CHAPTER 3

MATERIAL AND METHOD

MATERIALS AND METHODS

3.1 Animals

The original objective of this research work was to obtain 50 random samples from each of the populations in the study. However due to certain unavoidable difficulties, the study involved less than 50 samples in some of the populations (Table 3). While collecting blood samples, close relationship between sampled animals was carefully avoided.. A total of 600 katjang type goats were randomly selected for blood collection from Institutional herds and villages in Southeast Asia, Sri Lanka and Australia. In addtion to the random sampling, blood samples from 135 katjang type goats with pedigree data were also collected from the Goatry Farm at Musuan, Mindanao (the Philippines) and also at the Prince of Songkla University Goat farm, Hat Yai (Thailand) so that family studies could be done to establish firmly the genetic basis for the polymorphic systems. Blood samples were collected from animals of several locations.. Figure 2 shows the geographical locations from which populations were sampled.

3.2 Blood Samples

Blood samples from live animals were collected from jugular vein into heparinised 10 ml venoject tubes using 21G X 1.5 " (0.80 X 38 mm) venoject needles. The samples were immediately centrifuged at 3000 rpm for 10 -15 minutes. All samples were collected away from the laboratory. The collected blood samples were kept on ice and undisturbed until they were brought to the laboratory, quite close to the collection centres, for centrifugation.

Fig. 2 : *Geographical localities at which populations were sampled.*

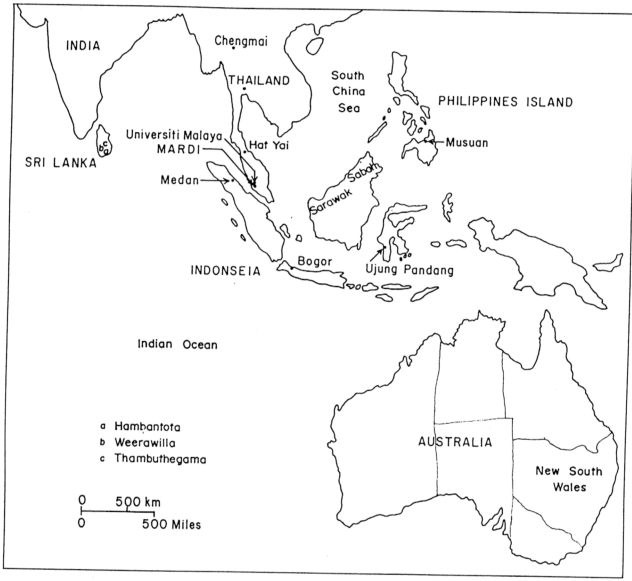


Table 3. Collection localities and Numbers of animals sampled at each of the locations.

COUNTRY	LOCALITY	AREA CODE	NO OF SAMPLES
MALAYSIA	MARDI / IPSR *	G 1	55
	Sabah	G 2	51
	Sarawak	G 3	71
INDONESIA	Bogor	G 4	50
	Ujung Pandang	G 5	48
	Medan	G 6	50
PHILIPPINES	Musuan	G 7	51
THAILAND	Chengmai	G 8	50
	Hat Yai	G 9	39
SRI LANKA	Hambantota	G 10	40
	Weerawilla	G 11	12
	Thambuthegama	G 12	31
AUSTRALIA	New South Wales**	G 13	52

* MARDI - Malaysian Agriculture Research and Development Institute

IPSR - Institute of Postgraduate Studies and Research, University of Malaya

** -Department of Agriculture and Fisheries, Wollangbar Research

The blood was separated into three components namely plasma, buffy coat and red cells. The separated samples were kept in duplicate in 2 labelled 1.8 ml cryovials. The cryovials were sealed with masking tape and marked again before dropping into liquid nitrogen (-196° C) for transportation. At the Biochemical Genetics Laboratory, Institute of Postgraduate Studies and Research (IPSR), University of Malaya, the samples were transferred to -70 °C freezer. Only required number of samples were taken out from the freezer and kept in a -20 °C freezer for analysis. Figure 3 shows the flow chart representing the various stages involved in the sample preparation for electrophoretic analysis.

