	252	100%	Un_JH121384	233613	233637	25
Cd00845	22	62	83	173	100%	9	93324886	93324907	22
Cd00846	40	134	290	431	79.6%	1	23236292	23236439	148
Cd00847	29	38	78	239	94.0%	11	11109619	11109665	47
Cd00848	45	49	116	264	83.6%	15	5181726	5181796	71
Cd00849	130	54	258	435	86.6%	4	118753014	118753221	208
Cd00850	32	105	137	211	100%	27	45997276	46331741	334466
Cd00851	34	55	136	306	97.3%	13	66004976	66014684	9709
Cd00852	53	21	86	225	90.8%	1	142303731	142303798	68
Cd00853	100	109	277	461	81.9%	6	50881280	50881447	168
Cd00854	88	1	187	282	78.9%	21	1599943	1600124	182
Cd00855	44	53	134	250	81.3%	22	55348889	55348961	73
Cd00856	23	266	294	445	89.7%	X	42605701	42605729	29
Cd00856	23	366	388	445	100%	2	15598394	15598416	23
Cd00856	23	334	358	445	96.0%	27	34934426	34934450	25
Cd00856	20	373	392	445	100%	1	3406491	3406510	20
Cd00857	62	57	281	322	82.9%	11	68700495	68700712	218
Cd00858	28	208	238	499	96.7%	5	121089630	121089663	34
Cd00859	55	121	181	287	98.4%	24	29259853	29260218	366
Cd00860	24	471	499	513	92.9%	16	40407501	40407531	31

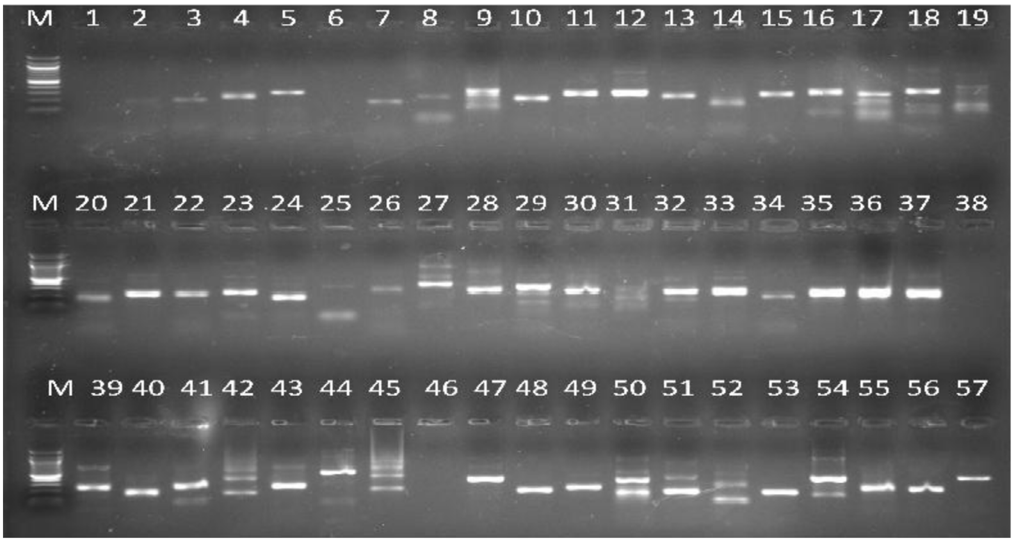

Figure 1. Screening of selected SSRs primers on pooled camel genomic DNA. M = 100-bp DNA ladder. Numbers 1–57 correspond to loci *Cd00801* and *Cd00857*, respectively.

Table 4. Characteristics of selected SSRs for genetic diversity in Saudi camels.

