ANTHROPOLOGICAL SCIENCE
Vol. 121(1), 75–83, 2013

Brief Communication

Y-chromosome haplotypes and clan structure of the Sherpa of the Solukhumbu (Nepal): preliminary ethnogenetic considerations

Irene AMORUSO[1]*, Luciana CAENAZZO[2], Silvano PRESCIUTTINI[3], Elisa GREGIANIN[4], Andrea BOZZATO[4], Gianumberto CARAVELLO[1]

[1]*University of Padua, Department of Molecular Medicine—Public Health Unit, Via Loredan 18, 35131 Padova, Italy*
[2]*University of Padua, Deptartment of Molecular Medicine—Legal Medicine Unit, Via Giustiniani 2, 35131 Padova, Italy*
[3]*University of Pisa, Department of Physiological Sciences, SS 12 Abetone e Brennero 2, 56127 Pisa, Italy*
[4]*University of Padua, Department of Biology, Via Bassi 58/B, 35131 Padova, Italy*

Received 14 October 2011; accepted 22 November 2012

Abstract The present-day Sherpa are thought to descend from a small number of ancestors that settled in Nepal several centuries ago, coming from the Eastern Tibetan region of Kham. A generally accepted ethnographic theory involves the out-of-Kham migration of four proto-clans between the 15th and 16th centuries. Traditional Sherpa society is still divided into clans, called ru, which are patrilineary transmitted. Ru therefore roughly correspond to the surnames of Western societies: males of the same ru are expected to share identical Y-chromosome haplotypes. However, multiple origins of ru and/or frequent gene flow of male lineages from neighbouring populations can complicate the genealogical structure. In the present work, 25 male Sherpas of the Solukhumbu district were typed for the 17 Y-chromosomal short tandem repeats included in the AmpF*l*STR® Yfiler™ kit. Seventeen different haplotypes were found; 12 were unique. A phylogenetic tree was then drawn from the pairwise mutational distance matrix with a neighbour-joining algorithm. Branching reliability was also assessed through bootstrap analysis. Two macro-clusters of haplotypes were found, ascribable on the whole to two out of four of the presumed Tibetan proto-ru, the Thimmi and the Minyagpa. However, the Minyagpa macro-cluster was found to be bipartite in terms of haplogroups, being composed by two distinct haplotype clusters. Clustering of the contributors by birthplace was also performed, suggesting a differential ru spatial distribution between upper Khumbu and lower Solu. Khumbu seems predominantly populated by newer clans and putative descendants of the Thimmi proto-ru, whereas Solu is mostly inhabited by members of the Minyagpa proto-ru.

Key words: Sherpa, clan structure, Solukhumbu, Y-STRs, AmpF*l*STR Yfiler

Introduction

The Sherpa are a Tibeto-Burman ethnic group from the Nepalese Himalayas, best known for their exceptional mountaineering skills as porters and guides. From the genetic point of view, Sherpa constitute a population isolate. Likely causes of isolation (Arcos-Burgos and Muenke, 2002) are ethnic, cultural, linguistic, religious, and geographical confinement. Sherpa DNA has been studied in detail in order to investigate the potential association between genes and adaptation to high altitude (Malacrida et al., 2007; Dromo et al., 2008). Sherpa villages and communities are scattered over an extremely wide area, ranging from the Helambu region, 72 km north-east of Kathmandu, to the Darjeeling district, in West Bengal. The core of the Sherpa's earliest settlement area lies in the neighbouring regions of Solu and Khumbu (Von Fürer-Haimendorf, 1964), located in the north-eastern Nepalese administrative districts of Solukhumbu and Okhaldhunga (Figure 1).

Traditional Sherpa society is divided into patrilineal clans, although the number of clans is debated: it ranges from 12 to 24 (Table 1), depending on the author (Von Fürer-Haimendorf, 1964; Oppitz, 1968; Krämer and Sherpa, 2002; Wangmo, 2005). The Sherpa word for clan is *ru*, which means 'bones,' since Sherpas actually believe that bones pass from father to son, thus establishing the natural propensities of a man. This pseudo-genetic paternal inheritance was considered superior to that of the mother, whose contribution is only transmitting the flesh, or *sha* (Wangmo, 2005). Ru are neither castes nor social classes, but rather wide familial clans with common roots in a certain village or in a common, and often legendary, ancestor (Fantin, 1973; Childs, 1997). After all, as in every human society, the addition of a heritable element facilitates identification and also marks lineages, providing a label of membership (King and Jobling, 2009). Sherpa can overall be considered as an endogamic ethnic group, yet exogamy between ru is compulsory: marrying within the same ru is forbidden and considered incestuous (Presciuttini et al., 2010).

* Correspondence to: Irene Amoruso, University of Padua, Department of Molecular Medicine—Public Health Unit, Via Loredan 18, 35131 Padova, Italy.
E-mail: irene.amoruso@gmail.com

Published online 7 February 2013
in J-STAGE (www.jstage.jst.go.jp) DOI: 10.1537/ase.121122