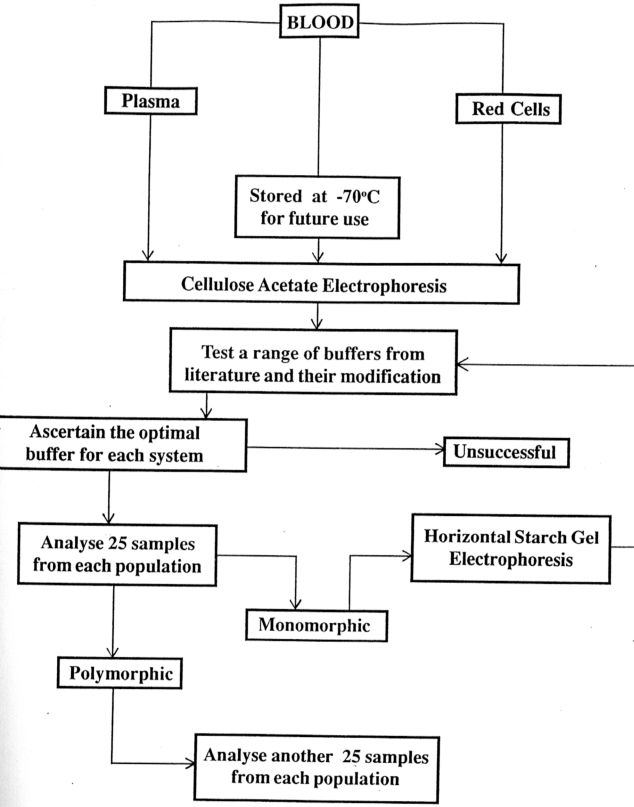
3.3 Preparation of Homogenates

For analysis, red blood cells or plasma along with their cryovials were removed from the -20 °C freezer and placed in a tray containing ice cubes to thaw. Thawing took about 15-20 minutes. After thawing, the required amount of red cells were transferred into a microwell plate and lysed with distilled water. The ratio of red blood cells to distilled water was 1:4 for analysis of haemoglobin. Plasma was diluted with distilled water to a ratio of 1:4 for albumin and transferrin and in the ratio of 1:2 for alkaline phosphatase and amylase analyses. The ratio of red blood cells to distilled water was 4:1 for all other red cell systems.

3.4 Electrophoresis

Cellulose acetate electrophoresis [CAE] was the main method employed in this study and horizontal starch gel electrophoresis [STAGE] was employed as a backup only

Figure 3. Diagrammatic representation of sample preparation for electrophoretic analysis



when CAE did not give clear resolution.. Almost all the genetic markers analysed (thirty-nine systems) used CAE for analysis and only one system used STAGE. In this study, the cellulose acetate electrophoresis set used is from Gelman Sciences Inc. Horizontal starch gel electrophoresis used here was conducted essentially as described by Kristjansson (1963). Bromophenol blue was initially used as the tracker dye so as to determine the distance moved by the solvent front during electrophoretic run.

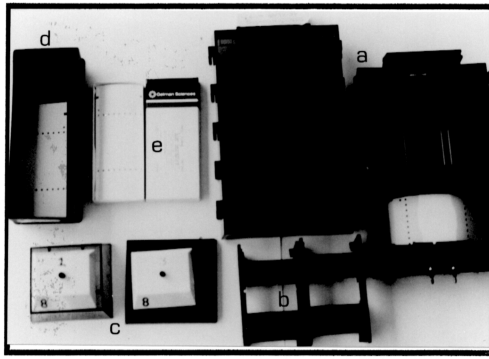
3.4.1 Cellulose Acetate Electrophoresis

Cellulose acetate gels are produced in a variety of forms by a number of different manufacturers. For the present study, the following steps were followed for the electrophoretic run, namely pre-loading and loading. The basic equipment needed consists of the electrophoresis chamber with the lid, three bridges, sepratek 8-sample applicator (Gelman Product No: 51118) and four gel storage tanks (Plate 2). The cellulose acetate strips (Product No: 51040) used were also from Gelman Sciences.

3.4.1.1 Pre-Loading

The super sepraphore cellulose acetate strips used are of size 5.7 cm X 14.4 cm, (Gelman Product No: 51040). The strip has porous surface and a plastic coated non-porous surface. Super sepraphore is a microporous, cellulosic mylar support which provides a flexible, easily handled electrophoresis medium. Handling was kept to a minimum and confined to the edges only. At the commencement of a run, the cellulose acetate strips were labelled with the date of run and enzyme / protein to be stained. The sepraphore strip was inserted slowly into the buffer, allowing it to wet completely as it

Plate 2. Gelman's Cellulose Acetate Electrophoretic Apparatus.



a: electrophoretic chamber with the lid, b: bridge
c: sepratek 8-applicator
d: gel storage tanks. e: cellulose acetate strips

enters (Plate 3). Buffer slowly advanced up the membrane as it enters into the buffer solution. Plunging the membrane into the buffer hindered complete wetting.

Care was taken in soaking the gel so as to ensure that no air bubbles were trapped on the gel strips. The gel must be completely immersed in the buffer by gently shaking the tray. The strip was soaked for at least 20 minutes to equilibrate the gel strip with the buffer.