Locus	Total alleles	Average number of fragments*	Total number of fragments	Polymorphism %**	H_o	H_e	PIC
Cd00811	2	2.00	32	0	1.00	0.52	0.50
Cd00812	2	2.00	32	0	1.00	0.52	0.50
Cd00815	3	1.56	25	100	0.56	0.67	0.66
Cd00816	3	2.00	32	67	1.00	0.55	0.53
Cd00818	2	1.94	31	50	0.94	0.51	0.50
Cd00824	2	1.00	16	100	0.00	0.44	0.43
Cd00827	3	1.00	16	100	0.00	0.56	0.54
Cd00828	2	1.00	16	100	0.00	0.52	0.50
Cd00829	3	1.19	19	100	0.19	0.28	0.35
Cd00832	2	1.00	16	100	0.00	0.51	0.49
Cd00833	3	2.00	32	67	1.00	0.59	0.58
Cd00835	2	1.44	23	100	0.64	0.52	0.50
Cd00836	2	2.00	32	0	1.00	0.52	0.50
Cd00837	2	0.88	14	100	0.00	0.51	0.49
Cd00839	1	1.00	16	0	0.00	0.00	0.00
Cd00840	1	1.00	16	0	0.00	0.00	0.00
Cd00841	1	1.00	16	0	0.00	0.00	0.00
Cd00843	3	0.81	13	100	0.00	0.49	0.47
Cd00844	3	1.00	16	100	0.00	0.57	0.55
Cd00847	2	1.00	16	100	0.00	0.44	0.43
Cd00848	2	1.00	16	100	0.00	0.51	0.49
Cd00849	1	1.00	16	0	0.00	0.00	0.00
Cd00850	1	0.88	14	0	0.00	0.00	0.00
Cd00851	1	1.00	16	0	0.00	0.00	0.00
Cd00852	2	1.00	16	100	0.00	0.23	0.22
Cd00853	2	1.00	16	100	0.00	0.51	0.49
Cd00854	3	1.00	16	100	0.00	0.69	0.66
Cd00855	2	1.31	21	100	0.50	0.39	0.44
Cd00856	1	1.00	16	0	0.00	0.00	0.00
Cd00860	2	1.00	16	100	0.00	0.44	0.43
Total	61	------	592	-------	-------	-------	-------
Mean	2.03	1.23	19.7	62.80	0.26	0.38	0.38

*: Average number of fragments scored per animal.
**: Polymorphism % equals number of polymorphic alleles divided by total alleles.

The latter was also the most abundant trimer in alpaca (21.8%), whereas the former was very rare (1.7%) (19).

Molecular markers have provided new opportunities to assess animal genetic variability at the DNA level. Microsatellite markers have been widely used, since they are polymorphic and randomly distributed in the genome. In this study, 30 microsatellite loci were characterized using 16 Saudi camels that represented 4 morphologically diverse breeds. Twenty SSRs produced polymorphic information for the animals under study. They revealed 61 amplified DNA fragments (alleles) that ranged from 1 to 3 alleles with an average of 2.03. This range is comparable with that observed by Mehta et al. (11) in 3 Indian camel populations, where the range was 2–6 alleles using 16 SSR

