IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)
e-ISSN: 2319-2380, p-ISSN: 2319-2372.Volume 7, Issue 2 Ver. II (Mar-Apr. 2014), PP 65-70
www.iosrjournals.org

Identification of SNPs and their validation in camel (*Camelus bactrianus* and *Camelus dromedarius*)

Sushma Prasad[1,2], Sharique A. Ali[1], P. Banerjee[2], Jyoti Joshi[2], Upasna Sharma[2], R. K. Vijh[2]

[1]*Depatment of Biotechnology and bioscience, Safia Science College, Barkatullah University, Bhopal, Madhya Pradesh, India.*
[2]*Molecular and Comparative genomics lab, National Bureau of Animal Genetic Resources, Karnal-132001, Haryana, India.*

Abstract: *The dromedary* (*Camelus dromedarius*) *and the bactrian camel (Camelus bactrianus) are among the last species that have been domesticated around 3000–6000 years ago. To understand relationship between genetic and phenotypic variations in camel, single nucleotide polymorphism (SNP) markers covering the coding part of genome were developed. These gene-associated SNPs can themselves be causative SNPs for traits. The main objective of this work was to identify SNPs from coding regions using high-throughput next generation sequencing. The data was generated on two tissues as 75bp paired end reads by using Illumina platform. These reads were generated were mapped on cattle genes (Ensemble gene version 69.0). The mapping was carried out separately for each tissue and camel species. The mapped reads were analysed for SNP identification based on coverage depth. The 374 SNPs were validated using a set of 672 camels using golden gate assay of Illumina. The SNPs identified in this report provides a much needed resource for genetic studies in camel and shall contribute to the development of a high density SNP array. Validation and testing of these SNPs using SNP arrays will form the material basis for genome association studies and whole genome-based selection in camel.*

Keywords*: Transcriptome, single nucleotide polymorphism (SNPs), synonymous, nonsynonymous, RNA-Seq, C. dromedarius, C. bactrianus*

I. Introduction

The development of "next-generation" sequencing technologies and high-throughput genotyping platforms have tremendously advanced whole-genome analysis in domestic animals and nonmodel species [1], enhancing our understanding of animal responses to external abiotic stresses that disturb the homeostatic equilibrium of the animal body. Thermal stress triggers a complex program of gene expression and biochemical adaptive responses [2, 3]. Biologically, the ability to survive and adapt to thermal stress appears to be a fundamental requirement of cellular life. The cell stress responses are ubiquitous among livestock species including camel which thrive in different climatic regions ranging from hot dry regions with a temperature of +45°C to 50°C in semi humid regions of Rajasthan, Gujarat and Madhya Pradesh. In contrast to this camels also survive in the cold dry regions of Leh and Laddakh with sub 0 temperature (-20°C). An in-depth analysis is required to exactly fish out the genes involved and the pathways that are shared by various stressors and of those that are unique to particular stressor. Adaptability to various thermal stresses is likely to be the result of changed cellular responses, changed protein structure among two species of camels. The changed proteins might result from non-synonymous SNPs and adapted to the specific climatic region.

A simpler and potentially more comprehensive way to measure transcriptome composition and to discover new exons or genes is by direct ultra-high-throughput sequencing (RNA-Seq approach). RNA-Seq is a powerful new method for mapping and quantifying transcripts developed to analyse global gene expression in different tissues and also used to detect unannotated transcriptional activity, to differentiate between different transcriptional or splicing isoforms and to provide digital measurements at single base resolution. Recently, this technique has also been used as an efficient and cost-effective method to systematically identify SNPs in transcribed regions in different species (4, 5, 6, 7, 8)] as higher throughput expressed sequence reads are needed to increase coverage and depth and ensure sequence accuracy. Taking this into account, we applied this novel approach for the identification of gene, and to identify polymorphism in camel at nucleotide level (SNPs).

II. Materials and methods

2.1 Sample Collection and RNA isolation

Kidney and heart tissues were collected from two different species viz. *camel dromedarius* and *camelus bactrianus*. The fresh tissues were steeped in liquid nitrogen immediately after collection and brought to lab and processed for RNA extraction.

RNA was isolated following the standard protocols of RNeasy Kit (Qiagen). The mRNA comprises only 1- 3% of total RNA samples it was not readily detectable even with the most sensitive of methods. Quantification of RNA was done using Agilent 2100 Bioanalyzer (Agilent, Foster city, USA) which provided ng RNA/µl values. The clear 28S and 18S rRNA bands were indicative of intact RNA. The Bioanalyzer 2100 was used for all the 4 isolated samples of RNA. The RNA samples were processed further if the RNA Integrity Number (RIN) was found to be greater than 8.5, for RNA-Seq library preparation.

