



Blood group typing in five Afghan populations in the North Hindu-Kush region: implications for blood transfusion practice.

S. Mazières, S. A. Temory, H. Vasseur, P. Gallian, J. Di Cristofaro, J. Chiaroni

► To cite this version:

S. Mazières, S. A. Temory, H. Vasseur, P. Gallian, J. Di Cristofaro, et al.. Blood group typing in five Afghan populations in the North Hindu-Kush region: implications for blood transfusion practice.. *Transfusion Medicine*, Wiley, 2013, 23 (3), pp.167-74. .

HAL Id: hal-00814792

<https://hal.archives-ouvertes.fr/hal-00814792>

Submitted on 23 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blood group typing in five Afghan populations in the North Hindu-Kush region:

Implications for blood transfusion practice

Stéphane Mazières,¹ Shah Aga Temory,² Hélène Vasseur,³ Pierre Gallian,¹ Julie Di Cristofaro,¹
Jacques Chiaroni¹

¹ Laboratoire Hématologie Moléculaire, ADES UMR 7268, Aix Marseille Université, 13392,
Marseille, France

² Centre National de Transfusion Sanguine de Kabul, Afghanistan

³ Etablissement Français du Sang Auvergne-Loire, Saint-Etienne, France

Correspondence to: Jacques Chiaroni, Etablissement Français du Sang Alpes-Méditerranée, 149
boulevard Baille - 13392 Marseille Cedex 5 | Phone: +33 (0)4 9118 9500 | Fax: +33 (0)4 9218 9350 |
e-mail: jacques.chiaroni@efs.sante.fr.

Grant support: This work was supported by the *Agence Nationale de la Recherche*, grant number
ANR-07-BLAN-0049 - CSD 9.

Conflict of interest: the authors state that there is no conflict of interest.

Running head: Blood group typing in Afghans

SUMMARY

Background and Objectives: Blood incompatibility arises from individual and ethnic differences in red blood cell (RBC) antigen profiles. This underlines the importance of documenting RBC antigen variability in various ethnic groups. Central Asia is an area with a long and complex migratory history. The purpose of this article is to describe key antigen frequencies of Afghan ethnic groups in the Hindu-Kush region of Afghanistan as a basis for improving blood transfusion practices in that area.

Material and Methods: The key ABO, Rh and Kell antigens were investigated in five Afghan populations. In order to depict accurately the blood group gene diversity in the area, DNA from eight additional Pakistani populations were included, and the entire sample set screened using two multiplex PCRs sensitive for 17 alleles in 10 blood group genetic systems (MNS, Kell, Duffy, Kidd, Cartwright, Dombrock, Indian, Colton, Diego and Landsteiner-Wiener).

Results: Phenotype and allele frequencies fell within the ranges observed in Western European and East Asian populations. Occurrence of *DI*01*, *IN*01*, *LW*07*, and *FY*02N.01* and prevalence of *ABO*B.01* were consistent with migratory history as well as with putative environmental adaptation in the subtropical environment Hindu-Kush region.

Conclusion: These findings expand current knowledge about key antigen frequencies. Regarding occurrence of viral markers, further blood transfusion in the region requires rigorous typing.

Key words: Hindu-Kush populations, multiplex genotyping, blood group antigens

INTRODUCTION

Transfusion institutions strive to provide blood products that are microbiologically safe and compatible with recipient blood characteristics. Achieving this goal requires stringent pre-transfusion screening at two levels. The first level involves blood cell antigens. To date, the International Society Blood Transfusion (ISBT) has recorded a total of 339 blood group specificities including 297 in 33 systems (Reid *et al.* 2012). Since the ABO and Rh antigens can be immunologically significant due to the presence of natural ABO antibodies and strong immunogenicity of the RHD antigen, systematic ABO-RhD typing is necessary. The second level of pre-transfusion screening involves identification of Rh-Kell antigenic profiles and, to a lesser extent, Duffy, Kidd and Ss phenotype. Screening for the latter phenotypes is necessary depending on the patient, pathology and risk of alloimmunization.