While the gels are being soaked, the homogenates, freshly prepared from the frozen red blood cells or plasma samples are transferred from the mircowell plate onto the Sepratek-8 wells unit. The whole process was done on ice (Plate 4).

3.4.1.2 Loading

After equilibration, the gel was removed from the soaking tray, laid flat on an absorbent pad and blotted gently with a second absorbent pad. The gel was than fixed to the bridge (Plate 5).

The Separatek-8 applicator (Gelman Product No: 51118) was used to transfer the hemolysates from the Sepratek-8 wells onto the gel strips. This was done by placing the applicator on top of the Sepratek-8 wells unit. By pressing down the knob of the applicator, samples were taken up by the teeth of the applicator by capillary action. The applicator was left to take up samples for at least 30 to 60 seconds (Plate 6).

The applicator was than immediately transferred and fitted into the fixed slots on the bridge and the samples were loaded onto the porous surface of the gel strips. The slots on the bridge ensured a fixed position of the origin on the gel strips which was especially important when more than one application is required (Plate 7).

Plate 3. Soaking of Cellulose Acetate Electrophoresis gel membrane.

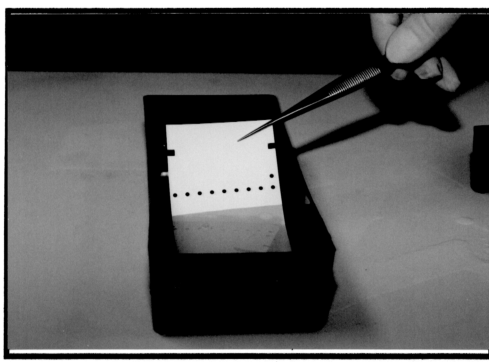


Plate 4. Preparation of homogenates into microwell plate.

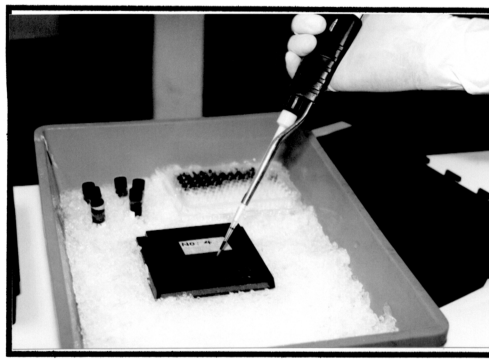


Plate 5. Fixing of Cellulose Acetate gel membranes to bridges.

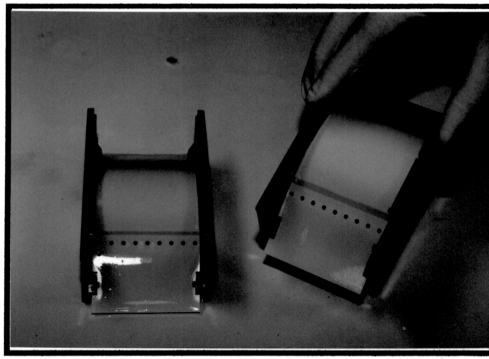


Plate 6. Transfer of samples to applicator.

Plate 7. Loading samples onto Cellulose Acetate gel membranes.

The number of applications varied from one to six, depending on the enzyme marker to be analysed. The optimum number of application for each system, which after preliminary runs is given in the appendix 1. Preliminary studies were made to ascertain the number of applications required for each marker. Each loading involved the application of 0.5 microlitre of each sample onto the porous surface of the gel strip in a straight line and the distance between the two samples being 0.2 cm.

The bridge was then positioned in the electrophoresis running chamber containing 100 ml of the cold bridge buffer in each component. The running chamber used was the Semi-Micro II electrophoresis chamber (Gelman, Product No: 51214) which comprised of two compartments. Two electrodes, the cathode and the anode, one in each compartment distributes the potential differences throughout the buffer (Plate 8).

It is ensured that both the ends of the gel strips were submerged into the buffer. The number of bridges used in each run varied from one to three bridges. In normal runs of three bridges, a total of 24 different samples can be analysed at a time in one running chamber.

Care was taken to ensure that the porous surface of the gel was not splashed with buffer after positioning . The level of the two compartments was stabilized using a spirit level. The electrophoresis chamber was then covered with the lid.

3.4.1.3 The Electrophoretic Run

The electrophoretic run was conducted in a glass doored refrigerator with the inside temperature maintained at 4 °C (Plate 9). The electrophoresis chamber was connected to the power supply by inserting the negative (black) lead into the cathodal compartment plug whilst the positive (red) lead was connected into the anodal compartment plug. The

Plate 8. Bridges with loaded gel membrane in electrophoretic chamber.

