International Journal of Innovations in Applied Sciences & Engineering http://www.ijiase.com

(IJIASE) 2016, Vol. No. 2, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

SEQUENCE CHARACTERISATION AND IDENTIFICATION OF SNP MARKERS IN CHICKEN MYOSTATIN GENE

Vanshika Goel and Sainyam Goel
Indraprastha Convent Senior Secondary School
Begumpura, Rohini, New Delhi

INTRODUCTION

Indigenous Breeds

The common country hen, the desi, is as a rule the best mother for hatching. She is a good forager. Some of the Indian fowls resemble the Leghorn in size and shape, but have poor laying qualities. They are found in various colours. One variety found in India resembles the Sussex or Plymouth Rock in shape, but is smaller. These birds lay fairly well and are more common in the eastern parts of the country.

The Indian birds are mostly non-descript, and are of very little value as layers. They have several local breed names such as Naked Neck, Punjab Brown, Ghagus, Lolab, Kashmiri Favorolla, Teri, Busra, Tellicherry, Danki, Nicobari and Kalahasti, Haringhata Black. There are only 4 pure breeds of fowls indigenous to India. They are the Chittagong, the Aseel, the Kadaknath and the Busra. The last occurs in western India. A large number of fowls of different sizes, shapes and colours, and for the most part resembling the Red jungle fowls, are found all over India. They vary in appearance according to the locality in which they have been bred.

According to 1992 censes, the total poultry population in India was 307 million (284 million chicken, 22 million ducks and 1 million other poultry), which increased to 489 million in 2003. The fowl population of India can be roughly classified into two types. (i) Desi or indigenous and (ii) Improved or exotic

A wide range of variations is found among these breeds in relation to body weight, plumage, skin colour, feathering and comb type. These breeds have acquired considerable adaptability to local climatic environment and considerable resistance to tropical disease. They are also considered to be the best suited to contribute economic benefits in their respective home tracts.

Of the total 734 breeds, 18 indigenous poultry populations have been reported in India. They are 15 breeds with defined distribution (Aseel, Kadaknath, Miri, Kashmir Favorolla, Punjab brown, Chittagong, Daothigir, Kalasthi, Ghagus, Telicherry, Busra, Nicobari, Ankaleshwar, Dunki and Haringhata Black) and 3 populations distributed throughout India (Desi, Tani and Titri). The

indigenous poultry distribution along with its utility in the various regions of the country has been given in table 1.

The native genetic resources are the goldmines of major genes for future poultry production. Genes and biological mechanisms involved in yielding better flavor of meat, egg production, disease resistance, body growth etc. are yet to be well established in avian populations, which can be done by characterizing the important genes related with above characters.

MYOSTATIN (GDF-8)

Myostatin, a candidate gene, member of TGF- β is a negative regulator of muscle growth encodes a protein, a member of a superfamily transforming growth factors beta (TGF-β). There are several TGF-β subtypes, which are based on their related structure. One such member is called growth and differentiation factors (GDF) and specifically regulates growth and differentiation. Myostatin (GDF-8) is the skeletal muscle protein associated with the double muscling in different species. Further exploration of genes present in skeletal muscle in the two breeds of double-muscled cattle revealed mutations in the gene that codes for Myostatin. The double-muscling trait of the Myostatin gene knockout mice and the double-muscled cattle demonstrates that Myostatin performs the same biological function in these two species. Apparently, Myostatin may control the growth of skeletal muscle. Myostatin gene consists of three exons, two introns, a promoter region along with 5'UTR and 3'UTR.

Promoter Region: In genetics, a promoter is a DNA sequence that enables a gene to be transcribed. The promoter is recognized by RNA polymerase, which then initiates transcription. In RNA synthesis, promoters are a means to demarcate which genes should be used for messenger RNA creation - and, by extension, control which proteins the cell manufactures. The perfect promoter is called a canonical sequence (A canonical sequence is a sequence of DNA, RNA, or amino acids that reflects the most common choice of base or amino acid at each position).

Promoters element: Promoters represent critical elements that can work in concert with other regulatory regions (enhancers, silencers, boundary elements/insulators) to direct the level of transcription of a given gene. As promoters are typically upstream from the gene in question on the DNA strand, positions in the promoter are denoted by numbering from +1, or the transcriptional start site, where transcription of RNA begins for a particular gene. Positions upstream are negative numbers counting back from +1, for example -100 is a position 100 base pairs upstream. The usage of canonical sequence for a promoter is often problematic, and can lead to misunderstandings about promoter sequences. Canonical implies perfect, in some sense. In the case of a transcription factor-binding site, then there may be a single sequence, which binds the protein most strongly under specified cellular conditions. This might be called canonical. However, natural selection may favor less energetic binding as a way of regulating transcriptional output. In this case, we may call the most common sequence in a population, the wild-type sequence. It may not even be the most advantageous sequence to have under prevailing conditions.

Recent evidence also indicates that several genes (including the proto-oncogene c-myc) have G-quadruplex motifs as potential regulatory signals.

Core promoter - the minimal portion of the promoter required to properly initiate transcription

- Transcription Start Site (TSS)
- Approximately -35
- A binding site for RNA polymerase
  - RNA polymerase I: transcribes genes encoding ribosomal RNA
  - RNA polymerase II: transcribes genes encoding messenger RNA and certain small nuclear RNAs
  - RNA polymerase III: transcribes genes encoding tRNAs and other small RNAs
- General transcription factor binding sites

Proximal promoter - the proximal sequence upstream of the gene that tends to contain primary regulatory elements

- Approximately -250
- Specific transcription factor binding sites

Distal promoter - the distal sequence upstream of the gene that may contain additional regulatory elements, often with a weaker influence than the proximal promoter.

- Anything further upstream (but not an enhancer or other regulatory region whose influence is positional/orientation independent)
- Specific transcription factor binding sites

A major question in evolutionary biology is how important tinkering with promoter sequences is to evolutionary change, for example, the changes that have occurred in the human lineage after separating from chimps. Some evolutionary biologists, for example Allan Wilson, have proposed that evolution in promoter or regulatory regions may be more important than changes in coding sequences over such time frames.

Diseases associated with aberrant promoter function

The relationship between mutations and natural variation in gene sequence and susceptibility to hundreds of diseases, it requires a sophisticated search strategy to extract those diseases that are associated with defects in transcriptional control where the promoter is believed to have direct involvement. Evidence suggests that some involvement of promoter malfunction, either through direct mutation of a promoter sequence or mutation in a transcription factor or transcriptional co-activator leading to diseases.

A gene that may influence development of disease or trait of interest is known as candidate gene. Positional candidate genes are selected for association studies on the basis of their location in a genomic region link to the disease of interest in humans' and/ or their location in a syntenic chromosomal region influencing the trait of interest. However the skeletal muscle growth was

considered to be a quantitative trait from the beginning, it has been lately found to be monogenic trait with the truncation of protein of Myostatin gene being the cause for double muscling in cattle with comparative effect on other organisms as well. The sequence analysis of this gene has found only one non-synonymous change in the Indian poultry population from the NCBI and other databases but is fixed in all the Indian chicken populations. The other changes were found only in the nucleotides present in Exon and introns, which do not have effect on the protein sequence. The promoter region of the gene is equally responsible for the expression of the protein and may be directly or indirectly effecting the quantitation of the gene product. The promoter region of the Myostatin gene was found to be having 5 nucleotide changes (SNPs) and these changes were typed in four breeds of Indian chicken with a view to estimated the frequency of the SNPs in indigenous chicken populations and subsequent analysis.

This work utilizes promoter region sequence variation to investigate molecular evolution, population genetics and natural selection in different poultry populations with a focus that the muscle development in the game birds has a different configuration with respect to Myostatin gene. The work on SNPs of Promoter region of the Myostatin gene was carried out with the following objectives:

1. To search for genetic polymorphism (SNPs) in the promoter region of the Chicken Myostatin gene
2. To estimate the gene frequency of the SNPs in four chicken populations.
3. To estimate the differences among the various populations with respect to the SNPs in Promoter of Myostatin gene.
4. To test the various population genetic hypothesis utilizing the SNP frequency data.
5. To study the various nucleotide parameters and the population demographic features they present.
6. To check for the recombination events occurring in the promoter region of Myostatin gene.
7. To find out the genetic relationship among the four populations utilizing the population genetic tools.
8. To draw phylogenetic relationship among all the individuals studied and carryout the multivariate analysis of the data for relationship.
9. To find out the minimum spanning tree and relationship among the various haplotypes arising of the mutations.
10. To test the populations for Hardy Weinberg Equilibrium and gene flow analysis.

MATERIALS AND METHODS

The four populations of interest selected for present study belongs to different regions of the country of which Ghagus is used for both gaming and meat purpose. The breeds considered in the study and their breeding tract are given below (Fig 2-6)

1. Busra
2. Ghagus
3. Haringhata black
4. Kashmir favorolla

Busra

Busra is a minor breed occurring in small numbers in parts of Gujarat and Maharashtra states. The birds have body confirmation typical of layers and are majestic in appearance.

Ghagus

Ghagus birds are distributed in Kolar and adjoining parts of Bangalore District of Karnataka and chittoor and Anantapur District of Andhra Pradesh. The birds might have derived its name by a peculiar sound and locally known as desi or Ghagus. Ghagus birds are mainly kept for meat and egg purpose Tantia M.S et. al, (2005).

Kashmiri Favorolla

Tantia M.S et al, (2005), Kashmiri favorolla are locally known as "Kashmir kukkar". Birds are basically reared as backyard poultry in the high altitude region of Kashmir valley of Jammu & Kashmir state. The birds are basically interbreeding group of different phenotypes, genotypes.

Haringhata Black

It is found in the state of West Bengal. Small bodied with the typical confirmation of layer, but a poor layer. Not much resistant to disease.

Methods Performed

Blood Sample Collection

Four breeds Busra (24 samples), Ghagus (24 samples), Haringhata Black (24 samples), Kashmiri Favorolla (32 samples) were selected for the study of Myostatin gene effect on poultry population, 1 ml of whole blood was collected from unrelated birds aseptically from wing vein using heparinised vacutainer tubes and transported to laboratory at 0-5^{o}C.

DNA Isolation, Storage, Quantification

DNA Isolation

1. For isolation of DNA from collected sample, the blood was transferred to autoclaved Oakridge centrifuged tubes.
2. The RBCs were lysed with lysis buffer (Ammonium chloride 155 mM, Potassium bicarbonate 10 mM and EDTA 0.1 mM). In all the blood samples, lysis buffer was added.
3. The samples were then mixed gently and kept in ice for 10 minutes.
4. The tubes were centrifuged at 12,000 rpm for 10 minutes at 4^{o}C in the refrigerated centrifuge.

5. The supernatant was carefully decanted and the pellet was redissolved in lysis buffer and washed three times as described above.
6. The pellet was resuspended in 10 ml digestion buffer (Sodium chloride 75mM, Tris-Cl 1 M, pH 8.0 and EDTA 0.5 mM).
7. Vortexed gently and to it added 20% Sodium lauryl sulphate (200 µl/sample) and Proteinase K (1 mg/sample) and incubated at 57^{o}C overnight in a water bath.
8. After incubation, digested solution was obtained to which equal amount of Tris equilibrated phenol (pH 8.0) was added.
9. This was mixed gently by moving the tubes gently in '8' fashion for 10 minutes and centrifuge at 12,000 rpm for 10 minutes at 25^{o}C.
10. After centrifugation the aqueous phase and organic phase were separated. The DNA remained dissolved in aqueous phase while the protein was retained at the interphase.
11. The aqueous phase was transferred carefully by Pasture pipette to another Oakridge tube without disturbing the interphase.
12. To the aqueous phase, phenol: chloroform: isoamyl alcohol (25:24:1) was added and mixed by moving the tubes gently in '8' fashions for 10 minutes at 25^{o}C to separate the aqueous phase and organic phase.
13. Again the aqueous phase was carefully pipetted to another tube without disturbing the interphase. The aqueous phase was mixed with chloroform: isoamyl alcohol (24:1).
14. The solution was subjected to centrifugation at 12,000 rpm for 10 minutes at 25^{o}C and aqueous phase was carefully transferred to glass culture tube.
15. The DNA was precipitated by adding 1/10th volume of Sodium acetate (3M, pH 5.2) and 2.5 volume of aqueous phase of chilled absolute alcohol.
16. The tubes were covered with paraffin film and mixed slightly, the DNA get precipitated and is visible as white stringy strands (Sambrook, J., Fritsch E.F. and Maniatis. T, 1989).

Washing and Storage of Isolated DNA

17. The DNA was spooled out into eppendrof tubes and washed twice with 70% ethanol to remove the salts.
18. After washing, the alcohol was allowed to evaporate and DNA was redissolved in 500 µl Tris EDTA buffer (Tris 10mM, EDTA 10mM, pH 8.0)
19. Kept out eppendrof tubes at 65oC for 1 hour. The stock DNA was stored at –20oC.

Preparing the sample

While the gel was cooling, prepared the DNA sample by adding 5 µl of tracking dye to 1 µl of each DNA sample. The tracking dye contains Bromophenol Blue and Xylene cynol FF and 50% glycerol. Adding tracking dye to the sample increased its density so it falls into the well of the gel and provided a visible marker to monitor the progress of electrophoresis. Also prepared molecular

size standard by mixing 5 µl of the 1 kb ladder with 1 µl of tracking dye. Bromophenol Blue migrated through agarose gel approximately 2.2 fold faster than Xylene cyanol FF dye, independent to agarose gel concentration.

Loading and running the gel

The gel on the tray, inserted horizontally into the electrophoresis chamber and flooded the top of the gel with fresh running buffer (1X TAE) to cover the gel to depth of about 1 mm. Sucked the solution (DNA sample with dye) into the pipette, placed the tip in the top of the well and gently expelled the solution into the well. The lid and power leads were placed on the apparatus, 40-80 V current is applied, and current flowing was confirmed by observing bubbles coming off from the electrodes. Run the gel until the Bromophenol blue and Xylene cynol FF was migrated the appropriate distance about three-fourth of the way across the gel.

Visualization of DNA

Bright orange coloured DNA bands were visualized by placing the tray onto High Performance UV Transilluminator.

Databases visited

On selecting Myostatin as candidate gene, search for myostatin gene sequence was carried out on online databases and AF346599 was selected as reference sequence from NCBI database. For reported SNPs, the SNP consortium was screened and few SNPs in coding sequences and non coding sequences were selected from chicken database.

Myostatin (Growth differentiation factor 8)

Myostatin gene, a negative regulator of skeletal muscle growth comprises of three exons and 2 introns coding for 376 amino acid propeptide, which give rise to 15 Kda active, processed and matured protein (Mc Pherron., et al., 1997). The locations of selected SNPs on promoter region selected from SNP database are given below in Figure 7 and table 2.

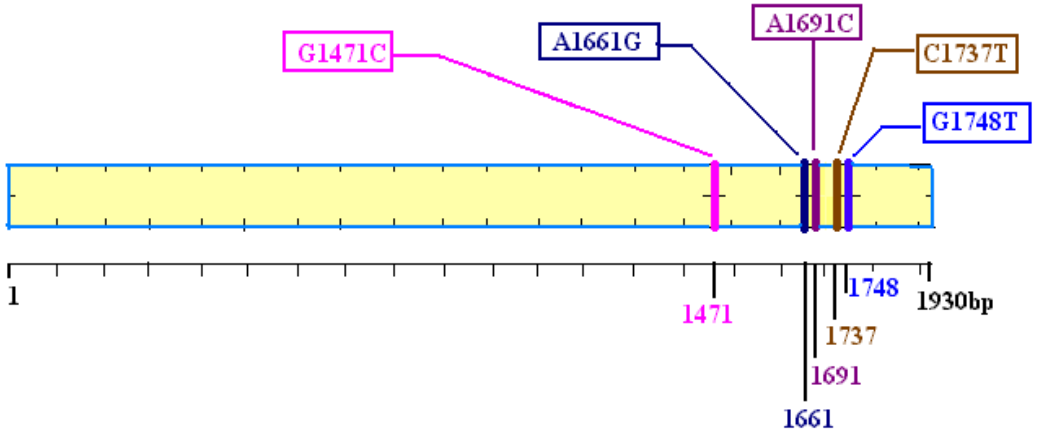


Fig: 1 Promoter region of Gallus Myostatin gene (1-1930 bp) and position of SNPs.

Region	Location (Location w.r.t. AF346599)	Sequence
Promoter	1471	TGG → TAG
Promoter	1661	TGA → TGG
Promoter	1691	TAG→ TCG
Promoter	1737	ACG → ATG
Promoter	1748	AGA → ATA

Table: 1 Location of SNPs selected and their variations

Designing of Primers

The primer was designed by using **Primer 3** software to amplify the region of interest in Myostatin gene targeting the SNPs in Promoter region. The sequences of the primer and region of amplification and targeted SNPs are mentioned in table 3.

Primers	Sequences	Targeted SNPs	From– To (w.r.t. AF346599)	Size (bp)
GMST 4	**F**-CAAAATgTTTATTCCTgCTCACC **R**-CTAAACAgATCCgggACAgC	G(1473) A A(1663) G A(1693) C C(1739) T G(1750) T	1373-1978	605

Table: 2 Primer for Myostatin promoter region screened in Chicken Breeds

Designing of PCR-RFLP

Using web cutter (www.neb.com/webcutter) restriction sites for targeted SNPs were screened. There were two Restriction enzymes (Nla III and Sca I) at two targeted SNPs (1471 and 1691) were present in the amplified promoter region. The selection of enzymes using web cutter and their expected product sizes / genotypes given below.

