Y-CHROMOSOME POLYMORPHISM

IN

THE NEPALESE POPULATION

1998

RAJU ADHIKARI

Department of Biochemistry

University of Oxford

SPECIAL THANKS TO

Dr. Chris Tyler-Smith, Dr. Fabricio R Santos, Miss Arpita Pandya

(Dept of Biochemistry, University of Oxford)

Prof. Mathura Pd. Shrestha, Dr. Madan Mani Dixit

(Nepal Health and Research Council, Ministry of Health, Nepal)

&

ALL BLOOD DONORS

Abstract:

Thirty-eight Nepalese blood samples were collected from male donors and DNA was extracted from them by a conventional method (for 36 samples) or by using Agarose plugs (2 samples). Each of the samples was scored for Y- Chromosome polymorphisms with 10 PCRable markers available. Y haplotypes found in Nepal were compared with those found in genetically and geographically distinct parts of the world using the data already obtained (1). A Neighbor-joining tree and a Maximum likelihood tree were constructed to assess the results (2) . It was found that the Nepalese are genetically closely related to the Indians and the Sri Lankans and two Kirants showed a new αh polymorphism, designated XXIX (3)

Introduction

The Y- Chromosome is passed down the paternal line (father-to-son). Except for the pseudoautosomal region, the chromosome does not undergo homologous recombination at meiosis and passes down the paternal line accumulating mutations (4). By analyzing and knowing the state of these polymorphisms as well as that of the Out Group, it is possible to build phylogenetic trees of DNA loci and hence get idea about ancestral history of a particular population. This has been used in the identification of the origin of anatomically modern humans (5-8).

Studies of human evolution have been based on archaeology, paleontology, anthropology and linguistics (9-11). All of these have limits and the results are not always clear and reliable. Genetic markers including blood groups (the detection of ABO, RH and MN blood -groups systems) , protein polymorphisms (especially of proteins present in Serum or Plasma or those in the RBCs) and Human Lymphocytes Antigen (HLA) polymorphisms also have been useful in such studies. However, these systems may not be selectively neutral and may not directly reflect variations at the genetic level. In more recent work, the DNA has been analyzed directly.

Mitochondrial DNA (mtDNA) has been used extensively to study variations at genetic level. The first DNA polymorphisms examined for evolutionary purposes were from mtDNA. Mitochondria pass down from one generation to the next through the maternal lineage (mother-to-daughter) (12). Because mtDNA does not recombine, its genetic pattern is usually an exact replica of the mother's except when mutations occur. mtDNA polymorphisms have been studied extensively by Wilson's group at Berkeley, California. Polymorphisms could be scored either by electrophoretic analysis of RFLP or by sequencing a segment. By estimating the number of mutations accumulated since its common origin, it was possible to draw an evolutionary tree of mtDNA.

mtDNA can only provide evolutionary history through maternal line. The Y-Chromosome does the same through paternal line and the data obtained from Y-Chromosome analysis could be used to compare with the results obtained from other analyses. Moreover, Y- specific polymorphisms present in the Y-Chromosome can be powerful in distinguishing ethnic origin.

Polymorphisms on Y

The polymorphisms found on the Y include base substitutions, rearrangements (insertions, deletions, duplications, translocations, inversions) and tandemly repeated loci including major satellites, mini satellites and micro satellites. These polymorphisms have different mutation rates which could be useful for studying evolution in different time scale.

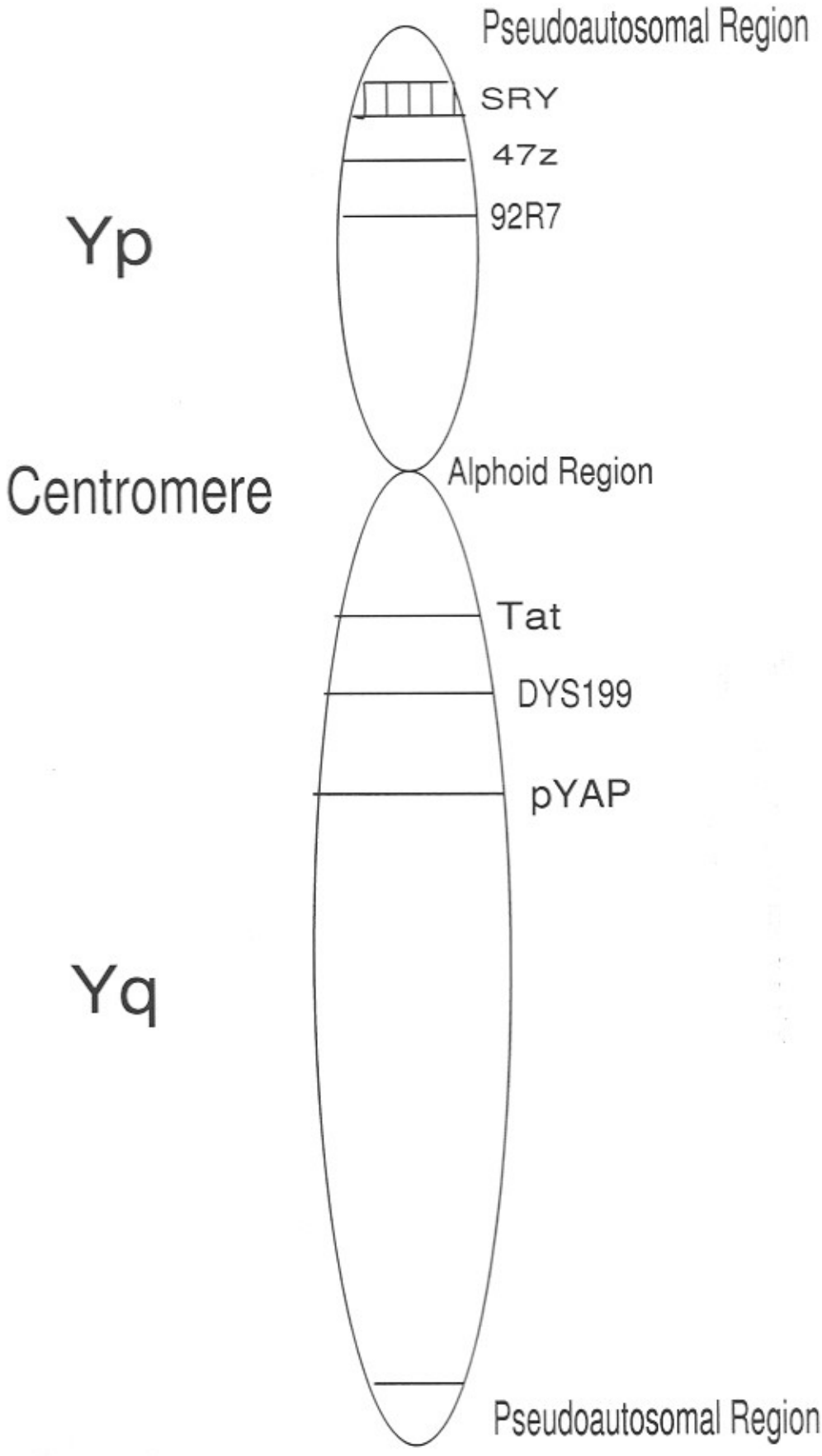
For the analysis of the Nepalese samples, this author used a collection of markers that could be scored using PCR.

