IDENTIFICATION OF GENOME-WIDE SNPS AND SELECTION SWEEPS IN SAHIWAL CATTLE

THESIS SUBMITTED TO THE
ICAR-NATIONAL DAIRY RESEARCH INSTITUTE, KARNAL
(DEEMED UNIVERSITY)
IN PARTIAL FULFILMENT OF THE REQUIREMENT
FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ANIMAL GENETICS AND BREEDING

BY

VINEETH M. R

M.V.Sc. (Animal Genetics and Breeding)

ANIMAL GENETICS & BREEDING DIVISION

ICAR - NATIONAL DAIRY RESEARCH INSTITUTE

(DEEMED UNIVERSITY)

KARNAL-132001 (HARYANA), INDIA

2019

Regn. No. 15-P-AG-04

IDENTIFICATION OF GENOME-WIDE SNPS AND SELECTION SWEEPS IN SAHIWAL CATTLE

By

VINEETH M. R.

THESIS SUBMITTED TO THE
ICAR-NATIONAL DAIRY RESEARCH INSTITUTE, KARNAL
(DEEMED UNIVERSITY)
IN PARTIAL FULFILMENT OF THE REQUIREMENT
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ANIMAL GENETICS AND BREEDING

Approved by:

( )
EXTERNAL EXAMINER

(I. D. GUPTA)
MAJOR ADVISOR & CHAIRMAN

Members, Advisory Committee

1. **Dr. A. K. Gupta**
   Principal Scientist (Retd.), AG&B Division

2. **Dr. Archana Verma**
   Principal Scientist, AG&B Division

3. **Dr. Anupama Mukherjee**
   Principal Scientist, AG&B Division

4. **Dr. Jayakumar S**
   Senior Scientist, NBAGR

5. **Dr. A. K. Mohanty**
   Principal Scientist, ABTC
   (Director's nominee)

ANIMAL GENETICS & BREEDING DIVISION
ICAR - NATIONAL DAIRY RESEARCH INSTITUTE
(DEEMED UNIVERSITY)
KARNAL - 132001 (HARYANA), INDIA

Dr. I. D. Gupta, Ph. D.
Principal Scientist

C E R T I F I C A T E

This is to certify that the thesis entitled, "Identification of genome-wide SNPs and selection sweeps in Sahiwal cattle" submitted by Vineeth M. R in partial fulfilment of the award of the degree of Doctor of Philosophy in Animal Genetics and Breeding of the ICAR-National Dairy Research Institute (Deemed University), Karnal (Haryana), India, is a bonafide research work carried out by her under my supervision, and no part of the thesis has been submitted for any other degree or diploma.

(I. D. Gupta)
Major Advisor & Chairman

Dated: 16-02-2019

To Almighty God,

For his showers of blessings

To my parents and my little sister,

For their love and support

To my teachers,

For their motivation and guidance

To my friends,

For always being there for me

Acknowledgement

It is my pleasure to express my sincere thanks to all those who contributed in many ways to the successful completion of this thesis.

I feel my foremost duty to express my deepest gratitude to my benevolent advisor Dr. I. D. Gupta, Principal Scientist, Animal Genetics and Breeding Division for his enthusiastic supervision, guidance and moral support. Sir, you will continue to be an eternal source of inspiration to me. Any success that I have or will receive in my journey through science will perpetually be dedicated, in part, to you. Thank you, sir.

I express my sincere gratitude to members of my advisory committee, Dr. A. K. Gupta, (Head, AG&B Division), Dr. Archana Verma, Principal Scientist, AG&B Division; Dr. Anupama Mukherjee, Principal Scientist, AG&B Division; Dr. Jayakumar S, Senior Scientist, ICAR-NBAGR and Dr. A. K. Mohanty, Principal Scientist, ABTC for their constructive criticism, discussion and constant inspiration during the entire period of research work.

My utmost respect to my teachers Dr. Avtar Singh, Dr. S. M. Deb, Dr. Anupama Mukherjee and Dr. Vikas Vohra, Principal Scientists, AG&B Division, who made me sound in the subject, I will remain grateful to each one of them. A formal line of appreciation would hardly meet the end of justice to record my regards and gratitude to Dr. Jancy Gupta, for her love and inspiration during my stay at ICAR-NDRI.

My sincere thanks are due to Dr. R.R.B Singh, Director, ICAR-NDRI, Karnal, Dr. Arjava Sharma (Former Director, ICAR-NBAGR), Dr. A. K. Chakravarty (Former Head, AG&B Division) and Dr. A. K. Gupta, (Head, AGB Division), for providing all the necessary facilities to carry out the research. Financial assistance from ICAR in the form of Institute Fellowship is gratefully acknowledged.

I express my grateful thanks to Dr. Jayakumar S, Senior Scientist, ICAR-NBAGR, Dr. S. P Dixit, Principal Scientist, ICAR-NBAGR and Dr. S. K Niranjan, Senior Scientist, ICAR-NBAGR for rendering their precious time and help during my research and analysis work.

I owe profound thanks to my seniors Santosh Kumari, Mahesh, Ankit Magotra, Ashwani Arya, Ramendra Das, Mohsin Ayoub, Arun Pratap and Manvendra Singh for their immense help, invaluable guidance at every stage of my research.

I take this opportunity to thank my labmates Anshuman, Beena, Ragini, Rebeka, Sushil, Nisha, Siddu, Ravi, Kaushalya, Megha, Surya, Linda, Anjali and Joel, my friends Amit Sharma, Shrobana and Dimpee for being with me all through the dissertation work and for the wholehearted help and moral support rendered to me.

I am extremely thankful to my wonderful and ever helping batch mates Anshuman, Dimpee, Varinder, Suchit, Manoj, Prajwalita, Poonam, Beena, Ragini, Rebeka, Uday and Manjunatha Sir.

I wish to extend my heartfelt love and thanks to all my lovely juniors, Saleem, Govind, Naseer, Shaba, Anand, Ravi, Aneet, Martina, Ekta, Surya, Mohan, Narendra, Subroto, Vivek, Nandini, Aruna, Namith, Durai, Pooja and Ankita. They were always there in need and with them I enjoyed every moment during my stay at ICAR-NDRI.

Also, I take the pleasure to thank all staffs of AG&B Division for their help, care and affection. The help and support rendered by Suresh Bhaiya, Baldev Ji and Mahender Ji is gratefully acknowledged.

Where emotions are involved, words cease to mean. There are no words but feelings to honorably pay my best regards to my parents and to my loving sister for their love and care. They are the sole source of inspiration for all my pursuits. I gratefully acknowledge the concern, affection and understanding of my friends Nihin, Sarika, Sarath, Indu and Sunita mam which can only be felt and not expressed in words.

Not everyone is mentioned, but none is forgotten.

Dated: 16/02/2019

Place: Karnal (Vineeth M. R)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

A B B R E V I A T I O N S

A	:	Adenine
AFLP	:	Amplified fragment length polymorphism
AI	:	Artificial insemination
bp	:	Base pair
C	:	Cytosine
CDS	:	Coding sequence
CLR	:	Composite likelihood ratio
cM	:	Centimorgan
CNV	:	Copy number variation
ddRAD	:	Double digestion restriction-associated DNA
DNA	:	Deoyxribo nucleic acid
EST	:	Expressed sequence tags
G	:	Guanine
GBS	:	Genotyping-by-sequencing
Gm	:	Gram
GO	:	Gene ontology
GWAS	:	Genome wide assosiation study
HSP	:	Heat Shock Proteins
KEGG	:	Kyoto encyclopedia of genes and genomes
LD	:	Linkage disequillibrium
MAS	:	Marker assisted selection
Mg	:	Milli gram
Ml	:	Millilitre
NCBI	:	National centre for biotechnology information
Ng	:	Nano gram

NGS	:	Next generation sequencing
OD	:	Optical density
PCR	:	Polymerase chain reaction
QC	:	Quality control
QTL	:	Quantitative trait loci
RAD	:	Restriction-associated DNA
RAPD	:	Randomly amplified polymorphic DNA
RBC	:	Red blood cell
RE	:	Restriction enzyme
RFLP	:	Restriction fragment length polymorphism
ROH	:	Runs of homozygosity
Rpm	:	Revolution per minute
SE	:	Standard error
SFS	:	Site frequency spectrum
SIFT	:	Sorting intolerant from tolerant
SNP	:	Single nucleotide polymorphism
SS	:	Selection signature
SSR	:	Simple sequence repeats
SW	:	Sahiwal
T	:	Thymine
UMD	:	University of Maryland
UV	:	Ultra violet
µg	:	Micro gram
µl	:	Micro litre

Identification of Genome-wide SNPs and Selection Sweeps in Sahiwal Cattle

A B S T R A C T

Research in cattle genetics has profoundly changed since the public release of the cattle genome sequence, with genomics being adopted as a foundational tool for livestock breeding, health, welfare and conservation. Cost effective discovery and genotyping of large number of genome wide markers like SNPs and SSRs is a prerequisite for the application of genomics tools in indigenous cattle. Hence, the present study was carried out with the following objectives- i) To identify and annotate the genome wide SNPs and SSRs in Sahiwal cattle, ii) To compare the efficacy of the ddRAD approach with bovine SNP chip and whole genome sequence and iii) To detect selection sweeps in the Sahiwal cattle. The genomic DNA of 10 Sahiwal animals were sequenced using a reduced representation (ddRADseq) approach and raw sequence reads were obtained. The coverage of the sequenced regions was found to be 3.27% of the whole genome. The reads were analysed using standardised bioinformatics workflows utilizing minimum computational power and open source tools and variants were identified. A total of 450431 and 25821 genome wide SNPs were identified in Sahiwal cattle with reference to *Bos taurus* and *Bos indicus* genomes respectively. A total of 14908 and 150231 novel SNPs were found with reference to *Bos taurus* and *Bos indicus* genomes respectively. About 90% of the SNPs were genotyped in half of the samples showing high genotyping efficiency. The missense to silent mutations ratio was found to be 0.5-0.6 while the transition to transversion ratio was 2.3-2.4. SIFT analysis revealed 89 SNPs to be deleterious affecting the protein structure and function. Mapping of the identified SNPs revealed 22762 SNPs in production trait QTLs while 42314 SNPS were mapped to reproduction traits QTLs. SNPs were mapped to mastitis (5765), udder and teat type (661), tick resistance (7689) and heat tolerance (2300) QTLs. Among the validation set of 25 SNPs, 22 SNPs were successfully validated. A total of 8266 genome wide SSRs were identified in the Sahiwal cattle and their primers were designed. Lesser than 1% of the SNP identified in the present study was mapped to existing bovine SNP chips. Selection sweeps were detected by finding the likelihood of site frequency spectrum along the genome. Total 1764 genes were found to be in selection sweep regions in Sahiwal cattle. The genes responsible for domestication and tropical adaptation were found to be selected in Sahiwal cattle. There is no evidence for fixation of candidate genes for production trait in Sahiwal cattle. Genome-wide SNP identification and annotation in Sahiwal cattle using *Bos indicus* as reference genome was successfully carried out for the first time. It is also the maiden study to identify genome wide SSRs using sequencing data in Sahiwal cattle. The variants identified in the present study can be used as baseline information to facilitate further studies on domestication history, population structure and genetic improvement by trait mapping, GWAS in indigenous cattle breeds.

सारांश

मवेशियों के जीनोम अनुक्रम के सार्वजनिक विमोचन के बाद से मवेशियों के आनुवांशिकी में अनुसंधान में गहरा बदलाव आया है, जीनोमिक्स को पशुधन प्रजनन, स्वास्थ्य, कल्याण और संरक्षण के लिए एक मूलभूत उपकरण के रूप में अपनाया गया है। एसएनपी और एसएसआर जैसे जीनोम वाइड मार्कर की बड़ी संख्या की लागत प्रभावी खोज और जीनोटाइपिंग स्वदेशी मवेशियों में जीनोमिक्स टूल के आवेदन के लिए एक शर्त है। इसलिए वर्तमान अध्ययन निम्नलिखित उद्देश्यों के साथ किया गया था- i) साहीवाल मवेशियों में जीनोम व्यापक एसएनपी और एसएसआर की पहचान करना, ii) बोवाइन एसएनपी चिप और पूरे जीनोम अनुक्रम और ii) के साथ ddRAD दृष्टिकोण की प्रभावकारिता की तुलना करना। साहीवाल मवेशियों में चयन झाड़ू का पता लगाएं। 10 साहीवाल जानवरों के जीनोमिक डीएनए को कम प्रतिनिधित्व (ddRADseq) दृष्टिकोण का उपयोग करके अनुक्रमित किया गया था और कच्चे अनुक्रम रीड प्राप्त किए गए थे। अनुक्रमित क्षेत्रों का कवरेज पूरे जीनोम का 3.27% पाया गया। न्यूनतम कम्प्यूटेशनल शक्ति और ओपन सोर्स टूल्स का उपयोग करते हुए मानकीकृत जैव सूचना विज्ञान वर्कफ़्लोज़ का उपयोग करके रीड्स का विश्लेषण किया गया था और वेरिएंट की पहचान की गई थी। 450431 और 25821 जीनोम चौड़े एसएनपी की पहचान क्रमशः साहीवाल मवेशियों में बॉश टौरस और बोस सिग्नस जीनोम के संदर्भ में की गई थी। 14908 और 150231 उपन्यास एसएनपी क्रमशः बॉस् टोरस और बॉस सिग्नस जीनोम के संदर्भ में पाए गए। एसएनपी के लगभग 90% नमूनों को उच्च जीनोटाइपिंग दक्षता दिखाने वाले आधे नमूनों में जीनोटाइप किया गया था। मौन उत्परिवर्तन अनुपात का प्रकोप 0.5-0.6 पाया गया जबकि अनुप्रस्थ अनुपात का संक्रमण 2.3-2.4 था। SIFT विश्लेषण से पता चला कि 89 एसएनपी प्रोटीन संरचना और कार्य को प्रभावित कर रहे हैं। पहचान किए गए एसएनपी के मैपिंग से उत्पादन विशेषता क्यूटीएल में 22762 एसएनपी का पता चला जबकि 42314 एसएनपीएस को क्यूटीएल के प्रजनन लक्षणों के लिए मैप किया गया था। 5765, 661, 7689 और 2300 एसएनपी क्रमशः मास्टिटिस, उडर और टीट प्रकार, टिक प्रतिरोध और गर्मी सहिष्णुता क्यूटीएल के लिए मैप किए गए थे। 25 एसएनपी के सत्यापन सेट के बीच, 22 एसएनपी सफलतापूर्वक सत्यापित किए गए थे। साहीवाल मवेशियों में 8266 जीनोम विस्तृत एसएसआर की पहचान की गई थी और उनके प्राइमरों को डिजाइन किया गया था। वर्तमान अध्ययन में पहचाने गए एसएनपी के 1% से कम को मौजूदा गोजातीय एसएनपी चिप्स के लिए मैप किया गया था। जीनोम के साथ साइट आवृत्ति स्पेक्ट्रम की संभावना का पता लगाकर चयन झाड़ू का पता लगाया गया था। 1764 जीन साहिवाल मवेशियों के चयन क्षेत्र में पाए गए। वर्चस्व और उष्णकटिबंधीय अनुकूलन के लिए जिम्मेदार जीन को साहिवाल मवेशियों में चुना गया था। साहीवाल मवेशियों में उत्पादन गुण के लिए उम्मीदवार जीन के निर्धारण के लिए कोई सबूत नहीं है। संदर्भ जीनोम के रूप में बॉश सिग्नस का उपयोग करने वाले साहीवाल मवेशियों में जीनोम-वाइड एसएनपी पहचान और एनोटेशन को पहली बार सफलतापूर्वक किया गया था। साहीवाल मवेशियों में अनुक्रमण डेटा का उपयोग करते हुए जीनोम विस्तृत एसएसआर की पहचान करना भी पहला अध्ययन है। वर्तमान अध्ययन में पहचाने गए वेरिएंट का उपयोग आधारभूत जानकारी के रूप में किया जा सकता है, ताकि देशी मवेशियों की नस्लों में GWAS द्वारा मानचित्रण, जनसंख्या संरचना और आनुवांशिक सुधार पर आगे के अध्ययन को सुविधाजनक बनाया जा सके।

CHAPTER –1

Introduction

1. Introduction

India has a vast population of livestock which plays a key role in nutritional security by providing protein rich food for human consumption. It also provides the raw material for industries and draft power for the resource poor farmers. However, the overall productivity of our livestock is low as compared globally. The major causes of low productivity of livestock in India are both intrinsic (low genetic potential) and extrinsic (poor nutrition/feed management). Resource maximization by selection of genetically superior breeding stock is a feasible strategy for increasing productivity and profits. In the traditional selection program, the breeding value of animal is estimated based upon the information from its related individuals. Moreover, in traditional selection programs the selection of individuals depends upon availability of phenotypic observations which further depends upon the heritability of the trait. For traits with low heritability, traditional selection may not be possible and it may not provide a clear picture of the value of the animal. This method of selection is difficult to follow for the sex limited traits, traits expressed later in the animal's life and for the traits that cannot be measured easily. The genomic information of the individuals may help in the selection of individuals for these traits.

Application of molecular biology tools in cattle breeding initiated the search for major genes controlling quantitative traits (often production related), but had limited success. This is mainly because the single gene effects tend to be small and the numbers of available genetic markers are insufficient for estimating effects accurately. Discovery and development of large number of genetic markers are therefore essential for characterization and mapping of quantitative traits in cattle.

Among the first markers used were microsatellites, AFLPs, RFLPs. The relative high cost and limited marker density of these methods led to the use of Single Nucleotide Polymorphism (SNP) as the preferred genetic marker system. The SNPs are nucleotide variations in the DNA sequence of individuals in a population and are the most abundant molecular markers in the genome.

The major challenge lies in genotyping large number of genome-wide SNPs. Cost efficiency is also an important element in generating high density genotypic data. The expected

value of information gained by genotyping must justify the cost of obtaining the genotypes. However, the introduction of single nucleotide polymorphism arrays made the collection of genome wide marker data feasible at fairly affordable costs. Common SNP genotyping platforms like the array based ones uses the SNPs that were previously discovered by DNA sequencing. These SNPs may not be geographically representative and tend to be at higher frequency than random SNPs. This method introduces substantial ascertainment bias and inherently excludes detection of rare or population-specific variants, a major source of information for both population history and genotype phenotype association. This may impair estimation of population parameters like diversity, population subdivision, recombination and the identification of causal mutations.

Moreover, Indian and African cattle which are having higher effective population size and higher sequence heterozygosity (The Bovine HapMap Consortium 2009) showed lower heterozygosity when genotyped using SNP arrays. This is because the SNP arrays are primarily constructed from SNP identified from taurine cattle.

The recent advances in DNA sequencing technologies have facilitated the development of more efficient and cost effective strategies that allow simultaneous discovery and genotyping of SNPs in multiple individuals. These methods collectively known as RADseq techniques (or GBS methods/a type of reduced representation approaches) combines the strength of next-generation sequencing (NGS) to produce enormous numbers of DNA sequences from the ends of genomic restriction fragments with DNA bar-coding for multiplexing of samples. A range of protocols including restriction-associated DNA (RAD), genotyping-by-sequencing (GBS), double digestion restriction-associated DNA (ddRAD) are currently available for obtaining subsets of genomic restriction fragments for NGS. These methods have advantages over other SNP genotyping platforms as it reduces the ascertainment bias and can be developed easily for genotyping the SNPs of predetermined area over many samples in cost effective manner. These methods originally developed for other species have been lately adopted for cattle also.

The SNPs genotyped by this method could be used for QTL mapping, genomic selection, genome wide association studies (GWAS), population studies and characterization of genetic resources and identification of selective sweeps.

Majority of the studies on genome wide SNP identification/genotyping has been carried out in taurine cattle and there is scanty literature available about such works in indigenous cattle breeds of India, mainly due to economics associated with it. Developing a cost effective method of whole genome SNP genotyping in the indigenous cattle breeds of India will facilitate further insights on domestication history, population structure and genetic improvement by trait mapping, GWAS and genomic selection in the future.

Hence, the present study has been carried out with the following objectives:

1. **To identify and annotate the genome wide SNPs and SSRs in Sahiwal cattle.**
2. **To compare the efficacy of the ddRAD approach with bovine SNP chip and whole genome sequence.**
3. **To detect selection sweeps in the Sahiwal cattle.**

CHAPTER –2

Review of Literature

2. Review of Literature

2.1 Genetic markers

Genetic marker can be defined as any stable and inherited variation that can be measured or detected by a suitable method, and can be used subsequently to detect the presence of a specific genotype or phenotype other than itself, which otherwise is non-measurable or very difficult to detect. The markers revealing variations at the DNA level are referred to as the molecular markers. On the basis of techniques used for their detection, these have been classified into two major categories: Hybridization-based markers and PCR-based markers.

Restriction Fragment Length Polymorphism (RFLP) (Botstein *et al*., 1980) is an example of hybridization based marker. The PCR based markers includes VNTRs (microsatellites and mini satellites) (Nakamura *et al*., 1987, Litt and Lutty, 1989), Single Strand Conformation Polymorphism (SSCP) (Orita *et al*., 1989), Randomly Amplified Polymorphic DNA (RAPD) (Williams *et al*., 1990) and Amplified Fragment Length Polymorphism (AFLP) (Vos *et al*., 1995).

2.2 Simple Sequence Repeats

Microsatellites or simple sequence repeats (SSRs) are tandemly repeated motifs of 1-6 bases found in prokaryotic and eukaryotic genomes. They are present in both coding and non-coding regions and are usually characterized by a high degree of length polymorphism (Sawaya *et al*., 2013). The origin of such polymorphism appeared most likely due to slippage events during DNA replication (Schltterer and Tautz, 1992). The SSRs were being widely employed in many fields soon after their first description (Litt and Luty, 1989; Tautz, 1989; Weber and May, 1989) because of the high variability which makes them very powerful genetic markers. Microsatellites are extremely valuable tool for genome mapping in many organisms, their applications span over different areas ranging from ancient and forensic DNA studies, to population genetics and conservation/management of biological resources (Heubl, 2010; Cavagnaro *et al*., 2011; Dalamu *et al*., 2012; Li *et al*., 2012a; Stolle *et al*., 2013) as well as commercial application like true hybrid identification (Manigbas and Villegas, 2004; Dongre *et al*., 2011).

Microsatellites are developed by the following methods:

a) Screening of genomic libraries
b) *In-silico* mining from DNA-sequence databases
c) Sequencing of whole genome or part of it using NGS (Senan *et al*., 2014).

2.3 Single Nucleotide Polymorphisms

The most abundant source of genetic polymorphisms is Single Nucleotide Polymorphisms (SNPs), representing a single base change between two individuals at a defined location, it is considered that the least frequent allele should have a frequency of 1% or greater (Vignal *et al*., 2001). There are two different categories of SNPs: transitions (C/T or G/A) and transversions (C/G, A/T, C/A, or T/G). Some studies consider one base pair indels (insertions or deletions) as SNPs, although they certainly occur by a different mechanism. The SNPs at any particular site could in principle be bi-, tri- or tetra-allelic, however tri- and tetra-allelic SNPs are rare and in practice SNPs are generally biallelic (Doveri *et al*., 2008). The life cycle of SNP can be divided into four phases:

a) Appearance of a new variant allele by nucleotide mutation
b) Survival against odds of the allele through early generations
c) Increase to substantial frequency by survival through population fluctuation
d) Fixation (Millar and Kwok, 2001)

SNP is the most widely used genetic marker due to the following reasons:

a) Most of the SNP markers are located in coding area of DNA; therefore, they affect protein function directly
b) SNPs are more suitable than microsatellites for high throughput genetic analysis
c) They are stably inherited than other DNA markers, making them more suited as long term selection markers
d) SNPs are more prevalent and provide potential markers near the locus of interest than other types of polymorphism.

However, SNPs are biallelic in nature, which means that there are only two alleles in a population. Consequently, the information content per SNP marker is lower than multi allelic markers such as microsatellite markers (Beuzen *et al*., 2000).

2.3.1 SNP Identification

SNPs are identified either by Sanger sequencing or Next Generation Sequencing (NGS) technologies. Whole genome SNP discovery is mainly by re-sequencing the entire genome and aligning it to reference genome.

2.3.2 SNPs genotyping

SNPs genotyping is performed using two main methods, the traditional and high throughput methods. The traditional gel-based approach uses standard molecular techniques, such as amplification refractory mutation system (ARMS) (Newton *et al*., 1989), allele specific PCR (AS-PCR) (Wu *et al*., 1989), restriction digests and various forms of gel electrophoresis like RFLP, denaturing gradient gel electrophoresis (DGGE) and single strand conformation polymorphism (SSCP) (Orita *et al*., 1989), PCR amplification of specific alleles (PASA) (Sarkar *et al*., 1990) and tetra-primer ARMS-PCR (T-ARMS) (Ye *et al*., 2001),

High throughput methods include allele discrimination methods (allele-specific hybridization, allele-specific single-base primer extension), high-throughput assay chemistry methods like flap endonuclease discrimination, oligonucleotide ligation etc., DNA arrays, pyrosequencing and melting curve analysis (Kwok, 2000; Koopaee and Koshkoiyeh, 2014). Commonly used SNP genotyping platforms are summarized in the table 2.1.

Table 2.1 Commonly used SNP genotyping platforms (Kumar *et al*., 2012)

Name	Assay type	Technology	Throughput	Multiplexing
Genechip	Hybridization	Oligonucleotide array	96/5 days	Up to18 X 10^6
Infinium II	Hybridization	Bead array	Upto 128/5 days	Up to 13 X10^6
Goldengate	Primer extension-ligation	Bead array	172/3 days	Up to 3,072
IPlex	Primer extension	Mass spectrometry (MALDI-TOF)	3840/2.5 days	Up to 40
Taqman	PCR	Taqman probe	Upto1536/ day	Up to 256
SNPlex	PCR	Capillary electrophoresis	Upto1536/ day	Up to 48
KASPar	PCR	FRET quenching oligos	Up to 96/day	-
Invader	Primer annealing/ endonuclease digestion	FRET quenching oligos	Up to 384/day	Up to 200000
HRM	PCR	Melting curve analysis	Up to 1536/ day	-

2.4 DNA arrays/SNP chips

The need for simultaneous genotyping of large number of SNPs on thousands of individual samples has lead to the invention of SNP genotyping platforms, such as oligonucleotide arrays (Affymetrix, Inc., Santa Clara, CA, USA) (Matsuzaki *et al*., 2004) and BeadArray microarrays (Illumina, Inc., San Diego, CA, USA) (Steemers and Gunderson, 2007). At present Affymetrix is offering genotyping arrays for livestock and aquaculture species (buffalo, cattle, chicken, pig, salmon and trout), crops (cotton, maize, soybean, strawberry and wheat) and model organisms (human, dog, mouse and *Arabidopsis thaliana*) (http://www.affymetrix.com), while Illumina is marketing whole genome genotyping BeadArrays for human and non-human species (cattle, dog, maize, pig and sheep) (http://www.illumina.com). The different SNP chips available for cattle have been given in the table 2.2 (Nicolazzi *et al*., 2015).

Table 2.2 Different SNP chips for cattle

Sl. No	Name	No. of SNPs
1	Illumina Bovine 3k Bead Chip	2900
2	Illumina Bovine LD Bead Chip	6909
3	Illumina Bovine LD v1.1 Bead Chip	6912
4	Illumina Bovine LD v.2 Bead Chip	7931
5	Illumina Bovine SNP50 v.1Bead Chip	54001
6	Illumina Bovine SNP50 v.2 Bead Chip	54609
7	Illumina Bovine HD Bead Chip	777962
8	GeneSeek Dairy Ultra LD v2	7049
9	GeneSeek Genomic Profiler LD v1	8610
10	GeneSeek Genomic Profiler LD v2	19721
11	GeneSeek Genomic Profiler LD v3	26151
12	GeneSeek Genomic Profiler HD	76879
13	GeneSeek Genomic Profiler HD v2	139480
14	ICBF International Dairy and Beef v2	17807
15	ICBF International Dairy and Beef v3	53262
16	Affymetrix Axiom Bovine	648875

Despite the fact that SNP genotyping technology has enabled successful genome-wide association studies (GWAS) in humans and in livestock species (Meuwissen *et al*., 2001; Goddard and Hayes, 2009; Hindorff *et al*., 2009). However, an ascertainment bias is sometimes introduced which is derived from the fact that the SNPs used are chosen to have a minimum rare allele frequency as well as to segregate in multiple breeds. The identification of rare causal mutations might be complicated due to failure to detect the disequilibrium between causal mutations and genotyped SNPs (Nielsen *et al*., 2004; Clark *et al*., 2005). The commercially available SNP arrays or chips cannot be easily modified to suit individual experimental designs. Moreover, relevant experiments cannot be conducted for species that do not have commercially available SNP arrays/chips. Although customized SNP arrays/chips can be manufactured, but they are cost-prohibitive to many researchers.

2.5 Next Generation Sequencing

Sequencing refers to the identification of the nucleotides in a polymer of nucleic acids, whether DNA or RNA. Since its inception in 1977 (Maxam and Gilbert, 1977; Sanger *et al*., 1977), sequencing has increased our understanding of the organization and composition of animal genomes.

Next generation sequencing refers to non Sanger based high-throughput DNA sequencing technologies. Next generation sequencing (NGS) technologies, like (a) Pyrosequencing/Roche (454) sequencing (Ronaghi *et al*., 1998), (b) SOLiD (Applied Biosystems) sequencing (Shendure *et al*., 2005) (c) Illumina/Solexa sequencing (Turcatti *et al*., 2008) generate millions of reads in a relatively short time, making them powerful tools for genome research. The high throughput capability of genome sequencing or re-sequencing projects can efficiently and accurately discover and genotype many thousands of genetic polymorphisms, mainly SNPs, which can be used to investigate quantitative, functional, and evolutionary genomics in human, animals, and plants.

2.6 Reduced representation approach for genome wide marker identification

In order to address the limitations associated with SNP arrays/chips and the high/low coverage whole genome sequencing/re-sequencing platforms described above, the genome research community has been developing alternative strategies to discover and genotype genetic variants in a cost effective manner. Basically, these alternative methods/techniques are NGS-

based, but different laboratory procedures can result in different data outcomes in terms of reduced genome complexities, reduced genome representations or selected genome targets. In contrast to whole genome sequencing/re-sequencing, a basic feature associated with these methods/techniques is to have a subset of a genome sampled and sequenced. Currently available reduced representation methods have mainly evolved from:

a) Reduced Representation (Library) Sequencing (RRS or RRLS) (Altshuler *et al*., 2000; Van Tassell 2008),
b) Complexity Reduction of Polymorphism Sequencing (CRoPS) (van Orsouw *et al*., 2007),
c) Restriction Site Associated DNA Sequencing (RADseq) (Baird *et al*., 2008)
d) Genotyping by Sequencing (GBS) (Elshire *et al*., 2011) methods.

However, cross-interpretations of different assays with the same terminologies or cross-labeling of the same assays with different terminologies have occurred frequently in the literature.

The term RADseq was originally used to describe one particular method (Baird *et al*., 2008), but has subsequently been adopted to refer to a range of related techniques that rely on restriction enzymes to determine the set of loci to be sequenced. These methods are also sometimes grouped under the term 'genotyping by sequencing' (GBS) techniques (Narum *et al*., 2013). As with RADseq, the term GBS was originally used to describe one specific method (Elshire *et al*., 2011); however, this term is less descriptive than RADseq, which captures the defining feature of these methods, that is, the use of restriction enzymes to obtain DNA sequence at a genome-wide set of loci.

2.7 GBS/RADseq family of methods

Traditionally, reduced representation/GBS/RADseq methods involve three common steps:

1) DNA digestion with restriction enzymes
2) Ligation with adapters that fit into the relevant sequencing platforms
3) PCR amplification to increase the yield of library products for sequencing.

The genome sampling process depends on the use of rare-cutter enzymes, size selection of products or selective amplification of products. Adaptors added during RADseq protocols may contain barcodes, which are used to identify individual samples that are sequenced together (multiplexed) in a single library. Depending on the enzyme(s) used, RADseq protocols also reduce and/or select the sizes of DNA fragments that are optimal for next-generation sequencing.

RADseq methods differ in the order and details of enzyme digestion, adaptor ligation, barcoding and size selection, as well as the type of sequence data that can be produced at each locus. These differences can be used to categorize techniques into major groups. Below, we discuss important variations among methods as classified by Andrews *et al*. (2016)

2.7.1. Methods that sequence fragments adjacent to single restriction enzyme cut sites

Original restriction site-associated DNA sequencing (RADseq) (Miller *et al*., 2007; Baird *et al*., 2008) digests genomic DNA with one restriction enzyme, followed by mechanical shearing to reduce fragments to the appropriate length for sequencing, which unlike other methods, creates variance in the fragment sizes at each locus. The 2bRAD (Wang *et al*., 2012; Guo *et al*., 2014) method uses type IIB restriction enzymes, which cleave DNA upstream and downstream of the recognition site, resulting in short fragments of uniform length (33–36 bp).

2.7.2 Methods that sequence fragments flanked by cut sites of two restriction enzymes

Single enzyme indirect size selection: Genotyping by sequencing (GBS) (Elshire *et al*., 2011) uses a common-cutter enzyme, and PCR preferentially amplifies short fragments. Sequence-based genotyping (SBG) (Truong *et al*., 2012) uses a rare cutter and one or two common cutters, and PCR preferentially amplifies short fragments.

Double enzyme indirect size selection: Complexity reduction of polymorphic sequences (CRoPS) (van Orsouw *et al*., 2007) uses two enzymes and a proprietary library preparation kit.

Single enzyme direct size selection: Reduced representation libraries (RRLs) (Van Tassell *et al*., 2007; Greminger *et al*., 2014) are unique in using a blunt-end common-cutter enzyme followed by a size selection step and a proprietary Illumina library preparation kit. Multiplexed shotgun genotyping (MSG) (Andolfatto *et al*., 2011) uses one common-cutter enzyme and a size selection step. The ezRAD uses one or more

common-cutter enzymes, and a proprietary kit for Illumina library preparation (Toonen *et al*., 2013).

Double enzyme direct size selection: Double-digest RAD (ddRAD) (Peterson *et al*., 2012) uses two restriction enzymes, with adaptors specific to each enzyme, and size selection by automated gel cut.

Variations on the above techniques include use of methylation-sensitive enzymes (EpiRAD) (Schield *et al*., 2015); adding more restriction enzymes to existing protocols to further reduce the set of loci (RESTseq) (Truong *et al*., 2012*;* Stolle and Moritz, 2013); adding a second digestion to eliminate adaptor dimers (Graham *et al*., 2015); adapting RADseq techniques to other sequencing platforms such as Ion Torrent (Recknagel *et al*., 2015; Pukk *et al., 2*015); and other minor technical modifications (GGRS, flexible and scalable GBS) (Chen *et al.,* 2013; Heffelfinger *et al*., 2014)

2.8 Double digest RAD sequencing (ddRADseq)

This is a modified form of RADseq method with the following improvements: (a) simultaneous restriction digestion with two different RE (a rare cutter and a frequent cutter) (b) Two-index combinatorial tagging (Peterson *et al*., 2012). Comparison of various RADseq methods/GBS methods have been summarized in table 2.3. These modifications have given ddRADseq the following advantages over other methods:

a) Lower cost and lesser time to prepare the sequencing libraries
b) High level of multiplexing
c) Precise size selection
d) Lesser reads to achieve high confidence SNP calling

2.9 SNP Chip v/s ddRAD

DeDonato *et al*. (2013) compared GBS method with 50K Bovine SNP chip and found that GBS gives similar results as that of 50K Bovine SNP chip at a one third of the cost. No study comparing the GBS methods with the bovine high density SNP chips are reported. The comparison of SNP chip with ddRAD has been given in table 2.4.

Table 2.3 Comparison of different RADseq methods

Method	Steps						Remarks
RAD	Digestion	Ligation	Pooling	Shearing	Size Selection	Ligation	DNA Loss
GBS	Digestion	Ligation	Pooling	No size selection			
RRL	Digestion	Pooling	Size Selection	Ligation	No multiplexing		
ddRAD	Digestion	Ligation	Pooling	Size Selection	Advantageous		

Table 2.4 SNP chip v/s ddRAD

Principle		**SNP chips**	**ddRAD**
		Array based	Sequencing
Use in non-model species		No	Yes
Variant calling	Novel SNPs	No	Yes
	SSRs	No	Yes
	Indels	No	Yes
Genotyping		Yes	Yes
Commercial Availability	Cost/sample	~Rs. 33000	~Rs. 15000
	Protocol	Illumina Bovine SNP HD DNA Analysis 48 samples = 1547680/-	ddRAD Sequencing & Analysis 24 samples = 360000/-
	Source	M/s. Sandor Life sci., Cat No. SPGT-BT-HD	M/s. SciGenome, Cat No. SCI-NGS-Hiseq-ddRAD-Seq-65

2.10 Comparison between reduced-representation and whole-genome sequencing for SNP identification

Davey *et al*. (2011) reviewed that sequencing cost of 100 diploid human genome samples (consisting of two 3-gigabase sequences) at 30X coverage was 35 times as compared to restriction-site-associated DNA sequencing (RADseq) covering nearly 200,000 markers.

Furthermore, whole genome sequencing/re-sequencing for SNP genotyping is technically unnecessary and is cost-prohibitive, because linkage disequilibrium relationships (r^2 values) can be as higher as 95% - 100% among genetic markers within a genomic region (Jiang *et al*., 2009; Zhang *et al*., 2012). Commonly, only one of the highly linked markers is selected as a tagged SNP in data analysis. The whole genome sequencing studies in cattle for SNP discovery have been summarized in the table 2.5.

Table 2.5 Whole genome sequencing studies in cattle for SNP discovery

Breed		Sample	SNPs	Genome coverage	Depth	Reference
1	Flekvieh	1bull	2443637	98%	7.4X	Eck *et al*., 2009
2	Holstein Friesian	1 bull	6239482	98.3%	14.8X	Zhan *et al*., 2011
3	Kuchinoshima-Ushi	1 cow	6303790	93%	15.8X	Kawahara-Miki *et al*., 2011
4	Black Angus	1 bull	3200000	-	21.9X	Stothard *et al*., 2011
5	Holstein	1 bull	3700000	-	18.6X	
6	Gir	4 bull	9990733	80-88%	2.8-4.4X	Liao *et al*., 2013
7	Holstein	1 cow	5923230	-	36.7	Koks *et al*., 2013
8	Hanwoo	1 cow	6469804	98.6%	25.5X	Choi *et al*., 2014
9	Jeju Hengu	1 cow	6484293	98.5%	29.6X	
10	Korean Holstein	1 cow	5814990	98.5%	29.5X	
11	Danish Jutland	1bull	6812198	98.9%	26.4X	Das *et al*., 2014
12	Holstein bull	1 bull	6362988	-	60X	Koks *et al*., 2014
13	1000 bull genomes	234 bulls	1600000	-	8.3X	Daetwyler *et al*., 2014

2.11 Application of GBS/RADseq

The GBS/RADseq methods have a wide range of applications such as:

1) SNP and SSR discovery

RAD tag sequencing was used for identification of SSR marker in egg plant (*Solanum melongea*) by Barchi *et al*. (2011). Jansen *et al*. (2016) developed microsatellite markers for Goldsinny wrasse fish (*Ctenolabrus rupestris*) using ddRAD sequencing data. The GBS of microsatellites using individual combinatorial barcoding was found to be faster and cheaper approach for genotyping microsatellites in large scale population genetic studies (Vartia *et al*., 2016). The different methods of reduced representation approaches used in livestock species for SNP discovery are listed in table 2.6.

Table 2.6 SNP discovery using reduced representation approaches in livestock

Species	Method	SNPs	Reference
Cattle	GBS	52748	DeDonato *et al*., 2013
		272103	Brouard *et al*., 2017
		8065	Gurgul *et al*., 2019
		107488	Malik *et al*., 2018
	RAD	238725	Wang *et al*., 2018
Buffalo	GBS	49607	Imartino *et al*., 2013
	ddRAD	130688	Surya *et al*., 2018
Chicken	RAD	75587	Zhai *et al*., 2014
Pig	GBS	185206	Tan *et al*., 2015
Sheep	GBS	300000	Clarke *et al*., 2016
		39846	Gurgul *et al*., 2019
Goat	GBS	56000	Wheeler *et al*., 2018
Camel	GBS	310311	Holl *et al*., 2016
Horse	GBS	22433	Gurgul *et al*., 2019

2) Development of SNP genotyping arrays

SNP discovery was performed using Reduced Representation sequencing (RRSeq) and Restriction site-Associated DNA sequencing (RADseq) for development and validation of a high density SNP genotyping array for Atlantic salmon (Houston *et al*., 2014).

3) Conservation study

Genotyping by sequencing was used to resolve the shallow population structure of Chinook salmon (*Oncorhynchus tshawytscha*) and thus providing valuable information for its conservation (Larson *et al*., 2014).

4) Development of Genetic maps

High-density genetic maps were developed for barley and wheat using two-enzyme Genotyping by Sequencing approach (Poland *et al*., 2012a).

5) GWAS

Genome Wide Association (GWAS) studies of milk fatty acids of Canadian Holstein cows was carried out using genotyping by sequencing (GBS) approach by Ibeagha-Awemu *et al*. (2014) and they found that GBS is a robust, cost effective method to identify population-based unique SNPs for GWAS. Genotyping by sequencing (GBS) approach was used to provide dense genome-wide marker coverage (>47,000 SNPs) for Genome-Wide Association study (GWAS) in soya bean (Sonah *et al*., 2015).

6) Association mapping

A study by Constant *et al*. (2015) validated a whole genome genotyping by sequencing (GBS) method to identify single nucleotide polymorphisms (SNPs) associated with bovine paratuberculosis.

7) Study of domestication pattern

Genotyping by sequencing strategy was used to study the domestication patterns in chickpea by Kujur *et al*. (2015).

8) Genomic selection

GBS was used to discover and genotype SNPs for genomic selection in wheat (Poland *et al.,* 2012). Gorjanc *et al.* (2015) used simulation studies to evaluate the potential of

GBS for genomic selection in livestock populations. It was demonstrated that GBS offers great potential for developing genomic selection in livestock populations as it can cover large fractions of the genome and can vary the sequence read depth per individual. Thus, the accuracy of predictions can be improved by increasing the size of training populations and the intensity of selection is increased by genotyping a larger number of selection candidates.

