

## The Himalayas as a Directional Barrier to Gene Flow

Tenzin Gayden, Alicia M. Cadenas, Maria Regueiro, Nanda B. Singh, Lev A. Zhivotovsky, Peter A. Underhill, Luigi L. Cavalli-Sforza, and Rene J. Herrera

High-resolution Y-chromosome haplogroup analyses coupled with Y-short tandem repeat (STR) haplotypes were used to (1) investigate the genetic affinities of three populations from Nepal—including Newar, Tamang, and people from cosmopolitan Kathmandu (referred to as “Kathmandu” subsequently)—as well as a collection from Tibet and (2) evaluate whether the Himalayan mountain range represents a geographic barrier for gene flow between the Tibetan plateau and the South Asian subcontinent. The results suggest that the Tibetans and Nepalese are in part descendants of Tibeto-Burman-speaking groups originating from Northeast Asia. All four populations are represented predominantly by haplogroup O3a5-M134-derived chromosomes, whose Y-STR-based age ( $\pm$  SE) was estimated at  $8.1 \pm 2.9$  thousand years ago (KYA), more recent than its Southeast Asian counterpart. The most pronounced difference between the two regions is reflected in the opposing high-frequency distributions of haplogroups D in Tibet and R in Nepal. With the exception of Tamang, both Newar and Kathmandu exhibit considerable similarities to the Indian Y-haplogroup distribution, particularly in their haplogroup R and H composition. These results indicate gene flow from the Indian subcontinent and, in the case of haplogroup R, from Eurasia as well, a conclusion that is also supported by the admixture analysis. In contrast, whereas haplogroup D is completely absent in Nepal, it accounts for 50.6% of the Tibetan Y-chromosome gene pool. Coalescent analyses suggest that the expansion of haplogroup D derivatives—namely, D1-M15 and D3-P47 in Tibet—involved two different demographic events ( $5.1 \pm 1.8$  and  $11.3 \pm 3.7$  KYA, respectively) that are more recent than those of D2-M55 representatives common in Japan. Low frequencies, relative to Nepal, of haplogroup J and R lineages in Tibet are also consistent with restricted gene flow from the subcontinent. Yet the presence of haplogroup O3a5-M134 representatives in Nepal indicates that the Himalayas have been permeable to dispersals from the east. These genetic patterns suggest that this cordillera has been a biased bidirectional barrier.

The Himalayan mountain range extends from Pakistan in the west to Burma in the east, along the frontiers of northern India, Nepal, Tibet, and Bhutan. It is home to most of the highest mountains of the world and forms a natural barrier between the Tibetan plateau and the Indian subcontinent. Whereas the Himalayas border Tibet to the south, the Tibetan plateau is bounded on the north by the Kunlun Mountains and on the west by the Karakoram ranges. These unique geographic characteristics of the Tibetan landscape may have provided some degree of genetic encapsulation.

Archaeological findings have revealed late Paleolithic inhabitation of the Tibetan plateau, dating the initial entry of modern humans at  $\sim 25$ – $30$  thousand years ago (KYA).<sup>1</sup> Yet, the discovery of Neolithic sites,<sup>1,2</sup> genetic data,<sup>3,4</sup> and linguistic studies<sup>5–7</sup> favor the peopling of the plateau during the Neolithic period.

Tibeto-Burman speakers are the major inhabitants of the Himalayas. They occupy the territories of present-day Bhutan, Burma, Nepal, northeastern India, and Tibet. This family of languages is also spoken in some parts of Southeast Asia.<sup>4,8</sup> According to historical records, Di-Qiang tribes of northwestern China migrated south  $\sim 3$  KYA, admixing with native residents on arrival.<sup>4,8–10</sup> Su and collaborators<sup>4</sup>

suggested that the Bodic and Baric branches<sup>11</sup> of the Tibeto-Burman subfamily populated Tibet and Nepal, respectively,  $\sim 5$ – $6$  KYA. However, van Driem<sup>6,7</sup> argues that the model presented by Su et al.<sup>4</sup> is flawed because of an inadequate representation of the Tibeto-Burman speakers in their study, as well as the problems associated with the linguistic phylogeny presented by Matisoff.<sup>11</sup>

Previous genetic studies that used classic markers,<sup>2,12</sup> autosomal microsatellites,<sup>13</sup> and mtDNA<sup>14</sup> have placed Tibetans along with Koreans, Japanese, and Mongolians in a Northeast Asian cluster, thereby suggesting a northern Mongoloid origin. Yet a report employing Y-chromosome biallelic markers argues for the peopling of Tibet, Nepal, and Bhutan by East Asians after they left the Yellow River region in China.<sup>4</sup> Other studies suggest that the high frequency of the Y *Alu* insertion (YAP) in the Tibetan population signals a significant genetic contribution from Central Asia.<sup>4,15,16</sup> Unfortunately, some of the aforementioned studies<sup>4,16</sup> suffer shortcomings because of the small number of binary markers employed, limited sample sizes, and the absence of Central Asian samples needed to assess their role in the genetic composition of the Tibetan population.

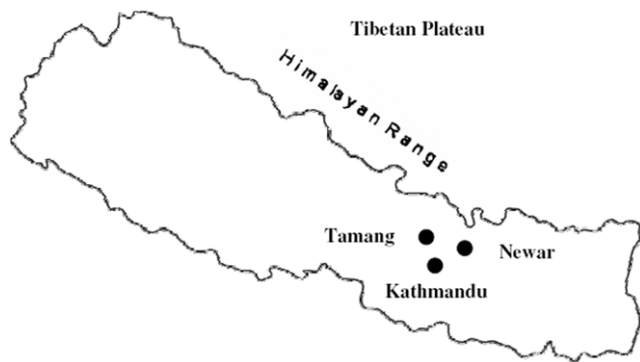
In contrast to the extensive anthropological and genetic

From the Department of Biological Sciences, Florida International University, Miami (T.G.; A.M.C.; M.R.; R.J.H.); Central Department of Zoology, Tribhuvan University, Kathmandu (N.B.S.); College of Medicine, Henan University, Kiefeng, China (N.B.S.); N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow (L.A.Z.); and Department of Genetics, Stanford University, Stanford (P.A.U.; L.L.C.-S.)

