Language and Genes of the Greater Himalayan Region

T. Kraayenbrink[1], Dr. P. de Knijff[1], Prof. Dr. G.L. van Driem[2], Dr. J.R.M.L. Opgenort[2], Dr. M.A. Jobling[3], Dr. E.J. Parkin[3], Dr. C. Tyler-Smith[4], Dr. D.R. Carvalho-Silva[4], Karma Tshering of Gaselô[2], Prof. Dr. G. Barbujani[5], I. Dupanloup[5], G. Bertorelle[5], Prof. Dr. Nirmal Man Tuladhar[6]
(authors listed in random order)

[1] MGC-Department of Human and Clinical Genetics, Leiden University Medical Centre, the Netherlands
[2] Himalayan Languages Project, Leiden University, the Netherlands
[3] Department of Genetics, University of Leicester, United Kingdom
[4] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom
[5] Department of Biology, University of Ferrara, Italy
[6] Professor of Linguistics, Centre for Nepal and Asian Studies (CNAS) of Tribhuvan University (TU), Kirtipur, Nepal

Summary

Isolation is crucial to both biological and linguistic change. Populations that are separated by barriers tend to diverge genetically because of genetic drift, and to independently undergo linguistic change, resulting in often-parallel patterns of genetic and cultural differentiation. Geographical as well as cultural barriers reduce population contacts, thus increasing isolation between populations. However, both biological and linguistic change are influenced by the size of the population. It is thus also important to infer reliable information on past human demography.

The greater Himalayan region is ethnolinguistically the most complex area of the Eurasian continent. This region includes the highest land barrier on the face of this planet, and linguistic evidence unambiguously indicates that topography has shaped and channelled prehistoric population movements. The intricate ethnolinguistic mosaic of this region holds many keys to the peopling of the Eurasian continent as a whole. Whereas most language communities in the Himalayan region belong either to the Tibeto-Burman or Indo-European family, there are also Austroasiatic, Dravidian, Daic and Altaic language communities settled in the mountains, foothills and periphery of the Himalayas. Moreover, there are two language isolates, Burushaski and Kusunda, in the region.

We have now sampled about 1000 individuals from Nepal and 1000 from Bhutan, representing 35 to 40 different groups identified on the basis of language and/or geography or caste. Results from studies on other populations have shown that past demographic expansions leave recognizable marks in the patterns of genetic diversity. Whereas mitochondrial data suggest rapid expansions for all populations, except those that did not practise agriculture, Y chromosome diversity is compatible with constant population sizes through much of our evolutionary history. One (not the only) way to reconcile these results is to take them as evidence that polygyny was the rule in humans up to a recent past, possibly until the Neolithic shift from mobile to sedentary communities. By means of genetic information from the autosomes, the Y chromosome and of the mtDNA genome in the samples we have collected, we will be able to compare and test alternative models of the peopling of Eurasia based on linguistic and archaeological evidence.

Planning of the project

Initially, the aim of the study was to organise three expeditions, to Nepal, Bhutan and northeastern India (specifically: Assam and Sikkim), to collect blood from the main ethnolinguistic groups of the greater Himalayan region.
Unfortunately, recently introduced Indian medical-ethical guidelines and bureaucratic restrictions imposed on foreign geneticists by Indian national and local authorities have made the collection of genetic samples in India virtually impossible.
Therefore, the project was restricted to the analyses of the Nepalese and Bhutanese ethnolinguistic groups, described in detail below. Fortunately, a plethora of manuscripts, all describing the rich Indian genetic heritage now appear in Journals. These data will still allow us to compare the Nepalese and Bhutanese data with those from India.

Expedition 1: Nepal

The first expedition aimed at collecting blood samples from Nepal's populations. During this expedition, held in December 2002 and January 2003, the team was assisted by Ivo van Asperen (Leiden University Medical Centre), for the blood-sampling and by several Nepali assistants for the screening.
The main areas of sample collection were the Kathmandu valley and the cities of Pokhara and Gorkha. The work in Nepal was carried out with the knowledge and cooperation of representatives of local groups and Tribhuvan University at Kirtipur. With the valuable assistance and guidance of Prof. Dr. Nirmal Man Tuladhar (Professor of Linguistics at Tribhuvan University), representatives of the ethnolinguistic groups were contacted and asked for cooperation in the project. The names of these representatives can be found in the detailed acknowledgements.