9) Marker assisted selection

The GBS derived SNPs were used as markers in marker-assisted breeding programs in cassava (Prochnik *et al*., 2012). He *et al.* (2014) reviewed GBS as a potential tool for developing markers for MAS in plants.

10) Genomics of adaptation

RADseq data for two butterfly species (*Heliconius melpomene aglaope* and *Heliconius melpomene amaryllis)* was used to demonstrate the genomics of adaptation; linking to the genomic regions affecting colour patterns to adaptation in hybrid zones. Selection on colour pattern was found to be the most important factor maintaining butterfly hybrid zones (Nadeau *et al*., 2014).

11) Inbreeding and genomic diversity

RADseq genomic diversity estimates were also used to characterize the influence of social structure on autosome versus sex chromosome diversity in Tonkean macaque monkeys (Evans *et al*., 2014). A study investigating heterozygosity - fitness correlations in seals found that genome-wide heterozygosity estimated using 14,585 RADseq SNPs had nearly five-fold higher correlation with fitness associated trait than did 27 microsatellite loci (Hoffman *et al*., 2014).

12) Effective population size (N_e)

Thousands of SNPs generated using RADseq were used to estimate *N*e in salmon and smelt from western North America (Larson *et al*., 2014; Candy *et al*., 2015).

13) Population structure

RADseq was used to develop a population informative SNP panel to monitor stock composition in salmon and to differentiate population units to harvest as discrete rather

than mixed stocks (Dann *et al*., 2013). Double digest restriction site-associated DNA sequencing technique was used to study the population structure of white perch (*Morone americana*) (Bian *et al*., 2016).

14) Genetic introgression

Hohenlohe *et al.* (2013) used RADseq to identify 3,180 species diagnostic SNPs and to calculate admixture between a native and an invasive trout species. Saarman and Pogson, (2015) used ddRADseq to genotype 1337 randomly selected SNPs and found weak levels of introgression in both directions in the central California Mytilus spp. hybrid zone. Chattopadhyay *et al*., 2016 studied cryptic diversity and genetic introgression in oriental cynopterine fruit bat species using ddRADseq.

15) Phylogenomics

RADseq data generated a highly resolved tree for 16 species of Lake Victoria cichlid fish, whereas, previous analyses using amplified fragment length polymorphism (AFLP), microsatellites failed to resolve species level relationships for these species (Wagner *et al*., 2013). Escudero *et al.* (2014) used Genotyping by Sequencing (GBS) approach to infer the phylogeny of seven closely related species in the genus Carex (Cyperaceae).

2.12 Selection signature

Selection not only affects the favoured mutation but it produces a “hitch- hiking” effect on the frequency of neutral alleles at linked loci (Smith and Haigh, 1974; Kaplan *et al*., 1989). A local reduction of genetic variation up and downstream of the beneficial mutation is caused by the rapid fixation of a beneficial mutation, leaving special patterns of DNA behind (Smith and Haigh, 1974). Selective sweep is the process by which a new beneficial mutation eliminates or reduces variation in linked neutral sites as it increases in frequency in the population (Braverman *et al*., 1995). Selection signatures are regions of the genome that contain a beneficial mutation which leave special patterns of DNA behind under natural or artificial selection (Qanbari and Simianer, 2014) (Figure 2.1).

The classic model of positive selection states that selection acts upon a newly arisen advantageous mutation, so that there is only one founding haplotype at the time of selection. Alternatively, selection could also act on pre existing genetic variation that was previously either

neutral or deleterious, but has become adaptive due to changes in the environment or genetic background.

Selection from standing variation has been referred to as a soft sweep, to distinguish it from the classic model or hard sweep. These footprints of selection, can be detected from SNP data by well-established methods (Hermisson and Pennings, 2005; Sabeti *et al*., 2006; Akey, 2009; Oleksyk *et al*., 2010) and can unravel past responses of the livestock genome to natural and artificial selection as well as evidences of loci and variants underlying adaptive and economically important traits. Detecting these selection signatures may not only help to study the key adaptive events that have generated the vast phenotypic variation between livestock breeds, but can also be of biotechnological relevance giving functional information about genes/genomic regions (Nielsen, 2001, 2005; Schlotterer, 2003).

In general, SS studies take the opposite direction: one starts with an evidence of selection in samples sharing geographical proximity, environmental factors or a common phenotype, and attempts to find selected mutations (Sabeti *et al*., 2006). Another advantage of SS studies is that they typically require over 10 fold smaller sample sizes in comparison to GWA studies. Moreover, SS can reveal signals on genes controlling traits that are difficult, expensive or even impossible to measure on a large population (e.g., disease resistance). This approach could also lead to the identification of genes related to ecological traits (e.g., genes related to tropical adaptation) that are difficult to identify through laboratory experiments. Selection signature (SS) studies are growing in popularity as they offer a complementary strategy to genome-wide association (GWA) studies on mapping variants impacting traits of interest, helping to link phenotypes to gene function. In a typical GWA analysis, one starts from a phenotype and scans genotypes identifying underlying large and moderate effect variants (Bush and Moore, 2012). Other advantages of SS studies are that they can pinpoint chromosomal segments sheltering large effect mutations, even if they no longer segregate in a population. In such cases, these variants cannot be detected by classical quantitative genetics methods unless linkage experiments are designed using crosses (Ramey *et al*., 2013).

2.13 Selection signatures in livestock

Domestication has resulted in considerable changes in the morphology and behaviour of livestock species. In the early stages of domestication, unconscious selection for behavioural

traits was applied. This early stage was followed by methodical selection in which specific traits were selected based on goals (Diamond, 2002; Gregory, 2009). The development of specialized breeds, improved to produce specific products or to reach a morphological standard, increased the differences between domesticated animals and their wild relatives and also generated an enormous variety of different populations, with specific traits related to their specialization. Some of these traits are controlled by several interacting genes with minor effects. This creates an exceptional opportunity to gain knowledge of the molecular basis of these traits, particularly since most economically important traits in livestock are quantitative (Andersson and Georges, 2004). The identification of genes targeted by selection in livestock can help to find and prove causal mutations in regions previously identified by QTL mapping experiments and can reveal genes related to ecological traits. Furthermore, these studies can help to identify the genes or gene networks that contribute to the same trait but that were selected differentially between breeds; they can also unveil genes responsible for genetic correlations and the domestication process (Schlotterer, 2003; Hayes *et al.*, 2008; Ojeda *et al.*, 2008; Flori *et al.*, 2009; MacEachern *et al.*, 2009). The selection signatures in different livestock species are described below:

1) Cattle

There are many signatures associated with domestication and early breed development in cattle. The coat colour patterns are one such trait which was selected during domestication because particular patterns may have been used as markers associated with improved individuals or because of cultural preferences (Fang *et al.*, 2009; Wiener and Wilkinson, 2011). The MC1R gene (Melanocyte Stimulating Hormone Receptor) regulates the production of eumelanin and phaeomelanin pigments (Werth *et al.*, 1996; Kijas *et al.*, 1998; Fang *et al.*, 2009; Li *et al.*, 2010b) and influence the coat colour pattern. These genes are found to be under selection in domestic cattle (Flori *et al.*, 2009; Stella *et al.*, 2010). Selection signatures around other genes affecting coat colour like KIT (V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog) have been reported for cattle (Stella *et al.*, 2010; Wiener *et al.*, 2011), while PMEL17 (Melanocyte Protein 17 Precursor) also known as the SILV (Silver gene) is suggested to be under selection in cattle (Gautier *et al.*, 2009; Wiener *et al.*, 2011).

The presence or absence of horns is another important characteristic in defining breeds in some livestock species. The RXFP2 (Relaxin-Like Receptor 2) gene was found to be associated with this trait (Johnston *et al.,* 2011). In cattle, the region surrounding the polled locus was found to be under selection (Drogemuller *et al.,* 2005; Li *et al.,* 2010; Stella *et al.,* 2010). Reduction in fear and anti-predator responses and an increase in sociability are behavioural changes due to domestication (Diamond, 2002; Amaral *et al.,* 2011; Wiener and Wilkinson, 2011). Numerous studies in livestock suggest selection signatures surrounding genes related to nervous system development and function (The Bovine HapMap Consortium, 2009; Gautier *et al.,* 2009; Stella *et al.,* 2010; Amaral *et al.,* 2011).

2) Beef Cattle

In centromeric region BTA 14 selection signals were found in beef cattle using Integrated Haplotype Score (iHS) and Fixation Index approaches (F_{ST}) [F_{ST} is a measure of population differentiation due to genetic structure. It is frequently estimated from genetic polymorphism data, such as SNP or microsatellites] (Hayes *et al.,* 2009; The Bovine HapMap Consortium, 2009; Wiener *et al.,* 2011) and this region is reported to be involved in the regulation of marbling and fatness traits (Barendse, 1999; Pannier *et al.,* 2010; Veneroni *et al.,* 2010). Another selection signal was detected as decrease in heterozygosity around the GDF-8 (Growth Differentiation Factor 8 or myostatin) gene has been demonstrated in double muscle breeds (Wiener *et al.,* 2003; Wiener and Gutierrez-Gil, 2009). The F_{ST} approach was used to detect a selection signature in the median region of BTA2 and this region contains R3HDMI (R3H Domain Containing 1) and ZRANB3 (Zinc Finger, RAN Binding Domain Containing 3) genes which were found to be associated with feed efficiency in beef cattle (Barendse *et al.,* 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.,* 2011).

3) Dairy Cattle

On BTA6, three QTL regions that have been linked with milk traits (Khatkar *et al.,* 2004; Ogorevc *et al.,* 2009; Weikard *et al.,* 2012), which were found to be under selection in dairy breeds (Hayes *et al.,* 2008; Barendse *et al.,* 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.,* 2010; Schwarzenbacher *et al.,* 2012). The first region

contains the ABCG2 [ATP Binding Cassette, sub family G (WHITE) Member 2] gene which has been reported to associated with milk yield and quality traits (Olsen *et al.*, 2007; Cole *et al.*, 2009; Weikard *et al.*, 2012). The second region contains the PPARGC1A (Peroxisome Proliferator-Activated Receptor Gamma, Co-activator 1 Alpha) gene that has been shown to mediate the expression of genes involved in adipogenesis and gluconeogenesis (Weikard *et al.*, 2005; Ogorevc *et al.*, 2009). The third region harbours the Casein cluster genes which are associated with milk and protein yield (Sodeland *et al.*, 2011). A high linkage disequilibrium using Extended Haplotype Homozygosity (EHH) and iHS in dairy breeds in the region surrounding DGAT1 showed that the gene was under selection (Hayes *et al.*, 2009; Qanbari *et al.*, 2010; Hosokawa *et al.*, 2012; Schwarzenbacher *et al.*, 2012). The gene was found to have a major effect on milk fat percentage (Khatkar *et al.*, 2004; Cole *et al.*, 2009; Hayes *et al.*, 2010; Jiang *et al.*, 2010). Two QTLs for milk traits were mapped to the BTA20. The first QTL surrounded the GHR gene (Growth Hormone Receptor Gene) and was associated with milk yield and fat percentage. The second QTL overlapping PRLR gene (Prolactin Receptor) was associated with fat and protein yield. Studies have shown these regions to be under selection (Khatkar *et al.*, 2004; Cole *et al.*, 2009; Ogorevc *et al.*, 2009; Jiang *et al.*, 2010). Another selection signal was found in the region surrounding STAT1 gene (Signal Transducer and Activator of Transcription 1) (Hayes *et al.*, 2009; Hosokawa *et al.*, 2012). This gene was involved in mammary gland development and was associated with milk, fat and protein yield (Cobanoglu *et al.*, 2006). Using iHS approach the region surrounding SIGLEC-5 (Sialic Acid Binding Ig- Like Lectin 5) and ZNF577 (Zinc Finger Protein 577) was found be under selection (Qanbari *et al.*, 2011). These genes were associated with net merit and related traits like conformation, longevity and calving ease in Holstein Friesian cattle (Cole *et al.*, 2009).

Dash (2016) identified 41, 29 and 60 selective sweep regions in Sahiwal-Tharparkar, Sahiwal-Gir and Tharparkar-Gir breed pairs, respectively. The main candidate genes for milk performance traits found within the selective sweep regions were, ACADL (Acyl-CoA dehydrogenase, Long Chain), SLC26A2 (Solute Carrier Family 26 Member 2), PLCB1 (Phospholipase C Beta 1), SYT9 (Synaptotagmin 9), ATPAF1 (ATP Synthase Mitochondrial F1 Complex Assembly Factor 1), LEF1 (Lymphoid Enhancer Binding

Factor 1), EIF6 (Eukaryotic Translation Initiation Factor 6), ACSS3 (Acyl-CoA Synthetase Short Chain Family Member 3), PLA2R1 (Phospholipase A2 Receptor 1), SCP2 (Sterol Carrier Protein 2), CACNA2D1 (Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 1) and SUMF1 (Sulfatase Modifying Factor 1)

4) Pigs

Ojeda *et al*. (2008) identified a selection signature in IGF2 gene in Pietrain, Hampshire and Duroc breeds using Tajimas D approach. This gene mutation was associated with increase in muscle mass and decrease in fat content (Van Laere *et al.*, 2003). The MC4R gene (Melanocortin 4 Receptor) associated with growth and fatness traits has also been shown to be under selection in pigs (Rubin *et al.*, 2012; Onteru *et al.*, 2013). The ESR gene (Estrogen Receptor) which was associated with litter size in pigs (Rothschild *et al.*, 1996; Short *et al.*, 1997) has been found to be under selection in Large White breed (Bonhome *et al*., 2010).

5) Sheep and goats

IL-32 (Interleukin) gene was found to be under selection in goats breeds of China (Asif *et al*., 2016). The GDF-8 gene was found to have strong selection signal in sheep using F_{ST} approach in Texel breed (Kijas *et al*., 2012). The region surrounding this gene was associated with carcass traits in the breed (Johnson *et al.*, 2005; Clop *et al.*, 2006).

2.14 Methods for detection of selection signature

Based on the hitch-hiking theory positive selection can leave a set of informative signatures such as:

1) Reduced local variability

2) Deviated spectrum of allele frequencies

3) A specific linkage disequilibrium pattern

Based on these signatures, a variety of statistical approaches are available for selection signature detection from SNP data (SNP-chip or sequence data). Qanbari and Simianer (2014) classified these methods in two main groups: intra-population statistics and inter-populations statistics. Intra-population statistics searches for informative signatures by comparison of

genomic data within populations. Intra-population statistics are focused on three neutrality theory:

1) **Site frequency spectrum (SFS)** is a class of tests which summarizes the allele frequency distribution of polymorphisms in a region of interest. A widely used statistic established in this class is Tajima's D (Tajima, 1989). A more recent statistic in this class is the maximum of composite likelihood ratio (CLR) (Nielsen *et al*., 2005).

2) **Linkage disequilibrium (LD)** refers to the non random association of alleles between two or more loci. An ongoing or incomplete selection signature has a high-frequency haplotype with extended LD, which is mainly because recombination does not (or rarely) occur during the rapid increase in frequency of a haplotype carrying a beneficial mutation. Popular LD based tests include relative Extended Haplotype Homozygosity (rEHH) (Sabeti *et al*., 2002), Integrated Haplotype Score (iHS) (Voight *et al*., 2006) and Linkage Disequilibrium Decay Test (LDD) (Wang *et al*., 2006).

3) Reduced local variability is a class of methods that identify genomic regions with a systematically reduced variation (e.g., nucleotide diversity or heterozygosity) relative to the average across the genome. Some tests in this class are Runs of Homozygosity (ROH) (McQuillan *et al*., 2008) and Pooled Heterozygosity (HP) (Rubin *et al*., 2010).

Inter-populations statistics compare genomic data between two or more populations to identify regions with informative signatures. Statistics in these methods focus on differentiation between populations. According to the theory that most populations exhibit some degree of population structure, comparison of genomic data between populations can reveal regions that have been under selection in different populations. Inter-populations statistics can be classified in two groups:

1) **Single site differentiation:** Single site differentiation is the simplest and most popular group. The statistic used to detect local increases in population stratification under selection is F_{ST} (Wright, 1949). A more recent statistic based on single site differentiation is FLK (Bonhomme *et al*., 2010)

2) **Haplotype based differentiation analyses:** Ascertainment bias of SNP has lesser effect while using haplotype clusters rather than SNPs. Methods in this class use haplotype information in multiple population comparisons. The popular methods in this class is cross population extended haplotype homozygosity (XP-EHH) (Sabeti *et al*., 2007) and hapFLK, a haplotype based extension of the FLK statistic (Fariello *et al*., 2013).

CHAPTER –3

Materials & Methods

3. Materials and Methods

3.1 Experimental animals

The present study was carried out in 10 Sahiwal cows. Blood samples were collected from unrelated animals from the field (Divya Jyoti Jagrati Sansthan, Noormahal, Jalandhar , Punjab) and Livestock Research Centre, ICAR-NDRI, Karnal.

3.2 Isolation of genomic DNA

The genomic DNA was isolated from blood samples by Phenol: Chloroform: Isoamyl alcohol extraction method as described by Sambrook and Russell (2001) with modifications to suit our laboratory condition.

The protocol followed for isolation of genomic DNA was as follows:

1. The blood samples were thawed to room temperature and 10 ml of blood was transferred to oak ridge tubes.
2. The tubes were filled with chilled RBC lysis buffer (1x), mixed end to end and placed in ice for 10 min.
3. The tubes were centrifuged at 4000 rpm for 15-20 min at room temperature and the reddish tinged supernatant containing plasma and lysed RBC were discarded.
4. Steps 2-3 were repeated, till the WBC pellet appeared nearly white in colour.
5. About 3 ml of DNA extraction buffer) was added in the tubes, tapped to disperse the WBC pellet in the extraction buffer and kept in incubator at 37°C for 30 minutes.
6. SDS (10%) at 200 µl per 10 ml of blood was added and was gently mixed in the tubes by inverting.
7. Twenty micro litre of Proteinase K (20 mg Proteinase K /ml) was added to the tube and was incubated at 50°C overnight.
8. After 24 hours, the contents were transferred into a clean, sterile, autoclaved polypropylene tube, containing equal amount of Tris saturated phenol, the contents were gently mixed by inverting the tubes for 15 min.

9. Centrifugation was done at 4000 rpm for 15-20 min.
10. The upper aqueous phase containing DNA was transferred into fresh 15 ml, clean, sterile, autoclaved polypropylene tubes using 1ml tip.
11. Similar extractions were repeated (as in steps 8, 9 and 10) again with equal volume of phenol: chloroform: isoamyl alcohol (25: 24: 1) and then with chloroform: isoamyl alcohol (24: 1).
12. To the final aqueous phase, 3M sodium acetate (pH-5.2) at 100 µl/per ml of aqueous phase was added and was mixed gently.
13. Two volumes of ethanol (chilled) were added, and the tubes were gently mixed by inversion and kept at room temperature to allow precipitation of DNA.
14. DNA along with 500 µl of ethanol was transferred into fresh eppendroffs (1.5 ml) by means of wide bore tips and centrifuged in micro centrifuge at 10,000 rpm for 10 min.
15. The supernatant was discarded by gentle inversion where DNA pellet was intact, otherwise aspirated when DNA pellet was loose.
16. Washed the DNA pellet twice in 500µl of 70% ethanol i.e. 500µl of 70% ethanol was added and eppendroffs were centrifuged at 10000 rpm for 10 min at room temperature.
17. The DNA pellet was air dried after inverting on blotting paper so that last traces of ethanol were removed.
18. Approximately 200 µl of TE buffer was added and put in water bath at 60°C for 2 hr.
19. The eppendroffs were stored at 4°C for a week so that DNA was dissolved properly and subsequently shifted to -20°C temperature for further use.

3.3 Quality and quantity checking of DNA

The quality of the genomic DNA was checked by agarose gel electrophoresis. The quantity of DNA was estimated by the digital nano-photometer. Optical density (OD) was also determined as the ratio of OD_{260} and OD_{280} for each sample. Quantity of DNA was calculated using the following formula.

$$\text{Quantity of DNA (µg/ml)} = OD_{260} \times \text{Dilution Factor} \times 50\text{µg/ml}$$

3.4 Library preparation

The samples meeting the required QC parameters of more than 100ng/ul concentration and having a OD_{260} by OD_{280} ratio of 1.7-1.9 was considered for library preparation using the standard protocols described for ddRAD by Peterson *et al*. (2012).

Library preparation workflow:

1. Double digestion of genomic DNA (1 microgram) was done using *Sph*I and *MluC*I restriction enzymes.
2. The digested products were cleaned-up using Ampure beads.
3. P1 (Barcoded) and P2 adaptors were ligated using T_4 DNA ligase.
4. The ligated products were pooled and cleaned up
5. Size selection of the products were done after 2% agarose gel electrophoresis.
6. The PCR amplification was carried to enrich and add the Illumina specific adapters and flow cell annealing sequences to the size selected sequences
7. Quality control was carried out using bioanalyzer
8. The libraries were pooled for sequencing.

3.5 Sequencing

Next generation sequencing of the libraries was carried out using Illumina HiSeq 2000 sequencing platform. The ddRAD library preparation and sequencing was carried out by SciGenom Labs Pvt Ltd, Kochi, Kerala

3.6 Identification of genome wide SNPs

3.6.1 Read processing

The raw reads were demultiplexed using custom perl scripts to obtain sample specific reads. Up to one mismatch was allowed to demultiplex the sample data. The low quality bases and regions showing base bias at the start or end and was removed from the reads using Prinseq lite version 0.20.4 (Schmieder and Edwards, 2011). The Illumina 5′and 3′ adapter sequences were also removed. The processed reads were analysed using FastQC and Process radtags of stacks (Catchen *et al*., 2013) for finding the RAD loci.

3.6.2 Alignment

The processed reads were then aligned to the reference genome. Both *Bos taurus* (Bos_taurus_UMD_3.1.1) and *Bos indicus* (Bos_indicus_1.0) genome was used.

3.6.3 Variant calling

The aligned samples and the reference genome sequence were used for variant calling. Samtools/bcftools version 1.6 (Li *et al*., 2009) were used for variant calling. The variants identified were filtered for minimum read depth (at 2, 5 and 10) maximum read depth (D $\leq$500) and base quality score (Q $\geq$30). The work flow for SNP identification is given in Figure 3.1.

3.7 SNP annotation

The SNPs discovered were annotated both structurally and functionally with the help of SnpEFF (Cingolani *et al*., 2012). The missense SNPs were analysed using SIFT (Sorting Intolerant from Tolerant) tool (Kumar *et al*., 2009) to find their effect on protein structure and function.

3.8 Novel SNP identification

The SNPs obtained after the variant calling were compared with the cattle SNPs in dbSNP (build 150) of NCBI to identify the novel SNPs.

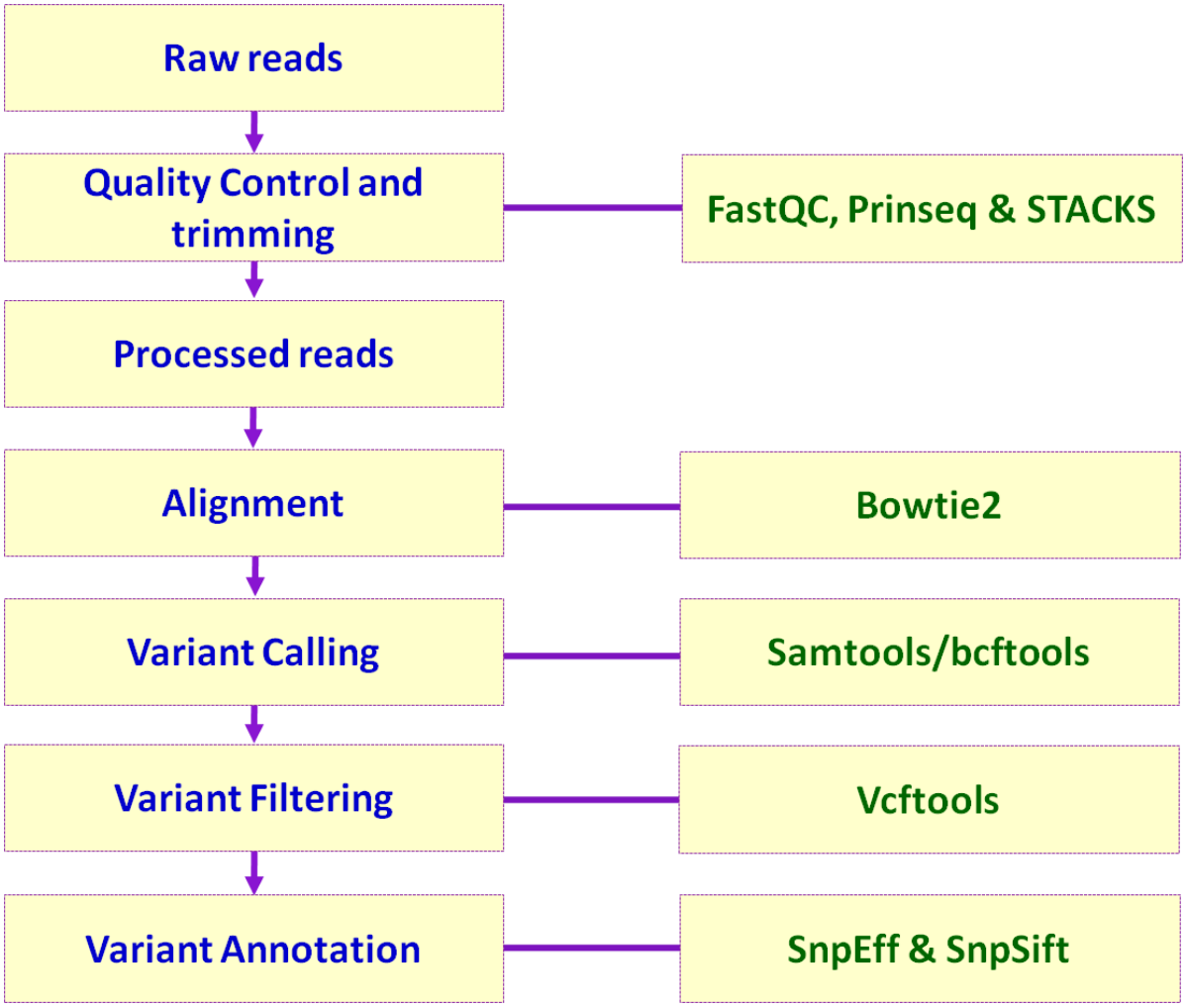
3.9 SNP validation

A set of 31 SNPs were selected so as to have at least one SNP per chromosome. Primers were designed using Primer3 software (http://www.primer3.ut.ee) (Untergrasser *et al*., 2012) for the target SNPs genotyping.

3.10 Mapping the SNPs to cattle QTLs

The SNPs identified were mapped to the cattle QTLs for production, reproduction, health and adaptation traits to find SNPs located in the QTL regions. The QTL information was downloaded from the cattle QTL database (http:// www. animalgenome. org/ cgi-bin/ QTLdb/ BT/ index).

Figure 3.1 Bioinformatics workflow for SNP identification and annotation



3.11 Genotyping efficiency

The genotyping efficiency of the ddRAD was studied by finding the number of SNP loci that got genotyped in 50% and 100% of samples at different read depths. The analysis was done using vcftools v 0.1.15.

3.12 Identification of genome wide SSRs

As the ddRAD sequencing reads were short and numerous, the μstacks script in Stacks v1.39 was used to construct consensus sequences for each sample. These consensus sequences were constructed from aligning a set of short-read sequences obtained from ddRAD sequencing into exactly-matching stacks.

Then, the pipe1.pl scripts in QDD_v3 (Meglecz *et al*., 2010) was used to identify the sequences containing microsatellites with di to hexa nucleotide motifs. The minimum repeat number for motifs to be identified as microsatellite was set to 5. Only simple/perfect SSR were selected. The sequences of the ten individuals with microsatellites were aligned using the QDD pipe2.pl script. This was done to identify polymorphic microsatellite loci among the ten samples. The pipe3.pl scripts in QDD_v3 utilizing the program Primer v3 was used to design SSR primers (http://bioinfo.ut.ee/primer3-0.4.0/). The expected products ranged from 90 to 300 bp. The primer size ranged from 18 to 23 bp with the optimal size of 20 bp. The optimum GC content was 50%, and the optimum melting temperature was set to 60°C. The SSR identified were annotated to different chromosomes using blastn tool of NCBI. The bioinformatics work flow for SSR identification and annotation is given in Figure 3.2

3.13 Comparison of ddRAD SNPs with SNP chip data

The identified SNPs from the ddRAD sequencing was compared with SNPs in the Bovine SNP chip (Illumina BovineSNP50 BeadChip and Illumina BovineHD BeadChip) data with respect to total number of SNPs and number of SNPs per chromosome.

3.14 Comparison ddRAD sequencing with whole genome sequence

The ddRAD sequencing reads were aligned with the cattle reference genome to find the genome coverage. The analysis was done using both reference genomes and for each sample separately

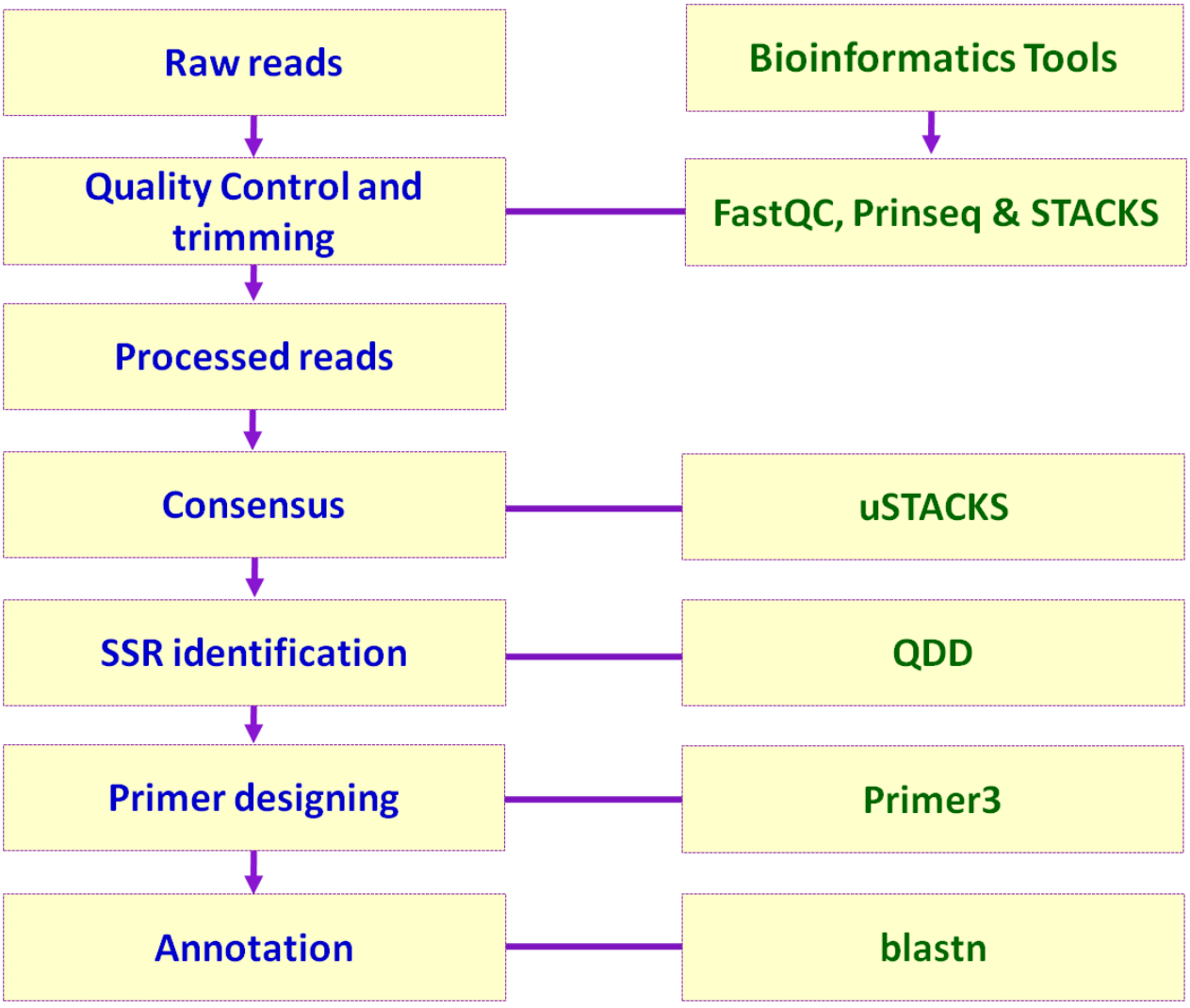
3.15 Detection of Selection sweeps

The selection sweeps were detected using SweeD 3.0 (Pavlidis *et al*., 2013) using the selective sweep model as described by Nielsen *et al.* (2005). The grid value was taken as 1000000. The window size along each chromosome ranged from ~40-150. SweeD works on a composite likelihood ratio test which identifies complete selective sweeps using Site frequency Spectrum patterns of SNPs in the genome. Hypothesis testing is done using a likelihood ratio test that compares the model (assuming neutrality) with alternative models. Higher CLR values denote a candidate region for selective sweep. The top 1% of the CLR values was taken as putative sweeps.

3.15.1 Annotation of Selection sweeps

The genes occurring in the selective sweep regions were identified by comparing it with the information provided by the UMD3.1 bovine genome assembly (Zimin *et al*. 2009). The genes identified in the selective sweep regions were submitted to DAVID (Database For Annotation, Visualization and Integrated Discovery) for gene enrichment analysis (Dennis *et al*., 2003). The clustering was done based on GO (gene ontology) BO (biological process) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Ogata *et al*., 1999; Ashburner *et al*. 2000).

Figure 2.2 Bioinformatics workflow for SSR identification and annotation



CHAPTER –4

Results and Discussion

4. Results and Discussion

4.1 Genome-wide variant identification in Sahiwal cattle

4.1.1 SNP identification using *Bos taurus* as reference sequence

In Sahiwal cattle, 450431 genome wide SNPs were identified with reference to *Bos taurus* genome at a minimum read depth of 2 (Figure 4.1). The number of SNPs reduced to 399711 and 324373 with read depth increasing to 5 and 10, respectively. Total 33085 indels were identified in Sahiwal genome with a minimum read depth of 2, while 29460 and 23595 indels were found with the read depth set to 5 and 10, respectively. The number of SNPs identified on bovine autosomes were highly correlated (Pearson's correlation coefficient = 0.992, $P \leq 0.001$) with the chromosome length (Figure 4.2 and Figure 4.3). The chromosome 1 was found to have the maximum number of SNPs and indels. Chromosome wise distribution of SNPs and indels at different read depths have been given in table 4.1

The number of SNPs identified in the present study was in the accordance with previous reports. De Donato *et al*. (2013) identified 63697 SNPs in 47 animals from both taurine and indicine breeds using single enzyme GBS method. Brouard *et al.* (2017) using a double enzyme GBS method identified 272103 SNPs in 48 dairy cows. Malik *et al*. (2018) identified 107488 SNPs by GBS approach in 24 animals belonging to seven Indian cattle breeds and high correlation between the length of the chromosome and the number of SNP was observed. Using the same method Gurgul *et al*. (2019) reported 8065 high confidence SNPs in 48 individuals of different cattle breeds. Yang *et al.* (2018) found 10058 SNPs in 40 dairy cows using a modified RAD sequencing method. Wang *et al*. (2018) identified 238725 high confidence SNPs in indigenous cattle breeds of China using RAD sequencing approach.

4.1.2 SNP identification using *Bos indicus* as reference sequence

Total 258231 genome wide SNPs were identified in Sahiwal cattle with reference to *Bos indicus* reference genome with a minimum read depth of 2 (Figure 4.4). At read depth 5 and 10, 232570 and 193803 SNPs were identified, respectively. Also 29868 indels were identified in Sahiwal genome with a minimum read depth of 2, while 26463 and 21232 indels were found with the read depth set to 5 and 10, respectively. The number of SNPs identified on bovine autosomes was highly correlated (Pearson's correlation coefficient = 0.982, $P \leq 0.001$) with the

chromosomes' length (Figure 4.5 and Figure 4.6). There are no previous reports available on SNP identification using *Bos indicus* reference genome to compare the present findings. Chromosome wise distribution of SNPs and indels at different read depths are given in table 4.2.