2.2 RNA-Seq library preparation and data generation

RNA library preparation was done by using Standard Illumina kit facilitated reading both the forward and reverse template strands of each cluster during one paired-end read. The unique paired-end sequencing protocol allowed us the length of the insert (200–300 bp), generating high quality, alignable sequence data. A typical paired-end run could achieve 2×76 bp reads and up to 35-60 million reads of the transcriptome data for each of the 4 RNA samples (2 each from kidney and heart). The heart and kidney tissues belonged to both the camels; *Camelus dromedarius* and *Camelus bactrianus*. The image analysis, base calling and quality score calibration were processed using the Illumina Pipeline Software v1.4.1 according to the manufacturer's instructions. Reads were exported in the FASTQ format and used for further analysis.

2.3 Sequence assembly and bioinformatics analysis

The 75 bp Illumia Solexa sequencing reads were first preprocessed by trimming adaptors and eliminating low quality reads and very short sequences. Trimmed RNA-Seq reads were aligned against *Bos taurus* genes downloaded from Ensemble Genome Browser, which is the most closely related species whose annotated genome sequence was available. We utilised RNA-Seq tool of CLC Genomics Workbench and all the default parameters were utilised for mapping of the reads of two tissues from camels of two geographical locations separately.

2.4 SNP identification

SNPs were identified utilizing SNP detection module included in CLC Genomics Workbench (CLC bio, Aarhus, Denmark). The central base quality score of ≥20 and average surrounding base quality score of ≥15 were set to assess the quality of reads at positions for SNP detection. We utilized the depth criteria of minimum allele frequency of 20% for the identification of SNPs. The SNPs were categorised into synonymous, non-synonymous and complex SNPs. Based on nucleotide changes the SNPs were categorized into transitions and transversions.

2.5 SNP validation

374 SNPs were selected based on depth and frequency. These SNPs were validated on 672 animals by Golden gate assay.

III. Results and discussion

3.1 Generation of data and assembly of transcriptome sequence

The Illumia Solexa sequencing of two tissues viz Kidney and heart of single humped camel and double humped camel were carried out. 62.6 million reads of heart and 59.1 million reads of Kidney of *C.dromedarius* while for *C. bactrianus* 34.2 million reads and 35.1 million sequence reads were generated (Table 1). After initial adapter trimming and quality filtering, the clean reads were aligned and assembled against annotated *Bos taurus* genes as reference. In *C. dromedarius* 8,328 reads were mapped for heart, while in *C. bactrianus* 9,426 reads were mapped for heart. In Kidney 16,836 reads were mapped for *C. dromedarius* and 90,865 reads were mapped in kidney for *C.bactrianus* (Table 1) (Figure 2).

3.2 SNP detection

From the RNA-Seq output, reducing false positive SNPs, we filtered potential SNPs using a stringent nucleotide depth cut off of 10. Large number of SNPs were detected in both the tissues shown in Fig.1

The analysis of total number of SNPs in Heart and kidney revealed that out of a total of 19162 SNPs in *C.bactrianus* heart, the total number of transversions was 8904 and the numbers of transitions were 6363. A total of 1439 SNPs detected in heart were complex. Similarly for kidney, the total number of 1354 SNPs were categorised into 258 transitions and 796 transversion. The total number of complex SNPs in kidney was 148. In *Camelus dromedarius* among 235 identified heart SNPs, 152 SNPs were transition and 71 were transversion while for kidney 22 SNPs of kidney were transition and 14 SNPs were transversion. It is generally assumed that the ratio of transitions to transversions is higher in animal nuclear genomes than the 1:2 ratio expected if all substitutions were equally likely. In present study it was observed that transitions were significantly higher in dromedarius compared to transversions.

SNPs in the coding region are of two types, synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the amino acid sequence while nonsynonymous SNPs change the amino acid sequence of protein. In the present study 24 and 10 nonsynonymous SNPs were identified in *C.dromedarius* for heart and kidney respectively, while in *C. bactrianus* 8344 SNPs were nonsynonymous in heart and 1354 SNPs in kidney. The details for the SNPs are given in Table 1 and Table 2.

Allele frequencies for heterozygous SNPs were obtained for the tissue samples by counting the number of reads representing each allele. Summarizing the information revealed that 3131 SNPs had 80/20 allele frequency, 808 SNPs had 75/25 and 553 SNPs had 76.9/23.1 allele frequency in kidney tissue of *C. bactrianus.* 326 SNPs had an allele frequency 80/20, 86 SNPs had an allele frequency of 75/25, 72 SNPs had an allele frequency 78/21 in heart tissue of bactrian camel.