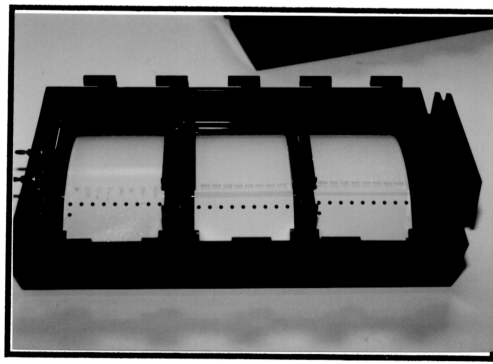


Plate 9. Glass doored refrigerator used for electrophoretic runs.

power supply was then set up to deliver the correct voltage to the system, with a check being made that the current being drawn is within the limits acceptable for that particular buffer.

3.4.2 Preparation of the gel mould for starch gel electrophoresis

A perspex frame of internal measurement of 18.5 cm x 15.0 cm x 0.6 cm was used for the enzyme analysed except that for the protein transferrin, a half frame of internal measurement of 18.5 cm x 15.0 cm x 0.3 cm was used. It was arranged on a glass plate. The whole set which would serve as a mould was than placed on a flat surface. A second glass plate was kept in the incubator (37°C) in readiness.

3.4.2.1. Preparation of starch gel

For enzymes analysis, 24 gm of hydrolysed starch was mixed with 200 ml of gel buffer in a clean 500 ml. Buchner flask. For transferrin, 13.2 gm of hydrolysed starch was mixed with 120 ml of gel buffer. The whole content of the flask was then swirled immediately to obtain an even suspension. The mouth of the flask was stoppered with a rubber bung carrying a short tube through it.

The suspension in the flask was heated in a hot water bath with constant swirling action by hand until a relatively non-viscous translucent solution was obtained. The flask was given a suction by a vacuum pump for 15-20 seconds or until frothing of the solution was minimal. The content of the flask was then poured into the mould. The second glass plate was gently lowered to cover the whole mould without trapping any air bubbles in between. A weight was then placed on the top glass plate and the gel was left to solidify for at least three hours.

3.4.2.2. Sample application

The top glass plate was removed gently from the top surface of the gel and a plastic ruler was placed perpendicular to the length of the gel about 6 cm. from one end. With the aid of a spatula with edge width of 1 cm., ten vertical slits were made on the gel along the ruler leaving 0.5 cm. between each slit. The ruler was then removed. The samples for analysis were absorbed onto pieces of 0.5 cm. x 0.9 cm. and 0.25 cm. x 0.9 cm. Whatman filter No.3 paper was used for the thick and thin gels respectively. Excess sample were blotted away. Using fine and clean forceps, these pieces of filter paper were inserted into the slits in the gel. The whole gel surface was then covered with a piece of thin polythene to prevent evaporation of the gel during electrophoresis.

3.4.2.3. Electrophoretic Run

Two perspex buffer tanks (18.0 cm. x 6.5 cm. x 6.0 cm.) were filled with bridge buffer to a level of 2 cm. below the top edges of the tanks. The two tanks were then bridged with the gel in a such a way so that the longer end of the gel from the origin is towards the anode. To obtain a complete electrical circuit, the gel was connected to the buffer by means of two 'vilene' wicks. The wicks covered about one centimetre of each of the ends of the gels whereas the rest of the wicks were dipped completely into the tanks containing the bridge buffer. Electrodes of platinum were placed into the tanks and were connected to a power unit. The whole electrophoresis set was placed in the glass doored refrigerator and the temperature maintained constantly at 4°C to avoid any denaturation of the enzymes. The electrophoretic run was conducted at 80 volts for 15 hours for the protein transferrin and for 5 hours at 200 volts for the enzymes.

3.4.3. Temperature of the Electrophoretic Run

For both starch gel (STAGE) and cellulose acetate (CAE) electrophoresis, a constant temperature of 4°C was found suitable for the run. The system is kept cool in a glass doored refrigerator with a fan circulating the air inside and maintaining the temperature (Plate 9).

Low temperature is important. Otherwise high molarity buffers can cause the cellulose acetate membrane strips to burn due to its heating effect. High voltage runs at 200 V to 350 V can also burn the strips easily if done at room temperature.

Low temperature also sustains the potential enzyme activity by preventing it from denaturing so that when stain buffer is added to the enzyme at room temperature after being removed from the refrigerator, the activity of the enzyme is triggered off and is best seen after incubating at 37°C.

3.5 Genetic Markers Investigated

The initial part of this research work was devoted to establishing the most appropriate buffer systems for each of the markers. This was established by trying out buffer systems for forty biochemical markers. These were done mainly on cellulose acetate electrophoresis. However, starch gel electrophoresis was also employed to establish the standard nomenclature for haemoglobin, transferrin, albumin and X-protein.

