

## **Ethno-anthropological and genetic study of the Yaghnobis: an isolated community in Central Asia. A preliminary study**

**Elisabetta Cilli, Paolo Delaini, Birgit Costazza, Laura Giacomello, Antonio Panaino & Giorgio Gruppioni**

*Dipartimento di Storie e Metodi per la Conservazione dei Beni Culturali, Università di Bologna, Via degli Ariani 1, 48121 Ravenna, Italy*

e-mail: elisabetta.cilli@unibo.it

The Yaghnob Valley (Fig. 1), situated in the Northern Tajikistan, represents an important, geographically isolated area that maintains a genetic, linguistic, cultural and environmental heritage, unique to the world. The Yaghnobis inhabit the vast mountains of this valley (2200 - 2800 metres above sea-level) and live in relative, geographical isolation due to the high mountains surrounding their homeland and the lack of roads in the valley (it takes two days by mule, or on foot, to cover the mule tracks that connect the last stretch of paved road at Maryib, suitable for vehicles, to the last Yaghnobi village).

They have been identified as descendants of the Sogdians, the ancient travelling merchants of the Silk Route, who, according to local traditions, took refuge in this valley after the Arab invasion in the VIII Century AD and the fall of the cities of Samarkand and Panjikent, but most probably were already located in such a remote area since earlier times, as the remarkable archaeological structures found in the close Valley of Matcha demonstrate. The Silk Road represents the most important and famous trade route, that played a pivotal role in the commerce and history of Asia for almost 3000 years.

The Yaghnobi people constitute a linguistic minority and speak an Eastern Iranian language (a branch of the Indo-European family) called "Yaghnobi", that is mainly spoken but not written and is closely related to the language of the ancient Sogdians. If we consider that the Sogdian language is, by now, documented only in manuscripts and archaeological testimonies we can appreciate the importance of a population that

still speaks a dialect variety of a dead language. The Yaghnobi community, which welcomed us during our study, live in small villages built on outcrops at high altitude and are isolated eight months of the year due to snow. Yaghnobi villages are poor, pastoral communities that raise sheep and keep dogs as guardians and for herding animals (Panaino, 2008).

In the 1970's, the Yaghnobi community (3000 - 4000 people) was displaced by the Soviet authorities and delocalised to a cotton-picking area (cotton is the most significant crop in the country) very far from their original settlement. This displacement, the change in their daily customs and child labour in the cotton fields caused a state of malnutrition and an outbreak of epidemic diseases. A few Yaghnobis (approximately 300), returned to their original lands in 1977 and today it is estimated that Yaghnobis in the valley number in the order of 500 people, dispersed throughout 16 villages (Delaini, 2008).

At the beginning of the 1990s, the collapse of the old Soviet system of assistance left mountain communities, such as the Yaghnobi, lacking in education, medical assistance (including a vaccination programme) and the sustenance of their agricultural and sheep farming work. A fifteen-year civil war resulted in keeping the community in a state of isolation, while the Tajikistan State became one of the poorest countries in the world.

The principal aims of the "Yaghnob Valley" Mission, organised by the University of Bologna in collaboration with the IsIAO and the Italian Foreign Ministry, are to study and preserve the ethno-linguistic and genetic heritage of the



**Fig. 1 - Yaghnob Valley in Tajikistan, topographic map. (July 97). In UNEP/GRID-Arendal Maps and Graphics Library, modified. Retrieved January 15, 2010 from [http://maps.grida.no/go/graphic/tajikistan\\_topographic\\_map](http://maps.grida.no/go/graphic/tajikistan_topographic_map)**

Yaghnobis, along with the conservation of their environment and historical traditions. Our project is based not only on historical, anthropological and linguistic research but also on sustaining, supporting and facilitating the Yaghnobi community in rediscovering their traditions, giving education to their children and receiving basic medical assistance.