Sequence digestion with NlaIII

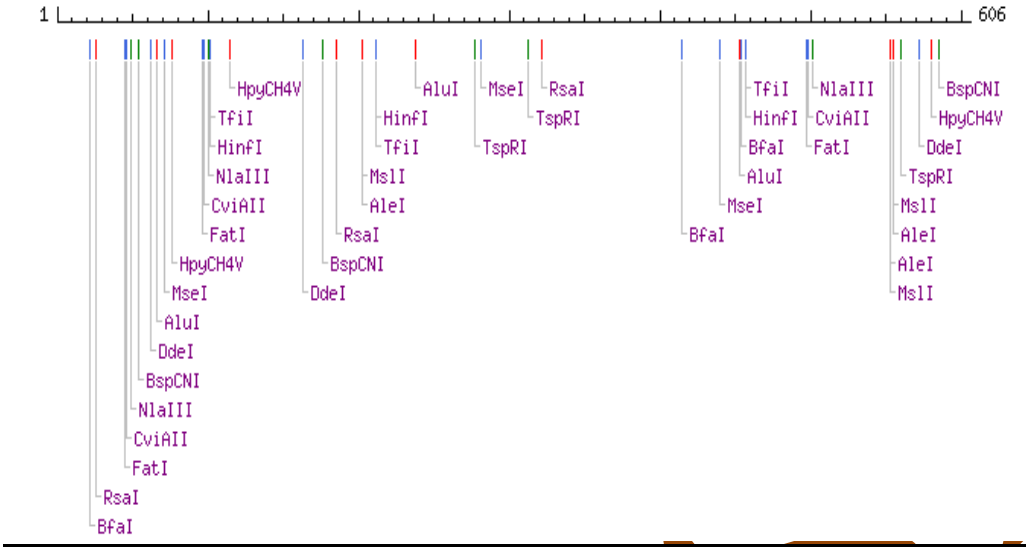


Fig: 2a Restriction sites of different Restriction Enzymes in amplified promoter region (607bp)

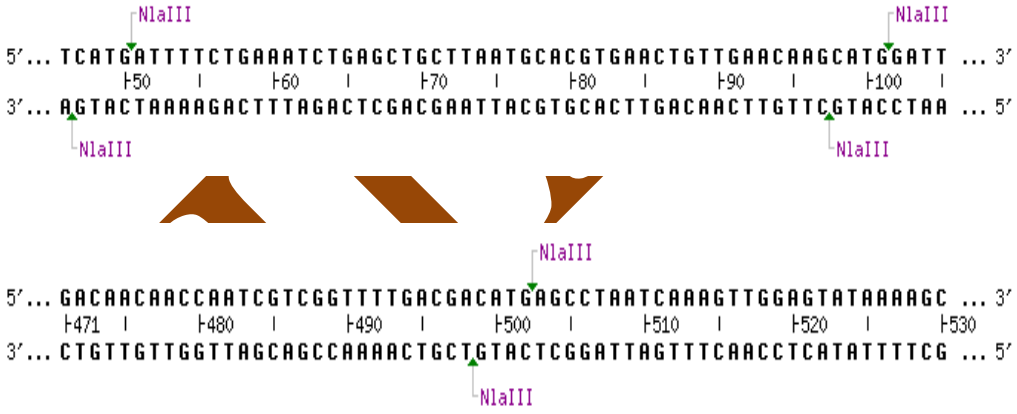


Fig: 2b Restriction sites for Restriction enzyme Nla III

Enzyme	No. of sites	Product size	Genotypes		
			Normal GG	**Mutant AA**	**Heterozygous AG**
NlaIII	Two	605	401 103 51 51	452 103 51	452 401 103 51

Table: 3 Expected genotypes of PCR-RFLP using enzyme Nla III

Sequence digested with ScaI enzyme

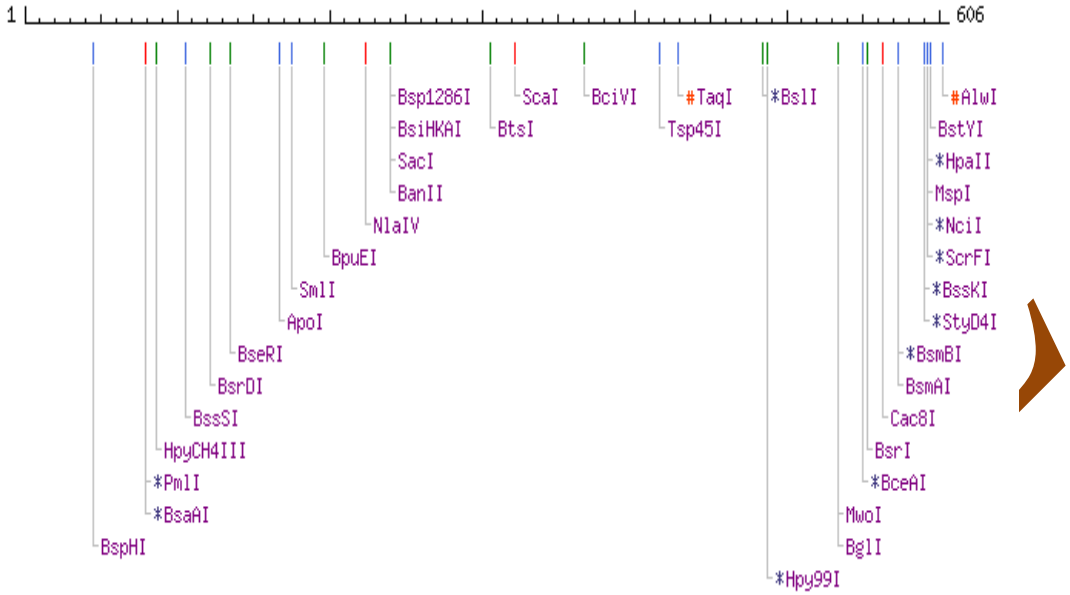


Fig : 3a Restriction sites of different Restriction Enzymes in amplified promoter region (607bp)

Fig: 3b Restriction sites for Restriction enzyme Sca I

Enzyme	No. of sites	Product size	Genotypes		
			Normal AA	Mutant CC	Heterozygous AC
ScaI	One	605	324 282	605	605 324 282

Table: 4 Expected genotypes of PCR-RFLP using enzyme Sca I

Pre-PCR Preparation

The PCR was performed under standard condition as described by Kaul et al., 2001. The genomic DNA was diluted so as to contain about 50 ng DNA/µl.

Cocktail Preparation

Component	Volume	Concentration
10X PCR Buffer	1.5 µl	-
DNTPs	0.2 µl	100 mM
Primer (Forward)	1.0 µl	4 pmol
Primer (Reverse)	1.0 µl	4 pmol
Taq Polymerase	0.2 µl	0.4 Unit
Milli Q water	10.1 µl	-
Total	14 µl	

Table 5: The cocktail for PCR reaction (14 µl):

PCR programme

1µl of the template DNA was directly added into the cocktail in each lane of the PCR plate to make final reaction volume of 15µl. The PCR was carried out in Applied Bio-system and Bio-Rad thermocycler.

Steps	Programme	Temperature	Time	No. of cycles
Step- 1	Initial Denaturation	95^{o}C	5 minutes	1 cycle
Step- 2	Denaturation	94^{o}C	45 seconds	35 cycles
Step- 3	Annealing	56^{o}C to 54^{o}C	45 seconds	
Step- 4	Polymerization	72^{o}C	45 seconds	
Step- 5	Final Extension	72^{o}C	5 minutes	1 cycle
Step- 6	Final Temperature	4^{o}C	Hold	

Table 6: Standardised PCR Program used for amplification

Agarose Gel Electrophoresis of PCR Amplified DNA

After completion of the PCR programme, the products were checked on 2% agarose gel for the amplification. Before loading into the well, gel-loading dye (xylene cyanol FF, bromophenol blue in glycerol) was added to the sample and the samples were run along with DNA ladder under constant voltage conditions (150 V) till the two dyes were separated. Amplified product appeared

as sharp orange color bands under UV Transilluminator due to the intercalation of ethidium bromide.

Restriction digestion

Enzyme: Nla III Stock concentration (10,000u/ml) and Nebuffer 4 with BSA.

Ingredients	Volume
Template	5 µl
BSA	0.1
Buffer	1µl
Enzyme (2units)	0.2 µl
D/W	3.7 µl

Table 7: Restriction Enzyme (Nla III) digestion composition

Incubated the sample at 37°C to overnight. The digested samples were run on 2.5% agarose gel at 120v for 30 min. The gel was visualized by staining EtBr. The gel was scored for genotypes and the data was recorded for further analysis.

Enzyme: Sca I Stock concentration (10,000u/ml) and Nebuffer3

Ingredients	Volume
Template	5 µl
Buffer	1µl
Enzyme(2 units)	0.2 µl
D/W	5 µl

Table 8: Restriction Enzyme (Sca I) digestion composition

Incubated the sample at 37°C to overnight. The digested samples were run on 2.5% agarose gel at 120v for 30 min. The gel was visualized by staining EtBr. The gel was scored for genotypes and the data was recorded for further analysis.

Genotyping of SNPs (1661,1737 and 1748) by SNAPSHOT:

The principle of Snap shot multiples kit in genotyping SNPs is based on Single Base Extension. The chemistry is based on the dideoxy single-base extension of an unlabeled oligonucleotide primer (or primers). Each primer binds to a complementary template in the presence of fluorescently labeled ddNTPs and AmpliTaq DNA Polymerase, The polymerase extends the primer by one nucleotide, adding a single ddNTP to its 3´ end.

Remaining SNPs in the promoter region (1661, 1737 and 1748) were genotyped by SnaPshot; a high throughput method of Genotyping of SNPs was used. The following primers were designed from the amplified sequence of the promoter region and targeting the selected SNPs.

Primer	Sequence	SNP	Size
GM2	Aaaaaaaatgcgttcagtgtatatttaagaat	A1661G	33
GM4	aaaaaaaaaaaaaaaaaaaaaaaaacagtatgtgaagacgtatcctcta	C1737T	50
GM5	Aaaaaaaagaagccacca	G1748T	20

Table 9: Primers used for genotyping SNPs in promoter region by snapshot

The criteria behind designing of primers is:

- Primers included in a single reaction need to differ significantly in lengths in order to avoid overlap between the final SnaPshot products. A difference of 4–6 nucleotides between primer lengths is recommended as a starting point.
- The length of a primer can be modified by the addition of nonhomologous polynucleotides at the 5′ end. Since the recommended annealing temperature for a SNaPshot control primer is 50 °C, the melting temperature for the complementary region between any primer and its corresponding template should be at least 50 °C.
- Poly (dT), poly (dA), poly (dC), and poly (dGACT) are 5′ non-homologous tails, which are predicted to have minimal secondary structures. They have all been used successfully.
- Generally the signal patterns are not affected by the kinds of tails that are used. The 5′ poly (dT) tails however may interfere with the addition of 3′ ddA. The mobility of an oligonucleotide in capillary electrophoresis is determined by its size, nucleotide composition, and dye. Thus the effect of nucleotide composition on mobility can be significant when the primer is short. We strongly recommend that primers shorter than 36 nucleotides be tested before being multiplexed to ensure that the final products are spatially resolved when analyzed on the instrument.
- Check primers for possible extendable hairpin structures within each primer and for extendable dimer formation between primers.

- HPLC purification of primers is recommended for oligonucleotides longer than 30 nucleotides. Heterogenous primer mixtures containing mixed molecular weight oligonucleotides may yield undesired products that will confuse analysis.
- Since SNP interrogation using primer extension does not permit any flexibility with respect to the location of the 3´ end of the primer, use primers that are complementary to the negative (–) DNA strand if the positive (+) DNA strand is difficult to assay.

PCR Amplification

PCR amplification of the promoter region using Gmst4F-4R primers and amplifying 605bp product.

Purification of PCR product:

Exo-SAP Digestion of PCR product: Make a master mix of Exonuclease I and shrimp alkaline phosphate as per the following

Component	Final Concentration (U/ul)
Exo.1 (20 U/ul)	0.5
SAP (1 U/ul	0.5
PCR buffer 10 x	1 ul
Milli Q water	Make up the
Total volume	final volume 10 ul

Table10: Exonuclease digestion composition

Add 1 ul of the master mix to 10 ul of PCR product (50 to 100 ng) and set up the following incubation protocol in the thermal cycler

Hold = 37˚C for 120 minutes

Hold = 85˚C for 15 minutes

Hold= 4˚C for infinity.

Purified PCR product can be used for snapshot reaction.

Preparing the SNaPshot Reactions

SNaPshot Multiplex Ready Reaction Mix	5 µl
Purified PCR products	3 µl
Pooled SNaPshot primers (GM2, 4 and 5)	1 µl
Deionized water	1 µl
Total	10 µl

Table11: SnaPshot reaction composition

Thermal cycling of Snapshot reactions:

The above sample reactions were placed in thermal cycler and following program was run at 25 cycles of

96 °C for 10 seconds
50 °C for 5 seconds
60 °C for 30 seconds and

final hold at 4 °C until ready for post-extension treatment.

Post extension treatment

To above each reaction mixture, 1.0 Unit of Shrimp Alkaline Phosphatase (SAP) or 1.0 Unit of Calf Intestinal Phosphatase (CIP) is added and incubated at 37 °C for 1 hour followed by deactivation of the enzyme by incubating at 75 °C for 15 minutes.

Preparing the SNaPshot Products for the ABI PRISM 3100 Genetic Analyzer

Size standard: Gene Scan- 120 LIZ™ Size Standard developed by Applied Biosystems was used for fragment sizing, LIZ size standard yield size fragments between 15-120bp providing 9 single – standard labeled fragments of 15, 20, 25, 35, 50, 62, 80, 110 and 120 bases. Each of the DNA fragments labeled with a proprietary fluorophore, which results in a single peak when run under denaturing condition. Internal lane size standard was run with every sample for accurate sizing

The genotyping reaction components were:

SnapShot Product - 0.5 µl
Hi-Di Foramide - *9.25 µl*
120-Liz Size standard - 0.25 µl

The above components were mixed well and denatured at 95ºC for 5 minutes. The 96 well plates were loaded in Automated DNA sequencer for genotyping.

Data Collection and Extraction

The data extracted from Automated DNA sequencer were sized using Gene Scan software and the sizes were extracted using Gene Mapper software version 3.0 The data was exported as text file and imported into excel sheet before submitting it to further Statistical Analysis.

Sequencing

Few representative samples were selected for amplification of targeted region of Myostatin and were sequenced for the confirmation of SNPs.

Cycle Sequencing

DNA sequencing is a powerful technique in molecular biology. It was first devised in 1975. DNA sequence allows analysis of genes at the nucleotide level. Furthermore a molecular biologist can utilize sequencing to identify the site of mutation. There are only a few examples illustrating the way in which DNA sequencing has revolutionized molecular biology. In this work we used

automated DNA sequencer model 3100 ABI PRISM DNA Sequencer. Cycle sequencing has emerged as a powerful alternative to conventional isothermal method.

Cycle sequencing is a simple method in which successive rounds of denaturation, annealing, and extension in a thermal cycler result in linear amplification of extension products (Fig). The products are then injected into a capillary. All current ABI PRISM DNA sequencing kits use cycle sequencing protocols (Marra.M.Weinstok1996).

Automated DNA sequencer is a variation of the traditional Sanger sequence utilizing the cycle sequencing method, in which fluorescence labeled ddNTPs, were used as a chain terminator. Using this method amplified target DNA sequences and obtained data within 24 hours. Only about 500-800bp can be sequenced in one go. And a PCR product of 99.3% accuracy is obtained (Devine, S.E.1994)

Template: Template purity and the optimum concentration are crucial in obtaining the good quality of DNA sequence. PCR products were used as a template for sequencing.

Template purification: PCR product should be clear of dimers and non specific amplifications. It should be free from excess reaction component like dNTPs, Taq polymerase etc. The purification can be done using commercially available columns or Alcohol purification or digestion using exonuclease and phosphotase (Table 10)

The purified PCR product can be used for sequencing

Template	Concentration
100-200 bp	1-3 ng
200-500 bp	3-10 ng
500-1000 bp	5-20 ng
1000-2000 bp	10-40 ng
72000 bp	40-100 ng

Table12: Concentration of PCR-Product used in sequencing

Cycle sequencing reaction/Ready reaction process: -

Reaction is set as per recommendation of AB1 using 8 ul of ready reaction mix in addition to template and primer.

Ready reaction mix:

ABI PRISM^{TM} Ready reaction mix /Dideoxy terminator premix (1000 ul) Applied Biosystems (ABI) consists of: 1.58 mM of ddATP, 94.74 mM ddTTP, 0 .42 uM ddGTP, 47.37 uM ddCTP, 98.95 uM dGTP, 15.79 uM dATP, 15.79 uM dCTP, 15.79 uM dTTP, 168.42 mM Tris-HCl (pH-9), 4.21 mM $(NH4)_2$ $S0_4$, 42.1 mM $MgCl_2$, and 0.42 U/ul ampli Taq DNA polymerase.