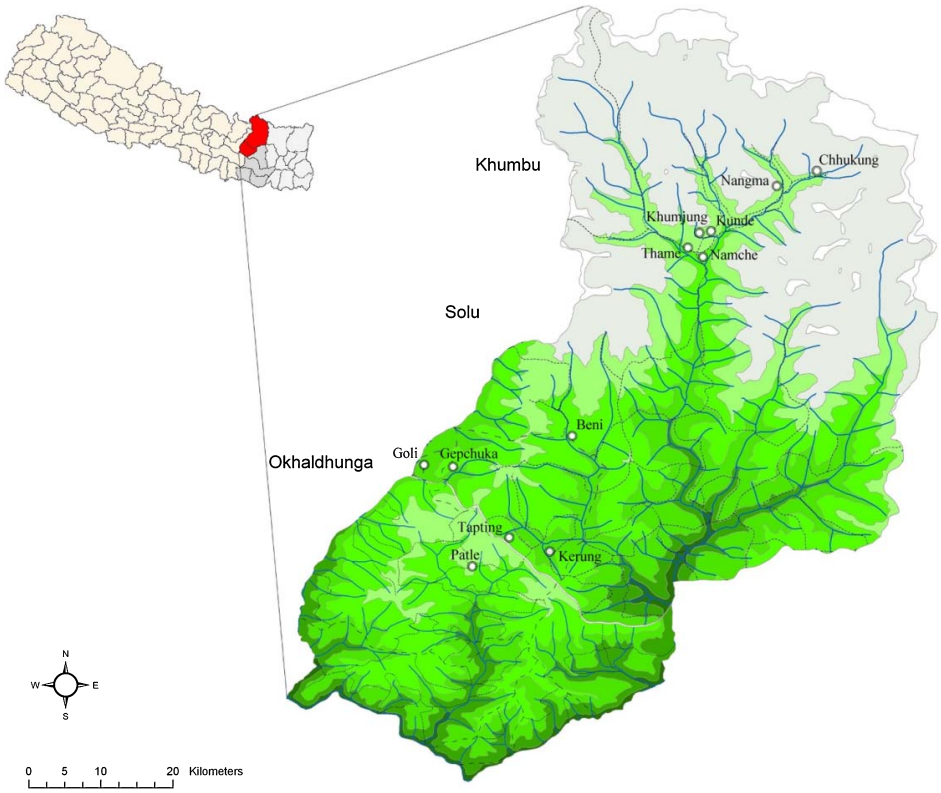

Figure 1. Sherpa main settlement area coincides with Solukhumbu and Okhaldunga Administrative Districts, in North-eastern Nepal. (Adapted from a GIS map by Cartography and Survey Engineering Consultant Kabindra Joshi, available at www.digitalhimalaya.com)

Sherpa tradition tells that all of the present Sherpa ru descend from some ancient Tibetan clans that came to Nepal from Eastern Tibet about 500 years ago. Many alternative stories about the very first arrival of the Sherpa to the Nepalese valleys exist, and most of these should probably be regarded as mere legends (Brower, 1991). There is not even agreement about the route of migration from Tibet to Solukhumbu, nor the exact Nepal entrance point (Von Fürer-Haimendorf, 1964).

Nevertheless, most Sherpa tales trace back the coming of their ancestors from the Salmo Gang (*Zal mo sgang*) area of Kham, in Eastern Tibet, during the 15th–16th centuries (Brower, 1991). The Salmo Gang area is presently part of the modern Ganzi Tibetan Autonomous Prefecture of the Chinese Sichuan Province. Significantly, the word *sherpa* itself is composed of two Tibetan words (*shar*, east and *pa*, person), together meaning 'Easterners,' and could thus be reflecting this migration (Stevens, 1993). The pioneering work of Oppitz (1968), derived from the analysis of a few written clan ancestry records retrieved in Solu, provides firmer historical evidence for the Kham homeland hypothesis, yet still has to face the dubious historical authenticity of the documents (Wangmo, 2005). On the whole, authors usually agree the broad terms of Oppitz's reconstruction (Von Fürer-Haimendorf, 1964; Brower, 1991; Krämer and Sherpa, 2002; Wangmo, 2005), which involves the out-of-Kham migration of four proto-ru at the turn of the 15th to the 16th century (Oppitz, 1968).

The four proto-ru were the *Lama-serwa* (or simply *Lama*), the *Chakpa* (or *Chiawa*), the *Minyagpa*, and the *Thimmi*. Probably, each proto-ru gave birth to several subclans through time, in accordance with the Sherpa assertion of their Tibetan ancestry. Indeed, Sherpa regard some of the present ru as brother-clans: this is probably due to their common descent from the same proto-ru. The supposed genealogy of present Sherpa ru and their relationship with the four proto-clans can be summarized based on a comparison of the works of the four main works (Von Fürer-Haimendorf, 1964; Oppitz, 1968; Krämer and Sherpa, 2002; Wangmo, 2005) that have already dealt with this issue (Table 1). An exact genealogy of modern ru cannot be traced for certain, since written sources are scarce and incomplete and oral tradition sometimes reports contrasting theories. Unfortunately, the degree of complexity has been increased over the years by the later arrival of other groups of Tibetan immigrants, particularly from the bordering Tingri area, after the first migration. The members of these more recent groups are also organized in patrilinear ru and are now considered as pure Khumbu Sherpa, even if none of them came from Kham nor had any kind of written tradition concerning its genealogy (Sherpa, 2007).