αh : This polymorphism has been described fully in Santos *et al.*, 1995. It detects the presence or absence of alphoid units and multiple point mutations one is an A ↔ G mutation which creates a *Hin*dIII site [equivalent to the presence of the 6 kb *Hin*dIII Y alphoid unit]. The ancestral allele is not known (3,13). The Y- location of this polymorphism is shown in Figure 1.

92R7: This occurs in a *Hin*dIII site on the Y- Chromosome. It is a T ↔ C transition on the Yp. The C mutation creates a *Hin*dIII site [equivalent to the 92R7 4.6 kb band]. The ancestral allele is not known for sure but is thought to be the C allele. The T allele is found in Asia, Europe and America and the C allele is found in these continents as well as in Africa. (14)

SRY-1532: This is found on Yp. Here, an A → G mutation creates a *Dra*III site. The ancestral allele is A which is found in Chimpanzee but most of the screened samples so far have had the G allele. The A allele has been found in Africa,

asic construction of the Human Y-Chromosome. The position of Markers shown is approximate and OT to scale.



1

(San populations) in association with the 92R7 C allele, and in Europe and Asian Populations in association with the 92R7 T allele. (15)

pYAP: This is an Alu insertion polymorphism on the Yq. The ancestral state is the YAP^- allele. The YAP^+ allele is frequently found in Africa and also in the Japanese. It is present at lower frequency in the European populations. (16, 17)

47z: This is an unique event marker where a mutation creates a *Stu*I site. It is part of an XY homologous region in the Yp. It is widespread in Japan (~ 28%) and is also found in Taiwan and Korea. (18-20)

DYS199: The polymorphism is found on the Yq. The transition from C ↔ T allele destroys an artificial *Mfe*I site created by a mismatched primer (for primer sequence see below), changing A → G. This polymorphism is found in Amerindians. (21)

SRY-2627: A C ↔ T allele transition destroys a *Ban*I site on Yp. This is equivalent to pSRY 373 described by Bianchi et al. (unpublished) (22)

SRY+465: This polymorphism is a C ↔ T transition and has been described by Shinka *et al*, 1996. The change to the T allele causes the destruction of a *Bso*FI site (*Fnu*4H site). (23)

Tat: This occurs in the Yq region of the chromosome. A T → C mutation destroys an *Hsp* 92II site as well as creating a *Mae*II site. This polymorphism is found in central Asia and Scandinavia. (24)

sY81: This is an A → G mutation in STS sY81. The presence of the G allele creates an *Hsp*92II site. The ancestral state is the A allele. This polymorphism has been seen mainly in the African population. The G allele is linked to the $pYAP^+$ allele. (25)

Nepal: A brief History

Nepal is a country with a rich ancient history. In the past, Nepal meant only the Kathmandu valley (now the capital). The first rulers of the valley are believed to be the Gopalas (meaning herdsmen). Buddhist chronicles and Hindu 'Purans' also describe the Gopalas. The period when they ruled Nepal is not known. After the Gopalas came the Kirants and this is believed to have taken place between 7^{th} - 8^{th} century BC. Siddhartha Gautam (Gautam Buddha) visited the valley during the reign of the seventh king who introduced Buddhism. Kiranti domination ended around 300 AD. Kirants have turned out to be interesting in terms of their origin. We do not have authentic records like inscriptions, manuscripts, colophons or coins and this has led history scholars to hypothesise their origin. Most old history scholars have the view that they came from Mongolia via Tibet and before invading the Kathmandu valley they settled in the Eastern part of Nepal. Recent Historians have a different view. Kirants race comes from the word 'Kiriath' meaning fort or town in the Mobite language of the Mediterranean region. The ancestors of the Kirants had a settlement in Nahor and Ur. A branch came to Mesopotamia (approx. 2400 BC) after they were expelled by the Hebrews. One of the groups within this branch migrated to the Himalayan region via Afghanistan. There is evidence of their existence in Media and Northern Persia. Even today, Kirants have a distinct culture, a special form of writing, peculiar laws of marriage (love marriage and divorce very common which is very unusual to other races in Nepal), social conducts, civil and administrative code that is connected with Kirat-Ashur people in Assyria. (26)

Lichhivis, the Indo - Aryan people coming from northern India , ruled Nepal after the Kirants. The Invasion took place around 300 AD. Along with agricultural technologies, the Aryans brought Hinduism and the Caste system. They were able to establish themselves in a very short time and are the longest rulers of Nepal in its history. Today, the majority of the Nepalese population

belongs to this race. There are other races in Nepal but their origin and migrations have not been described.

The samples collected were mostly from the Aryans. Two samples were from the Kirants race. Others were from different racial groups like the Sherpas, Newars, Gurungs etc.

Outline of the Project

The aim of my project was to screen the 36 Nepalese samples for Y-Chromosome polymorphisms and compare with the populations found in different parts of the world. The samples were screened with 10 PCRable markers available, both old and new. The frequency of the various markers in the Nepalese population was calculated. The frequencies of the same markers for other populations were available from (1). It was possible to construct unrooted evolutionary trees with these data using the computer program package PHYLIP which is described later.

Methods

Blood Collection:

The blood samples were taken with lithium heparin-coated Vacutainers. Approximately 10 ml of blood was collected from each donor. They were then stored in a cool place for an hour after which they were centrifuged at 10,000 rpm. The whitish -yellow buffy coat was then extracted and put in an eppendorf tube which was deep frozen for transport to the UK.

Isolation of DNA from frozen Blood:

The blood was defrosted by letting it sit at room temperature and aliquoted into volumes of approximately 700 µl. An equal volume of 1× SSC was added, mixed gently and centrifuged at 5000 $min.^{-1}$ for 5 minutes. The supernatant which contains the lysed Red Blood Cells was discarded and the washing was repeated two more times. The final pellet of white blood cells was resuspended in 700µl of 1×SSC, 10mM EDTA pH 7.6. SDS and Proteinase K were added to the final concentration of 0.5% and 150µg/ml. The solution was mixed gently and was incubated at 50^0 C overnight. After the incubation , the solution was cooled to room temperature and phenol extracted. Phenol extraction was done by adding equal volume of phenol, vortexing it and centrifuging at 14,000 $min.^{-1}$ for 3 minutes. Two phases are formed, the organic phase and the aqueous phase which contains the soluble DNA. The aqueous phase was transferred to a clean tube and phenol extraction repeated two more times. After the final extraction, the aqueous phase, containing the DNA was added to a tube containing 2 1/2 volumes of 96% ethanol. This would precipitate the DNA (a white formation is usually seen). The solution was then centrifused at 14,000 $min.^{-1}$ for 3 minutes to obtain white precipitate of DNA The pellet was resuspended in 200µl of sodium acetate 0.2 M pH 7.0 and the ethanol precipitation was again repeated as mentioned earlier. The final pellet was washed in 70% ethanol and left at room temperature for 10 minutes. The solution was centrifuged at 14,000 $min.^{-1}$ for 3 minutes. The ethanol was discarded and the DNA was left to dry at room temperature (approximately 10 minutes). The DNA was resuspended in an appropriate volume of TE (between 300µl - 500µl) depending on the size of the pellet. The DNA was run on an Agarose gel to measure the concentration.