Blood donation was on a voluntary basis and a blood-sample (10 ml) was only taken if the donor had read, understood and signed the informed consent form. When a donor was unable to read or write, the consent text (given in Figure 1) was read to the donor, after which one of the project's co-workers filled in his or her data (viz. group, name, date of birth, place of birth) on the form.

Figure 1: Informed consent for the Nepal expedition
(this text was also available in Nepali)

Informed consent
The greater Himalayan region is the area of the greatest phenotypic and linguistic diversity in the Eurasian continent. Yet a comprehensive genetic assay of Nepal has never been undertaken. Genetic research has already demonstrated the unity of all mankind. The findings of the present genetic research conducted by the Leiden University project Languages and Genes of the Greater Himalayan Region may have implications for principal theories of language relationship and the provenance of the many Nepalese population groups and language communities. From each individual we will collect 10 ml of blood. The blood specimens will be treated as anonymous samples. It will therefore not be possible for you to correspond with us about the research findings, but the results will be published in specialist literature in due time. The Ethical Protocol of the Human Genome Diversity Project will be strictly observed. Your blood specimens can therefore never be used for medical, pharmaceutical or commercial purposes. Your participation and consent will enable us to know more about the origin of the languages and the populations of Nepal.

To ascertain that a person belonged to a certain ethnolinguistic group or caste, the donor's name and place of birth was systematically checked against what is known about the names adopted by members of Nepal's diverse ethnic groups and the geographical spread of the group to which the person identified himself or herself to belong. In addition, several team-members and most of the project's co-workers speak one or more relevant Nepali languages. To ascertain that a person was ethnically unmixed, both parents had to belong to the same group.

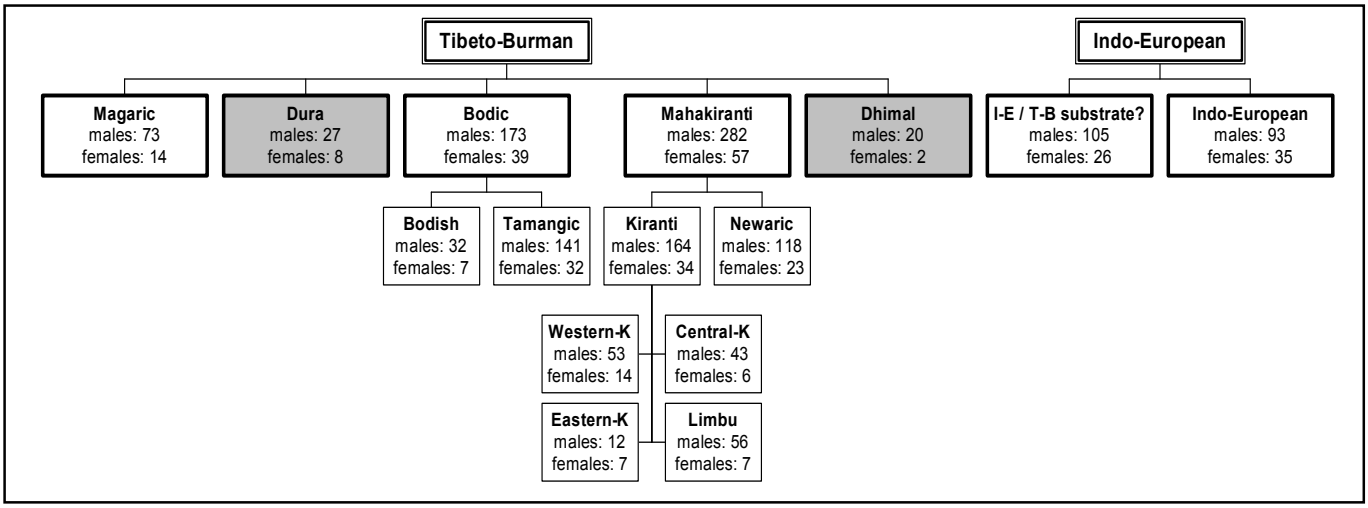
A detailed list of the numbers of samples collected per group is given in Table 1. The Magaric, Bodish, Tamangic, Kiranti and Newaric main groups consist of various groups speaking related Tibeto-Burman languages. However, the five Limbu groups represent four dialects and a general group for unclassified ethnic Limbu. Note also that a distinction is made between ordinary and high caste Newars, as the Newars are the sole Tibeto-Burman people to have adopted the Indo-Aryan caste system. In addition to main groups speaking Tibeto-Burman languages, there are also groups speaking Indo-European languages. The first main group consists of several tribes speaking an Indo-European language with a possible Tibeto-Burman substrate. The second Indo-European main group comprises ostensibly endogamous social groups more commonly known as castes.