Within this context, we have focussed our attention on the genetic analysis of the Yaghnobi people. The aim of the genetic analysis is to study the genetic legacy of Yaghnobis, to establish if they symbolise a genetic, other than a geographic and linguistic, isolation and also to study the relationship between the Yaghnobis and their neighbours. In order to answer these questions, mitochondrial DNA and Y-chromosome variation must be analysed. To date, only the

mitochondrial DNA analysis has been conducted and so, here, we present a preliminary preview of these results.

Buccal swabs together with anthropologic and bio-demographic data were collected from 62 unrelated individuals during three expeditions (2007-2008-2009). Information on birthplace, parents and grandparents, and also, informed consent were obtained from all the individuals sampled. In order to include only Yaghnobi people, all samples come from the most isolated villages in the valley.

DNA was extracted from all samples using a salting-out procedure (Miller *et al.*, 1998) and the first hypervariable region (HVR1) of mitochondrial DNA was amplified with primers L15996 and H16401 (Vigilant *et al.*, 1989). The same primers were used to sequence the fragments amplified.

Individuals with the “C-stretch” at the position 16189 were sequenced in each direction. Currently, 42 sequences have been obtained and analysed, spanning nucleotide positions 16024 - 16383.

In order to investigate the origins of the Yaghnobis and the relationships between the Yaghnobis and their neighbours, published mtDNA sequences were also included from Iranians (Arabs from Iran, Bakhtiari, Gilaki, Iranians from Isfahan, Iranians from Teheran and Mazandarani), Turks, Kurds (Kurmanji and Zazaki), Caucasians (Abkhazians, Avarians, Balkarians, Karachaians, Lezginians, North Ossetians and Rutulians) and Central Asian populations (Kazakhs, Kirghiz from Sary-Tash, Kirghiz from Talas, Tajiks and Uzbeks) (Comas *et al.*, 1998, 2004; Nasidze *et al.*, 2004a, 2004b, 2005, 2006, 2007, 2008; Derenko *et al.*, 2007).

Haplotype number, haplotype diversity and Tajima's D were calculated by DnaSP (Rozas & Rozas, 1995). The software package Arlequin 3.1 (Excoffier *et al.*, 2005) was run to estimate the Mean Number of Pairwise Differences (MNPD) and  $F_{st}$  indexes (10,000 permutations). Haplotype sharing analysis between populations was obtained by FaBox application (Villesen, 2007). The multidimensional scaling (MDS) analysis was performed by the software STATISTICA (StatSoft). Network 4.1 (Bandelt *et al.*, 1999) was run to estimate a median-joining network, using and the same weights for all loci.

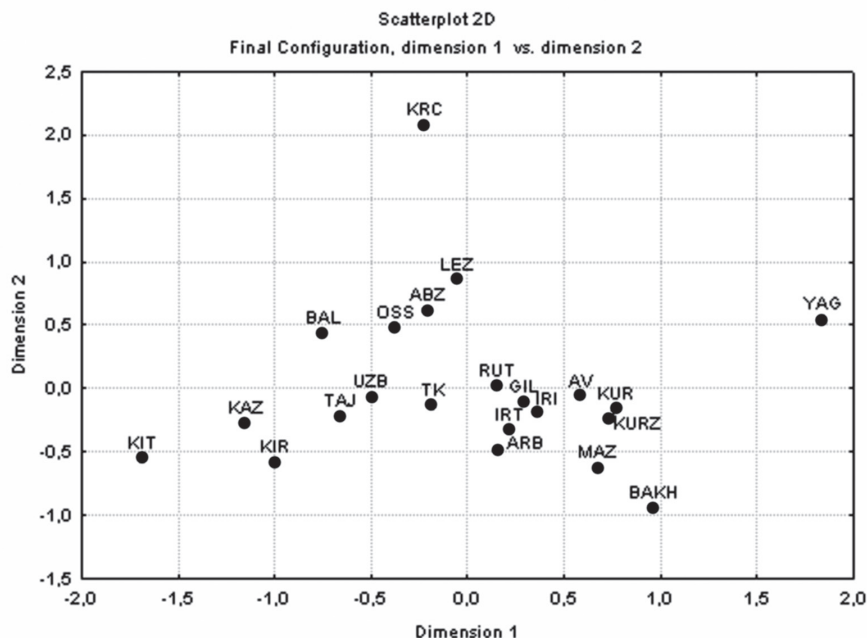
The population analysed showed a reduced number of haplotypes (19) and low haplotype diversity ( $0.930 \pm 0.023$ ), slightly lower with respect to the values that emerged from other Central Asian populations ( $0.984 - 0.995$ ; Comas *et al.*, 1998), but similar to the lower limit of European variations ( $0.936 - 0.973$ ; Comas *et al.*, 1997) (these values are not significantly different from one another as calculated by t-test,  $P \geq 0.05$ , two-tailed test). It is interesting to point out that the value of haplotype diversity which emerged from the Yaghnobi samples is lower than other Tajiks ( $0.995 \pm 0.006$ ; Derenko *et al.*, 2007), although the difference between these values is also not significant (t-test,  $P \geq 0.05$ , two-tailed test). The mean number of pairwise differences (MNPD) in the

