3.1 Sampling area

Tamil Nadu situated on the eastern side of the southern tip of the Indian peninsula is the 11th largest state and comprises of 4% of the country’s total area. Tamil Nadu has a very ancient history, which goes back more than 6000 years. The state represents the nucleus of Dravidian culture in India and along with the other south Indian states forms the repositories of Dravidian culture today. To the north of Tamil Nadu lie the Nilgiri hills, and to the west lie the Western Ghats and in the south, the Anaimalai hills. The Nilgiris, queen of hills literally meaning Blue Mountains, as they are called because of the blue haze surrounding them, is the smallest district of Tamil Nadu state in south India. These shimmering hills nestle at a point where the rugged beauty of the Eastern Ghats merges with the wooded verdancy of the Western Ghats (Mallah 1993). It covers an area of 2,549 sq.km, located at 11°10’ – 11°42’ north latitude and 76°14’ – 77°10’ east longitude and is situated at the trijunction between the Malabar region of Kerala state on the west, Karnataka state on the north and Tamil Nadu on the east and south. The topography of the Nilgiris consists of undulating green hills with an average altitude of 1,980 MSL, with the highest peak being the Doddabetta range 2,632 MSL. The climate is temperate with an average annual rainfall of 45” (Ghosh et al. 1977). The Anaimalai hills or the Elephant hills is a range of mountains in southern India, in the Coimbatore district of Tamil Nadu lying between 10° 13’ – 10° 31’ north latitude and between 76° 52’ – 77° 23’ east longitude forming a portion of the Western Ghats. This range is broken by the Palghat Pass, south of the Nilgiris. The heavily forested slopes of this range are the sources of the Kaveri, Vaigai and the Thamaraabarani rivers. The grasslands are thoroughly sheltered with bamboo forests. The Nilgiris and Anaimalai hills (Western Ghats) with its rich diversity of species are designated as one among the eighteen hot spots of the global mega diversities (Cherian 2000). The study populations (Fig 3.1) have attracted much attention because of certain unusual characteristics that has survived from pre-colonial times to the present day.
Fig 3.1: Map showing the sampling locations of the three tribal populations.
3.2 Study populations

3.2.1 Kadar

Kadar are a forest dwelling hill tribe inhabiting the Anaimalai ranges in Coimbatore district in Tamil Nadu and the great mountain range which extends southward into Travancore and Kerala. They are referred by the other tribes as ‘kattu al’ meaning people of the forest. Ananthakrishnan (1909) has reported two subdivisions, namely the Kadar and Malayans among them. The Kadar language is close to the Dravidian language Tamil and is locally known as Kadar bhasha which has no script. With others they converse in Malayalam and Tamil. The script used for writing is Malayalam. They are short-statured and long-headed people with a round or oval facial profile and a broad nose form suggesting Proto-Australoid strains (Ray et al. 1960) (Plate 3.1a). Extreme types have some Negrito character – frizzy hair instead of straight or wavy hair and especially dark skin – but it has been suggested that some observed examples of frizzy hair are due to rare admixture with Africans (Saha et al. 1974). They are nomadic; they do not farm but do not like to live as foragers. They work as laborers or as specialized collectors of commercial tropical plants. They are non-vegetarians and their staple food is rice. They consume different types of pulses, roots and tubers.

The Kadar observe community endogamy. Among them marriage between cross cousins is allowed. Widows, widowers and divorcees are free to remarry. Earlier the Kadar were hunters and gatherers. Nowadays, many of them have become settled cultivators, agricultural laborers, basket-makers, mat-weavers and shop owners. Their highly developed olfactory nerves, knowledge of the forests and skill at tree climbing make them very successful trackers, hunters, and honey and hill produce collectors (Plate 3.1b). Pathies (huts) are on slightly raised mud or stone foundations and covered on all sides with flattened bamboo and grass. Their primitive huts are very temporary structure and so low that one has to bend to enter (Plate 3.1c). The march of culture into their society has brought the desire for better homes.

A Kadar headman is known as Moopan. The institution of Moopan is hereditary. They worship Vanandevathai and she is worshipped anywhere as an invisible Goddess.
Plate 3.1: Kadar

a. Kadar Family

b. Hunter

c. Hut
They also worship Athuvacheri Amman. Once in a year, during the month of chittirai (April-May) a festival is held to honor their deity. Kadars worship their ancestors on a particular day during the Tamil month of Adi (July-August). They are also worshippers of Lord Ayyappa. Kadar folk-songs are sung and folk-dances are performed by both the men and women. They play percussion and wind instruments. They bury the dead in far-away graves. Burial is conducted in a solemn manner as expeditiously as possible after death. (Thurston 1909; Saha et al. 1974; Singh 1994).

3.2.2 Paniyan

The term Paniyan may have been derived from the word “hani” meaning work, and the term means the worker. They are mainly distributed in the hilly and forested tracks of the Wynad district of Kerala and also in the Nilgiri district of Tamil Nadu. They are dark skinned, long headed people of short stature with wavy or curly hair and show a broad nose form (Plate 3.2a, b). Some are of the opinion that they are an African tribe, who had come to India after a shipwreck on the west coast. At one time in the past, they were the principal stock-in-trade for slave trade on the west coast and it is possible that they were imported from Africa and sold in Malabar.

The Paniyans have exogamous lineages such as koyinouten, muutettan, naattilkapadam, padikan, anjili, parrier, pakkatti, vallatt, maniyankoden, pappali. Their traditional costumes are attractive. Women of older generation wear ear rings, nose rings, colored bangles and rolled palm leaves with beads inside from a plant called “Abrus precatorius” in their dilated ear lobes (Plate 3.2c, d). They prefer to marry after attaining adulthood. Marriage symbols are badges (thali), red waistbands and nose rings. Monogamy is the general rule among them. Widow re-marriage is allowed. The dead are buried and a period of death pollution is observed. Ancestor worship is a part of their death rites.

The Paniyan settlements are called paddies. Their language is primitive dialect of Malayalam with an admixture of Tamil and Tulu. They normally have a headman called
Plate 3.2: Paniyan

a. Paniya Family

b. Paniya Men

c. Paniya Woman

d. Paniya Woman
While their chief occupation is agricultural labor, they are allegedly subjected to "a subtle form of bonded labor by some of the local non-tribal landlords". They follow the religious cult of animism and worship Goddess "Kattu Bhagavathy". They also worship trees, banyan tree (*Ficus bengalensis*) is important for them. They will not cut such trees and believe if anyone attempts, for that, they face serious problems or fall sick. A few Paniyans have reportedly embraced Christianity. They do not share water sources and crematoria with others, but they visit the same religious shrines and participate in traditional festivals with their neighbors. In economic life, they maintain symbiotic relationships with the neighborhood (Thurston 1909; Breeks 1983; Singh 1994).

### 3.2.3 Todas

The Todas and the Nilgiris, the tribe and the land respectively are known for their exquisite qualities (Plate 3.3a). Their unique barrel vaulted huts and temple dairies, physical features (Caucasoid) and customs seen nowhere else on the subcontinent, gives rise to speculations of their origin. Many theories of their racial origin have been advanced; their peculiar features and the dress they wear gives an impression of Roman descent or a lost tribe of Israel or have a Sumerian links with ancient Mesopotamia or the refugees from the court of Ceylon. However the generally accepted theory states that they are a Dravidian race of Scythian origin, who were driven from the plains by Aryan invasions. They are also considered as proto-aryans who share a series of similarities with the Ainus of Japan. But with the prevailing antiquity and rich folklore, the Toda themselves believe that the Goddess Teikirsky, who with a tap of her cane, created out of a sacred pool called Nirykair, the first Toda and his buffalo. The influence of this legend on the Toda life has a profound effect on their entire social and religious life, which revolves around their sacred buffaloes and it is only with much reluctance that a Toda can be forced out of, what he believes, is his birthplace in the blue mountain (the Nilgiris).