Table 1: Overview of samples collected in Nepal

Main group	Group	Males	Females	Total
Magaric	Magar	40	6	46
	Kham (Magar)	13	1	14
	Chepang (Praja)	20	7	27
Dura	Dura	27	8	35
Bodish	Lhomi (Shingsaba)	10	2	12
	Sherpa	21	5	26
	Tibetan	1	0	1
Tamangic	Tamang	40	10	50
	Chantyal	21	2	23
	Ghale	20	5	25
	Gurung	40	6	46
	Thakali	20	9	29
Western Kiranti	Bahing	5	1	6
	Dumi	3	1	4
	Jero	7	2	9
	Khaling	2	0	2
	Sunwar	6	1	7
	Thulung	10	0	10
	Wambule	20	9	29
Central Kiranti	Bantawa	13	3	16
	Chamling	12	2	14
	Chintang	1	0	1
	Dungmali	1	0	1
	Kulung	5	0	5
	Nachiring	1	0	1
	Puma	1	0	1
	Sampang	9	1	10
Eastern Kiranti	Athpahariya	1	0	1
	Lohorung	3	1	4
	Yakkha	8	6	14
Limbu	Chathare	4	0	4
	Limbu (general)	12	1	13
	Pañcthare	9	1	10
	Phedappe	12	5	17
	Tamarkhole	19	0	19
Newaric	Newar	44	10	54
	High Caste Newar	24	6	30
	Barâm	32	5	37
	Jirel	2	0	2
	Thangmi	16	2	18
Dhimal	Dhimal	20	2	22
Indo-European (Tibeto-Burman Substrate)	Danuwar	26	4	30
	Darai	0	2	2
	Kachariya Danuwar	7	2	9
	Kumal	21	5	26
	Majhi (Bhote)	23	6	29
	Tharu	28	7	35
Indo-European	Brahmin	26	8	34
	Damai	9	1	10
	Giri	3	1	4
	Kshetriya	37	10	47
	Sarkí	12	11	23
	Sonar	3	1	4
	Vishvakarma	2	2	4
	Shah	0	1	1
	Rana	1	0	1