In addition to ensuring transfusion safety, awareness of antigen frequency in a given population is crucial to defining product requirements and minimum stock levels. Given the high variability of red cell antigens, transfusion institutes must determine antigen frequencies in the population that they serve. Such determination is especially important in population settings like Afghanistan (Witzel 1999). In Kabul, the Central Blood Bank (BCS, *Banque Centrale du Sang*) was established in the 1980s, but political unrest, lack of funding and policy matters forced closure between 1992 and 1995. Since 1995, the BCS has been working to maintain a sound basis for a self-sufficient blood transfusion system under the auspices of four government ministries: Health, Defense, National Security, and Home Affairs.

The *Etablissement Français du Sang Alpes-Méditerranée* (EFS-AM) has been deeply involved in the rehabilitation of the BCS in Kabul. In 1999, an initial study conducted in collaboration with the EFS-AM presented the serological profile and blood characteristics of the population living in the Afghan capital (Dupire *et al.* 1999). In 2008, the EFS-AM with the support of the *Agence Nationale de la Recherche* implemented the AFGHAPOP program in a larger sample of the Afghan population in surrounding areas. The twofold goal of AFGHAPOP was to describe genetic diversity based on anthropological genetic markers and improve the safety of blood transfusion in Afghanistan.

The present report describes results obtained in a cohort of blood donors originating from the main Afghan communities studied within the framework of the AFGHANPOP program. The primary study objective was to define the frequency of key antigens (ABO, Rh, Kell) and prevalence of other blood group genes for blood transfusion. The secondary objective was to gain better insight into the distribution of red cell antigens in human populations and highlight the correlated anthropological and environmental factors.

MATERIAL AND METHODS

Sample collection

This study was approved by the Afghan authorities and CNRS ethics committee. **Blood** collection was performed after obtaining informed consent in 5 Afghan populations living in the mountainous Hindu-Kush region: Hazara, Pashto, Uzbek, Tajik and Turkmen. To allow accurate interpretation of the Afghan dataset in the geographical and anthropological context, 168 additional **DNA** samples from 8 Pakistani populations (Hazara, Kalash, Makrani, Balochi, Brahui, Burusho, Pathan and Sindhi) were obtained from the Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) (Cann *et al.* 2002).

Surface antigen typing

A total of 507 **blood** samples were selected for red cell antigen typing of the ABO, Rh, and Kell blood group systems. Typing was carried out in Kabul with Diagast (France) reagents anti-A (Clone 9113D10), -B (Clone 9621A8), -AB (Clones 9113D10 + 152D12), and -D (Clones P3x61 + P3x21223B10 + P3x290 + P3x35); and Eurobio (France) reagents anti-C (Clones MS273), -c (Clone MS35), -E (Clones MS12 + MS260), -e (MS62 + MS69) and -K (Clone AEK4).

Blood group genotyping by two Singleplex PCR

Genomic DNA was extracted from whole blood samples using the QIAamp DNA Blood Mini Kit (Qiagen) and stored at -20°C (-4°F). Genotyping of both Afghan and Pakistani samples was performed using two multiplex PCRs at the EFS-AM in Marseilles, France with focus on 17 polymorphisms coding for MNS, Kell, Duffy, Kidd, Cartwright, Dombrock, Indian, Colton, Diego and Landsteiner-Wiener red cell antigens of transfusion and anthropological interest. The first one was designed for simultaneous detection of 11 SNPs of the *GYPB* (*MNS*), *KEL* (*KEL*), *FY* (*FY*), *SLC14A1* (*JK*), *ACHE* (*YT*), *DO* (*DO*), *AQP1* (*CO*) genes whose red cell antigens are strongly implicated in transfusion reactions (Daniels *et al.* 2004; Di Cristofaro *et al.* 2010).