Preparation of Agarose Gel:

Sigma I-A: Low EEO Agarose is available on Powder form. 1-1.5% Agarose gel was prepared by weighing 1-1.5g of powder agarose in a conical flask, adding 100 ml of 0.5× TAE and getting it to boil. The weight of the flask is taken before and after boiling , and if necessary $_{dd}$water is added after boiling in order to compensate for the loss of water by evaporation after boiling. The gel was cooled to approximately 55° C and poured on the gel tray. It was allowed to sit at least for half an hour before loading the sample.

Loading the gel:

10 µl of samples were prepared with 2µl DNA, 2µl Loading Buffer (30% Ficoll and Orange G was used) and 6 µl $_{dd}$water. 4µl of λ digested with *Hin*dIII was used as the molecular marker which contained approximately 200ng of DNA.

Staining was done with EtBr for about 20 minutes with constant agitation. Recipro - Shake, Luckham LTD, was used for agitation. The gel was then destained with $_{dd}$water for 2 minutes, observed on a UV transilluminator and a photograph taken. The concentration of the prepared DNA was calculated by comparing the intensity of the band of the DNA to a particular band of λ *Hin*dIII digest. [λ *Hin*dIII marker had the bands 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb and 0.6kb.]

Preparation of high molecular weight DNA in Agarose plugs:

DNA was prepared from agarose plugs for two Nepalese samples (852 and 853). DNA was prepared in agarose plugs at a concentration of approximately 1×10^7

WBCs/ml of 0.75 LGT Agarose. The plugs were stored at 4° C in NDS (500mM EDTA pH9.5, 1% lauryl sarcosine). The PCRable DNA was prepared by soaking a third of an agarose plug in TE (10mM Tris HCL pH 7.5, 1mM EDTA) for 30 minutes. Washing was repeated three times. The plug was then melted at 65° C for 10 minutes in 200µl TE and vortexed vigorously using MSI minishaker, IKA, at full speed. Melting and vortexing was repeated until a clear solutions (plug completely dissolved) was obtained. The concentration of the DNA was then obtained with the method which has been already described above. (27)

PCR

All the samples were scored with PCRable markers. An MJ Research PTC-200 Thermocycler was used for all the amplifications carried out. All the reactions were preformed in a final volume of 12.5 µl. The reagents and their amounts used for various PCR systems are tabulated below.

(All PCR Conditions and Reagents as suggested by Dr. Fabricio Santos)

Locus	Name	Buffer	MgCl2 (mM)	dNTPs (mM)	Primers (mM)	Tag Pol.(U)
aH		1x PROMEGA	1.5	200	1	1
	92R7	1x PROMEGA	1.5	200	1	0.5
SRY	-1532	1x PROMEGA	1.5	200	1	0.5
	YAP	1x PROMEGA	1.5	200	1	0.5
	47z	1x BIOLINE	1.5	200	1	0.5
	TAT	1x PROMEGA	1.5	200	1	1
DYS199		1x PROMEGA	1.5	200	1	1
SRY	-2627	1X BIOLINE	1.5	200	1	0.5
SRY	465	1X BIOLINE	1.5	200	1	0.5
sY81		1X BIOLINE	1.5	200	1	1

A sample PCR protocol used is given on the next page. The amounts of reagents used are shown. All except Taq Polymerase, are used at constant volume for every assay performed. Depending on the type of PCR system, the amount of Taq Polymerase used was also different.

PCR Garne Tarika

PCR TAT	Karyakram TAT-1		Chakra 30	Jodne Tapkram
Miti 08/08/96	**Upabhokta** Raju		Yantra PTC200	57 °C
Samanharu	Samanharu	Bholume (µl)	Antya	Dekhne
	ddH_2O	205.20		
10 X	Buffer 10X	30.00	1 X	Bioline
- mM	MgCl2	-	1.5 mM	
2.5 mM	dNTPs	24.00	200 µM	
25 µM	primers	12.00	1 µM	
10 u/µl	Taq Polimerase	4.80	1 u/tubo	Bioline
0 µM			0 µM	
0 ng/µl			0 ng/tubo	
0 %			0 %	
0 mM			0 mM	
Pura Bholume 300 µl	Pura Bholume	276	24 X 11.5 µl	
		ou →	tubos Bholume/tube	

DNA Samples Halne		24 X 1 µl
		samples Bholume/tube
Hala	stock concentration	amount
template	40 ng/µl	1 µl/tubo
	%	0 µl/tubo
	0	0 µl/tubo

Samples:

1 837 –	2 838 –	3 835 –	4 840 –	5 841 –
6 842 –	7 843 –	8 844 –	9 845 –	10 846 –
11 847 –	12 848 –	13 845 –	14 850 –	15 857 –
16 14 –	17 10 –	18 35 –	19 20 –	20 15 –
21 62 –	22 295 (true)	23 (-ve) ddH_2O	24	

The following conditions and Primer sequences were used for amplifications

Locus	Name	Primers 5' 3'	Conditions	Ref.
aH			35 Cycles:	
				3,13
			72o C for 10 min.- Final Extension	
			94o C for 30 sec, 72o C for 1 min.	
			65o C for 30 sec	
	92R7		30 Cycles:	
				unpublished
			72o C for 10 min.- Final Extension	
			94o C for 20 sec, 72o C for 30 sec	
			60o C for 20 sec	
SRY	*-1532*		30 Cycles:	
				15
			72o C for 5 min.- Final Extension	
			94o C for 20 sec, 72o C for 30 sec	
			60o C for 20 sec	
	pYAP		30 Cycles:	
				17
			72o C for 5 min.- Final Extension	
			94o C for 30 sec, 72o C for 60 sec	
			55o C for 30 sec	
	47z		30 Cycles:	
				23
			72o C for 5 min.- Final Extension	
			94o C for 30 sec, 72o C for 60 sec	
			55o C for 30 sec	
	Tat		30 Cycles:	
				24
			72o C for 5 min.- Final Extension	
			94o C for 20 sec, 72o C for 30 sec	
			61o C for 20 sec	
DYS199			30 Cycles:	
				21
			72o C for 7 min.- Final Extension	
			94o C for 20 sec, 72o C for 30 sec	
			61o C for 20 sec	
SRY	*-2627*		30 Cycles:	
				22
			72o C for 5 min.- Final Extension	
			94o C for 30 sec, 72o C for 60 sec	
			63o C for 30 sec	
SRY	*465*		30 Cycles:	
				23
			72o C for 5 min.- Final Extension	
			94o C for 20 sec, 72o C for 45 sec	
			58o C for 20 sec	
sY81			30 Cycles:	25
			72o C for 5 min.- Final Extension	
			94o C for 20 sec, 72o C for 45 sec	
			58o C for 20 sec	

All the references were kindly supplied by Dr. Fabricio R. Santos.