Except for its two pseudoautosomal regions found at the telomeric ends, the Y chromosome is exclusively inherited from father to son as a entire haplotype of physically linked loci. The portion of the chromosome which does not undergo homologous recombination is called the non-

Table 1. Lists of Sherpa ru and their presumed genealogy

Ref.	von Führer-Haimendorf (1964)	Oppitz (1968)	Krämer and Sherpa (2002)	Wangmo (2005)
No. of listed ru	22	24	12	20
Proto-clan	Minyakpa	Minyakpa	Minyakpa	Minyagpa
Subclans	Gole	Gole	Golela	Goleg
	Pinasa	Binasa	Binasa	Benasa
	Thaktu	Trakto	Takto	Tragtho
	Gardza	Gardza		Gartsa
	Pankarma	Pankarma	Pankarma	
	Shire	Shire	Shire	
		Yulgongma	Yulgongma	Yulgongma
		Kapa		
Proto-clan	Thimmi	Thimmi	Thimi	Thimmipa
Subclans	Salaka	Salaka	Salaka	Zalaka
	Paldorje	Paldorje	Paldorje	Paldorje
	Goparma	Gobarma		Gobarma
	Khambadze	Khambadze	Khampaje	Khampache
	Lakshindu	Lakshindo	Lakshindo	Labushingtog
Proto-clan	Lama	Serwa	Lamaserwa	Lama Serwa
Subclans	—	Lama	—	—
Proto-clan	Chiawa	Chakpa	Chakpa	Chagpa
Subclans	—	—	—	—
Recent clans	(no genealogy available)			
	Chuserwa	Chuserwa	—	Chuserwa
	Jongdomba	Jungdomba	—	Jongdongpa
	Lhukpa	Lhukpa	—	Lhugpa
	Mende	Mende	—	Mendewa
	Munming	Murmin	—	Murmin Tso
	Nawa	Nawa	—	Nawa
	Shangup	Shangup	—	Shangkhug
	Sherwa	Sherwa	—	Sherwa

The table lists the genealogical structure of present Sherpa ru. The four proto-clans (Minyagpa, Thimmi, Lamaserwa, and Chakpa) are followed by their respective subclans. Genealogies proposed by the four main works that have already dealt with the Sherpa ru history are given for comparison.

recombining region of the Y (NRY) and comprises male-specific genes only, known as holandric genes. The recent characterization of many new Y-chromosome markers has contributed to increasing their usefulness for studies of genetic anthropology, which, until recently, lagged far behind mitochondrial DNA and maternal lineage studies (Crawford, 2007). DNA polymorphisms of the NRY have already proved to be a useful tool for tracing ancient patrilinear lineages (Jobling and Tyler-Smith, 2003; Kwak et al., 2005; King and Jobling, 2009): all patrilinear relatives share the same NRY haplotype and thus inheritance, with absence of recombination, leads to the perpetration of polymorphisms through a simple paternal transmission pathway (Ljubković et al., 2008). The most widely used kind of NRY polymorphisms for phylogeny of recently diverged paternal lineages are the short tandem repeats (STRs). Y-STRs are highly polymorphic markers and thanks to a high level of variability due to variation in repeat number they can reveal further variation within haplogroups (Kwak et al., 2005). Isonymous individuals are expected to carry the same Y haplotype: observed incongruity can be ascribed to new mutational events, adoptions, illegitimacy, or polyphyletic origin of the surname.

In this work, the 17 AmpF*l*STR Yfiler polymorphic loci (Mulero et al., 2006) were amplified in a sample of 25 male Solukhumbu Sherpa. Haplogroups were also predicted and the emerging genetic tree was compared with ethnogenealogical hypotheses.

In other words, the present research is aimed at a preliminary definition of the true kinship and clan structure of the Sherpa population. The approach followed is that of establishing any existing correspondence between the Y-STR haplotypes and traditional patrilineal Sherpa clans.

Materials and Methods

Samples

Buccal swab samples were collected from 25 male volunteers with Whatman® OmniSwabs. Each ejectable head was put in a 1.5 ml sterile extraction tube. The sampling procedure was approved by the Social and Ethics Committee of the Department of Environmental Medicine and Public Health (EMPH), Padua University.

16 samples were collected in spring 2008, during a meeting of the Sherpa community held in Namche Bazaar, the main village of Khumbu Valley, and 9 samples were collect-

ed in spring 2009 in the villages of Namche, Kunde, and Khumjung. All individuals freely donated their saliva, after signing an informed consent form that explained the purpose of DNA analysis in the present research. The volunteers were questioned about their ru affiliation, age, birthplace, and dwelling place. All of the contributors reported as being unrelated to one another and those sharing the same ru were actually unaware of any kind of biological relationship between them.

Samples were packaged and transported to Padua (Italy) within 7 days and processed in the Forensic Genetics Laboratory of the Department of Environmental Medicine and Public Health (Legal Medicine Unit), Padua University. The laboratory achieved quality control certification by the Y chromosome Haplotype Reference Database (YHRD) on 23 November 2000 (YHRD c.n. YC000052).

DNA extraction

DNA extraction from buccal swabs was performed using the QIAGEN QIAmp® DNA Microkit according to the manufacturer's instructions (Qiagen, 2010). About 30–50 μg of DNA were recovered in 30 μl of final solution.

Multiplex PCR

Multiplex polymerase chain reaction (PCR) was performed on 0.5–1.0 ng of target DNA for each sample using the AmpF*l*STR® Yfiler™ PCR Amplification Kit, in accordance with the protocols described in its User's Manual (Applied Biosystems, 2006). The AmpF*l*STR® Yfiler™ Kit amplifies 17 Y-STR loci: they consist of the 9-marker European minimal haplotype (minHt) (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, and DYS393), the 11-marker Scientific Working Group DNA Analysis Methods (SWGDAM) recommended Y-STR panel (minHt plus DYS438 and DYS439) and the additional highly polymorphic loci DYS437, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4.1. Allele nomenclature follows ISFG recommendations (Gusmao et al., 2006). The locus DYS385 was excluded from further analysis, as its two twin loci show indistinguishable alleles.