Table 4.1 Chromosome-wise variants identified in Sahiwal with *Bos taurus* (Bos_taurus_UMD_3.1.1) as reference genome

Chr.	Unfiltered		Read Depth 2		Read Depth 5		Read Depth 10	
	SNP	INDEL	SNP	INDEL	SNP	INDEL	SNP	INDEL
1	59222	4004	26743	1974	23663	1763	19171	1415
2	46513	3404	22401	1760	19808	1553	15840	1249
3	39271	2796	19430	1431	17279	1264	13978	1011
4	43856	3173	20693	1645	18183	1453	14533	1189
5	41178	2841	18859	1386	16672	1224	13563	989
6	42697	3085	20306	1550	17870	1367	14452	1096
7	39753	2633	18201	1320	16197	1197	13164	949
8	41411	2875	18811	1432	16691	1289	13512	1016
9	36978	2618	17505	1325	15437	1177	12445	991
10	36810	2579	17472	1340	15560	1208	12655	991
11	37537	2672	18707	1325	16688	1195	13529	964
12	35437	2319	16245	1169	14358	1029	11576	806
13	28199	2021	14663	1042	13131	919	10828	730
14	31011	2108	14586	1063	13031	950	10659	754
15	32077	2086	15115	1020	13399	893	10845	706
16	29237	2071	14359	1066	12891	953	10481	756
17	27440	1964	13535	1003	12064	904	9862	719
18	22376	1564	11611	867	10382	790	8685	644
19	19958	1440	10824	805	9697	728	8020	577
20	27888	1886	14136	1003	12535	882	10011	692
21	26044	1676	13233	854	11783	774	9580	622
22	20758	1405	11084	721	9929	646	8116	529
23	19613	1475	10047	773	8917	683	7271	543
24	24216	1624	11948	880	10559	778	8644	625
25	14994	1074	8559	581	7705	521	6397	433
26	19414	1351	9870	735	8807	667	7213	542
27	20782	1250	9418	627	8251	563	6685	459
28	18550	1289	9356	696	8327	618	6748	516
29	22262	1362	10080	697	9010	612	7341	508
X	34298	2274	12591	994	10859	860	8562	641
Total	**939781**	**64919**	**450388**	**33084**	**399683**	**29460**	**324366**	**23595**

Figure 4.1 Genome-wide distribution of SNPs in Sahiwal cattle

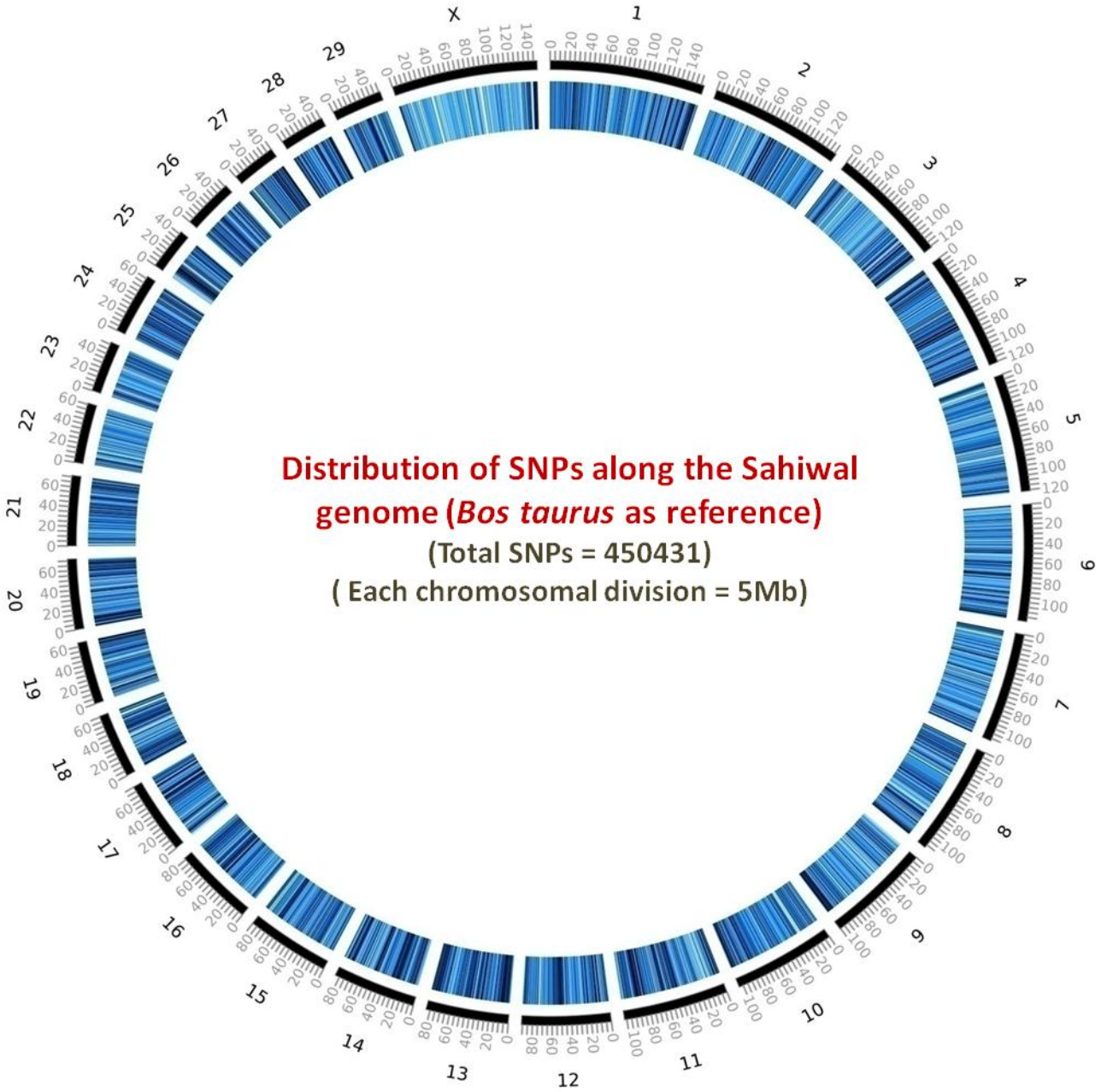


Figure 4.2 Correlation between chromosome length and number of ddRAD SNPs in Sahiwal cattle (*Bos taurus* as reference)

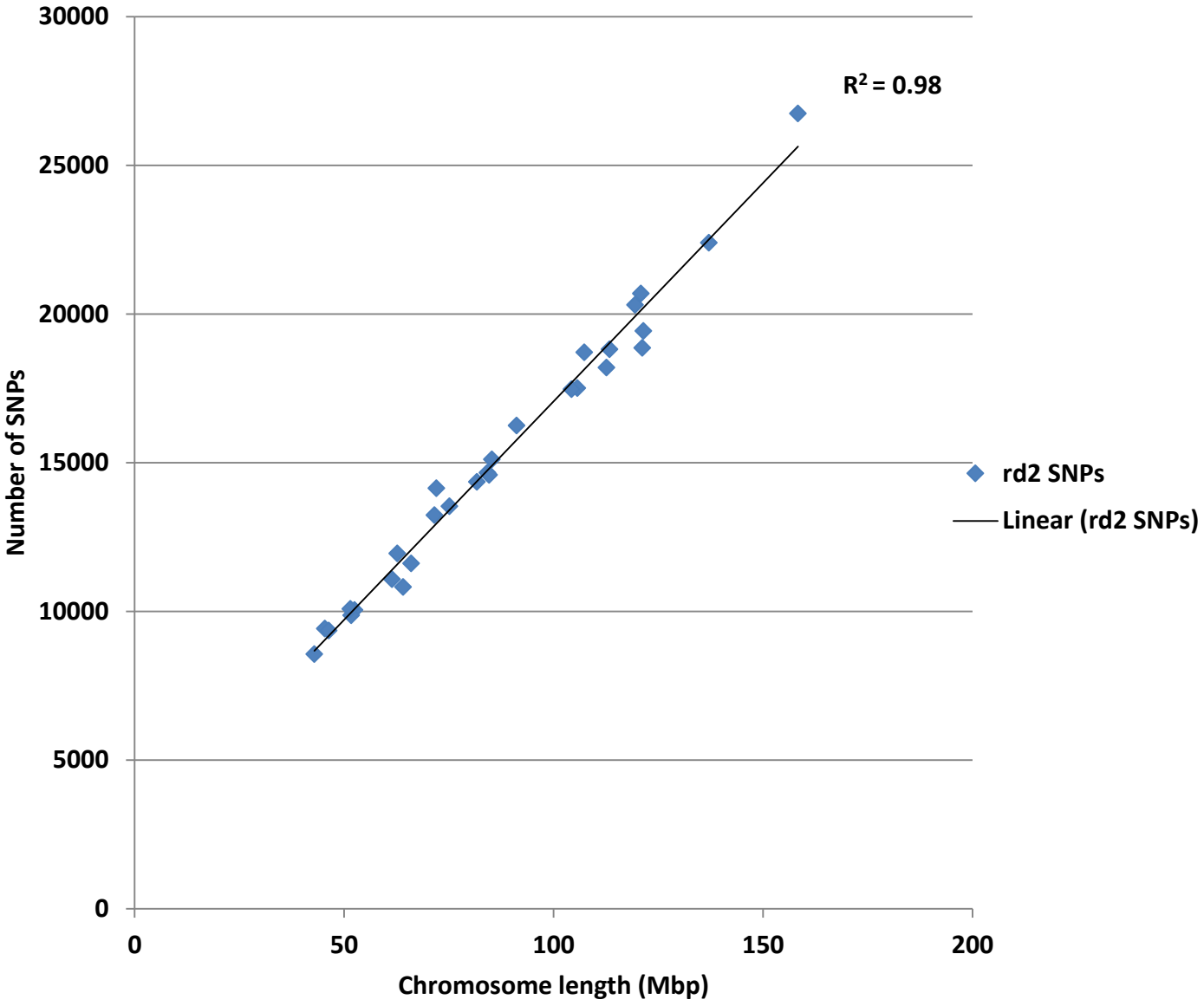


Figure 4.3 Chromosome wise distribution of Novel SNPs in Sahiwal (*Bos taurus* as reference)

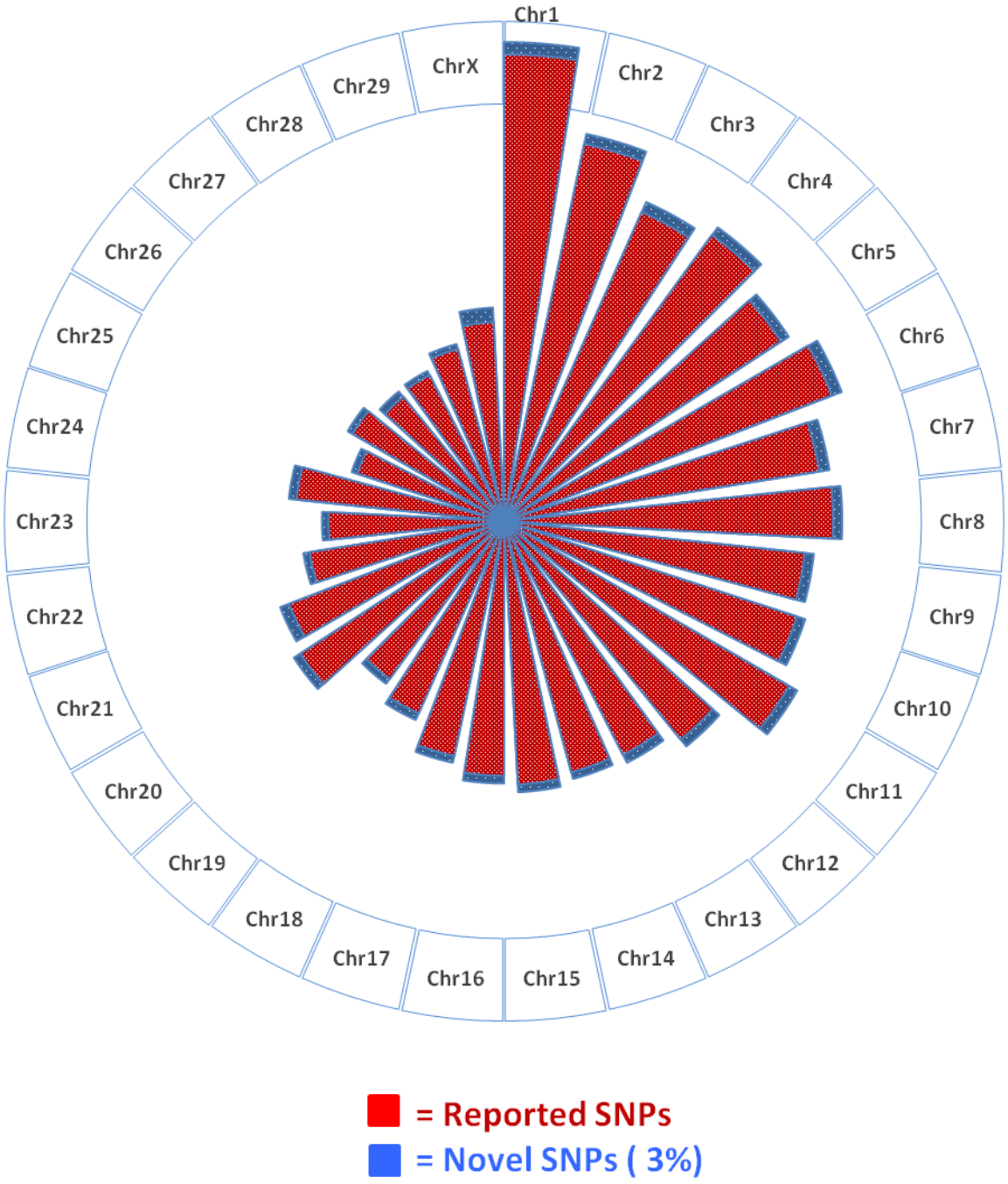


Figure 4.4 Genome-wide distribution of SNPs in Sahiwal cattle (*Bos indicus* as reference)

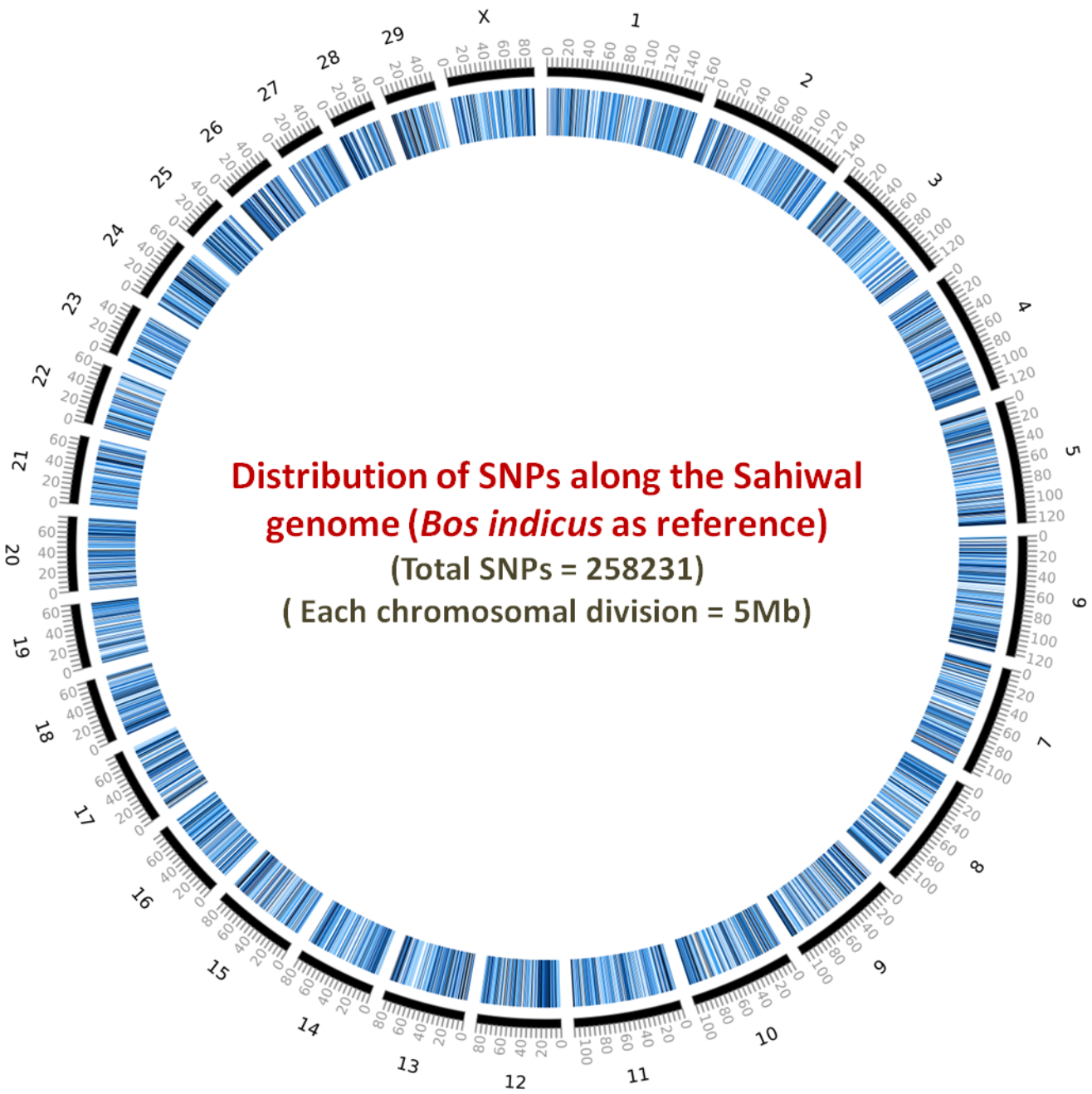


Table 4.2 Chromosome-wise variants identified in Sahiwal with *Bos indicus* (Bos_indicus_1.0) as reference genome

Chr.	Unfiltered		Read Depth 2		Read Depth 5		Read Depth 10	
	SNP	INDEL	SNP	INDEL	SNP	INDEL	SNP	INDEL
1	39598	3526	14851	1759	13346	1549	11089	1258
2	33463	3067	12602	1558	11357	1381	9345	1106
3	29846	2650	11411	1330	10291	1181	8566	948
4	31136	2910	12135	1509	10797	1334	8903	1078
5	29297	2615	10912	1291	9781	1131	8153	909
6	31492	2715	12147	1375	10935	1218	8969	963
7	27460	2355	11092	1191	10021	1078	8311	849
8	27768	2633	10688	1344	9652	1180	8046	950
9	26502	2350	10383	1187	9332	1052	7655	826
10	25904	2449	9806	1261	8851	1117	7387	904
11	26243	2450	10770	1272	9701	1136	8115	915
12	25523	2037	10227	1047	9180	926	7669	727
13	19450	1867	7929	966	7238	848	6084	678
14	20312	1835	7973	971	7238	858	6125	675
15	21632	1815	8591	901	7693	788	6487	623
16	19953	1802	8155	938	7356	840	6087	670
17	18535	1736	7560	926	6859	822	5762	659
18	15276	1439	6799	789	6116	709	5190	585
19	14338	1339	6129	749	5628	672	4701	539
20	20311	1733	8204	949	7388	833	6088	644
21	17635	1470	7387	746	6646	670	5549	544
22	14689	1274	6641	656	6023	583	5048	478
23	13618	1330	5753	695	5164	610	4340	491
24	17031	1520	7140	804	6396	704	5349	568
25	11145	1036	4997	534	4498	475	3850	388
26	14642	1221	5563	648	5026	574	4208	471
27	13323	1120	5498	594	4910	529	4087	432
28	12528	1167	5368	622	4783	560	3967	467
29	13385	1206	5638	646	5140	572	4323	468
X	21237	1386	5882	610	5224	533	4350	419
Total	**653272**	**58053**	**258231**	**29868**	**232570**	**26463**	**193803**	**21232**

4.1.3 Identification of novel SNPs

Novel SNPs were identified by comparing the ddRAD SNPs with those existing in public database (NCBI dbSNP build 150). The novel SNPs 14908, 11570 and 9828 were identified with respect to *Bos taurus* reference at read depth 2, 5 and 10, respectively. At read depth 2, 5 and 10 the number of identified novel SNPs were 150231, 134502 and 12443, respectively in Sahiwal cattle with reference to *Bos indicus* genome. Chromosome wise distribution of novel SNPs at different read depths compared to different reference genomes are given in the table 4.3

Gurgul *et al*. (2019) found 758 novel SNPs (9.4% of total SNPs identified) in taurine cattle using a single enzyme GBS method. Another study (Wang *et al*., 2018) on taurine cattle using single enzyme RAD method identified 103574 novel SNPs (43.4% of total SNPs identified).

The NCBI dbSNP (build 150) contains 104.3 million *Bos taurus* and 17.8 million *Bos indicus* reference SNPs. The large number of novel SNPs identified in *Bos indicus* as reference genome might be due to the recent inclusion of *Bos indicus* in SNP databases leading to low number of reported SNPs.

4.2 Genotyping efficiency

The genotyping efficiency of ddRAD sequencing was studied by identifying the number of SNPs getting genotyped in 50% and 100% of samples. The analysis was done separately for the different read depths of 2, 5 and 10 and also for the two reference genomes.

4.2.1 Genotyping efficiency of SNPs identified with *Bos taurus* as reference genome

In 50% of samples of Sahiwal cattle 314344, 314344 and 293394 SNPs were genotyped at read depth 2, 5 and 10, respectively. While the number of SNPs genotyped in all the 10 samples remained same (82648) regardless of read depth criteria (Table 4.4). About 90.44% and 25.40% of the high confidence SNPs identified were genotyped in 50% and 100% of the samples respectively. The sample wise genotyping efficiency is given in figure 4.7. These results showed that with other filtering parameters set at optimum levels, the read depths did not make any significant difference in genotyping efficiency.

Figure 4.5 Correlation between chromosome length and number of ddRAD SNPs in Sahiwal cattle (*Bos indicus* as reference)

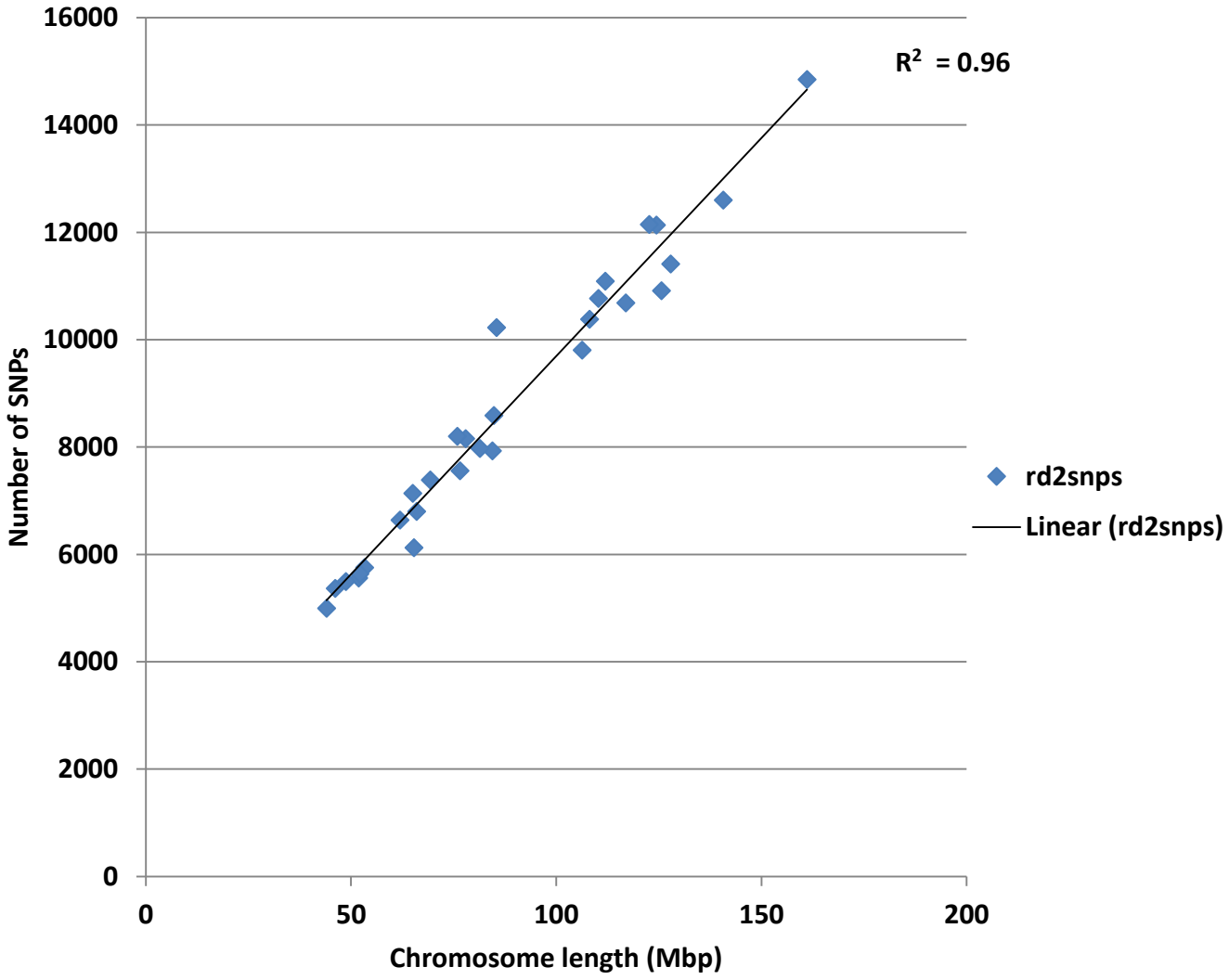


Figure 4.6 Chromosome wise distribution of Novel SNPs in Sahiwal (*Bos indicus* as reference)

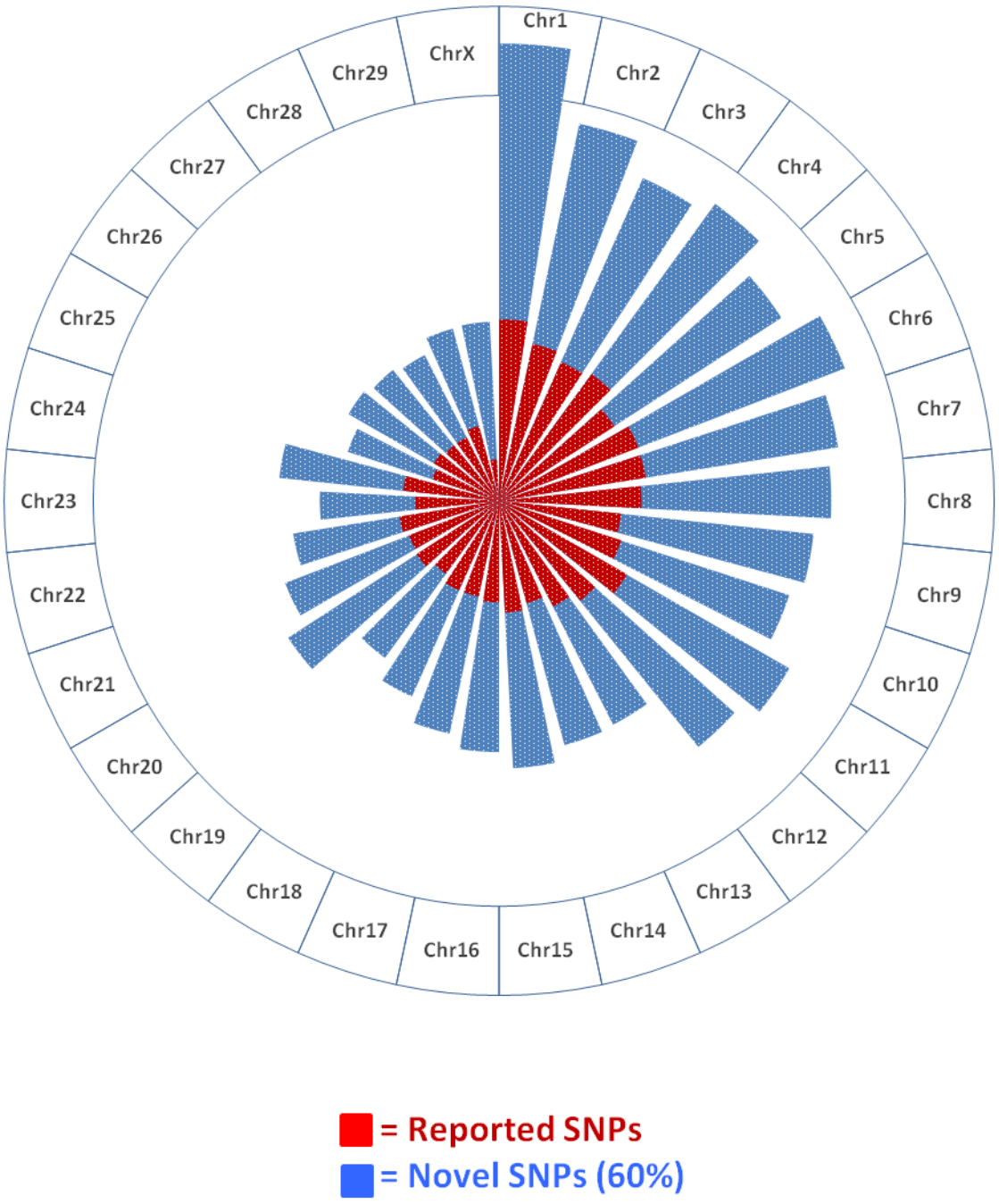


Figure 4.7 Genotyping efficiency of SNPs identified in Sahiwal cattle with *Bos taurus* as reference genome

Table 4.3 Novel SNPs identified in Sahiwal with *Bos taurus* (Bos_taurus_UMD_3.1.1) as reference genome

Chromosome	Novel SNPs at different Read Depths		
	Read Depth 2	**Read Depth 5**	**Read Depth 10**
1	827	615	505
2	626	512	417
3	583	499	443
4	634	471	383
5	627	443	356
6	646	480	407
7	592	457	395
8	580	454	342
9	558	403	350
10	519	381	333
11	597	440	365
12	544	414	354
13	466	382	351
14	407	315	276
15	489	377	314
16	425	345	310
17	451	357	277
18	412	314	283
19	301	254	231
20	494	351	290
21	496	404	364
22	353	275	237
23	346	275	242
24	398	320	290
25	294	242	220
26	318	256	220
27	448	335	278
28	273	223	187
29	304	257	212
X	900	719	591
Total	**14908**	**11570**	**9828**

Table 4.4 Novel SNPs identified in Sahiwal with *Bos indicus* (Bos_indicus_1.0) as reference genome

Chromosome	Novel SNPs at different Read Depths		
	Read Depth 2	**Read Depth 5**	**Read Depth 10**
1	8913	7969	6686
2	7407	6629	5455
3	6535	5845	4878
4	7029	6207	5102
5	6541	5834	4833
6	7277	6537	5366
7	6337	5692	4727
8	6140	5515	4606
9	6322	5613	4694
10	5671	5094	4287
11	6057	5410	4552
12	6014	5401	4552
13	4218	3813	3205
14	4599	4157	3557
15	5055	4488	3773
16	4849	4352	3625
17	4419	4001	3390
18	3655	3270	2815
19	3324	3021	2515
20	5049	4530	3713
21	4282	3808	3217
22	3480	3132	2616
23	3111	2778	2318
24	4075	3627	3023
25	2804	2498	2157
26	3174	2871	2394
27	3171	2814	2359
28	3018	2678	2239
29	3199	2911	2445
X	4506	4007	3344
Total	**150231**	**134502**	**112443**

DeDonato *et al*. (2013) reported 63.9% (40725) of the SNPs identified by GBS method in cattle were genotyped in ≥90% of the samples. Wang *et al.* (2018) identified that 45.32% (240924) of the SNPs identified in the 55 Chinese cattle using RAD sequencing was genotyped in 100% of the samples. Single enzyme based GBS method has found 95.4% (7694) of the high quality SNPs passing stringent filtering parameters were genotyped in 100% of samples (Gurgul *et al*., 2019). The lower percentage of SNPs genotyped in 100% of the samples in the present study might be because of use of different restriction enzymes and filtering parameters optimized for maximum variant identification.

4.2.2 Genotyping efficiency of SNPs identified with *Bos indicus* as reference genome

The genotyping efficiency analysis with respect to *Bos indicus* genome yielded similar results as that of *Bos taurus*. Total 188653, 188653 and 177767 SNPs were genotyped in 50% of samples at read depths of 2, 5 and 10, respectively; whereas, 52841 SNPs were genotyped in 100% of the samples irrespective of the read depth level (Table 4.6). Among the high confidence SNPs identified 91.86% and 27.30% were genotyped in 50% and 100% of the samples, respectively. The sample wise genotyping efficiency has been given in figure 4.8.

There were no previous reports available on reduced representation approaches using *Bos indicus* as reference genome to compare and contrast the present findings.

4.3 SNP annotation

The structural and functional annotation of the identified SNPs was carried out for both *Bos taurus* and *Bos indicus* as reference genomes. The SNP effects were classified based on their functions, impact and genomic region (Tables 4.7 to 4.9). The resulting nucleotide changes and amino acid changes were also identified and presented in tables 4.10 to 4.13.

The silent mutations were most frequent among the functional class. The missense to silent ratio was 0.52 (*Bos taurus* genome) and 0.62 (*Bos indicus* genome). The SNPs classification based on their impact revealed maximum SNPs in modifier class followed by low impact class. Among the genomic regions maximum SNPs were found in intergenic regions, consequently, more number of SNPs were found in the modifier class.

Table 4.5 Genotyping efficiency of identified SNPs (*Bos taurus* as reference genome)

Chromosome	Read Depth 2		Read Depth 5		Read Depth 10	
	Number of Samples (%)					
	50	100	50	100	50	100
1	18516	4538	18516	4538	17251	4538
2	15394	4021	15394	4021	14299	4021
3	13578	3674	13578	3674	12637	3674
4	14119	3644	14119	3644	13092	3644
5	13171	3378	13171	3378	12321	3378
6	14013	3398	14013	3398	13013	3398
7	12623	3192	12623	3192	11779	3192
8	13098	3409	13098	3409	12216	3409
9	11943	3095	11943	3095	11117	3095
10	12411	3128	12411	3128	11557	3128
11	13155	3477	13155	3477	12285	3477
12	11153	2756	11153	2756	10418	2756
13	10568	2955	10568	2955	9866	2955
14	10388	2971	10388	2971	9658	2971
15	10419	2922	10419	2922	9728	2922
16	10274	2756	10274	2756	9629	2756
17	9568	2541	9568	2541	8975	2541
18	8428	2466	8428	2466	7933	2466
19	7718	2184	7718	2184	7296	2184
20	9820	2422	9820	2422	9086	2422
21	9276	2625	9276	2625	8686	2625
22	7826	2074	7826	2074	7357	2074
23	7011	1954	7011	1954	6556	1954
24	8257	2201	8257	2201	7793	2201
25	6212	1728	6212	1728	5849	1728
26	6966	1769	6966	1769	6539	1769
27	6491	1653	6491	1653	6020	1653
28	6486	1774	6486	1774	6085	1774
29	7068	1945	7068	1945	6625	1945
X	8386	1998	8386	1998	7727	1998
Total	**314336**	**82648**	**314336**	**82648**	**293390**	**82648**

Table 4.6 Genotyping efficiency of identified SNPs (*Bos indicus* as reference genome)

Chromosome	Read Depth 2		Read Depth 5		Read Depth 10	
	Number of Samples (%)					
	50	100	50	100	50	100
1	10751	2944	10751	2944	10129	2944
2	9085	2562	9085	2562	8561	2562
3	8289	2420	8289	2420	7840	2420
4	8754	2425	8754	2425	8200	2425
5	7950	2200	7950	2200	7482	2200
6	8760	2216	8760	2216	8198	2216
7	7973	2138	7973	2138	7504	2138
8	7811	2095	7811	2095	7345	2095
9	7469	2033	7469	2033	7000	2033
10	7308	1980	7308	1980	6854	1980
11	7887	2178	7887	2178	7424	2178
12	7417	2010	7417	2010	7051	2010
13	5981	1790	5981	1790	5648	1790
14	5991	1844	5991	1844	5622	1844
15	6236	1923	6236	1923	5912	1923
16	6024	1698	6024	1698	5670	1698
17	5587	1532	5587	1532	5294	1532
18	5065	1594	5065	1594	4788	1594
19	4562	1338	4562	1338	4301	1338
20	5922	1589	5922	1589	5559	1589
21	5405	1585	5405	1585	5079	1585
22	4902	1365	4902	1365	4649	1365
23	4207	1209	4207	1209	3986	1209
24	5180	1432	5180	1432	4896	1432
25	3731	1076	3731	1076	3547	1076
26	4107	1036	4107	1036	3890	1036
27	4010	1076	4010	1076	3745	1076
28	3854	1082	3854	1082	3670	1082
29	4188	1258	4188	1258	3956	1258
X	4247	1213	4247	1213	3967	1213
Total	**188653**	**52841**	**188653**	**52841**	**177771**	**52841**

Table 4.7 Effects of SNP by functional class

Type	Number of SNPs	
	Bos taurus	***Bos indicus***
Missense	493	582
Nonsense	3	5
Silent	954	936
Missense/Silent	0.52	0.62

Table 4.8 Effects of SNP by impact

Type of Impact	Number of SNPs	
	Bos taurus	***Bos indicus***
High	12	15
Low	1119	1213
Moderate	493	579
Modifier	368574	478466

Table 4.9 Effect of SNPs by genomic region

Genomic region	Number of SNPs	
	Bos taurus	***Bos indicus***
Exon	1530	2153
Intergenic	242579	128819
Intron	97950	146115
Splice Site Acceptor	6	3
Splice Site Donor	3	4
Splice Site Region	173	222
Upstream	13467	25815
5′UTR	113	527
3′UTR	472	2080

The transition to transversion (Ts/Tv) ratio was calculated as an indicator of potential random sequencing errors. With *Bos taurus* as reference genome 1612452 transitions and 652162 transversions were identified in Sahiwal with a Ts/Tv ratio of 2.47. Similar results were obtained with *Bos indicus* reference genome with Ts/Tv ratio of 2.36 comprising of 712399 transitions and 302171 transversions. The above Ts/Tv ratios indicated higher accuracy of the SNPs identified in the present study.

The number of A/T and T/A substitutions was relatively lower than other types of substitutions (Table 4.10-4.11). This could be attributed to the fact that deeper sequencing was needed to discover SNPs in AT-rich genomic regions (Dohm *et al*., 2008).

These results are supported by previous RADseq study on cattle, which identified 163266 transitions and 75459 transversions, with an overall Ts/Tv ratio of 2.16 (Wang *et al*., 2018). Surya *et al.* (2018) reported 2.45 Ts/Tv ratio in buffaloes using ddRAD sequencing. Studies using whole genome sequencing for variant identification in cattle have reported Ts/Tv ratios of approximately 2.2 in *Bos taurus, Bos indicus* and their crossbreds (Stothard *et al*., 2011; Choi *et al*., 2014; Choi *et al*., 2015; Stafuzza *et al*., 2017). Higher Ts/Tv ratios were also observed in other reduced representation approaches for genome-wide SNP discovery (Kraus *et al*., 2011; Le and Durbin 2011; Ba *et al.,* 2017).

4.3.1 SIFT Analysis

The SIFT (Sorting Intolerant from Tolerant) predicts whether, an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. The mutations were found to be deleterious when the SIFT score ranges from 0.0 to <0.05 and considered to be tolerated when the score ranged 0.05 to 1.0 (Kumar *et al*., 2009). The SIFT analysis was carried out on the 443 non-synonymous SNPs identified. The SIFT scores were not available for 15 SNPs due the lower number of homologous sequences to compare. Out of 443 non-synonymous SNPs, 339 SNPs were found to be tolerated, with changed amino acid but not affecting the protein structure and function. On the other hand SIFT analysis predicted that remaining 89 SNPs were found to having deleterious effect on protein structure and function.

Table 4.10 Nucleotide changes in Sahiwal cattle with respect to *Bos taurus* reference genome

	A	C	G	T
A	0	12904	54202	8565
C	13274	0	13454	63103
G	63233	13436	0	13429
T	8729	54102	12827	0

Table 4.11 Nucleotide changes in Sahiwal cattle with respect to *Bos indicus* reference genome

	A	C	G	T
A	0	7834	30664	5698
C	8304	0	8110	39699
G	39644	8043	0	8457
T	5424	30282	7472	0

Table 4.12 Amino acid changes due to non synonymous SNPs in Sahiwal cattle (*Bos indicus* as reference genome)

	*	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
*																					
A		110		1	5		5							5			4	28	16		
C	1		12			1	9										2				2
D				43	2		2	1		12			9								
E					19		4				12				1						
F			2			26											3				1
G		2		1	8		73									4	3			1	
H								14					3		8	8					4
I						4			35	1	1	8	1				5	5	13		
K					3					13		1	4		2	10					
L						11			3		94	1		9							
M									3		2										
N										1			40								
P											12			111							
Q	1				1			2		2	3				27						
R	1		23				7	21	1	14		1			16	51				9	
S			1			2	3				4		14	3		1	92	2			
T		9							6	4		9	3	5		1	8	97			
V		23		1	1				18		2	7							25		
W	1		1													5					
Y	1					1		1													33

Table 4.13 Amino acid changes due to non synonymous SNPs in Sahiwal cattle (*Bos taurus* as reference genome)

	*	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
*																					
A		110		1	5		5							5			4	28	16		
C	1		12			1	9										2				2
D				43	2		2	1		12			9								
E					19		4				12				1						
F			2			26											3				1
G		2		1	8		73									4	3			1	
H								14					3		8	8					4
I						4			35	1	1	8	1				5	5	13		
K					3					13		1	4		2	10					
L						11			3		94	1		9							
M									3		2										
N										1			40								
P											12			111							
Q	1				1			2		2	3				27						
R	1		23				7	21	1	14		1			16	51				9	
S			1			2	3				4		14	3		1	92	2			
T		9							6	4		9	3	5		1	8	97			
V		23		1	1				18		2	7							25		
W	1		1													5					
Y	1					1		1													33

4.4 Validation of SNPs

Out of the set of 31 SNPs selected for validation, 25 SNPs were successfully amplified and Sanger sequenced. Heterozygous loci were detected in 22 loci out of the 25 sequenced (Figure 4.9). The details of the validated SNPs are summarized in table 4.14.

Table 4.14 Details of the SNPs validated by sanger sequencing

Chromosome		Position	Alleles
1	1	69513042	G/A
2	2	44729967	G/A
3	3	251534	C/T
4	5	80484794	A/G
5	7	3340499	T/G
6	9	34404747	T/C
7	11	7045033	A/G
8	12	79722429	A/T
9	13	31754433	T/C
10	15	78617028	T/C
11	15	54218636	T/C
12	17	40841339	G/C
13	18	24906224	T/G
14	20	6840788	G/A
15	21	6730554	A/G
16	22	963071	T/G
17	23	22136851	T/C
18	24	7927208	T/C
19	26	35436616	T/G
20	27	4437971	T/C
21	28	13324730	T/A
22	29	6840788	T/C

4.5 Mapping of SNP to various QTLs in sahiwal cattle

Mapping of the identified SNPs to the cattle QTLs retrieved from CattleQTL database found 87904 SNPs in different QTLs. Trait-wise QTLs have been discussed in the following sections:

4.5.1 Production trait QTLs

Total 22762 SNPs were mapped to production trait QTLs (Table 4.15). The chromosome wise distribution of SNPs in QTLs for milk yield, milk protein and milk fat have been presented in figures 4.10-4.12. All chromosomes except chromosome 12 exhibited SNPs for milk yield QTLs. Chromosome 3, 5, 6, 10, 14, 20 and 26 were having higher density of SNPs in milk yield QTL regions.

The chromosomes 5, 14 and 26 were found to have more SNPs for both milk fat yield and percentage QTLs. The chromosome 17 and 22 did not exhibited any SNPs for milk fat QTLs. SNPs in milk protein QTLs were mostly presented on chromosome 6, 10 and 23. The chromosomes 2 and 26 were found to have SNPs in lactation persistency QTLs.

Candidate genes for milk yield and composition like ANXA9 (Annexin 9 Protein) and SLC27A3 (Fatty Acid Transport Protein Type 3) in chromosome 3; OLR1 (Oxidized Low Density Lipoprotein Receptor 1), ATF4 (Activating Transcription Factor 4), SOCS2 (Suppressor Of Cytokine Signaling 2) and WNT10B (Wnt Family Member 10B) in chromosome 5; PPARGC1A (PPARG Coactivator 1 Alpha), AREGB (Amphiregulin), CSN2 (Casein Beta 2) and ABCG2 (ATP Binding Cassette Subfamily G Member 2) in chromosome 6; DGAT1 (Diacylglycerol O-Acyltransferase 1) and PTK2 (Protein Tyrosine Kinase 2) in chromosome 14; FASN (Fatty Acid Synthase), STAT5A (Signal Transducer And Activator Of Transcription 5A), STAT5B (Signal Transducer And Activator Of Transcription 5B) and STAT3 (Signal Transducer And Activator Of Transcription 3) in chromosome 19; GHR (Growth Hormone Receptor), LIFR (LIF Receptor Alpha), OSMR (Oncostatin M Receptor) and PRLR (Prolactin Receptor) in chromosome 20; PRL (Prolactin) in chromosome 23; SCD (Stearoyl-Coa Desaturase) and DKK1 (Dickkopf WNT Signaling Pathway Inhibitor 1) in BTA26 might be contributing to the higher density of variants in these chromosomes (Grisart *et al*., 2002; Brym *et al*., 2004; Cohen-Zinder *et al*., 2005; Calvo *et al*., 2006; Morris *et al*., 2007; Schennink *et al*., 2009; Kgwatalala *et al*., 2009; Raven *et al*., 2014).