PCR Analysis:

The PCR products were analyzed by polyacrylamide gel electrophoresis (PAGE) (3) and silver staining (29). The polymorphisms were read directly from the gel and scored by the method described below. A negative and positive controls were employed in order to avoid confusion with contamination. The gel can be dried immediately but is usually better if it is kept in fixation solution at 4° C overnight. The gel was dried using Promega Gel Drying film, 25.5 x 28 cm. The dried gel can be stored permanently at room temperature or photograph can be developed from it (X.Ref. gel photographs on next page)

Data Analysis:

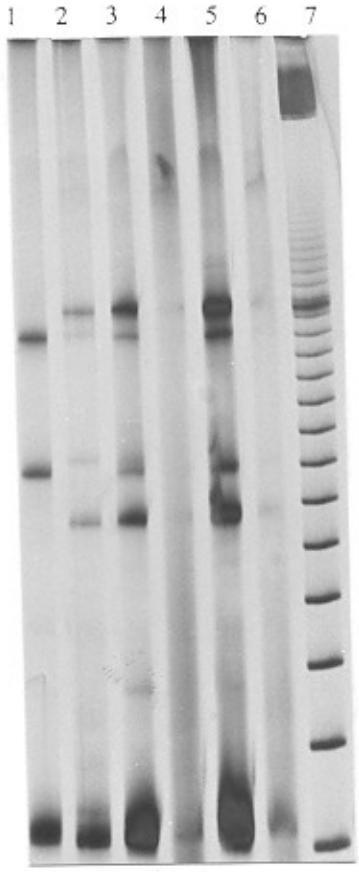
The data obtained were analyzed in different ways to compare and contrast the results obtained.

The frequency of each polymorphic locus was calculated for the Nepalese population and the frequencies for other parts of the world were obtained from (1). The gene frequency was then used to construct the unrooted neighbor-joining tree and maximum likelihood tree. " gendist" program within PHYLIP package (Joseph Felsenstein, Universiity of Washington, USA) was used to obtain Nei's genetic distance and " neighbor" and "treetool" (Mike Maciukenas, University of Illinois, USA) programs were used for the construction of unrooted Neighbor-Joining Tree. "contml" and "treetool" programs were similarly used for the construction of maximum likelihood tree.

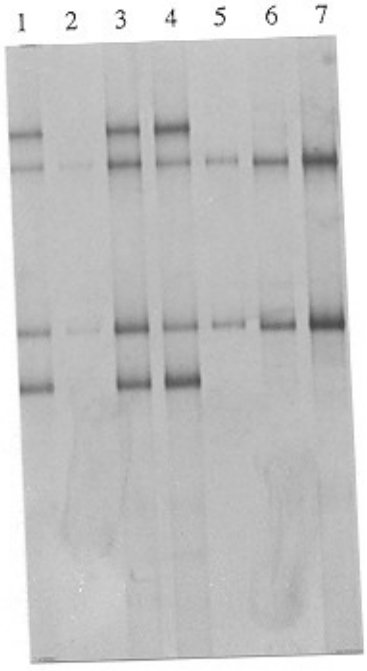
In order to find the diversity within populations, Nei's method was used. This is the most widely accepted method for calculating diversity and it has the formula

$$\text{Diversity} = 1 - \sum_i P^2$$

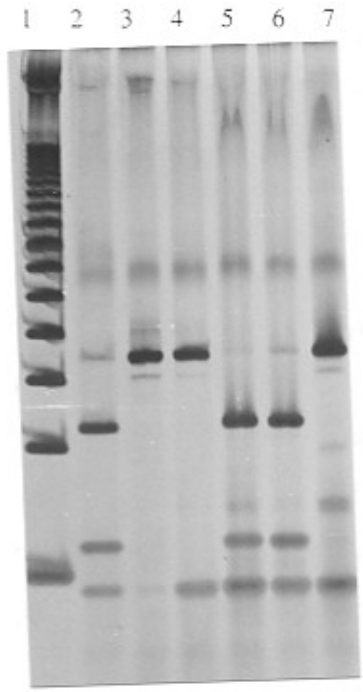
where, P is the allele frequency.



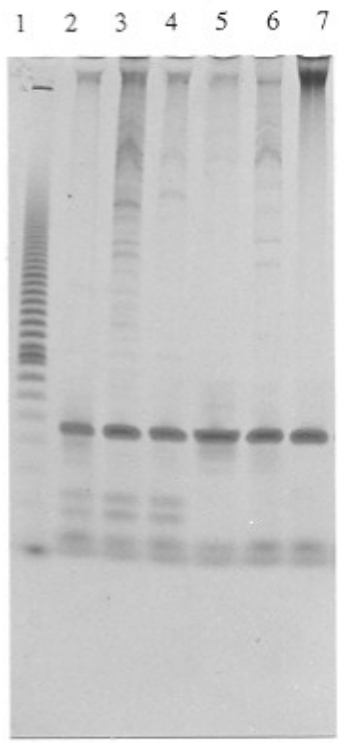
∝h: Lanes 1-6 are samples 835,836,837,838,839 and 840 respectively. Samples in lanes 4 and 6 didn't amplify. Lanes 3 and 5 have heteroduplex haplotype III where as lane 2 has II. The new variant, XXIX, is shown in lane 2. The Molecular Weight marker lane 7 forms 100bp ladder.



∝h: Lanes 1-6 are samples 826,827,846,847,848,849 and 849 respectively. Samples in lanes 1,3 and 4 have heteroduplex haplotype III where as those of lane 2,5,6 and 7 have II. 100 bp molecular weight marker ws used here.



SRY -1532: Lanes 2, 5,6 have G mutation and are sample No. 820, 828 and 829 respectively.
A → G mutation creates *Dra*III site.
Lanes 3,4 and 7 have ancestral A allele and are sample No. 822,823 and 830 respectively.
Lane 1 is a 50 bp Molecular Weight Marker.



92R7: Lanes 2-7 are samples 845-851 respectively. Lanes 2-4 have *Hind* III site (presence of a C allele).
10 bp molecular weight marker is shown in lane 1.

Polymorphism Scoring:

Polymorphisms were scored by indicating the polymorphic locus. The ancestral state is also indicated where it is known (as described earlier).

RESULTS:

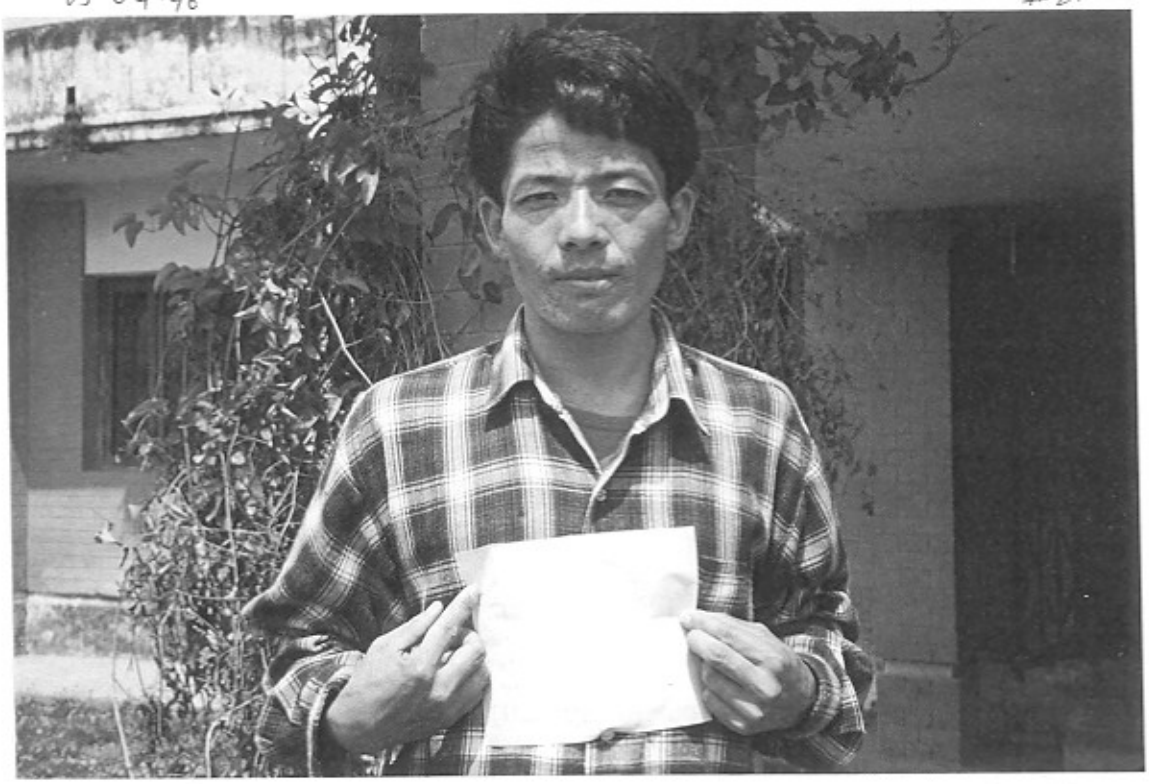
Samples were collected from as diverse a population as possible and included ones from different races/castes. A total of 36 samples were collected during my visit to Nepal in Easter 1996. This size would have increased if I had more time and didn't have to run after convincing the local Health Assistants to take blood from donors. Getting legal permission from the Nepalese government took a while. It wouldn't have been possible without the very generous support from Dr. Mathura Prasad Shrestha, Chairman, Nepal Health and Research Council (NHRC) and Dr. Madan Mani Dixit, Bir Hospital. A further two samples were collected from two other Nepalese at Oxford. A questionnaire for each donor was filed in; a sample questionnaire is shown on the next page. This was done especially to note the ancestral history of the donor.

Every effort was made to obtain good quality DNA during the extraction but this wasn't possible for all the samples. Only traces of DNA were obtained for some samples, while there was considerable DNA degradation for a few others. Luckily, this wasn't a problem for PCR amplification analysis. Data is incomplete for the samples whose DNA quality and quantity was poor and these samples have not been included in this analysis.

A total of 10 polymorphic markers, all PCRable, was chosen for this study. They fall into two groups:

QUESTIONAIRE FOR NEPALESE SAMPLES

05-04-96 #21 836



NAME: DIL BAHADUR LIMBU

AGE: 25

PLACE OF BIRTH TAPLEJUNG - FAR EAST

RESIDENCE: TAPLETNG

COMMUNITY: LIMBU

CASTE/TRIBE: MAINABU

LANGUAGE: LIMBU

DIALECT: -

PATERNAL ANCESTRY: Relation with the Kirati Dynasty which ruled Nepal (East-Nepal) ~1500 yr.

a. Group 1: These markers consist of pYAP and most point mutations on the Y and are called Unique Event Markers . The markers are: 92R7, 47Z, pYAP, sY81, Tat, DYS199, SRY-2627, SRY+465

b. Group 2: These markers describe 'Rare events': They include SRY-1532 and Alphoid heteroduplex.

Polymorphisms in thirty-eight Nepalese Y-Chromosomes:

All the thirty-eight Nepalese samples were analyzed at ten polymorphic loci. The results are shown in Table 1. Not all the samples could be scored. thirty-two samples could be scored fully at ten polymorphic loci. Based on the results from this scoring, five compound haplotypes could be constructed for the Nepalese population. It should be noted that out of ten polymorphic markers just three were polymorphic for the Nepalese population (Alphoid, 92R7, SRY-1532); the rest of the markers were non polymorphic. The realtionship between the five compound haplotypes is shown in Fig 2 and their basic composition is shown in table 2. The haplotypes are named as per Ref 1 with the inclusion of heteroduplex , e.g. 2(X) is a haplotype group two having heteroduplex variant X (3).

Haplotypes, based on the same ten polymorphic markers, were constructed in five other populations, and diversity based on Nei's method calculated. (Table 3). This diversity is also illustrated by means of a histogram (Fig 3). Basques and Amerindians have a low diversity. This is as expected from a small, isolated population. On the other hand Europeans, Mongolians and the Nepalese have a higer diversity. This is again as expected from a large, outbred population.

A Neighbour Joining tree was constructed based on genetic distances between various populations (Fig. 4). African, Algerian, Chinese, Indian and Sri_Lankan populations were also included for this analysis (data from ref. 1). On the tree Nepalese population lies between the Indian and the Sri_Lankan populations and

more broadly between the Europeans and the Mongolians. This is as expected genetically and on the basis of geographic location of the country. The Maximum Likelihood tree (not shown) looks very much similar to the Neighbour joining tree indicating the consistency of the results.

Sample	Amount of DNA	Heteroduplex Haplotype	92R7	SRY-1532	pYAP	47z	DYS199	SRY-2627	SRY+465	sY81	Tat
816	**		C	G	-	-	C		C	A	T
817	trace		T	A	-	?-	C		C	A	T
818	**	II	T	G	-	-	C	C	C	A	T
819	***	II	T	A	-	-	C	C	C	A	T
820	trace	III	C	G	-	-	C	C	C	A	T
821	trace	II	T	A	-	-	C	C	C	A	T
822	**	II	T	A	-	-	C	C	C	A	T
823	trace	II	T	A	-	-	C	C	C	A	T
824	trace		C	G	-	-	C	C	C	A	T
825	*	IV	C	G	-	-	C	C	C	A	T
826	**	III	C	G	-	-	C	C	C	A	T
827	trace	II	T	A	-	-	C	C	C	A	T
828	***	III	C	G	-	-	C	C	C	A	T
829	**	III	C	G	-	-	C	C	C	A	T
830	***	II	T	A	-	-	C	C	C	A	T
831	**	II	T	A	-	-	C	C	C	A	T
832	*	III	C	G	-	-	C	C	C	A	T
833	*	III	C	G	-	-	C	C	C	A	T
834	***	II	T	A	-	-	C	C	C	A	T
835	***	II	T	G	-	-	C	C	C	A	T
836	**	XXIX	C	G	-	-	C	C	C	A	T
837	**	III	C	G	-	-	C	C	C	A	T
838	*	III	C	G	-	-	C	C	C	A	T
839	***	III	C	G	-	-	C	C	C	A	T
840	trace	III	C	G	-	-	C	C	C	A	T
841	***	XXIX	C	G	-	-	C	C	C	A	T
842	***	II	T	G	-	-	C	C	C	A	T
843	**	II	T	A	-	-	C	C	C	A	T
844	trace	III	C	G	?-	-	C	C	C	A	T
845	**	III	C	G	-	-	C	C	C	A	T
846	***	III	C	G	-	-	C	C	C	A	T
847	*	III	C	G	-	-	C	C	C	A	T
848	***	II	T	A	-	-	C	C	C	A	T
849	**	II	T	A	-	-	C	C	C	A	T
850	***	II	T	A	-	-	C	C	C	A	T
851	***	II	T	A	-	-	C	C	C	A	T
852	***	II	T	A							
853	***	III	C	G							