Received January 24, 2007; accepted for publication February 23, 2007; electronically published April 4, 2007.

Address for correspondence and reprints: Dr. Rene J. Herrera, Department of Biological Sciences, Florida International University, University Park, OE 304, Miami, FL 33199. E-mail: herrerar@fiu.edu

*Am. J. Hum. Genet.* 2007;80:884–894. © 2007 by The American Society of Human Genetics. All rights reserved. 0002-9297/2007/8005-0008\$15.00  
DOI: 10.1086/516757



**Figure 1.** Map of Nepal: geographic locations of the Himalayas and populations studied.

studies undertaken to analyze the rich genetic heritage and cultural diversity of the Indian subcontinent,<sup>2,17–23</sup> limited genetic investigations had been performed on the neighboring Himalayan region of Nepal, just south of Tibet across the Himalayas.<sup>24,25</sup> Kraaijenbrink et al.<sup>24</sup> reported the allele-frequency distributions of 21 autosomal STR loci, whereas the other study, by employing Y-STR analysis, revealed isolation and genetic drift in the Nepalese population.<sup>25</sup> Nepal is located between two ancient cultural giants, India and Tibet, and their influence in shaping Nepal's contemporary genetic landscape is undeniably significant. India and Nepal share a long historical, cultural, and religious legacy resulting from geographic proximity and several pivotal historical events involving political control by Indian royalty.

The genetic diversity of populations inhabiting an area is often influenced by the geographic and physical features encompassing the region. Whereas the Hindu Kush Mountains and the arid deserts in Iran have served as obstacles to gene flow,<sup>26,27</sup> the Nile River Valley,<sup>28</sup> the strait of Babel Mandeb,<sup>29,30</sup> and Beringia<sup>31</sup> are examples of natural passageways for the migrations of modern humans. The Himalayan range, in addition to being a formidable barrier, provides for dramatically diverse climatic conditions on either side of it: arid and cold on the Tibetan side, as compared with monsoons and extreme dry spells, depending on the season of the year, in the Nepalese territory.

We decided to investigate the influence of the Himalayas in populating the region by use of high-resolution Y-chromosome SNP analyses of four geographically targeted populations from the area—namely, Newar, Tamang, and the general population of Kathmandu from Nepal (referred to as “Kathmandu” subsequently), as well as a collection from Tibet. In addition, 15 Y-STR loci were typed to provide information on the temporal origins of these four Himalayan groups. This study is the first of its kind reporting biallelic markers of the Y-chromosome haplogroup diversity in Nepalese populations. This work improves on earlier reports by analyzing a large number of

Y-chromosome binary and associated STR markers to define compound Y-chromosome lineages for a sizable number of individuals from informative Tibetan and Nepalese populations. In the process, we uncovered evidence that the Himalayas have acted as an obstacle modulating the skewed dispersal of human groups southward.

## Material and Methods

### Sample Collection and DNA Isolation

Blood samples were collected after receipt of informed consent from 344 males who comprise the general population of Tibet ( $n = 156$ ) and three populations from Nepal ( $n = 188$ ), which include Tamang ( $n = 45$ ), Newar ( $n = 66$ ), and Kathmandu ( $n = 77$ ). Figure 1 illustrates the geographical location of the above four populations. With the exception of the Kathmandu group who speak predominantly Nepali, an Indo-European language, the individuals in the remaining three collections are Tibeto-Burman speakers. Genealogical information of sampled individuals was recorded for at least 2 previous generations. Table 1 lists the sample size, geography, and linguistic affiliation of the populations examined. All ethical guidelines were followed, as stipulated by the institutions involved in the study. DNA was extracted using phenol-chloroform extraction as described elsewhere<sup>38</sup> and was stored in 10 mM Tris–1 mM EDTA (TE) (pH 8.0), as stock solutions, at  $-80^{\circ}\text{C}$ .

### Y Haplogrouping

A total of 103 binary markers were genotyped by standard methods, including PCR-RFLP,<sup>30</sup> the YAP polymorphic *Alu* insertion (PAI),<sup>39</sup> and allele-specific PCR.<sup>27,40</sup> The nomenclature followed for the Y-SNP haplogroup is as recommended by the Y Chromosome Consortium.<sup>41</sup> Five Kathmandu males failed to amplify for marker M12 because of an  $\sim 2.45$ -Mb deletion that also encompasses the amelogenin gene.<sup>42</sup> Marker P47 was genotyped by Tsp509I RFLP analysis, which identifies a C→T transition at position 171 within the 364-bp PCR fragment amplified using primers 5'-CTGATGTTGCAGTGTGAGC-3' (forward) and 5'-ACACAGCCAAATACCAGTCG-3' (reverse).

### Y-STR Haplotyping and Time Estimation

Y-STR haplotypes were assayed using the AmpFISTR Yfiler PCR amplification kit (Applied Biosystems) that coamplifies the following 17 Y-STR loci: *DYS19*, *DYS385 a/b*, *DYS389I/II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS437*, *DYS438*, *DYS439*, *DYS448*, *DYS456*, *DYS458*, *DYS635*, and Y-GATA H4. *DYS385a* and *DYS385b* were not used in time estimation and variance calculations because of the duplicated nature of the loci. The PCR products were separated on a 3100 Genetic Analyzer (Applied Biosystems), and the data generated were analyzed with the GeneScan 3.7 and Genotyper 3.7 NT software. We treated *DYS389CD* as equivalent to *DYS389I* and *DYS389AB* as equivalent to the numerical subtraction of *DYS389II* minus *DYS389I*.<sup>35,43,44</sup>

Microsatellite variances were calculated using the Vp equation of Kayser et al.<sup>45</sup> Haplogroup ages, based on Y-STR variation, were estimated using the linear expansion,<sup>46,47</sup> BATWING,<sup>48</sup> and SNP-STR coalescence methods.<sup>44</sup> All three methods assume an average Y-STR mutation rate of 0.00069 per locus per generation,<sup>44</sup> with an intergeneration time of 25 years.<sup>30,49,50</sup> With the exception of a prior distribution for the population growth rate ( $\alpha$ ) at