Yaghnobi population is  $4.36 \pm 2.20$  and lies within the range of European samples ( $3.15 - 5.03$ ; Comas *et al.*, 1997), whereas Central Asian populations presented higher haplotype diversities ( $5.91 - 6.64$ ; Comas *et al.*, 1998). Tajima's D was negative and it was not significantly different from zero.

Haplotype sharing analysis shows that Yaghnobis share 8 out of 19 haplotypes with other populations, whereas 11 haplotypes are exclusively found in Yaghnobi population. Among the haplotypes shared, one is represented by the Cambridge Reference Sequence, shared by all the populations examined. With regard to the eight common haplotypes, the Yaghnobis shared all these haplotypes with Iranians (in particular with Gilaki (4 haplotypes), Iranians from Isfahan (4), Iranians from Teheran (3), Mazandarani (3), Arabs from Iran (1) and Bakhtiari (1)). They also share four haplotypes with Kurds (Kurmanji (4) and Zazaki (3)), one haplotype with Turks and four haplotypes with Caucasians (Abkhazians (1), Avarians (4), Lezginians (1), North Ossetians (1) and Rutulians (1)). Moreover, the Yaghnobis share four haplotypes with Central Asian populations (Tajiks (1), Uzbeks (2), Kirghiz from Talas (1) and Kazakhs (2)). It is important to highlight that Yaghnobis share only one haplotype with other Tajiks.

Pairwise  $F_{st}$  comparisons indicate that Yaghnobis are more similar to Kurds (Kurmanji and Zazaki, respectively  $F_{st} = 0.039$ ,  $p = 0.000$ ;  $F_{st} = 0.040$ ,  $p = 0.000$ ) and also to Iranians (Gilaki, Iranians from Isfahan and Iranians from Teheran, respectively  $F_{st} = 0.040$ ,  $p = 0.000$ ;  $F_{st} = 0.040$ ,  $p = 0.000$ ;  $F_{st} = 0.043$ ,  $p = 0.000$ ). Respect to their neighbours, Yaghnobis showed closest affinities with Uzbeks ( $F_{st} = 0.055$ ,  $p = 0.002$ ), than to other Tajiks ( $F_{st} = 0.085$ ,  $p = 0.000$ ), Kirghiz from Sary-Tash and Kirghiz from Talas (respectively  $F_{st} = 0.096$ ,  $p = 0.000$ ;  $F_{st} = 0.126$ ,  $p = 0.000$ ). Moreover, the Yaghnobis appear significantly different from Karachaians ( $F_{st} = 0.120$ ,  $p = 0.000$ ) and Balkarians ( $F_{st} = 0.109$ ,  $p = 0.000$ ).