The name Toda, possibly was derived from the word ‘**TUD**’ – the sacred Teidr tree (*Maliosma simplicifolia*) of the Todas, which is also called as ‘**tode**’ by the neighboring Badagas. Contradictorily this word ‘Toda’ itself has an alien origin – they call themselves
'Awl' or simply the people. The upper Nilgiris are dotted with curious megalithic circles of pilled stones, earthen circles, cists and dolmens. Of these, the most fascinating are the stone circles that are generally found on the summits of prominent hills. It was at once apparent that these stone circles were used as ancient funerary sites. The one still existing tribe that could possibly have done so, could be the Todas. It appears that the Todas have dwelt on the upper Nilgiris for a thousand or more years and in all likelihood they were the stone circle builders. With these romanticized theories and their early history, the Todas have been regarded by other hill tribes as 'Lords of the Soil'.

The Todas number about 1400 according to the 1991 census of India. They are the smallest among the native dwellers of Nilgiris. The 'Toda' dialect is an independent language of the Dravidian family affiliated to Tamil/Malayalam. The Toda community consists of two endogamous divisions – the Taerthar Awl of ten clans and the Teovily Awl of five clans. They live in a settlement called 'Mund' and around 48 munds are spread in the upper Nilgiris with a maximum of 4 to 5 families. Their houses are Igloo-like oval-pent-shaped huts made of bamboo and dried grass fastened together with rattan and well thatched. The entrance is small and access is possible only by crawling, and it is so to protect from the wild animals and also to provide warmth (Plate 3.3b).

The Todas are most dignified in their bearing. They are well-built, tall and set up with finely proportioned muscular limbs, almost European contour with decidedly Roman noses, bright hazel eyes and an abundance of rather coarse but glossy black hair. They wear the traditional – 'Puthkuli', a unisex garment of coarse, white cloth embroidered in red, blue and black with 'Roman-Toga' style (Plate 3.3c). The expression of their countenance is open, fearless and agreeable. All these Caucasoid features inspired a lot of researchers to explore their origin.

The Todas are purely a pastoral tribe (Plate 3.3d), but have now taken up agriculture in a big way. However, caring for the buffaloes and breeding them continues, in their own imitable way and the wealth is measured by the buffaloes they own. They are traditionally lacto vegetarians and in general love liquor. The afterworld myth inspires them and buffaloes are sacrificed at funerals so that the dead person may continue to live in
Plate 3.3 : Toda

a. Toda Unique Tradition

b. Barrel-Vaulted hut

c. Toda Women

d. Toda & his Buffalo

e. Conical Temple

f. Barrel-Vaulted Temple
comfort in ‘Amunawdr’ (after world). The temples are barrel-vaulted and all major hamlets have such temples. There are also two very sacred temples of a unique conical structure (Plate 3.3e,f). They use natural materials like kwehtf or paarsh (Sideroxylon sp); tree poles; specified wooden planks; rattan cane (Calamus pseudotenuis); theff or bamboo reeds (Pseudoxytenanthera monadelpha) a swamp grass called awful and specified wood for the door and carved Kweghaishveil a totem atop the thatch. The sacred bells, earthen vessels, churning sticks are the main objects of worship. Besides their simple life style, the marriage system also attracts a lot. The man weds a woman on the new moon day of her seventh month of pregnancy in a quaint ceremony called ‘Purshutt’ by presenting a miniature of bow and arrow.

The Todas with unique appearance and cultural entity have attracted the anthropologists and demographers. They state that the population has a slow growth and at times even negative growth and the fact remains that they remained near static for a long time. By and large, the Toda are good-natured clan and beginning to catch up with the lure of modern world. But they are fiercely independent people, who are always proud of their heritage of history and culture (Marshall 1873; Thurston 1896, 1909; Walter 1986; Ananthakrishnan 1930; Emeneau 1937, 1974; Rivers 1951; Breeks 1963; Singh 1994).

3.3 Blood collection

A total of 102 autochthonous individuals with prior informed consent protocols were analyzed. They were interviewed for their personal details like name, age, sex, place and birth and if any known possible details of their ancestry and their medical histories were also recorded. The obtained information was kept confidential and preserved in our database. The samples comprised of purely unrelated healthy blood donors (males) consisting of Kadar (40 individuals), Paniyan (30 individuals) and Toda (32 individuals) tribes. From each individual 10 ml of whole blood was obtained by venipuncture in EDTA-containing evacuated sterile tubes and stored at 4° C in ice-cold container and transported to the genetics laboratory of Bharathiar University for the isolation of DNA.
3.4 DNA isolation

Total genomic DNA was extracted using standard and simple salting out procedure (Miller et al. 1988). The peripheral blood mononuclear cells were separated from the bulk erythrocytes by centrifugation (6000 g for 30 min) of anticoagulant treated non-frozen blood. The erythrocytes form a pellet, with the leukocytes (buffy coat cells) forming a layer on top. This white buffy coat was carefully removed and washed twice with lysis I (RBC buffer) until a clear white pellet was seen. To the pellet 3 ml of lysis II (WBC buffer) was added and aspirated well. Treatment of cells with 10% sodium dodecyl sulphate (SDS) and 500μl proteinase K at 37°C (overnight) helps to digest the residues of proteins and aids in the complete extraction of DNA. Addition of 1 ml of 6M NaCl lyses the cells and nuclei and liberates the DNA, which is tightly bound in chromatin. Following the extraction steps, high molecular weight DNA molecules are precipitated with cold ethanol and collected into a sterile micro centrifuge tube (Plate 3.4a). It was then washed in 70% ice cold ethanol and dried in speed vac. 0.5ml-1ml TE buffer (pH 8.0) was added to the DNA and allowed to dissolve for 24h before use. DNA quantity was assessed after agarose gel electrophoresis and visualized under UV light after ethidium bromide staining. Aliquots of DNA samples were kept at -20°C and the rest was frozen at -70°C. List of buffers and solutions used for DNA isolation are given in Table 3.1.

Table 3.1: Buffers and solutions used for isolation of Genomic DNA

<table>
<thead>
<tr>
<th>Buffer / Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysis I (RBC buffer)</td>
<td>7.46g NH₄Cl, 2.06g Tris-HCl in 1000ml water</td>
</tr>
<tr>
<td>Lysis II (WBC buffer)</td>
<td>1M Tris-HCl (pH-8), 0.5M EDTA (pH-8), 1M NaCl</td>
</tr>
<tr>
<td>Tris-EDTA (TE) buffer</td>
<td>0.01 M Tris-HCl (pH 7.4), 0.001 M Na₂-EDTA (pH 8)</td>
</tr>
<tr>
<td>Proteinase K solution</td>
<td>100mg Proteinase K, 5ml of 10%SDS, 0.2ml of 0.5M EDTA (pH-8)</td>
</tr>
<tr>
<td>6M NaCl</td>
<td>35g NaCl was dissolved in 100ml water</td>
</tr>
</tbody>
</table>
3.5 Genetic systems studied

3.5.1 mtDNA polymorphisms

Most often, population analyses have examined allele frequencies at autosomal (biparental) genetic markers. The incorporation of mtDNA during the 1980's added a powerful tool to the geneticists' tool kit, due to its unusual properties which include a strictly maternal mode of inheritance, high copy number and rapid rate of evolution making it particularly well-suited for use in evolutionary studies (Stoneking and Södyall 1996).