PCR Product	50-100 ng
RR Mix	8 ul
5 x Sequencing buffer	4 ul
Primer (Forward/Reverse)	4 pmol
Milli Q Water	Make up the volume 20 ul.

Table13: Pre-sequencing reaction composition

Above contents were mixed briefly and kept in a thermal cycler set at following reaction conditions

Initial Denaturation	-96ºC for 1 minute
Denaturation	-96ºC for 10 seconds
Annealing	-50ºC for 5 seconds
Extension	-60ºC for 4 minutes - 30 cycles
	Hold at 4ºC -ά

For short PCR product, number of cycles can be reduced (e.g., 20 cycles for a 300 bp or smaller fragment).

If the Tm of a primer is >55ºC , the annealing step can be eliminated

If the Tm of a primer is <50ºC, increase the annealing time to 30 seconds or decrease the annealing temperature.

For templates with high GC content (>70%), heat the tubes at 98ºC for 5 minutes before cycling to help denature the template.

Purification of the Sequencing product:

After the sequencing reaction, the products are purified by the following protocol

1. Added 2ul of 125mM EDTA to stop the reaction and mixed well
2. Added 2ul of 3 M Sodium acetate pH 4.6 to each reaction well. Ensure the proper mixing of the contents.
3. Added 50 ul of 95% ethanol to each well and incubate at room temperature for 15 minutes.
4. Spin at a speed of 1650g for 45 minutes at room temperature.
5. Invert the plate on paper towel and give a short spin at 180g for removing supernatant.
6. Added 200ul of 75% ethanol and spin at 1650 g for 5 minutes.
7. Invert the plate slowly on paper towel and spin at 180g for 1 minute.

Denaturation:

8. Add 10 ul of Hi Di Formamide and denature at 95ºC for 5 minutes and chilled on ice (snap chilling) immediately for 5 minutes.
9. The plate can be now loaded in automated DNA sequencer.

ABI 3100 Avant- Automated DNA Sequencer:

- Provided with 4 capillaries with different array sizes (22cm, 36cm, 50cm and 80cm). The large surface area of a capillary allows heat generated during electrophoresis to be dissipated efficiently, allowing high voltage electrophoresis. The result is rapid, high-resolution separation of DNA fragments.
- Polymer POP-6 can be used for sequencing (Performance Optimized Polymer 6, medium used to separate DNA fragments).
- Plate record to be loaded with appropriate model number and sequencing chemistry.
- 96 well plate to be linked and start the run.

Data collection and Extraction:

- After the completion of the electrophoresis (run) chromatogram drawn by data collection software is used to extract the sequence data.
- The sequences with chromatogram can be visualized, edited for vector sequences and further saved by ABI PRISM ® DNA sequencing Analysis Software.

Sequence Analysis

The sequences of representative samples were compared among themselves after sequencing and using Seqscape software the presence of selected SNPs were confirmed.

Statistical Analysis

The following parameters were estimated from the PCR- RFLP, DNA sequence data on the Myostatin gene (GDF-8).

1. The general information about the nucleotide content, polymorphism, nucleotide diversity parameters.
2. Total number of haplotypes detected along with number of haplotypes produced for each population.
3. Haplotype diversity and its sampling variance
4. The estimates of nucleotide diversity (Jukes and Cantor), the average number of nucleotide substitutions per site between two sequences (Lynch and Crease, 1990).
5. Theta per sites i.e. the Watterson estimator, The Eta (n) is the total number of mutations and S the number of sites, which are segregating in a population.
6. The average number of nucleotide differences, K (Tajima 1983) along with stochastic and sampling variances.
7. The Linkage disequilibrium (Kelly 199; Rozas et al., 2001) was also calculated.
8. The DNA Divergence between the two population and was calculated on the basis of average number of nucleotide differences (Tajima, 1983).
9. The genetic diversity analysis for each population was also estimated along with the total data for the four populations.
10. The genetic differentiation analysis based on Haplotype (Hudson et al., 1922) and testing their significance using X^2 (Hudson, 1992 and Nei, 1982) and permutations test (Hudson et al., 1992a).

11. The gene flow analysis was carried out using both haplotype data information and nucleotide sequence data information.
12. The linkage disequilibrium between nucleotide variants as per the methods and procedures given by Liewontin and Kajima 1964' Hill and Robertson, 1968.
13. The distribution of observed pairwise nucleotide site differences (mismatch distribution) was estimated in a stable population i.e. population with constant population size (Slatkin and Hudson 1991, Rogers and Harpending, 1992).
14. The statistical tests D and F proposed by Fu and Li (1993) for testing the selective neutrality of the various mutations was also carried out.
15. The various genetic distances were calculated and the dendrogram were prepared using UPGMA and Neighbor joining algorithm using the individual genetic distances.
16. The population genetic parameters were obtained by using haplotype frequency data. The relationship between the populations was estimated using the Cluster analysis.
17. The haplotypes and their minimum spanning tree for the populations were constructed and depicted.
18. The allelic pattern and allele frequency for all the five loci and four populations were calculated and graphically depicted.
19. The deviation from Hardy Weinberg Law shall be tested using the chi-square statistics.

The genetic distances shall be estimated to find the relationship among the populations based on SNP frequencies of Promoter region of Myostatin gene.

RESULTS AND DISCUSSION

The PCR-RFLP data (Fig 10 and 11) and Snapshot data (Fig 12) was recorded as genotypes and was also converted into sequence data for further analysis. Few representative samples were sequenced by automated DNA sequencer (3100 Avant, Applied Biosystem) and the SNPs were confirmed (fig13-15).

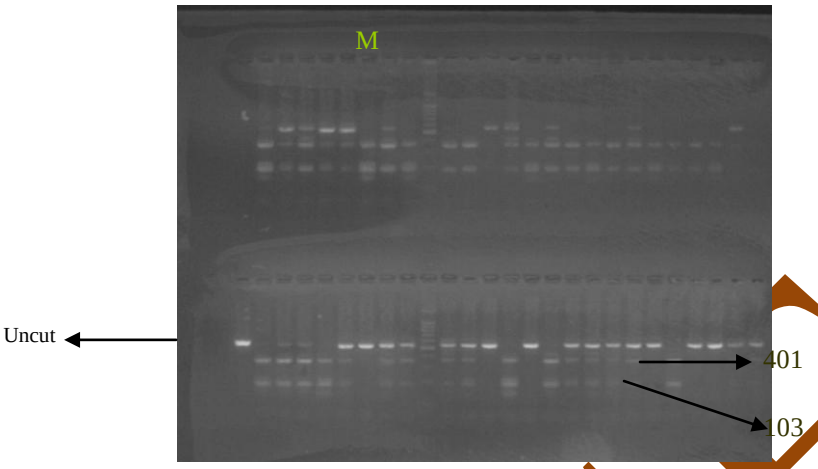


Fig 4: PCR-RFLP pattern of promoter region (G1473A) digested by Nla III & their genotypes are: GG-401, 103, 51 (wild) AA-452, 103, 51 (Mutant) AG-452, 401, 103, 51 (Hetero.)

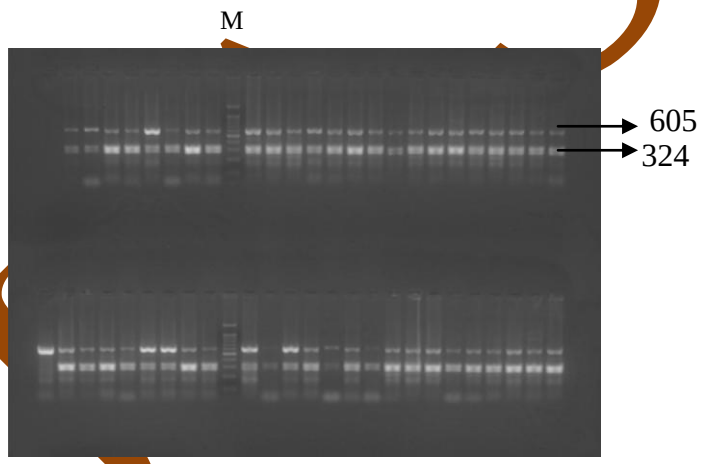
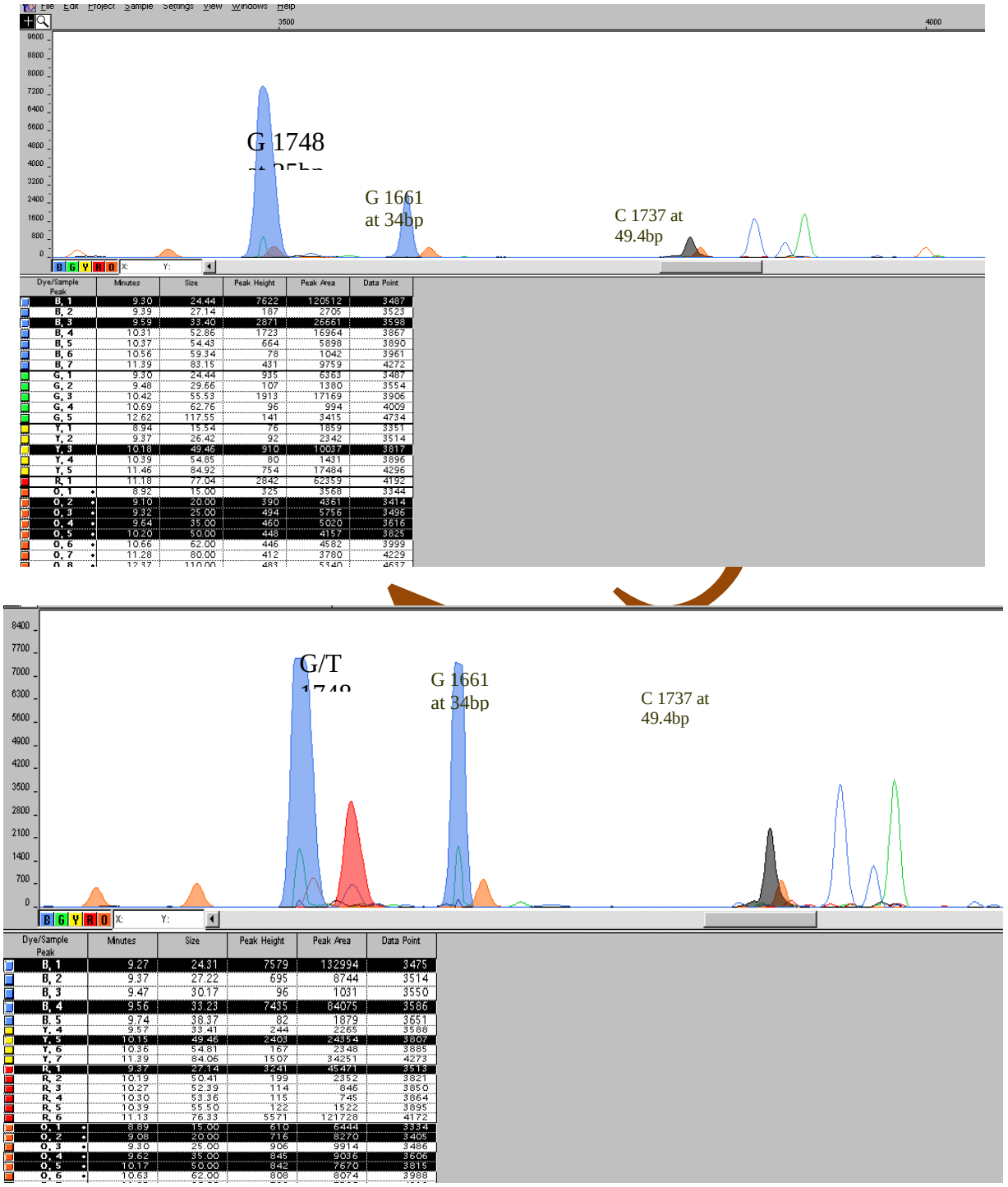


Fig 5: PCR-RFLP pattern of promoter region (C1693A) digested by Sca I & their genotypes are: AA-324, 282 (wild-W) CC- 605 (Mutant-M) and AC- 605, 324, 282 (Heterozygous-H)



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
B, 1	9.30	24.44	7622	120512	3487
B, 2	9.39	27.14	187	2705	3523
B, 3	9.59	33.40	2871	26661	3598
B, 4	10.31	52.86	1723	16964	3867
B, 5	10.37	54.43	664	5898	3890
B, 6	10.56	59.34	78	1042	3961
B, 7	11.39	83.15	431	9759	4272
G, 1	9.30	24.44	935	6383	3487
G, 2	9.48	29.66	107	1380	3554
G, 3	10.42	55.53	1913	17169	3906
G, 4	10.69	62.76	96	994	4009
G, 5	12.62	117.55	141	3415	4734
Y, 1	8.94	15.54	76	1859	3351
Y, 2	9.37	26.42	92	2342	3514
Y, 3	10.18	49.46	910	10037	3817
Y, 4	10.39	54.85	80	1431	3896
Y, 5	11.46	84.92	754	17484	4296
R, 1	11.18	77.04	2842	62359	4192
O, 1	8.92	15.00	325	3568	3344
O, 2	9.10	20.00	390	4361	3414
O, 3	9.32	25.00	494	5756	3496
O, 4	9.64	35.00	460	5020	3616
O, 5	10.20	50.00	448	4157	3825
O, 6	10.66	62.00	446	4582	3999
O, 7	11.28	80.00	412	3780	4229
O, 8	12.37	110.00	483	5340	4637

Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
B, 1	9.27	24.31	7579	132994	3475
B, 2	9.37	27.22	695	8744	3514
B, 3	9.47	30.17	96	1031	3550
B, 4	9.56	33.23	7435	84075	3586
B, 5	9.74	38.37	82	1879	3651
Y, 4	9.97	44.41	244	2265	3588
Y, 5	10.15	49.46	2403	24854	3807
Y, 6	10.36	54.81	167	2348	3885
Y, 7	11.39	84.06	1507	34251	4273
R, 1	9.37	27.14	3241	45471	3513
R, 2	10.19	50.41	199	2352	3821
R, 3	10.27	52.39	114	846	3850
R, 4	10.30	53.34	115	745	3864
R, 5	10.39	55.50	122	1522	3895
R, 6	11.13	76.33	5571	121728	4172
O, 1	8.89	15.00	610	6444	3334
O, 2	9.08	20.00	716	8270	3405
O, 3	9.30	25.00	906	9914	3486
O, 4	9.62	33.00	845	9036	3605
O, 5	10.17	50.00	842	7670	3815
O, 6	10.63	62.00	808	8074	3988

Fig6: Genescan views of Snapshot product run on automated DNA sequencer indicating three SNPs typed (G1661A, C1737T and G1748T) in single PCR product

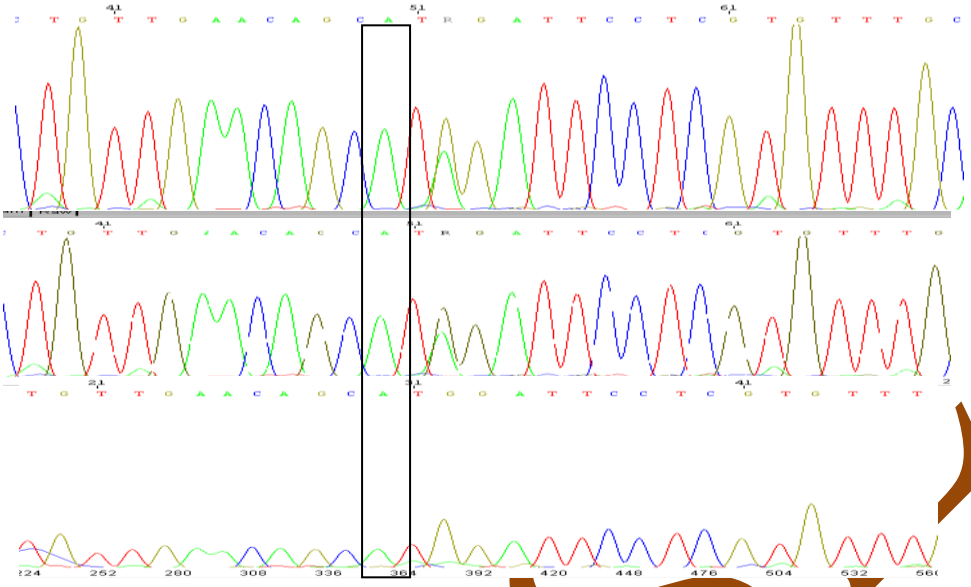


Fig7: G1473A SNP confirmed by sequencing

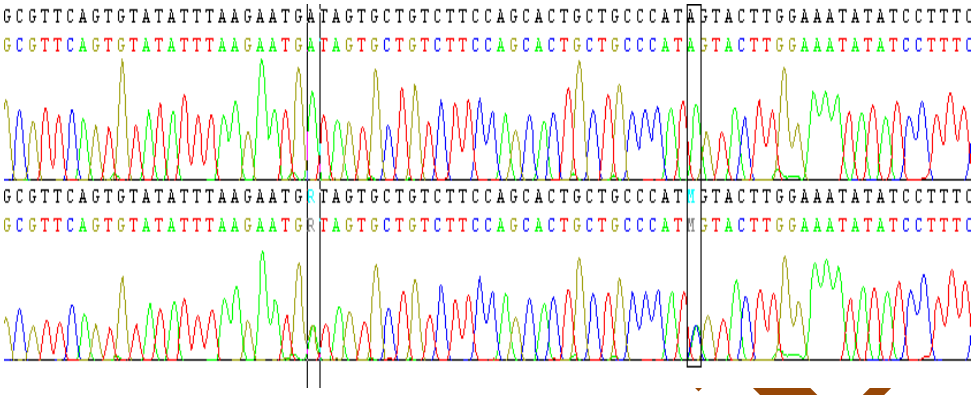


Fig8: G1663A and A1693C SNPs confirmed by sequencing

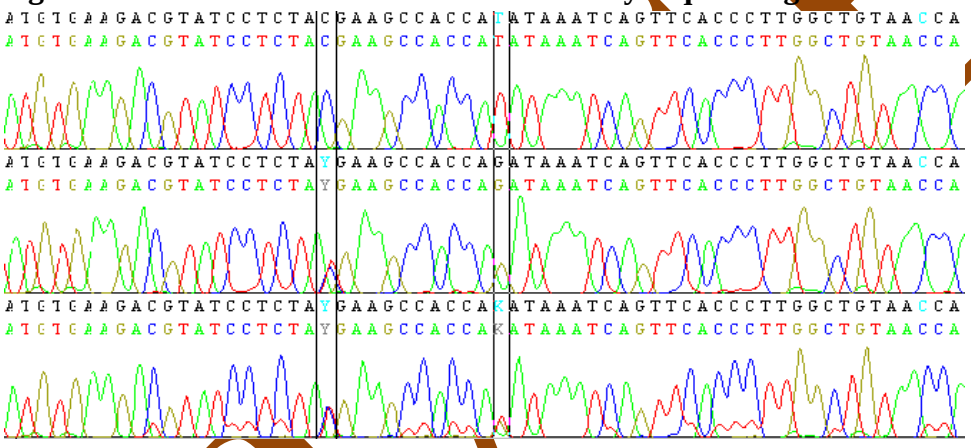


Fig9: C1737T and G1748T SNPs confirmed by sequencing

Nucleotide composition

The total length of the myostatin gene including the three exons and 2 introns and a promoter region in which Promoter region of 1930 nucleotide bases was subjected to data analysis. Five of the nucleotide bases were subject of analysis as these were the SNPs. The data of each individual for each of the five SNPs, which were located at the following positions with respect to the sequence of myostatin gene, was generated.