Figure 4.8 Genotyping efficiency of SNPs identified in Sahiwal cattle with *Bos indicus* as reference genome

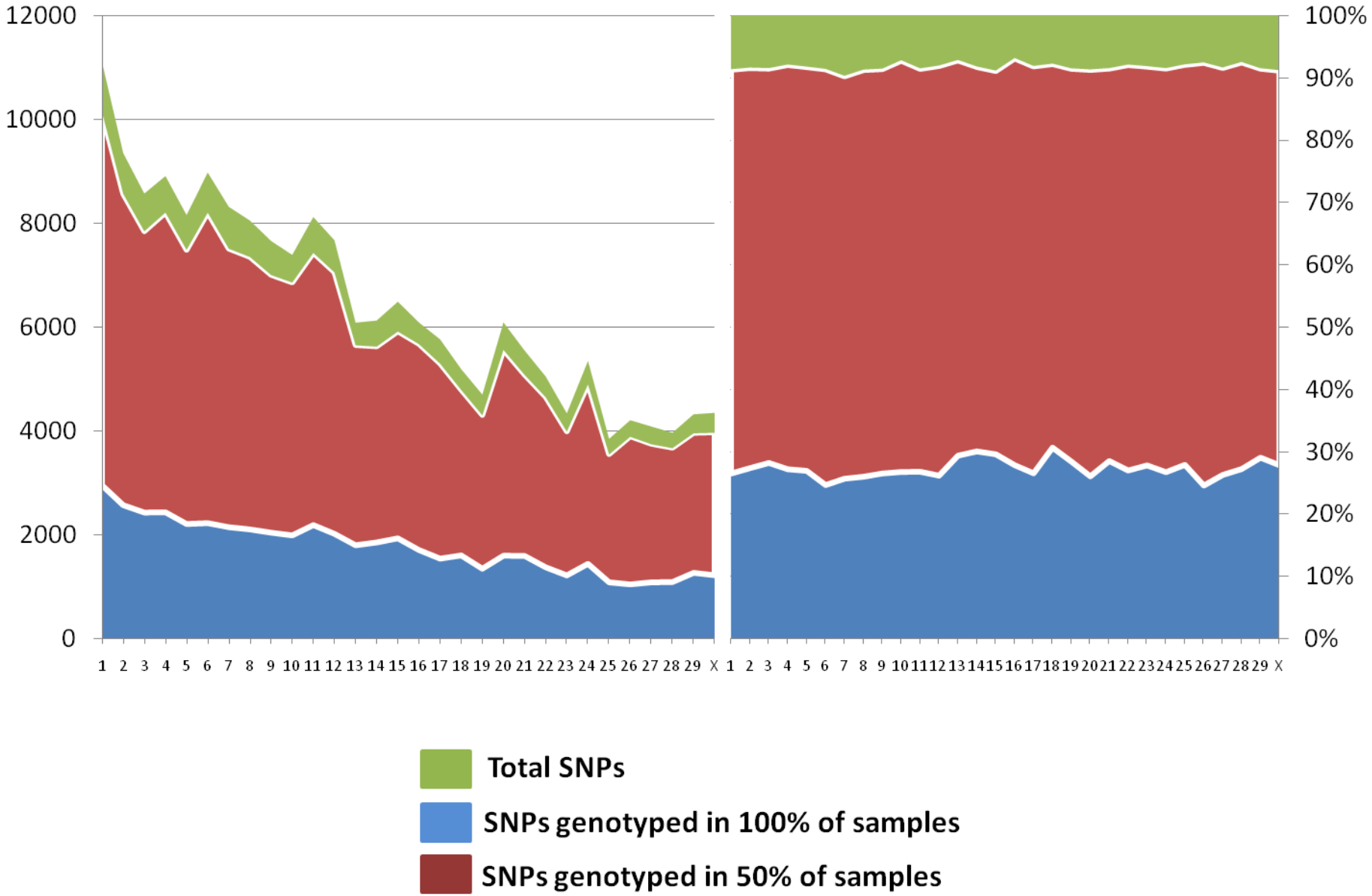
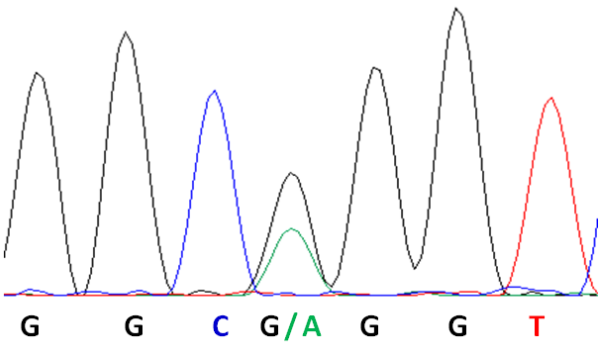
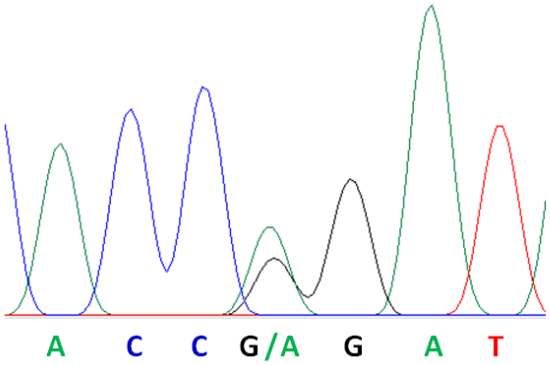


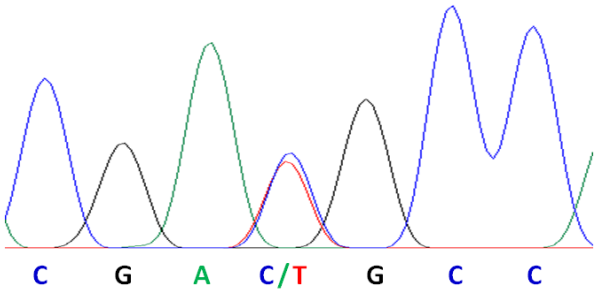
Figure 4.9 Chromatograms showing heterozygous loci in the validated SNPS



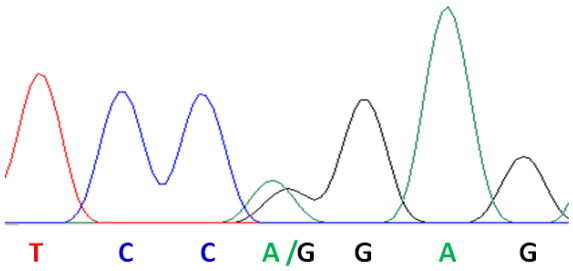
1: 69513042



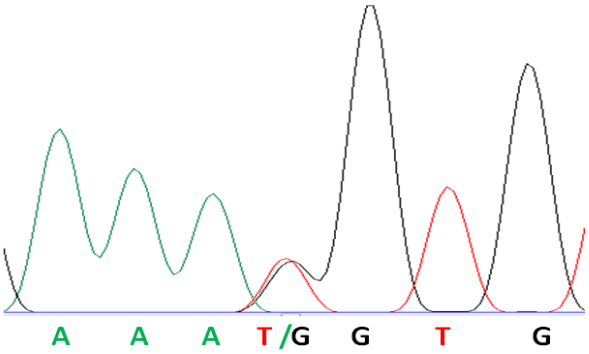
2: 44729967



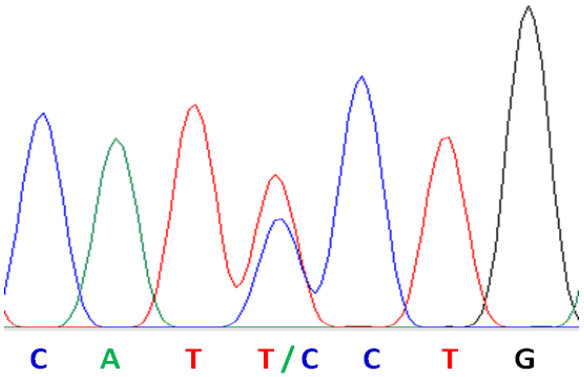
3: 251534



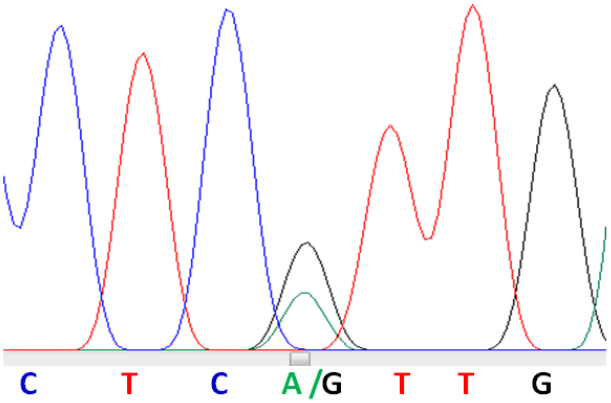
5: 80484794



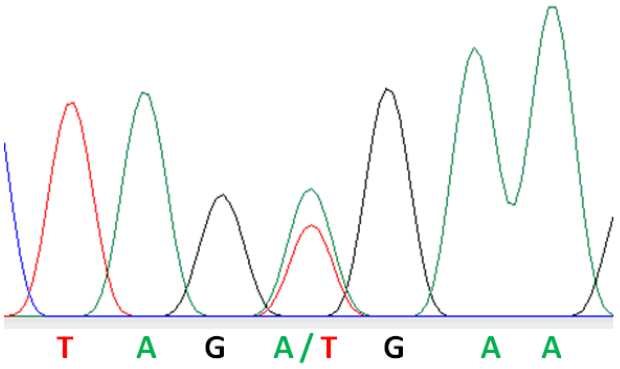
7: 3340499



9: 34404747

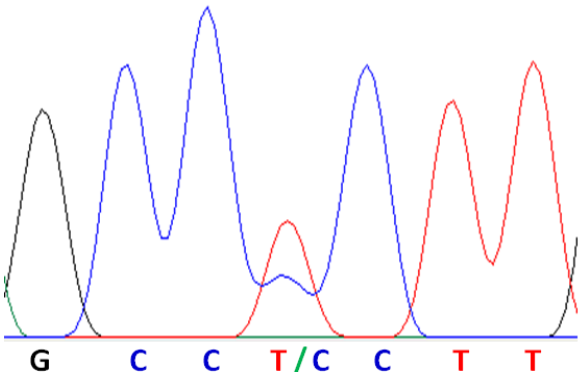


11: 7045033

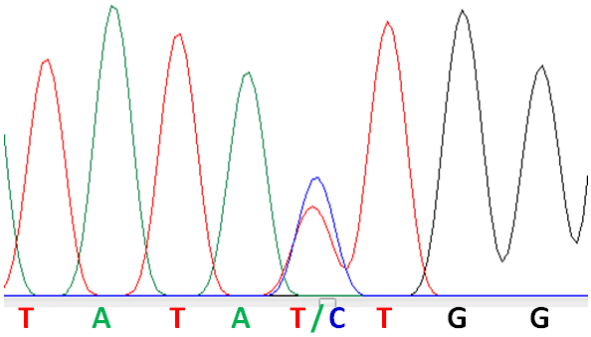


12: 79722429

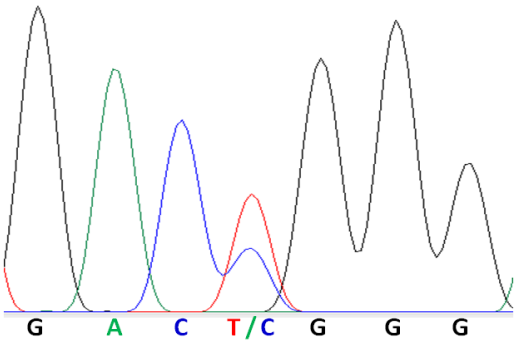
Figure 4.9 contd…



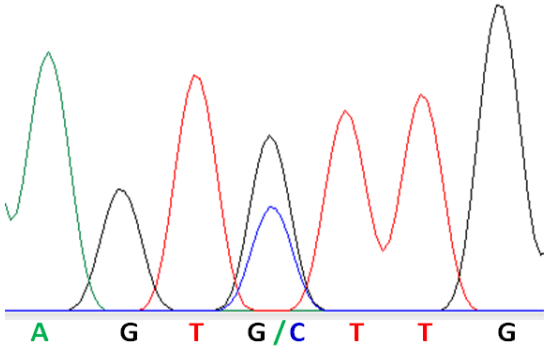
13: 31754433



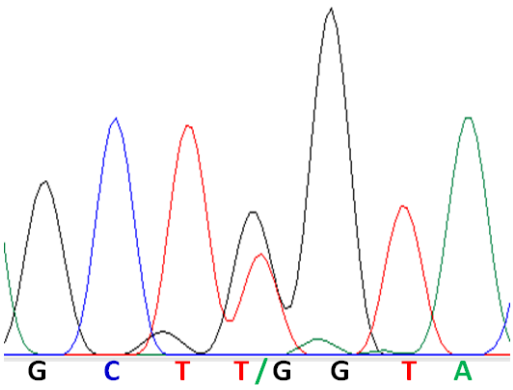
15: 78617028



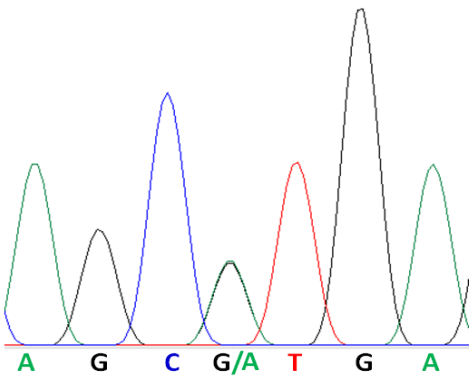
15: 54218636



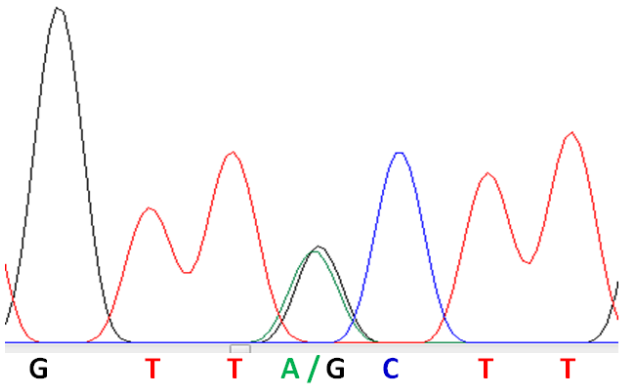
17: 40841339



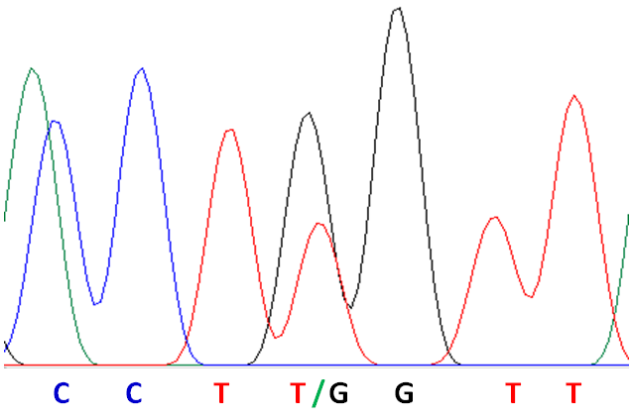
18: 24906224



20: 6840788

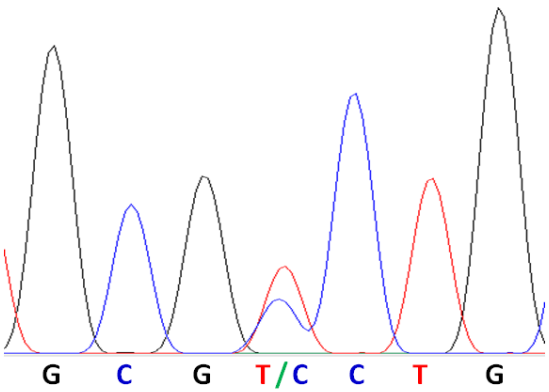


21: 6730554

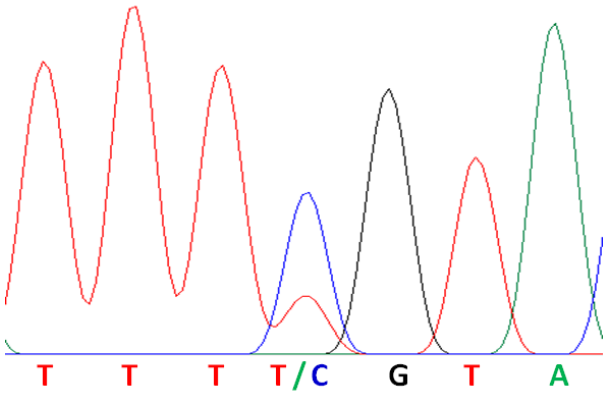


22: 963071

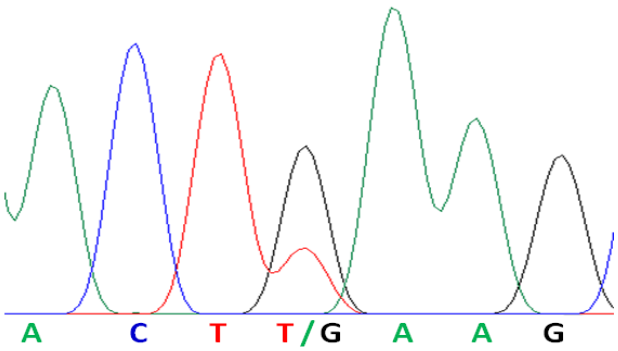
Figure 4.9 contd…



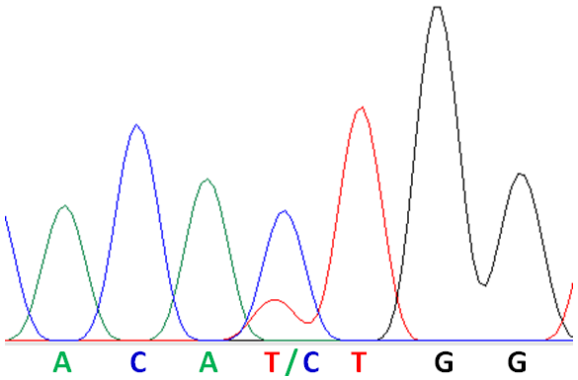
23: 22136851



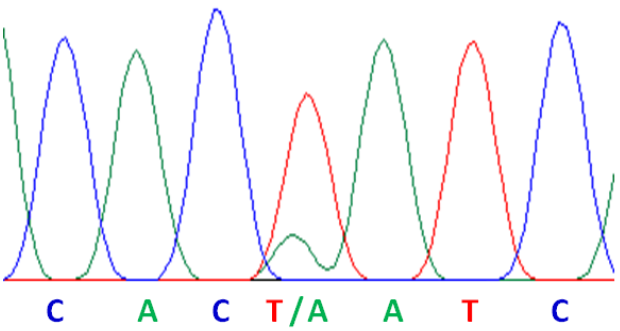
24: 7927208



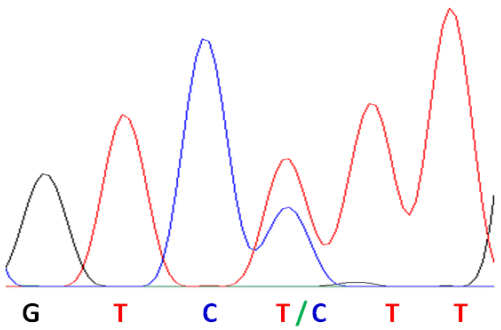
26: 35436616



27: 4437971



28: 13324730



29: 6840788

Figure 4.10 Genome-wide distribution of SNPs in milk yield QTLs

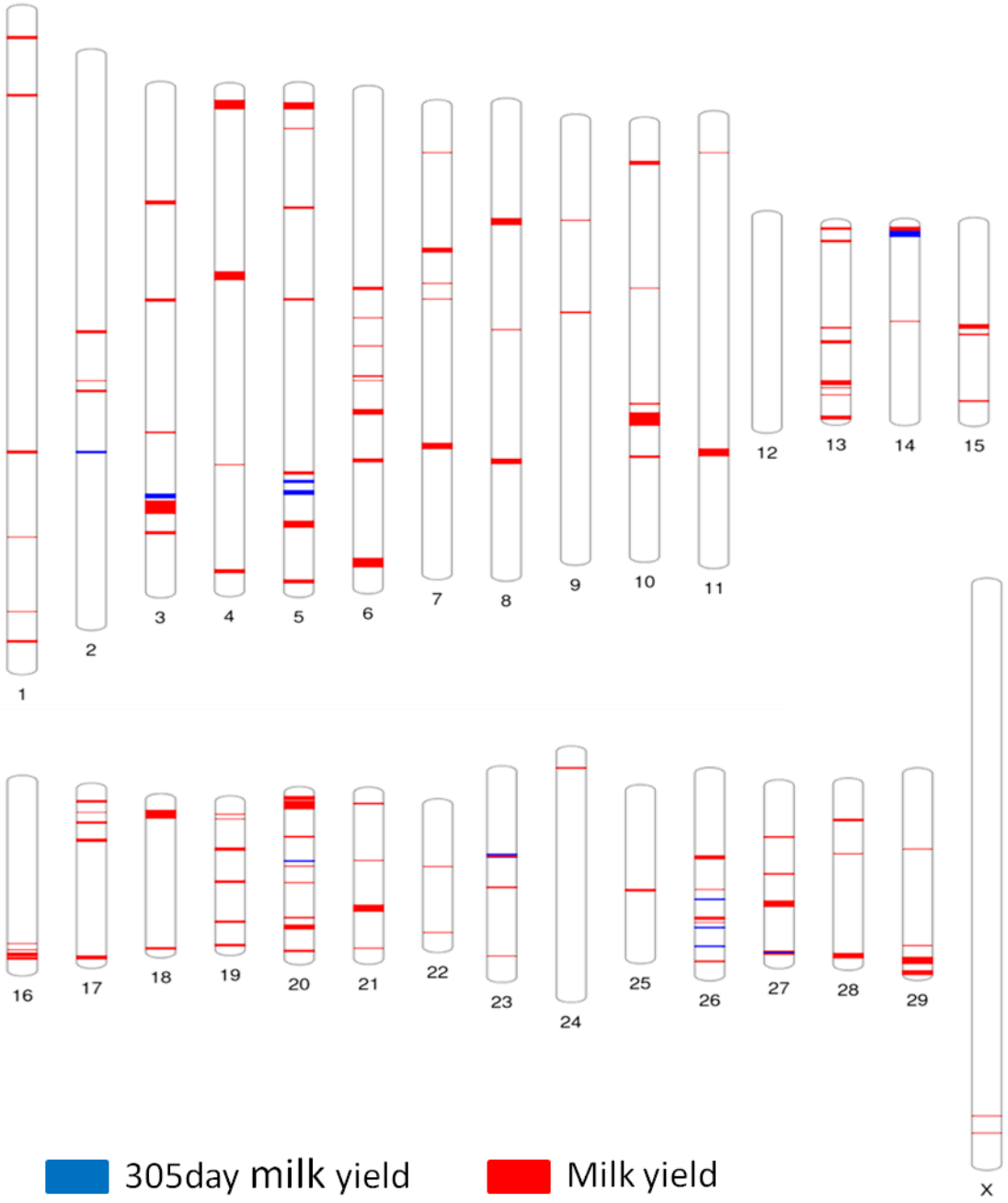


Figure 4.11 Genome-wide distribution of SNPs in milk fat QTLs

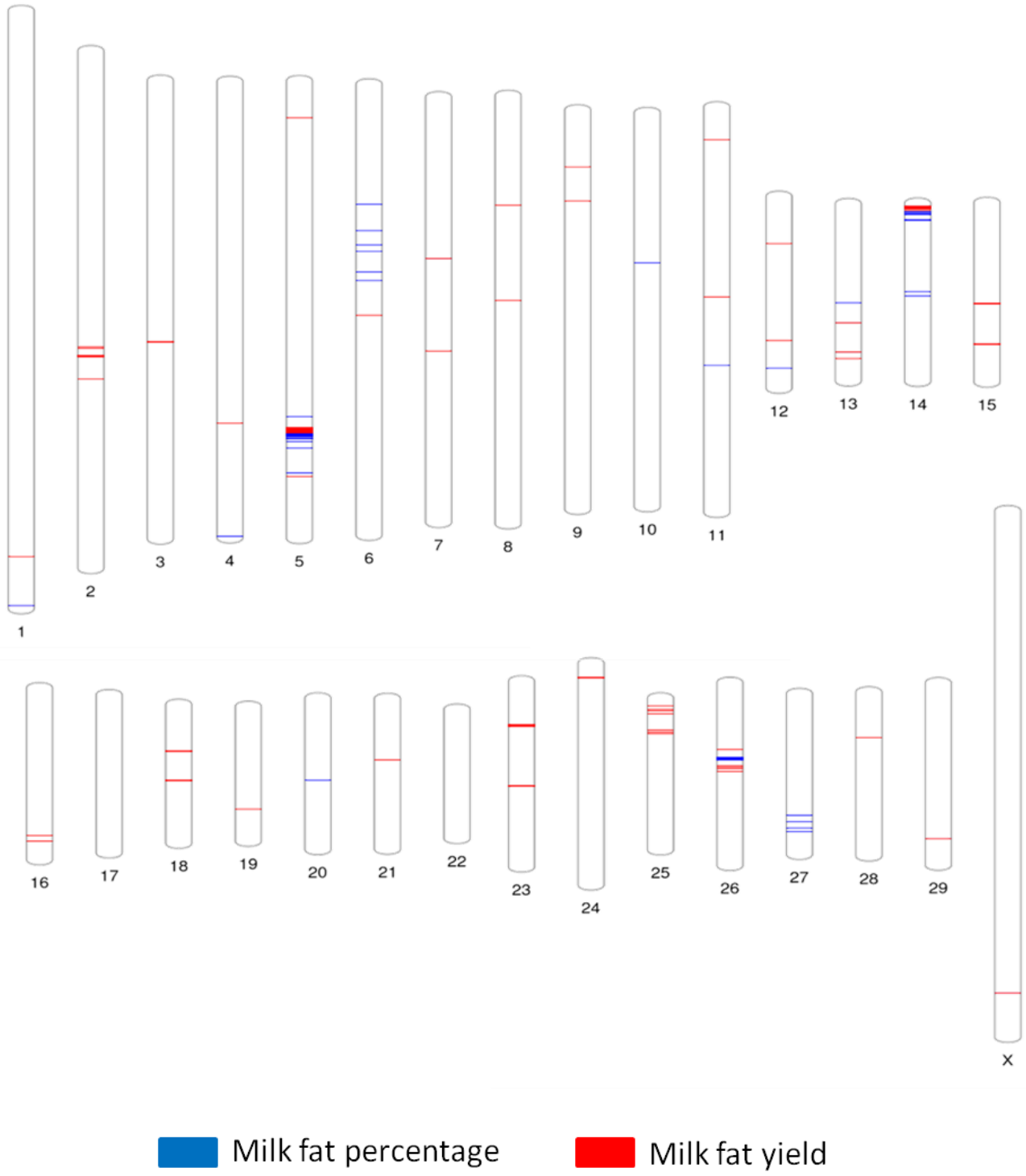
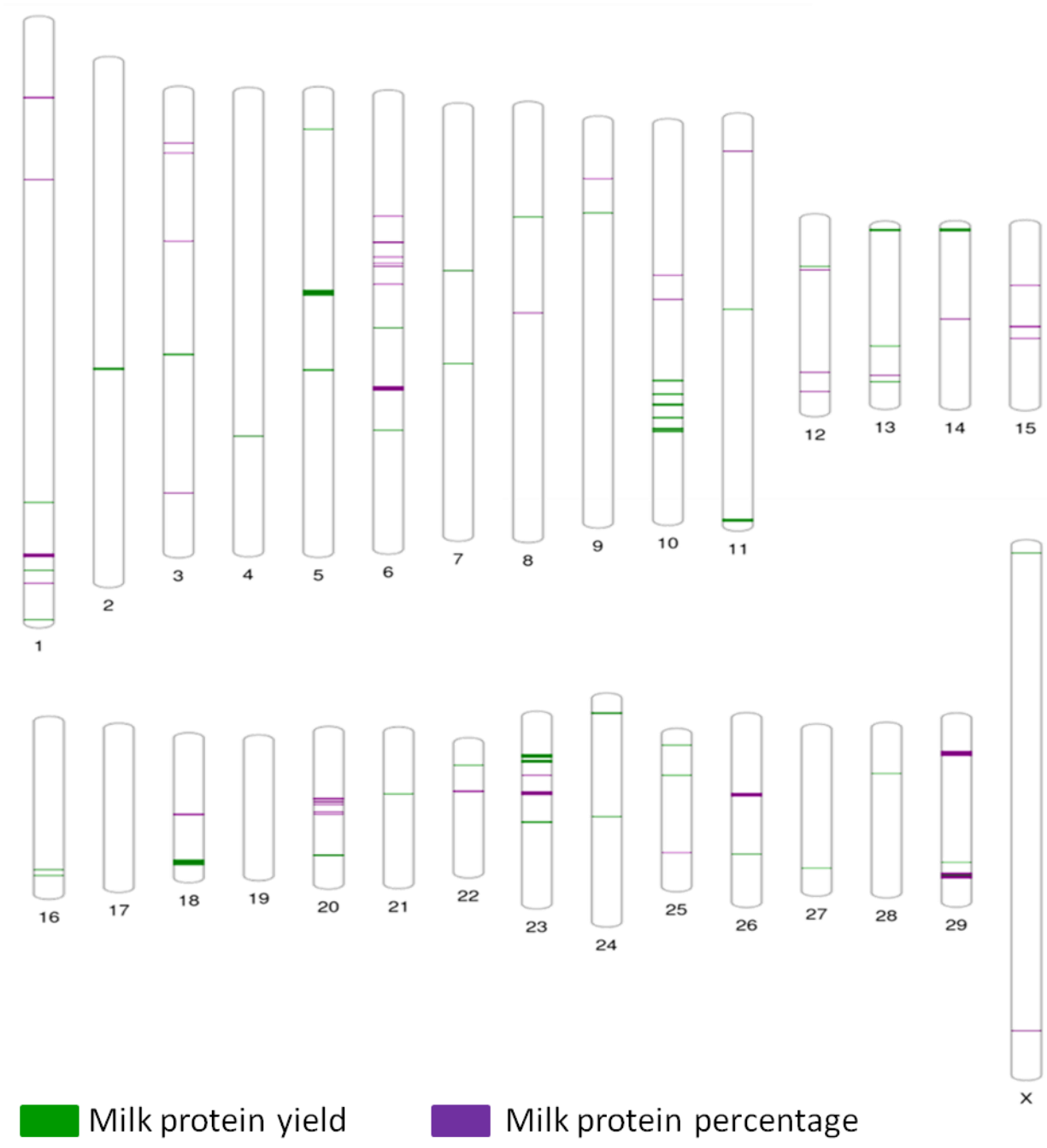


Figure 4.12 Genome-wide distribution of SNPs in milk protein QTLs



POLR3H (Polymerase (RNA) III subunit H), ACO2 (Aconitase 2) and ALG12 (Asparagine-Linked Glycosylation 12 Homolog) genes in BTA5; DGAT1 gene on chromosome 14 and SCD in chromosome 26 are reported to be associated with milk fat composition in dairy cattle (Grisart *et al*., 2002; Kgwatalala *et al*., 2009; Palombo *et al*., 2018)

Casein genes (CSN1S1, CSN1S2, CSN2, and CSN3) on chromosome 6; PRL on chromosome 23 are reported to be candidate genes for milk protein composition in dairy cattle (Schopen *et al*., 2011; Huang *et al*., 2012). The presence of these candidate genes might be the reason for increased variant density in these chromosomes.

Table 4.15 Mapping of SNPs to production trait QTLs

QTL	Number of SNPs
Milk yield	15912
305-day milk yield	1818
Milk fat yield	1231
Milk fat percentage	424
Milk protein yield	1717
Milk protein percentage	1651
Lactation persistency	9

4.5.2 Reproduction trait QTLs

The number of SNPs mapped to reproduction traits QTLs were 42314. The different trait wise SNPs distribution has been presented in table 4.16. The interval from first to last insemination was found to have the maximum SNPs followed by the interval to first estrus after calving trait. The SNPs in age at first calving (AFC) QTLs were found on chromosomes 2, 8, 9, 14, 16, 17 and 29. The maximum numbers of SNPs were located on chromosome 14. The SNPs located in age at puberty QTLs were maximum on chromosome X (Figure 4.13-4.15).

The genes located on BTA14 like PLAG1, NPBWR1, OPRK1, CRH, NCOA2 and FER1L6 were reported to be involved in the biological process for AFC. The NPBWR1 gene (Neuropeptide W (NPW)/Neuropeptide B (NPB) Receptor-1) acts on the brain to control

prolactin, corticosterone and growth hormone release (Baker *et al*., 2003). The OPKR1 (Opioid Receptor Kappa 1) reported as a potential candidate for puberty in Brahman cattle (Fortes *et al*. 2012). The CRH (Corticotropin-Releasing Hormone) gene was reported to be involved in ovarian steroidogenesis regulation and follicular maturation, ovulation and luteolysis (Kiapekou *et al*., 2010). The NCOA2 (Nuclear Receptor Coactivator) plays a key role in the development of puberty by acting as a transcription factor for multiple genes affecting the onset of puberty (Fortes *et al*. 2012). Camargo *et al*. (2015) found polymorphisms of NCOA2 associated with reproductive traits in Nellore cattle. These authors have reported significant effect of SNPs in NCOA2 gene for early pregnancy probability, days to first calving and AFC in Nellore females. The FER1L6 was found to be related with folliculogenesis (Stigliani *et al*., 2013). The PLAG1 gene was found to affect bovine stature (Karim *et al*., 2011) and could be related to age at puberty and AFC, due to the high genetic correlation between these traits.

Table 4.16 Mapping of SNPs to reproduction trait QTLs

QTL	Number of SNPs
Age at first calving	2897
Age at puberty	1101
Calving interval	1848
Calving to conception interval	5487
Conception rate	2943
Daughter pregnancy rate	391
First service conception	1997
Gestation length	405
Heifer pregnancy	3065
Inseminations per conception	1670
Interval from first to last insemination	7865
Interval to first estrus after calving	7325
Luteal activity	3552
Pregnancy rate	187
Still birth	1581

4.5.3 Mastitis QTLs

The number of SNPS mapped to clinical mastitis QTLs were 1410 SNPs, while the somatic cell score QTLs were having 4355 SNPs (Table 4.17). The chromosome 9, 25 and 27 did not exhibited any SNP for mastitis QTLs. While the chromosomes 12, 18 and 20 had dense mapping of SNPs for mastitis QTLs (Figure 4.16). Candidate genes for mastitis resistance like HPN (Hepsin), C5AR1 (Complement C5a Receptor 1), HP (Haptoglobin) and SPEF2 (Sperm Flagellar 2) were reported on these chromosomes. (Ogorevc *et al*., 2008, Cai *et al*., 2018)

Table 4.17 Mapping of SNPs to mastitis QTLs

QTL	Number of SNPs
Clinical mastitis	1410
Somatic cell score	4355

4.5.4 Udder and teat type QTLs

Among the udder and teat type QTLs, the udder cleft was having maximum SNPs in associated QTL regions. No SNPs were mapped to QTLs associated with teat number, udder texture and udder width (Table 4.18). The chromosome 22 was found to have maximum QTLs for udder and teat trait QTLs (Figure 4.17). The GWAS showed significant SNPs for fore udder attachment, udder depth and udder cleft and located on BTA22 (Cole *et al*., 2011). The TGFBR2 (Transforming Growth Factor Beta Receptor 2) gene which was involved with mammary gland development and mammary gland duct morphogenesis was mapped to BTA22 (Marete *et al*., 2018).

4.5.5 Disease susceptibility and tropical adaptation QTLs

Total 5068 SNPs were mapped to bovine tuberculosis susceptibility QTLs, while 1345 SNPs were mapped to *M. paratuberculosis* susceptibility. Among the tropical adaptation traits, tick resistance QTLs were having 7689 SNPs while, QTLs for heat tolerance exhibited 2300 SNPs (Table 4.18 and Table 4.19).

The chromosomes 3, 18 and 23 exhibited maximum SNPs for bovine tuberculosis susceptibility QTLs. The SNPs to *M. paratuberculosis* susceptibility QTLs were having

maximum on chromosome 23 (Figure 4.18). The SNPs to tick resistance QTLs were located in all chromosomes except 9, 18, 19, 21, 24 and X. The chromosome 12 carried the maximum SNPs for heat tolerance QTLs (Figure 4.19).

Table 4.18 Mapping of SNPs to udder and type QTLs

QTL	Number of SNPs
Teat length	29
Teat number	0
Teat placement	35
Udder attachment	29
Udder cleft	506
Udder depth	28
Udder height	30
Udder structure	4
Udder texture	0
Udder width	0

Table 4.19 Mapping of SNPs to disease susceptibility and tropical adaptation QTLs

QTL	Number of SNPs
Bovine tuberculosis susceptibility	5068
M. paratuberculosis susceptibility	1345
Tick resistance	7689
Heat tolerance	2300

Two QTLs were reported in BTA23 for *M. paratuberculosis* susceptibility (Kirkpatrick *et al*., 2011, Sallam *et al*., 2017). Functional gene like TDP2 (Tyrosyl-DNA Phosphodiesterase 2) in this region was likely candidate gene for *M. paratuberculosis* susceptibility trait (Gao *et al*., 2018). The chromosome 12 carried several heat shock protein (HSP) genes like HSP 90-alpha A2, DnaJ heat shock protein family (Hsp40), DnaJ heat shock protein family (Hsp40) member

C3 and heat shock protein family H (Hsp110) member 1(Ajayi *et al*., 2018). These HSP genes might have contributed to the increased SNP density on chromosome 12.

4.6 SSR identification

Total 242145 consensus sequences were obtained which were found to have 8266 perfect microsatellites in ten Sahiwal cattle. Aligning the sequences with the *Bos taurus* genome 5766 microsatellites were found mapped to different chromosomes. The di-nucleotides were the most abundant repeat motifs (4552; 78.94%), followed by tri-nucleotides (945; 16.26%), penta-nucleotides (216; 3.74%) and tetra-nucleotides (60; 1.04%) (Table 4.20 and Table 4.21).

The dominant di-nucleotide motifs were TG repeats and the least common repeat was CT. The AAC repeats were the most frequent tri-nucleotides motifs, whereas, AAAG and AACTG were the most common tetra- and penta-nucleotide motifs, respectively (Table 4.22).

Microsatellites were developed for fish using ddRAD sequencing (Jansen *et al*., 2016) and in plants using RAD seq data (Barchi *et al*., 2011). But no reports were available in cattle to compare the present results. Yan *et al*. (2008) identified 1831 EST based SSRs from the cattle gene sequences of NCBI. The di-nucleotide motif repeats were the most abundant SSR (54%) and tetra-nucleotide repeats the least (4%). Similar results pertaining to high frequency of di-nucleotide motif repeats, AC (57%) were obtained by Yan *et al*. (2008) and 35,402 AC repeats by Xu *et al*. (2017) in cattle.

Table 4.20 Classification of the identified SSRs

Motif length	Number of SSRs
Di-nucleotide	6593
Tri-nucleotide	1216
Tetra-nucleotide	156
Penta-nucleotide	301
Total	8266

Table 4.21 Chromosome wise distribution of SSRs

Chromosome	Motif type			
	Di-nucleotide	**Tri-nucleotide**	**Tetra-nucleotide**	**Penta-nucleotide**
1	284	74	4	11
2	200	49	1	15
3	137	39	3	07
4	201	48	5	08
5	156	37	4	11
6	182	53	2	12
7	186	41	0	09
8	201	41	3	07
9	153	37	1	07
10	258	31	0	04
11	156	31	6	13
12	136	31	3	08
13	261	22	0	05
14	145	29	3	03
15	166	36	0	06
16	208	16	0	03
17	111	19	1	12
18	155	18	2	08
19	112	33	1	02
20	118	26	2	08
21	108	38	2	11
22	83	31	4	02
23	114	10	3	04
24	118	28	1	02
25	73	16	2	06
26	64	19	0	04
27	62	12	1	02
28	80	21	1	07
29	81	14	1	05
X	243	45	6	14
Total	**4552**	**945**	**62**	**216**

Table 4.22 Classification of di-, tri-, tetra-and penta-nucleotide SSRs

Motif	Number of SSRs
Di-nucleotide SSRs	
AC	642
AG	509
AT	408
CA	555
CT	189
GA	414
GT	502
TA	269
TC	270
TG	794
Tri-nucleotide SSRs	
AAC	199
AAG	7
ACA	12
ACC	12
AGC	72
AGG	3
ATC	14
ATG	3
ATT	11
CAA	55
CAC	20
CAG	11
CAT	1
CCA	7
CCT	2
CTT	14
CTG	146
CTC	3
TTG	22
TTA	1
TGT	14
TGG	4
TGC	163
TGA	3
TCC	2
TCA	1
TAA	8

Table 4.22 contd…

…Tri-nucleotide SSRs	
Motif	**Number of SSRs**
GTT	17
GTG	9
GGT	10
GGA	1
GCT	90
GCC	1
GCA	4
GAT	1
Tetra-nucleotide SSRs	
AAAG	12
CTTT	1
TTCT	1
AAAT	10
TTTA	6
AACC	1
CATG	2
GGAT	5
CCAT	1
AGGG	1
GCCT	1
GGCA	1
CCAG	1
AATG	1
ATTC	2
GAAT	2
TGAA	1
TGTA	1
CATA	1
TTTG	1
TTGT	2
TGTT	4
GTTT	1
CAAA	3

Table 4.22 contd…

Penta-nucleotide SSRs	
Motif	**Number of SSRs**
AAAAC	3
GTTTT	2
CAAAA	1
ATCAG	8
CTGAT	11
GATCA	5
AACTG	131
CAGTT	53
TATGC	1
AATGG	1

4.7 Comparison of ddRAD SNPs with bovine SNP chips

Comparison of the SNPs identified in the present study with Illumina BovineSNP50 BeadChip and Illumina BovineHD BeadChip revealed 713 (1.3%) and 1021 (0.1%) common SNPs respectively. The chromosome 1 was found to have the most number of common SNPs in both the chips. The chromosome wise distribution of shared SNPs have been given in table 4.23

Brouard *et al*. (2017) found 1604 (2.9%) common SNPs between the Illumina BovineSNP50 BeadChip and those identified using a double enzyme GBS method in taurine cattle. Gurgul *et al.* (2019) identified at least 44 common SNPs between GBS-generated SNPs and Illumina BovineSNP50 BeadChip. The lesser number of common SNPs identified might be due to preselection and non-representation of SNPs in genotyping chips, while ddRAD identifies population specific SNPs including rare SNPs.

Many researchers have reported the polymorphic loci of Illumina BovineHD BeadChip in Sahiwal cattle. Dash *et al*. (2018) reported 343951 (44.21%) SNPs using Illumina BovineHD BeadChip genotyping of Sahiwal cows (n=19) of north India. A study by Nayee *et al*. (2018) in Sahiwal bulls (n=43) identified 620344 (64.39%) SNPs. Genotyping using Illumina BovineHD BeadChip chips identified 500968 SNPs in Sahiwal cattle (n=14) of Pakistan (Mustafa *et al*., 2018).

Table 4.23 Chromosome wise distribution of shared SNPs with bovine SNP chips

Chromosome	Illumina BeadChip	
	BovineSNP50 (50k)	BovineHD (777k)
1	52	66
2	44	44
3	42	44
4	40	49
5	19	45
6	32	48
7	42	48
8	32	52
9	28	41
10	27	40
11	31	48
12	18	33
13	19	31
14	20	28
15	17	40
16	18	34
17	21	33
18	21	26
19	17	20
20	13	24
21	15	36
22	9	22
23	15	23
24	23	24
25	17	16
26	18	15
27	22	22
28	13	24
29	15	19
X	13	26
Total	**713**	**1021**

Only less than 1% of the SNPs identified in the present study in Sahiwal cows were mapped to the existing bovine SNP chips. Hence, there is a scope for inclusion of more indigenous SNPs in the existing SNP chips for its efficient use in zebu cattle.