Genetic variation at the mtDNA has been studied with different techniques mainly through high-resolution RFLP analysis, which allows to define broad, continent-specific groups of sequences or haplogroups. A different approach consists of sequencing the hypervariable fragments of the control region. It has been shown that most haplogroups can be recognized by specific nucleotide motifs in the control region. Thus mtDNA "haplogroups" are defined by coding-region RFLPs, "haplotypes" are defined by hypervariable segment sequences, and mtDNAs defined by both RFLPs and hypervariable sequences are referred to as "lineages" (Gresham et al. 2001). The intergenic COII/trRNA_Lys region of human mtDNA usually contains two tandemly arranged copies of a 9-bp sequence. Length variation in this region was loss of one copy of the 9-bp repeat sequence CCCCCCTCTA. This deletion has been found at varying frequencies and is commonly referred to as an "Asian-specific" marker (Wrischnik et al. 1987).

The control region is an 1122 bp segment of non coding DNA that is most rapidly evolving and polymorphic region of the human mtDNA genome. Two hypervariable regions, HVS1 and HVS2, bracket a highly conserved 200-bp G-rich region and contain more than 90% of the polymorphisms reported in the mtDNA molecule (Vigilant et al. 1989). 60% of the polymorphisms in the control region are found in HVS1 and 40% in HVS2. As a consequence, analysis of HVS1 has been the emphasis of many recent investigations of human population history and structure. It affords the maximum resolution for distinguishing among even very closely related mtDNAs. Sequence information from the mtDNA control region (CR) which documents variation within and
between human populations, has been widely used to infer certain aspects of human population and demographic history (Cann et al. 1987; Di Rienzo and Wilson 1991; Stoneking et al. 1992; Horai et al. 1995; Richards et al. 1996; Thangaraj et al. 1999; Roychoudhury et al. 2001; Brakez et al. 2001).

In the present study each DNA sample was screened for ten mtDNA RSP's and one insertion/deletion polymorphism (IDP). They were also subjected to sequencing of the first hypervariable region (HVS1) of the mitochondrial D-loop. The loci screened along with the nucleotide change are given in table 3.2. The sites were chosen such that individuals could be classified into haplogroups that are most relevant for Indian populations. These markers have been used extensively in recent years to study the origin and dispersal of contemporary human populations (Torroni et al. 1993a, b, 1996; Roychoudhury et al. 2000, 2001; Deepa et al. 2002).

Table 3.2: Polymorphic loci examined and their locations

<table>
<thead>
<tr>
<th>Locus</th>
<th>Polymorphism</th>
<th>Polymorphic Site (np)</th>
<th>Nucleotide Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaeIII 663</td>
<td>RSP</td>
<td>663</td>
<td>A→G</td>
</tr>
<tr>
<td>HpaI 3592</td>
<td>RSP</td>
<td>3594</td>
<td>C→T</td>
</tr>
<tr>
<td>AluI 5176</td>
<td>RSP</td>
<td>5178</td>
<td>C→A</td>
</tr>
<tr>
<td>AluI 7025</td>
<td>RSP</td>
<td>7028</td>
<td>C→T</td>
</tr>
<tr>
<td>DdeI 10394</td>
<td>RSP</td>
<td>10398</td>
<td>A→G</td>
</tr>
<tr>
<td>AluI 10397</td>
<td>RSP</td>
<td>10398, 10400</td>
<td>A→G</td>
</tr>
<tr>
<td>HinfI 12308</td>
<td>RSP</td>
<td>12308</td>
<td>A→G</td>
</tr>
<tr>
<td>HincII 13259</td>
<td>RSP</td>
<td>13263</td>
<td>A→G</td>
</tr>
<tr>
<td>AluI 13262</td>
<td>RSP</td>
<td>13263</td>
<td>A→G</td>
</tr>
<tr>
<td>HaeIII 16517</td>
<td>RSP</td>
<td>16519</td>
<td>T→C</td>
</tr>
<tr>
<td>9-bp Del</td>
<td>Deletion</td>
<td>8272 - 8289</td>
<td>CCCCCTCTTA</td>
</tr>
<tr>
<td>HVS1</td>
<td>Sequencing</td>
<td>16024 -16380</td>
<td>Hypervariable region</td>
</tr>
</tbody>
</table>
3.5.2 Y-Chromosome polymorphisms

Y-Chromosome specific markers are of considerable importance for studies of microevolution, especially those of historical migration patterns and sex-specific gene flow. They are also valuable in measuring population affinities, genetic admixture and Y chromosome evolution. The non-recombining portions of the human Y chromosome have special features of a haploid and a father-to-son transmission pattern. The DNA sequence of these portions therefore, contain a genetic record of the mutational events that occurred in their past. Haplotypes constructed with Y chromosome alleles at multiple polymorphic loci can give us valuable information for inferring the paternal lineages and population history in humans.

The variety of polymorphic markers now available on the non-recombining portion of the Y chromosome ranges from base substitution and insertion/deletion polymorphisms, which are rare (probably even unique) events in evolution and which tend to be biallelic, to faster-mutating polymorphisms such as microsatellites - also known as short tandem repeats (STRs) – and the minisatellites (MSY1) (Bouzekri et al. 1998; Jobling et al. 1998). Microsatellites are tandemly repeated arrays of two to six bp units and it occurs as frequently as every 30,000 bp in the human genome. They show high mutation rates ranging from $10^{-3}$ (Jeffreys et al. 1988) to $10^{-4}$ (Levinson and Gutman 1987). The use of Y STRs allows the simple construction of highly variable haplotypes. With these haplotypes, it is possible to analyze differences in population structure by a comparison of haplotype diversity and of the number of population-specific haplotypes. The use of Y STRs also allows the simultaneous analysis of closely related and distantly related populations. Because of their relatively high mutation rate, Y STRs are polymorphic in potentially all human populations and allow human migration processes to be traced on historical timescale. Single nucleotide polymorphisms (SNPs), defined by the co-occurrence of at least two different bases at the same location, each with a population frequency >1% (Vogel and Motulsky 1979), represents the most common form of sequence variation in the human genome. It is estimated that, on average, one in 1000 nucleotide positions is polymorphic in humans (Wang et al. 1998) implying that $10^6 - 10^7$
SNPs can be expected to exist in an average population. Y chromosome SNPs have a lower mutation rate and hence are ideal for the study of human migration at an evolutionary rather than a historical timescale. The dual approach of using Y STRs as well as Y SNPs renders the maximum amount of information (de Knijff et al. 1997, 2000).

The microsatellites screened in the present study are DYS19, DYS388, DYS390, DYS391, DYS393, DYS389I, DYS389II, DYS425 and DYS426. DYS19, DYS391, DYS389II, DYS390, DYS391 and DYS393 are tetranucleotide repeats, while DYS388, DYS425 and DYS426 are trinucleotide repeats. The SNPs typed are 92r7, Tat, sY81, SRY465, SRY4064, M9, M13, M17, M20, and SRY 10831. One insertion polymorphism YAP was also typed.