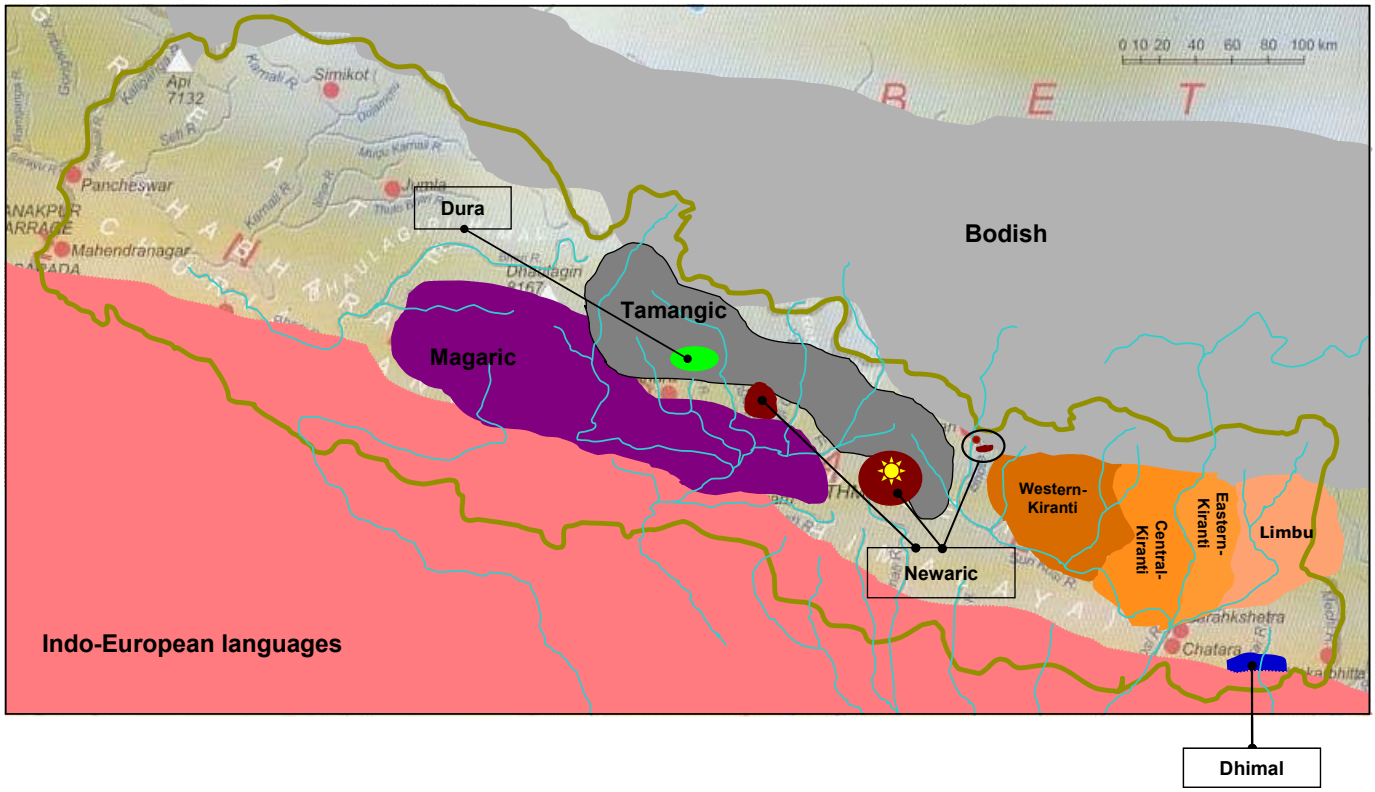
Figure 2 shows a pedigree (with total sample numbers per main group), linking the different Nepali groups based on the current linguistic theories. The Limbu main group is linguistically speaking closely related to the Eastern, Central and Western Kiranti groups. The Kiranti group plus Newaric form Mahakiranti. The Bodish and Tamangic main groups form Bodic.

Figure 2: "Pedigrees" linking the main ethnolinguistic groups of Nepal



A general overview of the distribution of these ethnolinguistic groups throughout Nepal is shown in Figure 3.

Figure 3: General distribution of the main ethnolinguistic groups throughout Nepal



A total of 954 blood samples from 20+ different ethnically and/or linguistically defined Nepali groups (11 main ethnolinguistic groups) were collected.

Expedition 2: Bhutan

The second expedition, which aimed at collecting blood samples from Bhutanese populations, was completed in October and November 2003. The work in Bhutan was carried out with the knowledge, permission and cooperation of the Royal Government of Bhutan at Thimphu and with the valuable assistance of Ivo van Asperen, Janine van Nes, Dr. Mariëtte Hoffer and Karma Tshering.

As in Nepal, blood donation was on a voluntary basis and sampling of 10 ml of blood was only carried out after the donor had read, understood and signed the consent form. Again, when a donor was unable to read or write, the consent text (given in Figure 4) was read to the donor and his or her data were filled in by one of the projects' co-workers.