**Table 1. Populations Analyzed**

Geographic Region and Population	Abbreviation	Language Family	No. of Subjects	References
<b>Middle East:</b>				
Northern Iran	NIR	Indo-European	33	Regueiro et al. <sup>27</sup>
Southern Iran	SIR	Indo-European	117	Regueiro et al. <sup>27</sup>
Iraq	IRQ	Afro-Asiatic	139	Al-Zahery et al. <sup>32</sup>
Lebanon	LEB	Afro-Asiatic	50	Wells et al. <sup>26</sup>
Syria	SYR	Afro-Asiatic	20	Semino et al. <sup>33</sup>
<b>Central Asia:</b>				
Kazak	KAZ	Altaic	54	Wells et al. <sup>26</sup>
Kyrgyz	KYR	Altaic	52	Wells et al. <sup>26</sup>
Karalkalpak	KAR	Altaic	44	Wells et al. <sup>26</sup>
Shugnan	SHU	Indo-European	44	Wells et al. <sup>26</sup>
Mongolia	MON	Altaic	24	Wells et al. <sup>26</sup>
Tibet	TIB	Sino-Tibetan	156	Present study
<b>South Asia:</b>				
Adi	ADI	Sino-Tibetan	55	Cordaux et al. <sup>20</sup>
Gujarat	GUJ	Indo-European	29	Kivisild et al. <sup>18</sup>
Punjab	PUN	Indo-European	66	Kivisild et al. <sup>18</sup>
Pakistan	PAK	Indo-European	176	Sengupta et al. <sup>23</sup>
Tamang	TAM	Sino-Tibetan	45	Present study
Newar	NEW	Sino-Tibetan	66	Present study
Kathmandu	KAT	Indo-European	77	Present study
<b>Northeast Asia:</b>				
Korea	KOR	Altaic	74	Karafet et al. <sup>34</sup>
Japan	JAP	Altaic	259	Hammer et al. <sup>35</sup>
Tuva	TUV	Altaic	42	Wells et al. <sup>26</sup>
Buryat	BUR	Altaic	81	Karafet et al. <sup>34</sup>
Manchu	MAN	Altaic	35	Xue et al. <sup>36</sup>
<b>Southeast Asia:</b>				
Philippines	PHI	Austronesian	48	Karafet et al. <sup>37</sup>
Malaysia	MAL	Austronesian	32	Karafet et al. <sup>37</sup>
Vietnam	VIE	Austro-Asiatic	70	Karafet et al. <sup>37</sup>
Bali	BAL	Austronesian	551	Karafet et al. <sup>37</sup>
Southern Han	SHA	Sino-Tibetan	166	Karafet et al. <sup>37</sup>

gamma (1.01, 1),<sup>30</sup> other prior assumptions for the BATWING analysis were as described by Cinnioglu et al.<sup>50</sup>

### Statistical and Phylogenetic Analyses

Twenty-four reference populations (table 1) from the published literature were included for statistical and phylogenetic analyses. Correspondence analysis (CA) was performed using the NTSYSpc-2.02i software.<sup>51</sup> The frequency data were analyzed at the phylogenetic resolution of major haplogroups that define clades A–R in the Y tree. Admixture estimation was performed<sup>52,53</sup> using the statistical package SPSS 14.0, with Northeast Asia, Southeast Asia, Central Asia, and India<sup>35</sup> as parental populations for the four Himalayan populations. Tibet is excluded from the Central Asian populations<sup>35</sup> in the admixture analysis.

## Results

### Phylogeography

Of the 103 binary markers analyzed, 14, 27, 28, and 29 loci were found to be polymorphic in Tamang, Newar, Tibet, and Kathmandu, respectively. Altogether, the 39 polymorphic loci in these four groups define 24 paternal haplogroups. Overall, Tamang exhibits high homogeneity with only 5 haplogroups, whereas Kathmandu shows the highest heterogeneity with 16 haplogroups, followed by

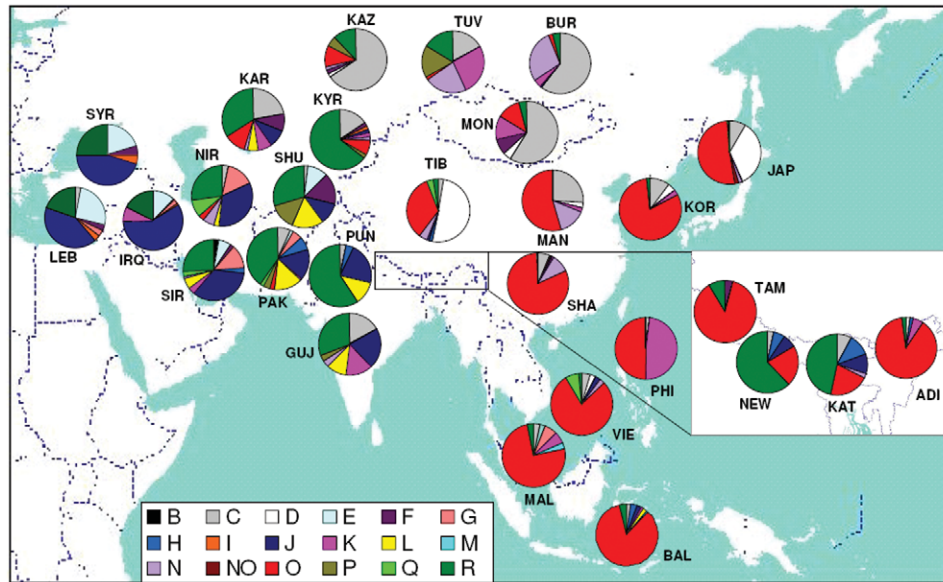
Tibet with 15 and Newar with 8. It is noteworthy that only three haplogroups—namely, O3a5a-M117, R1a1-M198, and R2-M124—are shared across the four groups. Neither haplogroup G-M201 nor haplogroup L-M20 was observed. Figure 2 displays the hierarchical phylogenetic relationships and frequencies of the 24 Y-chromosome haplogroups observed in the Tibetan and Nepalese collections. The geographic distribution of the major haplogroups is illustrated in figure 3.