4.8 Comparison of genome coverage of ddRAD sequencing with whole genome of cattle

The genome coverage of the samples ranged from 2.14 to 4.5%. The average coverage was found to be 3.27% (Figure 4.19) in the present study and similar results of genome coverage of 4.66% in chicken were obtained by Pertille *et al*. (2016), while sequencing with GBS method. Whereas, Zhu *et al*. (2016) reported mean genome coverage of 13% in ducks and the higher coverage might be due the many gaps existing in duck reference genome.

4.9 Identification of selection sweeps

Putative selection sweeps were identified in the Sahiwal genome by finding the outlier regions with high CLR value. Maximum numbers of putative selection sweep regions were identified in chromosome 17 (7 selection sweep regions). The average size of selective sweeps regions were 3481 ± 476 bp. The size of the putative selection sweeps regions ranged from 76 to 30649 bp. On the contrary, Ramey *et al*., (2013) reported an average breed-specific selection sweep of 336 ± 119 kb with a range of 207 to 702 kb. The higher value may be due to the different genotyping method and statistics used. The chromosome wise putative selection sweeps identified in the present study are summarized in table 4.24.

Table 4.24 Putative Selective sweep regions identified in Sahiwal

Chromosome	Region (UMD3.1 coordinates bp)
1	36714-39932
	4401053-4401857
	14961132-14963545
	19220079-19221206
	66081375-66088615
	143347749-143348553
2	21589233-21602594
	94645218-94648312
3	13026717-13057366
	14558412-14558540
	17178411-17180071
	104015107-104015873

Table 4.24 contd…

Chromosome	Region (UMD3.1 coordinates bp)
4	51229581-51239772
	57863677-57865665
5	512442-522485
	113055582-113057967
6	14083688-14083811
	18344319-18359029
	69619118-69622060
	111940035-111940770
7	10632052-10632723
	47355419-47357990
	88239653-88246358
	91883513-91887648
8	68319740-68331523
9	12486785-12488082
	76769380-76770353
	77299740-77299848
	78560304-78560412
	82139449-82144637
	83300685-83303279
10	15775189-15778266
	17022324-17023809
	72044448-72045827
11	29335720-29335830
	39398021-39406939
	49701974-49702304
	82415710-82416040
	14497737-14502471
12	30734419-30734590
	48986041-48991508
	59904179-59905802
	70127253-70128364
13	78466388-78478423
14	22078872-22079116
	42589853-42599287
	42590016-42599287
	43955777-43955858

Table 4.24 contd…

Chromosome	Region (UMD3.1 coordinates bp)
15	79785555-79792247
	80175300-80177756
16	12314797-12316276
	35581563-35582887
	45493185-45498012
17	10718598-10719744
	17371738-17371814
	43115975-43116052
	43166882-43167799
	47304715-47305555
	48082460-48085288
	49824685-49828048
18	44871913- 44877372
19	29695465-29696573
	58086740-58092669
20	32127079-32132691
	69357643-69359539
21	23985810-23986087
	24035562-24035909
	51465473-51471710
22	2798645-2804197
	31617271-31618073
23	39021791-39025833
	39026897-39028227
24	34609239-34610795
	35066542-35069396
	42193177-42195512
25	21419297-21421184
	34994026-34995123
	37474462-37485125
26	277506-278022
	17195006-17196142
	17589074-17590673
	18536084-18536446
	43166460-43167905

Table 4.24 contd…

Chromosome	Region (UMD3.1 coordinates bp)
27	48743050-48747970
28	601246-604696
	27502723-27503275
	33395123-33395628
29	14782162-14782578
	14937223-14942841

4.9.1 Candidates genes in selection sweep regions

The genes located in the putative selection sweeps were identified and were screened for candidate genes. Many genes associated with production, reproduction, domestication and tropical adaptation were found in putative selection sweep regions (Table 4.25 and Table 4.26). Candidate genes for production and reproduction traits like CLEC16A (C-type Lectin Domain Containing 16A), EPS8 (Epidermal Growth Factor Receptor Pathway Substrate 8), FBP2 (Fructose-bisphosphatase 2), FYB (FYN Binding Protein 1), IGFBP2 (Inositol 1,4,5-Trisphosphate Receptor Type 2), IGFBP7 (Insulin-like Growth Factor Binding Protein 7), ITPR2 (Inositol 1,4,5-Trisphosphate Receptor Type 2), LIFR (LIF Receptor Alpha), MGST1 (Microsomal glutathione S-transferase 1), MKL1 (Myocardin Related Transcription Factor A), NEURL1 (Neuralized E3 Ubiquitin Protein Ligase 1), RICTOR (RPTOR Independent Companion Of MTOR Complex 2), TRAPPC9 (Trafficking Protein Particle Complex 9), SEMA3C (Semaphorin 3C), SOX5 (SRY (sex determining region Y)-box 5), GPR125 (G Protein-Coupled Receptor 125), FSHR (Follicle Stimulating Hormone Receptor), CACNB2 (Calcium Voltage-Gated Channel Auxiliary Subunit Beta 2) , TOX (Thymocyte Selection Associated High Mobility Group Box), CSPP1 (Centrosome And Spindle Pole Associated Protein 1), MTOR (Mechanistic Target Of Rapamycin Kinase) and CACNA1D (Calcium Voltage-Gated Channel Subunit Alpha1 D) were found to be located in the putative selection sweep regions.

Higher expression levels of IGFBP7 were reported in pregnant equine endometrium (Merkl *et al*., 2010) and glandular epithelial cells during the secretory phase (Kutsukake *et al*., 2010). Upregulation of IGFBP2 expression was reported in early stages of pregnancy in bovine endometrium (Klein *et al*., 2006). The IGFBP2 also had an important role in activation of

Figure 4.13 Genome-wide distribution of SNPs in Reproduction traits QTLs

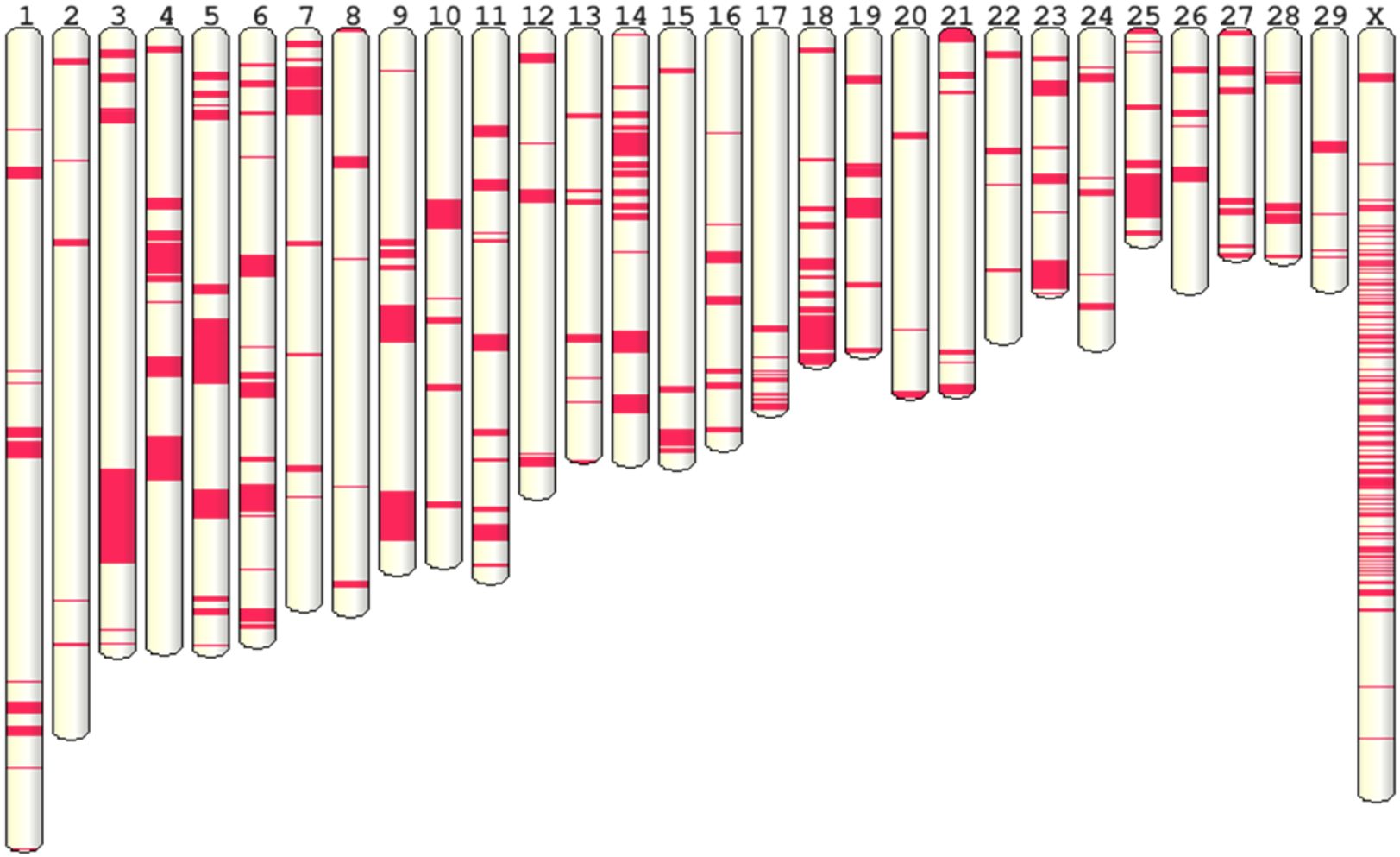


Figure 4.14 Genome-wide distribution of SNPs in age at first calving QTLs

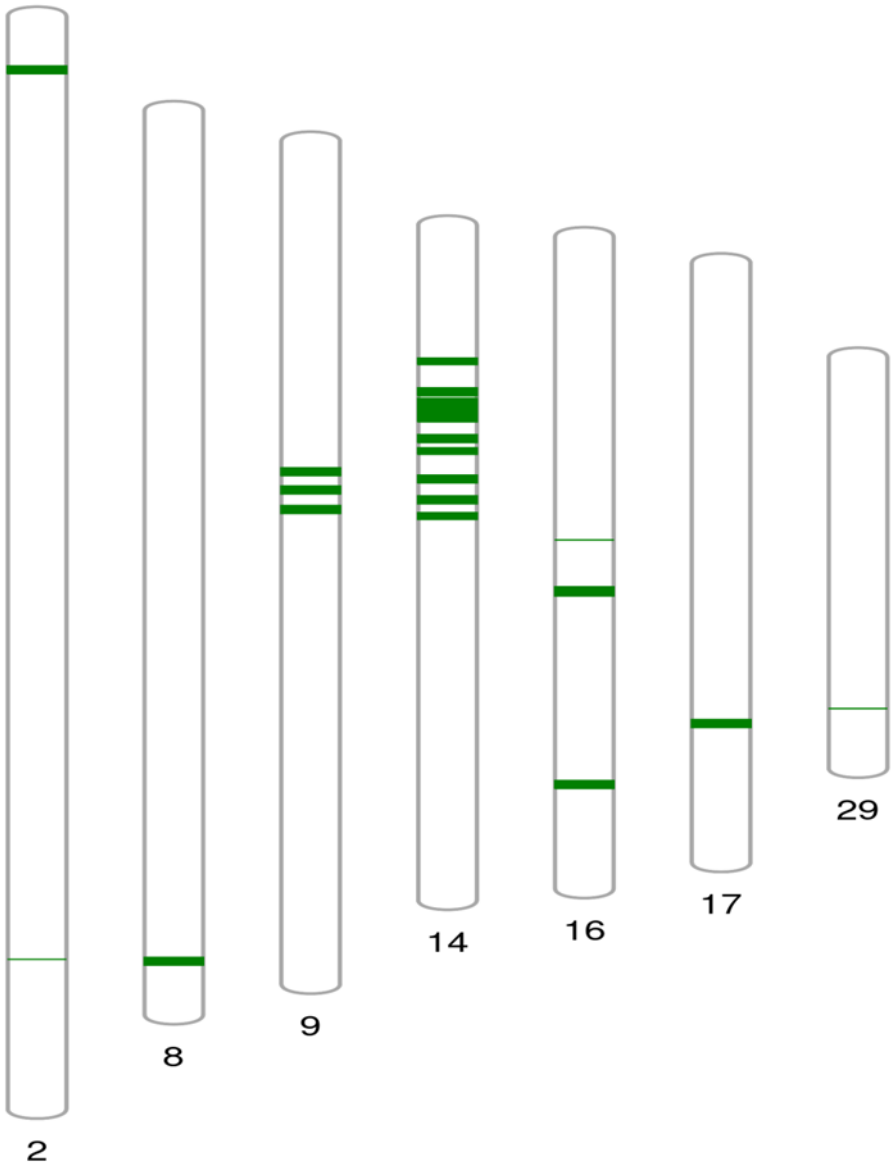


Figure 4.15 Genome-wide distribution of SNPs in age at puberty QTLs

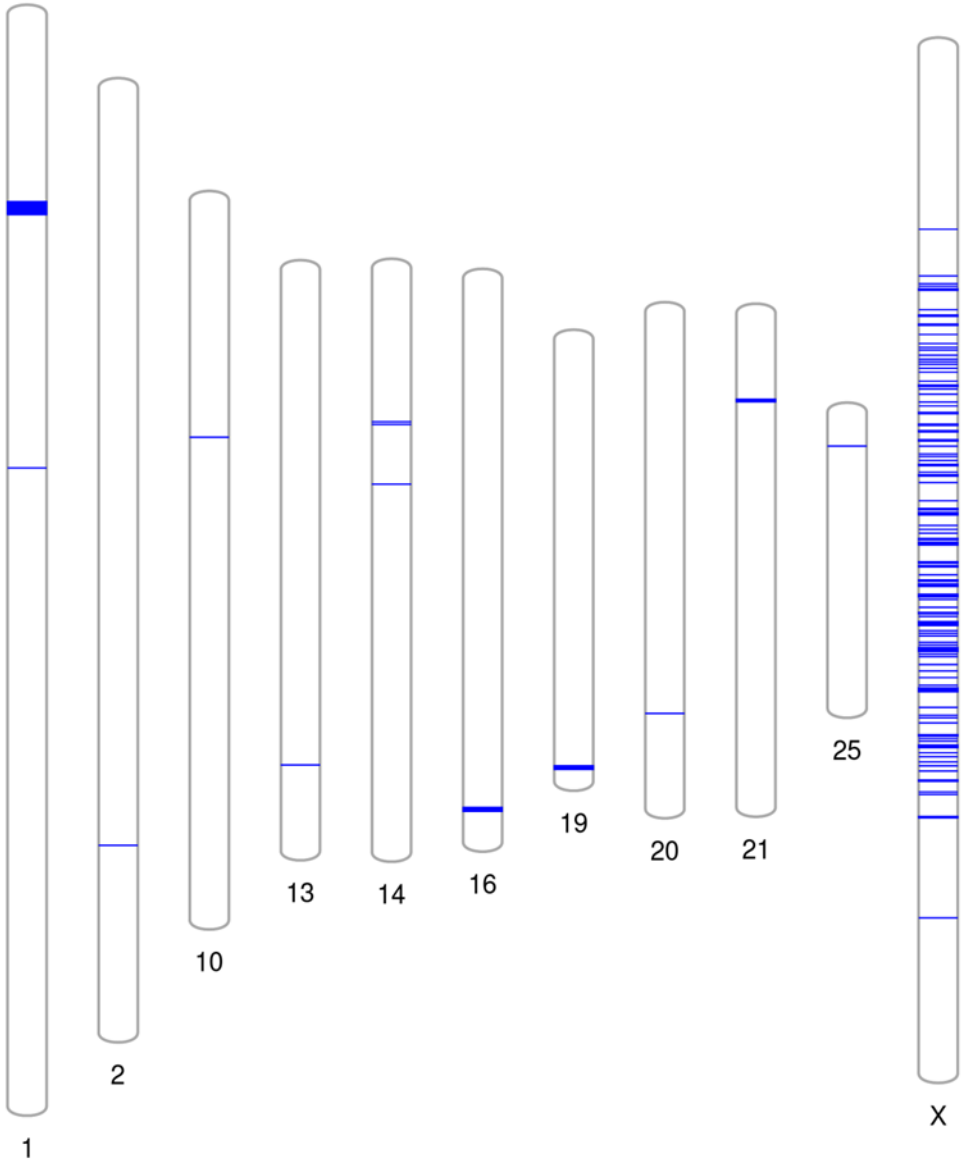


Figure 4.16 Genome-wide distribution of SNPs in mastitis QTLs

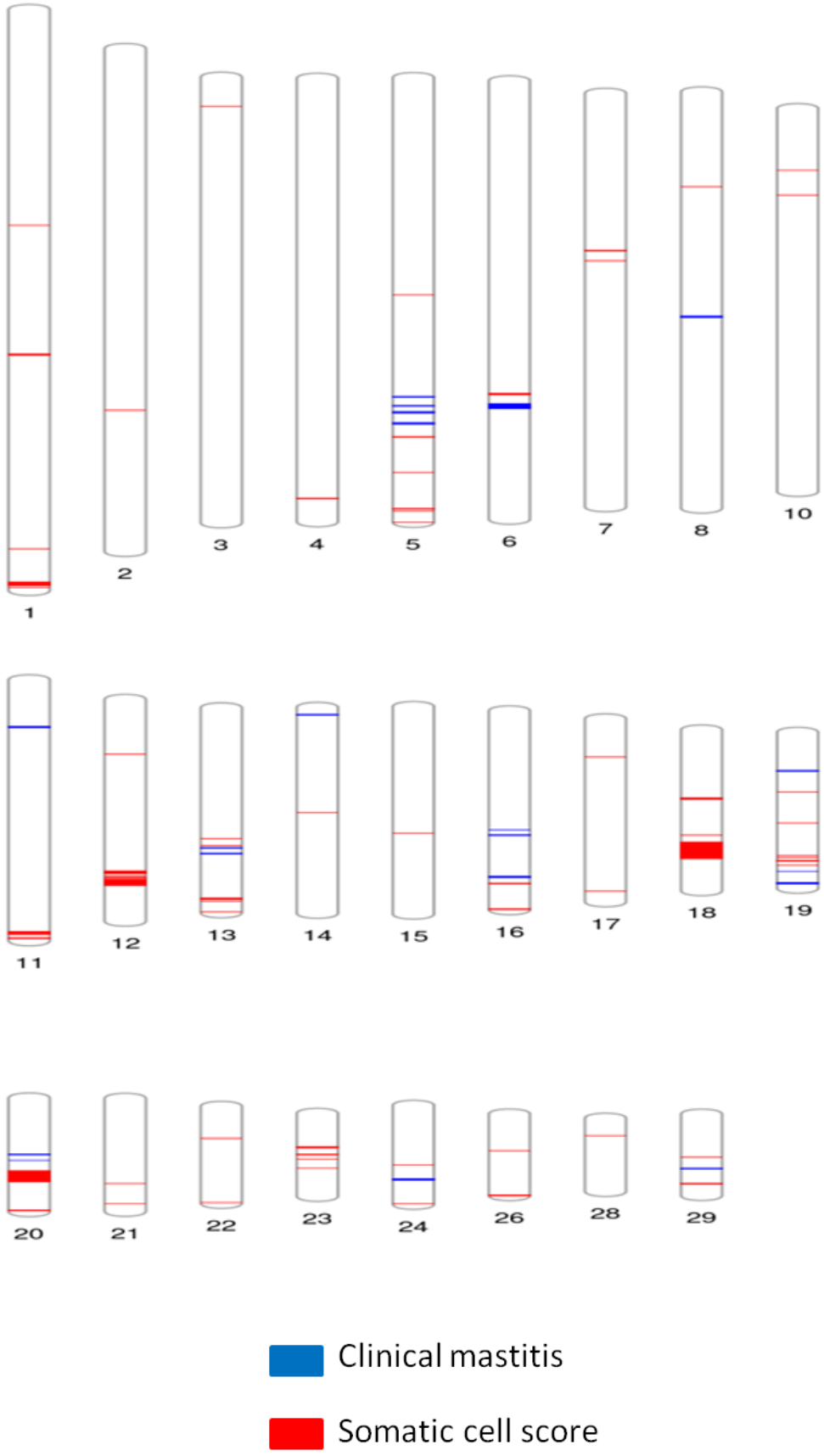


Figure 4.17 Genome-wide distribution of SNPs in udder and teat type QTLs

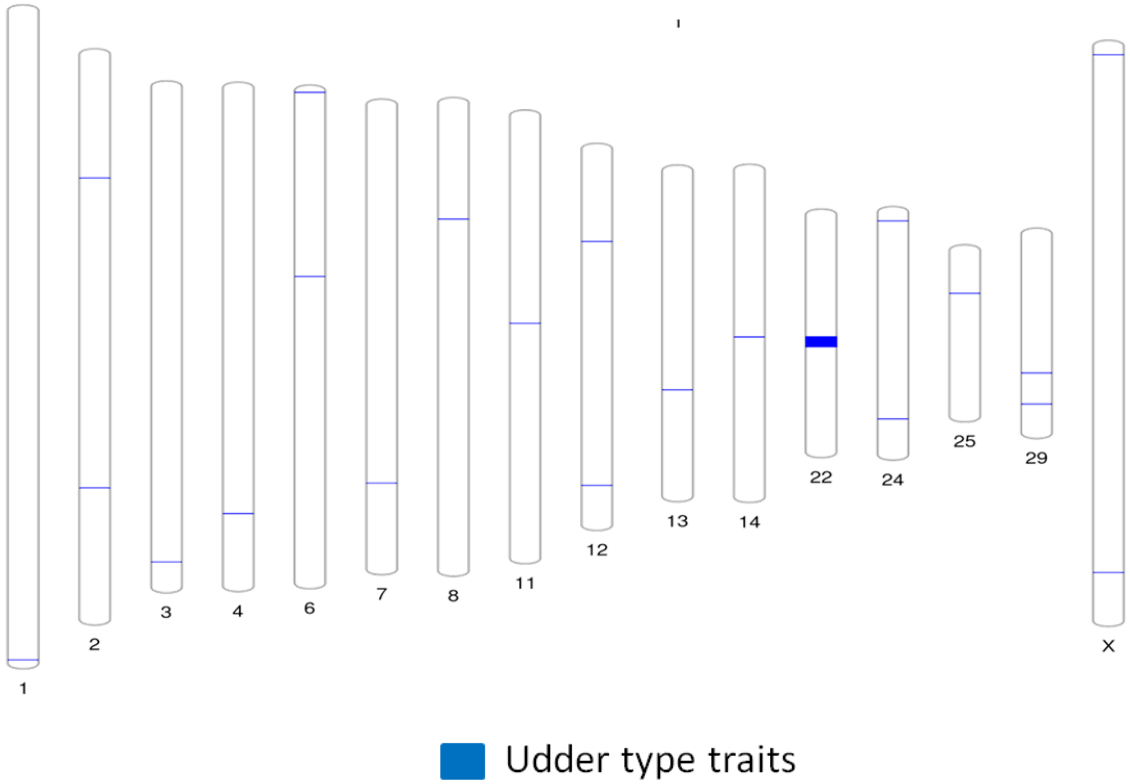


Figure 4.18 Genome-wide distribution of SNPs in disease susceptibility QTLs

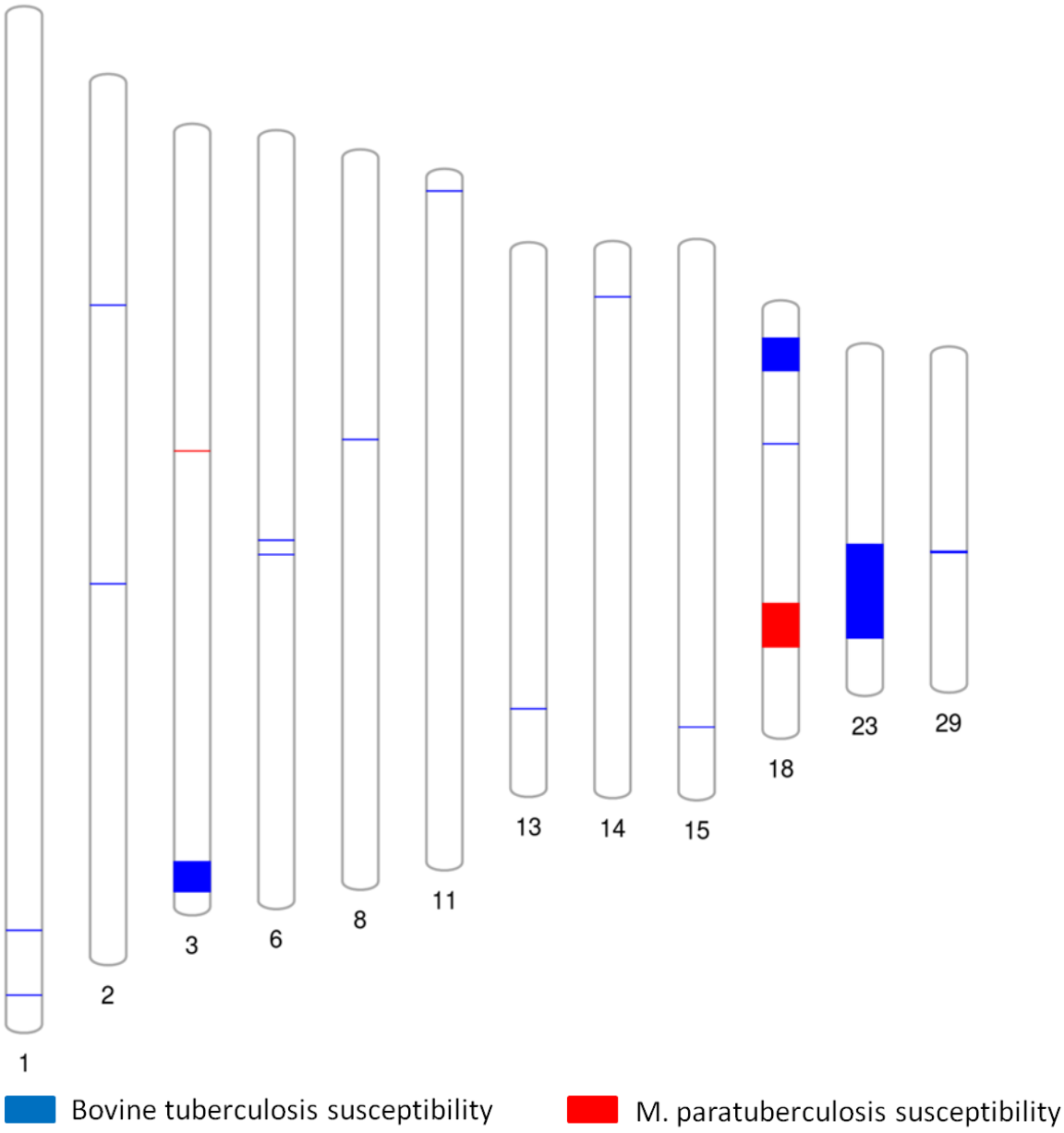


Figure 4.19 Genome-wide distribution of SNPs in tropical adaptation trait QTLs

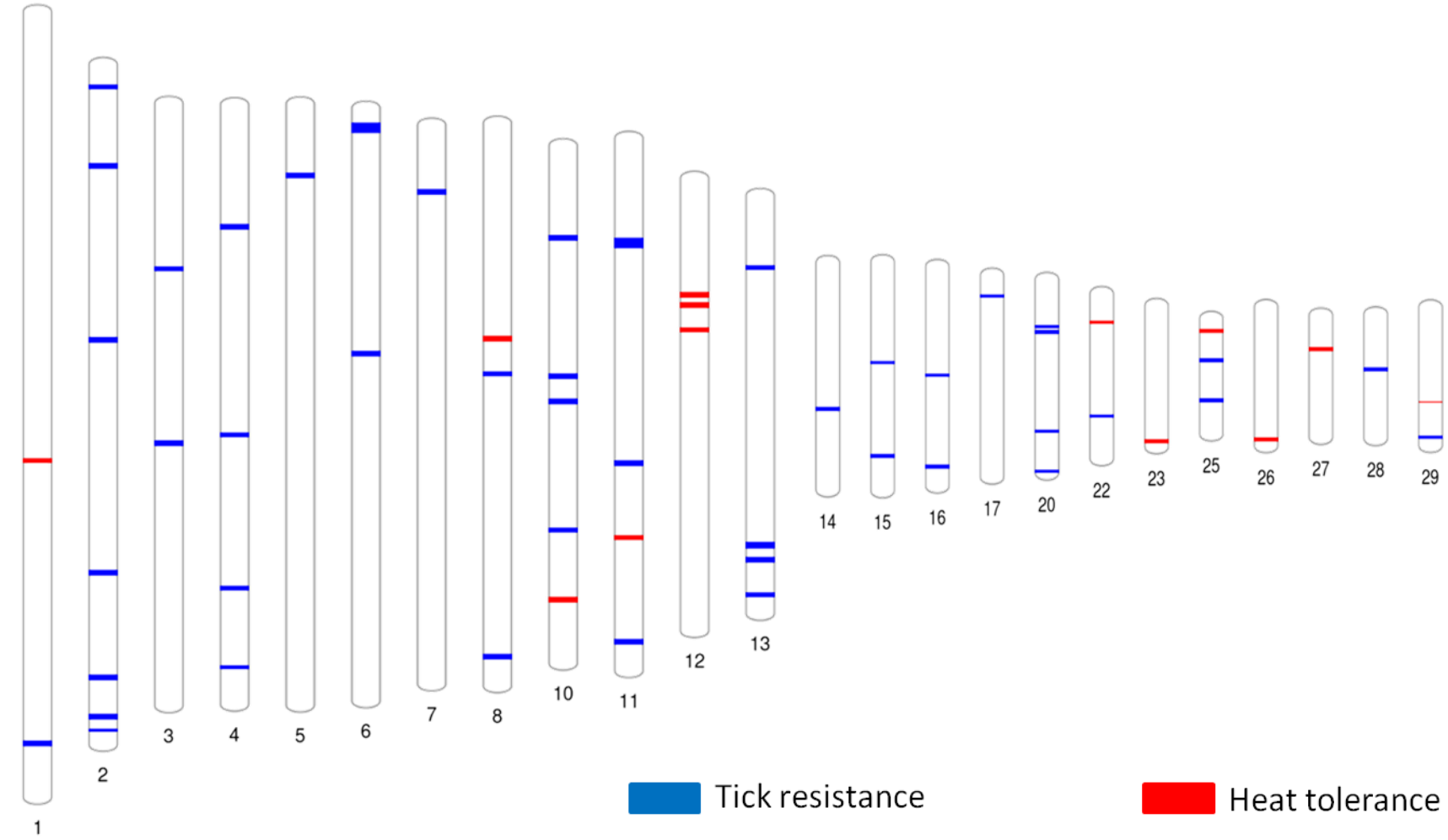
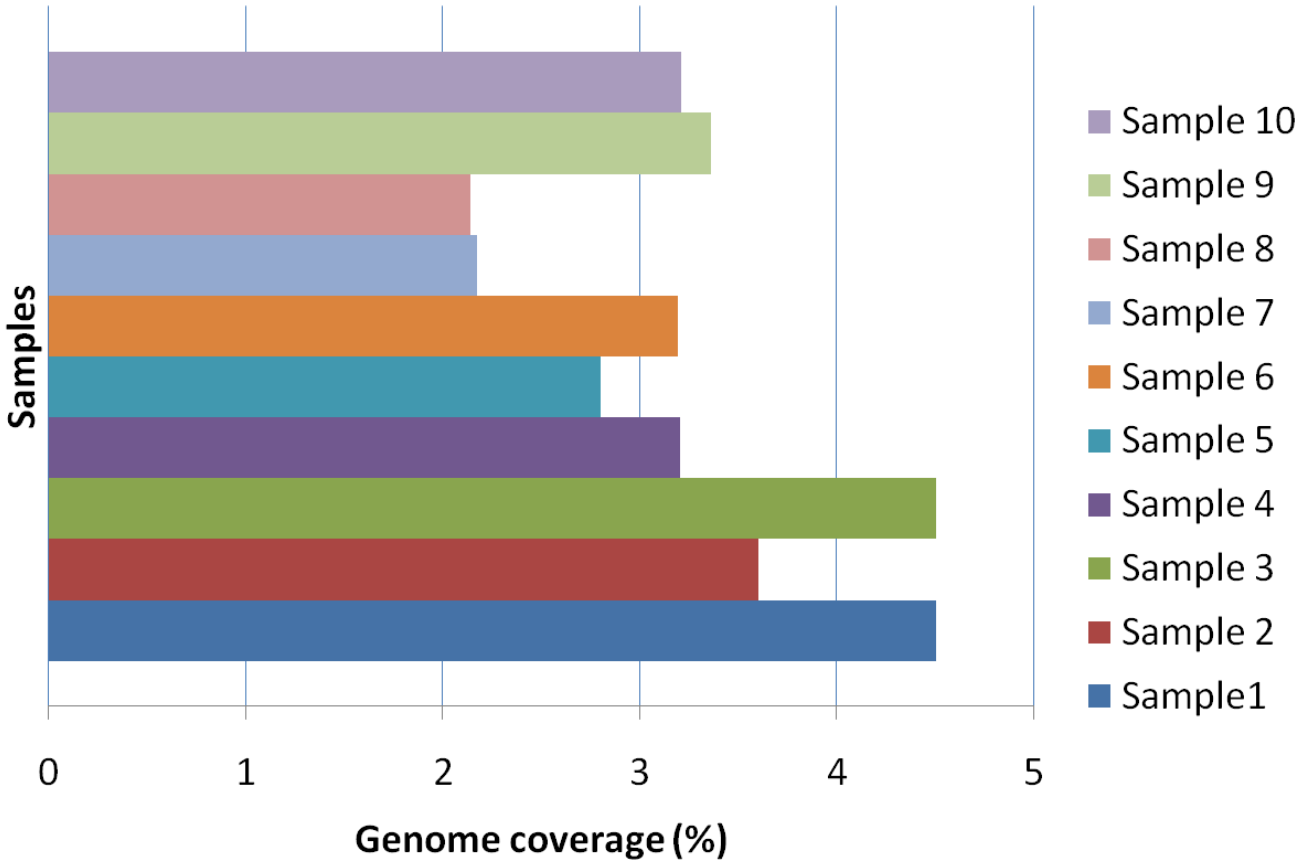


Figure 4.20 Genome coverage of ddRADseq reads from different samples



metabolic processes (glycolysis, storage, and sequestering of glycogen) to meet the increased metabolic demands of the cows under of high milk production (Rhoads *et al*., 2008). The LIFR was involved in mammary gland development and involution pathways and in prolactin signaling pathways in dairy cattle (Raven *et al*., 2014b). The MGST1 was an inflammation response gene which was highly expressed throughout pregnancy and lactation (Church, *et al*., 2012). The FBP2 gene's effect on milk production was mainly due to its association with hexokinases (HK2), which acted as control points in glycolysis and was expressed throughout mammary gland development. The HK2 had a functional role in the mammary gland as a consequence of the increased energy production associated with lactation (Kaselonis *et al*., 1999). The FYB was reported to be strong candidate gene for udder health in cattle (Jardim *et al*., 2018). The SEMA3C activated the RAC1 mediated NFKB signaling pathway (Man *et al*., 2014) and regulated uterine remodeling and implantation events (Bagci *et al*., 2014; Tu *et al*., 2016). The CLEC16A, ITPR2 and MKL1 genes were reported to be located in production QTLs regions in cattle genome (Kolbehdari *et al*., 2009, Awad 2011; Shopen *et al*., 2011).

The EPS8 gene acts as a receptor tyrosine kinase substrate for epidermal growth factor receptor and thus increases the signaling response to epidermal growth factor (Fazioli *et al*., 1993) and affects mammary gland development. Previous studies have identified EPS8 as a candidate gene for milk fat percentage in cattle (Wang *et al*., 2012). The NEURL1 gene is reported to be associated with milk yield traits in cattle (Iso-Touru *et al*., 2016). The MTOR is involved in the leptin regulation of Kiss1 expression in the hypothalamus and regulates turn controls ovarian cycling through GnRH release (Roa *et al*., 2010). Gene TOX acts as a transcription factor responsible for the molecular regulation of puberty (Fortes *et al*., 2011). The CSPP1 gene plays a role in spindle formation and cytokinesis (Asiedu *et al*., 2009). Gene CACNA1D is involved in reproductive system development and function (Ortega *et al*., 2017). The CACNB2 gene controls the secretion of FSH (Sugimoto *et al*., 2013). The Sox5 activates the Catsper1 genes and is linked to bull fertility (Mata-Rocha *et al*., 2014). The FSHR gene is associated with female fertility as the influence of FSH hormone on target cells are be mediated by them.

Among the candidate genes ITPR2, TOX and TRAPPC were already reported to be under selection in indicus cattle (Taye *et al*., 2017, Maiorano *et al*., 2018), while CSPP1 was selected in taurine cattle (Zhao *et al*., 2015). Except FSHR, no other major candidate genes affecting

production and fertility in dairy cattle were identified in putative selection sweep regions in the present study.

Table 4.25 Candidate genes for production traits identified in putative selective sweep regions in Sahiwal cattle

Gene	Chromosome	Trait
CLEC16A	25	Fat yield
EPS8	5	Milk yield, Fat percentage
FBP2	8	Milk performance
FYB	20	Protein percentage
IGFBP2	2	Lactation, establishment of pregnancy
IGFBP7	6	Milk yield, 56-day non-return rate, interval from first service to successful insemination (heifer)
ITPR2	5	Fat percentage
LIFR	20	Mammary development pathways, involution pathways, prolactin signalling pathways
MGST1	5	Fat yield, fat percentage
MKL1	5	Milk yield
NEURL1	26	Fat yield
RICTOR	20	Protein percentage
TRAPPC9	14	Test day fat yield (milk fat%)
SEMA3C	4	Number of insemination per conception; days from first to last insemination; 56-day non-return rate; the length in days of the interval from calving to first insemination
SOX5	5	Fertility
GPR125	6	Female fertility index
FSHR	11	Conception rate (heifer), productive life, superovulation response
CACNB2	13	Number of inseminations, 56-day non-return rate, days from first to last insemination, the interval from calving to first insemination
TOX	14	Age at puberty
CSPP1	14	Daughter pregnancy rate
MTOR	16	Reproduction rate (regulation of GnRH release before the initiation of ovarian cycling)
CACNA1D	22	Daughter pregnancy rate

Table 4.26 Candidate genes for tropical adaptation and domestication identified in selective sweep regions in Sahiwal cattle

Gene	Chromosome	Trait
DNAJC3	12	Heat shock response
DNAJC8	2	
PLCB1	13	
TYRP1	8	Brown coat colour
VWA3A	25	Temperament
ZBTB20	1	
NCKAP5	2	
DOCK1	26	
CSMD2	3	Gastro intestinal nematode resistance
RSAD2	11	
ABCC2	26	
HS6ST3	12	
AADAC	1	
PLB1	11	
BST1	6	
SLC6A20	22	
RBP2	1	
ACAD11	1	
COL11A1	3	Tick resistance
COL19A1	9	
COL23A1	7	
COL25A1	6	
COL26A1	25	
COL4A3BP	10	
COL4A4	2	
COL5A1	9	
COL8A1	1	
KRT20	19	
SMIM12	3	
FER	7	
CACFD1	11	
PRKG1	26	

Screening the putative selection sweeps for genes associated with the domestication process revealed genes for coat colour and temperament. The TYRP1 (Tyrosinase Related Protein 1), vWA3A (von Willebrand Factor A Domain Containing 3A), ZBTB20 (Zinc Finger and BTB Domain Containing 20), NCKAP5 (NCK Associated Protein 5), DOCK1 (Dedicator of Cytokinesis 1) genes were found to be located in putative selection sweep regions.

The TYRP1 gene was reported to be linked with brown coat colour in cattle. Coat colour in cattle resulted from the relative presence of eumelanin (black-brown pigment) and phaeomelanin (red-yellow pigment), the two basic pigments produced by melanocyte cells (Adalsteinsson *et al*., 1995). Gene TYRP1 appeared in the melanin synthesis pathway after the branch point between eumelanin and phaeomelanin synthesis (Kobayashi *et al*. 1998) and affected coat colours associated with eumelanin. The TYRP1 might interact and stabilize tyrosinase and decreased its activity in the melanin pathway (Kobayashi *et al*. 1994; Manga *et al*., 2000), and thus, could affect both pigments.

Genes vWA3A, ZBTB20 were found to be associated with reactivity, a phenotype of temperament in Guzerat cattle (Santos *et al*., 2017). Proteins containing von Willebrand domains were involved in basal membrane formation, cell migration, cell differentiation, adhesion, haemostasis, signaling, chromosomal stability. Gene vWA3A was also reported to be differentially expressed in the brain (Marchler-Bauer *et al*., 2012). Ectopic Zbtb20 expression was found to cause behavioral abnormalities in mice (Nielsen *et al*., 2007). The Zbtb20 also had impacts on the development of hippocampus (Rosenthal *et al*., 2012), affecting behavioral traits (Bannerman *et al*., 2004).

Gene NCKAP5 was found to be associated with human attention deficit hyperactivity disorders (Lasky *et al*., 2008) and bipolar disorders (Smith *et al*., 2009) schizophrenia and bipolar disorders (Wang *et al*., 2010). The DOCK1 gene was associated with brain development and played an important role in regulation of axon guidance, dendritic spine morphogenesis in hippocampal neurons (Kim *et al*., 2011). The dendritic spine is the major postsynaptic site of glutamatergic synapses and is thought to be important for synaptogenesis, synaptic regulation, and cognition. Considering the influence of brain system on temperament traits (Robinson, 2001, Tuominen *et al*., 2013), these genes might affect temperament in cattle. Gene DOCK1 and NCKAP5 were found to be associated with temperament in Nellore cattle (Valente *et al*., 2015).