The second multiplex PCR was developed to screen 10 polymorphisms defining 15 alleles of anthropological interest: *KEL**02.06, *FY**01/*FY**02, *FY**02M.01, *FY**02N.01, *DI**01/*DI**02, *DO**01/*DO**02, *DO**02.-04, *DO**01.-05, *LW**05/*LW**07 and *IN**01/*IN**02 (Table1). Some of these markers are redundant with the first multiplex PCR devoted to transfusion purposes (Di Cristofaro *et al.* 2010). It was performed on 250 ng of genomic DNA in a final volume of 25 µl containing 0.05 unit of TaqDNA-polymerase and 2X of the provided PCR buffer (Invitrogen, Cergy Pontoise, France), 3.4 mmol/L MgCl₂, 0.2 mmol/L of each dNTP, and a suitable concentration of each primer (see table 1 for final concentration). Amplification was performed under stringent conditions: 96°C for 5 minutes for 1 cycle; 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 1 minute for 30 cycles, followed by one 10-minute-long cycle at 72°C. PCR amplicon purification were described elsewhere (Di Cristofaro *et al.* 2010). Probe primer extensions for the SNaPshot reaction are presented in Table 1. Primer extension reactions were segregated by capillary electrophoresis on an automated fluorescence-based ABI PRISM 3130 XL genetic analyzer according to manufacturer's protocol.

In order to check accuracy and reproducibility of the multiplex SNaPshot assay presented here, the genotypes not included in the first multiplex (Di Cristofaro *et al.* 2010) were validated with known phenotypes obtained by serology and sequencing. For the crossmatch of the results using serology, only the anti-Fya (Clone DG-FYA-02, Diagast, France) and anti-Fyb (polyclonal, Diagast, France) antibodies were available. In addition, the validation was enhanced using 41 representative homozygotes and heterozygotes from our collection to act as control samples by direct sequencing.

However, six genotypes could not be checked because of their scarcity: heterozygote for *DO*02.-04* or *DO*01.-05*; homozygote for the total absence of *DO*02.-04* or *DO*01.-05*; *IN*01/01*; and *LW*07/07*.

Statistical analysis

Red cell antigen frequencies were calculated by direct counting. ABO allele frequencies were estimated from the ABO phenotypes using Bernstein's method that calls for direct counting in the case of genotypes with recessive alleles (Bernstein 1925). By default we labeled the traditional *ABO*A.01*, *ABO*B.01* and *ABO*O.01.01* alleles as the ones coding for the observe phenotypes. RH haplotype frequencies were estimated from the RH phenotypes using the Expectation-Maximization (EM) algorithm in the GENE[RATE] package (Nunes 2006). Assuming the Hardy-Weinberg equilibrium, this procedure enables to estimate the maximum-likelihood haplotype frequencies from multi-locus genotype data and considers all possibilities of genotypic ambiguities, that is when several genotypes can lead to the same phenotype (Dempster *et al.* 1977). The EM algorithm was done with 1000 iterations of all possible haplotype combinations.

Population comparison was performed between data obtained using Bernstein's method and the EM algorithm and previous ABO, Rh and Kell data including Western Europeans, Sub-Saharan, 983 West Asians, circa 43,000 South Asians and an undefined number of East Asians described in the literature (Cavalli-Sforza *et al.* 1994; Nanu and Thapliyal 1997; Dupire *et al.* 1999; Asadov *et al.* 2003; Periyavan *et al.* 2010; Thakral *et al.* 2010; Keramati *et al.* 2011; Radzhabov *et al.* 2011; Reid *et al.* 2012).

RESULTS

Data obtained by serological ABO, Rh and Kell antigen testing observed in the Afghan samples are given in Table 2, and the inferred allele frequencies in Table 3. The last five columns of Table 3 present the allele frequencies observed in the main Western European, Sub-Saharan, and Asian populations. A striking feature of the ABO system in the Afghan populations studied here was

the predominance of the phenotype B (range: 25-32%) and frequency of the *ABO*B.01* allele (range, 21-25%). Higher AB phenotype frequencies were observed in the Uzbek, Tajik, Turkmen and Pashto ethnic groups (13-15%) than in the Hazara group (4%). The O phenotype was found in 26 to 35% of the individuals serotyped and the estimated *ABO*O.01.01* allele frequency ranged between 50 and 60%.