Haplogroup O predominates in the Himalayan collections examined, accounting for an average frequency of 35.46% (frequency range of 20.8%–86.6%) (fig. 2). The majority of haplogroup O individuals are represented by subclade O3-M122 and its derived subtype O3a5-M134, which includes the terminal mutation O3a5a-M117 (fig. 2).

The second-most abundant major clade is haplogroup R, which occurs at an average frequency of 24.70% (frequency range of 2.5%–62.1%). It includes most of the Newar (62.1%) and Kathmandu (46.8%) groups but is significantly less represented in Tamang (8.8%) and Tibet (2.5%). Haplogroup R1b1a-M269 is found only in Newar (10.6%), whereas R1a1-M198 is present in all four collections (fig. 2). Similarly, haplogroup R2-M124 occurs in all four







**Figure 3.** Geographic distributions of major Y-chromosome haplogroup frequencies for the 28 populations listed in table 1

tion of  $10.8 \pm 5.6$  KYA determined using the coalescent method of Zhivotovsky et al.<sup>44</sup> Unless otherwise stated, the time estimates followed throughout this text are those calculated according to Zhivotovsky et al.<sup>44</sup> Y-STR genotypes are provided in tables 3–6.

The divergence times calculated for the haplogroups on the basis of Y-STR data resulted in few clusters of individuals with different age estimates. When analyzed across all four populations, O3a5-M134 resulted in one major cluster consisting of 79 individuals with an age ( $\pm$ SE) of  $8.1 \pm 2.9$  KYA and a minor cluster comprising only 9 Tamang samples, which exhibits a relatively young age of  $1.3 \pm 1.3$  KYA. The rest of the individuals do not cluster and can be interpreted as part of an ancient component of the O3a5-M134 haplogroup whose age is estimated at  $22.1 \pm 4.4$  KYA.

It is noteworthy that the haplogroup D-M174 subclades D1-M15 and D3-P47 segregate into two distinct groups at locus *DYS392*. Those with seven repeats are entirely represented by individuals belonging to haplogroup D3-P47, whereas the cluster with 11 repeats is made up exclusively of D1-M15 chromosomes. Both D1-M15 and D3-P47, in turn, form two subclusters each whose ages are  $5.1 \pm 1.8$  and  $2.3 \pm 1.0$  KYA and  $11.3 \pm 3.7$  and  $7.0 \pm 4.2$  KYA, respectively.

#### Admixture Proportions

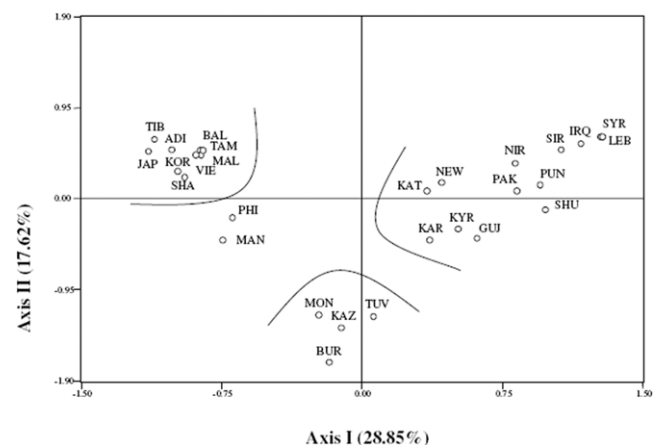
Table 7 displays the admixture proportions for Tibetan and Nepalese populations estimated using the weighted-least-squares method.<sup>52,53</sup> The admixture proportions were calculated using four parental populations, including Northeast Asia, Southeast Asia, Central Asia, and India. The results indicate that Tibet and Tamang have been the recip-

ients of significant Y-chromosome influence from Southeast Asia (60.4% and 66.2%, respectively) compared with Central Asia (26.2% and 28.7%, respectively), Northeast Asia (8.9% and 5.1%, respectively), and India (4.5% and 0.0%, respectively). In contrast, Central Asia and India account for the total contribution to Newar (56.6% and 43.4%, respectively) and for the vast majority of the Kathmandu Y chromosomes (51.4% and 48.6%, respectively), except for the negligible influence from Southeast Asia (1.3%).

## Discussion

### East Asian Influence on Himalayan Populations

The Himalayan populations examined in the present study exhibit high frequencies of haplogroup O3-M122, with its



**Figure 4.** CA of Y-haplogroup frequency data from 28 populations

**Table 2. Y-Chromosome Haplogroup Variance, Expansion, and Coalescence Time, Based on 15 Y-STR Loci**

Country and Haplogroup	No. of Subjects	Haplogroup Variance	BATWING Expansion Time <sup>a</sup>					Mean ( $\pm$ SE) Divergence Time <sup>b</sup> (KY)
			Time <sup>c</sup> (KY)	Mean (KY)	Median (KY)	.025 Quantile	.975 Quantile	
Tibet:								
Q3a5-M134	46	.29	10.5	12.3	9.1	1.3	79.6	7.6 $\pm$ 2.3
D1-M15	41	.65	23.2	5.7	3.8	.6	51.7	5.1 $\pm$ 1.8 and 2.3 $\pm$ 1.0 <sup>d</sup>
D3-P47	30	.24	8.6	23.8	8.8	.2	36.2	11.3 $\pm$ 3.7 and 7.0 $\pm$ 4.2 <sup>d</sup>
Kathmandu:								
R1a1-M198	27	.29	10.3	9.5	7.6	2.0	38.4	4.8 $\pm$ 2.1
Q3a5-M134	15	.35	12.6	13.5	4.4	.1	268.6	6.6 $\pm$ 3.6
Newar:								
R1a1-M198	17	.16	5.9	12.0	.3	.0	336.3	1.0 $\pm$ 1.0
Q3a5-M134	14	.17	6.3	5.4	7.9	.0	221.3	7.6 $\pm$ 2.8
Tamang:								
Q3a5-M134	39	.27	9.7	12.6	.6	.0	521.8	10.8 $\pm$ 5.6

NOTE.—Locus *DYS392* is excluded from the analysis for all Q3a5-M134 samples because of allele dropouts.