Figure 4: Informed consent for the Bhutanese expedition
(this text was also available in Dzongkha, the national language of Bhutan)

Informed consent
The greater Himalayan region is the area of the greatest phenotypic and linguistic diversity in the Eurasian continent. Yet a comprehensive genetic assay of Bhutan has never been undertaken. Genetic research has already demonstrated the unity of all mankind. The findings of the present genetic research conducted by the Dzongkha Development Commission of the Royal Government of Bhutan and the Leiden University project Languages and Genes of the Greater Himalayan Region may have implications for principal theories of language relationship and the provenance of the many Bhutanese population groups and language communities. From each individual we will collect 10 ml of blood. The blood specimens will be treated as anonymous samples. It will therefore not be possible for you to correspond with us about the research findings, but the results will be published in specialist literature in due time. The Ethical Protocol of the Human Genome Diversity Project will be strictly observed. Your blood specimens can therefore never be used for medical, pharmaceutical or commercial purposes. Your participation and consent will enable us to know more about the origin of the languages and the populations of Bhutan.

Donors had been pre-selected by representatives of the Royal Government of Bhutan, based on the same criteria as were used in Nepal (the only difference being that not all ethnolinguistic groups in Bhutan use group-specific names). Therefore, during donation, only the name, age and place of birth of the donors needed to be crosschecked.

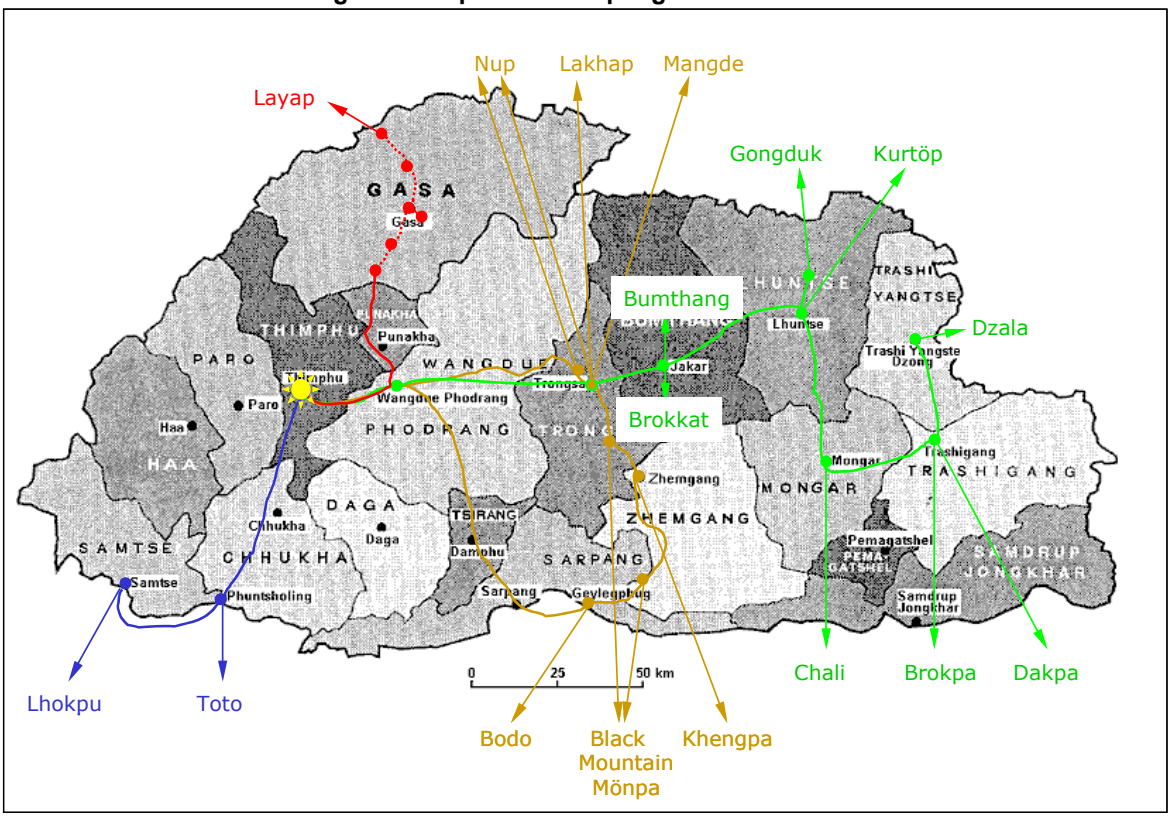
Blood was collected throughout Bhutan during four expeditions. A map with the sampling locations of the four expeditions, including the names of the groups collected at the locations, is shown in Figure 5.

