Am. J. Hum. Genet. 75:1113–1116, 2004

Report

The Genetic or Mythical Ancestry of Descent Groups: Lessons from the Y Chromosome

Raphaëlle Chaix,[1] Frédéric Austerlitz,[3] Tatyana Khegay,[4] Svetlana Jacquesson,[5]
Michael F. Hammer,[6] Evelyne Heyer,[1] and Lluís Quintana-Murci[2]

[1]Unité d'Eco-Anthropologie, Centre National de la Recherche Scientifique (CNRS) UMR 5145/Université Paris 7, Musée de l'Homme, and [2]CNRS URA 1961, Unit of Molecular Prevention and Therapy of Human Diseases, Institut Pasteur, Paris; [3]Laboratoire Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, Orsay, France; [4]Institute of Immunology, Academy of Sciences, and [5]Institut Français d'Etudes sur l'Asie Centrale, Tashkent, Uzbekistan; and [6]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson

Traditional societies are often organized into descent groups called "lineages," "clans," and "tribes." Each of these descent groups claims to have a common ancestor, and this ancestry distinguishes the group's members from the rest of the population. To test the hypothesis of common ancestry within these groups, we compared ethnological and genetic data from five Central Asian populations. We show that, although people from the same lineage and clan share generally a recent common ancestor, no such common ancestry is observed at the tribal level. Thus, a tribe might be a conglomerate of clans who subsequently invented a mythical ancestor to strengthen group unity.

Many societies described as "traditional" are organized into so-called "descent groups": the population is divided into tribes, the tribe into clans, and the clan into lineages (Ghasarian 1996; Maquet 2003). These descent groups usually claim to have a distinctive common ancestor, and this ancestry is traced back through the male or female line. Patrilineal populations, in which a father transmits his patronymic and his group affiliation to his offspring, are roughly twice as frequent as matrilineal populations (Burton et al. 1996). The extent to which the claimed ancestry of descent groups is real or socially constructed remains to be elucidated. To distinguish between these two hypotheses, we have taken advantage of the paternal inheritance of the nonrecombining region of the Y chromosome (Jobling and Tyler-Smith 2003). If descent-group organization in patrilineal societies corresponds to a genetic reality, there should be a correlation between Y-chromosome diversity and group affiliation.

We compared ethnological (descent-group affiliation) and Y-chromosome data from 247 men of five different patrilineal populations from the Karakalpakia region of Uzbekistan: Kazakhs ($n = 50$), Turkmen ($n = 51$), Uzbeks ($n = 40$), Karakalpaks Qongïrat (referred to as "Qongïrat") ($n = 53$), and Karakalpaks On Tört Uruw (referred to as "On Tört Uruw") ($n = 53$). Each sample was composed of unrelated healthy donors from whom appropriate informed consent was obtained. These five populations consist of former pastoral nomads who have recently adopted a seminomadic way of life, balancing between agriculture and pastoralism. The rigor with which these Central Asian peoples have maintained their patrilineal social organization has few parallels worldwide. Some of their oral-tradition genealogies comprise up to 60 generations (Krader 1966). The populations are organized into well-defined lineages, clans, and tribes—with the exception of the Qongïrat and the Uzbeks, who have mostly abandoned social considerations about the lineage level. In some populations, intermediary levels of descent groups between the tribe and the clan (phratries and subtribes in the Khazaks and clan groups in the Uzbeks) are observed, but, since they are population-specific, they were not taken into account in this study.

To assess Y-chromosome variation, we defined Y-chromosome haplotypes by analyzing 12 Y-linked STR loci (*DYS19, DYS385I, DYS385II, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426*, and *DYS439*), as described elsewhere (Redd et al. 2002). The highly polymorphic nature of STR markers allowed us to investigate population genetic struc-

Received August 3, 2004; accepted for publication September 9, 2004; electronically published October 1, 2004.