<sup>a</sup> With a 25-year generation.

<sup>b</sup> Divergence time is based on the SNP-STR coalescence method.<sup>44,55</sup>

<sup>c</sup> Linear expansion time.

<sup>d</sup> Subcluster time estimates.

derivative O3a5-M134 accounting for the majority of samples. O3-M122 is present at low frequencies throughout Central Asia, Siberia,<sup>26</sup> and Pakistan,<sup>23</sup> whereas it is widely distributed in East Asia and among Tibeto-Burman groups from northeastern India.<sup>4,20,22,23</sup> The high percentage of O3a5-M134 in Tamang (86.6%) is comparable to the frequencies observed in Tibeto-Burman-speaking groups from Northeast India (~85%), including Adi, Naga, Apatani, and Nishi.<sup>4,20</sup> The fact that O3a5-M134 is also present (28.9%) as the M117 subclade in Tibet may be indicative of a common ancestry for this language family, as suggested by Su et al.<sup>4</sup> This affinity is also reflected in the CA graph (fig. 4). Notable is the observation that O3a5a-M117 is common in Tibeto-Burman speakers from Southeast Asia.<sup>10</sup>

On the basis of historical records, the Tibeto-Burman people are believed to have originated from Di-Qiang tribes that migrated south from the Yellow River valley in the central plain of East Asia. A central East Asian origin is further supported by the complete absence of Southeast Asian markers O1-M119 and O2-M268 in all populations from the present study. In contrast, van Driem<sup>6,7</sup> argues for Sichuan, in southwestern China, as the homeland of the Tibeto-Burman language family, with subsequent northward migration to the fertile loess plains of the Yellow River basin. He classified the Tibeto-Burman language into an eastern and western group. The former is further subdivided into northern (Sino-Bodic) and southern languages. The Bodic group, one of the two variants of the Sino-Bodic language, is believed to have spread from Gansu with the dispersal of the Majiayao Neolithic culture southward into eastern Tibet, Bhutan, and Sikkim in the 3rd millennium B.C.<sup>7</sup>

Contrary to the Tamang and Tibetans, the Newar and Kathmandu groups exhibit relatively lower frequencies of O3a5-M134, perhaps because of subsequent admixture with Indians, as suggested by the distributions of haplo-

group R and H profiles. Given their geographic proximity and the influences India has historically exercised over Nepal, a significant Indian contribution to the Newar and Kathmandu gene pool is plausible. Admixture analysis further supports this, since Newar and Kathmandu both show high admixture proportions from India (43.4% and 48.6%, respectively), whereas the opposite is observed in Tamang, which has a high Southeast Asian contribution (66.2%) and null contribution from India.

The divergence times obtained for O3a5-M134 with use of the method described by Zhivotovsky et al.<sup>44,55</sup> generated an older age for Tamang (10.8  $\pm$  5.6 KYA), whereas similar values were estimated for Newar (7.6  $\pm$  2.8 KYA), Tibet (7.6  $\pm$  2.3 KYA), and Kathmandu (6.6  $\pm$  3.6 KYA). Whereas the median BATWING values are comparable for the latter three populations, the age estimate for O3a5-M134 in Tamang is extremely recent (0.6 KYA). When the large lower and upper quantiles (0.0 and 521.8 KYA, respectively) are considered, the coalescence time from the BATWING analysis may be subject to inaccuracy. The overall age estimated for haplogroup O3a5-M134 in all four populations (8.1  $\pm$  2.9 KYA) is more recent than its Southeast Asian counterpart (25.36  $\pm$  1.16 KYA).<sup>10</sup> Therefore, the data support an immigration scenario in which this haplogroup dispersed from Southeast Asia into the Himalayan region more recently. The age of haplogroup O3a5-M134 obtained here (8.1  $\pm$  2.9 KYA) is within the range of the age estimated by Su et al.<sup>4</sup> for the same marker (~5–6 KYA), supporting the earlier observation that it may have been introduced to Tibet during the Neolithic expansion from the Yellow River basin in China.<sup>4</sup> However,

**Table 3. Y-STR Data for Haplogroup D1-M15**

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

**Table 4. Y-STR Data for Haplogroup D3-P47**

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

the more ancient age ( $22.1 \pm 4.4$  KYA) obtained for O3a5-M134 individuals that do not form clusters may represent remnants of an earlier Southeast Asian dispersal that arrived in the Himalayas before the Neolithic period. Evidence of an early migration is provided by archaeological sites at Xiao Qaidam in the Central Qaidam basin and Chusang near Lhasa, dating to  $\sim 33.0 \pm 4.4$  KYA and  $21.7 \pm 2.1$  KYA, respectively, which points to a late Paleolithic entry and occupation of the plateau by modern humans.<sup>1</sup>

The presence of haplogroups N and Q in Tibet, haplogroups which are commonly found in Siberia and Mongolia,<sup>34,36,37,56-58</sup> also reflect gene flow from the north and/or east. A genetic affinity between Tibet and the northern populations of Siberia and Mongolia is consistent with previous mtDNA studies.<sup>14</sup> Recently, an extensive phylogeographic survey<sup>59</sup> of haplogroup N-M231 indicated that this marker originated in Southeast Asia but failed to spread into South Asia.<sup>22,23</sup> Our results are consistent with that interpretation.