PCR reaction was performed in a GeneAmp® PCR System 9700 following the manufacturer's instructions (Applied Biosystems, 2010) except for minimal modifications of the final reaction volume, which was 12.5 μl (9.5 μl of master mix + 3.0 μl of DNA solution) instead of 25.0 μl. In addition, six supplementary amplification cycles were performed for some samples. PCR thermocycling parameters were the following: initial hot-start incubation step at 95°C for 11 min, 30 amplification cycles (denaturing at 94°C for 1 min, annealing at 61°C for 1 min, and extension at 72°C for 1 min) and final extension at 60°C for 80 min.

Electrophoresis

Amplified STRs were separated by capillary electrophoresis using an ABI Prism® 3130 Genetic Analyzer (Applied Biosystems). GeneScan-500 Internal Lane Size Standard LIZ-500 was the internal standard of choice. The size of the PCR products was classified with GeneMapper v.2 software, with the Yfiler Allelic Ladder (Applied Biosystems) as comparison.

Data analysis

The genetic distance for each pair of haplotypes was calculated as the sum, over all loci, of the absolute difference in the number of STRs. Genetic distances were then reported in a pairwise distance matrix. Further data analysis was performed with Arlequin v. 3.5.1.2 software (Excoffier and Lischer, 2010). Allelic frequencies, allelic ranges, haplotype diversity (Nei, 1987), and gene diversity per locus (*h*) were also calculated. A phylogenetic tree was extrapolated from the pairwise distance matrix, with a neighbour-joining (NJ) algorithm. Statistical reliability of branching nodes was assessed through bootstrap (Felsenstein, 1985), performing 100 repetitions. The tree was finally drawn with the TreeGen tool (CBRG, 2012). Multi-copy marker DYS385 was excluded from the phylogenetic analysis, as its two loci, DYS385a and DYS385b, were indistinguishable from each other and thus the result from these was not informative.

After haplotype typology, haplogroup was predicted for each sample with the free Haplogroup Predictor tool v. 21 (Athey, 2007). The software identifies the most probable haplogroup, assigning a prediction confidence score expressed as a percentage. Haplogroups obtaining the highest score were chosen and will subsequently be reported.

Results and Discussion

First, statistics of general genetic interest for the Sherpa population were calculated. 17 different haplotypes were found out of 25 samples: 5 of these were shared and 12 were unique. The allelic profiles of the haplotypes are listed in full in Table 2: neither null alleles nor duplicated loci were found. DYS448 showed the 20.2 microvariant in samples M1 and M12. Allelic ranges and frequencies, along with gene diversity, are also schematically reported (Table 3).

Prediction of haplogroups identified three main groups (Table 2): 9 individuals belong to haplogroup Q (M10, M11, M16, M17, M18, M20, M21, M24, M29), 6 to haplogroup E1b, subclade E1b1 (M4, M6, M13, M23, M26, M27), 9 to haplogroup R1 (M1, M3, M12, M14, M15, M19, M25 are R1a; M8, M28 are R1b; M19 had an equal score between R1a and R1b). Individual M22 alone belongs to haplogroup H. Probability scores for each predicted haplogroup are reported in Table 2 as well. Overall haplotype diversity for the Sherpa population assumes a value of 0.9633 ± 0.0208. Haplotype diversities previously reported for a mixed Nepalese and two Tibetan population samples are respectively 0.9970 (Parkin et al., 2007), 0.9998 (Tian-Xiao et al., 2009) and 0.9981 (Gayden et al., 2010). Since haplotype diversity of Sherpas is not so different from neighbouring populations, it can be supposed that their ancestors had substantial diversity in their male lineages.

The NJ tree (Figure 2) represents the presumable kinship of the 25 typed Sherpa. In addition, the tree reports the stated ru kinship of each contributor, the proto-ru and the predicted haplogroup. Unfortunately, contributors M1 and M8 simply asserted themselves to be 'Sherpa' and of 'Tibetan origin', respectively. The topology of the phylogenetic tree clearly presents two main macro-clusters of haplotypes, branched with a bootstrap value of 100 (Figure 2). The first macrocluster results further partitioned into two minor haplotype

Table 2. The 17 amplified haplotypes and their extended profiles

Haplotype	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Sample id	M1 M12	M3	M4 M23 M26	M6	M8	M10	M11	M13	M14 M15 M25	M16 M20	M17 M18 M24	M19	M21	M22	M27	M28	M29
Absolute frequency	2	1	3	1	1	1	1	1	3	2	3	1	1	1	1	1	1
Relative frequency	0.08	0.04	0.12	0.04	0.04	0.04	0.04	0.04	0.12	0.08	0.12	0.04	0.04	0.04	0.04	0.04	0.04
DYS19	15	15	14	14	15	14	14	14	15	14	14	16	14	16	14	15	14
DYS389-I	13	13	10	10	14	12	12	10	13	12	12	14	12	12	10	13	12
DYS389-II	29	29	27	27	30	28	28	27	29	28	28	30	30	29	28	27	28
DYS390	23	25	23	23	23	23	23	23	25	23	23	23	23	27	23	23	23
DYS391	10	11	10	10	10	10	10	10	10	10	10	10	10	10	10	11	10
DYS392	11	7	11	11	7	14	14	11	7	14	14	7	14	10	11	7	14
DYS393	13	13	14	14	13	13	12	14	13	13	12	13	12	12	14	13	13
DYS437	14	14	14	14	14	16	15	14	14	16	14	14	15	14	14	14	15
DYS438	11	11	10	10	11	11	11	10	11	11	11	11	11	10	10	11	11
DYS439	12	12	11	12	12	13	12	12	12	12	12	12	12	11	11	12	12
DYS448	20.2	19	19	19	18	21	20	19	19	21	20	18	21	18	19	19	20
DYS456	16	16	14	14	17	15	15	13	16	15	15	16	15	15	14	16	15
DYS458	16	16	15	15	17	17	18	15	16	17	18	17	17	15	15	16	18
DYS635	21	21	21	21	21	21	20	20	21	21	20	21	20	20	21	21	20
YGATAH4	11	11	11	11	11	12	11	11	11	12	12	11	12	11	11	11	12
DYS385	13–15	11–11	15–16	15–16	11–11	14–18	13–19	15–16	11–11	14–18	13–19	11–12	18–19	17–17	15–16	11–11	13–19
Haplogroup	R1a	R1a	E1b1	E1b1	R1b	Q	Q	E1b1	R1a	Q	Q	R1a/R1b	Q	H	E1b1	R1b	Q
Prediction % score	42.0	98.9	84.8	96.7	90.3	94.8	95.9	99.3	98.4	94.0	52.8	32.0/30.0	99.2	99.3	98.0	91.1	92.7