Address for correspondence and reprints: Dr. Evelyne Heyer, Unité d'Eco-Anthropologie, Musée de l'Homme, 17 place du Trocadéro, 75 116 Paris, France. E-mail: heyer@mnhn.fr

0002-9297/2004/7506-0016$15.00

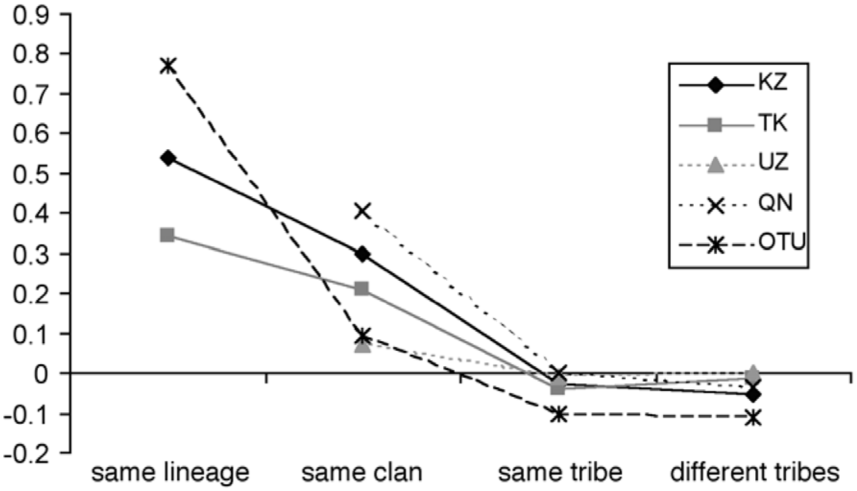

Figure 1 Mean genetic kinship coefficient for each ethnological class of the five populations examined in this study. KZ = Kazakhs; TK = Turkmen; UZ = Uzbeks; QN = Qongïrat; OTU = On Tört Uruw.

ture at a very small scale, which facilitated our focus on the descent-group relationships. The software SPAGeDi, version 1.1 (Hardy and Vekemans 2002), was used to estimate a mean kinship coefficient (Loiselle et al. 1995) from Y-haplotype data from each population. This software actually estimates a relative kinship coefficient standardized by the mean probability of identity by descent observed in the sample. We computed the average kinship coefficient of the individuals who belonged to each ethnological distance class: same lineage, same clan but different lineages, same tribe but different clans, and different tribes. Using the permutation approach implemented in SPAGeDi, we computed the level of significance (*P* value) for each kinship coefficient—the null hypothesis being a nonexcess of kinship within the descent group, compared with the global population. The individuals with missing genealogical data at a given descent-group level (6 for the tribe affiliation, 36 for the clan affiliation, and 66 for the lineage affiliation) were not taken into account in the estimation of the kinship coefficient at this level.

The mean kinship coefficients for people from the same lineage exhibited high values in the three populations that have maintained the lineage level: 0.54 ($P < .001$) for the Kazakhs, 0.34 ($P < .01$) for the Turkmen, and 0.77 ($P < .001$) for the On Tört Uruw. People who belong to the same lineage are therefore significantly much more related than individuals selected at random in the global population. In other words, people from the same lineage share a common ancestor who is more recent than the common ancestor of the whole population. Consequently, this observation supports the oral tradition of a recent common ancestry at the lineage level.

The kinship coefficients for people from the same clan (but from different lineages) were lower than those observed at the lineage level but still significant in three populations: 0.30 ($P < .01$) for the Kazakhs, 0.21 ($P < .001$) for the Turkmen, and 0.40 ($P < .001$) for the Qongïrat. Thus, in these three populations, men from the same clan are significantly more genetically related than men selected randomly from the population, which again supports the claim of common ancestry at the clan level. However, two populations (the Uzbeks and the On Tört Uruw) did not exhibit a high kinship coefficient at this level (0.07 [$P > .05$] and 0.09 [$P > .05$], respectively), which highlights how differences in historical and cultural lifestyles between populations may account for the heterogeneity in social structure. In this context, the Uzbek population illustrates well the historical-genetic correlation. They always exhibit low kinship coefficients in comparison with the other populations and never show kinship coefficients >0.08 (fig. 1). This observation is clearly supported by the historically well-documented progressive dissolution of their descent groups after the early loss of their nomadic way of life in the 17th and 18th centuries (Jacquesson 2002). Today, they have forgotten their lineages, and descent groups no longer play their previous social roles.