The names, abbreviations, enzyme commission numbers, structure and function of protein for the 40 loci studied in the present study is given (Table 4).

Table 4

NAMES, ABBREVIATION, ENZYME COMMISION NO, STRUCTURE AND FUNCTION OF PROTEINS INCLUDED IN THE PRESENT ELECTROPHORETIC STUDY

	Enzyme	locus Symbol [a]	E.C.No [b]	Quarternary [c] Structure	Supposed[d] Physiological Function
1.	Acid phosphatase	ACP	3.1.3.2	monomeric	Non-specific diphosphorylation Ca^{2+} metabolism in bone
2.	Albumin	ALB	-	dimeric	to maintain an osmotic equilibrium between the blood and tissue fluids
3.	Adenylate kinase	AK	2.7.4.3	monomeric	modulates ATP/ADP ration thus affecting metabolism generally
4.	Amylase	AMY	3.2.1.1	monomeric	converts hydrated starch and glycogen hydrolyzed to maltose
5.	Alkaline phosphatase	AP	3.1.3.1	monomeric	-
6.	Biliverdin reductase	BLVR	1.3.1.24	monomeric	biliverdin is reduced to bilirubin
7.	Carbonic anhydrase	CA	4.2.1.1	monomeric	to catalyse the reversible hydration of CO_2 to carbonic acid
8.	2,3 Diphosphoglycermutase	DPGM	2.7.5.4	dimeric	interconversion of glycerate-3-phosphate and glycerate-2-phosphate
9.	Fructokinase	FK	2.7.1.4	tetrameric	fructose converted into fructose-6- phosphate
10.	Fructose-1, 6-diphosphatase	FDP	3.1.3.11	tetrameric	links glycolysis synthesis

a - Recommended symbol

b - Enzyme Commission Number (International Unit of Biochemistry, 1975)

c - Quaternary structure as found in sheep/goat/human

d - Sources: Harris and Hopkinson (1973), McDermid *et al.*, (1975) and Di Stasio *et al.*, (1984).

(Contd. Table 4)

NAMES, ABBREVIATION, ENZYME COMMISION NO, STRUCTURE AND FUNCTION OF PROTEINS INCLUDED IN THE PRESENT ELECTROPHORETIC STUDY

	Enzyme	locus Symbol [a]	E.C.No [b]	Quarternary [c] Structure	Supposed[d] Physiological Function
11.	Fumarase	FUM	4.2.1.2	tetrameric	catalyses addition of H_2O to fumarate to form malate
12.	Glucose dehydrogenase	GLDH	1.1.1.47	monomeric	formation of NADPH in the pentose phosphate pathway
13.	Esterase-2	Est-2	3.1.1.1	monomeric	-
14.	Glucosephosphate isomerase	GPI	5.3.1.9	dimeric	glucose-6-phosphate -> fructose-6-phosphate
15.	Glutamate oxaloacetate transaminase	GOT	2.6.1.1	dimeric	soluble: cytoplasmic transamination; mitochondrial: glutamate metabolism
16.	Glutamate pyruvate transaminase (alanineamino transaminase)	GPT	2.6.1.2	dimeric	alanine converted to pyruvate and glutamate after reacting with α-beta glutarate
17.	Glycerophosphate-3-phosphate dehydrogenase	GPDH	1.1.1.8	dimeric	links glycolysis and fats synthesis
18.	Glutathione reductase	GSR	1.6.4.2	dimeric	oxidized glutathione is reduced to glutathione
19.	Haemoglobin	HB	-	monomeric	to transport oxygen to the tissues and CO_2 to the lungs
20.	Hexokinase	HK	2.7.1.1	monomeric	converts free sugars to hexose phosphate

a - Recommended symbol

b - Enzyme Commission Number (International Unit of Biochemistry, 1975)

c - Quaternary structure as found in sheep/goat/human

d - Sources: Harris and Hopkinson (1973), McDermid *et al.*, (1975) and Di Stasio *et al.*, (1984).