Table 3. Allelic frequencies, allelic range, gene diversity per locus and G-W index at the 17 Y-STR loci

DYS	19	389-I	389-II	390	391	392	393	437	438	439	448	456	458	635	GATAH4.1	385	Mean
No. of alleles	3	4	4	3	2	4	3	3	2	3	4	5	4	2	2	8	3.5
Allelic range	2	4	3	4	1	7	2	2	1	2	3	4	3	1	1	8	3.1
Gene diversity *h*	0.553	0.727	0.750	0.347	0.153	0.717	0.640	0.410	0.420	0.397	0.717	0.723	0.777	0.453	0.453	0.853	0.568

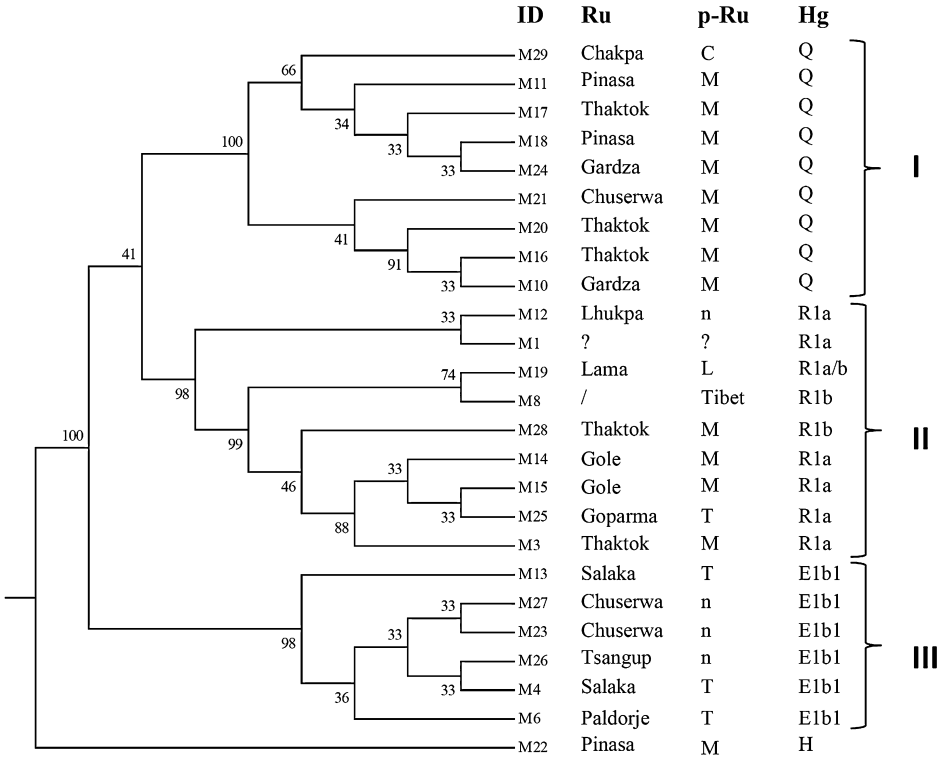

Figure 2. Neighbour-joining tree (NJ) for the Sherpa population sample. For each of the 25 Sherpa contributors, self-stated ru affiliation has been listed along with the respective ancestral proto-clan. Abbreviations for the four proto-ru used in the phenogram are: Lama (L), Chakpa (C), Minyagpa (M), and Thimmi (T). Lower case 'n' stands for 'new' and represents recent Sherpa clans that are supposed to have reached Solukhumbu after the pristine out-of-Kham migration. The three main cluster of haplotypes are named I, II and III in the figure. Contributor M1 unfortunately simply claimed to be Sherpa but could not provide his Ru, which in the figure is replaced by a question mark "?." Moreover, M8 stated to be of Tibetan origin and therefore he had no Sherpa Ru kinship: in the figure this is represented by a slash "/."

clusters, named Cluster I and II, that are mostly composed by individuals affiliated to ru all descending from the Minyagpa proto-clan (Thaktok, 5 individuals; Pinasa, 2 individuals; Gardza, 2 individuals; Gole, 2 individuals; Chuserwa, 1 individual). The Minyagpa proto-clan could therefore have comprised ancestors with two different Y-chromosome lineages, consistently with the two different haplogroups predicted, Q and R1, that respectively characterize Cluster I and Cluster II. Moreover, the Thaktok ru, being common to both Clusters I and II, is indeed still heterogeneous in terms of Y haplotypes and haplogroups. A similar polyphyletic origin for the Minyagpa clan is thus quite plausible. On the other hand, individual M29 falls as outsiders in Cluster I, since he belongs to the Chakpa lineage and not to the Minyagpa one. Similarly, the Tibetan contributor M8 and the Lama ru member M19, are part of Cluster II but they do not belong to the Minyagpa proto-ru. In fact, Lama and Chakpa ru were expected to form two distinct clusters, since they are numbered among the four original Tibetan proto-ru and are described as unsplit clans. Moreover, M25, a Goparma from the Thimmi proto-clan, unexpectedly figures in Cluster II. In order to explain these discrepancies, illegitimacies, adoptions, and personal data uncertainty should be considered.