3.6 DNA analysis

3.6.1 RSP's / Insertion-deletion (InDel) polymorphisms

DNA samples were subjected to PCR amplifications. All the polymorphic loci studied were genotyped by amplifying 50-100 ng of DNA in a standard 30-cycle three step PCR (Perkin Elmer 2400 thermal cycler, Plate 3.4b). Appropriate annealing temperatures and additives were optimized for each system. PCR products of the restriction site polymorphisms were digested with the appropriate restriction enzymes (Gibco BRL/Promega) in the respective buffers for 2-4 hours at the required and suitable temperatures (Detailed locus wise protocols presented in Sec 3.7.1). After PCR, the samples were mixed with 1/5th volume of 5x gel loading dye and subjected to electrophoresis (Horizon 11.14, GIBCO-BRL) at 120V for 45 minutes. For InDel polymorphisms the samples were electrophoresed in 6% polyacrylamide gel at 700V for two hours (GIBCO-BRL gel apparatus 18cm x 40cm). Ethidium bromide stained gels were visualized under UV trans-illuminator (Bio-Rad) and were documented. Band sizes were compared to λ DNA molecular weight (φ 174 HaeIII) standards, and with previously defined patterns. Allelic status was assigned by two independent observers.
who had no disagreement in allelic assignment. List of all buffers and solutions for electrophoresis are given in Table 3.3.

### Table 3.3: Buffers and solutions used for electrophoresis

<table>
<thead>
<tr>
<th>Buffer / Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x TBE buffer</td>
<td>0.089 M Tris-borate, 0.089 M boric acid, 0.002 M Na₂-EDTA</td>
</tr>
<tr>
<td>5x Gel loading buffer</td>
<td>1.6% bromophenolblue, 1.6% xylene cyanol, 1.6% SDS, 3 mM EDTA (pH-8) in 10% glycerol</td>
</tr>
<tr>
<td>Agarose gel electrophoresis buffer</td>
<td>2% agarose in 1x TBE buffer</td>
</tr>
<tr>
<td>Agarose gel running buffer</td>
<td>1x TBE buffer</td>
</tr>
<tr>
<td>40% polyacrylamide stock for PAGE gel</td>
<td>38% acrylamide and 2% bis-acrylamide</td>
</tr>
<tr>
<td>Polyacrylamide gel electrophoresis buffer</td>
<td>6% polyacrylamide in 0.5 x TBE buffer</td>
</tr>
<tr>
<td>PAGE running buffer</td>
<td>0.5 x TBE buffer</td>
</tr>
<tr>
<td>10% Ammonium persulphate (APS)</td>
<td>100mg of APS was dissolved in 1ml of water and stored at 4°C in dark</td>
</tr>
<tr>
<td>10% Sodium dodecyl sulphate (SDS)</td>
<td>2.5g SDS was dissolved in 20ml of water by stirring at low heat and the volume was made up to 25ml.</td>
</tr>
<tr>
<td>Ethidum-Bromide (10mg/ml)</td>
<td>100mg ethidum-bromide was dissolved in 10ml water by stirring overnight. The solution was filtered and stored at 4°C in dark</td>
</tr>
</tbody>
</table>
3.6.2 mtDNA sequencing

The HVS1 region (nt 16033 – nt 16370) of the mitochondrial D-loop was amplified using standard primers (Vigilant et al. 1991). Standard hot-start PCR reactions were carried out with 50-100ng of DNA in 20μl reaction volume. Sequencing of the PCR product was carried out using Big Dye terminator ready reaction kit (Perkin Elmer) and reconstituted primer (3.2 pm). The extended products were purified by adding 3M sodium acetate and 95% alcohol, centrifuged at 13,000 rpm and the supernatant discarded. The pellets were washed with 70% alcohol twice and air-dried. The pellets were again dissolved in 15μl Hi-Di formamide (Applied Biosystems), heat denatured for five minutes and chilled immediately (Detailed protocol given in Sec 3.7.2). The products were run on an ABI-3100 capillary based genetic analyzer (Plate 3.4c) using POP-6 polymer (supplied Applied Biosystems). Sequencing analysis and auto-assembler software were used for further analysis. In the HVS1 region, the nucleotide position from 16184 to 16193 is CCCCCCTCCCC (Anderson et al. 1981). In some individuals the ‘T’ at np 16189 is polymorphic and is changed to ‘C’ creating a stretch of 10 C’s. In those samples the sequencing reaction could not proceed after 10 C’s thus giving a truncated sequence. Therefore, to obtain sequences after the ‘C’ stretches the sequencing reaction was performed using reverse primer. Full length sequences were then generated by joining the two truncated sequences.

3.6.3 Y- chromosome analysis

The multiplex PCR “kits” that permit high throughput analysis of Y chromosome polymorphisms contain oligonucleotide primers labeled with three different ABI dyes (HEX, TET and FAM). Microsatellite kit 1 (MS1) contains five primer pairs that amplify DYS19, DYS388, DYS390, DYS391 and DYS393 (Table 3.4). Primer for loci DYS391 was redesigned to ensure that all PCR products from this kit fall within a limited size range (100-230bp) and that the five microsatellites can be discriminated using a combination of size and fluorescent dye label. Microsatellite kit 2 (MS2) contains primer pairs that amplify loci DYS388, DYS389I, DYS389II, DYS425 and DYS426 (Table 3.5). The
DYS388 locus is amplified by both MS1 and MS2 kits providing a degree of internal control (Detailed protocol given in sec 3.7.3).

Multiplex PCR kits to type biallelic polymorphisms (unique event polymorphism) at 11 separate loci consist of two kits namely UEP1 kit and UEP2 kit. UEP1 kit contains primers to amplify loci 92r7, sY81, SRY+465, SRY4064, Tat and YAP (Table 3.6). All the makers except YAP are single nucleotide substitution. The YAP marker is an Alu insertion polymorphism. An insertion deletion polymorphism p12f2 was also typed. The UEP2 kit contains primers, which amplify regions containing the SNPs M9, M13, M20, SRY10831 and the single base pair deletion polymorphism M17 (Table 3.7). All the loci of UEP1 and UEP2 kit were amplified by PCR, digested with a cocktail of restriction enzymes in multiplex conditions and then typed by the presence or absence of DNA fragments of specific size and dye label. The status of the YAP polymorphism was determined by PCR product size only and does not require digestion with a restriction endonuclease. Typing of p12f2 was based on the absence or presence of an 88-bp PCR product. As an internal control a 148-bp product encompassing the M172 polymorphism was also co-amplified (Detailed protocol presented in Sec 3.7.3).