A MDS plot based on the pairwise  $F_{st}$  values was obtained (Fig. 2). Stress value was lower than the critical stress value calculated by Sturrock & Rocha (2000). In particular, in this figure the



**Fig. 2 - MDS plot based on pairwise  $F_{st}$  values (stress value 0.121). Names of populations are abbreviated as follows: ABZ, Abkhazians; ARB, Arabs from Iran; AV, Avarians; BAKH, Bakhtiari; BAL, Balkarians; GIL, Gilaki; IRI, Iranians from Isfahan; IRT, Iranians from Teheran; KAZ, Kazakhs; KRC, Karachaians; KIR, Kirghiz from Sary-Tash; KIT, Kirghiz from Talas; KUR, Kurmanji; KURZ, Zazaki; LEZ, Lezginians; MAZ, Mazandarani; N\_OSS, North Ossetians; RUT, Rutulians; TD, Tajiks; TK, Turks; UZB, Uzbeks; YAG, Yaghnobis.**

Yaghnobis occupy a peripheral position, distant from all other populations, demonstrating that Yaghnobis are not close to the other groups examined. Moreover, the Karachaians appear distant from all other populations, probably because of their geographical location, as previously showed by Nasidze *et al.* (2004a).

The phylogenetic relationships of the sequences of Yaghnobis and their neighbours (Tajiks and Uzbeks) were summarised in a median-joining network (Fig. 3). No clear pattern can be detected, even if a large part of Yaghnobis tend to cluster together. When data about haplogroups identification will be available, the analysis will be more exhaustive.

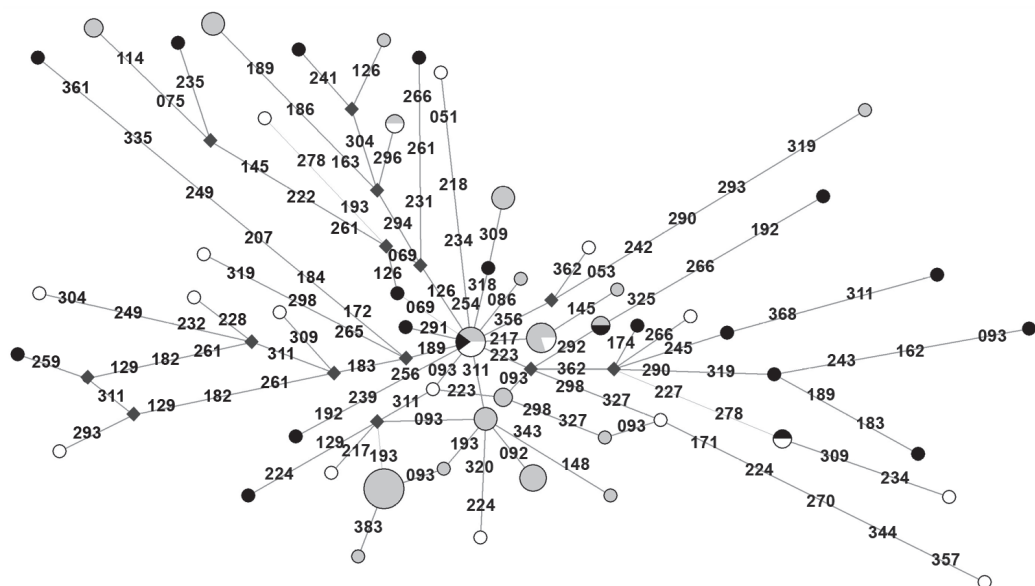
The preliminary analysis of mitochondrial DNA shows that Yaghnobis have a unique biological identity, distinct from other populations examined in this study. It appears that they show a reduced variability, probably resulting from

their prolonged isolation and the reduction of population size as a consequence of deportation. This study represents a preliminary analysis and it is necessary to extend the analysis to the study of mitochondrial DNA (coding regions) and also Y-chromosome, to understand the contribution of both sexes to population dynamics.

The Yaghnobi people represent a living museum of history, culture, language and genetic heritage and maintaining the ethno-anthropological and ethnographic features of the Yaghnobis is of vital importance.