The second macro-cluster, named Cluster III, comprises members of the Salaka (2 individuals) and Paldorje (1 individual) ru: both clans belong to the Thimmi proto-ru group (Table 1). Cluster III also comprises 3 individuals belonging to the Chuserwa and Tsangup ru, that Oppitz (1968) and Wangmo (2005) regard as recent clans and unrelated to the four Kham proto-ru. Nevertheless, the genetic evidence could be revealing an unreported kinship of these two ru with the Thimmi proto-clan. Moreover, Cluster III is mainly composed by haplotypes belonging to the E1b1 haplogroup. Cluster III could be then regarded as the putative Thimmi proto-ru tree cluster. In addition, assuming the authenticity of the reconstructed genealogies (Oppitz, 1968), several marriages between members of the Minyagpa, Chakpa, and Thimmi clans seem to have taken place during the 15th century (Figure 3). As discussed above, kinship incongruities observed in some NJ tree branches might thus be linked to events of haplotype interchange that occurred among the families of the Sherpa ancestors.

In fact, there is only one clustering outlier represented by M22, a Pinasa Sherpa, thus coming from proto-ru Minyagpa, with the unique property of belonging to the H haplogroup.

Samples were further grouped by considering the birthplaces of contributors. Home villages were located on a map (Figure 1) and assigned either to the Khumbu Valley (North cluster) or to the Solu and Okhaldunga area (South cluster) (Table 4). This kind of approach was undertaken as a preliminary evaluation of two contrasting hypothesis, the

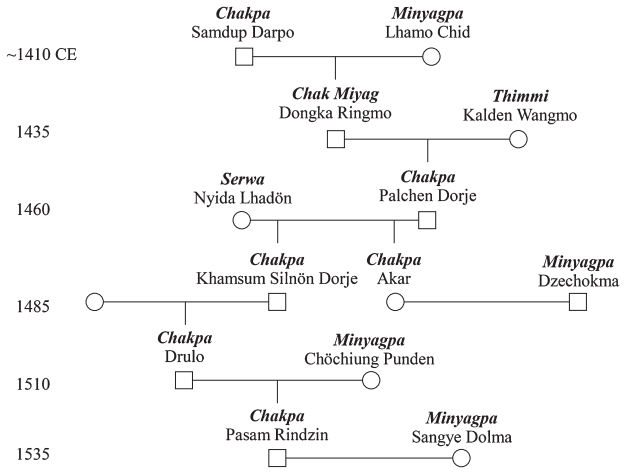

Figure 3. The Chakpa genealogy: an example of intermarriages among the four proto-ru of the Sherpa. The picture reproduces part of a wider and more complex tree of the Chakpa ru genealogy drawn by Oppitz based on the translation of some Sherpa clan records (Oppitz, 1968). From the tree it can be inferred that various marriages between members of the four proto-clans took place during the 15th century.

'Khumbu first' and the 'Solu first' theories, that describe the history of Sherpa settlement in the Solukhumbu area. Briefly, authors supporting the 'Khumbu first' theory agree that Sherpa entered Nepal from a high mountain pass, notably the Nangpa La, and immediately settled Khumbu. In particular, they state that the Thimmi proto-clan originally settled in the Bhote Kosi valley, in Western Khumbu (Oppitz, 1968; Krämer and Sherpa, 2002), while Minyagpa families at first settled in the Phortse village area, also in Khumbu, but later on moved to the Solu lowlands (Oppitz, 1968; Krämer and Sherpa, 2002). Moreover, Lama and Chakpa people supposedly went straight to Solu and directly settled there (Krämer and Sherpa, 2002).

On the other hand, the 'Solu first' theory suggests that the settlement of the Sherpa in Nepal did not even begin in Khumbu. In fact, the higher Khumbu valley was probably settled only after Solu, since at the time of their arrival prolonged periods of frost still sequestered its cultivable lands. Frost periods faded only with the ending of Little Ice Age, due to major climate changes (Muehlich, 1998). Moreover, in Khumbu only 'mixed' Sherpa settlements are found, whereas in Solu true Sherpa-clan territories and villages can be identified, even nowadays (Muehlich, 1996). This fact again strongly supports that the first Sherpa settlement area was indeed Solu.

Indirectly, the statement of Krämer and Sherpa (2002) about Lama and Chakpa families heading straight to Solu supports the 'Solu first' hypothesis as well. Genealogies provided by Oppitz (1968) too suggest that upper Khumbu was more recently and primarily settled by newer ru, as well as by Tibetans coming from the bordering Tingri region. Newcomers were then accepted and integrated by Sherpa already resident in their society (Muehlich, 1998).

Interpretation of genetic evidence and ru spatial distribution seems favourable to elements of both theories: it suggests that all but two (71.4%) of the Thimmi cluster members, which also comprises most of the new ru, were born in upper Khumbu and eight out of nine (88.9%) members of the major Minyagpa cluster were born in the Southern Solu and Okhaldhunga areas (Table 4). An ancient proto-clan settlement area could thus be roughly identified both for the Thimmi (Khumbu Valley) and for at least one family branch of the Minyagpa (Solu).