The location of the 5 SNPs was as follows

Position I in the myostatin gene	1471	Promoter
Position II in the myostatin gene	1661	Promoter
Position III in the myostatin gene	1691	Promoter
Position IV in the myostatin gene	1737	Promoter
Position V in the myostatin gene	1748	Promoter

All the SNPs studied were transitions ie; purines were substituted by purines and pyrimidines by pyrimidines except 3rd and 5th in which transversion SNPs were detected. The nucleotide composition of the sequence of interest was found to be as follows.

Nucleotide composition (Relative values)

C : 20.53%
T : 29.26%
A : 33.14%
G : 17.07%

The composition shall slightly change with substitutions of nucleotides at five mutation sites.

Gene and Nucleotide Diversity

The haplotype (gene) diversity and its sampling variance were estimated using the standard formula (Nei 1987). The values for the haplotype diversity for all the four population data were found to be 0.856. The nucleotide diversity, Pi (p), for the overall data was found to be 0.00079 (Fig 16), the average number of nucleotide differences k was found to be 1.515 with stochastic variance of k as 0.00017 (Nei 1987; Nei and Miller 1990), and its sampling variance as 0.013 (Nei 1987).

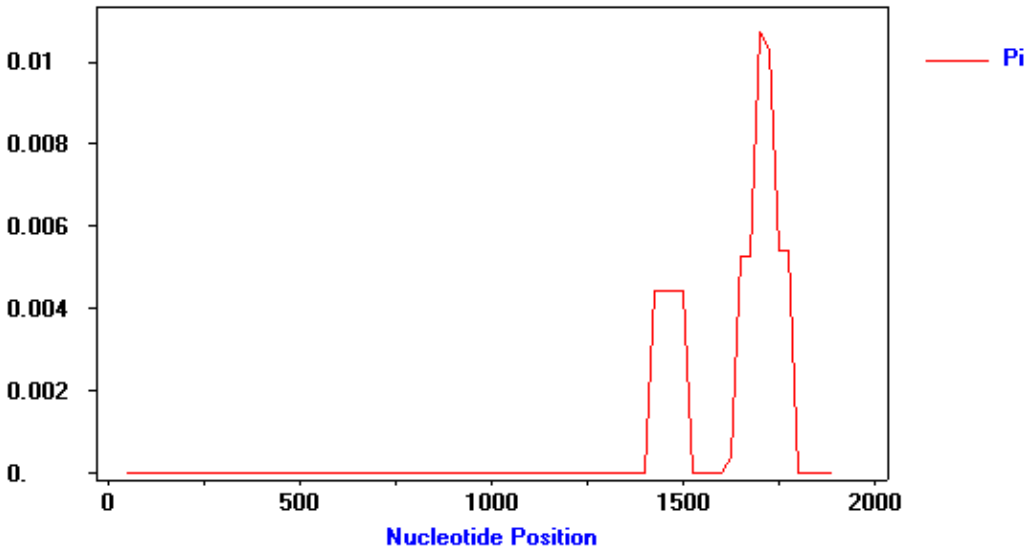


Fig10: Nucleotide diversity in promoter region of Myostatin

When the analysis was carried out population wise the haplotypic diversity was found to be 0.89744,.83784,0.90338 and 0.90687 for Busra, Ghagus, Hblack and Kashmir Favorolla respectively. The nucleotide diversity p$_i$ I values for each of the above mentioned populations were obtained as 0.00024,0.00017,0.00020,0.00021 respectively for the four populations. The values obtained were same for Jukes and Cantor differences. The average number of differences (k) was obtained as 2.12051, 1.52553, 1.72077, and 1.84770 respectively for Busra, Ghagus, Hblack and Kashmir Favorolla.

Linkage Disequilibrium:

The linkage disequilibrium test was conducted using the five positions where we carried out the PCR-RFLP testie; at Locations 1471, 1691, 1737, 1748 nucleotide position of the Myostatin gene. These locations were digested with appropriate restriction enzymes and the PCR- RFLP data was generated. We conducted the Exact test for testing the linkage disequilibrium that existed between the five positions for which the data was generated. The analysis of linkage disequilibrium was carried out separately for each population. The tables 15-18 given below present the significance for the five loci in the populations. The + sign indicates the presence of the linkage disequilibrium (that is the loci are not assorting independently of one another but move together or are tightly linked).

Locus	1471	1691	1737	1748
1471	*	-	-	-

1691	-	*	+	-
1737	-	+	*	+
1748	-	-	+	*

Table14: Significant linkage disequilibrium (significance level= .05) at 4 different sites in Busra.

In Busra poultry none of the nucleotide positions are significantly linked to one another and assort independently except 1691 and 1737, the number itself shows closeness to one another. However in case of Ghagus poultry there is no statistically significant linkage disequilibrium.

Locus	1471	1661	1691	1737	1748
1471	*	+	-	-	-
1661	+	*	-	-	-
1691	-	-	*	-	-
1737	-	-	-	*	-
1748	-	-	-	-	*

Table15: significant linkage disequilibrium (significance level=0.0500) in Ghagus poultry

In Haringhata Black poultry the positions 1661,1737 and 1748 are linked to one another and the linkage is statistically significant. Significant linkage disequilibrium (significance level=0.0500) in Haringhata Black poultry are given in table 16. In Kashmir Favorolla there is statistically significant

Locus	1471	1661	1691	1737	1748
1471	*	-	-	-	-
1661	-	*	-	+	+
1691	-	-	*	-	-
1737	-	+	-	*	+
1748	-	+	-	+	*

Table16: Significant linkage disequilibrium (significance level=0.0500) in Haringhata Black poultry

Locus	1471	1691	1737	1748
1471	*	-	-	-
1691	-	*	+	-
1737	-	+	*	-
1748	-	-	-	*

Table17: Significant linkage disequilibrium (significance level=0.0500) in Kashmir favorolla poultry

Population differentiation on the basis of Nucleotide Data

We carried out the population differentiation analysis (F_{ST} test) while the confidence intervals were obtained using permutation test to find out the significance of the values obtained. The values obtained for the F_{ST} between populations and their statistical significance is depicted below in table 19.

Distance method: Pairwise difference

	Busra	**Ghagus**	**Hblack**	**K Favorolla**
Busra	0.000			
Ghagus	0.103*	0.000		
HBlack	0.116*	0.116*	0.000	
K Favorolla	0.110*	0.042*	0.091*	0.000

Table18: Significance values of population differentiation analysis

* Statistically significant based on 1100 permutations for the estimation of probability values.

The table18 and the significance obtained reveal that all the four population of indigenous chicken undertaken for this study have clear distinctivene ss.

Haplotypes obtained

The data generated using the PCR- RFLP technique was used for completing the sequence at the five locations on which the experiment was carried out. The alignment of sequences was done and the data was fed into the software to calculate the total number of haplotypes obtained in the study. A total of 18 haplotypes were obtained in the present study. The data was however also subjected to analysis for the determination of haplotypes in each of the four poultry populations under study. The combinations of the five-nucleotide positions responsible for the generation of number of haplotypes have been given in the Table 19.

Haplotype No.	Nucleotide (5)	Number of individuals with the Haplotype in four populations.
Hap_1	GGATG	10
Hap_2	AGACT	24
Hap_3	GGCCT	52
Hap_4	GGCTG	18
Hap_5	GGACT	40
Hap_6	AGCCT	5
Hap_7	AACCG	3
Hap_8	AGATT	16
Hap_9	AGCTT	4
Hap_10	AAACG	4
Hap_11	AACTT	4
Hap_12	AAATT	8
Hap_13	AGCTG	6
Hap_14	GGACG	1
Hap_15	GGCCG	1
Hap_16	AGATG	1
Hap_17	AGACG	1
Hap_18	AGCCG	1

Table20: Details of different haplotypes obtained

The population wise study revealed the distribution of the haplotypes and all the haplotypes were not present in all the populations. A total of 9 haplotypes were found in Busra and Ghagus also had 9 haplotypes present in its population. The values for HBlack and Kashmir Favorolla were 11 haplotypes each. The haplotype distribution giving number of individuals with for each of the four poultry populations is given in table 21. The Interhaplotypic distances between different breeds are mentioned in table 22-25.

Haplotype:	Busra	HBlack	KashFav	Ghagus
	44	42	63	48
Hap_1	9	1	0	0
Hap_2	10	7	6	1
Hap_3	13	6	23	10
Hap_4	1	13	2	2
Hap_5	1	5	22	12
Hap_6	0	2	0	3
Hap_7	1	3	0	0
Hap_8	5	0	0	11
Hap_9	3	0	0	1
Hap_10	1	1	2	0
Hap_11	0	1	0	7
Hap_12	0	2	3	1
Hap_13	0	1	0	0
Hap_14	0	0	1	0
Hap_15	0	0	1	0
Hap_16	0	0	1	0
Hap_17	0	0	1	0
Hap_18	0	0	1	0

Table20: Haplotypes and their distribution among populations

	Hap_1	Hap_2	Hap_3	Hap_4	Hap_5	Hap_7	Hap_8	Hap_9	Hap_10
Hap_1		0.9997	1.4135	1.4135	1.7307	1.7307	0.9997	1.4135	1.9979
Hap_2			0.9997	0.9997	1.4135	1.4135	1.4135	0.9997	1.7307
Hap_3				1.4135	0.9997	1.7307	0.9997	1.4135	1.4135
Hap_4					0.9997	0.9997	1.7307	1.4135	1.4135
Hap_5						1.4135	1.4135	1.7307	0.9997
Hap_7							1.9979	0.9997	0.9997
Hap_8								1.7307	1.7307
Hap_9									1.4135

Table21: Interhaplotypic distances in Busra breed of Poultry.

	Hap_3	Hap_4	Hap_5	Hap_6	Hap_8	Hap_9	Hap_11	Hap_12
Hap_2	0.9997	0.9997	1.4135	1.4135	1.4135	0.9997	1.7307	1.4135
Hap_3		1.4135	0.9997	1.7307	0.9997	1.4135	1.4135	0.9997
Hap_4			0.9997	0.9997	1.7307	1.4135	1.4135	1.7307
Hap_5				1.4135	1.4135	1.7307	0.9997	1.4135
Hap_6					1.4135	1.7307	0.9997	1.9979
Hap_8						1.7307	0.9997	1.4135
Hap_9							1.9979	0.9997
Hap_11								1.7307

Table22: Interhaplotypic distance in Ghagus population

	Hap_1	Hap_2	Hap_3	Hap_4	Hap_5	Hap_6	Hap_7	Hap_10	Hap_11	Hap_12	Hap_13
Hap_1		0.9997	1.4135	1.4135	1.7307	0.9997	1.7307	1.9979	1.4135	1.7307	1.9979
Hap_2			0.9997	0.9997	1.4135	1.4135	1.4135	1.7307	1.7307	1.4135	1.7307
Hap_3				1.4135	0.9997	1.7307	1.7307	1.4135	1.4135	0.9997	1.4135
Hap_4					0.9997	0.9997	0.9997	1.4135	1.4135	1.7307	1.4135
Hap_5						1.4135	1.4135	0.9997	0.9997	1.4135	0.9997
Hap_6							1.4135	1.7307	0.9997	1.9979	1.7307
Hap_7								0.9997	1.7307	1.4135	1.7307
Hap_10									1.4135	0.9997	1.4135
Hap_11										1.7307	1.4135
Hap_12											1.7307
Hap_13											

Table23: Interhaplotypic distance in Hblack breed of Poultry.

	Hap_3	Hap_4	Hap_5	Hap_10	Hap_12	Hap_14	Hap_15	Hap_16	Hap_17	Hap_18
Hap_2	0.9997	0.9997	1.4135	1.7307	1.4135	1.7307	1.9979	1.4135	2.2332	1.9979
Hap_3		1.4135	0.9997	1.4135	0.9997	1.4135	1.7307	0.9997	1.9979	1.7307
Hap_4			0.9997	1.4135	1.7307	1.9979	1.7307	1.7307	1.9979	2.2332
Hap_5				0.9997	1.4135	1.7307	1.4135	1.4135	1.7307	1.9979
Hap_10					0.9997	1.4135	0.9997	1.7307	1.4135	1.7307
Hap_12						0.9997	1.4135	1.4135	1.7307	1.4135
Hap_14							0.9997	1.7307	1.4135	0.9997
Hap_15								1.9979	0.9997	1.4135
Hap_16									1.7307	1.4135
Hap_17										0.9997

Table24: Interhaplotypic distance in Kashmir Favorolla breed

The examination of the presence and absence of the haplotypes in the four poultry populations reveals that Haplotypes 14,15,16,17 and 18 are present only in Kashmir favorolla while Haplotype 13 are present only in Haringhata black poultry.

An examination of the sharing of haplotypes among the four populations shall reveal that Ghagus and Busra share 6 haplotypes while 7 are shared between HBlack and Busra. Kashmir Favorolla and Busra share 5 haplotypes among themselves. Ghagus shares 7 haplotypes each with HBlack and 5 haplotypes with Kashmir Favorolla. Kashmir Favorolla and HBlack share 6 haplotypes among themselves. The distribution of the haplotypes and their relative frequency are important as they form the data for further analysis of the population genetic data and to find the genetic distances and relationship among the poultry populations. The relative frequencies of the haplotypes in the four poultry populations are given in table 26. Relative Frequency of each haplotype in the four poultry populations and their graphical representation is given below.