(Contd. Table 4)

NAMES, ABBREVIATION, ENZYME COMMISION NO, STRUCTURE AND FUNCTION OF PROTEINS INCLUDED IN THE PRESENT ELECTROPHORETIC STUDY

	Enzyme	locus Symbol [a]	E.C.No [b]	Quarternary [c] Structure	Supposed[d] Physiological Function
21.	Isocitrate dehydrogenase	IDH (ICD)	1.1.1.42	dimeric	mitochondrial: rate limiting step in TCA cycle
22.	Lactate dehydrogenase	LDH	1.1.1.27	tetrameric	control of NAD: NADH ratio
23.	Malic enzyme	ME	1.1.1.40	tetrameric	to catalyse the NADP dependent decarboxylation of malate to pyruvate
24.	Malate dehydrogenase	MDH	1.1.1.37	dimeric	soluble (cytoplasm): NAD/NADP shuttle; mithochondrial: TCA cycle gluconegenesis
25.	Mannose phosphate isomerase	MPT	5.3.1.8	monomeric	monosaccharide interconversion pathway
26.	Purine nucleoside phosphorylase	NP	2.4.2.1	trimeric	to catalyse the phosphorolytic cleavage of certain purine nucleosides
27.	NADH-Diaphorase 1 Zone 2	Dia 1-1	1.6.2.2	monomeric	to prevent the accumulation of high concentration of methaglobin in the red cell
28.	NADH-Diaphorase 1 Zone 2	Dia 1-2	1.6.2.2	monomeric	- as above-
29.	NADH-Diaphorase 2	Dia 2	-	monomeric	unknown
30.	Peptidase - A	Pep-A	3.4.11/13	dimeric	protein catabolism

a - Recommended symbol

b - Enzyme Commission Number (International Unit of Biochemistry, 1975)

c - Quaternary structure as found in sheep/goat/human

d - Sources: Harris and Hopkinson (1973), McDermid *et al.*, (1975) and Di Stasio *et al.*, (1984).

(Contd. Table 4)

NAMES, ABBREVIATION, ENZYME COMMISION NO, STRUCTURE AND FUNCTION OF PROTEINS INCLUDED IN THE PRESENT ELECTROPHORETIC STUDY

	Enzyme	locus Symbol [a]	E.C.No [b]	Quarternary [c] Structure	Supposed[d] Physiological Function
31.	Peptidase - B	Pep-B	3.4.11/133	monomeric	protein catabolism
32.	Peptidase - C	Pep-C	3.4.11/13	monomeric	protein catabolism
33.	Peptidase - D	Pep-D	3.4.11/13	dimeric	protein catabolism
34.	Phosphoglucomutase - 2	PGM-2	2.7.5.1	monomeric	interrconverrts G-1-P and G-6-P
35.	Pyruvate kinase	PK	2.7.1.40	tetrameric	phosphorylattion of pyruvate
36.	6-Phosphogluconate dehydrogenase	6PGDH	2.7.1.44	dimeric	pentose shunt
37.	Sorbitol dehydrogenase	SORDH	1.1.1.14	tetrameric	fructose mettabolism
38.	Superoxide dismutase	SOD	1.15.1.1	dimeric	removal of cytotoxix superoxide radical (O_2)
39.	Transferrin	TF	-	monomeric	to transport iron from the plasma to the bone marrow and tissue storage compartment
40	X-protein	XP	-	monomeric	unknown

a - Recommended symbol

b - Enzyme Commission Number (International Unit of Biochemistry, 1975)

c - Quaternary structure as found in sheep/goat/human

d - Sources: Harris and Hopkinson (1973), McDermid *et al.*, (1975) and Di Stasio *et al.*, (1984).

3.6 Buffers and Staining

Many different buffers are prepared for use in the analysis. The practical purpose of a buffer is to resist electrolytically induced pH changes (thus maintaining uniform electrophoretic conditions through out a run). All solutions are prepared using deionized water.

Up until this point in the procedure , all gels received similar treatments. It is in the enzyme visualisation or staining step that divergence occurred, and indeed different gels received very different, enzyme-specific treatment . The mixture of stains were prepared immediately after removing the cellulose acetate strips from the electrophoretic chamber. The strips were lightly blotted dry before pouring the mixture of stains with 2% agarose. Similar steps were also taken for the starch gels that were stained for enzymes. The stained gels were incubated at 37°C to increase the rate of enzyme reaction.

The details of electrophoretic buffer and stain mixture preparations for all the loci investigated are given in the appendix 1. The details of the reagents , substrate and stock solution used are given in the appendices 2, 3 and 4 respectively.

3.7 The use of Photo-Flo 600 solution (Kodak, Cat No.4055786)

The presence of KODAK photoflo-600; which contains ethylene glycol , in the cellulose acetate electrophoresis (CAE) buffers was found to enhance the activities and sharpen the bands formed by several goat blood enzymes on gels. This solution also overcomes the problems encountered with overheating in some buffer systems . It's presence in the gel buffer also made it easier to soak the gel strips without any entrapment of air bubbles (Selvaraj *et al.*, 1991).

3.8 Guidelines used for Electrophoretic Analysis and the Naming of Electromorphs

The following methodology was developed before the commencement of project for genetic analysis of swamp buffaloes and goats.