Finally, the genetic kinship coefficients of people from the same tribe (but from different clans and lineages) were all slightly negative (Kazakhs: −0.02 [$P > .05$]; Turkmen: −0.04 [$P < .001$]; Uzbeks: −0.07 [$P > .1$]; Qongïrat: −.0011 [$P > .1$]; and On Tört Uruw: −0.10 [$P < .01$]). Their genetic kinship is not significantly higher than the mean kinship of the whole population, and, in two populations, it is even significantly lower. Thus, our results show that a tribe does not correspond to a real genetic entity, and its claimed common ancestry is likely to be socially constructed.

There are two processes that lead to the birth of the so-called “descent groups”: either a subgroup increases in size and becomes independent from the original group or an existing group assimilates other groups, creating a bigger entity. The oral tradition in these Central Asian populations supports the former scenario, in which a tribe gives birth to its clans and a clan to its lineages by fission events. Our thorough comparison of the ethnological and genetic data also supports this first scenario at the lineage level and, in three of the five populations studied, at the clan level. Conversely, we demonstrate here that the tribe level follows the second model, in which a tribe is a conglomerate of clans of diverse origins. This phenomenon could be due to clan fusions that resulted from economic, demographic, and/or military considerations (Krader 1966). Such fusion events were indeed observed in the 19th century in Central Asian populations (Krader 1966), as well as in other populations like the South Sinai Bedouins (Baily 1977). After a fusion event, the genealogy was frequently amended to restore the filiation of the descent groups by the creation of a fictitious common ancestor to enhance group unity (Krader 1966). In this context, the observation that individuals of the same tribe are sometimes less related than the average population (i.e., significantly negative kinship coefficients at the tribal level are observed in the Turkmen and the On Tört Uruw) suggests that these fusions could involve clans with an even more remote common ancestor than that expected at random.

Table 1

Estimates of the Age of the Descent Groups

Descent Group and Population	No. of Descent Groups[a]	No. of Generations (95% CI[b])	Age in Years (95% CI)
Clan (case A):			
KZ	2	101 (43–355)	3,042 (1,304–10,648)
TK	5	440 (188–1,539)	13,194 (5,655–46,181)
QN	3	457 (196–1,598)	13,697 (5,870–47,940)
Clan (case B):			
KZ	1	20 (9–69)	595 (255–2,083)
TK	4	102 (44–356)	3,051 (1,307–10,677)
QN	2	35 (15–123)	1,058 (454–3,704)
Lineage (case A):			
KZ	2	37 (16–130)	1,114 (477–3,899)
TK	3	17 (7–60)	516 (221–1,806)
OTU	4	202 (86–706)	6,052 (2,594–21,181)
Lineage (case B):			
KZ	2	14 (6–48)	415 (178–1,451)
TK	3	17 (7–60)	516 (221–1,806)
OTU	4	13 (6–46)	397 (170–1,389)

NOTE.—In case A, all haplotypes were included in the analysis, whereas, in case B, the analysis was restricted to the individuals who belonged to the modal haplogroup of the descent group. KZ = Kazakhs; TK = Turkmen; QN = Qongïrat; OTU = On Tört Uruw.

[a] We excluded the descent groups for which the number of individuals was <3.

[b] CIs represent uncertainty in the mutation rate.