Candidate genes for thermo tolerance and parasitic resistance were identified which is an important aspect of tropical adaptation. Among the heat tolerance genes, DNAJC3 (Dnaj Heat Shock Protein Family (Hsp40) Member C3), DNAJC8 (Dnaj Heat Shock Protein Family (Hsp40) Member C8), PLCB1 (Phospholipase C beta 1) were found to be among the genes under selection. These genes were already reported to be under selection in cattle (Taye *et al*., 2018). The PLCB1 gene was involved in energy metabolism and was reported to be associated with adaptation to hot arid environments in sheep and goats (Kim *et al*., 2016). The HSP40 heat shock protein family genes (DNAJC3, DNAJC8) and PLCB1 identified in this study might contribute to the superior heat tolerance ability of indigenous cattle breeds.

Candidate genes for gastro intestinal nematode resistance (Li *et al*., 2011) like CSMD2 (CUB And Sushi Multiple Domains 2), RSAD2 (Radical S-Adenosyl Methionine Domain Containing 2), ABCC2 (ATP Binding Cassette Subfamily C Member 2), HS6ST3 (Heparan Sulfate 6-O-Sulfotransferase 3), AADAC (Arylacetamide deacetylase (esterase)), PLB1 (Phospholipase B1), BST1 (Bone Marrow Stromal Cell Antigen 1), SLC6A20 (Solute Carrier Family 6 Member 20), RBP2 (Retinol Binding Protein 2), ACAD11 (Acyl-Coa Dehydrogenase Family Member 11) were found to be located in the putative selection sweeps regions identified in the present study. However, further studies are required to identify its role in conferring resistance or susceptibility to nematode resistance.

The SMIM12, FER, CACFD1, PRKG1 genes were reported to be associated with tick resistance in African cattle (Mapholi *et al*., 2016). No clear mechanism of action of these genes on tick resistance is known. The genes associated with tick resistance like SMIM12 (Small Integral Membrane Protein 12), FER (Fps/Fes Related Tyrosine Kinase), CACFD1 (Calcium Channel Flower Domain Containing 1), PRKG1 (Protein Kinase, Cgmp-Dependent, Type I) and collagen and keratin related genes were found in the putative selection sweeps in the present study.

Collagen genes (COL11A1, COL19A1, COL23A1, COL25A1, COL26A1, COL4A3BP, COL4A4, COL5A1, COL8A1) and Keratin associated gene (KRT20) were found to be under selection in Sahiwal cattle. Keratin genes coded for heteropolymeric structural proteins in skin and hair cells and contributed to tick resistance by acting as a barrier to the external environment (Nakamura et al., 2013). Keratinocytes in the epidermis also secreted cytokines that initiate local

inflammatory responses (Nakamura *et al*., 2013). Keratin associated gene was reported to be over represented in tick resistant cattle (Kongsuwan *et al*., 2010). The positive selection of keratin-related genes was previously reported in indicus cattle of Africa (Bahbahani *et al*., 2015; Chan *et al*., 2010; Makina *et al*., 2015; Taye *et al*., 2018). In Brahman cattle structural integrity was found to be enriched in differentially expressed genes in tick resistant cattle (Kongsuwan *et al*., 2008). Collagen genes were found to under selection in indicine cattle (Taye *et al*., 2018). Selection of the parasitic resistance genes may have contributed to better tropical adaptability in Sahiwal cattle.

4.9.2 Gene enrichment analysis

The gene enrichment analysis for biological process revealed that 79 process were enriched (Table 4.27). Most of the enriched processes were associated with brain and nervous system development. This might be due to modifications in neurological development leading to tameness and facilitating human handling of cattle (Jensen, 2014; Jensen and Wright, 2014). Changes in behaviour like reduction in fear and increased sociability are key indicators of animal domestication (Diamond, 2002; Wiener and Wilkinson, 2011). The behavioral fear response was another term which was enriched in the present investigation.

Biological processes like neuronal action potential, regulation of Rho protein signal transduction, actin cytoskeleton reorganization, brain development, establishment of protein localization, lamellipodium assembly were found to be enriched in the present study were also reported by Maiorano *et al*. (2018) in Gir cattle. Biological process like Glutamatergic synaptic transmission and regulation of glutamatergic synaptic transmission which had effect on behavioral adaptation of stress and fear responses were also found to enriched in this study.

The pathway analysis showed wide variety of pathways to be enriched with genes located in selective sweeps regions (Table 4.28). The pathways which were previously reported and proposed to have role in tropical adaptation of cattle have been discussed below:

The MAPK signaling pathway which was reported by Taye *et al*. (2017) to be under selection in cattle was also found in the present study. The MAPK signaling pathways also were found to be involved in residual feed in-take (Rolf *et al*., 2012). The pathways associated with feeding like salivary secretion was also found to enriched. This pathway was reported to be under in selection in Gir cattle (Maiorano *et al*., 2018). Adrenergic signaling in cardiomyocytes

Table 4.27 List of enriched biological process

GO Biological process		Fold Enrichment	Genes
1	GO:0016561~protein import into peroxisome matrix, translocation	14.75	PEX6, PEX5, PEX14
2	GO:0050806~positive regulation of synaptic transmission	10.54	SYT1, SLC1A3, CLSTN2, CLSTN3, GRIK2
3	GO:0070100~negative regulation of chemokine-mediated signaling pathway	8.85	ROBO1, SLIT2, SLIT3
4	GO:2000369~regulation of clathrin-mediated endocytosis	8.85	SNAP91, SMAP1, AAK1
5	GO:0046632~alpha-beta T cell differentiation	8.85	TCF7, LEF1, ABL1
6	GO:0097264~self proteolysis	8.43	TENM4, CAPN7, TENM2, TENM3
7	GO:0043586~tongue development	8.43	NKX2-6, BNC2, LEF1, PRDM16
8	GO:0048268~clathrin coat assembly	8.20	EPS15, SNAP91, CALY, NSG1, HIP1
9	GO:0048714~positive regulation of oligodendrocyte differentiation	7.38	TENM4, RHEB, MTOR
10	GO:0060087~relaxation of vascular smooth muscle	7.38	KCNMA1, GUCY1A3, PRKG1
11	GO:0035385~Roundabout signaling pathway	6.71	ROBO1, EPYC, SLIT1, SLIT2, SLIT3
12	GO:0042327~positive regulation of phosphorylation	6.56	ITGA6, GLMN, PPARGC1B, DSCAM
13	GO:0050861~positive regulation of B cell receptor signaling pathway	6.32	PRKCH, NFAM1, PRKCB
14	GO:0055001~muscle cell development	6.32	SGCG, SGCD, AXIN1

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
15	GO:0033365~protein localization to organelle	6.15	COG3, COG7, CNTLN, CSRP3, BICD1
16	GO:1901379~regulation of potassium ion transmembrane transport	5.90	KCNN2, DPP6, KCNIP1, KCNIP4
17	GO:0051056~regulation of small GTPase mediated signal transduction	5.53	RALGAPA1, SIPA1L1, SIPA1, SIPA1L2, CHN2, RAP1GAP2
18	GO:0060999~positive regulation of dendritic spine development	5.53	NLGN1, NRG1, NEURL1
19	GO:0030516~regulation of axon extension	5.53	ABL1, TTL, DNM2
20	GO:0051639~actin filament network formation	5.53	COBL, PLS1, ACTN1
21	GO:0090129~positive regulation of synapse maturation	5.53	NRXN3, RELN, NEURL1
22	GO:0048846~axon extension involved in axon guidance	5.36	EPYC, SLIT1, SLIT2, SLIT3
23	GO:0050804~modulation of synaptic transmission	5.36	HRH1, NLGN1, PPP3CA, BTBD9
24	GO:0031290~retinal ganglion cell axon guidance	5.21	NRCAM, PTPRM, BMPR1B, SLIT1, EPHB1, SLIT2
25	GO:0031047~gene silencing by RNA	4.92	SND1, ASZ1, PIWIL2, TDRD1
26	GO:0045880~positive regulation of smoothened signaling pathway	4.66	EVC, IFT80, IFT172, PRRX1, CTNNA1, INTU
27	GO:0071320~cellular response to cAMP	4.54	HCN1, EZR, RAP1B, CFTR, AKAP9, RAPGEF2, ITPR2, IGFBP5
28	GO:0007029~endoplasmic reticulum organization	4.43	LMAN1L, ATL2, COL4A3BP, PEX5, JAGN1, VMP1

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
29	GO:0050770~regulation of axonogenesis	4.43	SIPA1L1, LRRC4C, EPYC, SLIT1, SLIT2, SLIT3
30	GO:0002053~positive regulation of mesenchymal cell proliferation	4.34	WNT2, TGFBR2, PRRX1, FOXP1, FOXP2
31	GO:0061028~establishment of endothelial barrier	4.22	EZR, MARVELD2, RAP1B, RAPGEF2
32	**GO:0001662~behavioral fear response**	**4.02**	**ALS2, LYPD1, BRINP1, GRIK2, GRM7, KALRN**
33	GO:0060045~positive regulation of cardiac muscle cell proliferation	3.93	WNT2, ERBB4, MAPK14, ZFPM2
34	**GO:0035249~synaptic transmission, glutamatergic**	**3.93**	**ALS2, CLSTN3, GRIK2, UNC13B**
35	GO:0035269~protein O-linked mannosylation	3.93	ISPD, TMEM5, DPM3
36	GO:0030534~adult behavior	3.88	RNF180, BBS2, GABRG2, GRM7, PCDH17
37	**GO:0030032~lamellipodium assembly**	3.69	ABLIM1, CDH13, NCK2, SH2B1, VAV2, PTPRO, FGD4
38	**GO:0045184~establishment of protein localization**	3.69	NPNT, NLGN1, SMYD3, MCC, ABL1, PHLDB2, USH2A
39	GO:0051209~release of sequestered calcium ion into cytosol	3.69	CHERP, RYR3, RYR1, RYR2, HTR1E, ITPR2
40	GO:0045773~positive regulation of axon extension	3.69	LIMK1, MAP1B, NRG1, CDH4
41	GO:0045907~positive regulation of vasoconstriction	3.69	HRH1, HRH2, OXTR, SMTNL1
42	**GO:0051966~regulation of synaptic transmission, glutamatergic**	**3.69**	**SYT1, GRM3, GRM7, KALRN**

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
43	**GO:0019228~neuronal action potential**	3.35	KCNMA1, GRIK2, SCN9A, SCN8A, SCN5A
44	GO:0040018~positive regulation of multicellular organism growth	3.33	BBS2, EZR, NIPBL, SLC6A3, PLS1, PEX5, CELF1
45	**GO:0031532~actin cytoskeleton reorganization**	3.24	PHACTR1, EZR, EPS8, SIPA1L1, S100A9, ANTXR1, FER, RICTOR, PARVA
46	**GO:0007595~lactation**	**3.21**	**XDH, ERBB4, SLC6A3, OXTR, NEURL1**
47	GO:0021766~hippocampus development	3.13	EPHA5, BBS2, NF2, DCLK2, RELN, XRCC1, KIRREL3
48	GO:0000902~cell morphogenesis	3.12	TENM4, EGFR, LIPA, TENM2, TFCP2L1, COL4A3BP, MAPK14, TENM3, SART3, NRG1, PTPRO
49	GO:0007416~synapse assembly	3.07	CLSTN3, NRXN3, NLGN1, NRG1, KIRREL3
50	GO:0055088~lipid homeostasis	3.05	ACADSB, ACADM, COL4A3BP, ACAD11, ACAD9, ABCA12
51	GO:0043087~regulation of GTPase activity	3.04	EPHA5, SBF2, PLXNB1, CHN2, RICTOR, MTOR, PRKG1
52	GO:0006869~lipid transport	2.95	APOL3, LOC527460, OSBPL6, OSBPL10, OSBPL11, ANO4, ABCA5
53	GO:0086091~regulation of heart rate by cardiac conduction	2.95	CACNA2D1, CACNB2, AKAP9, CACNA1D, CTNNA3
54	GO:0051017~actin filament bundle assembly	2.84	EZR, EPS8, PLS1, ACTN1, PCDH15
55	GO:0050853~B cell receptor signaling pathway	2.84	CD19, NCKAP1L, NFAM 1, ABL1, PRKCB
56	GO:0072659~protein localization to plasma membrane	2.83	FYB, MYO5A, EPB41L3, DENND4C, RAPGEF2, DPP6, RAB10, SKAP1, ABCA12

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
57	GO:0001570~vasculogenesis	2.77	MYO18B, ZMIZ1, TGFBR2, GLMN, HEG1, ZFPM2, HAS2, RASA1, GJC1
58	GO:0006635~fatty acid beta-oxidation	2.77	PPARD, ACADM, EHHADH, PEX5, HIBCH, HSD17B4
59	GO:0008217~regulation of blood pressure	2.60	EDNRA, TRHDE, PPARG, GUCY1A3, ERAP1, CACNA1B
60	GO:0051453~regulation of intracellular pH	2.60	SLC9A9, SLC9A8, SLC4A10, SLC9C1, SLC26A11, SLC4A5
61	GO:0001558~regulation of cell growth	2.46	EPB41L3, CLSTN3, IGFBP7, RB1, IGFBP2, IGFBP5
62	GO:0021987~cerebral cortex development	2.46	BBS2, TRAPPC9, BTBD3, PLCB1, TACC2, FOXP2
63	GO:0007156~homophilic cell adhesion via plasma membrane adhesion molecules	2.43	PTPRM, CLSTN2, CLSTN3, CADM2, CDHR3, DSCAML1, PCDH15, PCDH17, CDH4, CDH12, CDH13, FAT3, CDH16, ROBO1, CDH11, KIRREL3
64	GO:0007605~sensory perception of sound	2.43	KCNMA1, THRB, LRIG1, TMPRSS3, EYA4, SLC1A3, NIPBL, MARVELD2, CLIC5, GRM7, CEMIP, ATP6V0A4, OTOA, CACNA1D, USH2A
65	GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling	2.36	PPARD, ERBB4, PLXNB1, RELN, PDGFC, PDGFD, NRG1, FSHR
66	GO:0006919~activation of cysteine-type endopeptidase activity involved in apoptotic process	2.29	XDH, DLC1, ANP32B, ROBO1, PPARG, S100A9, HIP1
67	GO:0018105~peptidyl-serine phosphorylation	2.11	CSNK1G2, TGFBR2, PRKCH, SMTNL1, ATR, RICTOR, PRKCB, PRKCQ, STK32A, CAMK4, RPS6KA2, MAPK14, DCLK2, MTOR, DCLK1, AKT3

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
68	**GO:0035023~regulation of Rho protein signal transduction**	2.11	ALS2, DLC1, TIAM2, TIAM1, ITSN2, VAV2, SPATA13, KALRN, FGD4, ARHGDIB
69	GO:0007411~axon guidance	2.01	EPHA5, KLF7, CRMP1, ROBO1, ISPD, SEMA3C, CNTN4, RELN, UNC5D, PTPRO, CDH4
70	GO:0098779~mitophagy in response to mitochondrial depolarization	1.98	C5, LMCD1, HK2, DZANK1, ITPKC, ANXA5, PRKG1, KIAA1549L, SLC1A3, TOMM7, SERPINB10, TBC1D5, SNTG1, SLC22A3, CHAF1B
71	GO:0009791~post-embryonic development	1.95	GABRG2, TAPT1, SLC4A10, ACADM, SCN9A, HEG1, SEMA3C, MTOR, FOXP2
72	GO:0042787~protein ubiquitination involved in ubiquitin-dependent protein catabolic process	1.94	RNF144B, HERC4, HERC3, KLHL3, BTBD9, RNF43, AREL1, UBR5, BTBD2, BTBD3, RNF34, RNF122, CUL1, KLHL20, KLHL32
73	**GO:0007420~brain development**	1.93	UTP3, IMMP2L, NIPBL, TGFBR2, IFT172, DSCAML1, RELN, CDK5RAP2, FAS, ZIC1, SPATA5
74	GO:0016192~vesicle-mediated transport	1.80	NCK2, MLC1, STX2, GOSR2, TGFBRAP1, JAGN1, ARFGEF2, RAB10, GGA2, MYO5B, VPS39
75	GO:0007155~cell adhesion	1.79	COL4A4, TLN2, NRXN3, IGFBP7, NLGN1, DSCAML1, FER, NRXN1, CTNNA3, CTNNA2, ITGA9, NCAM2, RGMB, CNTN4, HAS2, RELN, AATF, CNTN3, NTM, DSCAM, PARVA
76	GO:0035556~intracellular signal transduction	1.66	DSTYK, AKAP13, ARFGEF2, MYO9A, PLCL1, RGS11, STK32A, DGKB, TIAM2, GUCY1A3, DCLK2, SH2B1, NRG1, PLCB1, DCLK1, AKT3, PRKCH, DGKH, NFAM1, NEK11, PRKCB, PRKCQ, TULP4, RPS6KA1, CAMK4, RPS6KA2, PSEN2, ASB1, CHN2, UNC13B, SPATA13

Table 4.27 contd…

GO Biological process		Fold Enrichment	Genes
77	GO:0007165~signal transduction	1.55	DLC1, ZNF536, PPP4R1, PPP2R5D, SLC39A12, ARHGAP19, PDE11A, AKAP9, SPOCK1, STARD13, LOC613909, RASAL2, LINGO1, SMOC2, NDRG3, ARHGAP44, COL4A3BP, LANCL2, ARHGAP42, GUCY1A3, UNC5D, SH2B1, RASA1, AXIN1, FYB, PTPRM, TRHDE, LOC788750, OR51E1, ARHGAP24, GABRR2, TENM4, TENM3, OR52E8, CRH, RIN2, AKAP5, TGFBRAP1, INPP4B, INPP4A, ZNF219, RIN3
78	GO:0006915~apoptotic process	1.45	DLC1, RNF144B, OPA1, S100A9, TGFBR2, GAS2, KRT20, RB1, BRAT1, CIAPIN1, PRKCB, CASP6, PRUNE2, ZDHHC16, CASP4, FAS, RNF34, CUL1, FAIM2, MAP2K6
79	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	1.27	RNASEL, THRB, ELF5, ARID4B, PPARG, RORA, SKAP1, WNT2, SPX, NRG1, EGFR, PID1, CTBP2, RELA, CDK7, RB1, ARNTL, LPIN2, PPARGC1B, NCK2, DCAF6, NCOA2, MTF2, ZMIZ1, ZFPM2, ZNF382, CAMTA2, FHL5, NIPBL, AATF, PPP3CA, TCF4, NFATC1, IKZF2, CCPG1, KLF12, ZNF567, MAML2, SMYD3, TET1, CSRP3, FOXP1, TP73, CDH13, ITGA6, RPS6KA1, ETS1, MAPK14, BMPR1B, TCF12

Table 4.28 List of enriched pathways

KEGG pathway		Fold Enrichment	Genes
1	bta05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	4.07	TCF7, CACNA2D1, LEF1, CACNB2, CACNG3, CTNNA1, CACNA2D3, ITGB1, TCF7L1, CTNNA3, ITGA9, SGCG, ITGA6, RYR2, SGCD, CACNA1D
2	bta00120:Primary bile acid biosynthesis	3.89	CYP7B1, CYP46A1, HSD17B4, SCP2
3	bta00534:Glycosaminoglycan biosynthesis - heparan sulfate / heparin	3.44	XYLT1, HS3ST2, HS6ST3, EXT2, HS2ST1
4	bta00640:Propanoate metabolism	3.18	ACADM, SUCLG2, SUCLG1, EHHADH, HIBCH
5	bta02010:ABC transporters	3.15	ABCC9, ABCG5, ABCC1, CFTR, ABCC2, LOC515333, ABCA5, ABCA12
6	bta04520:Adherens junction	2.92	EGFR, PARD3, TCF7, PTPRM, SORBS1, TGFBR2, LMO7, LEF1, ACTN1, CTNNA1, TCF7L1, CTNNA3
7	bta05213:Endometrial cancer	2.92	EGFR, TCF7, GSK3B, LEF1, CTNNA1, TCF7L1, AKT3, CTNNA3, AXIN1
8	bta04360:Axon guidance	2.91	ABLIM1, DCC, PLXNA4, PLXNA1, PLXNB1, LIMK1, LRRC4C, ITGB1, SLIT1, EPHB1, SLIT2, SLIT3, EPHA5, EPHA6, ROBO1, GSK3B, SEMA3D, SEMA3C, UNC5D, PPP3CA, ABL1, RASA1
9	bta04730:Long-term depression	2.80	LOC536367, RYR1, CRH, GUCY1A3, PLA2G4F, PRKG1, PLCB1, PLA2G4E, ITPR2, PRKCB
10	bta04921:Oxytocin signaling pathway	2.62	EGFR, CACNA2D1, ADCY2, OXTR, CACNB2, CACNG3, CACNA2D3, ITPR2, PRKCB, MYL9, CAMKK2, CAMK4, RYR3, RYR1, RYR2, GUCY1A3, PLA2G4F, PPP3CA, PLCB1, CACNA1D, PLA2G4E, NFATC1, CAMK1D
11	bta05414:Dilated cardiomyopathy	2.50	CACNA2D1, ADCY2, MYBPC3, CACNB2, CACNG3, CACNA2D3, ITGB1, ITGA9, SGCG, ITGA6, SGCD, RYR2, CACNA1D
12	bta05410:Hypertrophic cardiomyopathy (HCM)	2.48	ITGA9, CACNA2D1, ITGA6, SGCG, MYBPC3, RYR2, CACNB2, SGCD, CACNG3, CACNA2D3, ITGB1, CACNA1D

Table 4.28 contd…

KEGG pathway		Fold Enrichment	Genes
13	bta05221:Acute myeloid leukemia	2.36	TCF7, PPARD, RELA, SPI1, LEF1, MTOR, TCF7L1, AKT3
14	bta05231:Choline metabolism in cancer	2.34	EGFR, PLD1, DGKH, PRKCB, DGKB, PLA2G4F, SLC22A3, RHEB, PCYT1A, PDGFC, MTOR, PDGFD, PLA2G4E, AKT3
15	bta05215:Prostate cancer	2.31	EGFR, TCF7, RELA, GSK3B, LEF1, PDGFC, CREB5, RB1, PDGFD, MTOR, TCF7L1, AKT3
16	bta04150:mTOR signaling pathway	2.24	RPS6KA1, RPS6KA2, RHEB, RICTOR, MTOR, AKT3, RRAGC, PRKCB
17	**bta04724:Glutamatergic synapse**	2.19	PLD1, ADCY2, GRIK2, ITPR2, PRKCB, GRM3, SLC1A3, GRM8, GNG10, GRM7, PLA2G4F, PPP3CA, PLCB1, CACNA1D, PLA2G4E
18	bta04976:Bile secretion	2.19	FXYD2, ABCG5, ADCY2, KCNN2, EPHX1, CFTR, ABCC2, LOC515333, SLC4A5
19	bta04670:Leukocyte transendothelial migration	2.14	SIPA1, ACTN1, CTNNA1, VAV2, ITGB1, MMP2, CTNNA3, PRKCB, MYL9, EZR, MAPK14, RAP1B, TXK, JAM3, RHOH
20	bta04912:GnRH signaling pathway	2.11	EGFR, PLD1, ADCY2, MAPK14, PLA2G4F, PLCB1, CACNA1D, MMP2, MAP2K6, PLA2G4E, ITPR2
21	bta04024:cAMP signaling pathway	2.09	FXYD2, PLD1, ADCY2, PTGER3, RELA, HTR4, OXTR, CFTR, CREB5, PDE4D, LOC531747, VAV2, FSHR, MYL9, EDNRA, CAMK4, TIAM1, RYR2, RAP1B, LOC515333, CACNA1D, AKT3, LIPE, HTR1E, NFATC1
22	bta04270:Vascular smooth muscle contraction	2.08	KCNMA1, ADCY2, CALD1, PRKCH, PRKG1, ITPR2, PRKCB, MYL9, EDNRA, PRKCQ, GUCY1A3, PLA2G4F, PLCB1, CACNA1D, PLA2G4E
23	**bta04728:Dopaminergic synapse**	2.07	CALY, PPP2R5D, SLC6A3, CREB5, ARNTL, ITPR2, PRKCB, GSK3B, MAPK14, GNG10, PPP3CA, PLCB1, PPP2R2B, CACNA1D, AKT3, CACNA1B
24	bta04713:Circadian entrainment	2.04	ADCY2, GNG10, RYR3, RYR1, RYR2, GUCY1A3, PER3, LOC531747, PRKG1, PLCB1, CACNA1D, PRKCB

Table 4.28 contd…

KEGG pathway		Fold Enrichment	Genes
25	bta04720:Long-term potentiation	2.03	RPS6KA1, CAMK4, RPS6KA2, RAP1B, PPP3CA, PLCB1, ITPR2, PRKCB
26	bta04070:Phosphatidylinositol signaling system	2.02	CDS2, DGKB, PIK3C2G, DGKH, INPP4B, PI4K2B, INPP4A, ITPKC, PIP4K2A, PLCB1, ITPR2, PRKCB
27	bta04015:Rap1 signaling pathway	2.01	PARD3, ADCY2, TLN2, FGF14, SIPA1, FGF12, ITGB1, SKAP1, TIAM1, PDGFC, PDGFD, PLCB1, RAPGEF2, LOC509513, MAP2K6, AKT3, FYB, EGFR, MAGI2, MAGI1, SIPA1L2, DOCK4, PRKCB, SIPA1L1, MAPK14, RAP1B
28	bta04020:Calcium signaling pathway	2.01	EGFR, ADCY2, PTGER3, ERBB4, TACR1, TRHR, HTR4, OXTR, ITPKC, ITPR2, PRKCB, EDNRA, HRH1, CAMK4, HRH2, RYR3, RYR1, RYR2, PPP3CA, LOC509513, PLCB1, CACNA1D, CACNA1B
29	bta04911:Insulin secretion	2.01	KCNMA1, FXYD2, ADCY2, SLC2A2, KCNN2, RYR2, CREB5, PLCB1, CACNA1D, PRKCB
30	**bta04970:Salivary secretion**	1.99	KCNMA1, FXYD2, ADCY2, BST1, RYR3, GUCY1A3, PRKG1, PLCB1, ITPR2, PRKCB
31	bta04722:Neurotrophin signaling pathway	1.98	ARHGDIG, RELA, TP73, NTRK3, CAMK4, RPS6KA1, RPS6KA2, GSK3B, MAPK14, PSEN2, RAP1B, SH2B1, ABL1, AKT3, ARHGDIB
32	bta01200:Carbon metabolism	1.97	ME1, ACADM, SUCLG2, EHHADH, SUCLG1, ESD, HK2, IDH3B, FBP2, GOT1, SDHC, HAO2, HIBCH
33	bta04611:Platelet activation	1.95	ADCY2, TLN2, PRKG1, ITGB1, COL5A1, ITPR2, RASGRP1, MAPK14, GUCY1A3, PLA2G4F, RAP1B, PLCB1, COL11A1, AKT3, PLA2G4E
34	**bta04010:MAPK signaling pathway**	1.94	TRAF2, FGF14, CACNB2, FGF12, RASGRP1, FAS, PPP3CA, RAPGEF2, AKT3, RASA1, MAP2K6, NFATC1, EGFR, CACNA2D1, RELA, TGFBR2, NF1, CACNG3, CACNA2D3, MECOM, PRKCB, MAP4K4, RPS6KA1, RPS6KA2, MAPK14, PLA2G4F, RAP1B, CACNA1D, PLA2G4E, CACNA1B

Table 4.28 contd…

KEGG pathway		Fold Enrichment	Genes
35	bta04514:Cell adhesion molecules (CAMs)	1.93	GLG1, PTPRM, NRXN3, NLGN1, NEO1, LRRC4C, NRXN1, ITGB1, CDH4, LOC509006, NRCAM, ITGA9, NCAM2, ITGA6, CNTNAP2, LOC509513, NEGR1, JAM3
36	bta04152:AMPK signaling pathway	1.93	RAB2A, PPP2R5D, PPARG, CREB5, CFTR, FBP2, CAMKK2, RHEB, MTOR, PPP2R2B, RAB10, TBC1D1, AKT3, LIPE
37	bta04012:ErbB signaling pathway	1.92	EGFR, NCK2, NRG3, ERBB4, GSK3B, MTOR, ABL1, NRG1, AKT3, PRKCB
38	bta04750:Inflammatory mediator regulation of TRP channels	1.91	PRKCQ, HRH1, ADCY2, MAPK14, ASIC2, PRKCH, PLA2G4F, PLCB1, MAP2K6, PLA2G4E, ITPR2, PRKCB
39	bta04512:ECM-receptor interaction	1.90	COL4A4, ITGA9, ITGA6, HSPG2, RELN, LAMC2, SV2B, ITGB1, COL11A1, COL5A1
40	bta04727:GABAergic synapse	1.90	GABRR2, GABRG2, PLCL1, GAD2, GPHN, ADCY2, GNG10, CACNA1D, CACNA1B, PRKCB
41	bta04660:T cell receptor signaling pathway	1.89	PRKCQ, NCK2, MAPK14, RELA, RASGRP1, GSK3B, MALT1, PPP3CA, LOC509513, VAV2, AKT3, NFATC1
42	bta04972:Pancreatic secretion	1.89	KCNMA1, FXYD2, ADCY2, BST1, RYR2, RAP1B, CFTR, PLCB1, CPB2, ITPR2, PRKCB
43	bta00564:Glycerophospholipid metabolism	1.87	CDS2, PLD1, PLA2G16, DGKB, PLB1, PLA2G4F, DGKH, PCYT1A, PTDSS1, LPIN2, PLA2G4E
44	bta04390:Hippo signaling pathway	1.86	FZD9, TCF7, PARD3, LOC505383, NF2, TGFBR2, LEF1, CTNNA1, TP73, TCF7L1, CTNNA3, WNT2, FRMD6, GSK3B, PPP2R2B, BMPR1B, AXIN1
45	**bta04261:Adrenergic signaling in cardiomyocytes**	1.82	FXYD2, CACNA2D1, ADCY2, PPP2R5D, CACNB2, CACNG3, CREB5, CACNA2D3, MAPK14, RYR2, PLCB1, PPP2R2B, CACNA1D, SCN5A, AKT3
46	**bta04310:Wnt signaling pathway**	1.81	FZD9, TCF7, PPARD, CTBP2, LEF1, DAAM1, TCF7L1, PRKCB, WNT2, GSK3B, PPP3CA, PLCB1, CUL1, AXIN1, NFATC1
47	**bta04071:Sphingolipid signaling pathway**	1.79	TRAF2, PLD1, SPTLC3, PPP2R5D, MAPK14, RELA, CERS5, ABCC1, SGMS1, PPP2R2B, PLCB1, AKT3, PRKCB

Table 4.28 contd…

KEGG pathway		Fold Enrichment	Genes
48	bta04723:Retrograde endocannabinoid signaling	1.75	GABRR2, GABRG2, ADCY2, GNG10, ABHD6, MAPK14, PLCB1, CACNA1D, ITPR2, CACNA1B, PRKCB
49	bta04726:Serotonergic synapse	1.71	GNG10, KCNN2, HTR4, PLA2G4F, ALOX5, PLCB1, CACNA1D, PLA2G4E, HTR1E, 50ITPR2, CACNA1B, PRKCB
50	bta04510:Focal adhesion	1.67	EGFR, COL4A4, TLN2, ACTN1, VAV2, ITGB1, COL5A1, PRKCB, MYL9, ITGA9, DOCK1, ITGA6, GSK3B, LAMC2, PDGFC, RAP1B, RELN, PDGFD, COL11A1, AKT3, PARVA
51	bta05200:Pathways in cancer	1.62	DCC, TRAF2, PPARD, ADCY2, FGF14, ARNT2, PPARG, SPI1, FGF12, MMP2, ITGB1, TCF7L1, EDNRA, WNT2, RASGRP1, FAS, PLCB1, AKT3, AXIN1, COL4A4, EGFR, FZD9, TCF7, PTGER3, CTBP2, RELA, TGFBR2, LEF1, RB1, CTNNA1, MECOM, CTNNA3, PRKCB, ITGA6, GNG10, GSK3B, LAMC2, MTOR, ABL1
52	bta04014:Ras signaling pathway	1.62	EGFR, PLD1, PLA2G16, FGF14, RELA, NF1, FGF12, RGL1, PRKCB, RASAL2, TIAM1, ETS1, RASGRP1, GNG10, PLA2G4F, PDGFC, RAP1B, PDGFD, LOC509513, ABL1, AKT3, RASA1, PLA2G4E

which was reported to be under selection in cattle (Edea *et al*., 2018) was also found in the present study.

Sphingolipid signaling pathways which were involved in heat shock response were also found to be under selection in the present study. These pathways activate hydrolysis of sphingomyelin and also their biosynthesis as reported by Jenkins (2003). The two main metabolites of Sphingolipid signaling pathway, the ceramide and sphingosine-1-phosphate played important role in heat stress response. Ceramides signaled cell apoptosis in response to severe heat stress (Jenkins, 2003). The genes in Sphingolipid signaling pathways were found to be enriched in positively selected genes in african indigenous (*Bos indicus*) cattle (Taye *et al*., 2018).

Another pathway which was under selection was Wnt signaling. It was reported to be involved in development of sweat glands and sweating (Cui *et al*., 2014). Heat tolerance mechanisms involved sweat glands through evaporative cooling. (Jian *et al*., 2014). The genes in these two pathways like PLCB1, MAPK14 had been reported to be in involved thermo tolerance. Gene PLCB1 was involved in adaptation to hot arid environments. (Kim *et al*., 2016; Jin *et al*., 2017). Mitogen-activated protein kinases (MAPKs) regulated cellular response with respect to apoptosis (Sugimoto *et al*., 2012). The Wnt signalling pathway were reported to be under selection in indicine and taurine cattle (Gurgul *et al*., 2016; Taye *et al*., 2018).

Dopaminergic and glutamatergic synapse pathways genes were found to be enriched in the present study. Dopamine is an organic chemical of the catecholamine and phenethylamine families that mediates a wide range of brain functions by acting as a neurotransmitter. It was implicated in the process of domestication of animals (Nikulina, 1990). Glutamatergic synapse pathway played a key role in the behavioral adaptation of stress and fear responses (Kamprath *et al*., 2009). These pathways might also be related to domestication, which affected tameness and fear response in domestic animals (Mirkena *et al*., 2010; Zeder, 2012). Glutamate and dopamine regulated foraging and feeding behaviors of animals (Hills, 2006). The dopaminergic system helped free-ranging animals to identify the nutritious and non-poisonous grass/forage species from others (Jensen, 2002; Berthoud, 2007). The pastoral mode of livestock production in the early periods in the country might have lead to the positive selection of these genes involved in above pathways. The findings of the present study highlights the tropical adaptive mechanisms of *Bos indicus* cattle .

CHAPTER –5

Summary and Conclusions

5. Summary and Conclusions

Research in cattle genetics has been profoundly changed since the public release of the cattle genome sequence. Genomics has been adopted as a foundational tool for livestock breeding, health, welfare and conservation. Discovery and genotyping of large number of genome wide markers (SNPs) is a prerequisite for application of genomics in cattle breeding. Scanty information is available on genome wide marker data for indigenous cattle breeds mainly due to the high cost associated with it.

The present study was carried out in 10 Sahiwal cows from Livestock Research Centre, ICAR-National Dairy Research Institute, Karnal (Haryana) and Divya Jyoti Jagrati Sansthan, Noormahal, Jalandhar, Punjab to identify genome-wide SNPs, SSRs and selection sweeps.

Genomic DNA was isolated from blood samples by phenol-chloroform method and the quality and quantity was checked. The genomic DNA samples were sequenced using ddRADseq method. The raw reads obtained were passed through stringent quality control and the processed reads were analyzed through standardized bioinformatics workflows for identification of genome-wide SNPs, SSRs and selection sweeps. The identified variants were annotated structurally and functionally.

Bioinformatics workflows for SNP, SSR and single population selection sweep identification and annotation were standardized for the criteria of minimum computational power and open source tools. Total 450431 and 258231 genome wide SNPs were identified with reference to *Bos taurus* and *Bos indicus* genomes, respectively and 8266 genome wide SSRs were identified in the Sahiwal cattle. About 90% of the SNPs were genotyped in at least half of the samples showing high genotyping efficiency. The novel SNPs with reference to *Bos taurus* were 14908 and 150231 with *Bos indicus* genomes, respectively. The missense to silent ratio was found to be 0.5 to 0.6, while the transition to transversion ratio was 2.3-2.4. The SIFT analysis revealed 89 SNPs to be deleterious affecting the protein structure and function. Total 22762 SNPs were mapped to production trait QTLs while 42314 SNPs were mapped to QTLs associated with reproduction traits. The mastitis QTLs had 5765 SNPs mapped within the QTLs. Tick resistance and heat tolerance QTLs had 7689 and 2300 SNPs.

Among the validation set of 25 SNPs, 22 SNPs were successfully validated. Less than 1% of SNPs identified in the present study in Sahiwal cows were mapped to the existing bovine SNP chips. The selection sweep regions in Sahiwal cattle comprised of 1764 genes. The genes responsible for domestication and tropical adaptation were found to be selected in Sahiwal cattle

5.3 Conclusions:

On the basis of results observed in the present study, the following conclusions were drawn:

- Genome-wide SNPs were identified for the first time in Sahiwal cattle using the *Bos indicus* genome as reference.
- It is the maiden study to identify genome-wide SSRs from sequencing data in Sahiwal cattle.
- Genome-wide novel SNPs were identified in Sahiwal cattle (450431 and 258231 SNPs) with reference to *Bos taurus* and *Bos indicus* genomes respectively.
- The reduced representation approach was found to have higher efficiency for simultaneous discovery and genotyping of SNPs.
- Total 22762 SNPs were mapped to production trait QTLs while 42314 SNPS were mapped to QTLs associated with reproduction traits.
- Lesser than 1% of SNPs identified in the present study on Sahiwal cows were mapped to the existing bovine SNP chips, indicating scope for inclusion of indigenous SNPs in the existing SNP chips for its efficient use in zebu cattle.
- The genes responsible for domestication and tropical adaptation were found to be selected in Sahiwal cattle

Recommendations:

The variants identified in the present study can be used as baseline information to study domestication history, population structure and GWAS. The SNPs mapped to different QTLs and the gene pathways enriched in selective sweeps may be further explored for studying the underlying genetic variation and molecular mechanism of production, reproduction and tropical adaptation traits in indigenous cattle breeds.

Bibliography

Bibliography

Adalsteinsson, S., Bjarnadottir, S., Vage, D.I. and Jonmundsson, J.V., 1995. Brown coat color in Icelandic cattle produced by the loci Extension and Agouti. *The Journal of Heredity*, 86(5), pp.395-398.

Ajayi, O.O., Peters, S.O., De Donato, M., Sowande, S.O., Mujibi, F.D., Morenikeji, O.B., Thomas, B.N., Adeleke, M.A. and Imumorin, I.G., 2018. Computational genome-wide identification of heat shock protein genes in the bovine genome. *F1000Research*, 7.

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Research*, 19, pp. 711–722.

Al Abri, M., Holl, H.M., Miller, D., Abdalla, S., Shykind, B., Malek, Mohamoud, J.Y., Pasha, K., Khalili, K., Antczak, D.F. and Brooks, S., 2017. Proceedings of the 36th international conference on animal genetics, At: Dublin, Ireland Volume: 36.

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L. and Lander, E.S., 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 407, pp. 513–516.

Amaral, A.J., Ferretti, L., Megens, H.J., Crooijmans, R.P.M.A., Nie, H., Ramos-Onsins, S.E., Perez-Enciso, M., Schook, L.B. and Groenen, M.A.M., 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLOS One*, 6, p. e14782.

Andersson, L. and Georges, M., 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics*, 5, pp. 202–212.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T. and Stern, D.L., 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research*, 21, pp. 610–617.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. and Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17, pp. 81–92.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. and Harris, M.A., 2000. Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), p.25.

Asiedu, M., Wu, D., Matsumura, F. and Wei, Q., 2009. centrosome/spindle pole–associated protein regulates cytokinesis via promoting the recruitment of MyoGEF to the central spindle. *Molecular Biology of the Cell*, 20(5), pp.1428-1440.

Asif, A.R., Ijaz, N., Javed, R., Ansari, A.R., Awais, M. and Riaz, H., 2016. Genetic signature of strong recent positive selection at IL-32 gene in goat. *Asian-Australasian Journal of Animal Sciences*, 30(7), p.912.

Awad, A.F.S., 2011. Mapping of quantitative trait loci for milk yield traits on bovine chromosome 5 in the Fleckvieh cattle (Doctoral dissertation, lmu).

Ba, H., Jia, B., Wang, G., Yang, Y., Kedem, G. and Li, C., 2017. Genome-wide snp discovery and analysis of genetic diversity in farmed Sika deer (*Cervus nippon*) in northeast China using double-digest restriction-site associated DNA sequencing. *G3: Genes, Genomes, Genetics*, pp.g3-300082.

Bagci, H., Laurin, M., Huber, J., Muller, W.J. and Cote, J.F., 2014. Impaired cell death and mammary gland involution in the absence of Dock1 and Rac1 signaling. *Cell Death & Disease*, 5(8), p.e1375.

Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M.N., Van Tassell, C., Sonstegard, T., Woolhouse, M. and Hanotte, O., 2015. Signatures of positive selection in East African Shorthorn zebu: A genome-wide single nucleotide polymorphism analysis. *Scientific Reports*, 5, p.11729.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS One*, 3, p. e3376.

Baker, J.R., Cardinal, K., Bober, C., Taylor, M.M. and Samson, W.K., 2003. Neuropeptide W acts in brain to control prolactin, corticosterone, and growth hormone release. *Endocrinology*, 144(7), pp.2816-2821.

Bannerman, D.M., Rawlins, J.N.P., McHugh, S.B., Deacon, R.M.J., Yee, B.K., Bast, T., Zhang, W.N., Pothuizen, H.H.J. and Feldon, J., 2004. Regional dissociations within the

hippocampus-memory and anxiety. *Neuroscience & Biobehavioral Reviews*, 28(3), pp.273-283.

Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Valè, G., Toppino, L. and Rotino, G.L., 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, 12, pp. 1-9.

Barendse, W., 1999. Assessing lipid metabolism. International patent application PCT/AU98/00882, International Patent Publication WO 99/23248.

Barendse, W., Harrison, B.E., Bunch, R.J., Thomas, M.B. and Turner, L.B., 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. *BMC Genomics*, 10, pp. 1-15.