Table 1

SYMBOLS	MEANING
*	Quality of DNA; *** as the best quality DNA
II - XXIX	Heteroduplex Haplotype
T, A, C, G	Allele Type
-	absence of Alu Insertion for pYAP and absence of *StuI* site for 47z systems
?-	results uncertain

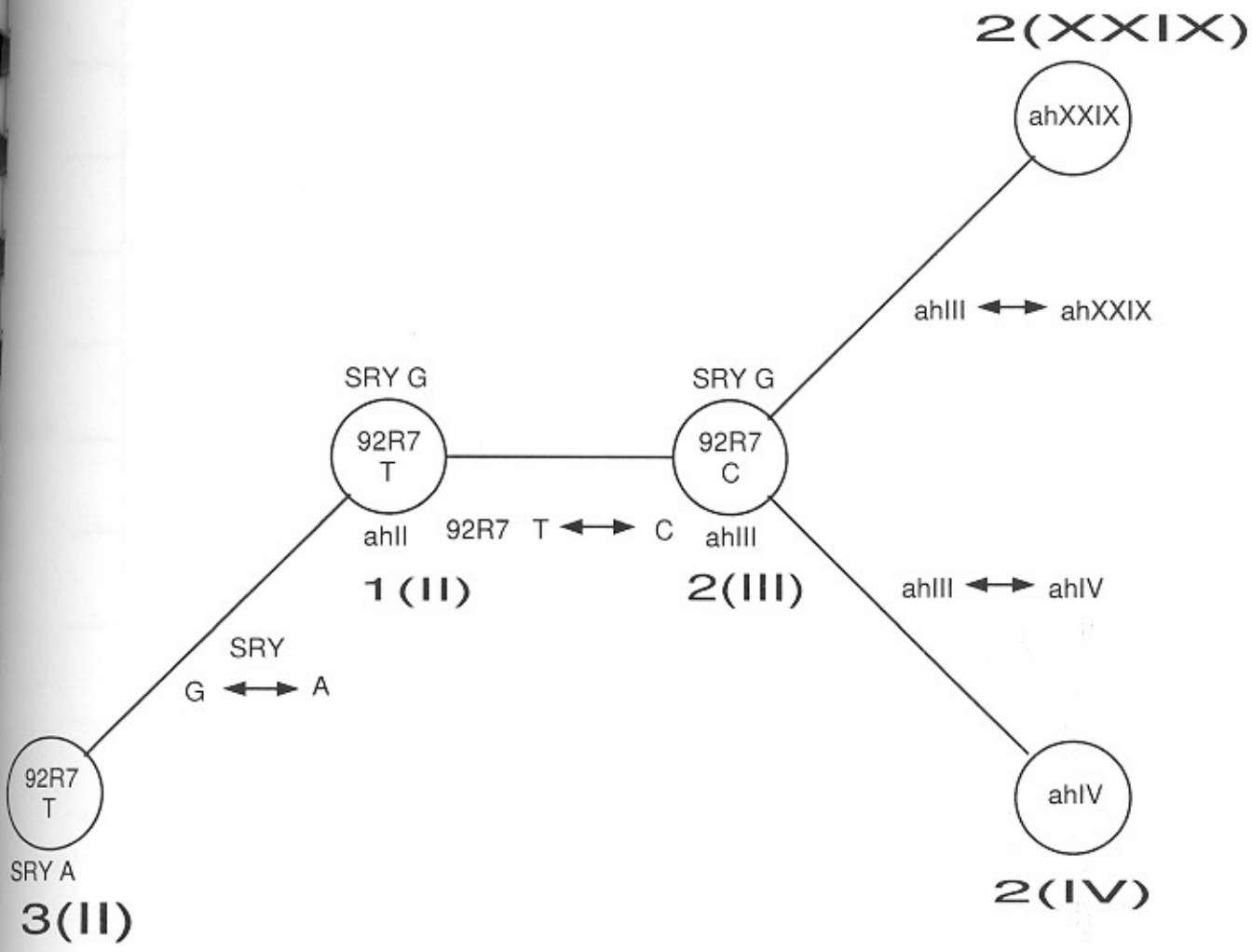
Some polymorphic locus could not be amplified for various samples and some amplifications were dubious (where I have marked with a question mark). For these samples , the amplification will be carried out again.

Compound	92R7	SRY-1532
Haplotype		
2(III)	C	G
1(II)	T	G
3(II)	T	A
2(IV)	C	G
2(XXIV)	C	G

Table 2

Compound Haplotype	Europeans	Basques	Amerindians	Mongolians	Nepalese	*Total/Group*
1 (I)	0	0	0	1	0	1
1 (II)	20	19	6	1	3	49
1 (III)	3	0	0	0	0	3
1 (VI)	1	0	0	0	0	1
2 (I)	0	0	0	2	0	2
2 (II)	1	0	0	2	0	3
2 (III)	9	2	0	6	14	31
2 (IV)	0	0	0	1	1	2
2 (V)	0	0	0	2	0	2
2 (XVIII)	0	0	1	20	0	21
2 (XIX)	0	0	0	1	0	1
2 (XX)	0	0	0	5	0	5
2 (XXIX)	0	0	0	0	2	2
3 (II)	2	0	0	4	12	18
3 (III)	1	0	0	0	0	1
Total/Popln.	37	21	7	45	32	142
Nei's Diversity	0.64	0.17	0.24	0.76	0.65	0.79