In *camelus bactrianus,* maximum numbers of SNPs were identified on ENSBTAG00000017122 gene which is related with integral component of basement membranes and it plays essential roles in vascularization. Critical for normal heart development and for regulating the vascular response to injury. The gene is also responsible for vascular cartilage development. In kidney maximum SNPs were observed for ENSBTAG00000002485 gene and it is involved in fibrillar adhesion formation, involved in cell migration, cartilage development and in signal transduction pathways of cytoskeleton.

In heart tissue of dromedary camel, maximum numbers of SNPs were mapped to members of the perilipin family; it includes coat intracellular lipid storage.It plays a role in triacylglycerol packaging into adipocytes, while in kidney tissue maximum number of SNPs were identified on gene ENSBTAG00000040053, is a protein-coding gene which involve in immune response, signaling in eosinophils.

How these SNPs change the function of the proteins in a matter of study, science these nonsynonymous changes change the amino acid sequences of the proteins.

3.3 Validation of identified SNPs

The SNP genotyping was carried out using 672 camel genomic DNA samples. 374 SNPs were identified. The total percentages of SNP which were not been called in the samples were 5.81% (having zero value) and 1.17% were (NaN), 93.02 % of the SNPs were called in the samples.

The SNPs have been deposited in NCBI and the NCBI Accession number from NCBI_ss_947844623 to NCBI_ss_947845390