Regarding Rh phenotypes, the frequency of D- ranged from 6% in the Pashto, and Tajik groups to 16% in the Hazara group. Three D- phenotypes were observed: *D-C+E-c+e+* in two Hazaras and one Turkmen, *D-C-E+c+e+* in one Tajik, and the ubiquitous *D-C-E-c+e+* that ranged from 6% (n=5) in the Pashto to 13% (n= 10) in the Hazara. A total of four D+ phenotypes were encountered in the five Afghan populations. The most frequent D+ phenotype was *D+C+E-c-e+* (range, 31-37%) followed by *D+C+E-c+e+* (range, 22-33%), *D+C+E+c+e+* (range 16-19%) and *D+C-E+c+e+* (range, 6% to 14%).

Based on the estimates inferred from the RH phenotypes using the EM algorithm of GENE[RATE] (Dempster *et al.* 1977; Nunes 2006), the three most prevalent haplotypes in the five Afghan populations were *DcE*, *dce* and *DcE*, accounting for 51-54%, 24-29% and 13-19% respectively. The *Dce* haplotype ranged from 3 to 4% in the Pashto, Tajik and Uzbek groups and the *dCe* haplotype ranged from 3 to 5% in the Hazara and Turkmen groups. The *DCE* haplotype, which is characteristic of Asian populations (1-3%), was observed only in the Pashto (1%) and Tajik (2%) groups. Only the Tadjik group carried the European-specific *dcE* haplotype (1%). The *dCE* haplotype was not detected in any of the five Afghan populations.

With regard to the Kell system, the K antigen was detected in all five ethnic groups. Its frequency ranged from one case (1%) each in the Tajik and Turkmen groups up to seven cases in the Uzbek group (6%).

Table 4 shows the frequencies of 15 biallelic polymorphisms successfully amplified in 5 Afghan and 8 Pakistani populations. In the samples under study, the *DO*02.-04* and *DO*01.-05* alleles were ubiquitously found, and conversely the *GYPB*03N.03* and *GYPB*03N.02* alleles were not encountered. The two latter were not included in Table 4. Seven alleles were sporadically detected

with a frequency of less than 4%: *KEL*01.1*, *KEL*02.06*, *FY*02M.01*, *DI*01*, *IN*01*, *LW*07* and *CO*02*. *KEL*02.06* coding for the Js^a antigen was observed in one heterozygous Tajik, and *IN*01* responsible for the In^a antigen was detected in one heterozygous Tajik and one heterozygous Burusho. *DI*01* was observed in 6 populations (Pashto, Tajik, Turkmen, Balochi, Hazara and Sindhi) between 1% and 3%. *CO*02* was spotted in the Hazara (1%), the Turkmen (1%) and the Burusho (2%). The frequency of the *FY*02N.01* allele was low in Afghan samples (range, 1 to 5%) but interestingly reached 12% in the Baloshi and 18% in the Makrani from Pakistan. After cross-validation with the raw data and using the EM algorithm supplied in the GENE[RATE] package (Nunes 2006), we could ensure that they all FY silent mutations were associated with the Fy^b-specific polymorphism. Lastly, *JK*01* and *YT*01* ranged respectively between 48% (Burusho) and 73% (Kalash), and 84% (Pathan) and 100% (Brahui).

DISCUSSION

The present study provides serologic frequencies in order to ensure safe blood transfusion in Afghanistan. We especially performed two multiplex molecular analyses which amplified simultaneously alleles coding for RBC antigens either major for transfusion and anthropological purposes in the same Afghan populations enriched with DNA samples from Pakistan. These molecular arrays turned to be preferentially employed because they were low-cost, fast, and avoided the clinical shortcomings of the hemagglutination-based assays (Di Cristofaro *et al.* 2010). Importantly, no discordance was observed between the phenotype inferred from the SNaPshot genotyping and the control samples, so we assume that the method herein presented covered accurately the DNA regions of interest.