Haplotype:	Busra	HBlack	KashFav	Ghagus
	44	42	63	48
Hap_1	0.205	0.0238	0	0
Hap_2	0.227	0.167	0.0952	0.0208
Hap_3	0.295	0.143	0.365	0.208
Hap_4	0.0227	0.31	0.0317	0.0417
Hap_5	0.0227	0.119	0.349	0.25
Hap_6	0	0.0476	0	0.0625
Hap_7	0.0227	0.0714	0	0
Hap_8	0.114	0	0	0.229
Hap_9	0.0682	0	0	0.0208
Hap_10	0.0227	0.0238	0.0317	0
Hap_11	0	0.0238	0	0.146
Hap_12	0	0.0476	0.0476	0.0208
Hap_13	0	0.0238	0	0
Hap_14	0	0	0.0159	0
Hap_15	0	0	0.0159	0
Hap_16	0	0	0.0159	0
Hap_17	0	0	0.0159	0
Hap_18	0	0	0.0159	0

Table25: **Relative frequencies of each haplotypes in different poultry populations**

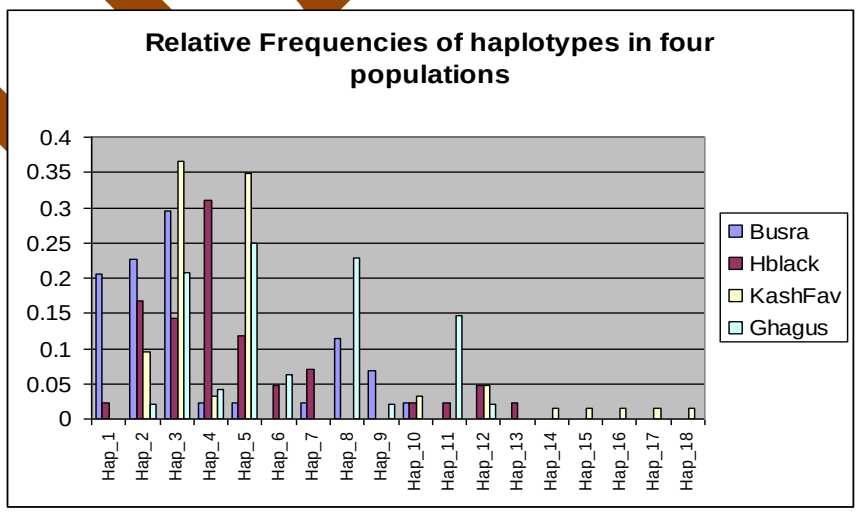


Fig11: Graphical presentation of relative frequency of different haplotypes

Phylogenetic tree construction: We constructed the phylogenetic tree based on the nucleotide differences obtained. The Neighbour joining algorithm was utilized for the purpose. We utilized the software for phlylogenetics for arriving at a tree. The visualization of the tree gave all the individuals belonging to the population clustered together giving credence to the tree thus obtained. The diagrammatic representation of the tree is as follows (Fig 18)

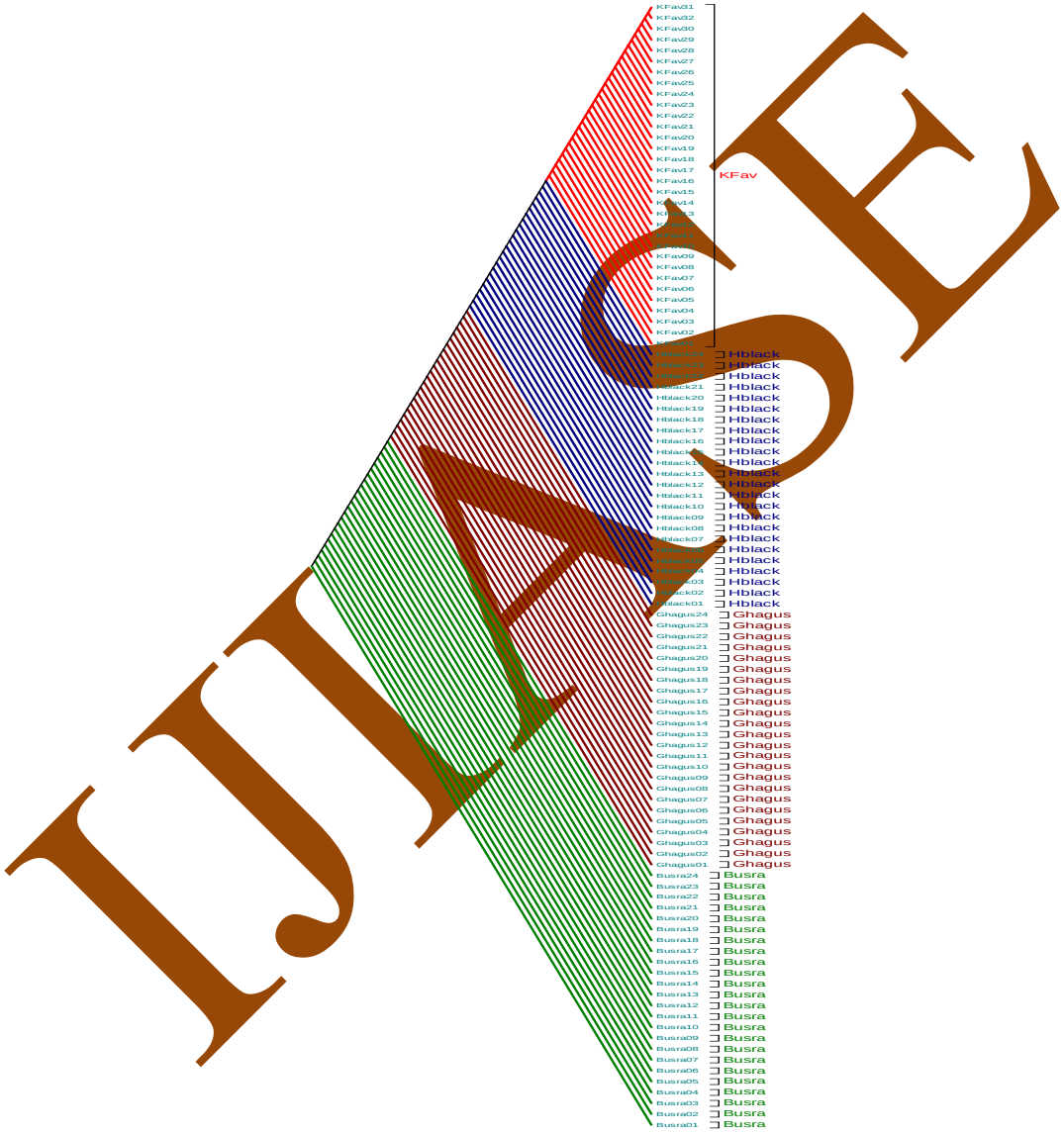


Fig12: Diagrammatic representation of the tree

The phylogenetic tree was also prepared using UPGMA algorithm and the tree thus prepared has been depicted below (fig 19)

Fig13: UPGMA TREE OF FOUR POPULATIONS.

Haplotype and Minimum Spanning Tree

The data for the four populations was subjected for further analysis to find out the minimum spanning tree, which gives the relationship among the haplotypes based on the mutations, and the changes, which are responsible for making them distinctive. As already mentioned there were in total 18 haplotypes while the number of haplotypes for each of the populations varied one from the other. The pictures given below (fig 20-23) provide the minimum spanning tree for each of the four populations for which the data was generated.

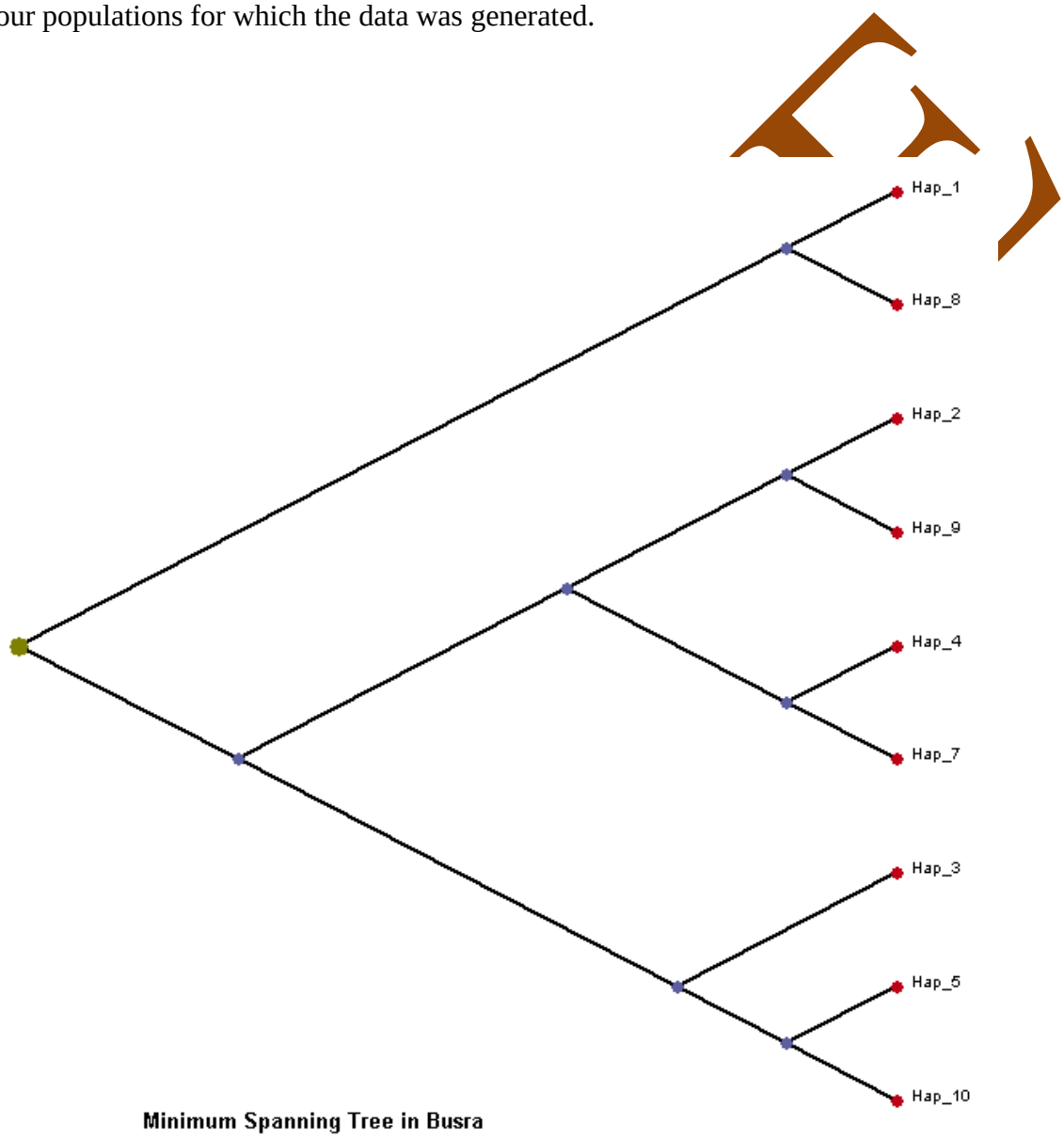
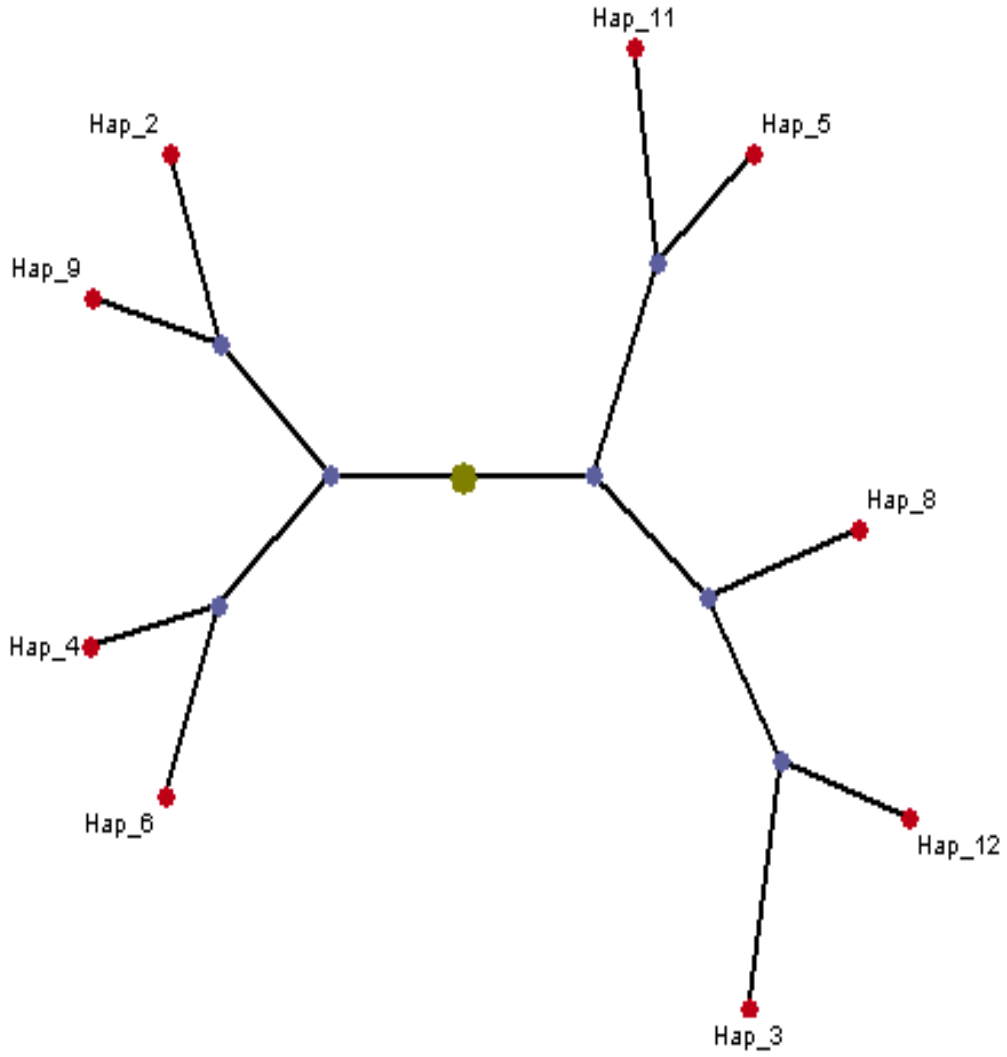


Fig14: Minimum Spanning Tree of Busra population.



Minimum Spanning Tree of haplotypes in Ghagus

Fig15: Minimum Spanning Tree of Ghagus.

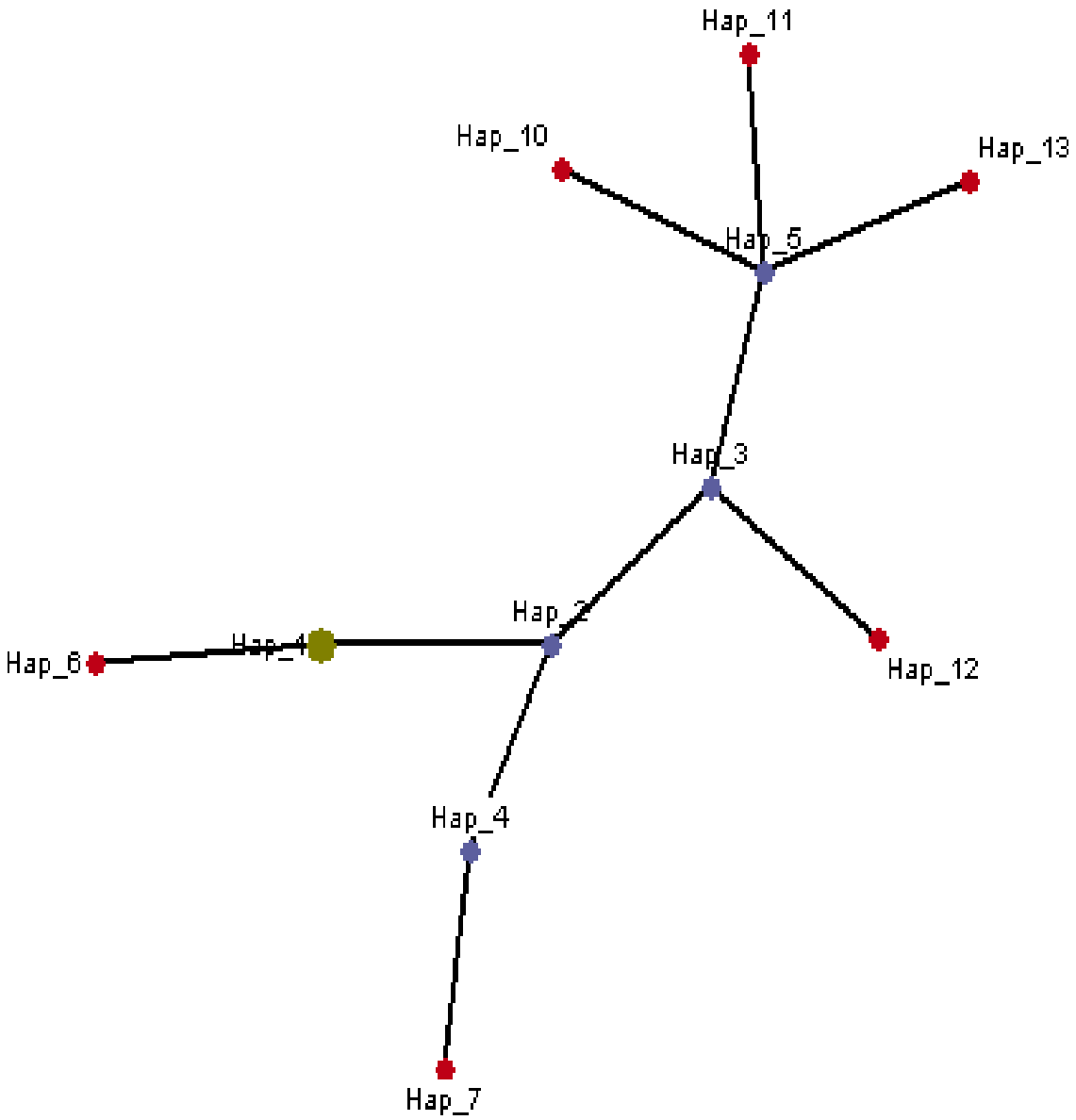


Fig16: Minimum Spanning Tree for Hblack population.