IV. Figures and tables

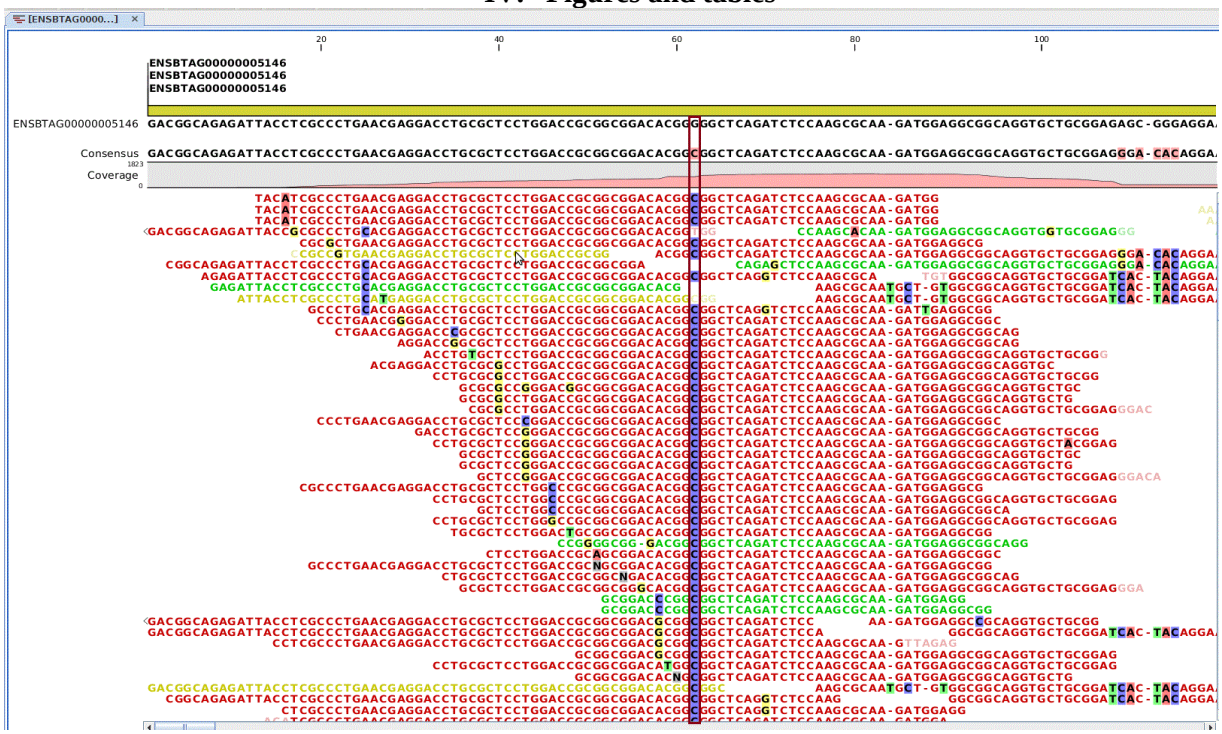
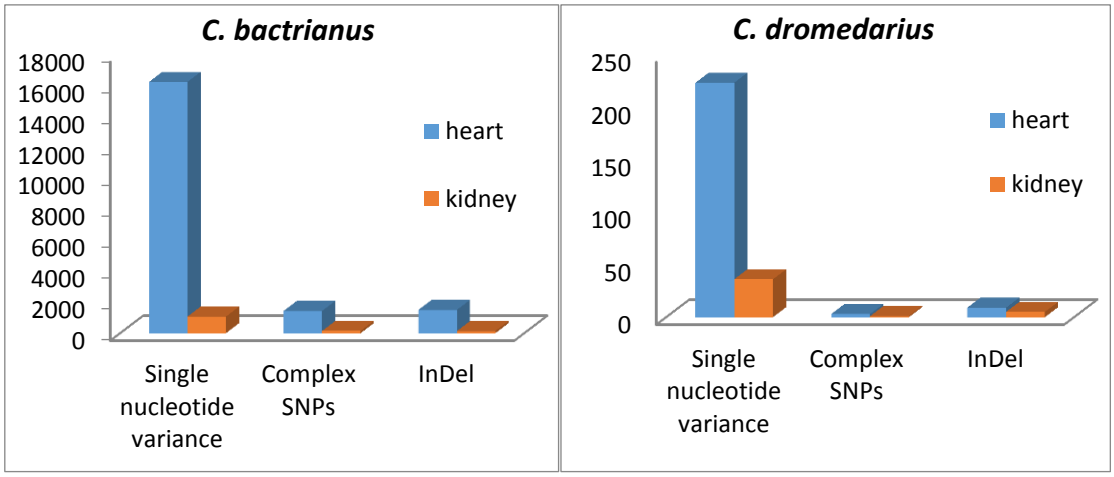
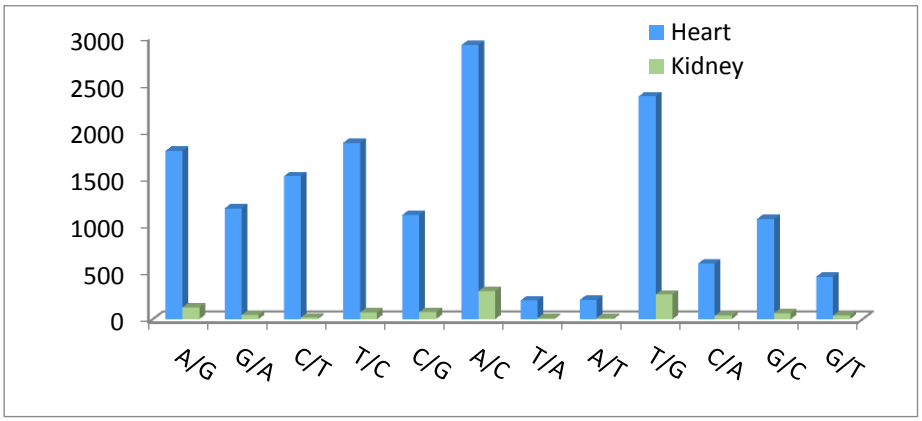


Fig. 1 Screenshot of CLC genomics workbench software showing SNP detection with *Bos taurus* as reference

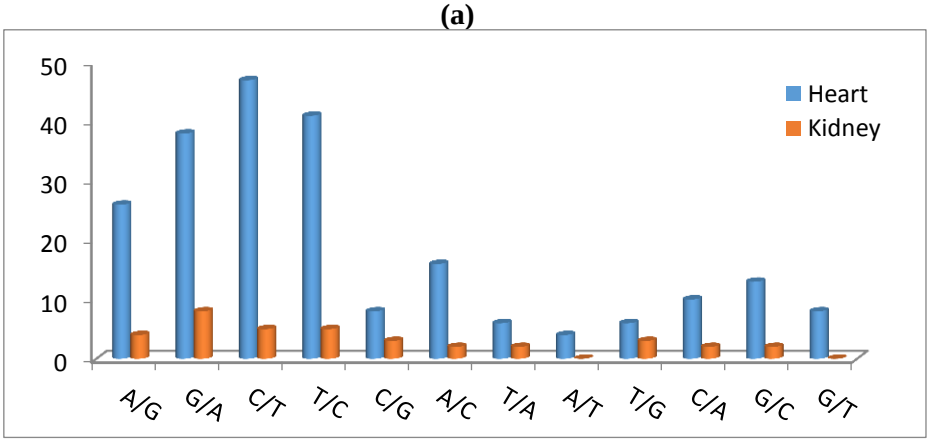


A B

Fig.2 Graphical representation of SNPs for two tissues in two camel species (A) *C. bactrianus*, (B) *C. dromedarius*.