3.6.3.1 Genescan analysis of microsatellite and biallelic polymorphism

The microsatellite PCR products and UEP digestion products were run on ABI-3100 capillary based genetic analyzer. 2 μl of the microsatellite PCR products or UEP digestion products were mixed with 0.18 μl of fluorescent dye labeled size standard (500 Rox) and 12 μl of deionized formamide. Samples were then run using POP-4 polymer and a 36cm POP-4 capillary. Both microsatellite kits and UEP2 kit samples required electrophoresis for approximately 22 minutes, while the UEP1 kit samples required electrophoresis for approximately 26 minutes. The amplified DNA fragments generated are all under 430 bp and most are smaller than 230 bp. Allele sizes were obtained by analyzing the data with Gene Scan software.
Plate 3.4

a. Genomic DNA

b. Thermal Cycler

c. ABI-3100 Genetic Analyzer
3.7 PCR protocols

3.7.1 mtDNA RSP/Insertion deletion polymorphisms

*HaeIII 663*

Template: Genomic DNA 50-100 ng / reaction volume of 20 µl

Primer 1 (nt 577-598): 5′ - GTT TAT GTA GCT TAC CTC CTC - 3′  50ng/reaction

Primer 2 (nt 743-721): 5′ - GAT CGT GGT GAT TTA GAG GGT G - 3′  50ng/reaction

dNTPs Final concentration 200 µM each

Buffer: Final concentration 1.5mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 µl

PCR cycling profile: 94°C (4.0′)

94°C (30″), 58°C (30″), 72°C (30″) × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of *HaeIII* enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size : 167 bp

Digested product size: 80, 87 bp


* Plate 4.1a

*HpaI 3592*

Template: Genomic DNA 50-100 ng / reaction volume of 20 µl

Primer 1 (nt 3388-3409): 5′ – CTA GGC TAT ATA CAA CTA CGC – 3′

Primer 2 (nt 3717-3700): 5′ – GGC TAC TGC TCG CAG TG – 3′

dNTPs Final concentration 200 µM each

Buffer: Final concentration 2.25 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 µl

PCR cycling profile: 94°C (4.0′)

94°C (30″), 56°C (30″), 72°C (30″) × 30 cycles
PCR amplified segment is subjected to digestion with 3 U of *HpaI* enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size: 330 bp

Digested product size: 124, 206 bp


*Plate 4.1b

### AluI 5176

Template: Genomic DNA 50-100 ng / reaction volume of 20µl

Primer 1 (nt 099-5122): 5' – CCT AAC TAC TAC CGC ATT CCT AC – 3'

Primer 2 (nt 5274-5251): 5' – CTT CGA TAA TGG CCC ATT TGG GC – 3'

dNTPs Final concentration 200 µM each

Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 µl

PCR cycling Profile: 94°C (4.0')

94°C (30''), 61°C (30''), 72°C (30'') × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of *AluI* enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size: 175 bp

Digested product size: 78, 98 bp


*Plate 4.1c

### AluI 7025

Template: Genomic DNA 50-100 ng / reaction volume of 20µl

Primer 1 (nt 6890-6910): 5' – AAG CAA TAT GAA ATG ATC TG – 3'

Primer 2 (nt 7131-7114): 5' – CGT AGG TTT GGT CTA GG – 3'
dNTPs Final concentration 200 μM each
Buffer: Final concentration 2.25 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20 μl
PCR cycling profile: 94°C (4.0’)
   94°C (30’), 50°C (30’), 72°C (30’) × 30 cycles
PCR amplified segment is subjected to digestion with 3 U of *AluI* enzyme in appropriate buffer for 3 h at 37°C.
Gel: 2% Agarose in 1 x TBE buffer
Product size : 242 bp
Digested product size: 30, 75, 137 bp
*Plate 4.1d

*DdeI* 10394

Template: Genomic DNA 50-100 ng / reaction volume of 20μl
Primer 1 (nt 10284 - 10309): 5’- CCA TGA GCC CTA CAA ACA ACT AAC C - 3’
Primer 2 (nt 10484 - 10458): 5’- GTA AAT GAG GGG CAT TTG GTA AAT AT - 3’
dNTPs Final concentration 200 μM each
Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20 μl
PCR cycling profile: 94°C (4.0’)
   94°C (30’), 61°C (30’), 72°C (30’) × 30 cycles
PCR amplified segment is subjected to digestion with 3 U of *DdeI* enzyme in appropriate buffer for 3 h at 37°C.
Gel: 2% Agarose in 1 x TBE buffer
Product size : 200 bp
Digested product size: 72, 128 bp
   38, 72, 90 bp
*Plate 4.1e
**AluI 10397**

Template: Genomic DNA 50-100 ng / reaction volume of 20μl

Primer 1 (nt 10284 - 10309): 5' - CCA TGA GCC CTA CAA ACA ACT AAC C - 3'
Primer 2 (nt 10484 - 10458): 5'- GTA AAT GAG GGG CAT TTG GTA AAT AT - 3'

dNTPs Final concentration 200 μM each

Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 μl

PCR cycling profile: 94°C (4.0')

94°C (30''), 61°C (30''), 72°C (30'') × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of AluI enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size : 201 bp

Digested product size: 70, 131 bp


**HinfI 12308**

Template: Genomic DNA 50-100 ng / reaction volume of 20μl

Primer 1 (nt 12104 -12124): 5' - CTC AAC CCC GAC ATC ATT ACC -3'
Primer 2 (nt12338-12309): 5'- ATT ACT TTT T TGG AGT TGC ACC AAG ATT-3'

dNTPs Final concentration 200 μM each

Buffer: Final concentration 1.0 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 μl

PCR cycling profile: 94°C (4.0')

94°C (1.0'), 63°C (1.0'), 72°C (1.0') × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of HinfI enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size : 235 bp
Digested product size: 66, 169 bp

38, 66, 131 bp


*Plate 4.1f

**Hin**cII 13259

Template: Genomic DNA 50-100 ng / reaction volume of 20μl
Primer 1 (nt 13208 - 13232): 5' - CGC CCT TAC ACA AAA TGA CAT CAA -3'
Primer 2 (nt 13413 – 13393): 5’ - ATT TTT CGA ATA TCT TGT TC -3’
dNTPs Final concentration 200 μM each
Buffer: Final concentration 1.5 mM MgCl2, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20 μl
PCR cycling profile: 94°C (4.0’)

94°C (30”), 49°C (30”), 72°C (30”) × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of **Hin**cII enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 × TBE buffer

Product size: 206 bp

Digested product size: 53, 153 bp


*Plate 4.1g

**Alu**I gain at np 13262

Template: Genomic DNA 50-100 ng / reaction volume of 20μl
Primer 1 (nt 13208 - 13232): 5’ – CGC CCT TAC ACA AAA TGA CAT CAA -3’
Primer 2 (nt 13413 – 13393): 5’– ATT TTT CGA ATA TCT TGT TC -3’
dNTPs Final concentration 200 μM each
Buffer: Final concentration 1.5 mM MgCl2, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20 μl
PCR cycling profile: 94°C (4.0')

94°C (30”), 49°C (30”), 72°C (30”) × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of AluI enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size: 206 bp

Digested product size: 55, 151 bp


_HaeIII gain at np 16517_

Template: Genomic DNA 50-100 ng / reaction volume of 20μl

Primer 1 (nt 16453 - 16473): 5’ – CCG GGC CCA TAA CAC TTG GG -3’

Primer 2 (nt 48 - 25) : 5’ – GCA TGG AGA GCT CCC GTG AGT GG -3’

dNTPs Final concentration 200 μM each

Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3

Taq DNA polymerase: 0.5U/reaction volume of 20 μl

PCR cycling profile: 94°C (4.0’)

94°C (30”), 58°C (30”), 72°C (30”) × 30 cycles

PCR amplified segment is subjected to digestion with 3 U of HaeIII enzyme in appropriate buffer for 3 h at 37°C.