The Bhutanese expeditions:

1. From 9 to 14 October 2003: Lhokpu and Toto
2. From 19 to 24 October 2003: Bodo, Black Mountain Mönpa, Mangde, Nup and Lakha
3. From 31 October 2003 to 8 November 2003: Gongduk, Chali, Kurtöp, Brokpa, Dakpa, Dzala, Brokkat and Bumthang
4. From 12 to 19 November 2003 and 28 February 2004: Layap

Members of other Bhutanese groups were sampled in and around Thimphu (the capital city of Bhutan) and when encountered during any of the four expeditions.

Figure 5: Map of the sampling locations in Bhutan

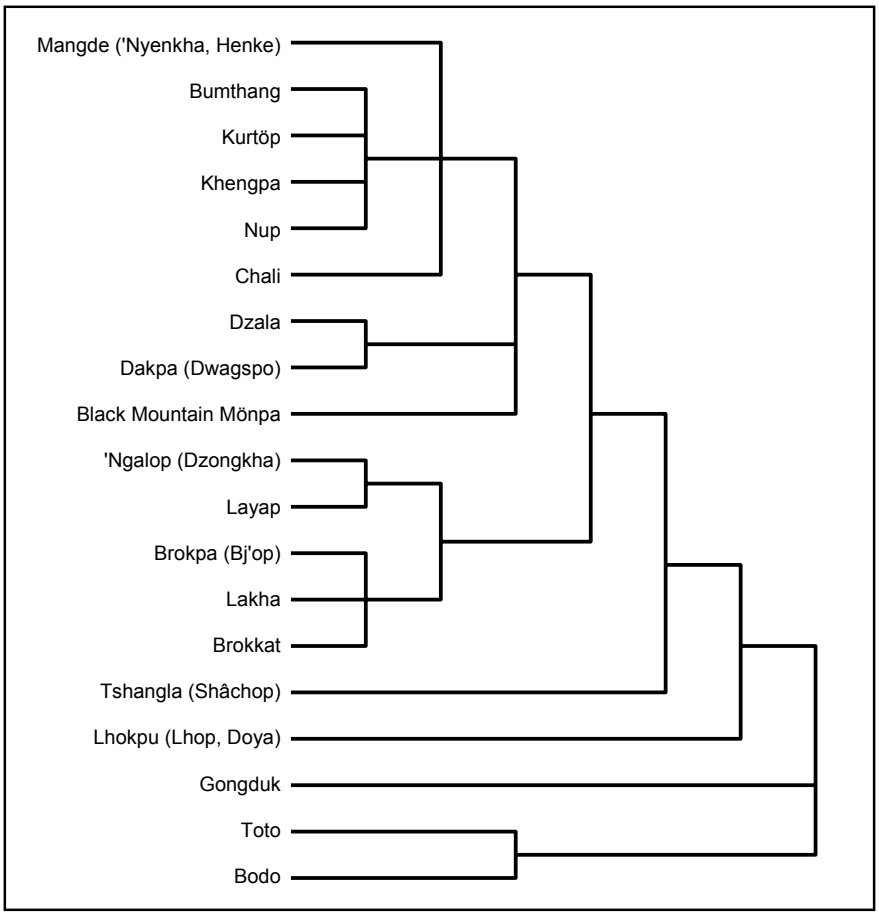


A total of 1050 samples from 19 ethnolinguistic groups were collected, all of which are sub-groups of Tibeto-Burman. A detailed list of the numbers of samples collected per group is given in Table 2 and a pedigree linking all Bhutanese ethnoliguistic groups, based on current linguistic theories is shown in Figure 6.