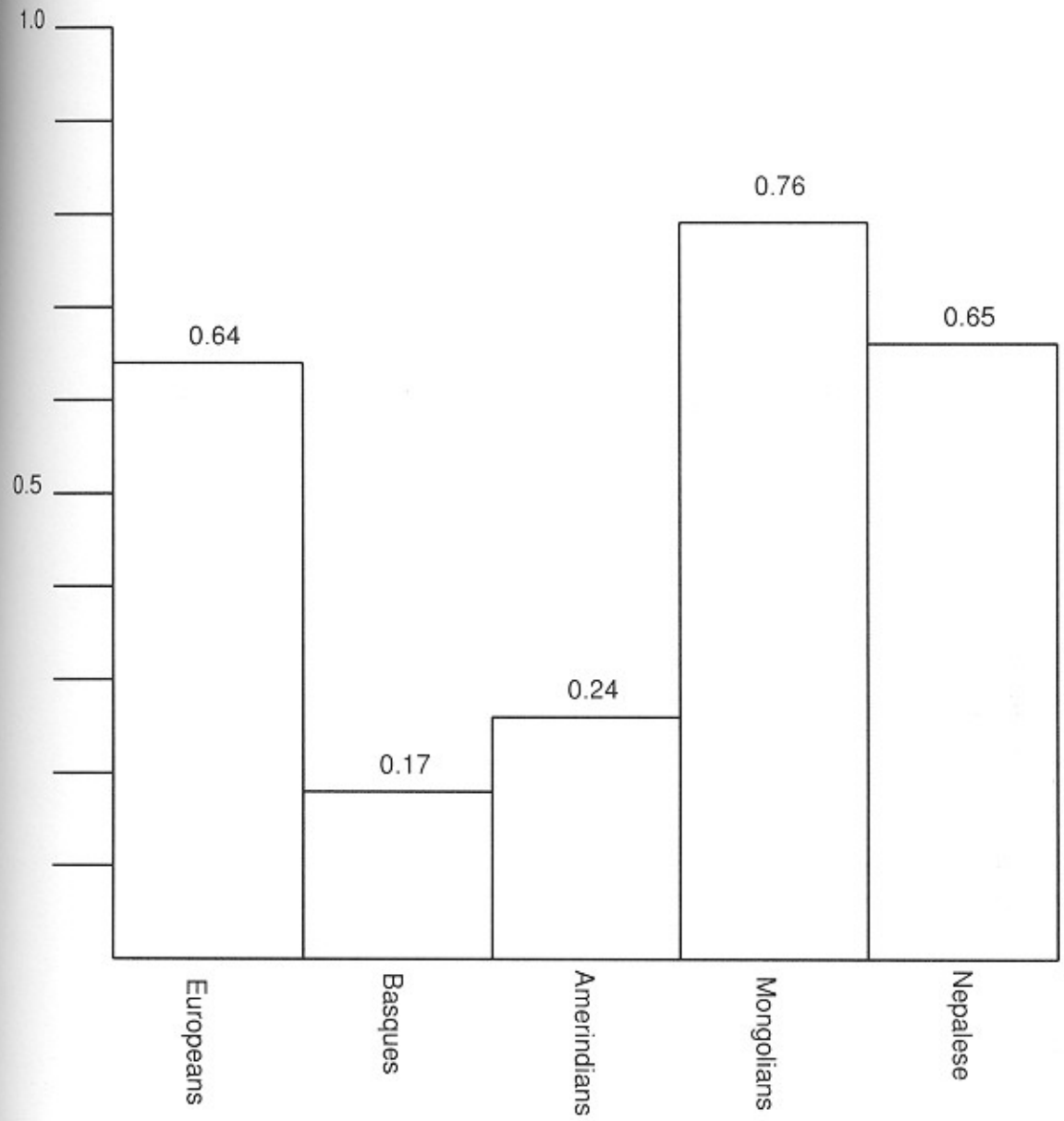
Table 3



A simple Y Chromosome tree showing relationship between compound haplotypes.

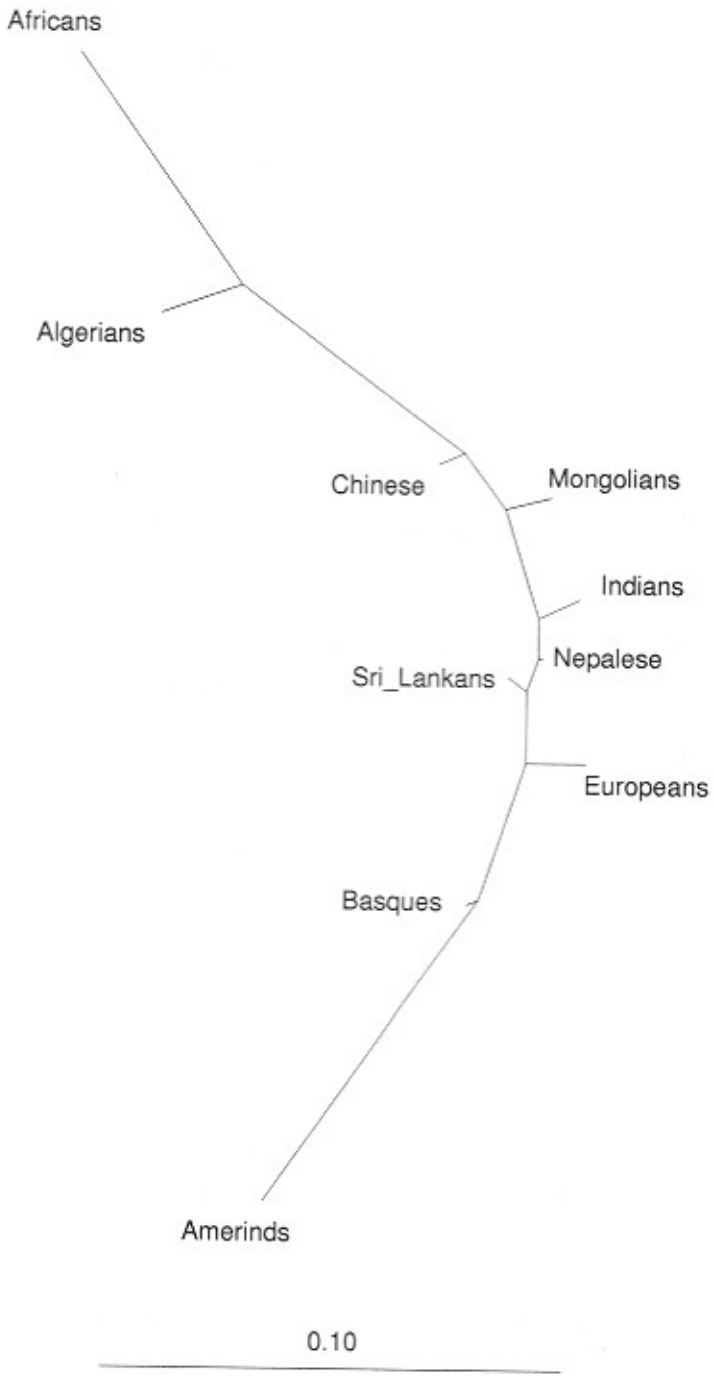
Polymorphic loci are shown between nodes.

Fig. 2



A histogram showing Y haplotype diversity within five populations calculated using Nei's method

Fig 3



An unrooted Neighbour Joining tree of 10 populations based on genetic distances according to Saitov and Nei's method.

Fig. 4

DISCUSSION

There are various points raised by this kind of analysis. I will consider the points and suggest possible solutions in turn.

Sampling

For a good survey, it is very important to obtain samples which are representative of a particular area concerned. My samples, in true sense, do not represent Nepal to the full extent. Most of my samples are from Kathmandu, the central region of the Kingdom. Because of lack of time and other problems I couldn't go to different parts of the country to get the blood samples. Nepal is also a land of many tribes. 'Rautes' and 'Chepangs' still live in the forests and live by hunting. I couldn't get samples from any of these tribes. There are still other strange tribes existing in different parts of the country. The caste system still has a strong impact in the Nepalese society and different castes have different ancestral background. I couldn't manage to get samples representative of all the castes. For future work, sampling of a large number of Nepalese population will produce more interesting results and will be more meaningful.

Data from (1) I used are more representative of a single population but are not fully representative. The sample size varied from a single Saudi Arabian, 3 Australian aborigines to 52 Indians and 65 Mongolians and sub-Saharan African samples from several areas of the continent had to be combined

There is a degree to which a marker search can be biased towards a particular population. A marker which is found only in one or two populations (eg. 47z is biased for the Japanese population) is not very useful for analyzing other populations and can lead to rather uninformative haplotype.