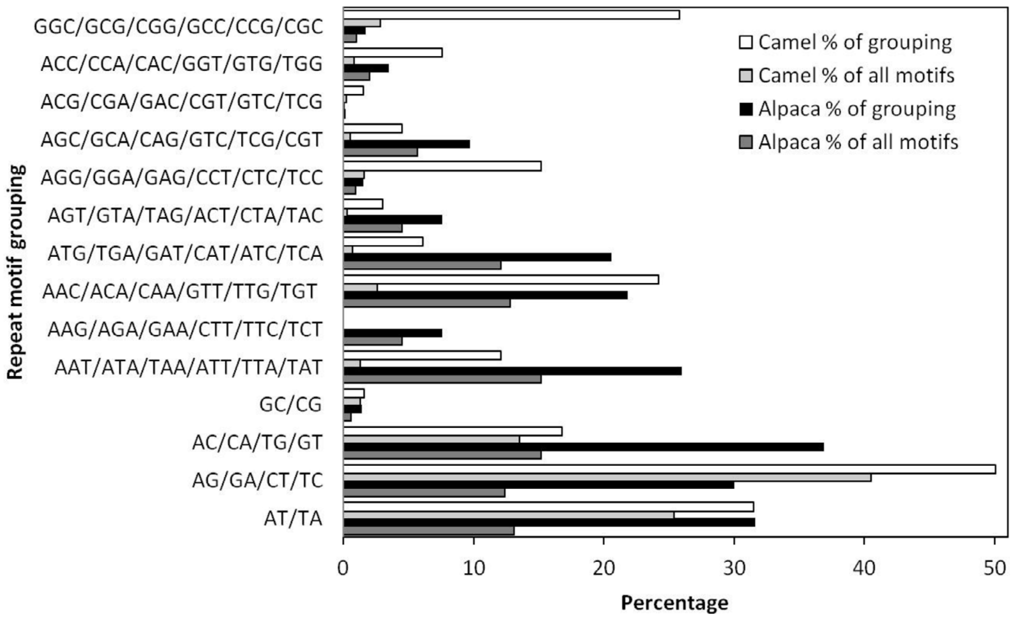

Figure 2. Abundance of SSR dimer and trimer motif groupings in camel (this study) and alpaca (24).

primers, and by Al-Swailem et al. (12) in 3 Saudi camel populations, where the range was 1–7 alleles. However, this number of alleles is considered low compared to earlier studies (14,15). Generally, the number of alleles is highly associated with sample size and the number of unique alleles in the population. As the sample size increases, the total number of expected alleles also increases. In a study on Saudi camels, Al-Swailem et al. (12) showed that 61 alleles were generated with an average of 3.81 alleles per locus, using 99 Saudi camels. Mburu et al. (9) found that a total of 115 alleles were observed at 14 loci in 332 camels from a study of 7 dromedary populations. Spencer and Woolnough (14) generated 185 alleles from 28 loci using 484 Australian camels belonging to 6 sampling locations.

PIC value is another important measure of polymorphism. The calculated PIC value in this study indicates relatively low polymorphism in the investigated population. The average PIC value was 0.38, which is close to the reported values of related studies using microsatellite markers in camel genetic diversity. The reported values were 0.48 (11), 0.51 (14), and 0.58 (15). Considerable polymorphism was detected among the investigated Saudi camels, which reflects their potential for future breeding purposes. In this study, H_o averaged 0.26, while H_e averaged 0.38. These values are considered low compared to reported data for Saudi camels, where H_e was 0.633, while H_o was 0.665, 0.605, and 0.662 for Majaheem, Maghateer, and Sofr breeds, respectively (15). Schulz et al. (13) recorded a value of 0.633 for Arabian camels from different regions. Conversely, Mburu et al. (9) recorded a value of 0.51 for camels from the United Arab Emirates, which could indicate narrow genetic selection for many generations. The low heterozygosity values in our study could be attributed to the small population size, which was used for characterization purposes.

The developed camel SSRs had a high score of BLAT matches, reflecting good synteny between bovine and camel genomes. Such synteny is helpful in comparative analyses of genetic maps. In conclusion, the present study developed insights into camel genomic SSR abundance and polymorphism. Thirty SSR markers were experimentally characterized and can be potentially utilized in genetic diversity analyses for both dromedary and Bactrian camels. The developed camel SSRs are expected to expand the available molecular marker toolbox and be further utilized for genetic mapping, identification of important QTLs, and breeding.

Acknowledgments

We gratefully acknowledge the financial support of the National Plan for Science and Technology at King Saud University, Saudi Arabia (project number 09-BIO855-02). The authors would also like to thank Dr Kalid Abdoun and Mr Emad MA Samara for providing blood and hair samples.