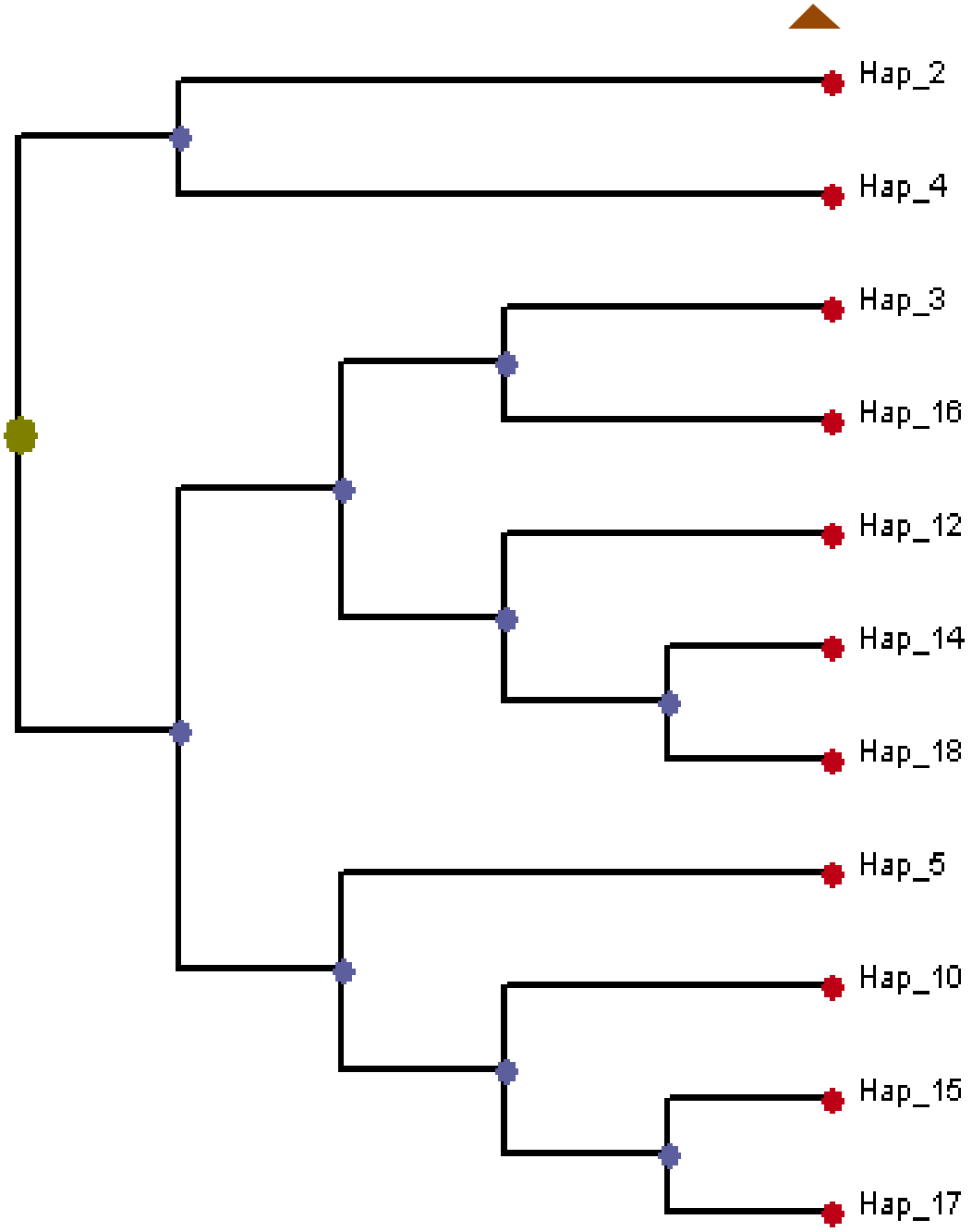


Fig 17: Minimum Spanning Tree for Kashmir Favorolla population.

Mismatch Distribution based on Nucleotide differences:

Episodes of population growth and decline leave characteristic signatures in the distribution of nucleotide site differences between pairs of individuals. These signatures appear in the histograms. The differences in the expected and observed shall tell us the demographic features of the populations. The analysis was carried out using suitable softwares and the diagrams are depicted below (Fig 24).

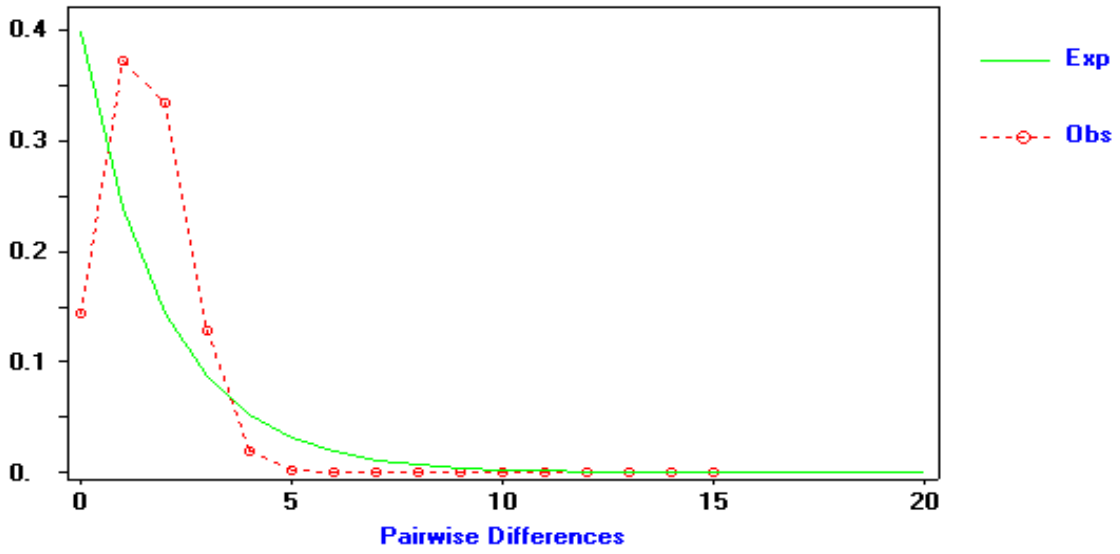


Fig 18: Mismatch Distribution based on Nucleotide differences based on constant population size.

The data reveals that there is no major significant difference among the expected and the observed values and a minor bump at the initial stage just signifies a rapid increase in the population which was not sustained and the population remained constant thereafter. When it was expected that the population underwent growth and then declined the graphical representation was found as follows.

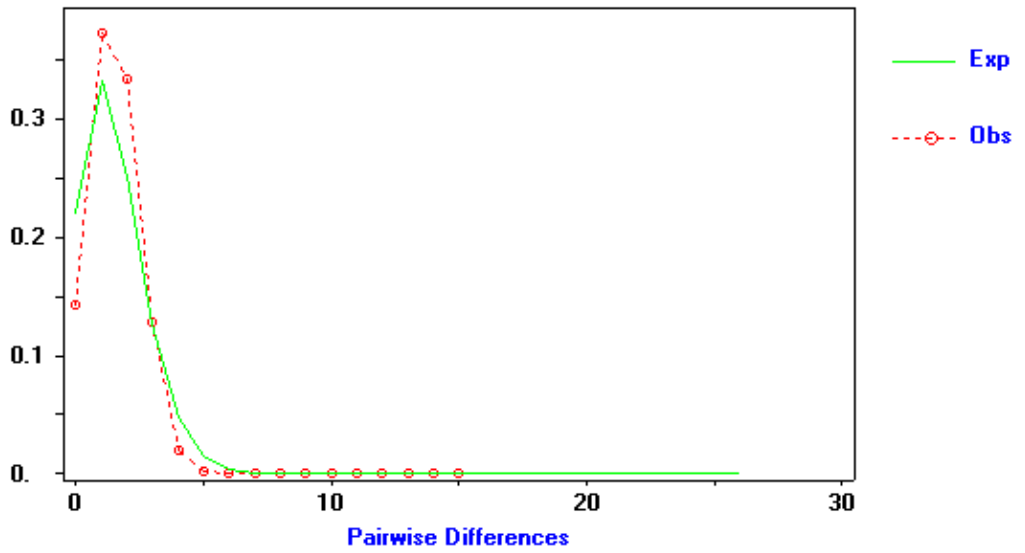


Fig 19: Mismatch Distribution based on Nucleotide differences based on growth decline population.

The figure represents a past growth followed by constant population being maintained in the population tracts. There has been no rapid changes in the demography of the populations ie; Bottleneck or rapid expansion of the population.

Data generation and Allelic distribution of Five SNPs of Myostatin gene

We utilized two different enzymes Nla III and Sca I for the two targeted SNPs on promoter region while for other three SNPs SNAPSHOT method was used . On resolution on Agarose gel we obtained all the three genotypes for the populations under consideration. The data obtained for each individual of each population was subjected to statistical analysis. The allelic frequency of each locus and each population was estimated and has been presented graphically in figure 26.

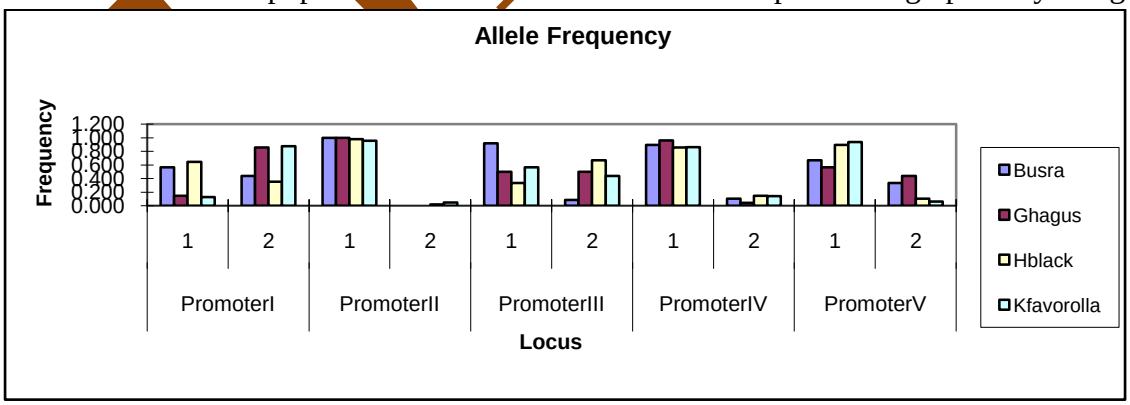


Fig20: Allele frequency of five selected SNPs in promoter region

As far as the allelic pattern for the loci (Fig 27) and their gene frequency was concerned the statistical analysis revealed the following pattern in an overall manner over all the loci and populations studied.

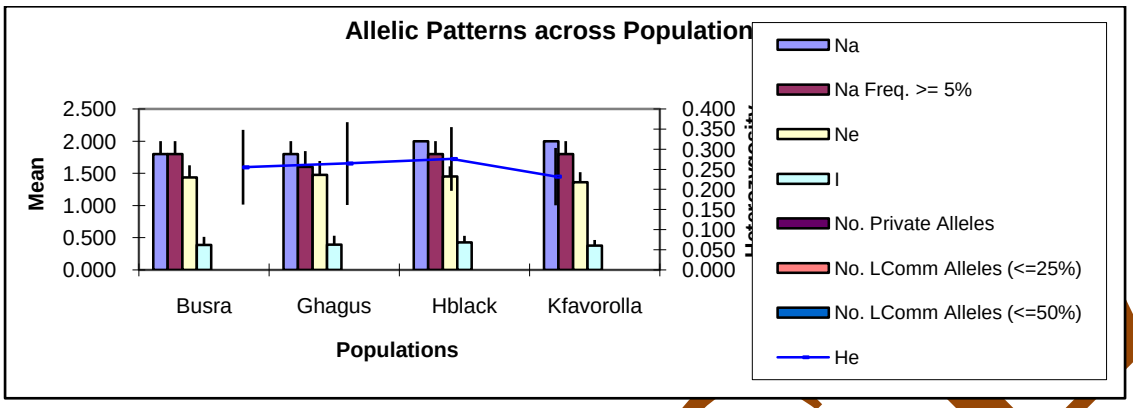


Fig21: Allelic pattern across four different populations

The data obtained for the four populations and over all loci was subjected to statistical analysis and heterozygosity values were estimated on the basis of the genotypes obtained. The expected heterozygosity values were estimated from the allele frequencies. The observed and the expected heterozygosity values in different populations are depicted graphically in figure 28-31.

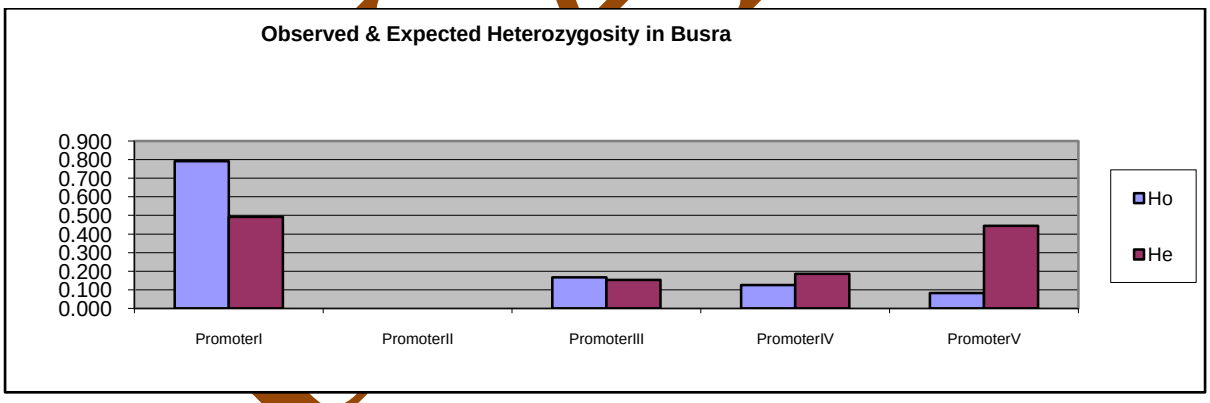


Fig22: Observed and expected heterozygosity in Busra populations

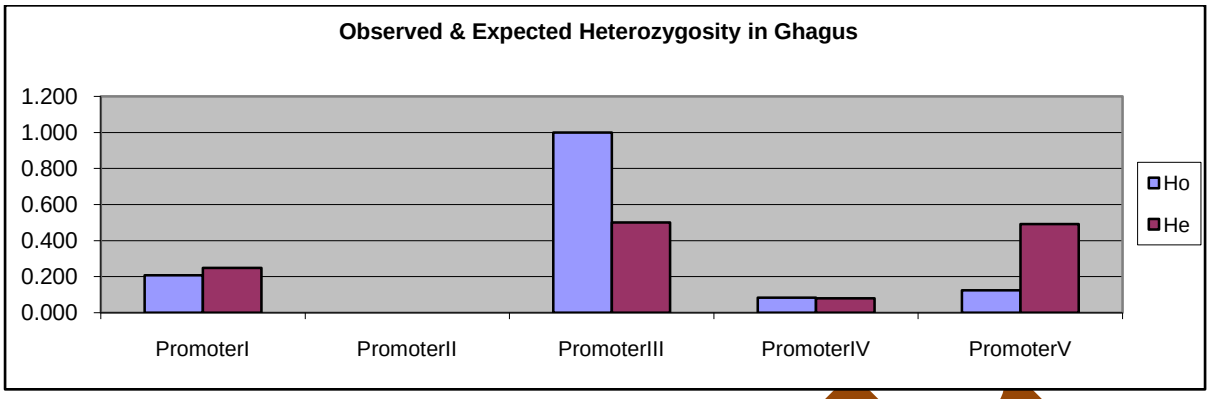


Fig23: Observed and expected heterozygosity in Ghagus populations

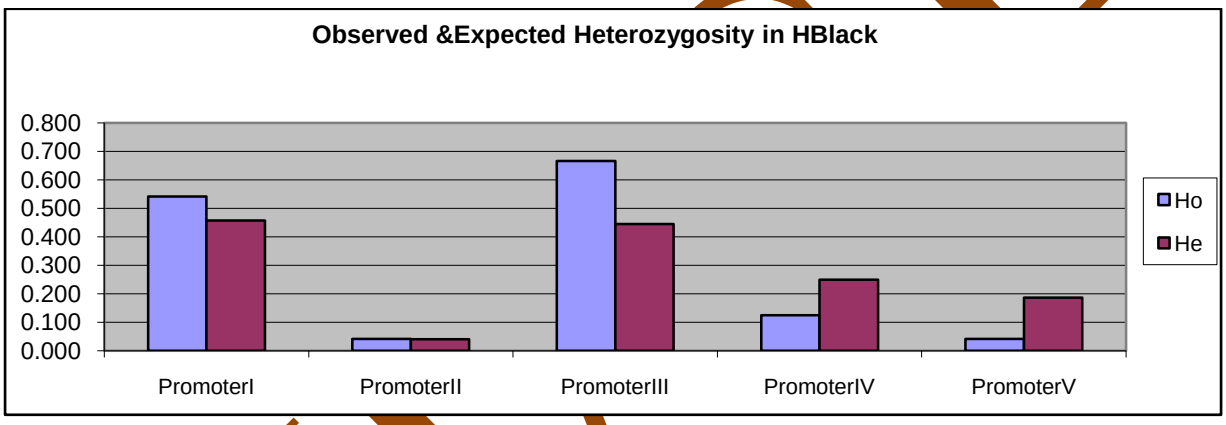


Fig24: Observed and expected heterozygosity in Hblack populations

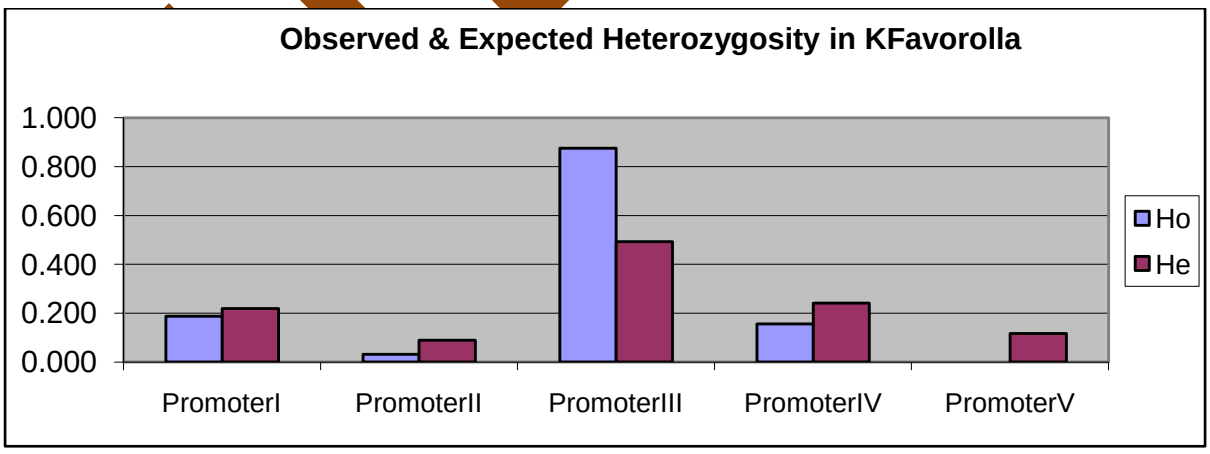


Fig25: Observed and expected heterozygosity in Kashmir favorolla populations

Hardy Weinberg Equilibrium:

The allele frequency data obtained was subjected to testing for Hardy Weinberg Equilibrium using the normal chi-square statistics using the differences between observed and expected number of genotypes. The locus-wise, population-wise data has been graphically represented below (Fig 32-35).