Table 2: Overview of samples collected in Bhutan

Group	Alternative group-names	Males	Females	Total
Black Mountain Mönpa		42	18	60
Bodo		37	2	39
Brokkat		24	5	29
Brokpa	Bjokha, Bj'op, Merak, Sakteng	41	10	51
Bumthang		50	10	60
Chali		54	10	64
Dakpa	Dwagspo	49	10	59
Dzala		51	11	62
Gongduk	Gongdukpa	51	10	61
Khengpa		52	10	62
Kurtöp	Zhâke	51	13	64
Lakha		50	10	60
Layap		25	5	30
Lhokpu	Lhop, Doya	45	11	56
Mangde	'Nyenkha, Henke	54	10	64
'Ngalop	Dzongkha	50	10	60
Nup	Trongsap	27	10	37
Toto	Totopara, Bhutan duars	54	17	71
Tshangla	Shâchop	51	11	62

Figure 6: "Pedigree" linking the ethnolinguistic groups of Bhutan



DNA research

After blood sampling, a small amount of blood was spotted in duplicate onto FTA filter-paper cards for archival purposes (4 spots of ~1 cm diameter per FTA card).

The blood samples were sent to the Netherlands via DHL as soon as possible after sampling (tubes and FTA cards were sent separately). One set of FTA cards is currently stored in Leiden and one set in Leicester for future reference.

The blood in the tubes was used for DNA isolation, using the Autopure LS® from Gentra Systems, according to the manufacturer's specifications. All blood samples yielded a sufficient amount of good quality DNA. Aliquots of all samples were shipped to the United Kingdom for Y chromosome and mitochondrial analyses.

The initial marker set for analysis of the samples comprises 21 autosomal STRs, 23 Y chromosomal STRS, Y chromosomal and mitochondrial SNPs (the number depending on the results of the typings) and the mitochondrial D-loop sequences HVR1 and HVR2.
Thus far, analysis of the autosomal STRs has been completed for all Nepali and all Bhutanese samples and analysis of the Y chromosomal STRs has been completed for all Nepali samples (an overview of the STR loci is given in table 3).

The choice of further markers to be tested will depend mainly on the data available for comparison in relevant neighbouring populations.

Table 3: Overview of STR loci

Autosomal STR loci	CSF1PO, D13S317, D16S539, D18S51, D19S433, D21S11, D2S1338, D3S1358, D5S818, D7S820, D8S1179, HumF13A01, HumF13B, HumFES/FPS, HumFGA, HumTH01, HumVWA, LPL, Penta D, Penta E, TPOX
Y chromosomal STR loci	DYS19, DYS389I / II, DYS390, DYS391, DYS392, DYS393, DYS385, YCAII, DYS388, DYS425, DYS426, DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, DYS447, DYS448, DYS460, DYS461, DYS462, H4