1. When the enzyme showed more than one region of banding and if each region is coded for by one locus, the locus whose bands appear closest to the origin anodally was called locus 1., the next more anodal region was coded for by locus 2 and so on. For cathodal migrating regions, the same system of naming was used except that the locus name was prefixed by the letter C.
2. The most common allele for any locus in the University of Malaya's goat herd situated at the Institute of Postgraduate Studies and Research Farm was named as allele "100" for all systems except for those cases mentioned in point 6. All other alleles were named with reference to allele 100. The distance of migration in mm. between the band of allele 100 and the band produced by the variant allele was the criterion used for naming the variant allele. For example , allele 102 has a band which migrates 2 mm more anodal than the band of allele 100 while allele 98 has a band which migrates 2 mm less anodally than the allele 100 band. Bands that migrate cathodally are prefixed by the negative sign. For protein that are not monomeric, the hybrid bands are ignored in naming their alleles.
3. All new systems tried out will have gels with Bromophenol Blue as the tracker dye for the buffer front and length of time needed for the run is noted.

4. A standard goat that is homozygous for the most common allele of each polymorphic system was identified in the University of Malaya's goat herd and a blood sample of this animal is present on every gel used to analyse a particular polymorphic system.
5. For all new polymorphic system discovered, a provisional scoring system, based upon a proposed genetic hypothesis is devised and used to score the gels as soon as they are run. Gels were scored while they are still fresh i.e., as soon as the bands stain up.
6. For polymorphic systems that have already been well studied by others (including pedigree data), the existing nomenclature in the literature is used. If the gel running system in the literature reports is different from that used in the present study, test runs are conducted to define comparative band patterns. One sample of all the variants observed in the present study gel system must be run on the original buffer/gel systems, so that the banding patterns obtained in the present study system can be compared with the photographs and / or diagrams found in the original paper. Examples of such systems are transferrin, albumin, haemoglobin and X-protein (Larsen *et al.*, 1992)

3.9. Family analysis

The genetic hypothesis were tested by running family samples in an attempt to obtain meaningful pedigree data that could support the proposed genetic interpretation of the electromorphs that were observed. Blood samples from many families was obtained to maximise the probability of getting families that are segregating meaningfully for each new polymorphic system. Not all families that are analysed are going to be useful for this purpose because they may not be segregating for the variant alleles. The parents of the families used in the pedigree analysis also found part of the population data for any locality provided they were not inbreds or crossbreds but their progeny were not used for this purpose.

3.10. Analysis of Electrophoretic Data

3.10.1. Calculation of Allele Frequencies

The allele frequencies were calculated by direct counting method at the loci with codominant alleles and from frequencies of the recessive homozygote for the complete dominant systems.

a) Codominant allele

The frequency of an allele is given by

$$\frac{2\,Ho + He}{2\,N}$$

where Ho = number of homozygotes for that allele

He = number of heterozygotes for that allele

N = number of individuals examined

The 'standard deviation' of the frequency of an allele(p) is estimated by

$$\frac{p\,(1-p)}{2\,N}$$

b) Simple Dominant System

$$f = \sqrt{\frac{R}{N}}$$

where R = Number of recessive homozygote.

N = Total number of individuals examined

3.10.2. Calculation of Expected Genotypes

The calculation of expected genotype are as follows:

Expected number of genotypes An An = Pn Pn X N

Expected number of genotype An Am = 2 Pn Pm X N

Expected number of genotype Am Am = Pm Pm X N

where

Pn = frequency of allele An

Pm = frequency of allele Am

N = total number of individual examined.

3.10.3. Estimation of Heterozygosity

The amount of genetic variation present in each locus was calculated by the heterozygosity index (Nei and Roychoudhury,1974a).

$$H = \frac{2N\,(1 - \sum X_i^2)}{(2\,N - 1)}$$

where

H = heterozygosity

X_I = frequency of *i*th allele at a locus

N = total number of samples

The mean heterozygosity per locus $\overline{H}$ is the sum of H over all loci (including monomorphic loci where H = 0) divided by the total number of loci studied.

3.10.4. Levene's Correction for small sample size

When sample size is small, more accurate values of the expected number of genotypes are obtained by:

$$\tilde{n}_I = \frac{nx_I\,(2n\,x_i - 1)}{2n\text{-}1} \quad , \quad \tilde{n}_{ij} = \frac{4\,n^2 x_i\,x_j}{2n - 1}$$

where x_i , x_j = gene frequency and $\tilde{n}$ = number of individuals

3.10.5. X^2 test for Hardy-Weinberg Equilibrium

If a population is in Hardy-Weinberg equilibrium, then the frequencies of genotypes will be in the ratio of p^2: 2pq and q^2, for a two allele polymorphism, where p is the frequency of allele A and q is the frequency of allele B.