We then compared the frequencies obtained with that of the main geographical areas of interest (Table 3), and the following patterns were outstanding. The mean frequency of the *ABO*B.01* allele was higher in Afghan populations than in Sub-Saharan (13%) and Eurasian (7-20%) populations. Similar frequencies have been reported in South and Central Asia (25-35%) (Mourant *et al.* 1958; Abawi *et al.* 2001; Keramati *et al.* 2011). In addition, the D-C-E-c+e+ phenotype and the

*KEL*01.1*, *FY*01*, *DI*01* and *YT*02* allele frequencies in the Afghan groups were intermediate between those observed in Western Europeans and East Asians in agreement with earlier reports describing the clinal frequencies of the *D-C-E-c-e+* phenotype and K, Fy^b, Di^a and Yt^b antigens across Eurasia (Cavalli-Sforza *et al.* 1994; Reid *et al.* 2012).

Previous analysis of gene flow patterns in Afghanistan that is located at the crossroad between West and East Asia have documented waves of population migration resulting in substantial gene admixture (Quintana-Murci *et al.* 2004). In this view, we also detected in addition to the *DI*01* allele, classically related to Mongols expansion, the *LW*07*, and *FY*02N.01* alleles unusually encountered in Central Asian but in Northern European and Sub-Saharan populations (M'Baya *et al.* 2010; Howes *et al.* 2011; Reid *et al.* 2012). Another puzzling finding with respect to blood group alleles involves *IN*01* coding for In^a whose current geographical distribution is split into two cores: Middle Eastern (Cavalli-Sforza *et al.* 1994; Reid *et al.* 2012) and South Asian (Badakere *et al.* 1974) populations. Curiously the geographical distribution of *IN*01* matches that of the Elamo-Dravidian language family, a past linguistic cluster linking modern Dravidian languages spoken in southern India to the extinct Elamite language formerly spoken in southwestern Iran. This suggests that diffusion of *IN*01* into India took place about 4,500 years ago by Elamo-Dravidian populations (Cavalli-Sforza *et al.* 1994). The presence of *IN*01* in Afghanistan and in northern Pakistan supports the possibility of an ancient settlement of Elamo-Dravidian populations in Central Asia and adds to our understanding of the current distribution area of the In^a antigen.

Furthermore, Central and South Asia are known to be the original reservoir of cholera (Sack *et al.* 2004) and plague bacteria (Achtman *et al.* 2004), and the southern coast of Pakistan has been defined as a malaria transmission area (WHO 2011). Since RBC antigens can act as fixation sites for infectious organisms (Telesmanich *et al.* 2008; Howes *et al.* 2011), it is plausible that directive selection played a role in the evolution of the different human antigenic profiles, notably in our region of study. Previously attempts suggested that individuals who can make anti-H were likely to be more resistant to the plague and cholera (Garraty 2005; Harris *et al.* 2005; Telesmanich *et al.* 2008). The high frequency of the B phenotype in our samples does not contradict this plausible hypothesis. Also,

the higher presence in the Makrani of the FY silent allele could offer a selective advantage against *Plasmodium vivax*. Nonetheless, we detected in the Makrani samples 14% of the Fya-/Jkb-/S-phenotype; a component found in 31.5% in Africans as opposed to only 4.5% of Europeans (Noizat-Pirenne 2003), and the Makrani were further documented by 39.4% of mitochondrial haplotypes of African origin (Quintana-Murci *et al.* 2004). In this regard, our data are consistent with the notion of overlapping blood group allele frequencies between Western Eurasia and East Asia, gene exchanges, and probably positive selection.