## Acknowledgements

*This work was supported by Fondazione del Monte di Bologna e Ravenna and Fondazione Flaminia di Ravenna.*



**Fig. 3 - Median-joining network constructed for mtDNA sequences of Yaghnobis and their neighbours. The size of the circles is proportional to the number of sequences. Yaghnobis are represented in light grey, Tajiks in black and Uzbeks in white. Median vectors are represented by dark grey rhombus. Mutated sites (minus 16000) are indicated along the lines.**

## References

- Bandelt H.J., Forster P. & Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, 16: 37-48.
- Comas D., Calafell F., Mateu E., Pérez-Lezaun A., Bosch E. & Bertranpetit J. 1997. Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.*, 99: 443-449.
- Comas D., Calafell F., Mateu E., Pérez-Lezaun A., Bosch E., Martínez-Arias R., Clarimon J., Facchini F., Fiori G., Luiselli D., Pettener D. & Bertranpetit J. 1998. Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am. J. Hum. Genet.*, 63: 1824-1838.
- Comas D., Plaza S., Wells R.S., Yuldaseva N., Lao O., Calafell F. & Bertranpetit J. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.*, 12: 495-504.
- Delaini P. 2008. A small-scale pharmacy at Gharman – 2800 m. in Tajikistan Mountains. In Panaino A. (ed): *A Scuola di Pace vol. II. Nuove risposte alle sfide autoidentitarie, economiche e giuridiche del XXI secolo*, pp.105-114. Mimesis, Milano.
- Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Dambueva I., Perkova M., Dorzhu C., Luzina F., Lee H.K., Vanacek T., Villems R. & Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am. J. Hum. Genet.*, 81: 1025-1041.
- Excoffier L., Laval G. & Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, 1: 47-50.
- Miller S.A., Dykes D.D. & Polesky H.F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells *Nucleic Acids Res.*, 16: 1215.
- Nasidze I., Ling E.Y., Quinque D., Dupanloup I., Cordaux R., Rychkov S., Naumova O.,

- Zhukova O., Sarraf-Zadegan N., Naderi G.A., Asgary S., Sardas S., Farhud D.D., Sarkisian T., Asadov C., Kerimov A. & Stoneking M. 2004a. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann. Hum. Genet.*, 68: 205-221.
- Nasidze I., Quinque D., Dupanloup I., Rychkov S., Naumova O., Zhukova O. & Stoneking M. 2004b. Genetic evidence concerning the origins of South and North Ossetians. *Ann. Hum. Genet.*, 68: 588-599.
- Nasidze I., Quinque D., Ozturk M., Bendukidze N. & Stoneking M. 2005. MtDNA and Y-chromosome variation in Kurdish groups. *Ann. Hum. Genet.*, 69: 401-412.
- Nasidze I., Quinque D., Rahmani M., Alemohamad S.A. & Stoneking M. 2008. Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. *Ann. Hum. Genet.*, 72: 241-252.
- Nasidze I., Quinque D., Udina I., Kunizheva S. & Stoneking M. 2007. The Gagauz, a linguistic enclave, are not a genetic isolate. *Ann. Hum. Genet.*, 71: 379-389.
- Panaino A.C.D. 2008. Attraversando la valle dello Yaghnob: passato, presente e (possibile) futuro di una minoranza etno-linguistica. In G.P. Basello, D. Guizzo & P. Ognibene (eds.): *Sulla punta di uno spillo*, pp. 3-77. Mimesis, Milano.
- Rozas J. & Rozas R. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Applic. Biosci.*, 11: 621-625.
- Sturrock K. & Rocha J. 2000. A multidimensional scaling stress evaluation table. *Field Methods*, 12: 49-60.
- Vigilant L., Pennington R., Harpending H., Kocher T.D. & Wilson A.C. 1989. Mitochondrial DNA sequences in single hairs from a southern African population *Proc. Natl. Acad. Sci. U.S.A.*, 86: 9350-9354.
- Villesen P. 2007. FaBox: an online toolbox for FASTA sequences. *Mol. Ecol. Notes*, 7: 965-968.

Editor, Giovanni Destro-Bisol