#### *Central Asian Signature in the Himalayas*

Haplogroup D lineages have been reported at high frequencies in both Japan and in previous studies of Tibet.<sup>3,4,15,35,60</sup> These two populations exhibit contrasting distribution of D-M174 subclades. Tibet is characterized by D1-M15, whereas Japan is represented by haplogroup D2-M55.<sup>3,4,35</sup> In addition to D1-M15, the Tibetan population also possesses high frequencies of D3-P47<sup>35</sup> but displays relatively low proportions of D\*-M174.<sup>8</sup> The latter is reported in high proportion in Jarawa and Onge males from the Andaman Islands, consistent with an early migration via the southern coastal route.<sup>61-63</sup>

In the present study, 50.6% of the Tibetan Y chromosomes are represented by haplogroup D lineages, which, in descending frequencies, are D1-M15 (28.2%), D3-P47 (18.6%), and D\*-M174 (3.8%). The close genetic affinity portrayed between the Tibetans and Japanese in the CA plot (fig. 4) may suggest an ancient common ancestry, in agreement with earlier observations.<sup>35,64,65</sup> However, unlike the Paleolithic founder for the Japanese DE-YAP chromosomes, the DE-YAP haplogroup in Tibet appears Neolithic in origin. Previous studies of Y-chromosome polymorphisms in Tibet argue for a Central Asian contribution to account for the haplogroup DE-YAP chromosomes in their gene pool.<sup>3,4,16</sup> Divergence time estimations for D1-M15 ( $5.1 \pm 1.8$  KYA) and D3-P47 ( $11.3 \pm 3.7$  KYA) in Tibet suggest a bipartite penetration, on different occasions, into the Tibetan plateau. Nevertheless, the ancient components of these two lineages that do not form clusters with either 7 (D3-P47) or 11 (D1-M15) *DYS392* repeats were found to possess ages of  $13.8 \pm 5.0$  KYA and  $27.3 \pm 9.8$  KYA, re-

spectively, corroborating the earlier report of older shared ancestry between the Japanese and Tibetan populations.

On the basis of Y-chromosome and linguistic analyses, Su et al.<sup>4</sup> proposed that the Baric people, a branch of the Tibeto-Burman language family, were the first inhabitants of the Himalayan region, whereas Tibetans, who belong to the Bodic branch, were recent immigrants who arrived subsequent to admixture with Central Asians. Complete absence of haplogroup DE-YAP in Tamang, Newar, and Kathmandu adds support to the peopling of Nepal and Tibeto-Burman northeastern India by the Baric group from East Asia, who were shown to lack the haplogroup DE-YAP.<sup>4,20</sup>

#### *Gene Flow from India*

The distribution of haplogroup R1a1-M198 has been suggested elsewhere to be associated with the spread of the Kurgan culture from the Central Asian Steppes into northern Europe.<sup>33</sup> This haplogroup is also found at high frequencies in Central Asia<sup>26,66</sup> and South Asia<sup>18,23,67</sup> but is rarely observed in East Asia.<sup>3</sup> Quintana-Murci et al.<sup>67</sup> attributed the high proportion of R1a1-M198 in India to a Neolithic demic diffusion of Indo-European speakers from southwestern Asia. However, a recent study revealed an early Holocene expansion of R1a1-M198 from the northwestern region of India before the arrival of the Indo-European speakers.<sup>23</sup> Although a much earlier entry of R1a1-M198 into India has been postulated on the basis of its high microsatellite variance and estimated age (14 KYA), multiple events resulting from subsequent migrations from southwestern Asia may have also contributed.<sup>23</sup>

In the present study, haplogroup R1a1-M198 is found at low frequencies in Tibetan (1.9%) and Tamang (4.4%) populations, whereas higher percentages are observed in the Kathmandu (35.1%) and Newar (25.7%) populations, implying significant gene flow from the Indian subcontinent for the latter two populations, particularly from northern India, where its presence is high (40%).<sup>18</sup> The CA also suggest strong Indian influence, since Kathmandu and Newar group closer to the Indian populations (fig. 4). In addition, the CA graph (fig. 4) displays close genetic association of Central Asian populations (Kyrgyz and Karalpak) with Newar and Kathmandu groups, in agreement with the admixture results (56.6% and 51.4%, respectively). Microsatellite-based age estimates for haplogroup R1a1-M198 in Kathmandu ( $4.8 \pm 2.1$  KYA) and Newar ( $1.0 \pm 1.0$  KYA) are considerably younger than in India (14 KYA).<sup>23</sup> Taken together, these results indicate that R1a1-M198 most likely arrived in Nepal from Central Asia and/or northern India. When the obstacle presented by the Himalayas is considered, it is possible that the penetration of R1a1-M198 into Nepal occurred indirectly from

**Table 5. Y-STR Data for Haplogroup O3a5-M134**

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

**Table 6. Y-STR Data for Haplogroup R1a1-M198**

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Central Asia, possibly as Indo-Aryans migrated north from India ~3–3.5 KYA.

With the exception of the Newar, in whom R2-M124 accounts for 25.7% of Y chromosomes, the other Nepalese populations and Tibetans exhibit lower frequencies similar to those observed in India (<10%).<sup>21,23</sup> Furthermore, Newar is the only population that exhibits R1b1a-M269 (10.6%), despite the low levels found throughout South Asia.<sup>23</sup> The relatively high proportion of these two markers in the Newar may be the result of genetic drift.

Previous studies have shown that haplogroup H\*-M69 and its subclades H1-M52 and H2-APT originated within India.<sup>18,23,68</sup> Haplogroup H1-M52 is found at higher frequencies in southern India,<sup>26,69</sup> with Punjab in the northwest displaying the lowest frequency.<sup>18</sup> In the present study, H1-M52 occurs in the Newar and Kathmandu populations at 6.1% and 11.7%, respectively. The indigenous Indian marker H\*-M69 is also found at a low frequency (1.9%) in the Tibetan gene pool, which may represent a limited gene flow from India. In historical times, the spread of Buddhism from India to Tibet may account for the presence of signature Indian chromosomes in the region.