Table 4. Birth-village, ru and proto-ru

Sample id	Birthplace	Ru	Proto-ru	MST Cluster
Khumbu area				
M4	Nangma	Salaka	T	I (71.4%)
M23	Namche	Chuserwa	n	
M26	Thame	Tsangup	n	
M27	Chukkhung	Chuserwa	n	
M6	Shigatze	Paldorje	T	
M21	Khumjung	Chuserwa	n	II (11.1%)
M1	Namche	?	?	III (71.4%)
M3	Namche	Thaktok	M	
M28	Namche	Thaktok	M	
M12	Namche	Lhukpa	n	
M25	Kunde	Goparma	T	
Solu and Okhaldunga area				
M13	Okhaldhunga	Salaka	T	I (28.6%)
M22	Okhaldhunga	Pinasa	M	
M20	Goli	Thaktok	M	II (88.9%)
M11	Kerung	Pinasa	M	
M24	Okhaldhunga	Garja	M	
M16	Patle	Thaktok	M	
M17	Patle	Thaktok	M	
M18	Patle	Pinasa	M	
M29	Solu	Chawa	C	
M10	Tapting	Garja	M	
M14	Gepchuka	Gole	M	III (42.8%)
M15	Gepchuka	Gole	M	
M19	Beni	Lama	L	

Samples are grouped both by geographical provenance (Khumbu vs. Solu and Okhaldunga area) and by NJ cluster affiliation (clusters I, II, and III). The last column reports percentages of cluster members born in the corresponding geographical area.

In summary, three main clusters of haplotypes were found, ascribable on the whole to two of the presumed proto-clans, the Thimmi and the Minyagpa Tibetan ru. Chakpa and Lama appeared as sub-clans of the Minyagpa rather then self-standing families. Ethnogenetic findings about the clan structure of the Sherpa, do not in fact contradict the reported theories and also confirm an interesting association between Y-haplotypes and social kinship. The oral Sherpa tradition indeed gave birth to a family identification system similar to the one adopted by societies with a formalized patrilineal transmission of surnames. A major weakness of the present preliminary research is the small sample size. As it stands, members of every ru were not typed and consequently some clans are still missing from the tree. However, the preliminary results strongly encourage the ongoing pursuit of ethnogenetic research on Sherpa history: it is likely that Sherpa true clan genealogy and history will be fully elucidated by future sampling campaigns, aimed at enlarging the sample size, and in loco investigations. Moreover, among our future intentions stands out the achievement of a distinct sampling

campaign in some Sichuan prefectures, where the legendary Sherpa homeland should lie, in order to assess potential genetic similarity between the Solukhumbu Sherpa and the autochthonous population. Promisingly, in the present Dartsedo (Kangding) County, in the balkanized area of the so-called Western Sichuan Ethnic Corridor (Sun, 1990), a small ethnic group called Minyag is still dwelling (Ikeda, 1998). Some authors sustain that the Minyag people of Sichuan could be the descendants of some dynasty that from 1038 to 1227 CE ruled a North Minyag kingdom northeast of Lake Kokonor, also identified as part of Xi-Xia or Tangut Kingdom, in what later became the Tibetan province of Amdo (Stein, 1951; Balikci, 2008). Sichuan Minyag possibly migrated to their current dwelling area after the Mongols conquered the former Minyag kingdom in 1227.

Surprisingly, it is curious to find that the Namgyal dynasty of the Sikkim kings, the Chogyals, also claims to descend from a Minyag clan member, named Guru Tashi. Oral histories report that, during the 13th century, Guru Tashi left his homeland in Kham Minyag and, via Lhasa, eventually crossed a Himalayan pass and reached the Chumbi Valley in Sikkim. This strong analogy with Sherpa oral traditions about their migration route along with the quite common recurrence of the Minyag appellation in apparently unrelated ethnic groups, definitely represent an interesting research trail that should be further investigated.

In conclusion, in the firm belief that in the end only a multidisciplinary research approach can yield an exhaustive explanation of the true origin of the Solukhumbu Sherpa, we plan to draw together a team capable of juxtaposing clues from linguistics, social sciences, and ethnoclimatology to the genetic data.

Acknowledgments

Revealing ethnohistory through the juxtaposed investigation of manuscripts, oral tradition, and genetic pedigrees is a captivating and engrossing path: the authors wish to thank everybody who, like them, became fascinated with the Solukhumbu Sherpa tradition and especially all the Sherpa volunteers that contributed in person to the present research. The authors are also grateful to the University of Padua for a grant (no. CPDR100470).

References

Applied Biosystems (2006) AmpFlSTR® Yfiler™ PCR Amplification Kit User's Manual. http://www3.appliedbiosystems.com (accessed 3 December 2010)

Applied Biosystems (2010) GeneAmp® PCR System 9700 User's Manual. http://www3.appliedbiosystems.com (accessed 3 December 2010)

Arcos-Burgos M. and Muenke M. (2002) Genetics of population isolates. Clinical Genetics, 61: 233–247.

Athey W. (2007) Haplogroup Predictor ver. 21: Y Haplogroup Prediction from Y-STR Values. http://www.hprg.com/hapest5/index.html (accessed 19 April 2012)

Balikci A. (2008) Lamas, shamans and ancestors: village religion in Sikkim. Koninklijke Brill NV, Leiden, the Netherlands.