For the last two decades, there have been several blood group surveys in the urban populations of Afghanistan which have mainly highlighted major failures in the blood handling coupled with a worrying microbiological situation (Todd *et al.* 2007; Ayyoubi *et al.* 2010; Khan and Attaullah 2011). Compatibility screening in Afghanistan is currently restricted to the ABO, RHD, and K key antigens (Dupire *et al.* 1999; Abawi *et al.* 2001). In emergency situations, e.g. after bombing, it is limited to ABO compatibility only without detection of HIV, HBV, or HCV viral markers. These practices must be considered as extremely hazardous, noticeably because of the vaccination rate against hepatitis in the country, a high level of drug use in urban areas, and several high-risk cultural practices including tattooing (Todd *et al.* 2007; Khan and Attaullah 2011). Fewer than half of emerging countries perform pre-donation medical interviewing to identify risk factors and post-collection testing to detect viral markers (Dupire *et al.* 1999). To further enhance safety, the commercially ELISA combo test should be used in order to prevent HIV and HCV contamination by blood products.

The present study provided a wide array of RBC antigens in Afghan donors and Pakistani populations in the ethnographic and blood transfusion contexts of Afghanistan. These study's findings, which were forwarded to the NBC in Kabul, emphasized how crucial the estimation of biological diversity is in a panel of Afghan donors in defining standards for safe blood collection and release in the area.

ACKNOWLEDGEMENTS

The authors are grateful to the donors who consented to participate in the present study and to Leslie Lison who performed genetic screening of the Afghan samples within the framework of the *Master de Génétique et Santé, Université Aix-Marseille*. Funding for this study was obtained from the *Agence National de la Recherche* (Grant #BLAN07-3_222301, CSD 9 - Sciences humaines et sociales). S. M. analyzed the data and wrote the paper, S. A. T. provided blood samples, H. V. performed and validated the molecular analyses, P. G. analyzed the serological data, J. D. C. analyzed the data, validated and performed molecular analyses, and J. C. designed the research and wrote the paper. The authors state that they have no competing interest.

REFERENCES

- Abawi, A.K., Chiaroni, J., Dupire, B., Ganteaume, C., Lam, T., Truze, P., Micco, P.D. and Boetsch, G. (2001). Genetic Polymorphism (ABO, Rh, Kell) in Kabul (Afghanistan). *International Journal of Anthropology*, **16**, 275-280.
- Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., Chenal-Francisque, V., Worsham, P., Thomson, N.R., Parkhill, J., Lindler, L.E., Carniel, E. and Keim, P. (2004). Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci U S A*, **101**, 17837-17842.
- Asadov, D.A., Durmanov, B.D., Ismailov, S.I. and Dzhamalutdinova, I. (2003). [Reforming of the health care system in Uzbekistan]. *Probl Sotsialnoi Gig Zdravookhranennii Istor Med*, 54-57.
- Ayyoubi, M.T., Konstenius, T., McCullough, J.C., Eastlund, T., Clay, M., Bowman, R., Rahmani, A.M., Riley, W. and McCullough, J. (2010). Status of blood banking and the blood supply in Afghanistan. *Transfusion*, **50**, 566-574.
- Badakere, S.S., Parab, B.B. and Bhatia, H.M. (1974). Further observations on the Ina (Indian) antigen in Indian populations. *Vox Sang*, **26**, 400-401.
- Bernstein, F. (1925). Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Z. Indukt. Abstammungs. Vererbungslehre*, **37**, 237-370.
- Bruce, L.J., Anstee, D.J., Spring, F.A. and Tanner, M.J. (1994). Band 3 Memphis variant II. Altered stilbene disulfonate binding and the Diego (Dia) blood group antigen are associated with the human erythrocyte band 3 mutation Pro854-->Leu. *J Biol Chem*, **269**, 16155-16158.
- Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., Chu, J., Carcassi, C., Contu, L., Du, R., Excoffier, L., Ferrara, G.B., Friedlaender, J.S., Groot, H., Gurwitz, D., Jenkins, T., Herrera, R.J., Huang, X., Kidd, J., Kidd, K.K., Langaney, A., Lin, A.A., Mehdi, S.Q., Parham, P., Piazza, A., Pistillo, M.P., Qian, Y., Shu, Q., Xu, J., Zhu, S., Weber, J.L., Greely, H.T., Feldman, M.W., Thomas, G., Dausset, J. and Cavalli-Sforza, L.L. (2002). A human genome diversity cell line panel. *Science*, **296**, 261-262.
- Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994). *The history and geography of human genes*, Princeton University Press, Princeton.
- Daniels, G., Finning, K., Martin, P. and Soothill, P. (2004). Fetal blood group genotyping from DNA from maternal plasma: an important advance in the management and prevention of haemolytic disease of the fetus and newborn. *Vox Sang*, **87**, 225-232.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1-38.
- Di Cristofaro, J., Silvy, M., Chiaroni, J. and Bailly, P. (2010). Single PCR multiplex SNaPshot reaction for detection of eleven blood group nucleotide polymorphisms: optimization, validation, and one year of routine clinical use. *J Mol Diagn*, **12**, 453-460.
- Dupire, B., Abawi, A.K., Ganteaume, C., Lam, T., Truze, P. and Martet, G. (1999). Establishment of a blood transfusion center at Kabul (Afghanistan). *Sante*, **9**, 18-22.
- Garraty, G. (2005). Relationship of blood groups to disease: do blood group antigens have a biological role? *Rev Med Inst Mex Seguro Soc*, **43**, 113-121.
- Gubin, A.N., Njoroge, J.M., Wojda, U., Pack, S.D., Rios, M., Reid, M.E. and Miller, J.L. (2000). Identification of the dombrock blood group glycoprotein as a polymorphic member of the ADP-ribosyltransferase gene family. *Blood*, **96**, 2621-2627.
- Harris, J.B., Khan, A.I., LaRocque, R.C., Dorer, D.J., Chowdhury, F., Faruque, A.S., Sack, D.A., Ryan, E.T., Qadri, F. and Calderwood, S.B. (2005). Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. *Infect Immun*, **73**, 7422-7427.