Gel: 2% Agarose in 1 x TBE buffer

Product size: 164 bp

Digested product size: 65, 100 bp


*Plate 4.1h

_9-bp deletion_

Template: Genomic DNA 50-100 ng / reaction volume of 20μl

Primer 1 (nt 8195 - 8215): 5’ – ACA GTT TCA TGC CCA TCG TC -3’

Primer 2 (nt 8317 - 8297): 5’ – ATG CTA AGT TAG CTT TAC AG -3’
dNTPs Final concentration 200 µM each
Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20 µl
PCR cycling profile: 94°C (4.0')
  94°C (30''), 55°C (30''), 72°C (30'') × 30 cycles
Gel: 6% Polyacrylamide in 0.5 x TBE buffer
Product size: 123 bp

*Plate 4.2

3.7.2 mtDNA HVS1 sequence polymorphism

Template: Genomic DNA 50-100ng/reaction volume of 20µl
Primer 1 (nt15996-5996): 5'- CTC CAC CAT TAG CAC CCA AAG C -3' - 1pm/reaction
Primer 2 (nt 16409) : 5'- TGA TTT CAC GGA GGA TGG TG -3' - 1pm/reaction
dNTPs Final concentration 20µM each
Buffer: Final concentration 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris HCl pH 8.3
Taq DNA polymerase: Hot start 0.5U/reaction of 20µl
PCR cycling profile: 94°C (4.0')
  94°C (1'), 60°C (1'), 72°C (1') × 30 cycles
Sequencing protocol
Template: 1µl PCR product / reaction volume of 20 µl
Primer: Forward/Reverse 3.2pm/reaction
Big Dye: 1.2 µl
5X Tris HCl: 3.0 µl
Sequencing PCR profile: 94°C (30.0'')
  90°C (10.0''), 50°C (0.05''), 60°C (4') × 25 cycles
3.7.3 Y chromosome polymorphisms

3.7.3.1 MSI kit

Template: Genomic DNA 100 ng / reaction volume of 10μl
dNTPs Final concentration 200μM each
Buffer: Final concentration 2.2 mM MgCl₂, 50 mM KCl, 10mM Tris-HCl (pH 9.0), 0.1% triton x 100, 0.01% gelatin
Taq DNA polymerase: 0.13 U/reaction of 10μl
PCR cycling profile: 95°C (10')
               94°C (30''), 57°C (30''), 72°C (30'') × 38 cycles
               72°C (20')

Table 3.4: Primer sequences and concentrations for MSI kit

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Dye label</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19-L</td>
<td>CTA CTG AGT TTC TGT TAT AGT</td>
<td>TET</td>
<td>0.236</td>
</tr>
<tr>
<td>DYS19-R</td>
<td>ATG GCA TGT AGT GAG GAC A</td>
<td></td>
<td>0.236</td>
</tr>
<tr>
<td>DYS388-L</td>
<td>GTG AGT TAG CCG TTT AGC GA</td>
<td>TET</td>
<td>0.318</td>
</tr>
<tr>
<td>DYS388-R</td>
<td>CAG ATC GCA ACC ACT GCG</td>
<td></td>
<td>0.318</td>
</tr>
<tr>
<td>DYS390-L</td>
<td>TAT ATT TTA CAC ATT TTT GGG CC</td>
<td></td>
<td>0.127</td>
</tr>
<tr>
<td>DYS390-R</td>
<td>TGA CAG TAA AAT GAA CAC ATT GC</td>
<td>FAM</td>
<td>0.127</td>
</tr>
<tr>
<td>DYS391-L</td>
<td>CTA TTC ATT CAA TCA TAC ACC CAT AT</td>
<td>FAM</td>
<td>0.384</td>
</tr>
<tr>
<td>DYS391-R</td>
<td>ACA TAG CCA AAT ATC TCC TGG G</td>
<td></td>
<td>0.384</td>
</tr>
<tr>
<td>DYS393-L</td>
<td>GTG GTC TTC TAC TTG TGT CAA TAC</td>
<td></td>
<td>0.180</td>
</tr>
<tr>
<td>DYS393-R</td>
<td>AAC TCA AGT CCA AAA AAT GAG G</td>
<td>HEX</td>
<td>0.088</td>
</tr>
</tbody>
</table>

3.7.3.2 MS2 kit

Template: Genomic DNA 100 ng / reaction volume of 10µl
dNTPs Final concentration 200µM each
Buffer: Final concentration 2.2 mM MgCl₂, 50 mM KCl, 10mM Tris-HCl (pH 9.0), 0.1% triton x 100, 0.01% gelatin
Taq DNA polymerase: 0.13 U/reaction of 10µl
PCR cycling profile: 95°C (10’)
94°C (30”), 55°C (30”), 72°C (30”) × 38 cycles
72°C (20’)

Table 3.5: Primer sequences and concentrations for MS2 kit

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Dye label</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS388-L-TET</td>
<td>GTG AGT TAG CCG TTT AGC GA</td>
<td>TET</td>
<td>0.185</td>
</tr>
<tr>
<td>DYS388-R</td>
<td>GGC GAC AGA GCG AGA GTC</td>
<td></td>
<td>0.185</td>
</tr>
<tr>
<td>DYS389-L-TET</td>
<td>CCA ACT CTC ATC TGT ATT ATC TAT GTG</td>
<td>TET</td>
<td>0.350</td>
</tr>
<tr>
<td>DYS389-R</td>
<td>CCT GAG TAG CAG AAG AAT GTC ATA</td>
<td></td>
<td>0.350</td>
</tr>
<tr>
<td>DYS425-L</td>
<td>TGG AGA GAA GAA GAG AGA AAT</td>
<td></td>
<td>0.150</td>
</tr>
<tr>
<td>DYS425-R-FAM</td>
<td>AGC TCT ACA AGC CAT TGT GAT CT</td>
<td>FAM</td>
<td>0.150</td>
</tr>
<tr>
<td>DYS426-L-HEX</td>
<td>GGT GAC AAG ACG AGA CTT TGT G</td>
<td>HEX</td>
<td>0.060</td>
</tr>
<tr>
<td>DYS426-R</td>
<td>CTC AAA GTA TGA AAG CAT GAC CA</td>
<td></td>
<td>0.060</td>
</tr>
</tbody>
</table>


3.7.3.3 UEPI kit (biallelic polymorphisms)

Template: Genomic DNA 100 ng / reaction volume of 10µl
dNTPs Final concentration 200µM each
Buffer: Final concentration 1.5mM MgCl₂, 50 mM KCl, 10mM Tris-HCl (pH 9.0), 0.1% triton x 100, 0.01% gelatin

Taq DNA polymerase: 0.13 U/reaction of 10µl

PCR cycling profile: 95°C (10')
94°C (30''), 56°C (30''), 72°C (30'') x 38 cycles
72°C (20')

Table 3.6: Primer sequences and concentrations for UEP1 kit

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Dye label</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>92R7-L</td>
<td>TGC ATG AAC ACA AAA GAC GTA</td>
<td></td>
<td>0.125</td>
</tr>
<tr>
<td>92R7-R-HEX</td>
<td>GCA TTG TTA AAT ATG ACC AGC</td>
<td>HEX</td>
<td>0.125</td>
</tr>
<tr>
<td>Tat-L</td>
<td>GAC TCT GAG TGT AGA CTT GTG A</td>
<td></td>
<td>0.078</td>
</tr>
<tr>
<td>Tat-R-TET</td>
<td>GAA GGT GCC GTA AAA GTG TGA A</td>
<td>TET</td>
<td>0.078</td>
</tr>
<tr>
<td>sY81-L-FAM</td>
<td>ATG GGA GAA GAA CGG AAG GA</td>
<td>FAM</td>
<td>0.125</td>
</tr>
<tr>
<td>sY81-R</td>
<td>TGG AAA ATA CAG CTC CCC CT</td>
<td></td>
<td>0.125</td>
</tr>
<tr>
<td>SRY+465-L</td>
<td>GCC GAA GAA TTG CAG TTT GC</td>
<td></td>
<td>0.055</td>
</tr>
<tr>
<td>SRY+465-R-HEX</td>
<td>GTT GAT GGG CGG TAA GTG GC</td>
<td>HEX</td>
<td>0.055</td>
</tr>
<tr>
<td>SRY4064-L-TET</td>
<td>GGT ATG ACA GGG GAT GAT GTG A</td>
<td>TET</td>
<td>0.095</td>
</tr>
<tr>
<td>SRY4064-R</td>
<td>CCA CGC CCA GCT AAT TTT TTT GC</td>
<td></td>
<td>0.095</td>
</tr>
<tr>
<td>YAP-c-TET</td>
<td>AGG ACT AGC AAT AGC AGG GGA AGA</td>
<td>TET</td>
<td>0.100</td>
</tr>
<tr>
<td>YAP-d</td>
<td>CAG GGC CAA CTC CAA CCA AG</td>
<td></td>
<td>0.100</td>
</tr>
</tbody>
</table>