Choice of markers is not a big problem now. In this study, all suitable markers were used but as new markers are being discovered at quite a high rate, this means we will be having a lot more markers to use. At this time, a choice will

have to be made. Consideration of particular place, linguistic values, age and distribution of the marker and and any genetic information available about the population may be used as a guide line to select more useful markers.

Construction of Compound haplotypes:

With the polymorphisms obtained, five compound haplotypes were constructed for Nepalese population. Data from ref.1 were used to construct similar haplotypes for the other world population which include Europe, Asia, Africa, America and Oceania. Different populations often have their own unique haplotypes which are different from the rest of the world population. It is seen from the figures in the earlier pages that different haplotypic groups are confined to particular geographical regions.

The Maximum likelihood tree and the Neighbor Joining tree look the same suggesting the reliability of the data. One has to have in mind that these trees were constructed with the help of two different programs based on different principles.

Diversity

Using these markers, diversity values from 0.17 to 0.76 were found for different populations (Table 3). Nepalese population had Nei's diversity value of 0.65. The population comprises of five different compound haplotypes with most of the population falling into either group 1or 2. The two Kiranti samples (836, 841) show a completely new αoid heteroduplex . This will be given the name XXIX (3).

Implications for Nepalese Population History

The unrooted Neighbour Joining tree very nicely places the Nepalese population close and in between the Indian and the Sri_Lankan populations and more broadly between the Europeans and the Mongolians. This is what we expect considering genetics as well as the geographic location of the country.

It was also found that the two Kirants analyzed had a new heteroduplex haplotype (designated XXIX). B. J. Harsat in his book- History of Nepal- has mentioned that the Kirants might have come from Afaghanistan, their ancestors being the derivatives of the people who lived in Babylon. He also mentions the archeological evidence in support of his statement. It could be that our general belief about the Kirants being Mongolian derivatives (also as suggested by quite a lot of historians) could be wrong. This analysis could go further to provide a marker which could be useful in the search of this peculiar race.

We have not done genetic analysis of the existing many tribes within Nepal. They might prove to be even more interesting in terms of their origin and ancestral migrations.

Nepal, although, geographically a small country, has diverse population types. Genetic analysis of this diversity would require more samples from throughout the country. I have to indicate that this analysis has been biased towards the Aryan race as more blood samples were obtained from people who belong to this race. In any case, such genetic analysis should help anthropologists, archeologists and historians to study peopling in Nepal and track their ancestral routes.

The general Nepalese society is based on caste system . It could be interesting to analyze samples on the basis of different caste system to see if they differ in the Y- haplotype distribution. The language differs from race to race and from tribe to tribe. Linguistic and Genetic relationship has also been useful in the study of

human population relationships. We can see how this might vary within Nepal and the rest of the world.

With more and more markers available, this field looks exciting as ever. More ideal and PCRable markers would help analyze polymorphisms quickly and produce more reliable results. Scientists are working on the various ways of presenting a set of data in order to study population relationships and evolution in general. This would add to more accurate conclusions being made than what we can make now. Cooperating with Archeologists, Paleontologists and historians will enable geneticists to achieve better and reliable results.

REFERENCES:

1. Pandya, A. (1996) Human Population Relationships deduced from Y Chromosomal DNA Polymorphisms. Change of Status Report, Oxford University.
2. Jobling, M. A. and Tyler-Smith, C. (1995) Fathers and Sons: The Y-Chromosome and the human Evolution.*Trends Genet.* **11:** 449-456
3. Santos, F. R. *et al.* (1995) PCR haplotypes for the human Y chromosome based on αoid satellite DNA variants and heteroduplex analysis. *Gene* **165:** 191-198
4. Ellis, N. (1991). The Human Y Chromosome. *Semin.Deo.Biol.* **2:** 231-240
5. Wolpoff, M.H. (1989) Multiregional Evolution: the fossil alternative to Eden. In: The human revolution; Behavioural and Biological perspectives on the origin of Modern Humans. Edinburgh University Press, Edinburgh. 62-108.
6. Stringer, C.B. (1990). The emergence of Modern Humans. *Scientific American:* 68-74
7. Stringer, C. B. and McKie, R. (1996) African Exodus. *The Origins of Modern Humanity* (Random House (UK) Ltd. London).
8. Lahr, M. M. and Foley, R. (1994). Multiple dispersals and modern human origins. *Evol. Anthropol.* **3**: 48-60
9. Jones, S., Martin, R. and Pilbeam, D. eds. (1992) *The Cambridge Encyclopedia of Human Evolution* (Cambridge University Press).
10. Cavali-Sforja, L. L., Menozzi, P. and Piazza, A. (1994) *The History and Geography of Human Genes* (Princeton University Press).
11. Fagan, B. M. (1995) *People of the Earth* (Harper-Collins, New York).
12. Cann, R. L., Stoneking, M. and Wilson, A. C. (1987)Mitochondrial DNA and Human Evolution. *Nature.* **325:** 31-36
13. Santos, F. R. *et al.* (1996) Worldwide Distribution of Human Y Chromosome Haplotype. *Genome Research.*
14. Mathias, N., Bayes, M. and Tyler-Smith, C. (1994) Highly informative compound Haplotypes for the Human Y chromosome

15. Whitfield, L.S., Sulston, J.E. and Goodfellow, P. N. (1995) Sequence variation of the Human Y- Chromosome. *Nature* **378**: 379-380.

16. Hammer, M. F. (1995) A recent common ancestry for human Y chromosomes. *Nature* **378:** 376-378

17. Hammer, M. F., and Horai, S. (1995) Y chromosomal DNA variation and the peopling of Japan. . *Am. J. Hum. Genet.* **56:** 951-962

18. Nakahori, Y. *et al.* (1989) Two 47z [DXYS5] RFLPs on the X and Y chromosome *Nucleic Acids Res.* **17:** 2152

29. Nakagome, Y. *et al.* (1992) A Y-associated allele may be characteristic of certain ethnic groups in Asia. *Ann. Hum. Genet.* **56:** 311-314

20. Lin, S. J. *et al.* (1994) A Y-associated allele is shared among a few ethnic groups of Asia. *Jpn. J. Hum. Genet* **39:** 299-304

21. Underhill, P. A. *et al.* (1996) A pre-Columbian Y Chromosome-specific translation and its implications for Human Evolutionary History. *Proc. Natl. Acad. Sci. USA* **93**:196-200

22. Bianchi, N. O. *et al.* Submitted

23. Shinka *et al.* Submitted

24. Zerjal, T. *et al.* in preparation

25. Seielsted *et.al.* published in 1994.

26. Harsat, B. J.(1970) *History of Nepal-as told by its own and contemporary chronicles* . VV Research Institute Press: pXXIV-XXVII

27. Anand, R. (1986) Pulsed Field Gel Electrophoresis : A technique for fractionating large molecules. *Trends Genet.* **2:** 278-283

28. Santos, F. R. et al. (1993) *Hum. Genet.* **90:** 655- 656

29. Brookfield, J. F.Y. (1994) A new Molecular view of Human Origins. *Curr. Biol.***4:** 651-652 .