Acknowledgements

The research in Nepal was conducted in association with the Centre for Nepal and Asian Studies (CNAS) at Kirtipur under the Bilateral Agreement for Academic Cooperation between Tribhuvan University (TU) and Leiden University. The research team received the organisational support and enthusiastic assistance of many grass-roots community service organisations and from the many informed volunteers of indigenous language communities who took an active interest in the research and discussed with interest the ramifications of genetic investigations for an enhanced understanding of our shared prehistorical past. Special gratitude is due the Arjun Limbu, Yograj Limbu and the Kirat Yakthung Chumlung headquartered at Mahalakshmisthan in Lalitpur; the Praja Capacity Development Programme (Praja Samudayik Vikash Karyakram) at Shaktikhor in Chitwan district; to Ajay Praja, Santa Bahadur Praja and Dambar Bahadur Chepang and the Nepal Chepang (Praja) Sangh at Kathmandu; to our many Tharu friends at Sauraha; to Bal Gopal Shrestha of Sankhu and the Newar community organisation Friends of Sankhu; to Bharat Rai of the National Youth Service Trust, our dear friend Ashok Rai and the Danuwar Rai community; to Dilendra Subba and the Limbu Literary Development Association; to the Kirat Yakkha Chuma headquartered at Mahalakshmisthan in Lalitpur; to Shree Mani Chand Chantyal and the Nepal Chantyal Samaj headquartered in Samakhushi in Kathmandu; to Buddhiman Dura, Lt. Col. John P. Cross, Ritu Kumar Dura and the Dura Seva Samaj Sampark Karyalaya at Ram Bajar in Pokhara; to Kishor Dura, Singh Raj Dura and the Dura Seva Samaj Samiti at Vasundhara in Kathmandu; to the Kshetriya Karyalaya of the Thakali Seva Samiti at Nadipur in Pokhara; to Dashrath Rai and the Kendriya Karyalaya of the Kirat Rai Yayokkha at Kathmandu; to Lila Thangmi, Kavi Raj Thangmi of Lapilang and the Nepal Thami Samaj; to Avinath Rai and Ganesh Rai of the Wambule Rai Samaj Nepal; to our many kind Baram friends of Gorkha district, for whom we are still completing a grammatical description of the language; to our wonderful friend Temba Bhote and the Buddhist half of the Lhomi (Shingsaba, Bhote) community; to Lt. Col. Michael Roe, Capt. Simon Garside and those troops of the British Gurkhas at Pokhara who came forward to volunteer blood; to our old friend Vishva 'Bishow' Bhatt of Gorkha; to Tek Bahadur Kumal and the Kumal Service Organisation at Chevetar in Gorkha district; to the kind director and staff of the Youth Awareness Environmental Forum and the Environmental Library at Badegaon at Godawari. Gratitude is also due to the Nepal Janjati Mahasangh at Anamnagar for their scholarly interest and kind advice. Furthermore, we owe gratitude to Cas F. de Stoppelaar, Honorary Consul-General of Nepal in the Netherlands and to Kari Cuelenare of the Dutch Consulate in Nepal. Last but not least, we thank our old and dear friends Narayan Prasad 'Yangsarumba' Pangyangu Subba and Gram Bahadur 'Sarumba' Pangyangu Subba of Tamphula in Tehrathum district and our dear friend Surendra Raj Dhakal of Gorkha.

The research in the Kingdom of Bhutan was conducted as part of the long-standing bilateral cooperation between the Royal Government of Bhutan and Leiden University. The field campaign was carried out in accordance with the Memorandum of Understanding concluded between the Dzongkha Development Authority (DDA) and Leiden University. Much gratitude is due to the Chairmen of the Dzongkha Development Authority, Their Excellencies 'Lönpo Sangay Ngedup and 'Lönpo Thinley Gyamtsho, as well as to the Home Minister, His Excellency 'Lönpo Jigmi Yoezer Thinley, for encouraging and facilitating the research in every way. Logistics and preparations at district and village level were coordinated in a timely and thorough fashion by Dr'âsho Sangye Dorji, the Honourable Secretary of Dzongkha Development Authority, and by Tshewang Dorji, the Chief Research Officer. Much gratitude is due to all the village headmen and district officers who assisted the research team on site. Particular gratitude is due to our friends 'Adap Dôji and Seta of the Lhokpu language community at Loto Kucu, Karma Chen and Dhani Ram Toto at Phüntsho'ling, Tandri and friends at Riti in the Black Mountains, 'Ap Drakpa of Phajong Pam and friends in the Gongduk language community. Furthermore, we thank Cecila Keijzer, Peter Newsum and Yanchen Doma of SNV (Netherlands Development Organisation), Bhutan. Special thanks are due to our old friend Karma Tshering of Gaselô, who really made the work happen. Karma involved himself and made himself indispensable at every level, anticipated and prevented all possible difficulties, and facilitated every aspect of the research programme with unparalleled perspicacity.

Finally we would like to thank mr. H.N.W. van Gent of the Netherlands Ministry of Foreign Affairs in The Hague for his assistance in sending the materials to Nepal and Bhutan.