Digestions were performed in micro titer plates in a final volume of 8 µl. Each reaction contained 2 µl of PCR product, NEB buffer 4 (New England Biolabs, Beverly) to 1x concentration, 0.01µg/µl acetylated BSA, 0.3U Bsr B1, 0.3U Fnu 4HI, 0.3U NlaIII and 1.8U HindIII. Plates were incubated at 37°C for 3h. Predicted sizes and associated polymorphic status for each dye-labelled PCR product is given in table 3.8.

3.7.3.4 UEP2 kit

Template: Genomic DNA 100 ng/reaction volume of 10µl
dNTPs Final concentration 200µM each
Buffer: Final concentration 1.5mM MgCl₂, 50 mM KCl, 10mM Tris-HCl (pH 9.0), 0.1% triton x 100, 0.01% gelatin
Taq DNA polymerase: 0.13 U/reaction of 10µl
PCR cycling profile: 95°C (10')
94°C (30''), 56°C (30''), 72°C (30'') × 38 cycles
72°C (20')

Table 3.7: Primer sequences and concentrations for UEP2 kit

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Dye label</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>M9-L-TET</td>
<td>TCA GGA CCC TGA AAT ACA GAA CT</td>
<td>TET</td>
<td>0.125</td>
</tr>
<tr>
<td>M9-R</td>
<td>TTG AAG CTC GTG AAA CAG ATT AG</td>
<td></td>
<td>0.125</td>
</tr>
<tr>
<td>M13-L-HEX</td>
<td>TAG TTT ATG CCC AGG AAT GAA C</td>
<td>HEX</td>
<td>0.078</td>
</tr>
<tr>
<td>M13-R</td>
<td>ATC CAA CCA CAT TTG CAA AA</td>
<td></td>
<td>0.078</td>
</tr>
<tr>
<td>M17-L</td>
<td>GTG GTT GCT GGT TGT TAC GT</td>
<td></td>
<td>0.125</td>
</tr>
<tr>
<td>M17-R-TET</td>
<td>AGC TGA CCA CAA ACT GAT GTA GA</td>
<td>TET</td>
<td>0.125</td>
</tr>
<tr>
<td>M20-L-FAM</td>
<td>AGT TGG CCC TTT GTG TCT GT</td>
<td>FAM</td>
<td>0.055</td>
</tr>
<tr>
<td>M20-R</td>
<td>CAT GTT CAG TGC AAA TGC AAC</td>
<td></td>
<td>0.055</td>
</tr>
<tr>
<td>SRY10831-L-FAM</td>
<td>TCA TTC AGT ATC TGG CCT CTT G</td>
<td>FAM</td>
<td>0.095</td>
</tr>
<tr>
<td>SRY10831-R</td>
<td>CAC CAC ATA GGT GAA CCT TGA A</td>
<td></td>
<td>0.095</td>
</tr>
</tbody>
</table>

Digestions were performed in 384 well micro titer plates in a final volume of 8µl. Each reaction contained 2µl of PCR product, NEB buffer 3 (New England Biolabs, Beverly) to 1x concentration, 0.01µg/µl acetylated BSA, 0.32 U Hinfl, 0.32 U Bsp1431, 0.32 U AflIII, 0.32 U Sspl and 0.32 U DraIII. Plates were incubated at 37°C for 3h. Predicted sizes and associated polymorphic status for each dye-labelled PCR product is given in table 3.8.

Reference: Thomas et al. 1999
### Table 3.8: UEP kits - expected product sizes after multiplex restriction enzyme digestion

<table>
<thead>
<tr>
<th>Kit</th>
<th>Polymorphism</th>
<th>Discriminating enzyme</th>
<th>Non-discriminating enzyme that also cuts</th>
<th>Labelled primer</th>
<th>Dye label</th>
<th>PCR product size</th>
<th>Labelled fragment size when discriminating enzyme cuts PCR product</th>
<th>Labelled fragment size when discriminating enzyme does not cut PCR product</th>
</tr>
</thead>
<tbody>
<tr>
<td>UEP1</td>
<td>92R7</td>
<td><em>Hind</em>III</td>
<td><em>Nla</em>III</td>
<td>Right</td>
<td>HEX</td>
<td>55</td>
<td>28 (C)</td>
<td>53 (T)</td>
</tr>
<tr>
<td>Tat</td>
<td><em>Nla</em>III</td>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td>112</td>
<td>83 (T)</td>
<td>112 (C)</td>
</tr>
<tr>
<td>sY81</td>
<td><em>Nla</em>III</td>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td>142</td>
<td>105 (A)</td>
<td>142 (G)</td>
</tr>
<tr>
<td>SRY+465</td>
<td><em>Fnu</em>4HI</td>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td>148</td>
<td>98 (C)</td>
<td>148 (T)</td>
</tr>
<tr>
<td>SRY4064</td>
<td><em>Bsr</em>BI</td>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td>225</td>
<td>135 (G)</td>
<td>174 (A)</td>
</tr>
<tr>
<td>YAP</td>
<td>NA</td>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td>99/413</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>UEP2</td>
<td>M9</td>
<td><em>Hinfl</em></td>
<td><em>Afl</em>III,<em>Ssp</em>I</td>
<td>Left</td>
<td>TET</td>
<td>214</td>
<td>48 (C)</td>
<td>76 (G)</td>
</tr>
<tr>
<td>M13</td>
<td><em>Bsp</em>1431</td>
<td>Left</td>
<td></td>
<td></td>
<td>HEX</td>
<td>119</td>
<td>56 (G)</td>
<td>119 (C)</td>
</tr>
<tr>
<td>M17</td>
<td><em>Afl</em>III</td>
<td>Right</td>
<td></td>
<td></td>
<td>TET</td>
<td>124</td>
<td>101 (-G)</td>
<td>124 (+G)</td>
</tr>
<tr>
<td>M20</td>
<td><em>Ssp</em>I</td>
<td>Left</td>
<td></td>
<td></td>
<td>FAM</td>
<td>108</td>
<td>62 (A)</td>
<td>102 (G)</td>
</tr>
<tr>
<td>SRY10831</td>
<td><em>Dra</em>III</td>
<td>Left</td>
<td></td>
<td></td>
<td>FAM</td>
<td>73</td>
<td>41 (G)</td>
<td>73 (A)</td>
</tr>
</tbody>
</table>