- Hernand, P., Gane, P., Mattei, M.G., Sistonen, P., Cartron, J.P. and Bailly, P. (1995). Molecular basis and expression of the LWa/LWb blood group polymorphism. *Blood*, **86**, 1590-1594.
- Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Menard, D., Williams, T.N., Weatherall, D.J. and Hay, S.I. (2011). The global distribution of the Duffy blood group. *Nat Commun*, **2**, 266.
- Keramati, M.R., Shakibaei, H., Kheiyami, M.I., Ayatollahi, H., Badiei, Z., Samavati, M. and Sadeghian, M.H. (2011). Blood group antigens frequencies in the northeast of Iran. *Transfus Apher Sci*.
- Khan, S. and Attaullah, S. (2011). Share of Afghanistan populace in hepatitis B and hepatitis C infection's pool: is it worthwhile? *Virology*, **8**, 216.
- Lee, S., Wu, X., Reid, M. and Redman, C. (1995). Molecular basis of the K:6,-7 [Js(a+b-)] phenotype in the Kell blood group system. *Transfusion*, **35**, 822-825.
- Mourant, A.E., Kopec, A.C. and Domaniewska-Sobczak, K. (1958). *The ABO blood groups: comprehensive tables and maps of world distribution*, Blackwell Scientific Publications, Oxford.
- Nanu, A. and Thapliyal, R.M. (1997). Blood group gene frequency in a selected north Indian population. *Indian J Med Res*, **106**, 242-246.
- Noizat-Pirenne, F. (2003). Immunohematologic characteristics in the Afro-caribbean population. Consequences for transfusion safety. *Transfus Clin Biol*, **10**, 185-191.
- Nunes, J.M. (2006). Generate: tools for analysis and handling of data with ambiguities. Laboratory of Anthropology, Genetics and Peopling History, University of Geneva, Switzerland, Geneva.
- Periyavan, S., Sangeetha, S.K., Marimuthu, P., Manjunath, B.K. and Seema, D.M. (2010). Distribution of ABO and Rhesus-D blood groups in and around Bangalore. *Asian J Transfus Sci*, **4**, 41.
- Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Qasim Mehdi, S., Torroni, A. and McElreavey, K. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet*, **74**, 827-845.
- Radzhabov, M.O., Shamov, I.A., Gasaev, D.G., Shil'nikova, I.N., Shneider Iu, V. and Goldenkova-Pavlova, I.V. (2011). The Dagestan gene pool: polymorphism of immunogenetic and biochemical markers in Kumyks. *Genetika*, **47**, 260-267.
- Reid, M.E., Lomas-Francis, C. and Olsson, M.L. (2012). *Blood Group Antigen Factsbook*, Elsevier Academic Press, New York.
- Rios, M., Hue-Roye, K., Oyen, R., Miller, J. and Reid, M.E. (2002). Insights into the Holley- and Joseph- phenotypes. *Transfusion*, **42**, 52-58.
- Sack, D.A., Sack, R.B., Nair, G.B. and Siddique, A.K. (2004). Cholera. *Lancet*, **363**, 223-233.
- Telen, M.J., Udani, M., Washington, M.K., Levesque, M.C., Lloyd, E. and Rao, N. (1996). A blood group-related polymorphism of CD44 abolishes a hyaluronan-binding consensus sequence without preventing hyaluronan binding. *J Biol Chem*, **271**, 7147-7153.
- Telesmanich, N.R., Koliakina, A.V., Lomov Iu, M., Men'shikova, E.A. and Mironova, A.V. (2008). Characterization of the adhesive activity of cholera vibrios in mammalian red blood cells as an additional test for assessment of their epidemic significance. *Klin Lab Diagn*, **7**, 45-48.
- Thakral, B., Saluja, K., Sharma, R.R. and Marwaha, N. (2010). Phenotype frequencies of blood group systems (Rh, Kell, Kidd, Duffy, MNS, P, Lewis, and Lutheran) in north Indian blood donors. *Transfus Apher Sci*, **43**, 17-22.
- Todd, C.S., Abed, A.M., Strathdee, S.A., Scott, P.T., Botros, B.A., Safi, N. and Earhart, K.C. (2007). HIV, hepatitis C, and hepatitis B infections and associated risk behavior in injection drug users, Kabul, Afghanistan. *Emerg Infect Dis*, **13**, 1327-1331.
- Tournamille, C., Colin, Y., Cartron, J.P. and Le Van Kim, C. (1995a). Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*, **10**, 224-228.
- Tournamille, C., Le Van Kim, C., Gane, P., Cartron, J.P. and Colin, Y. (1995b). Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum Genet*, **95**, 407-410.