Brower B. (1991) Sherpa of Khumbu: people, livestock and landscape. Oxford University Press, Delhi.

CBRG (2012) TreeGen: Tree generation tool from distance data. Computational Biochemistry Research Group (CBRG), Swiss Federal Institute of Technology (ETH), Zurich. www.cbrg.ethz.ch/services/TreeGen (accessed 8 October 2012)

Childs G. (1997) A note on the Tibetan Origin of the Sherpa Serwa lineage. Himalayan Research Bulletin, 17: 23–25.

Crawford M.H. (2007) Anthropological Genetics: Theory, Methods and Applications. Cambridge University Press, Cambridge.

Dromo Y., Ota M., Hanaoka M., Katsuyama Y., Basnyat B., Neupane P., Arjyal A., Pandit A., Sharma D., Ito M., and Kubo K. (2008) Two hypoxia sensor genes and their association with symptoms of acute mountain sickness in Sherpas. Aviation Space and Environmental Medicine, 79: 1056–1060.

Excoffier L. and Lischer H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10: 564–567.

Fantin M. (1973) Sherpa Himalaya Nepal. Arti grafiche Tamari, Bologna.

Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39: 783–791.

Gayden T., Chennakrishnaiah S., La Salvia J., Jimenez S., Regueiro M., Maloney T., Persad P.J., Bukhari A., Perez A., Stojkovic O., and Herrera R.J. (2010) Y-STR diversity in the Himalayas. International Journal of Legal Medicine, 124: 295–300.

Gusmao L., Butler J.M., Carracedo A., Gill P., Kayser M., Mayr W.R., Morling N., Prinz N., Roewer L., Tyler-Smith C., and Schneider P.M. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Science International, 157: 187–197.

Ikeda T. (1998) Some phonological features of modern Munya (Minyak) language. Nairiku Ajia Gengo no Kenkyuu, 13: 83–91 (in Chinese).

Jobling M.A. and Tyler-Smith C. (2003) The human Y chromosome: an evolutionary marker comes of age. Nature Reviews—Genetics, 4: 598–612.

King T.E. and Jobling M.A. (2009) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends in Genetics, 25: 351–360.

Krämer K.H. and Sherpa L.D. (2002) Die Sherpa von Solu-Khumbu. Nepal Information, 89: 12–13.

Kwak K.D., Jin H.J., Shin D.J., Kim J.M., Roewer L., Krawczak M., Tyler-Smith C., and Kim W. (2005) Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. International Journal of Legal Medicine, 119: 195–201.

Ljubković J., Stipišić A., Sutlović D., Definis-Gojanović M., Bučan K., and Anđelinović Š. (2008) Y-chromosomal short tandem repeat haplotypes in southern Croatian male population defined by 17 Loci. Croatian Medical Journal, 49: 201–206.

Malacrida S., Katsuyama Y., Droma Y., Basnyat B., Angelini C., Ota M., and Danieli G.A. (2007) Association between human polymorphic DNA markers and hypoxia adaptation in Sherpa detected by a preliminary genome scan. Annals of Human Genetics, 71: 630–663.

Muehlich M. (1996) Traditionelle Opposition: Individualitaet und Weltbuild der Sherpa. Pfaffenweiler.

Muehlich M. (1998) Rejoinder and response to 'Solukhumbu and the Sherpa' on Himalayan Research Bulletin 17:2. Himalayan Research Bulletin, 18:1.

Mulero J.J., Chang C.W., Calandro L.M., Green R.L., Li Y., Johnson C.L., and Hennessy L.K. (2006) Development and validation of the AmpFlSTR® Yfiler™ PCR Amplification Kit: a male specific, single amplification 17 Y-STR multiplex system. Journal of Forensic Sciences, 51: 64–75.

Nei M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Oppitz M. (1968) Geschichte und Sozialordnung der Sherpa. Universitäts Wagner, Innsbruck-München.

Parkin E.J., Kraayenbrink T., Opgenort J.R., Van Driem G.L., Tuladhar N.M., De Knijff P., and Jobling M.A. (2007) Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. Forensic Science International, 166: 176–181.

Presciuttini S., Caenazzo L., Amoruso I., and Caravello G.U. (2010) Ru affiliation in the Sherpa of the Khumbu Valley: ethnographic survey and preliminar analysis of Y-chromosomes haplotypes. In: Jha P.K. (ed.), Contemporary Research in Sagarmatha (Mt. Everest) Region. Nepal Academy of Science and Technology, Lalitpur.

Qiagen (2010) QIAamp® DNA Micro Handbook. http://www.quiagen.com (accessed 12 January 2011)

Sherpa A.T. (2007) Thesis on the origin of Sherpa peopole. Unpublished.

Stein R.A. (1951) Mi-ñag et Si-hia. Géographie historique et légendes ancestrales. Bulletin de l'Ecole française d'Extrême-Orient, 44: 223–265.

Stevens S.F. (1993) Claiming the High Ground: Sherpas, Subsistence, and Environmental Change in the Highest Himalaya. University of California Press, Berkeley.

Sun H. (1990) Languages of the ethnic corridor in Western Sichuan. Linguistics of the Tibeto-Burman Area, 13: 1–31.

Tian-Xiao Z., Li Y., and Sheng-Bin L. (2009) Y-STR haplotypes and the genetic structure from eight Chinese ethnic populations. Legal Medicine, 11: S198–S200.

Von Fürer-Haimendorf C. (1964) The Sherpas of Nepal: Buddhist Highlanders. Murray, London.

Wangmo J. (2005) The Lawudo Lama. Wisdom Publications, Boston, MA.