Berthoud, H.R., 2007. Interactions between the "cognitive" and "metabolic" brain in the control of food intake. *Physiology & Behavior*, 91(5), pp.486-498.

Beuzen, N.D., Stear, M.J. and Chang, K.C., 2000. Molecular markers and their use in animal breeding. *The Veterinary Journal*, 160, pp. 42–52.

Bian, Z., Feng, Y., Xue, Y., Hu, Y., Wang, Q., Zhou, L., Liu, Z., Zhang, J., Yin, Y., Gu, B. and Huang, Z., 2016. Down-regulation of SNX1 predicts poor prognosis and contributes to drug resistance in colorectal cancer. *Tumour Biology*, 37(5), pp.6619-6625.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. and San-Cristobal, M., 2010. Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*, 186, pp. 241–262.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32, pp. 314–331.

Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H. and Stephan, W., 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, 140, 783–796.

Brouard, J.S., Boyle, B., Ibeagha-Awemu, E.M. and Bissonnette, N., 2017. Low-depth genotyping-by-sequencing (GBS) in a bovine population: Strategies to maximize the selection of high quality genotypes and the accuracy of imputation. *BMC Genetics*, 18(1), 32.

Brym, P., Kamiñski, S. and Rusc, A., 2004. New SSCP polymorphism within bovine STAT5A gene and its associations with milk performance traits in Black-and-White and Jersey cattle. *Journal of Applied Genetics*, 45(4), pp.445-452.

Bush, W.S. and Moore, J.H., 2012. Genome-wide association studies. *PLOS Computational Biology*, 8, p. e1002822.

Cai, Z., Guldbrandtsen, B., Lund, M.S. and Sahana, G., 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics*, 19(1), p.656.

Calvo, J.H., Martínez-Royo, A., Silveri, L., Floriot, S., Eggen, A., Marcos-Carcavilla, A. and Serrano, M., 2006. Isolation, mapping and identification of SNPs for four genes (ACP6, CGN, ANXA9, SLC27A3) from a bovine QTL region on BTA3. *Cytogenetic and Genome Research*, 114(1), pp.39-43.

Camargo, G.M., Costa, R.B., Lucia, G., Regitano, L.C., Baldi, F. and Tonhati, H., 2015. Polymorphisms in TOX and NCOA2 genes and their associations with reproductive traits in cattle. *Reproduction, Fertility and Development*, 27(3), pp.523-528.

Candy, J.R., Campbell, N.R., Grinnell, M.H., Beacham, T.D., Larson, W.A. and Narum, S.R., 2015. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus, Osmeridae*), an anadromous pacific smelt. *Molecular Ecology Resources*, 15, pp. 1421–1434.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A., 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), pp.3124-3140.

Cavagnaro, P.F., Chung, S.M., Manin, S., Yildiz, M., Ali, A., Alessandro, M.S., Iorizzo, M., Senalik, D.A. and Simon, P.W., 2011. Microsatellite isolation and marker development in carrot - genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across Apiaceae. *BMC Genomics*, 12, pp. 1-20.

Chan, E.K., Nagaraj, S.H. and Reverter, A., 2010. The evolution of tropical adaptation: Comparing taurine and zebu cattle. *Animal Genetics*, 41(5), 467–477.

Chattopadhyay, B., Garg, K.M., Kumar, A.K.V., Doss, D.P.S., Rheindt, F.E., Kandula, S. and Ramakrishnan, U., 2016. Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. *BMC Evolutionary Biology*, 16, 41.

Chen, Q., Ma, Y., Yang, Y., Chen, Z., Liao, R., Xie, X., Wang, Z., He, P., Tu, Y., Zhang, X., Yang, C., Yang, H., Yu, F., Zheng, Y., Zhang, Z., Wang, Q. and Pan, Y., 2013. Genotyping by genome reducing and sequencing for outbred animals. *PLOS One*, 8, p. e67500.

Choi, J.W., Liao, X., Stothard, P., Chung, W.H., Jeon, H.J., Miller, S.P., Choi, S.Y., Lee, J.K., Yang, B., Lee, K.T., Han, K.J., Kim, H.C., Jeong, D., Oh, J.D., Kim, N., Kim, T.H., Lee, H.K. and Lee, S.J., 2014. Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. *PLOS One*, 9, p. e101127.

Church, P.C., Goscinski, A. and Lefevre, C., 2012. EXP-PAC: Providing comparative analysis and storage of next generation gene expression data. *Genomics*, 100(1), pp.8-13.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2), pp.80-92.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. and Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15, pp. 1496–1502.

Clarke, S.M., van Stijn, T.C., Brauning, R. and Dodds, K.G., 2014. Genotyping-by-sequencing (GBS) in sheep: Comparison to SNP chips and whole genome sequencing. *International Plant and Animal Genome Conference XXII*, Volume: P610.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. and Georges, M., 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38, pp. 813–818.

Cobanoglu, O., Zaitoun, I., Chang, Y.M., Shook, G.E. and Khatib, H., 2006. Effects of the Signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *Journal of Dairy Science*, 89, pp. 4433–4437.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.H., Drackley, J.K., Band, M.R., Hernandez, A.G., Shani, M. and Lewin, H.A., 2005.

Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research*, 15(7), pp.936-944.

Cole, J.B., VanRaden, P.M., O'Connell, J.R., Van Tassell, C.P., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. and Wiggans, G.R., 2009. Distribution and location of genetic effects for dairy traits. *Journal of Dairy Science*, 92, pp. 2931–2946.

Cole, J.B., Wiggans, G.R., Ma, L., Sonstegard, T.S., Lawlor, T.J., Crooker, B.A., Van Tassell, C.P., Yang, J., Wang, S., Matukumalli, L.K. and Da, Y., 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics*, 12(1), p.408.

Constant, E., Ibeagha-Awemu, E.M., Miglior, F., Robitaille, G. and Bissonnette, N., 2015. Improving the genotyping-by-sequencing (GBS) approach for the identification of SNPs associated with Johne's disease. In proceedings of ASDA-ASAS, Joint Annual Meeting, July 12-16, Orlando, Florida

Cui, M., Siriwon, N., Li, E., Davidson, E.H. and Peter, I.S., 2014. Wnt signaling functions in embryonic gene networks. *Proceedings of the National Academy of Sciences*, 111 (47), pp e5029-e5038.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.J., Anderson, C., VanTassell, C.P., Hulsegge, I., Goddard, M.E., Guldbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R. and Hayes, B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46, pp. 858–865.

Dalamu, T.K., Gaikwad, A.B., Saxena, S., Bharadwaj, C. and Munshi, A.D., 2012. Morphological and molecular analyses define the genetic diversity of Asian bitter gourd (*Momordica charantia L.*). *Australian Journal of Crop Science*, 6(2), pp .261-267.

Dann, T.H., Habicht, C., Baker, T.T. and Seeb, J.E., 2013. Exploiting genetic diversity to balance conservation and harvest of migratory salmon. *Canadian Journal of Fisheries and Aquatic Sciences*. 70, pp. 785–793.

Das, A., Panitz, F. and Holm, L., 2014. Identification and annotation of genetic variants (SNP/Indel) in Danish Jutland cattle. *In Vancouver: 10th World Congress on Genetics Applied to Livestock Production*, pp. 17-22.

Dash, S., 2016. Genome-wide SNP mining and diversity analysis of indigenous dairy cattle breeds. Ph.D. Thesis, ICAR-National Dairy Research Institute, Karnal, India.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, pp. 499–510.

DeDonato, M.D., Peters, S.O., Mitchell, S.E., Hussain, T. and Imumorin, I.G., 2013. Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLOS One*, 8, p. e62137.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology*, 4(9), p.R60.

Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature*, 418, 700–707.

Dohm, J.C., Lottaz, C., Borodina, T. and Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, 36(16), p.e105.

Dongre, A.B., Raut, M.P., Bhandarkar, M.R. and Meshram, K.J., 2011. Identification and genetic purity testing of cotton F1 hybrid using molecular markers. *Indian Journal of Biotechnology*, 10, pp. 301-306.

Doveri, S., Lee, D., Maheswaran, M., Powell, W., Kole, C. and Abbott, A.G., 2008. Molecular markers-history, features and applications. *Principles and Practices of Plant Genomics, Volume 1: Genome Mapping*, pp.23-67.

Drögemüller, C., Wöhlke, A., Mömke, S. and Distl, O., 2005. Fine mapping of the polled locus to a 1-Mb region on bovine chromosome 1q12. *Mammalian Genome,* 16, pp. 613–620.

Eck, S.H., Benet-Pagès, A., Flisikowski, K., Meitinger, T., Fries, R. and Strom, T.M., 2009. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biology*, 10, pp. 1-8.

Edea, Z., Dadi, H., Dessie, T., Uzzaman, M.R., Rothschild, M.F., Kim, E.S., Sonstegard, T.S. and Kim, K.S., 2018. Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. *Animal Genetics*, 49(6), pp.550-563.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS One*, 6, p. e19379.

Escudero, M., Eaton, D.A.R., Hahn, M. and Hipp, A.L., 2014. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in Carex (Cyperaceae). *Molecular Phylogenetics and Evolution*, 79, pp. 359–367.

Evans, B.J., Zeng, K., Esselstyn, J.A., Charlesworth, B. and Melnick, D.J., 2014. Reduced representation genome sequencing suggests low diversity on the sex chromosomes of Tonkean Macaque monkeys. *Molecular Biology and Evolution*, 31, pp. 2425–2440.

Fang, M., Larson, G., Ribeiro, H.S., Li, N. and Andersson, L., 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLOS Genetics*, 5, p. e1000341.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M. and Servin, B., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193, pp. 929–941.

Fazioli, F., Minichiello, L., Matoska, V., Castagnino, P., Miki, T., Wong, W.T. and Di Fiore, P.P., 1993. Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals. *The EMBO Journal*, 12(10), pp.3799-3808.

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. and Gautier, M., 2009. The genome response to artificial selection: A case study in dairy cattle. *PLOS One*, 4,p. e6595.

Fortes, M.R.S., Reverter, A., Nagaraj, S.H., Zhang, Y., Jonsson, N.N., Barris, W., Lehnert, S., Boe-Hansen, G.B. and Hawken, R. J., 2011. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle *Journal of Animal Science*, 89: 1669-1683.

Fortes, M.R.S., Lehnert, S.A., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N.J., Whan, V., Hawken, R.J. and Reverter, A., 2012. Finding genes for economically important traits: Brahman cattle puberty. *Animal Production Science*, 52(3), pp.143-150.

Gao, Y., Jiang, J., Yang, S., Cao, J., Han, B., Wang, Y., Zhang, Y., Yu, Y., Zhang, S., Zhang, Q. and Fang, L., 2018. Genome-wide association study of *Mycobacterium avium subspecies Paratuberculosis* infection in Chinese Holstein. *BMC Genomics*, 19(1), p.972.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K. and Foulley, J.L., 2009. A whole genome bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, 10, pp. 1-18.

Goddard, M.E. and Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10, pp. 381–391.

Gorjanc, G., Cleveland, M.A., Houston, R.D. and Hickey, J.M., 2015. Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genetics Selection Evolution*, 47, pp. 1-14.

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J.A., Pierson, T., Rogers, S.M., Wilson, J.Y. and Somers, C.M., 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, 15, pp. 1304–1315.

Gregory, T.R., 2009. Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. *Evolution: Education & Outreach*, 2, pp. 5–27.

Greminger, M.P., Stölting, K.N., Nater, A., Goossens, B., Arora, N., Bruggmann, R., Patrignani, A., Nussberger, B., Sharma, R., Kraus, R.H.S., Ambu, L.N., Singleton, I., Chikhi, L., van Schaik, C.P. and Krützen, M., 2014. Generation of SNP datasets for orangutan population genomics using improved reduced-representation sequencing and direct comparisons of SNP calling algorithms. *BMC Genomics*, 15, pp. 1-15.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P. and Spelman, R., 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, 12(2), pp.222-231.

Guo, Y., Yuan, H., Fang, D., Song, L., Liu, Y., Liu, Y., Wu, L., Yu, J., Li, Z., Xu, X., Zhang, H., 2014. An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (*Oryza sativa L.*) F2 population. *BMC Genomics*, 15, pp. 1-13.

Gurgul, A., Szmatoła, T., Ropka-Molik, K., Jasielczuk, I., Pawlina, K., Semik, E. and Bugno-Poniewierska, M., 2016. Identification of genome-wide selection signatures in the L imousin beef cattle breed. *Journal of Animal Breeding and Genetics*, 133(4), pp.264-276

Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, A., Fornal, A., Semik-Gurgul, E. and Bugno-Poniewierska, M., 2019. Genotyping-by-sequencing performance in selected livestock species. *Genomics*.111(2), pp. 186-195.

Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L. and Goddard, M.E., 2009. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Animal Genetics*, 40, pp. 176–184.

Hayes, B.J., Lien, S., Nilsen, H., Olsen, H.G., Berg, P., Maceachern, S., Potter, S., Meuwissen, T.H.E., 2008. The origin of selection signatures on bovine chromosome 6. *Animal Genetics*, 39, pp. 105–111.

Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J. and Goddard, M.E., 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLOS Genetics*, 6, p. e1001139.

He, J., Zhao, X., Laroche, A., Lu, Z.X., Liu, H. and Li, Z., 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, 5. p. 484.

Heffelfinger, C., Fragoso, C.A., Moreno, M.A., Overton, J.D., Mottinger, J.P., Zhao, H., Tohme, J. and Dellaporta, S.L., 2014. Flexible and scalable genotyping-by-sequencing strategies for population studies. *BMC Genomics*, 15, pp. 1-23.

Hermisson, J. and Pennings, P.S., 2005. Soft Sweeps. *Genetics*, 169, pp. 2335–2352.

Heubl, G., 2010. New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Planta Medica*, 76, pp. 1963–1974.

Hills, T.T., 2006. Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, 30(1), pp.3-41.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS*, 106, pp. 9362–9367.

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacy, R.C. and Dasmahapatra, K.K., 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *PNAS*, 111, pp. 3775–3780.

Hohenlohe, P.A., Phillips, P.C. and Cresko, W.A., 2013. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Science*, 171, 1059–1071.

Hosokawa, D., Ishii, A., Yamaji, K., Sasazaki, S., Oyama, K. and Mannen, H., 2012. Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50k SNP array. *Animal Science Journal*, 83, pp. 7–13.

Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., Bishop, S.C., Archibald, A.L., Bron, J.E., Penman, D.J., Davassi, A., Brew, F., Tinch, A.E., Gharbi, K. and Hamilton, A., 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). *BMC Genomics*, 15, pp. 1-13.

Huang, W., Peñagaricano, F., Ahmad, K.R., Lucey, J.A., Weigel, K.A. and Khatib, H., 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. *Journal of Dairy Science*, 95(1), pp.440-449.

Iamartino, D., Strozzi, F., Nicolazzi, E.L., Capoferri, R., Ramelli, P., Pasquariello, R., Stella, A. and Williams, J.L., 2013. Genotyping by sequencing (GBS) applied to the water buffalo. In Proceedings of the 10th ASPA Congress, Bologna, June 11-13 2013.

Ibeagha-Awemu, E.M., Peters, S.O., Akwanji, K.A., Imumorin, I.G. and Zhao, X., 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Scientific Reports*, 6, p.31109.

Iso-Touru, T., Sahana, G., Guldbrandtsen, B., Lund, M.S. and Vilkki, J., 2016. Genome-wide association analysis of milk yield traits in Nordic Red cattle using imputed whole genome sequence variants. *BMC Genetics*, 17(1), p.55.

Jansson, E., Taggart, J.B., Wehner, S., Dahle, G., Quintela, M., Mortensen, S., Kvamme, B.O., Glover, K.A., 2016. Development of SNP and microsatellite markers for

goldsinny wrasse (*Ctenolabrus rupestris*) from ddRAD sequencing data. *Conservation Genetics Resources*, 8, pp. 201–206.

Jardim, J.G., Guldbrandtsen, B., Lund, M.S. and Sahana, G., 2018. Association analysis for udder index and milking speed with imputed whole-genome sequence variants in Nordic Holstein cattle. *Journal of Dairy Science*, 101(3), pp.2199-2212.

Jenkins, G.M., 2003. The emerging role for sphingolipids in the eukaryotic heat shock response. *Cellular and Molecular Life Sciences*, 60, 701–710.

Jensen, P. and Wright, D., 2014. Behavioral genetics and animal domestication. *In Genetics and the Behavior of Domestic Animals* (Second Edition) (pp. 41-79).

Jensen, P., 2014. Behavior genetics and the domestication of animals. *Annual Review of Animal Biosciences*. 2, 85–104.

Jian, W., Duangjinda, M., Vajrabukka, C. and Katawatin, S., 2014. Differences of skin morphology in *Bos indicus*, *Bos taurus*, and their crossbreds. *International Journal of Biometeorology*, 58(6), pp.1087-1094.

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y. and Zhang, Q., 2010. Genome wide association studies for milk production traits in Chinese Holstein population. *PLOS One*, 5(10), p.e13661.

Jiang, Z., Wang, H., Michal, J.J., Zhou, X., Liu, B., Woods, L.C.S. and Fuchs, R.A., 2016. Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development. *International Journal of Biological Sciences*, 12, pp. 100–108.

Jin, Y., Zhou, T., Geng, X., Liu, S., Chen, A., Yao, J. and Liu, Z., 2017. A genome-wide association study of heat stress-associated SNPs in catfish. *Animal Genetics*, 48 (2), 233–236.

Johnson, P.L., McEwan, J.C., Dodds, K.G., Purchas, R.W. and Blair, H.T., 2005. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. *Journal of Animal Science*, 83, pp. 1988-2000.

Johnston, S.E., McEwan, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G., Pemberton, J.M. and Slate, J., 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology Resources*, 20, pp. 2555–2566.

Kamprath, K., Plendl, W., Marsicano, G., Deussing, J.M., Wurst, W., Lutz, B. and Wotjak, C.T., 2009. Endocannabinoids mediate acute fear adaptation via glutamatergic neurons independently of corticotropin releasing hormone signaling. *Genes, Brain and Behavior*, 8(2), pp.203-211.

Kaplan, N.L., Hudson, R.R. and Langley, C.H., 1989. The "hitchhiking effect" revisited. *Genetics*, 123, pp. 887–899.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J.A., Baurain, D., Cambisano, N., Davis, S.R., Farnir, F., Grisart, B. and Harris, B.L., 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics*, 43(5), p.405.

Kaselonis, G.L., McCabe, E.R.B. and Gray, S.M., 1999. Expression of hexokinase 1 and hexokinase 2 in mammary tissue of nonlactating and lactating rats: Evaluation by RT–PCR. *Molecular Genetics And Metabolism*, 68(3), pp.371-374.

Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., Oda, S.I., Ebihara, S., Yajima, S., Yoshikawa, H. and Kono, T., 2011. Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. *BMC Genomics*, 12(1), p.103.

Kgwatalala, P.M., Ibeagha-Awemu, E.M., Mustafa, A.F. and Zhao, X., 2009. Stearoyl-CoA desaturase 1 genotype and stage of lactation influences milk fatty acid composition of Canadian Holstein cows. *Animal Genetics*, 40(5), pp.609-615.

Khatkar, M.S., Thomson, P.C., Tammen, I. and Raadsma, H.W., 2004. Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. *Genetics Selection Evolution*, 36, pp. 1-28.

Kiapekou, E., Zapanti, E., Mastorakos, G. and Loutradis, D., 2010. Update on the role of ovarian corticotropin releasing hormone. *Annals of the New York Academy of Sciences*, 1205(1), pp.225-229.

Kijas, J.M.H., Wales, R., Törnsten, A., Chardon, P., Moller, M. and Andersson, L., 1998. Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics*, 150, pp. 1177–1185.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R.P., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K. and Paiva, S., 2012. Genome-wide analysis of

the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLOS Biology*, 10(2), p.e1001258.

Kim, E.S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B., Sayre, B., Mwacharo, J.M. and Rothschild, M.F., 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*,116(3), p.255.

Kim, J.Y., Bernard, L.P., Macara, I.G. and Zhang, H., 2011. The RhoG/ELMO1/Dock180 signaling module is required for spine morphogenesis in hippocampal neurons. *Journal of Biological Chemistry*, 286(43), pp.37615-37624.

Kirkpatrick, B.W., Shi, X., Shook, G.E. and Collins, M.T., 2011. Whole-genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Animal Genetics*, 42(2), pp.149-160.

Klein, C., Bauersachs, S., Ulbrich, S.E., Einspanier, R., Meyer, H.H., Schmidt, S.E., Reichenbach, H.D., Vermehren, M., Sinowatz, F., Blum, H. and Wolf, E., 2006. Monozygotic twin model reveals novel embryo-induced transcriptome changes of bovine endometrium in the preattachment period. *Biology of Reproduction*, 74(2), pp.253-264.

Kobayashi, T., Urabe, K., Winder, A. J., Tsukamoto, K., Brewington, T., Imokawa, G., Potterf, S. B. and Hearing, V. J., 1994. The DHICA oxidase activity of TRP1 and interactions with other melanogenic enzymes. *Pigment Cell & Melanoma Research*, 7, 227-234.

Kobayashi, T., Imokawa, G., Bennett, D.C. and Hearing, V.J., 1998. Tyrosinase stabilization by Tyrp1 (the brown locus protein). *Journal of Biological Chemistry*, 273(48), pp.31801-31805.

Koks, S., Lilleoja, R., Reimann, E., Salumets, A., Reemann, P., Jaakma, Ü., 2013. Sequencing and annotated analysis of the Holstein cow genome. *Mammalian Genome*, 24, pp. 309–321.

Koks, S., Reimann, E., Lilleoja, R., Lättekivi, F., Salumets, A., Reemann, P., Jaakma, Ü., 2014. Sequencing and annotated analysis of full genome of Holstein breed bull. *Mammalian Genome*, 25, 363–373.

Kolbehdari, D., Wang, Z., Grant, J.R., Murdoch, B., Prasad, A., Xiu, Z., Marques, E., Stothard, P. and Moore, S.S., 2009. A whole genome scan to map QTL for milk

production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics*, 126(3), pp.216-227.

Kongsuwan, K., Piper, E.K., Bagnall, N.H., Ryan, K., Moolhuijzen, P., Bellgard, M., Lew, A., Jackson, L. and Jonsson, N.N., 2008. Identification of genes involved with tick infestation in *Bos taurus* and *Bos indicus*. *In Animal Genomics for Animal Health* (Vol. 132, pp. 77-88). Karger Publishers.

Kongsuwan, K., Josh, P., Colgrave, M.L., Bagnall, N.H., Gough, J., Burns, B. and Pearson, R., 2010. Activation of several key components of the epidermal differentiation pathway in cattle following infestation with the cattle tick, Rhipicephalus (Boophilus) microplus. *International Journal for Parasitology*, 40(4), pp.499-507.

Koopaee, H.K. and Koshkoiyeh, A.E., 2014. SNPs genotyping technologies and their applications in farm animals breeding programs: review. *Brazilian Archives of Biology and Technology*, 57, pp. 87–95.

Kraus, R.H., Kerstens, H.H., Van Hooft, P., Crooijmans, R.P., Van Der Poel, J.J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H.H. and Groenen, M.A., 2011. Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). BMC genomics, 12(1), p.150.

Kujur, A., Bajaj, D., Upadhyaya, H.D., Das, S., Ranjan, R., Shree, T., Saxena, M.S., Badoni, S., Kumar, V., Tripathi, S. and Gowda, C.L.L., 2015. Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Frontiers in Plant Science*, 6, p.162.

Kumar, P., Henikoff, S. and Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), pp.1073-1081.

Kumar, S., Banks, T.W., Cloutier, S., Kumar, S., Banks, T.W. and Cloutier, S., 2012. SNP discovery through next-generation sequencing and its applications. *International Journal of Plant Genomics* p. e831460.

Kutsukake, M., Tamura, K., Yoshie, M. and Tachikawa, E., 2010. Knockdown of IGF-binding protein 7 inhibits transformation of the endometrial gland in an in vitro model. *Molecular Reproduction and Development: Incorporating Gamete Research*, 77(3), pp.265-272.

Kwok, P.Y., 2000. High-throughput genotyping assay approaches. *Pharmacogenomics*, 1, pp. 95–100.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), p.357.

Larson, W.A., Seeb, L.W., Everett, M.V., Waples, R.K., Templin, W.D. and Seeb, J.E., 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*,7, pp. 355–369.

Lasky-Su, J., Neale, B.M., Franke, B., Anney, R.J., Zhou, K., Maller, J.B., Vasquez, A.A., Chen, W., Asherson, P., Buitelaar, J. and Banaschewski, T., 2008. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 147(8), pp.1345-1354.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), pp.2078-2079.

Li, J., Yang, H., Li, J., Li, H., Ning, T., Pan, X.R., Shi, P. And Zhang, Y.P., 2010b. Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. *Heredity*, 105, pp. 274–281.

Li, J.W., Zhou, S.S. and Huang, F., 2012. Microsatellite markers and their application in genetic diversity research of Plasmodium spp. *Chinese Journal of Schistosomiasis Control*, 24(2), pp. 232-234.

Li, M.H., Iso-Touru, T., Laurén, H. and Kantanen, J., 2010a. A microsatellite-based analysis for the detection of selection on BTA1 and BTA20 in northern Eurasian cattle (*Bos taurus*) populations. *Genetics Selection Evolution*, 42, pp. 1-14.

Li, R.W., Rinaldi, M. and Capuco, A.V., 2011. Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. *Veterinary Research*, 42(1), p.114.

Liao, X., Peng, F., Forni, S., McLaren, D., Plastow, G. and Stothard, P., 2013. Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. *Genome*, 56, 592–598.

Litt, M. and Luty, J.A., 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*., 44, pp. 397–401.

MacEachern, S., Hayes, B., McEwan, J. and Goddard, M., 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. *BMC Genomics*, 10, pp. 1-19.

Maiorano, A.M., Lourenco, D.L., Tsuruta, S., Ospina, A.M.T., Stafuzza, N.B., Masuda, Y., Vercesi Filho, A.E., Cyrillo, J.N.D.S.G., Curi, R.A. and de Vasconcelos Silva, J.A.I., 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. *PLOS One*, 13(8), p.e0200694.

Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L. and Maiwashe, A., 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution*, 47(1), 1–14.

Malik, A.A., Sharma, R., Ahlawat, S., Deb, R., Negi, M.S. and Tripathi, S.B., 2018. Analysis of genetic relatedness among Indian cattle (*Bos indicus*) using genotyping by sequencing markers. *Animal Genetics*, 49(3), pp.242-245.

Man, J., Shoemake, J., Zhou, W., Fang, X., Wu, Q., Rizzo, A., Prayson, R., Bao, S., Rich, J.N. and Jennifer, S.Y., 2014. Sema3C promotes the survival and tumorigenicity of glioma stem cells through Rac1 activation. *Cell Reports*, 9(5), pp.1812-1826.

Manga, P., Sato, K., Ye, L., Beermann, F., Lamoreux, M. L. and Orlow, S. J., 2000. Mutational analysis of the modulation of tyrosinase by tyrosinase-related proteins 1 and 2 in vitro. *Pigment Cell & Melanoma Research*. 13, 364-374.

Manigbas, N.L. and Villegas, L.C., 2004. Microsatellite markers in hybridity tests to identify true hybrids of sugarcane. *The Philippine Journal of Crop Science*, 29(2), pp.23-32.

Mapholi, N.O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S.C., MacNeil, M.D., Banga, C., Taylor, J.F. and Dzama, K., 2016. Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks and Tick-Borne Diseases*,7(3), pp.487-497.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J. and Lu, F., 2012. CDD:

conserved domains and protein three-dimensional structure. *Nucleic Acids Research*, 41(D1), pp.D348-D352.

Marete, A., Lund, M.S., Boichard, D. and Ramayo-Caldas, Y., 2018. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. *PLOS One*, 13(7), p.e0199931.

Mata-Rocha, M., Hernández-Sánchez, J., Guarneros, G., de la Chesnaye, E., Sánchez-Tusié, A.A., Treviño, C.L., Felix, R. and Oviedo, N., 2014. The transcription factors Sox5 and Sox9 regulate Catsper1 gene expression. *FEBS Letters*, 588(18), pp.3352-3360.

Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., Law, J., Berntsen, T., Chadha, M., Hui, H., Yang, G., Kennedy, G.C., Webster, T.A., Cawley, S., Walsh, P.S., Jones, K.W., Fodor, S.P.A. and Mei, R., 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods*, 1,pp. 109–111.

Maxam, A.M. and Gilbert, W., 1977. A new method for sequencing DNA. *PNAS* 74, pp. 560–564.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H. and Wilson, J.F., 2008. Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83, pp. 359–372.

Meglecz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N. and Martin, J.F., 2009. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*, 26(3), pp.403-404.

Merkl, M., Ulbrich, S.E., Otzdorff, C., Herbach, N., Wanke, R., Wolf, E., Handler, J. and Bauersachs, S., 2010. Microarray analysis of equine endometrium at days 8 and 12 of pregnancy. *Biology of Reproduction*, 83(5), pp.874-886.

Meuwissen, T.H., Hayes, B.J. and Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, pp. 1819–1829.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. and Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*. 17, pp. 240–248.

Miller, R.D. and Kwok, P.Y., 2001. The birth and death of human single-nucleotide polymorphisms: New experimental evidence and implications for human history and medicine. *Human Molecular Genetics*, 10, pp. 2195–2198.

Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A.,Wurzinger, M., and Sölkner, J., 2010. Genetics of adaptation in domestic farm animals: A review. *Livestock Science,*132(1), 1–12.

Morris, C.A., Cullen, N.G., Glass, B.C., Hyndman, D.L., Manley, T.R., Hickey, S.M., McEwan, J.C., Pitchford, W.S., Bottema, C.D. and Lee, M.A., 2007. Fatty acid synthase effects on bovine adipose fat and milk fat. *Mammalian Genome,* 18(1), pp.64-74.

Mustafa, H., Khan, W.A., Heather, H.J., Kuthu, Z. H., EuiSoo, K., Ajmal, A., Ain, N. U. and Sonstegard, T.S., 2018, Performance of bovine high density SNPs genotyping array in indigenous Pakistani cattle breeds. *Pure and Applied Biology*. 7(1), pp221-226.

Nadeau, N.J., Ruiz, M., Salazar, P., Counterman, B., Medina, J.A., Ortiz-Zuazaga, H., Morrison, A., McMillan, W.O., Jiggins, C.D. and Papa, R., 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research*, 24, pp. 1316–1333.

Nakamura, Y., Kanemarum, K. and Fukami, K., 2013. Physiological functions of phospholipase Cδ1 and phospholipase Cδ3. *Advances in Biological Regulation*,53(3), pp.356-362.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and Et, A., 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, 235, pp. 1616–1622.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. and Hohenlohe, P.A., 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, pp. 2841–2847.

Nayee, N., Sahana, G., Gajjar, S., Sudhakar, A., Trivedi, K., Lund, M.S. and Guldbrandtsen, B., 2018. Suitability of existing commercial single nucleotide polymorphism chips for genomic studies in *Bos indicus* cattle breeds and their *Bos taurus* crosses. *Journal of Animal Breeding and Genetics*, *135*(6), pp.432-441.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. and Markham, A.F., 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research*, 17, pp. 2503–2516.

Nicolazzi, E.L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H., Evans, G., Simpson, B., Tosser-Klopp, G., Brauning, R., Williams, J.L., Stella, A., 2015. SNPchiMp v.3: Integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*, 16, pp. 1-6.

Nielsen, J.V., Nielsen, F.H., Ismail, R., Noraberg, J. and Jensen, N.A., 2007. Hippocampus-like corticoneurogenesis induced by two isoforms of the BTB-zinc finger gene Zbtb20 in mice. *Development*, 134: 1133-1140.

Nielsen, R., 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity*, 86, pp. 641–647.

Nielsen, R., 2005. Molecular signatures of natural selection. *Annual Review of Genetics*, 39, pp. 197–218.

Nielsen, R., Hubisz, M.J. and Clark, A.G., 2004. Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, 168, pp. 2373–2382.

Nikulina, E.M., 1990. The brain catecholamines during domestication of the silver fox Vulpes fulvus. *Zhurnal Evoliutsionnoi Biokhimii i Fiziologii*, 26(2), pp.156-160.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M., 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1), pp.29-34.

Ogorevc, J., Kunej, T. and Dovč, P., 2008. An integrated map of cattle candidate genes for mastitis: A step forward to new genetic markers. *Acta Agriculturae Slovenica*, 2, pp. 85-91.

Ogorevc, J., Kunej, T., Razpet, A. and Dovc, P., 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, 40, pp. 832–851.

Ojeda, A., Huang, L.S., Ren, J., Angiolillo, A., Cho, I.C., Soto, H., Lemús-Flores, C., Makuza, S.M., Folch, J.M., Pérez-Enciso, M., 2008. Selection in the making: A worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. *Genetics*, 178, pp. 1639–1652.

Oleksyk, T.K., Smith, M.W., O'Brien, S.J., 2010. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365, pp.185–205.

Olsen, H.G., Nilsen, H., Hayes, B., Berg, P.R., Svendsen, M., Lien, S. and Meuwissen, T., 2007. Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genetics*, 8, pp.1-9.

Onteru, S.K., Gorbach, D.M., Young, J.M., Garrick, D.J., Dekkers, J.C.M. and Rothschild, M.F., 2013. Whole genome association studies of residual feed intake and related traits in the pig. *PLOS One*, 8, p. e61756.

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K., 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, 5, pp. 874–879.

Ortega, M.S., Denicol, A.C., Cole, J.B., Null, D.J., Taylor, J.F., Schnabel, R.D. and Hansen, P.J., 2017. Association of single nucleotide polymorphisms in candidate genes previously related to genetic variation in fertility with phenotypic measurements of reproductive function in Holstein cows. *Journal of Dairy Science*, 100(5), pp.3725-3734.

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., Ajmone-Marsan, P., Pilla, F., Stefanon, B. and D'Andrea, M., 2018. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. *Journal of Dairy Science*, 101(12), pp.11004-11019.

Pannier, L., Mullen, A.M., Hamill, R.M., Stapleton, P.C. and Sweeney, T., 2010. Association analysis of single nucleotide polymorphisms in DGAT1, TG and FABP4 genes and intramuscular fat in crossbred *Bos taurus* cattle. *Meat Science*, 85, pp. 515–518.

Pavlidis, P., Zivkovic, D., Stamatakis, A. and Alachiotis, N., 2013. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30(9), pp.2224-2234.

Pértille, F., Guerrero-Bosagna, C., Da Silva, V.H., Boschiero, C., da Silva Nunes, J.D.R., Ledur, M.C., Jensen, P. and Coutinho, L.L., 2016. High-throughput and cost-effective chicken genotyping using next-generation sequencing. *Scientific Reports*, 6, p.26929.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS One*, 7, p. e37135.

Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.L., 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS One*, 7, p.e32253.

Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., Rokhsar, D.S. and Rounsley, S., 2012. The cassava genome: current progress, future directions. *Tropical Plant Biology*, 5, pp. 88–94.

Pukk, L., Ahmad, F., Hasan, S., Kisand, V., Gross, R. and Vasemägi, A., 2015. Less is more: extreme genome complexity reduction with ddRAD using ion torrent semiconductor technology. *Molecular Ecology Resources*, 15, pp. 1145–1152.

Qanbari, S. and Simianer, H., 2014. Mapping signatures of positive selection in the genome of livestock. *Livestock Science*., 166, pp. 133–143.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. and Simianer, H., 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*., 12, pp. 1-12.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R., Simianer, H., 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genet*ics, 41, pp. 377–389.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D., Taylor, J.F., 2013. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics*, 14, pp. 1-18.

Raven, L.A., Cocks, B.G. and Hayes, B.J., 2014a. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics*, 15(1), p.62.

Raven, L.A., Cocks, B.G., Goddard, M.E., Pryce, J.E. and Hayes, B.J., 2014b. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. *Genetics Selection Evolution*, 46(1), p.29.

Recknagel, H., Jacobs, A., Herzyk, P. and Elmer, K.R., 2015. Double-digest RAD sequencing using ion proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular Ecology Resources*,15, pp. 1316–1329.

Rhoads, M.L., Meyer, J.P., Kolath, S.J., Lamberson, W.R. and Lucy, M.C., 2008. Growth hormone receptor, insulin-like growth factor (IGF)-1, and IGF-binding protein-2 expression in the reproductive tissues of early postpartum dairy cows. *Journal of Dairy Science*, 91(5), pp.1802-1813.

Roa, J. and Tena-Sempere, M., 2010. Energy balance and puberty onset: Emerging role of central mTOR signaling. *Trends in Endocrinology & Metabolism*, 21(9), pp.519-528.

Robinson, D.L., 2001. How brain arousal systems determine different temperament types and the major dimensions of personality. *Personality and Individual Differences*, 31(8), pp.1233-1259.

Rolf, M.M., Taylor, J.F., Schnabel, R.D., McKay, S.D., McClure, M.C., Northcutt, S.L., Kerley, M.S. and Weaber, R.L., 2012. Genome-wide association analysis for feed efficiency in Angus cattle. *Animal Genetics*, 43(4), pp.367-374.

Ronaghi, M., Uhlén, M. and Nyren, P., 1998. A sequencing method based on real-time pyrophosphate. *Science*, 281(5375), p.363.

Rosenthal, E.H., Tonchev, A.B, Stoykova, A. and Chowdhury, K., 2012. Regulation of archicortical arealization by the transcription factor Zbtb20. *Hippocampus* 22: 2144-2156.

Rothschild, M., Jacobson, C., Vaske, D., Tuggle, C., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren, D., Southwood, O., Steen, H. van der, Mileham, A. and Plastow, G., 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *PNAS*, 93, pp. 201–205.

Rubin, C.J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.M. and Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. *PNAS*, 109, pp. 19529–19536.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010.

Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464, pp. 587–591.

Saarman, N.P. and Pogson, G.H., 2015. Introgression between invasive and native blue mussels (genus Mytilus) in the central California hybrid zone. *Molecular Ecolology*, 24, pp. 4723–4738.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. and Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419, 832–837.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D. and Lander, E.S., 2006. Positive natural selection in the human lineage. *Science*, 312, pp. 1614–1620.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.-F., Galver, L.M., Fan, J.-B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.-Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., Bakker, P.I.W. de, Barrett, J., Chretien, Y.R.,

Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Sham, P.C., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Johnson, T.A., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449, pp.913–918.

Sallam, A.M., Zare, Y., Alpay, F., Shook, G.E., Collins, M.T., Alsheikh, S., Sharaby, M. and Kirkpatrick, B.W., 2017. An across-breed genome wide association analysis of susceptibility to paratuberculosis in dairy cattle. *Journal of Dairy Research*, 84(1), pp.61-67.

Sambrook, J., and Russell, D.W., 2001. Molecular cloning: A laboratory manual (3-volume set). *Immunology*, *49*, pp.895-909.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *PNAS*, 74, pp. 5463–5467.

Santos, F.C., Peixoto, M.G.C.D., de Souza Fonseca, P.A., Pires, M.D.F.A., Ventura, R.V., Rosse, I.D.C., Bruneli, F.A.T., Machado, M.A. and Carvalho, M.R.S., 2017. Identification of candidate genes for reactivity in Guzerat (*Bos indicus*) cattle: A Genome-wide association Study. *PLOS One*, 12(1), p.e0169163.

Sarkar, G., Cassady, J., Bottema, C.D.K. and Sommer, S.S., 1990. Characterization of polymerase chain reaction amplification of specific alleles. *Analytical Biochemistry*, 186, pp. 64–68.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A. and Gemmell, N., 2013. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLOS One*, 8, p. e54710.

Schennink, A., Bovenhuis, H., Léon-Kloosterziel, K.M., Van Arendonk, J.A. and Visker, M.H., 2009. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk fat composition. *Animal Genetics*, 40(6), pp.909-916.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H. and Castoe, T.A., 2016. EpiRADseq: Scalable analysis of genome-wide patterns of methylation using next generation sequencing. *Methods in Ecology and Evolution*, 7, pp. 60–69.

Schlötterer, C. and Tautz, D., 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20, pp. 211–215.

Schlötterer, C., 2003. Hitchhiking mapping – functional genomics from the population genetics perspective. *Trends in Genetics*, 19, pp. 32–38.

Schmieder, R. and Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), pp.863-864.

Schopen, G. C., M. H. Visker, P. D. Koks, E. Mullaart, J. A. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. *Journal of Dairy Science*. 94:3148–3158..

Schwarzenbacher, H., Dolezal, M., Flisikowski, K., Seefried, F., Wurmser, C., Schlötterer, C. And Fries, R., 2012. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. *BMC Genomics*, 13, pp. 1-19.

Senan, S., Kizhakayil, D., Sasikumar, B. And Sheeja, T.E., 2014. Methods for development of microsatellite markers: An overview. *Notulae Scientia Biologicae*, 6, pp. 1–13.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309, 1728–1732.

Short, T.H., Rothschild, M.F., Southwood, O.I., McLaren, D.G., de Vries, A., van der Steen, H., Eckardt, G.R., Tuggle, C.K., Helm, J., Vaske, D.A., Mileham, A.J. and Plastow, G.S., 1997. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. *Journal of Animal Science*, 75, pp. 3138-3142.

Smith, E.N., Bloss, C.S., Badner, J.A., Barrett, T., Belmonte, P.L., Berrettini, W., Byerley, W., Coryell, W., Craig, D., Edenberg, H.J. and Eskin, E., 2009. Genome-wide association study of bipolar disorder in European American and African American individuals. *Molecular Psychiatry*, 14(8), p.755-763.

Smith, J.M., and Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetics Research*, 23, pp. 23–35.

Sodeland, M., Grove, H., Kent, M., Taylor, S., Svendsen, M., Hayes, B.J. and Lien, S., 2011. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. *BMC Genetics*, 12, pp.1-10.

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I. and Belzile, F., 2015. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, 13, pp.211–221.

Stafuzza, N.B., Zerlotini, A., Lobo, F.P., Yamagishi, M.E.B., Chud, T.C.S., Caetano, A.R., Munari, D.P., Garrick, D.J., Machado, M.A., Martins, M.F. and Carvalho, M.R., 2017. Single nucleotide variants and InDels identified from whole-genome re-sequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. *PLOS One*, 12(3), p.e0173954.