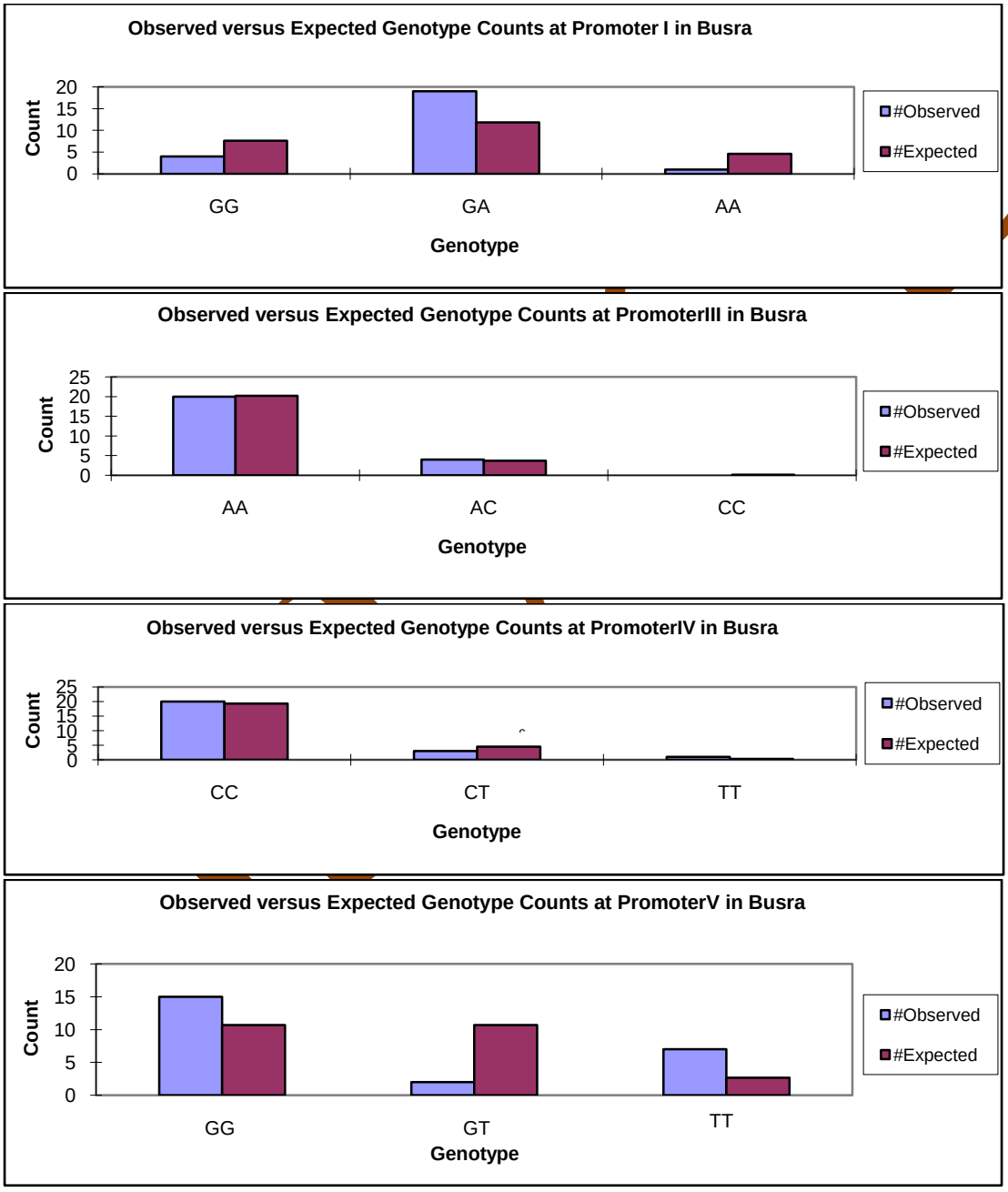


Fig26: Observed and expected genotypes counts in Busra population

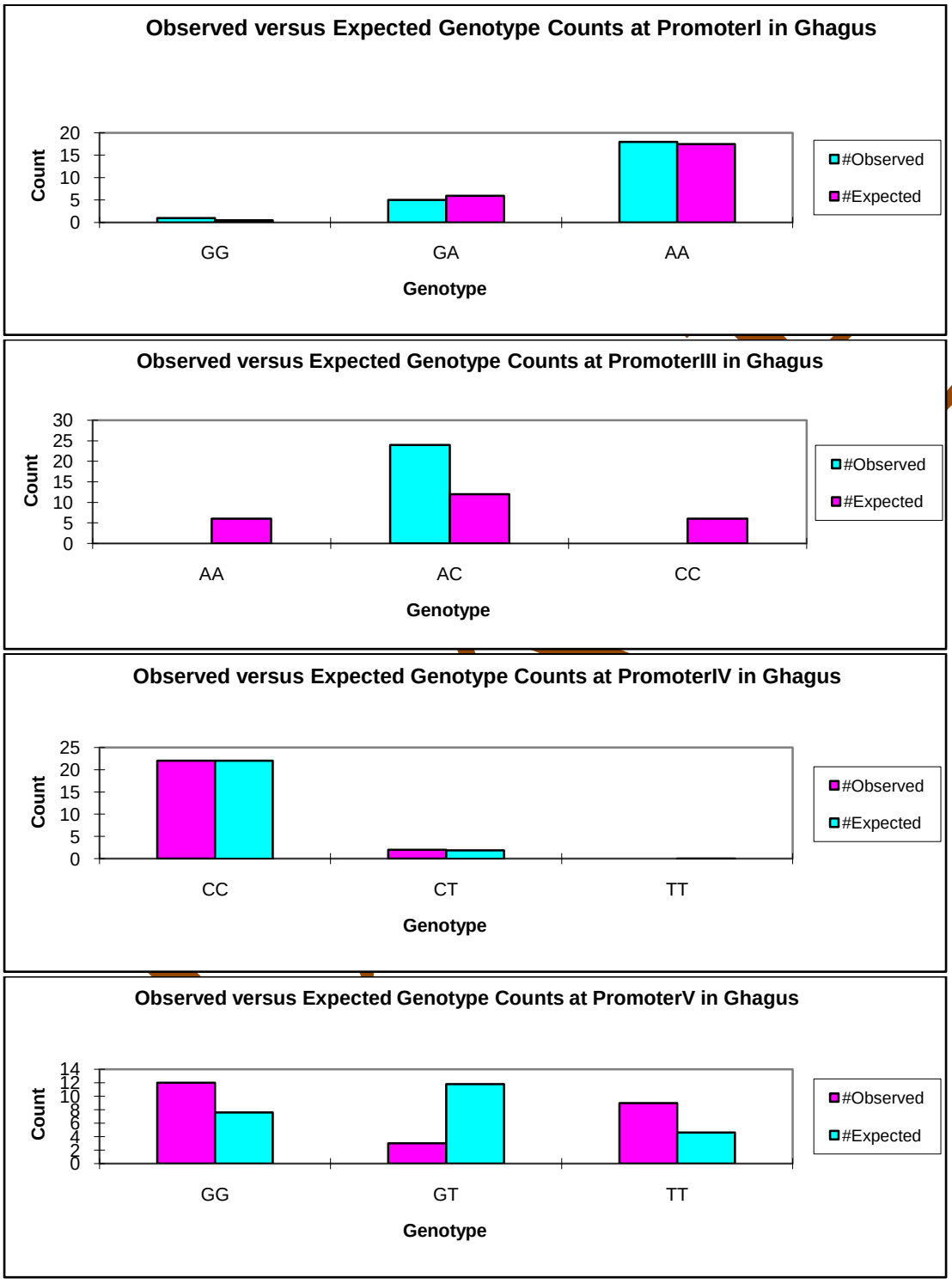
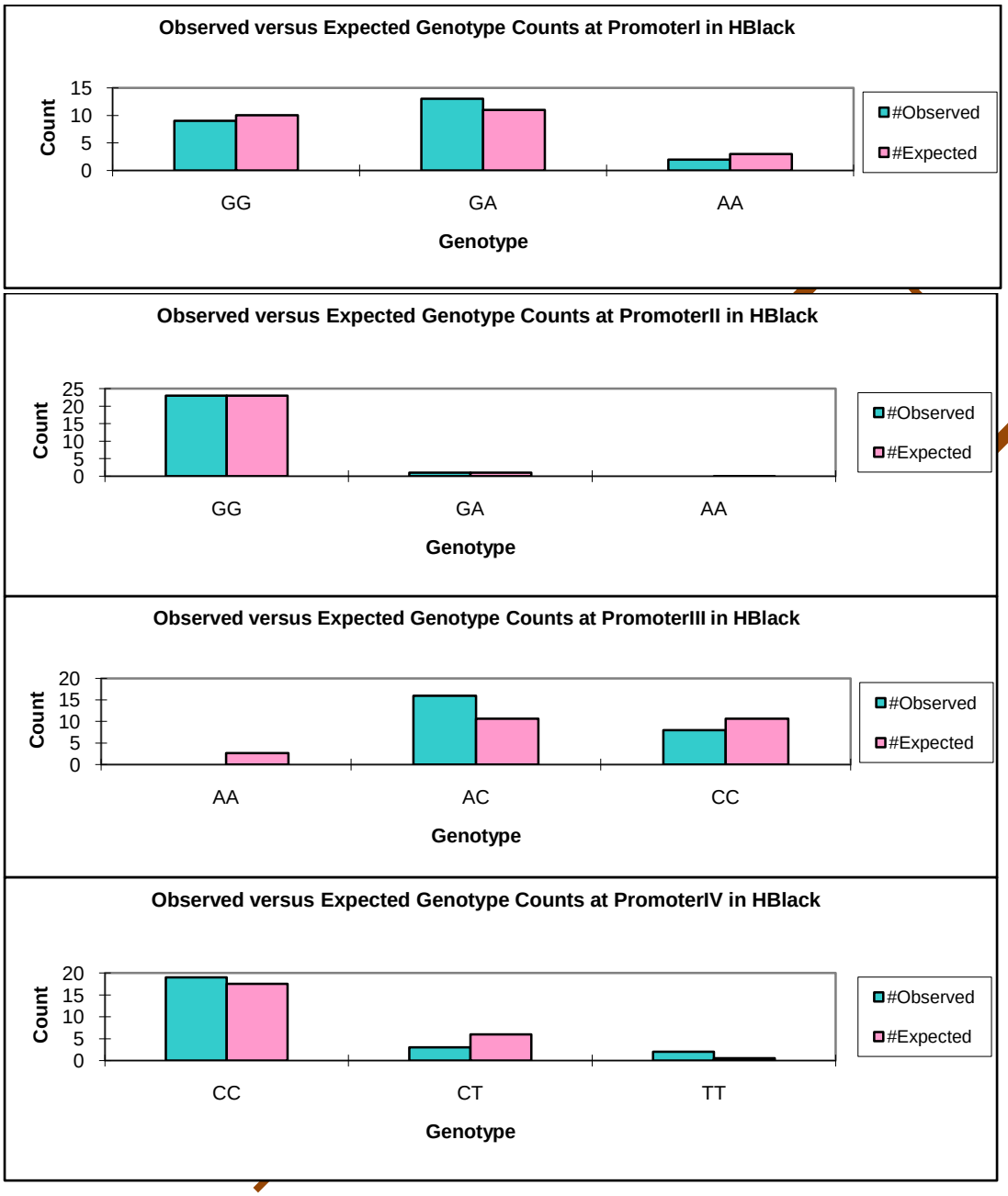


Fig27: Observed and expected genotypes counts in Ghagus population



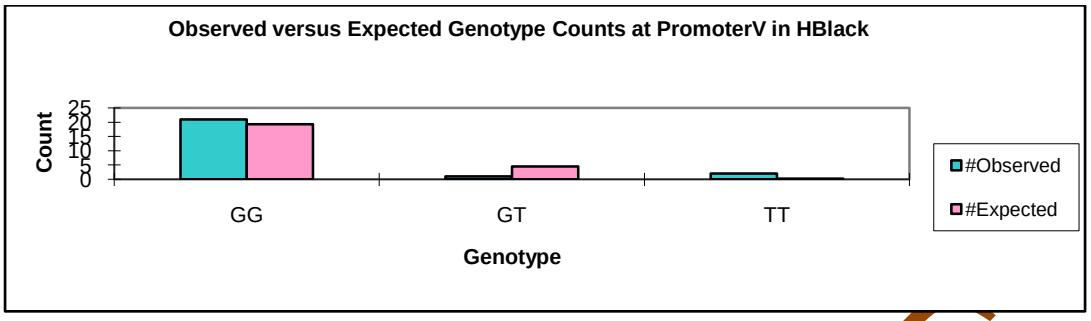
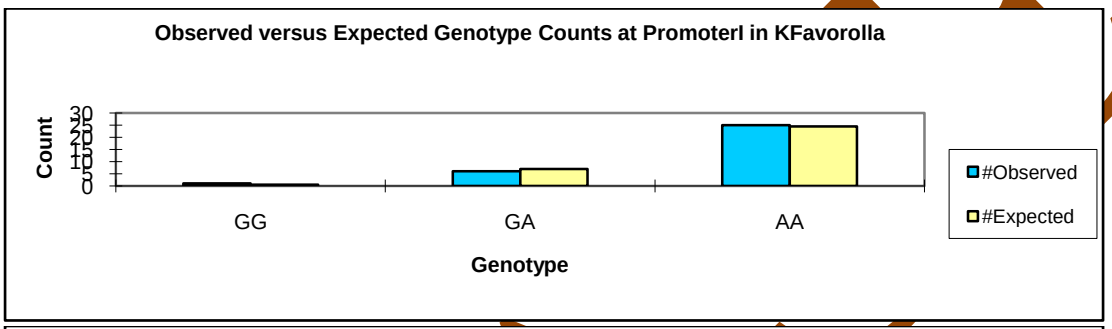
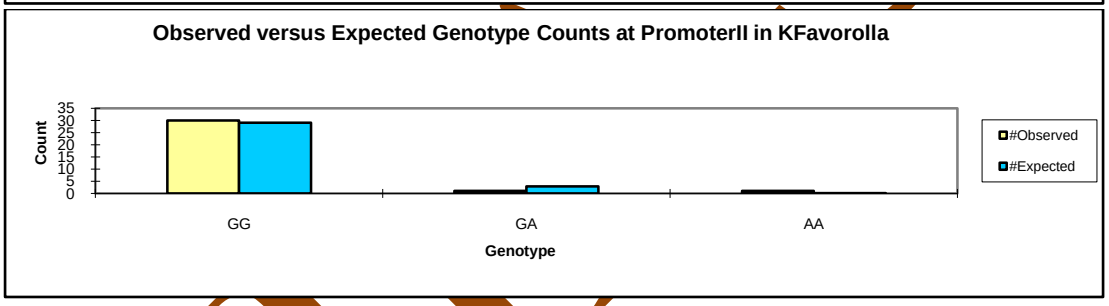