References

1. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E et al. Genetic diversity in farm animals – a review. Anim Genet 2010; 41: 6–31.
2. Al-Haidary A. Seasonal variation in thermoregulatory and some physiological responses of Arabian camel (*Camelus dromedaries*). J Saudi Soc Agr Sci 2006; 5: 30–41.
3. Kregel KC. Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. J Appl Physiol 2002; 92: 2177–2186.
4. Collier RJ, Collier JL, Rhoads RP, Baumgard LH. Genes involved in the bovine heat stress response. J Dairy Sci 2008; 91: 445–454.
5. Beuzen ND, Stear MJ, Chang KC. Molecular markers and their use in animal breeding. Vet J 2000; 160: 42–52.
6. Tautz D. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res 1989; 17: 6463–6471.
7. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991; 49: 746–756.
8. Teneva A. Molecular markers in animal genome analysis. Biotech Anim Husbandry 2009; 2: 1267–1284.
9. Mburu DN, Ochieng JW, Kuria SG, Jianlin H, Kaufmann B, Rege JEO, Hanotte O. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classifications. Anim Genet 2003; 34: 26–32.
10. Nolte M, Kotzé A, van der Bank FH, Grobler JP. Microsatellite markers reveal low genetic differentiation among southern African *Camelus dromedarius* populations. S Afr J Anim Sci 2005; 35: 152–161.
11. Mehta A, Goyal A, Sahani MS. Microsatellite markers for genetic characterization of Kachchhi camel. Indian J Biotechnol 2007; 6: 336–339.
12. Al-Swailem AM, Shehata MM, Al-Busadah KA, Fallatah MH, Askari E. Evaluation of the genetic variability of microsatellite markers in Saudi Arabian camels. J Food Agric Environ 2009; 7: 636–639.
13. Schulz U, Tupac-Yupanqui I, Martínez A, Méndez S, Delgado JV, Gómez M, Dunner S, Cañón J. The Canarian camel: a traditional dromedary population. Diversity 2010; 2: 561–571.
14. Spencer PBS, Woolnough AP. Assessment and genetic characterisation of Australian camels using microsatellite polymorphisms. Livest Sci 2010; 129: 241–245.
15. Mahmoud AH, Alshaikh MA, Aljumaah RS, Mohammed OB. Genetic variability of camel (*Camelus dromedarius*) populations in Saudi Arabia based on microsatellites analysis. Afr J Biotechnol 2012; 11: 11173–11180.
16. Lang KDM, Wang Y, Plante Y. Fifteen polymorphic dinucleotide microsatellites in llamas and alpacas. Anim Genet 1996; 27: 293.
17. Penedo MC, Caetano AR, Cordova KI. Eight microsatellite markers for South American camelids. Anim Genet 1999; 30: 166–167.
18. Evdotchenko D, Han Y, Bartenschlager H, Preuss S, Geldermann H. New polymorphic microsatellite loci for different camel species. Mol Ecol 2003; 3: 431–434.
19. Reed K, Chaves L. Simple sequence repeats for genetic studies of alpaca. Anim Biotechnol 2008; 19: 243–309.
20. Adelson DL, Raison JM, Edgar RC. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. P Natl Acad Sci USA 2009; 106: 12855–12860.
21. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11: 1441–1452.
22. Nei, M. Molecular Evolutionary Genetics. 1st ed. New York, NY, USA: Columbia University Press; 1987.
23. Anderson J, Churchill G, Autrique J, Tanksley S, Sorrells M. Optimizing parental selection for genetic linkage maps. Genome 1993; 36: 181–186.
24. Chaves LD, Knutson TP, Krueth SB, Reed KM. Using the chicken genome sequence in the development and mapping of genetic markers in the turkey (*Meleagris gallopavo*). Anim Genet 2006; 37: 130–138.
25. Cohen H, Danin-Poleg Y, Cohen CJ, Sprecher E, Darvasi A, Kashi Y. Mono-nucleotide repeats (MNRs): a neglected polymorphism for generating high density genetic maps *in silico*. Hum Genet 2004; 115: 213–220.
26. Katti MV, Ranjekar PK, Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 2001; 18: 1161–1167.
27. Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, Sibly RM. Likelihood-based estimation of microsatellite mutation rates. Genetics 2003; 164: 781–787.
28. Kalinowski ST. How many alleles per locus should be used to estimate genetic distances? Heredity 2002; 88: 62–65.
29. Bakhtiarizadeh MR, Arefnejad B, Ebrahimie E, Ebrahimi M. Application of functional genomic information to develop efficient EST-SSRs for the chicken (*Gallus gallus*). Genet Mol Res 2012; 11: 1558–1574.
30. Zhang W, Wang Z, Zhao Z, Zeng X, Wu H, Yu P. Using bioinpormcotics methods to develop EST-SSR makers from sheep's ESTs. J Anim Vet Adv 2010; 9: 2759–2762.