Steemers, F.J. and Gunderson, K.L., 2007. Whole genome genotyping technologies on the BeadArray platform. *Plant Biotechnology*., 2, pp. 41–49.

Stella, A., Ajmone-Marsan, P., Lazzari, B. and Boettcher, P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*, 185, pp. 1451–1461.

Stigliani, S., Anserini, P., Nicoletti, A.J., Di Luca, M., Sozzi, F., Venturini, P.L. and Scaruffi, P., 2013. Insight into the genomics of premature ovarian failure. *Journal of Molecular and Genetic Medicine*, 7, p.7.

Stolle, E. and Moritz, R.F.A., 2013. RESTseq--efficient benchtop population genomics with RESTriction Fragment SEQuencing. *PLOS One*, 8, p. e63960.

Stothard, P., Choi, J.W., Basu, U., Sumner-Thomson, J.M., Meng, Y., Liao, X. and Moore, S.S., 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*, 12, pp. 1-14.

Sugimoto, M., Sasaki, S., Gotoh, Y., Nakamura, Y., Aoyagi, Y., Kawahara, T. and Sugimoto, Y., 2013. Genetic variants related to gap junctions and hormone secretion influence conception rates in cows. *Proceedings of the National Academy of Sciences*, 110(48), pp.19495-19500.

Sugimoto, N., Shido, O., Matsuzaki, K., Ohno-Shosaku, T., Hitomi, Y., Tanaka, M., Sawaki, T., Fujita, Y., Kawanami, T., Masaki, Y. and Okazaki, T., 2012. Cellular heat acclimation regulates cell growth, cell morphology, mitogen-activated protein kinase activation, and expression of aquaporins in mouse fibroblast cells. *Cellular Physiology and Biochemistry*,30(2), pp.450-457.

Surya, T., Vineeth, M.R., Sivalingam, J., Tantia, M.S., Dixit, S.P., Niranjan, S.K. and Gupta, I.D., 2018. Genome-wide identification and annotation of SNPs in *Bubalus bubalis*. *Genomics.* In Press

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), pp.585-595.

Tan, C., Wu, Z., Ren, J., Huang, Z., Liu, D., He, X., Prakapenka, D., Zhang, R., Li, N., Da, Y. and Hu, X., 2017. Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing. *Genetics Selection Evolution*, 49(1), p.35.

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17, pp. 6463–6471.

Taye, M., Kim, J., Yoon, S.H., Lee, W., Hanotte, O., Dessie, T., Kemp, S., Mwai, O.A., Caetano-Anolles, K., Cho, S. and Oh, S.J., 2017. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genetics*, 18(1), p.11.

Taye, M., Lee, W., Caetano-Anolles, K., Dessie, T., Cho, S., Jong Oh, S., Lee, H.K. and Kim, H., 2018. Exploring the genomes of East African indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits. *Cogent Food & Agriculture*, 4(1), pp.1-16.

The Bovine HapMap Consortium, 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324, pp. 528–532.

Toonen, R.J., Puritz, J.B., Forsman, Z.H., Whitney, J.L., Fernandez-Silva, I., Andrews, K.R., and Bird, C.E., 2013. ezRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, p. e203.

Truong, H.T., Ramos, A.M., Yalcin, F., de Ruiter, M., van der Poel, H.J.A., Huvenaars, K.H.J., Hogers, R.C.J., van Enckevort, L.J.G., Janssen, A., van Orsouw, N.J. and van Eijk, M.J.T., 2012. Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. *PLOS One*, 7, p. e37565.

Tu, Z., Wang, Q., Cui, T., Wang, J., Ran, H., Bao, H., Lu, J., Wang, B., Lydon, J.P., DeMayo, F. and Zhang, S., 2016. Uterine RAC1 via Pak1-ERM signaling directs normal luminal epithelial integrity conducive to on-time embryo implantation in mice. *Cell Death and Differentiation*, 23(1), p.169.

Tuominen, L., Salo, J., Hirvonen, J., Någren, K., Laine, P., Melartin, T., Isometsä, E., Viikari, J., Cloninger, C.R., Raitakari, O. and Hietala, J., 2013. Temperament, character and serotonin activity in the human brain: A positron emission tomography study based on a general population cohort. *Psychological Medicine*, 43(4), pp.881-894.

Turcatti, G., Romieu, A., Fedurco, M., Tairi, A.-P., 2008. A new class of cleavable fluorescent nucleotides: Synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research.*, 36, p. e25–e25.

Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Research*, 40(15): e115.

Valente, T.S., Sant'Anna, A.C., Baldi, F., Albuquerque, L.G. and da Costa, M.J.P., 2015. Genetic association between temperament and sexual precocity indicator traits in Nellore cattle. *Journal of Applied Genetics*, 56(3), pp.349-354.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G. and Georges, M., Andersson, L., 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, 425, pp. 832–836.

Van Orsouw, N.J., Hogers, R.C.J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., Schneiders, H., Poel, H. van der, Oeveren, J. van, Verstegen, H. and Eijk, M.J.T., 2007. Complexity Reduction of Polymorphic Sequences (CRoPS^{TM}): A novel approach for large-scale polymorphism discovery in complex genomes. *PLOS One*, 2, p. e1172.

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. and Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5, pp. 247–252.

Vartia, S., Villanueva-Cañas, J.L., Finarelli, J., Farrell, E.D., Collins, P.C., Hughes, G.M., Carlsson, J.E.L., Gauthier, D.T., McGinnity, P., Cross, T.F., FitzGerald, R.D., Mirimin, L., Crispie, F., Cotter, P.D. and Carlsson, J., 2016. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society Open Science*, 3.

Veneroni, G.B., Meirelles, S.L., Grossi, D.A., Gasparin, G., Ibelli, A.M.G., Tizioto, P.C., Oliveira, H.N., Alencar, M.M. and Regitano, L.C.A., 2010. Prospecting candidate SNPs for backfat in Canchim beef cattle. *Genetics and Molecular Research*, 9, pp. 1997–2003.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), pp.1-31.

Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K., 2006. A map of recent positive selection in the human genome. *PLOS Biology*, 4, p. e72.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*., 23, pp. 4407–4414.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. And Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, pp. 787–798.

Wang, E.T., Kodama, G., Baldi, P. and Moyzis, R.K., 2006. Global landscape of recent inferred darwinian selection for *Homo sapiens*. *PNAS*, 103, pp. 135–140.

Wang, K.S., Liu, X.F. and Aragam, N., 2010. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophrenia Research*, 124(1-3), pp.192-199.

Wang, S., Meyer, E., McKay, J.K., Matz, M.V., 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, 9, pp. 808–810.

Wang, W., Gan, J., Fang, D., Tang, H., Wang, H., Yi, J. and Fu, M., 2018. Genome-wide SNP discovery and evaluation of genetic diversity among six Chinese indigenous cattle breeds in Sichuan. *PLOS One*, 13(8), p.e0201534.

Wang, X., Wurmser, C., Pausch, H., Jung, S., Reinhardt, F., Tetens, J., Thaller, G. and Fries, R., 2012. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. *PLOS One*, 7(7):e40711.

Weber, J.L. and May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, 44, pp. 388–396.

Weikard R., Kuhn C., Goldammer T., Freyer G. and Schwerin M., 2005. The bovine PPARGC1A gene: Molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiological Genomics*, 21, 1–13.

Weikard, R., Goldammer, T., Brunner, R.M. and Kuehn, C., 2012. Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiological Genomics*, 44(14), pp.728-739.

Weikard, R., Widmann, P., Buitkamp, J., Emmerling, R. And Kuehn, C., 2012. Revisiting the quantitative trait loci for milk production traits on BTA6. *Animal Genetics*, 43, pp. 318–323.

Werth, L.A., Hawkins, G.A., Eggen, A., Petit, E., Elduque, C., Kreigesmann, B. and Bishop, M.D., 1996. Rapid communication: melanocyte stimulating hormone receptor (MC1R) maps to bovine chromosome 18. *Journal of Animal Science*, 74, pp. 262–262.

Wheeler, M., Dodds, K.G., Brauning, R., Clarke, S.M., Van Stijn, T.C., Foote, J.F., Cameron, A.W.N. and McEwan, J.C., 2018. Genotyping-by-sequencing for genetic improvement in dairy goats. *In Proceedings of the 11th World Congress on Genetics Applied to Livestock Production*, 11. p 358

Wiener, P. and Gutiérrez-Gil, B., 2009. Assessment of selection mapping near the myostatin gene (GDF-8) in cattle. *Animal Genetics*, 40, pp. 598–608.

Wiener, P. and Wilkinson, S., 2011. Deciphering the genetic basis of animal domestication. *Proceedings of the Royal Society B: Biological Sciences*, 278(1722), pp.3161-3170.

Wiener, P., Burton, D., Ajmone-Marsan, P., Dunner, S., Mommens, G., Nijman, I.J., Rodellar, C., Valentini, A., Williams, J.L., 2003. Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. *Heredity*, 90, pp. 350–358.

Wiener, P., Edriss, M.A., Williams, J.L., Waddington, D., Law, A., Woolliams, J.A. and Gutiérrez-Gil, B., 2011. Information content in genome-wide scans: Concordance between patterns of genetic differentiation and linkage mapping associations. *BMC Genomics*, 12, pp. 1-10.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*., 18, pp. 6531–6535.

Wright, S., 1949. The genetical structure of populations. *Annals of Eugenics*, 15(1), pp.323-354.

Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B., 1989. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National Academy of Sciences*, 86(8), pp. 2757-2760.

Xu, L., Haasl, R.J., Sun, J., Zhou, Y., Bickhart, D.M., Li, J., Song, J., Sonstegard, T.S., Van Tassell, C.P., Lewin, H.A. and Liu, G.E., 2017. Systematic profiling of short tandem repeats in the cattle genome. *Genome Biology And Evolution*, 9(1), pp.20-31.

Yan, Q., Zhang, Y., Li, H., Wei, C., Niu, L., Guan, S., Li, S. and Du, L., 2008. Identification of microsatellites in cattle unigenes. *Journal of Genetics And Genomics*, *35*(5), pp.261-266.

Yang, J., Li, Y., Zhu, S., Chen, W., Li, J., Xue, H. and Li, X., 2018. Development and characterization of 26 SNP markers in *Ochetobius elongatus* based on restriction site-associated DNA sequencing (RAD-seq). *Conservation Genetics Resources*, pp.1-3.

Ye, S., Dhillon, S., Ke, X., Collins, A.R. and Day, I.N.M., 2001. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*., 29, p.e88.

Zeder, M. A., 2012. The domestication of animals. *Journal of Anthropological Research*, 68(2), 161–190.

Zhai, Z., Zhao, W., He, C., Yang, K., Tang, L., Liu, S., Zhang, Y., Huang, Q. and Meng, H., 2015. SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens. *Animal Genetics*, 46(2), pp.216-219.

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F. and Bendixen, C., 2011. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*, 12(1), pp.1-20.

Zhang, L., Michal, J.J., O'Fallon, J.V., Pan, Z., Gaskins, C.T., Reeves, J.J., Busboom, J.R., Zhou, X., Ding, B., Dodson, M.V. and Jiang, Z., 2012. Quantitative genomics of 30 complex phenotypes in Wagyu x Angus F 1 progeny. *International Journal of Biological Sciences*, 8, 838–858.

Zhao, F., McParland, S., Kearney, F., Du, L. and Berry, D.P., 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution*, 47(1), p.49.

Zhu, F., Cui, Q.Q,. and Hou, Z.C., 2016. SNP discovery and genotyping using genotyping-by-sequencing in Pekin Ducks. *Scientific Reports*, 6:36223.

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S. and Marçais, G., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology*, 10(4), p.R42.

Appendix- I & II

Appendix- I

Details of the missense SNPs identified in Sahiwal cattle using *Bos taurus* as reference genome

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
1	4893671	rs210799260	A	G	KRTAP27-1
1	6530487	rs43216839	C	G	RWDD2B
1	53295811	rs722584349	C	T	IFT57
1	56164129	rs1114914336	A	G	ENSBTAG00000038651
1	58593684	rs208584543	T	C	SPICE1
1	58593817	rs209696343	C	T	SPICE1
1	58768744	rs207583037	G	A	SIDT1,SIDT1
1	66575981	rs109032777	C	T	POLQ
1	67095399	rs379269630	T	C	ILDR1
1	71525073	rs523262845	G	A	ENSBTAG00000040161
1	74601989	rs525616158	C	T	ATP13A4
1	78694914	rs720790594	T	C	TPRG1
1	109823001	rs110509495	T	C	RSRC1
1	112997870	rs108998256	C	G	PLCH1,PLCH1
1	112997925	rs136063343	C	T	PLCH1,PLCH1
1	112997928	rs134587613	C	T	PLCH1,PLCH1
1	112997930	rs135749474	C	G	PLCH1,PLCH1
1	112997932	rs133938613	C	G	PLCH1,PLCH1
1	112997933	rs137306906	C	A	PLCH1,PLCH1
1	119938805	rs522544402	C	T	CP
1	136656325	-	C	T	SRPRB
1	145688456	rs208050383	T	C	ENSBTAG00000014880
1	147694619	rs134567976	G	C,A	MCM3AP
1	147797980	rs136469092	G	A	PCNT
1	147894602	rs719383767	G	A	DIP2A
1	147894635	rs43251562	A	C	DIP2A
1	147894665	rs208601955	G	A	DIP2A
2	1247411	rs523088580	A	C	ENSBTAG00000015841

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
2	7343971	rs451388817	C	T	COL3A1
2	18295356	rs109471436	A	G	ENSBTAG00000026986
2	27060317	rs720881676	A	T	LRP2
2	27407988	rs527121912	C	A	CERS6
2	89559731	rs211188056	T	C	AOX1
2	97026129	rs137828325	C	T	PIKFYVE
2	104541249	-	C	T	MREG
2	122571644	rs453049058	C	T	COL16A1
2	122571653	rs516327598	C	T	COL16A1,COL16A1
2	127267002	rs482056952	G	A	DHDDS
2	127504199	rs516044852	T	C	ENSBTAG00000047148,
3	251568	rs716467195	C	T	TBX19
3	251575	rs719694580	A	G	TBX19
3	9119516	rs440362324	T	C	SLAMF1
3	9119568	-	C	T	SLAMF1
3	10462519	rs209571442	T	C	ENSBTAG00000020765
3	10462538	rs110234269	T	C	ENSBTAG00000020765
3	11195202	rs517634532	G	A	SPTA1
3	16487733	rs518421517	G	A	NUP210L
3	21824834	rs720713184	G	A	GPR89,GPR89
3	21824840	rs458141513	A	G	GPR89,GPR89
3	25301371	rs381014304	G	A	SPAG17
3	26319687	rs207563521	G	A	CD101
3	49470802	-	G	A	ARHGAP29
3	80974075	rs522204299	T	G	RAVER2
3	85136688	-	G	T	NFIA
3	86291572	rs383424493	G	T	C1orf87
3	90001415	rs482696022	C	T	C1orf168
3	98577879	rs525440995	G	A	SLC5A9
3	106497583	rs43357656	G	A	RLF

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
3	108631323	-	C	T	INPP5B
4	7433857	rs134409316	G	A	ABCA13
4	7433954	rs136573701	T	C	ABCA13
4	7434014	rs135076666	A	C	ABCA13
4	7434067	rs723805103	C	T	ABCA13
4	9225923	rs516436203	G	C	AKAP9
4	9675995	rs446165914	C	T	PEX1
4	9676053	rs714017693	T	C	PEX1
4	9728061	rs207744653	G	A	RBM48,PEX1
4	9728079	rs437174559	C	T	RBM48,PEX1
4	9728088	rs455873222	G	A	RBM48,PEX1
4	9728094	rs474297785	G	T	RBM48,PEX1
4	28836005	rs434343925	G	A	MACC1
4	40433074	rs433466232	G	A	ENSBTAG00000047646
4	47372006	rs522648540	C	T	CDHR3
4	49392575	rs208032372	T	C	LAMB4
4	49392590	-	A	G	LAMB4
4	49392659	rs210656974	A	G	LAMB4
4	50899209	rs466271649	A	G	CTTNBP2
4	77882404	-	T	G	AEBP1,AEBP1,POLD2
4	78821597	-	G	A	MRPL32
4	87245518	rs134309914	A	G	PTPRZ1
4	88622400	rs439919935	T	C	IQUB,U7
4	94943415	rs1117301205	C	T	CPA1
4	99773325	rs520518297	T	C	STRA8,STRA8
4	103379637	rs209768289	A	G	KIAA1549
4	104352147	rs381492870	C	T	PARP12
4	104374422	rs110475969	T	G	KDM7A
4	104374473	rs519709482	C	T	KDM7A
4	105911646	rs211096605	A	G	PRSS37
4	107345117	rs481269201	A	G	ENSBTAG00000031212

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
5	4769085	rs452899825	T	C	CAPS2
5	4798432	rs517747990	G	A	CAPS2
5	27488999	rs1114473708	C	T	KRT74,KRT74
5	32106221	rs715400649	C	T	ZNF641
5	49679650	rs716381934	A	T	C12orf56
5	57244998	rs210561380	G	C	GLS2
5	57329238	rs209517641	G	A	STAT2,IL23A
5	63198645	rs719079404	C	G	APAF1
5	71283368	rs210165018	C	G	PWP1
5	83238202	rs381468431	G	A	TM7SF3
5	101161778	rs719553197	A	C	ENSBTAG00000005243
5	101161779	rs724024072	C	A	ENSBTAG00000005243
5	101543683	rs109099432	C	T	A2ML1,bta-mir-2284r
5	104223841	rs17871972	C	T	NOP2,IFFO1
5	105563512	rs134295224	C	T	ENSBTAG00000026522
5	105563556	rs716828138	T	C	ENSBTAG00000026522
5	118201831	rs721766648	A	G	TBC1D22A
5	119672924	rs210410555	T	A	ENSBTAG00000030185
5	119673026	rs207609316	T	C	ENSBTAG00000030185
5	119673034	rs1116140705	G	A	ENSBTAG00000030185
6	14290917	rs109070284	T	C	ALPK1
6	14290968	rs717902187	G	A	ALPK1
6	20269577	rs723187158	G	A	TBCK
6	20454074	rs386120488	T	A	NPNT
6	68332109	rs110593127	T	C	CNGA1
6	90654038	rs517796311	G	A	ENSBTAG00000024080
6	92557322	rs384936737	G	A	SDAD1
6	92809913	rs135580435	C	G	SCARB2
6	94869241	rs523452194	A	G	FRAS1
6	99459354	rs210185307	C	T	SEC31A
6	99533404	rs478018056	A	C	LIN54

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
6	103192943	rs137737610	G	A	ENSBTAG00000045839
6	103192952	rs723709512	C	T	ENSBTAG00000045839
6	105329471	rs517378271	T	A	EVC2
6	110478191	rs43489014	G	C	CLNK
6	115987828	rs42164968	G	A	PROM1
6	115987834	rs42164969	T	C	PROM1
6	116957949	rs523311247	G	A	ENSBTAG00000045917,LDB2
6	117749231	rs719474953	G	T	CLRN2,QDPR
7	3165828	rs720750757	C	T	PRSS38
7	3340531	rs136817567	C	T	SNAP47
7	6050080	rs715681396	T	A	CPAMD8
7	8399378	rs521379793	G	A	ENSBTAG00000039337
7	14353679	rs110670316	T	C	ENSBTAG00000037693
7	14644791	rs133734929	A	G	ENSBTAG00000038714
7	17760838	rs721210187	G	A	FCER2
7	19579296	rs720716740	C	T	RFX2
7	22110828	rs451955214	G	C	ZNF554
7	30447005	rs721896622	A	G	ZNF608
7	40574703	rs457755236	G	A	N4BP3
7	43970660	rs516759789	G	A	ZNF692 ,ZNF672
7	48921837	rs208209126	C	T	IL9
7	53501998	rs109574176	A	T	HARS2,HARS
7	55024761	-	C	T	ENSBTAG00000016789
7	55024762	rs441490904	A	G	ENSBTAG00000016789
7	55024766	-	A	G	ENSBTAG00000016789
7	55024772	rs460028536	C	T	ENSBTAG00000016789
7	55024786	rs478540972	C	T	ENSBTAG00000016789
7	55024790	-	C	T	ENSBTAG00000016789
7	55024791	-	A	T	ENSBTAG00000016789
7	55024796	rs721081163	C	G	ENSBTAG00000016789

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
7	55024856	rs721902151	T	C	ENSBTAG00000016789
7	55024916	rs445235991	G	A	ENSBTAG00000016789
7	62409071	rs517433759	A	G	SH3TC2
7	73928645	rs520302563	C	T	CCNJL
7	73928659	rs380876537	T	C	CCNJL
7	82835870	rs724004974	G	A	FBLL1
7	105009784	rs723063910	G	T	ENSBTAG00000027727
8	8629191	rs455248987	C	T	ENSBTAG00000040038
8	38045382	rs208711111	G	A	ENSBTAG00000043987
8	45696078	rs43547009	A	G	ENSBTAG00000011770
8	51149805	rs109415375	C	G	TRPM6
8	51149806	rs460641549	C	G	TRPM6
8	51149816	rs479080650	A	T	TRPM6
8	51149817	rs723947908	G	C	TRPM6
8	51149932	rs43557145	G	T	TRPM6
8	58238692	rs136484794	A	G	ENSBTAG00000005667
8	60093100	rs134136586	A	G	ENSBTAG00000011402
8	60114140	rs137122367	G	A	ENSBTAG00000011402
8	62436746	rs459850209	C	T	ENSBTAG00000014936
8	66592991	rs137621244	A	G	ENSBTAG00000006446
8	66593045	rs134851176	T	C	ENSBTAG00000006446
8	82246236	rs522003708	C	T	DAPK1
8	85659461	rs523687170	C	A	IPPK
8	87006885	rs209569108	C	T	ENSBTAG00000031359
8	87006954	rs720818184	C	A	ENSBTAG00000031359
8	90602026	rs715319860	C	T	SHC3
8	95550153	rs137258780	A	G	ENSBTAG00000047350
8	95550874	rs210159742	C	T	ENSBTAG00000047350
8	95551131	rs441898901	A	C	ENSBTAG00000047350
8	96290639	rs17871351	G	T	ABCA1

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
8	105033055	rs43578588	C	G	KIF12
9	687820	rs518979191	C	T	PHF3
9	687892	rs522684210	C	T	PHF3
9	33719368	-	G	A	DCBLD1
9	41378822	rs464111914	T	C	ENSBTAG00000047024
9	50228767	rs466382918	G	A	SIM1
9	52922680	-	C	T	ENSBTAG00000017170
9	61340884	rs43601306	A	G	MDN1
9	75244584	rs135848896	C	A	MTFR2
9	90484553	rs715278963	A	G	SYNE1
9	90484619	rs43607008	T	C	SYNE1
9	93153512	rs210337249	A	C	TIAM2
9	95206925	-	G	C	ARID1B
9	95233801	rs797662818	C	T	ARID1B
9	95233834	rs437298777	A	G	ARID1B
10	6973628	rs716227151	C	T	ANKDD1B
10	9994150	rs718178570	C	G	DMGDH
10	9994172	rs43611602	A	G	DMGDH
10	10132553	-	G	A	BHMT
10	32645010	-	G	T	MEIS2
10	37396877	rs436264020	A	C	SPTBN5
10	44952664	rs208810879	T	A	NID2
10	61431826	rs461984504	A	G,C	SECISBP2L
10	76151393	rs464586684	C	T	WDR89
10	76506172	rs135117708	C	T	SYNE2
10	76526211	rs210717092	C	A	SYNE2
10	76635682	rs525951799	A	G	SYNE2
10	79469128	rs719496570	G	A	FAM71D
10	80207862	rs716040186	C	G	ZFYVE26
10	80207867	rs516499159	T	A	ZFYVE26

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
11	9246648	rs384763482	A	C	C2orf49
11	21052895	rs720469206	A	G	GALM
11	36567435	rs456831156	G	A	ENSBTAG00000024019
11	71152369	rs519906008	C	T	PLB1
11	72102508	rs442310074	T	C	C2orf16
11	72102628	-	G	A	C2orf16
11	93584827	rs721105087	A	G	ENSBTAG00000037542
11	93610397	rs715060812	A	G	OR1J2
11	102992990	-	T	C	TSC1,TSC1
12	21676996	rs459360883	C	T	THSD1
12	23341041	-	T	G	FREM2
13	3165171	-	T	C	ANKEF1
13	7526692	rs722858401	C	T	ESF1
13	23165493	rs133611160	A	G	MLLT10
13	32440945	rs522844486	A	C	STAM
13	46798276	rs524835741	A	G	LARP4B
13	59102204	rs520726976	A	G	ZBP1
13	60231196	rs516393048	A	G	ENSBTAG00000006539
13	62929304	rs137122057	G	A	BPIFB4
13	62929506	rs135759048	G	T	BPIFB4
13	62929549	rs133895802	T	G	BPIFB4
13	65698555	rs449336579	T	C	PHF20
13	76854220	rs110355490	G	C	NCOA3
13	76854360	rs448740130	C	G	NCOA3
13	76875244	rs382490055	G	T	NCOA3,SULF2
14	2540474	rs717214047	A	G	ENSBTAG00000003606
14	62349535	rs207991631	C	T	DPYS,DCSTAMP
14	62349538	rs719553274	C	A	DPYS,DCSTAMP
14	69160876	rs208078360	A	G	ENSBTAG00000046135
14	69160916	rs209417727	G	A	ENSBTAG00000046135

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
15	415839	rs381025813	G	A	ENSBTAG00000048221
15	7165253	rs210247939	A	G	ANGPTL5
15	24628632	rs208693042	C	T	ZW10
15	24628641	-	G	A	ZW10
15	31227381	rs526078610	C	A	TRIM29
15	45022967	rs137639739	A	G	RIC3
15	45652959	rs42443111	G	A	CYB5R2
15	45653001	rs210735791	G	A	CYB5R2
15	47126285	rs41767670	T	C	DNHD1,APBB1
15	48687480	rs524096795	T	A	ENSBTAG00000046762
15	48708405	rs518818200	C	T	UBQLN3
15	48708442	rs211254548	G	A	UBQLN3
15	48708463	rs43031470	T	C	UBQLN3
15	52195136	rs470175207	G	A	CHRNA10
15	52605239	rs384729165	T	A	INPPL1
15	54218636	rs383354138	A	G	UCP3
15	78617028	rs41785299	C	T	MTCH2
15	79895816	-	G	A	ENSBTAG00000046471
15	81955560	rs136395049	G	C	ENSBTAG00000040320
15	82970889	rs110077200	C	T	ENSBTAG00000039592
16	2707165	rs516378737	G	C	RBBP5
16	4539381	rs799273174	T	G	PIGR
16	19787226	rs209563370	A	G	USH2A
16	27161656	rs209571348	G	A	ENSBTAG00000013463
16	35950391	rs208970203	T	G	CHML
16	35950516	rs210074075	C	G	CHML
16	58051695	rs208825889	C	G	TNR
16	67418253	rs136378705	A	G	FAM129A

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
16	77945655	rs523191831	G	A	ASPM
16	77945676	rs110838111	T	C	ASPM
16	77945699	rs133471320	G	A	ASPM
17	18324306	rs109269119	A	G	MAML3
17	25034116	rs719448784	A	G	ENSBTAG00000035012
17	25034122	-	C	A	ENSBTAG00000035012
17	25121848	rs715374299	G	A	PRAME
17	40841379	-	T	G	TNIP3
17	55054337	rs110085274	G	A	ENSBTAG00000047904
17	62921568	rs469758678	G	C	RBM19
17	63630428	rs724034854	G	C	OAS2
17	65985922	rs444960873	T	C	MYO1H
17	74988396	rs518307328	C	T	DGCR8
18	4571709	rs110791021	T	C	ADAMTS18
18	15079115	rs465418931	A	C	ENSBTAG00000023726,ORC6
18	19190676	rs518261693	G	A	NOD2
18	19190700	rs521828585	G	A	NOD2
18	24775168	rs384246907	G	A	CES5A
18	26535237	rs133508179	G	T	GOT2
18	35101662	rs41870951	G	C	LRRC36
18	40289244	rs383603596	G	A	HYDIN
18	43324811	rs476158640	C	G	RGS9BP,ANKRD27
18	47264627	rs447764383	T	G	ZNF566
18	47702641	rs210970338	G	T	ZNF420
18	47702721	rs42733526	G	A	ZNF420
18	48807692	rs453233026	G	A	LGALS4,ECH1
18	49281305	rs464827950	T	C	ENSBTAG00000047691
18	49793748	rs520830789	G	A	ZNF607

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
18	49887632	rs523065605	G	T	MAP3K10,TTC9B
18	50316611	-	G	T	ENSBTAG00000009079
18	51705870	rs722281544	T	C	LYPD4
18	51754558	rs211256933	A	G	ENSBTAG00000006859
18	54081566	rs211261651	C	G	PNMAL1
18	57732982	-	G	A	SIGLECL1
18	63987431	rs209788736	A	G	ENSBTAG00000030348
19	9676529	rs447592139	C	T	MTMR4
19	9676673	rs109649580	T	C	MTMR4
19	9891722	rs134278863	C	T	RAD51C
19	18143377	rs109447105	G	C	C17orf75
19	18442359	rs383932645	C	T	ATAD5
19	19046288	rs519422156	A	G	OMG,NF1
19	25047373	rs132886157	G	A	ITGAE
19	26647740	rs520759450	T	C	NLRP1
19	27369592	rs211598908	G	A	ENSBTAG00000031933
19	31982727	rs715507464	G	T	ELAC2
19	36932792	rs208042859	T	C	EME1,MRPL27
19	36932842	rs211691707	T	G	EME1,MRPL27
19	36932845	rs210808467	T	C	EME1 ,MRPL27
19	36932876	rs209161619	C	T	EME1,MRPL27
19	39637071	rs110390311	A	G	GPR179
19	40895001	rs43718843	T	C	GSDMB
19	41608230	rs134248980	T	C	KRT25
19	42947580	rs517151568	G	A	GHDC
19	43758609	rs479824112	C	T	BRCA1
19	48338310	rs722212946	C	T	TANC2
19	50559008	rs723007116	C	T	FN3KRP

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
19	56684049	-	C	A	TMEM94,CASKIN2
19	57556339	rs41638028	A	C	ENSBTAG00000008328
19	57556347	rs715585181	T	G	ENSBTAG00000008328
20	7995245	rs521679263	A	G	ANKRA2,UTP15
20	9076066	-	G	A	ZNF366
20	9973396	rs41934002	C	T	BDP1
20	13424129	rs209960553	G	A	SREK1
20	23047819	rs110459128	G	A	ANKRD55
20	23103142	rs456474560	A	G	ANKRD55
20	35130722	rs480475412	C	T	C9
20	35130755	rs449113407	A	C	C9
20	36011203	rs210775931	T	C	EGFLAM
20	39165075	rs434524775	G	A	AGXT2
21	26111502	rs1117671118	C	T	KIAA1024
21	27994284	rs137583196	G	A	TRPM1
21	31492517	rs717579712	G	T	CHRNA5
21	34314667	rs520161744	G	T	CYP1A2
21	34314689	rs209535296	T	C	CYP1A2
21	34610231	-	C	T	UBL7
21	34743622	rs1116315674	C	G	CCDC33
21	43350963	rs523266130	G	A	AKAP6
21	56567083	rs41584199	T	C	C14orf159
22	12276006	rs434263600	G	A	SCN11A
22	15307271	rs42000457	C	A	VIPR1
22	16878172	rs517746812	C	T	IL17RC,CRELD1
22	26145191	rs209166597	T	A	CHL1
22	26168366	rs454951643	G	A	CHL1
22	44546314	rs135271368	G	C	IL17RD
22	44546336	rs209280012	G	A	IL17RD

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
22	49542581	rs208953431	C	T	PARP3,RRP9
22	49603731	rs524059271	C	T	ENSBTAG00000037410
22	52211915	rs452420297	T	C	ENSBTAG00000020077
23	6975416	rs471077573	C	G	ENSBTAG00000015565
23	10235754	rs136678875	C	T	C6orf222
23	16982051	rs207804619	T	C	DLK2,ABCC10
23	17645467	rs519559009	G	A	MRPL14,AARS2
23	19958629	rs379185758	G	A	ANKRD66
23	19958704	rs718429005	G	A	ANKRD66
23	20231431	rs517579344	C	T	ADGRF1
23	22136739	rs109994318	T	C	CRISP2
23	22136811	rs136858089	C	A	CRISP2
23	49412959	rs721744807	G	C	RPP40
23	49412960	rs715892133	G	A	RPP40
23	49944132	rs42034438	C	T	FAM217A
23	50643634	rs109841416	A	G	SERPINB9
24	21301528	rs520616024	A	C	ELP2
24	26038722	rs722965902	A	G	DSG4
24	26038831	rs208939947	C	T	DSG4
24	33270556	rs461173190	G	A	LAMA3
24	41962472	-	A	G	ANKRD12
24	50021298	rs718719296	C	T	MYO5B
25	1865871	rs385239431	G	C	ENSBTAG00000034643
25	2692173	rs109091197	C	G	MEFV
25	2745742	rs379845466	G	C	TIGD7,ENSBTAG00000040381
25	3578898	rs109510085	T	C	DNAJA3
25	3899920	-	T	C	ROGDI,GLYR1
25	9647490	rs434491717	T	C	CIITA
25	11495702	rs133398852	G	A	CPPED1
25	17147233	rs456125090	C	T	TMC5
25	17147286	rs474692759	C	T	TMC5

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
25	18728299	rs134663557	C	T	ENSBTAG00000006449
25	19659591	rs456460108	G	C	ENSBTAG00000035915
25	19659635	rs719836772	A	G	ENSBTAG00000035915
25	20377536	-	C	A	OTOA
25	41256057	rs448060259	A	G	IQCE
26	11089314	rs208272356	G	A	IFIT3
26	11089439	rs210110362	G	A	IFIT3
26	11430005	rs517496673	C	T	ENSBTAG00000005708
26	15953558	rs462911685	G	A	CYP2C18
26	16361883	rs716181276	G	T	ENSBTAG00000023549
26	17250827	rs451751111	G	A	ENSBTAG00000039132
26	17250836	rs42088670	G	C	ENSBTAG00000039132
26	19164731	rs483278185	A	C	R3HCC1L
26	19164817	rs462312438	A	G	R3HCC1L
26	19164827	rs479358276	A	G	R3HCC1L
26	19164849	rs447958979	T	C	R3HCC1L
26	19164857	rs464576040	T	G	R3HCC1L
26	24262416	-	C	T	PDCD11
26	35436472	-	C	G	ENSBTAG00000013820
26	35436473	rs384434967	A	G	ENSBTAG00000013820
26	35436508	rs378461758	C	A	ENSBTAG00000013820
26	35436509	rs381387919	A	G	ENSBTAG00000013820
26	36086533	rs469115875	C	T	ATRNL1
26	42795640	rs455000035	C	T	ENSBTAG00000022715
26	42795654	rs723268945	G	A	ENSBTAG00000022715
26	42795708	rs519066685	A	G	ENSBTAG00000022715
27	12763716	rs525477945	G	A	TENM3
27	13683472	rs41650218	G	C	STOX2
27	13749517	rs521358398	T	C	ENPP6
27	15504423	-	G	A	FAT1
27	15504562	rs516172936	T	C	FAT1

CHROM	POS	ID	ALLELE		GENE
			REF	ALT	
27	34152988	rs434192465	C	G	ENSBTAG00000040445
27	34771699	rs209311358	G	A	IDO2
27	37198463	rs210661823	C	T	CHRNA6
28	2710848	rs137101680	A	G	OR5D14
28	9483679	rs109257602	G	A	MTR
28	42364043	rs467166388	T	C	ANTXRL,ANTXRL
28	43930932	rs443896996	G	A	C10orf71
28	44284834	rs717862287	C	T	PARG
28	46168873	rs521225504	C	T	COG2
28	46168897	rs717791507	A	C	COG2
28	46168918	rs110309555	C	T	COG2
29	5187242	rs109060372	G	A	TRIM77
29	19112732	rs520227645	A	G	ENSBTAG00000034536
29	29491569	rs451201478	C	A	ENSBTAG00000023300
29	44777249	rs524149933	A	G	CATSPER1
29	46765406	rs517065521	G	A	GAL
29	46917250	rs109769588	C	T	MRPL21
X	21969277	rs720966022	C	T	ENSBTAG00000018323
X	75151693	rs463142333	G	C	ENSBTAG00000013077
X	89907997	-	C	A	ENSBTAG00000047411
X	89908014	-	C	T	ENSBTAG00000047411
X	89908020	rs446746200	C	G	ENSBTAG00000047411
X	112234657	rs456298560	A	T	ENSBTAG00000024662
X	116042372	rs519769055	T	A	ENSBTAG00000046838
X	129542185	rs136304410	A	G	ENSBTAG00000013625
X	138887232	rs482876617	G	A	MXRA5,MXRA5,MXRA5
X	143913993	rs440235430	T	C	ENSBTAG00000000211

Appendix- II

The chemicals used in this study were of molecular biology grade. The list of chemicals and the preparation of reagents are as follows:

Preparation of Reagents

Tris (1 M) pH 8.0

Tris base (121.10 gm) was dissolved in 800 ml of double distilled waterand the pH was adjusted to 8.0 by adding concentrated HCl (40 ml). The solution was allowed to cool to room temperature before making final volume to 1 liter and stored at 4 ^{0} C.

Sodium chloride (5 M)

Sodium chloride (29.40 gm) was dissolved in 80 ml of double distilled water and the final volume made up to 100 ml. Sterilized by autoclaving and stored at 4 ^{0}C.

Sodium acetate (3 M)

24.60 gm of Sodium acetate (anhydrous) was dissolved in 80 ml of double distilled water. pH was adjusted to 5.2 with glacial acetic acid and the final volume was adjusted to 100 ml. Sterilized by autoclaving and stored at room temperature.

EDTA (0.5 M)

Disodium salt of EDTA (18.60 gm) was dissolved in 80 ml of double distilled water by adding NaOH pellets. pH was adjusted to 8.0 and final volume made up to 100 ml Sterilized by autoclaving and stored at 4 ^{0}C.

Ammonium chloride (1 M)

Ammonium chloride (5.35 gm) was dissolved in 80 ml of double distilled water and final volume was made up to 100 ml. Sterilized by autoclaving and stored at 4 ^{0}C.

Sodium bicarbonate (1 M)

10.00 gm of sodium bicarbonatewas dissolved in 80 ml of double distilled water and made up to a final volume of 100 ml. Sterilized by autoclaving.

Sodium dodecyl sulphate (SDS) 10 %

Dissolved 10.00 gm of Sodium dodecyl sulphatein 80 ml of autoclaved double distilled water and made up to a final volume of 100 ml.

RBC Lysis buffer (1X prepared freshly everytime)

NH_4Cl (155 mM)	:	155 ml
$KHCO_3$ (10 mM)	:	10 ml
EDTA (0.1mM)	:	2 ml

Final volume was made up to 100ml by adding double distilled water. Autoclaved and stored at 4^0C

RBC Lysis buffer (1X prepared freshly everytime)

NaCl (75 mM)	:	75 ml
Tris (10 mM)	:	10 ml
EDTA (2 mM)	:	4 ml

Final volume was made up to 100ml by adding double distilled water. Autoclaved and stored at room temperature

Phenol equilibration (Tris saturation)

Phenol crystals (500 gm) stored at –20 ^{0}C was liquefied by keeping in a water bath maintained at 65 ^{0}C for 1 hour. 8-hydroxyquinoline was added to a liquefied phenol at a final concentration of 0.1 %. Then equal volume (500 ml) of 0.5 M Tris (pH 8.0) was added and stirred for 4 hours on magnetic stirrer and pH was checked repeatedly till it reached 8.0. Finally, 0.1 M Tris was added to an equilibrated phenol and stirred well and stored in amber colored bottles at (4 ^{0}C).

Proteinase-K

Proteinase-k (20 mg) was dissolved in 1ml of double distilled water. Aliquots of 400µl per vial made and stored at -20 ^{0}C.

Ethanol 70 %

Ethanol 99.9 %	:	70 ml
Distilled water	:	30 ml

Phosphate Buffer Saline (1X) Ca^{++} and Mg^{++} free

NaCl	:	4.0 gm
KCl	:	0.1 gm
Na_2HPO_4	:	0.7 gm
KH_2PO_4	:	0.1gm

The pH was adjusted to 7.4 and final volume made up to 500 ml with double distilled water. Autoclaved and stored the buffer at room temperature.

TE buffer (10 mM)

Tris (1M, pH 8.0)	:	1.00 ml
EDTA (0.5M, pH 8.0)	:	200 µl
Final volume made up to	:	100 ml

Sterilized by autoclaving and stored the buffer at room temperature.

Agarose gel:

Agarose 0.6 %

Agarose	:	0.60 gm
TBE (10 X)	:	10 ml
Double distilled water	:	90 ml

Agarose 1.8 %

Agarose	:	1.80 gm
TBE (10X)	:	10 ml
Double distilled water	:	90 ml

Agarose 2.0 %

Agarose	:	2 gm
TBE (10X)	:	10 ml
Distilled water	:	90 ml

Ethidium bromide

Ethidium bromide (0.25 gm) was dissolved in 25 ml of double distilled water. Stirred on a magnetic stirrer for several hours to ensure that the dye is properly dissolved. Wrapped the container with aluminum foil and stored at 4 ^{0}C.

Preparation of 6X gel loading dye

Bromo phenol blue	:	0.25 % (w/v)
Xylene cyanol FF	:	0.25 % (w/v)
Glycerol in water	:	30 % (v/v)

TBE buffer (10X)

Tris base	:	108.0 gm
Boric acid	:	55.0 gm
EDTA (0.5 M, pH 8.0)	:	40 ml

Double distilled water was added to make the final volume 1000ml, filtered and autoclaved.

TAE buffer (50X)

Tris base	:	242.0 gm
Glacial acetic acid	:	57.1 ml
EDTA (0.5 M, pH 8.0)	:	100 ml

Double distilled water was added to make the final volume 1000ml, filtered and autoclaved.