To estimate the age of the descent-group levels for which genetic data support the existence of a real common ancestor, we calculated the average squared difference (Goldstein et al. 1995; Slatkin 1995) in allele size among all current chromosomes and the inferred ancestral haplotype, averaged over loci. This statistic was computed for the lineage and clan levels with the software YMRCA (Stumpf and Goldstein 2001), with the assumption of a mutation rate of 2.1×10^{-3} (95% CI $0.6–4.9 \times 10^{-3}$) (Heyer et al. 1997; Kayser et al. 2000) and a generation time of 30 years (Tremblay and Vezina 2000). The confidence limits for the age estimates were calculated, with consideration of the 95% CI of the mutation rate estimate. The genealogical depth of the three populations in which clans followed the scenario of recent common ancestry (table 1) was as follows: 101 generations (95% CI 43–355) for the Kazakhs, 440 generations (95% CI 188–1,539) for the Turkmen, and 457 generations (95% CI 196–1,598) for the Qongïrat. We performed a second estimation in which the analysis was restricted to the individuals who belonged to the most common haplogroup in the descent group (haplogroups were defined by analysis of unique-event polymorphisms [data not shown]). This procedure excluded ~17% of the individuals. We then estimated ages of 20 (95% CI 9–69), 102 (95% CI 44–356), and 35 (95% CI 15–123) generations for the Kazakhs, the Turkmen, and the Qongïrat, respectively. These estimated ages were much lower than those obtained using all individuals of each descent group. This is due to the sensitivity of the dating method to the presence of “outsider” haplotypes introduced by recent immigration. By excluding the individuals who belonged to an “outsider” haplogroup, we minimized the presence of haplotypes introduced into the population through recent migration. The high ages still observed in the Turkmen clans are due to the presence of three individuals who exhibited the modal haplogroup but a very different haplotype (i.e., more than five steps away from the ancestral core haplotype, whereas the remaining haplotypes matched the ancestral haplotype or were one-step derivatives). It is difficult to discern if these three individuals are recent immigrants who have adopted the clan affiliation or if such a difference in STR haplotype within the modal haplogroup is due to naturally occurring STR divergence over time. As to the lineage level, we obtained mean lineage ages of 37 (95% CI 16–130), 17 (95% CI 7–60), and 202 (95% CI 86–706) generations for the Kazakhs, the Turkmen, and the On Tört Uruw, respectively, when all individuals were included in the analysis, and values of 14 (95% CI 6–48), 17 (95% CI 7–60), and 13 (95% CI 6–46) generations, respectively, when only individuals who belonged to the modal haplogroup were considered (10% of the individuals were excluded). It is worth noting that the mean age of the lineage level (immigrants excluded) is ~15 generations, a value that is of the same order of magnitude as the ages claimed by oral tradition (5–10 generations, depending on the population) (Krader 1963*b*; Ghasarian 1996; Jacquesson 2002).

In conclusion, our results show that, although people from the same lineage and clan share generally a recent common ancestor, no such common ancestry is observed at the tribal level, which is likely to be socially constructed. Further studies of other traditional societies are needed to evaluate the extent to which the pattern observed in Central Asia can be extrapolated to other world regions. In any case, our study demonstrates that the resolution of modern genetic markers allows us to make historical investigations at the scale of kinship groups and to practice a kind of "ethnogenetics." It explores the roots of the descent groups in patrilineal populations and reveals the mythical nature of the genealogical links between people of a tribe and their claimed ancestor. As anthropologist Lawrence Krader pointed out, "genealogy is at once ideology and history" (Krader 1963*a*, p. 157).

Acknowledgments

We warmly acknowledge all the blood donors who kindly agreed to participate in this study, as well as Patricia Baldacci and Jacqui Shykoff for stimulating remarks. This work was supported by the CNRS grant project OHLL and the European Science Foundation project OMLL. Financial support was also provided by North Atlantic Treaty Organization collaborative linkage grant LST.CLG.977507 (to M.F.H. and L.Q.-M.).

References

Baily Y (1977) The Bedouins in Sinai. In: Zohar E (ed) The desert. Reshafim, Tel Aviv, pp 240–247

Burton ML, Moore CC, Whiting JWM, Kimball Romney A (1996) Regions based on social structure. Curr Anthropol 37:87–123

Ghasarian C (1996) Introduction à l'étude de la parenté. Editions du Seuil, Paris

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Jacquesson S (2002) Parcours ethnographiques dans l'histoire des deltas. In: Jacquesson S, Fourniau V (eds) Karakalpaks et autres gens de l'Aral: entre rivages et déserts. Vol 10. Edisud, Tashkent–Aix-en-Provence, pp 51–92

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598–612

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Krader L (1963*a*) Peoples of Central Asia. Vol 26. Indiana University Publications, Bloomington

——— (1963*b*) Social organization of the Mongol-Turkic pastoral nomads. Vol 20. Indiana University Publications, Bloomington

——— (1966) Peoples of Central Asia, 2nd ed. Vol 26. Indiana University Publications, Bloomington

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82:1420–1425

Maquet J (2003) Clan. Encyclopedia Universalis, Paris

Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF (2002) Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 130:97–111

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Stumpf MP, Goldstein DB (2001) Genealogical and evolutionary inference with the human Y chromosome. Science 291: 1738–1742

Tremblay M, Vezina H (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am J Hum Genet 66:651–658