(a)



(b)

Fig. 3 nucleotide variation type for two camel species (a) *C. bactrianus* (b) *C. dromedarius*, in two different tissues viz heart and kidney

Table 1 Summary statistics of SNPs for two tissues- total number of SNPs, complex SNPs, non-synonymous SNPs, transitions, transversions.

	Camelus bactrianus		*Camelus dromedarius*	
Tissues	**Heart**	**Kidney**	**Heart**	**Kidney**
Total reads (million)	**34.2**	**35.1**	**62.6**	**59.1**
Total reads mapped using RNA-Seq	**9,426**	**90,865**	**8,328**	**16,836**
Single nucleotide variation	**16221**	**1070**	**223**	**36**
Complex SNPs	**1439**	**148**	**3**	**1**
InDel	**1502**	**136**	**9**	**5**
Total SNPs observed	**19162**	**1354**	**235**	**42**
Non-synonymous SNPs	**8377**	**1354**	**24**	**10**
Transitions	**6363**	**258**	**152**	**22**
Tranversions	**8904**	**796**	**71**	**14**

Table 2 Summary of nucleotide variation

Type of Nucleotide variation	*Camelus bactrianus*		*Camelus dromedarius*	
	Heart	**Kidney**	**Heart**	**Kidney**
Transition				
A/G	**1790**	**125**	**26**	**4**
G/A	**1179**	**44**	**38**	**8**
C/T	**1520**	**17**	**47**	**5**
T/C	**1874**	**72**	**41**	**5**
Total	**6363**	**258**	**152**	**22**
Transversion				
C/G	**1109**	**75**	**8**	**3**
A/C	**2915**	**299**	**16**	**2**
T/A	**199**	**10**	**6**	**2**
A/T	**204**	**12**	**4**	**-**
T/G	**2370**	**262**	**6**	**3**
C/A	**591**	**36**	**10**	**2**
G/C	**1066**	**64**	**13**	**2**
G/T	**450**	**38**	**8**	**-**
Total	**8904**	**796**	**71**	**14**

V. Conclusion

In this study we estimated SNPs from an individual dromedary camel and bactrian camel transcriptome. It was observed that the transitions were higher in bactrian than dromedary camel, few SNPs were nonsynonymous and it provides a template for future genome-wide association and comparative studies to define underlying genotypes of specific favoured phenotypes. The Subsequent individual-based detailed analysis will have the purpose of locating regions with unusual high or low levels of nucleotide diversity that may point to loci and genomic regions under selection during the process of camel domestication in two extreme climatic regions. Thus using next generation sequencing we were not only in a position to identify the SNPs but also validated them using golden gate assay of Illumina.

Reference

Journal Papers:

[1]. T. Nawy, Rare variants and the power of association, *Nat Methods 9*, 2012, 324.
[2]. J. Fujita, Cold shock response in mammalian cells, *J. Mol. Microbiol. Biotechnol, 1*, 1999, 243-255.
[3]. S. Lindquist, The heat-shock response, *Annu. Rev. Biochem., 55*, 1986, 1151-1191.
[4]. I. Chepelev, G. Wei, Q. Tang, K. Zhao, Detection of single nucleotide variations in expressed exons of the human genome using RNASeq, *Nucl. Acid. Re, 37*, 2009, 106
[5]. E.T. Cirulli , A. Singh, K.V. Shianna , D Ge , J.P. Smith , J.M. Maia , E.L. Heinzen , J.J. Goedert , D.B. Goldstein, Screening the human exome: a comparison of whole genome and whole transcriptome sequencing *Genome. Biol., 11(5)*, 2010, 57.

[6]. N. Cloonan,., A. Forrest, G. Kolle, B. Gardiner, G. Faulkner,. Stem cell transcriptome profiling via massive -scale mRNA sequencing, *Nat. Methods. (5)*, 2008, 613-619.
[7]. Morin, P.A., G. Luikart, R.K. Wayne and S.W. Grp, SNPs in ecology, evolution and conservation. *Trends Ecol. Evol., 19(4)*, 2008a, 208-216.
[8]. R. Morin, M. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney. Application of massively parallel sequencing to micro RNA profiling and discovery in human embryonic stem cells. *Genome. Res., 18*, 2008b, 610-621.