For the loci with codominant alleles, expected genotypes can be calculated as:

Homozygous genotypes = f^2_1 (N) and f^2_2 (N)

Heterozygous genotypes = $2 f_1 f_2$ (N)

where f_1 , f_2 = Frequency of allele 1 and 2 respectively,

N = Total number of individuals examined.

The differences between the observed and expected values can be tested for statistical significance using a X^2 - test for goodness of fit.

$$X^2 = \sum \frac{\text{(Observed - Expected)}}{\text{Expected}}$$

3.10.6. Estimation of F-Statistics

Wahlund (1928) showed that the gene frequency differences among subdivided populations cause a deficiency of heterozygotes compared with the case of single random mating population. Wright (1943, 1951, 1965, 1969) presented subdivided populations in term of three parameters F_{IS} , F_{IT} and F_{ST} which are often called *F-statistics* or *allelic correlations*. F_{IT} and F_{IS} are often called fixation indices (F_I).

The three parameters are related by the following formula:

$$\mathbf{1-}F_{IT} = \mathbf{(1-}F_{IS}\mathbf{)\ (1-}F_{ST}\mathbf{)}$$

F_{IS} = Fixation index of individuals relative to their locality populations, or the correlation between genes within individuals within locality populations (inbreeding within each locality).

F_{IT} = Fixation index of individuals relative to the total set of populations, or the correlation between genes within individuals

F_{ST} = Correlation of genes between individuals within the same locality populations , which is interpreted as a measure of the amount of differentiation among locality populations .

3.10.7. Calculation of Genetic Variability

The measure of genetic variability based on allozyme data are the mean number of alleles per locus and the percentage of polymorphic loci (*p*). A locus is defined as polymorphic if the frequency of the most common allele is less than 0.99 or 0.95. The defination is arbitrary if sample size or the number of loci examined are small (Nei,1978). Estimates of genetic variability based on frequency values of polymorphic enzyme in natural population ranges from 0.0 to 0.8.

The *p* and *h* were calculated by using the following formulae:

1. Percentage of polymorphic loci (*p*)

$$p = \frac{\text{number of polymorphic loci}}{\text{total number of loci analysed}}$$

2. When the population is in Hardy-Weinberg equilibrium and X_i is the frequency of the i^{th} allele at the locus, the heterozygosity (*h*) is calculated as :

$$h = 1 - \sum_{i=1}^{m} x_I^2$$

3.10.8. Genetic Identity and Genetic Distance

The concept of genetic distance was first used by Sanghvi (1953) for an evolutionary study. There are many ways of expressing the genetic distance between populations but despite the controversies which have arisen over the best genetic distance, many authors have shown that comparison of different methods generally leads to very similar results (Gregorius ,1974; Chakraborty & Tateno ,1976; Aupetit ,1985). For the present study, the genetic distances model suggested by Nei (1972) has been used.

3.10.8.1. Genetic Identity

Nei's coefficient of genetic identity (I) between two populations is given by:

$$I = \frac{\Sigma\, xi\ yi}{(\Sigma\, x_i^{\,2}\, \Sigma\, y_i^{\,2})}$$

where

I = Nei's coefficient

x_i = frequency of the ith allele in population X.

y_i = frequency of the ith allele in population Y.

I = 1,when X and Y are monomorphic for the same allele, and

I = 0, when X and Y are monomorphic for different alleles.

The mean genetic identity ($\overline{I}$) is the mean over all loci studied (including monomorphic ones) and is calculated as:

$$\overline{I} = \frac{I_{XY}}{I_X\, I_Y}$$

where I_{XY}, I_X and I_Y are the arithematic means, over all the loci of the $\Sigma x_i y_i$, Σx_i^2 and Σy_i^2 respectively.

3.10.8.2. Standard Genetic Distance

The genetic distance calculated from the normalised identity of genes (Nei,1972) was used; that is the distance between two populations is estimated by:

$$\mathbf{D = -ln\, I}$$

3.10.9. Statistical Analysis Using the BIOSYS-1

The contingency X^2 tests, the heterozygosity indices, the proportion of polymorphic loci, the calculation of allele frequencies and expected genotypes, the tests for Hardy-Weinberg equilibrium and the determination of the genetic similarities were carried out using the BIOSYS-1 programme release 1.7 (A computer programme for the analyses of allelic variation in population genetics and biochemical systematics, Swafford and Selander,1989). The results were later verified manually using the method proposed by Nei (1972).

The genotype frequencies data input prepared for the Biosys-1 analysis for all polymorphic loci except alkaline phosphatase and nucleoside phosphorylase is shown in appendix 5. The allelic frequencies data input for all loci studied is given in appendix 6.