NA - Not Applicable

Reference: Thomas *et al.* 1999
3.7.3.5 *p12f2* marker

Template: Genomic DNA 50-100 ng / reaction of 20 μl
Primer (12F2D): 5'-CTG ACT GAT CAA AAT GCT TAC AGA TC-3' – 0.2μM /reaction
Primer (12F2G): 5'-GGA TCC CTT CCT TAC ACC TTA TAC -3' - 0.2μM /reaction
For internal control
Primer (M172 F): 5' – TCC CCC AAA CCC ATT TTG ATG CAT -3' – 0.3μM/reaction
Primer (M172 R): 5' – GGA TCC ATC TTC ACT CAA TGT TG - 3T- 0.3μM/reaction
dNTPs: Final concentration 200μM each
Buffer: Final concentration 2.5mM MgCl₂, 50mM, KCl, 10mM Tris HCl pH 8.3
Taq DNA polymerase: 0.5U/reaction volume of 20μl
PCR cycling profile: 94°C (5.0')
94°C (30") , 58°C (45") , 72°C (45") × 30 cycles
Gel: 3% LMP + LE Agarose (3:1) in 1x TBE buffer
Product size: 88bp
Reference: Rosser *et al.* 2000; Nebel *et al.* 2001; Mukherjee *et al.* 2001
*Plate 4.3

3.8 Data analysis

3.8.1 Allele frequency estimation

At a polymorphic locus, the proportion of a particular allele in a population is called
the allele frequency. For a population of haploid individuals, an estimate of the frequency
of the allele A₁, at a biallelic locus with alleles A₁ and A₂, in a sample of size N is:

\[ x₁ = \frac{N₁}{N}, \]

where \( N₁ \) is the number of individuals carrying. The estimate of the frequency of the allele
A₂ is:

\[ x₂ = 1 - x₁. \]
3.8.2 Haplotype and haplogroup frequencies

A haplotype is a combination of specific alleles present on a chromosome at a set of linked loci. For \( k \) linked biallelic loci, theoretically \( 2^k \) haplotypes are possible. However, in a specific population, the observed number of haplotypes, say \( l \), is generally smaller than \( 2^k \).

The estimate of the frequency of the \( i^{th} \) haplotype (\( i = 1, 2, \ldots, l \)) in a sample of \( N \) chromosomes is \( \frac{N_i}{N} \),

where \( N_i \) denotes the number of chromosomes with haplotype \( i \). Based on specific allelic configurations, haplotypes can often be phylogenetically grouped into disjoint classes, called haplogroups. The sum of the proportions of all haplotypes belonging to a haplogroup is called the haplogroup frequency.

3.8.3 Haplotype diversity analysis

The simplest measure of polymorphism of a short DNA segment is the number of haplotypes observed in the sample and it depends strongly on sample size. A more appropriate measure of polymorphism is haplotype diversity (\( h \)) defined as (Nei and Tajima 1981):

\[
h = n\left(1 - \sum x_i^2\right) / (n - 1)
\]

where \( x_i \) is haplotype frequency, \( n \) = number of individuals sampled; the summation is taken over all haplotypes present in the sample.

3.8.4 Analysis of Molecular variance (AMOVA)

The Analysis of Molecular Variance approach (AMOVA), (Excoffier et al. 1992) takes into account the number of mutations between molecular haplotypes. A particular genetic structure is imposed by grouping populations. A hierarchical analysis of variance partitions the total genomic variance into components ascribable to intra-individual differences, inter-individual differences and/or inter-population differences.
In the haploid case, the assumption is that the \( i^{\text{th}} \) haplotype frequency vector of the \( j^{\text{th}} \) population in the \( k^{\text{th}} \) group is a linear equation:

\[
X_{ijk} = X + a_k + b_{jk} + c_{ijk}
\]

where \( i = 1, 2, \ldots, n \), \( j \) = number of individuals sampled from the \( j^{\text{th}} \) population,

\( j = 1, 2, \ldots, t \), \( k \) = number of populations in \( k^{\text{th}} \) group,

\( k = 1, 2, \ldots, G \) = total number of groups.

\( X \) is the unknown expectation of \( X_{ijk} \) averaged over all sampled individuals. The group effects are designated as \( a \), population effects within groups as \( b \) and haplotype effects within a population within a group as \( c \). These effects are assumed to be additive, random, and independent. The variances associated with these effects are defined by \( \sigma_a^2 \), \( \sigma_b^2 \) and \( \sigma_c^2 \), respectively. The total molecular variance \( (\sigma^2) \) is therefore, the sum of the variance component due to differences among haplotypes within a population \( (\sigma_c^2) \), the variance component due to differences populations within a group \( (\sigma_b^2) \), and the variance component \( (\sigma_a^2) \) due to differences among groups. AMOVA analysis was performed using ARLEQUIN version 2.0 (Schneider et al. 2000) for both mtDNA RSP's and HVS1 nucleotide data.

3.8.5 Nucleotide sequence alignment

Mitochondrial hypervariable segment-1 sequences were aligned using program Clustal W (Thompson et al. 1994) The Cambridge Reference sequence (Anderson et al. 1981) was used as the reference sequence during alignment.

3.8.6 Nucleotide polymorphism and nucleotide diversity analysis

Sequence data were used to estimate the level of genetic variation at the nucleotide level. The level of nucleotide polymorphism, symbolized \( \theta \), is the proportion of nucleotide sites that are expected to be polymorphic in any sample from the particular region of the genome (Hartl and Clark 1997).
It is defined as: \( \theta = 4N_e \mu \), where \( N_e \) is the effective size of a diploid population and \( \mu \) is the mutation rate per site per generation.

\( \theta \) can be estimated from the total number of segregating sites (\( S \)) (Watterson 1975) as well as from the average number of mismatches between two sequences (\( k \)).

The estimate \( \theta \) equals the proportion of nucleotide polymorphism observed in sample, symbolized as \( S \), divided by:

\[
a = \sum_{i=1}^{\frac{n(n-1)}{2}} \frac{1}{l}
\]

Where \( n \) is the sample size. The estimate of \( \theta \), per nucleotide site is therefore:

\[
\theta = \frac{S}{al}
\]

\( \theta \) was also estimated from the average number of mismatches between sequences (\( k \)). For a pair of DNA sequences, it is more informative to consider the number of nucleotide differences between the two sequences rather than just whether they are different. Thus, for DNA sequence data, a more appropriate measure of polymorphism is the average number of nucleotide differences between two sequences randomly chosen from the population. When \( n \) sequences are taken randomly from the population, \( k \) is estimated by:

\[
k = \frac{\sum_{i<j} k_{ij}}{\frac{n(n-1)}{2}}
\]

where \( k_{ij} \) is the number of nucleotide differences between the \( i^{th} \) and \( j^{th} \) sequences and \( n(n-1)/2 \) is the number of possible pairs.

Under the infinite sites model (Nei 1987), which assumes that the number of nucleotide sites on the sequence is so large that each new mutation occurs at a site that has not been mutated before and under the assumption of random mating, the mean of \( k \) is
given by $E(k) = \theta$. ARLEQUIN version 2.0 (Schneider et al. 2000) and DnaSP (version 3.51; Rozas et al. 1999) was used to estimate $\theta$.

3.8.7 Nucleotide diversity ($\pi$)

This quantity is used to assess polymorphisms at the DNA level and is typically denoted as $\pi$. Nucleotide diversity is the number of nucleotide differences per site between two randomly chosen sequences (Nei 1987). This can be estimated as:

$$\pi = \frac{k}{l}$$

where $l = \text{length of the sequence}$ and $k = \text{the average number of mismatches between sequences}$. Nucleotide diversity was calculated in this study using ARLEQUIN version 2.0 (Schneider et al. 2000).