With further molecular resolution achieved within haplogroup C in the form of mutation C5-M356, a portion of the previously classified C\* (xC3-M217) paragroup are now characterized as representatives of this newly defined marker.<sup>23</sup> Its occurrence in Newar (2.6%) is double the frequency observed in Kathmandu (1.3%) and India (1.4%)<sup>23</sup> and most likely reflects gene flow from India into Nepal. Whereas haplogroups C3-M217 and C\* (xC3-M217) are prevalent in East Asia and Central Asia,<sup>26,34,64,65,70</sup> the former has not yet been detected in India. Therefore, the presence of C3-M217 in Kathmandu (2.6%) and Tibet (2.6%) is most likely attributable to gene flow from Mongolia and/or Siberia; the marker is found at its highest frequencies (55.0%) in Mongolia.<sup>26,70</sup>

#### *The Himalayas: Disproportional Barrier of Gene Flow*

The Himalayas stand between the Tibetan plateau and India, creating a formidable obstruction to population movements, as reflected in the Y-haplogroup composition of Tibet, where haplogroup R lineages, a signature of Indian penetration, are negligible. Similarly, haplogroup L-M20

did not penetrate upward from India. Despite this obstacle, since the 7th century A.D., India has exercised a strong influence on Tibet, especially with respect to the spread of Buddhism. However, historical records document the spread of Buddhism from northwestern India to Pakistan and Afghanistan and, later, into East Asia along the northern Silk Road. In fact, the younger time estimates calculated for the D-M174 chromosome derivative D1-M15 ( $2.3 \pm 1.0$  KYA) may be the result of gene flow connected to Silk Road migrations from Central Asia.

Conversely, Su et al.<sup>4</sup> proposed that the peopling of Nepal by the Baric-speaking group(s) occurred when they crossed the Himalayas from the Tibetan plateau. This scenario is consistent with the extensive sharing of haplogroup-O lineages among the Tibeto-Burman speakers in the region. The results of the present study suggest that population movements across the Himalayas from Tibet to Nepal have occurred periodically, whereas dispersals in the opposite direction from India into the Tibetan plateau have been limited. The observed preferential gene flow southward could be accounted for, at least in part, by the unfavorable conditions experienced on the Tibetan side of this vast mountain range, where higher altitudes, colder climates, and physiological stress would have discouraged human migrations to and settlements in the region.

#### **Conclusion**

In conclusion, the array of data presented supports the Tibeto-Burman affinities of the Himalayan populations examined in the present study. Whereas Tibet and Nepal share high levels of haplogroup O3a5-M134 (especially O3a5a-M117), commonly found in East Asia, they display an absence of Southeast Asia-specific markers (O1-M119 and O2-M268). In combination with the relatively younger age generated for O3a5-M134, East and Central Asian ancestry for the Himalayan groups is likely. However, subsequent gene flow from the Indian subcontinent into Nepal is signaled by the presence of haplogroups R and H, as well as from the results of the admixture analysis, in which Indian contribution is >40% in both Newar and Kathmandu. In contrast, Indian influence in Tamang is null.

The Tibetan gene pool reflects significant contributions from East and/or Southeast Asia. The elevated presence of haplogroup D-M174 chromosomes in Tibet and Japan indicates prolonged geographic isolation of these regions, since it is sparsely distributed in the rest of East Asia. The

**Table 7. Admixture Proportions**

Parental Group	Proportion ( $\pm$ SE) of Parental Group in			
	Tibet	Tamang	Newar	Kathmandu
Northeast Asia (P1)	.089 $\pm$ .305	.051 $\pm$ .561	.000 $\pm$ .305	.000 $\pm$ .203
Southeast Asia (P2)	.604 $\pm$ .254	.662 $\pm$ .467	.000 $\pm$ .254	.013 $\pm$ .169
Central Asia (P3)	.262 $\pm$ .315	.287 $\pm$ .579	.566 $\pm$ .315	.514 $\pm$ .210
India (P4)	.045 $\pm$ .206	.000 $\pm$ .379	.434 $\pm$ .206	.486 $\pm$ .137



age of Tibetan D1-M15 and D2-M55 lineages are considerably younger than the Paleolithic Japanese D2-M55, indicative of a recent bottleneck. Haplogroups Q-M242, C-M216, and N-M231 in Himalayan populations point to gene flow from Mongolia and Siberia, underscoring the geographic accessibility of the Tibetan plateau for human dispersals from the north. The Himalayas served as a strong barrier to gene flow from the south into the Tibetan plateau, although the same is not true for population movements occurring in the opposite direction. This preferential migrational direction may be associated with the physiological stress imposed on emigrants from lower altitudes.

## Acknowledgments

We gratefully acknowledge Laisel Martinez, Sheyla Mirabal, Charlie Haz, Jason Somarelli, and Tanya Simms for their technical assistance. We thank all sample donors and those who helped with the sample collection. T.G. is a recipient of the Tibetan Fulbright scholarship administered by the Tibet Fund, NY.

## References

- Aldenderfer M, Yinong Z (2004) The prehistory of the Tibetan plateau to the seventh century A.D.: perspectives and research from China and the West since 1950. *J World Prehist* 18:1–55
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, et al (1999) Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last Ice Age. *Am J Hum Genet* 65:1718–1724
- Su B, Xiao C, Deka R, Seielstad T, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, et al (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 107:582–590
- van Driem G (2001) Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan region containing an introduction to the symbiotic theory of language. Brill, Leiden, The Netherlands
- van Driem G (2002) Tibeto-Burman phylogeny and prehistory: languages, material culture, and genes. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. McDonald Institute for Archaeological Research, Cambridge, United Kingdom, pp 233–249
- van Driem G (2005) Tibeto-Burman vs Indo-Chinese: implications for population geneticists, archeologists and prehistorians. In: Sagart L, Blench R, Sanches-Mazas A (eds) The peopling of East Asia: putting together archaeology, linguistics and genetics. Routledge Curzon, London, pp 81–106
- Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, Xiao C, Jin J, Su B, et al (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* 74:856–865
- Deng W, Shi B, He X, Zhang Z, Xu J, Li B, Yang J, Ling L, Dai C, Qiang B, et al (2004) Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *J Hum Genet* 49:339–348
- Shi H, Dong Y-I, Wen B, Xiao C-J, Underhill PA, Shen P-d, Chakraborty R, Jin L, Su B (2005) Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet* 77:408–419
- Matisoff JA (1991) Sino-Tibetan linguistics: present state and future prospects. *Annu Rev Anthropol* 20:469–504
- Nie M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 10:927–943
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, et al (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763–11768
- Torrioni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 93:189–199
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkinson T, et al (1997) The geographic distribution of Y chromosome variation. *Genetics* 145:787–805
- Qian Y, Qian B, Su B, Yu J, Ke Y, Chu Z, Shi L, Lu D, Chu J, Jin L (2000) Multiple origins of Tibetan Y chromosomes. *Hum Genet* 106:453–454
- Mukherjee N, Nebel A, Oppenheim A, Majumder PP (2001) High-resolution analysis of Y-chromosomal polymorphisms reveals signatures of population movements from Central Asia and West Asia into India. *J Genet* 80:125–135
- Kivisild T, Roosti S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, et al (2003) The genetic heritage of earliest settlers persist in both the Indian tribal and caste populations. *Am J Hum Genet* 72:313–332
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, et al (2003) Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* 13:2277–2290
- Cordaux R, Weiss G, Saha N, Stoneking M (2004) The north-east Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol* 21:1525–1533
- Cordaux R, Aunger R, Brentley G, Nasidze I, Sirajuddin SM, Stoneking M (2004) Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* 14:231–235
- Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaxmi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, et al (2006) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci USA* 103:843–848
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow C-ET, Lin AA, Mitra M, Sil SK, Ramesh A, et al (2006) Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 78:202–221
- Kraaijenbrink T, van Driem GL, Opgenort JRML, Tuladhar NM, de Knijff P (2006) Allele frequency distribution for 21 autosomal STR loci in Nepal. *Forensic Sci Int* (electronically published ahead of print March 15, 2006; accessed March 13, 2007)
- Parkin EJ, Kraaijenbrink T, Opgenort JRML, van Driem GL, Tuladhar NM, de Knijff P, Jobling MA (2006) Diversity of 26-

- locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. *Forensic Sci Int* 166:176–181
26. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Smith JB, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249
  27. Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ (2006) Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered* 61:132–143
  28. Krings MA, Salem A-eH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermaan G, et al (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or barrier for migration? *Am J Hum Genet* 64:1166–1176
  29. Stringer CB, Andrew P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–1268
  30. Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 74:532–544 (erratum 74:788)
  31. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of native American Y chromosomes into the Americas. *Mol Biol Evol* 21:164–175
  32. Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 28:458–472
  33. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, et al (2000) The genetic legacy of Palaeolithic *Homo sapiens sapiens* in extant Europeans: a Y-chromosome perspective. *Science* 290:1155–1159
  34. Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628
  35. Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51:47–58
  36. Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurler ME, et al (2006) Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 172:2431–2439
  37. Karafet TM, Lansing JS, Redd AJ, Reznikova S, Watkins JC, Surata SPK, Arthawiguna WA, Mayer L, Bamshad M, Jorde LB, et al (2005) Balinese Y chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter gatherers, Austronesians farmers, and Indian traders. *Hum Biol* 77:93–114
  38. Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer MF, Yunis JJ, Yunis EJ, Damodarn C, Marinez de Pancorbo M, Caeiro JL, et al (2002) Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 23:3346–3356
  39. Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962
  40. Martinez L, Reategui EP, Fonesca LR, Sierra-Montes JM, Terros MC, Pereira Simon S, Herrera RJ (2005) Superimposing polymorphism: the case of a point mutation within a polymorphic *Alu* insertion. *Hum Hered* 59:109–117
  41. Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y chromosomal binary haplogroups. *Genome Res* 12:339–348
  42. Cadenas AM, Regueiro M, Gayden T, Singh N, Zhivotovsky LA, Underhill PA, Herrera RJ (2007) Male amelogenin dropouts: phylogenetic context, origins and implications. *Forensic Sci Int* 166:155–163
  43. Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* 5:1759–1766
  44. Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, et al (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74:50–61
  45. Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, et al (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 68:990–1018
  46. Di Rienzo, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91:3166–3170
  47. Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 62:1171–1179
  48. Wilson I, Weale M, Balding D (1998) Genealogical inference from microsatellite data. *Genetics* 150:499–510
  49. Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, et al (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 70:1197–1214
  50. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GI, Lillie AS, Roseman CC, Lin AA, Prince K, et al (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–148
  51. Rohlf F (2002) NTSYSpc. Exter Publishing, Setauket, NY
  52. Long JC, Williams RC, McAuley JE, Meids R, Partel R, Tregellas M, South SF, Rea AE, McCormick B, Iwaniec U (1991) Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84:141–157
  53. Perez-Miranda AM, Alfonso-Sanchez MA, Pena JA, Herrera RJ (2006) Qatari DNA variation at a crossroads of human migrations. *Hum Hered* 61:67–79
  54. Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, Moysse-Faurie C, Rutledge RB, Schiefenhoefel W, Gil D, et al (2006) Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol* 23:2234–2244
  55. Zhivotovsky LA (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol Biol Evol* 18:700–709

56. Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, et al (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73: 524–539
57. Seielstad M, Yuldasheva N, Singh N, Underhill PA, Oefner P, Shen P, Wells RS (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* 73:700–705
58. Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I, Dorzhu C, Luzina F, Mixcicka-sliwka D, Zakharov I (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet* 118:591–604
59. Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumäe A, et al (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* 15:204–211
60. Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, Cavalli Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62
61. Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, Hagelberg E (2003) Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol* 13:86–93
62. Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L (2005) Reconstructing the origin of Andaman Islanders. *Science* 308:996
63. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, et al (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036
64. Tajima A, Pan IH, Fucharoen G, Fucharoen S, Matsuo M, Tokunaga K, Juji T, Hayami M, Omoto K, Horai S (2002) Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. *Hum Genet* 110:80–88
65. Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49:187–193
66. Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am J Hum Genet* 71:466–482
67. Quintana-Murci L, Krausz C, Zerjal T, Sayar H, Hammer MF, Mehdi Q, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, et al (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 68: 537–542
68. Pandya A, King TE, Santos FR, Taylor PG, Thangaraj K, Singh L, Jobling MA, Tyler-Smith C (1998) A polymorphic human Y-chromosomal G→A transition found in India. *Ind J Hum Genet* 4:52–61
69. Ramana GV, Su B, Jin L, Singh L, Wang N, Underhill PA, Chakraborty R (2001) Y chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur J Hum Genet* 9:695–700
70. Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungereel G, Chae GT, Han H, Jia GJ, Tokunaga K, et al (2005) Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene* 346:63–70