Fig28: Observed and expected genotypes counts in Hblack population





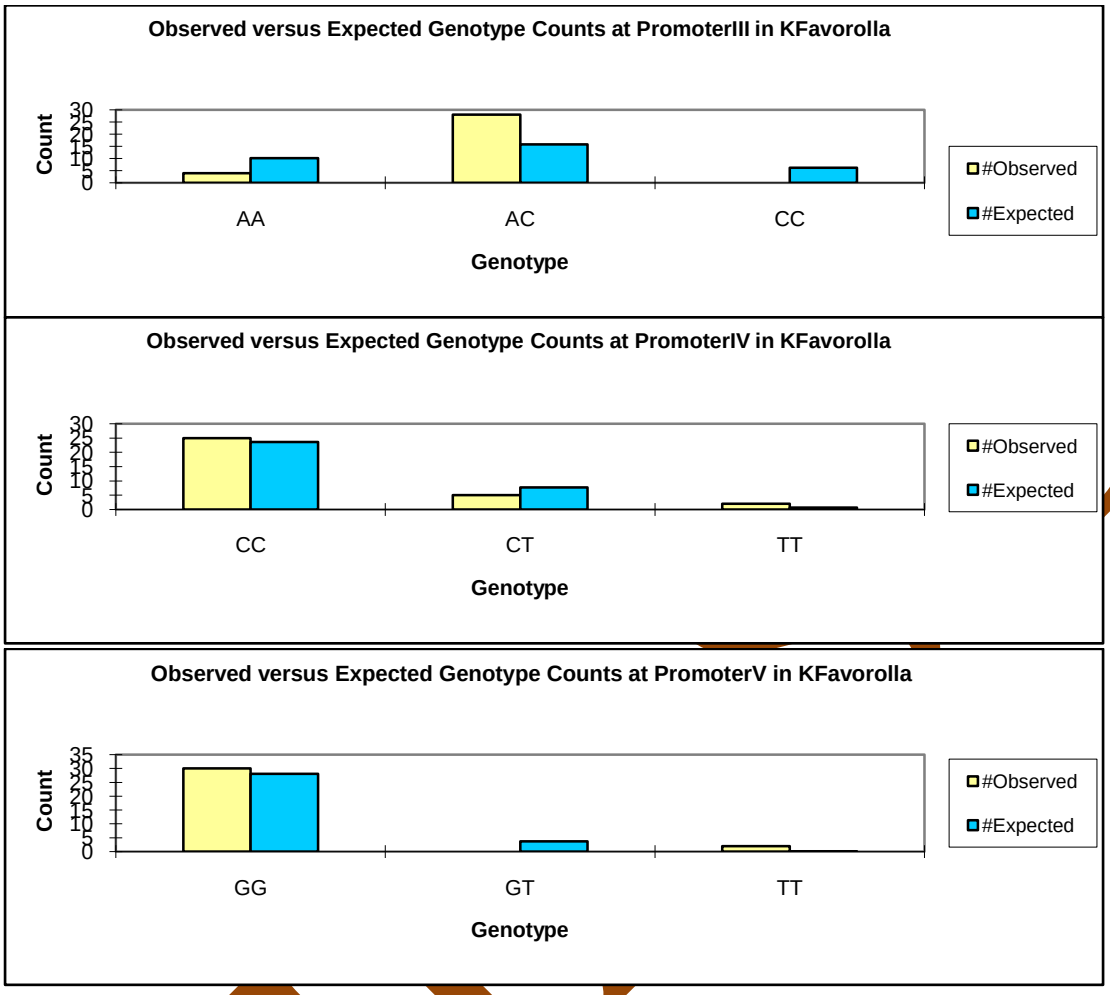


Fig29: Observed and expected genotypes counts in Kfavorolla population

The chi-square values were obtained by using the observed and expected genotypes based on the allele frequency obtained for each of the five loci in each of the populations. The probability values were estimated for testing the statistical significance. The analysis of data gave the following chi-square values and the probability values along with the significance level. The analyzed data has been depicted in the form of a table 27 given below.

Pop	Locus	DF	ChiSq	Prob	Signif
Busra	PromoterI	1	8.886	0.003	**
Busra	PromoterII	Monomorphic			
Busra	PromoterIII	1	0.198	0.656	Ns
Busra	PromoterIV	1	2.617	0.106	Ns
Busra	PromoterV	1	15.844	0.000	***
Ghagus	PromoterI	1	0.644	0.422	Ns

Ghagus	PromoterII	Monomorphic			
Ghagus	PromoterIII	1	24.000	0.000	***
Ghagus	PromoterIV	1	0.045	0.831	Ns
Ghagus	PromoterV	1	13.358	0.000	***
Hblack	PromoterI	1	0.813	0.367	Ns
Hblack	PromoterII	1	0.011	0.917	ns
Hblack	PromoterIII	1	6.000	0.014	*
Hblack	PromoterIV	1	5.958	0.015	*
Hblack	PromoterV	1	14.480	0.000	***
Kfavorolla	PromoterI	1	0.653	0.419	ns
Kfavorolla	PromoterII	1	13.531	0.000	***
Kfavorolla	PromoterIII	1	19.358	0.000	***
Kfavorolla	PromoterIV	1	4.000	0.046	*
Kfavorolla	PromoterV	1	32.000	0.000	***

Table26: Chi-square and probability values with significance levels

The SNP of promoter (II) of both Busra and Ghagus was monomorphic and thus the analysis could not be proceeded further. In case of Busra SNPs of Promoter (I and V) were significantly deviating from the HWE. In case of Ghagus SNP of Promoter region (III and V) were found to be deviating from the HWE. In case of Haringhata Black SNP of Promoter region (I and II) were in HWE while all the other three promoter SNPs were deviating from HWE. In Kashmir Favorolla all the Promoter SNPs except the 1st SNP were in HWE.

Genetic distance: The allele frequency data was utilized to find the Nei's Standard Genetic distances and relationship among the populations. The perusal of the genetic distances revealed that the Kashmir Favorolla population is a distinctive identity and joins the phylogenetic tree lastly. It is supported by the fact that at least four of the Haplotypes are private to this population and thus not available in any other population. The genetic distances along with the similarities have been given below in Table 28 and the dendrogram prepared on the basis of genetic distances is presented graphically.

pop ID	Busra	Ghagus	HBlack	KFavorolla
Busra	****	0.9022	0.8907	0.8960
Ghagus	0.1029	****	0.8904	0.9586
HBlack	0.1158	0.1160	****	0.9130

KFavorolla	0.1098	0.423	0.0911	****

Table27: Nei's Genetic distance among four poultry breeds of India.

The values above diagonal are the genetic similarities while the values below the diagonal are genetic distances.

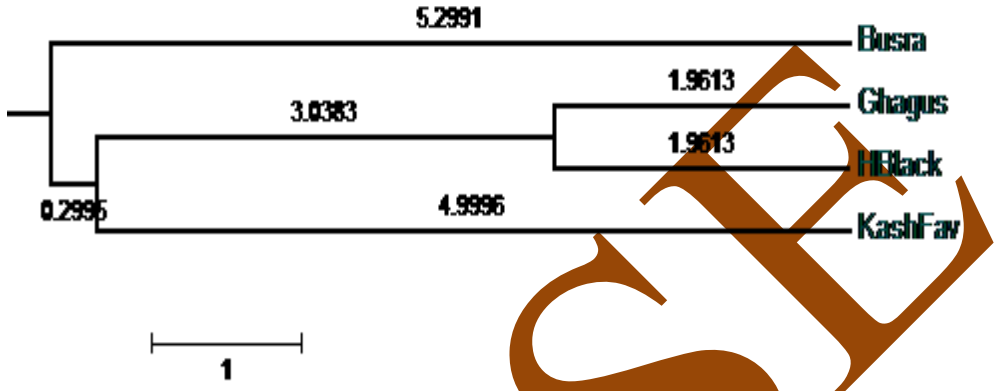


Fig30: Dendrogram construction based on Nei's genetic distance

The phylogenetic tree reveals closeness among the Ghagus and Haringhata Black populations, which are joined by Kashmir Favorolla at Node 2 while Busra is a distinct population joining at Node3.

We also carried out cluster analysis on the basis of gene frequency data on the five loci in the four populations. We also carried out the Principal coordinate analysis and the results obtained from this multivariate data analysis are presented below in fig 37 and 38.

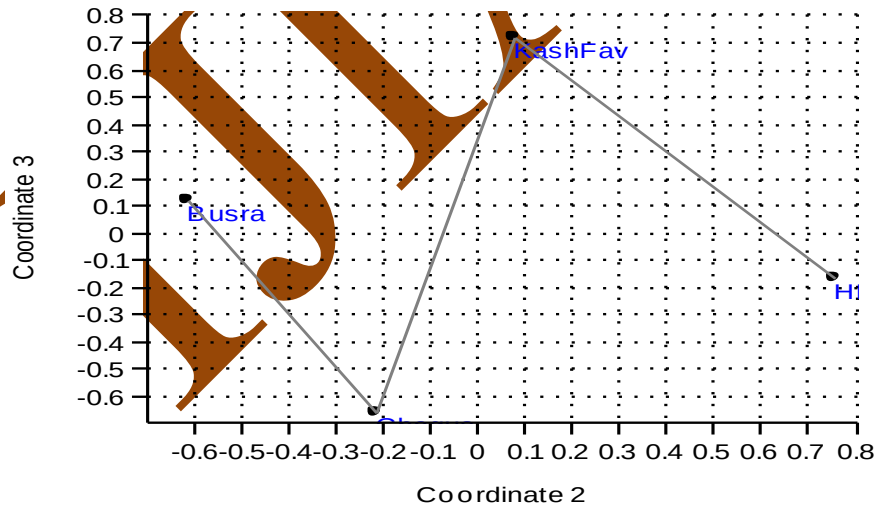


Fig 31: Principal coordinate analysis

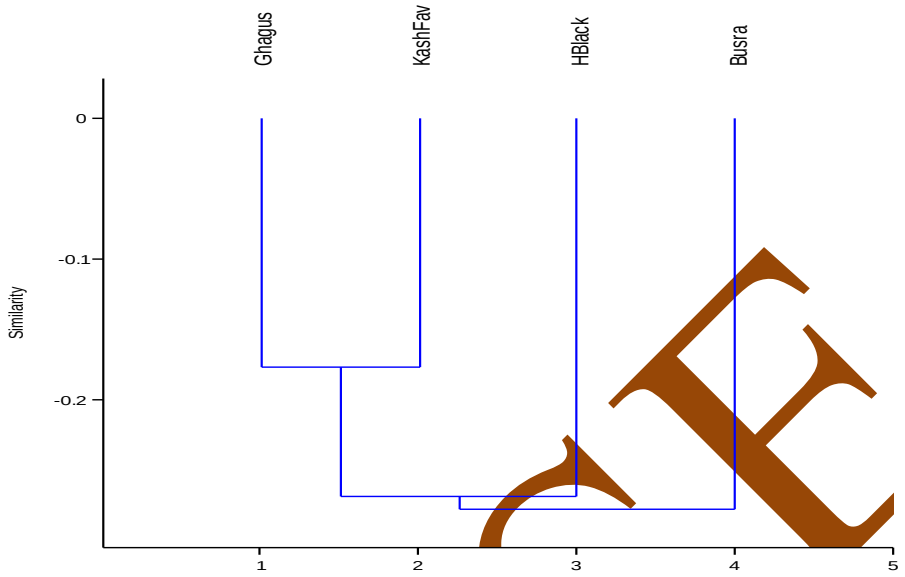


Fig32: Multivariate data analysis

The results from multivariate analysis match with the Nei's Standard genetic distances and dendrogram prepared on the basis of these. The cluster analysis gave a similar pattern, which was expected as the Nei's Standard Genetic distances are in fact the correlation of the Gene frequencies among the populations estimated from the covariance values. The diagram obtained from the Principal Component analysis (Fig 39) places the four populations as follows and has the same inferences with respect to the four populations studied.

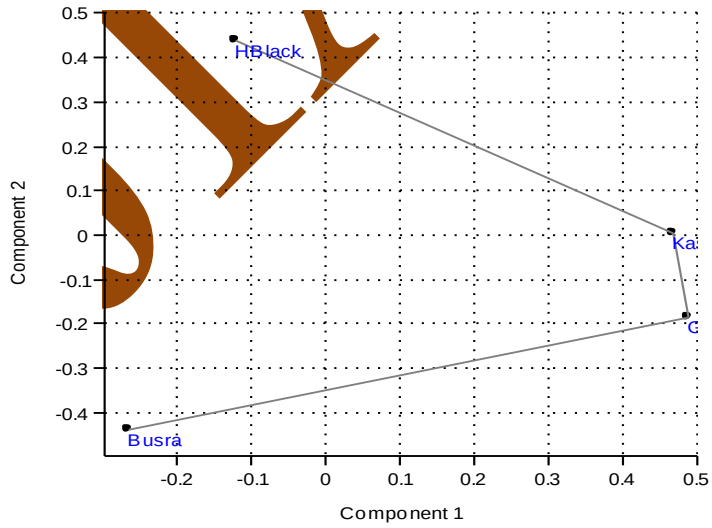


Fig33: Principal component analysis

Gene flow and F statistic analysis

The allele frequency data was used to calculate the F statistics and the population differentiation was found to be 16.45% (F_{ST} value of 0.1645). The Fis values were negative only for SNP on promoter region1 and Promoter region III (table 29). The number of migrants was estimated to be 1.2696, which is moderate, and points towards minor gene flow among the four populations.

Table28: Fis and Fst values obtained on F statistic analysis

Locus	Sample	Fis	Fit	Fst	Nm*
Promoter I	208	-0.2198	0.0725	0.2397	0.7931
Promoter II	208	0.4398	0.4523	0.0223	10.9543
Promoter III	208	-0.704	-0.3881	0.1854	1.0984
Promoter IV	206	0.3516	0.3628	0.0172	14.2497
Promoter V	208	0.7985	0.8259	0.1359	1.5895
Mean	208	-0.0264	0.1424	0.1645	1.2696

The F_{ST} values were utilized for estimating the number of migrants per generation among the four poultry populations for which the SNP data was generated. The relationship between number of migrants and F_{ST} values were as follows.

Nm = Gene flow estimated from Fst = 0.25(1-Fst)/Fst.

Among the five loci that were utilized for the purpose of F statistics it is SNP in the first region of promoter I which could differentiate among the populations to the largest extent that is 23.97%.

Other work carried out at lab:

Candidate genes PKM89 (pyruvate kinase muscle isozyme gene, exons 8, 9) and Hexa910 (hexosaminidase A alpha polypeptide gene, exons 9,10) on BTA-10 were selected for the study. Primers were designed based on Bos Taurus sequences available in database. The products were amplified, purified and was sequenced. The sequences were submitted to NCBI database and the accession numbers EF057417 & EF057418 and were obtained (Enclosed).

SUMMARY

1. The gene of interest was selected for its role in the muscle growth. The gene Myostatin is a member of TGF Beta family and is a negative regulator of muscle growth. The whole of the gene was sequenced and the 5 SNPs were selected for the study from the Promoter region. The basic aim of the study was estimation of population genetic parameters and gene frequency of the SNPs in the Indian populations of Poultry. The four populations of

poultry were selected for the study. One of the populations taken for the study was Ghagus, which is a fighting bird and kept by the tribals for game purposes and have distinctive musculature.

2. The various molecular biology tools were utilized for finding out the restriction enzymes, which can screen the SNPs for their genotyping in the four populations. Two enzymes one for each of the SNPs having restriction enzyme sites were selected for the study. In the promoter region three SNP sites did not have any restriction enzyme site and was thus genotyped using SNAP SHOT using Automated DNA sequencer AVANT 3100.
3. Heterozygosity was observed on all the five sites (SNPs) and the values were estimated for all the five loci and four populations for which the study was undertaken.
4. The data was analyzed for by using the gene frequency of each of the loci as well as nucleotide data obtained for each of the individual by incorporating the genotyped and observed SNP nucleotide.
5. The various haplotypes were formed using the various statistical softwares like ARLEQUIN VERSION 3.1 (Excoffier etal 2006) and a total of 18 haplotypes were observed in the four populations. Nine haplotypes were found in Busra and Ghagus chicken populations. The values for HBlack and Kashmir Favorolla were 11 haplotypes each.
6. The total number and their relative frequency obtained for each of the haplotypes were assessed for recombination and linkage disequilibrium studies. However when all the populations were simultaneously analysed for these populations there was no linkage disequilibrium among the SNP sites viz 1471, 1661, 1691 1737, 1748 respectively.
7. The minimum spanning tree and the relationship of each haplotype in each of the population was estimated on the basis of number of mutations and the diagrammatic representation has been given.
8. The pair wise differences among the nucleotides were estimated and the mismatch distribution analysis was carried out to find out some demographical parameters of the populations. There has been no rapid changes in the demography of the populations ie; Bottleneck or rapid expansion of the population.
9. The analysis of the data for the population being in Hardy Weinberg Equilibrium was carried out and the SNP data revealed most of the loci undertaken in the Study deviated from the HWE and pointing towards the presence of genetic structure in the four populations.
10. The Nei's genetic distance and the dendrogram prepared using these genetic distances reveal closeness between Ghagus and Haringhata Black poultry. The next population joining them is Kashmir Favorolla while Busra came out to be a distinctive population among the four studied.

11. The phylogenetic tree using the nucleotide sequence of each individual in the four populations was utilized for the construction of tree following the NJ and UPGMA algorithm.
12. The multivariate analysis was also carried out and the minimum spanning tree depicting the relationship among haplotypes, which are caused by a change in the nucleotides resulting in different haplotype.