- WHO (2011). World Malaria Report. In: *WHO Press*. (ed. Library-Cataloguing-in-Publication), 259. World Health Organization, Geneva, Switzerland.
- Witzel, M. (1999). Early Sources for South Asian Substrate Languages. *Mother Tongue*, **extra number**, 1-76.

Legends

Table 1. Primer and probe sequences used for blood group amplification in the 10-pair multiplex PCR (Bruce *et al.* 1994; Hermand *et al.* 1995; Lee *et al.* 1995; Tournamille *et al.* 1995a; Tournamille *et al.* 1995b; Telen *et al.* 1996; Gubin *et al.* 2000; Rios *et al.* 2002).

Table 2. Occurrence and frequencies of the ABO, Rh and Kell phenotypes in five Afghan populations

Table 3. Allele and haplotype frequencies of the ABO, Rh and Kell blood group systems in five Afghan populations compared to published data. (Cavalli-Sforza *et al.* 1994; Nanu and Thapliyal 1997; Dupire *et al.* 1999; Asadov *et al.* 2003; Periyavan *et al.* 2010; Thakral *et al.* 2010; Keramati *et al.* 2011; Radzhabov *et al.* 2011; Reid *et al.* 2012).

Table 4. Frequencies of 15 biallelic polymorphisms in five Afghan and eight Pakistani populations.

Table 1. Primer and probe sequences used for blood group amplification in the 10-pair multiplex PCR (Bruce *et al.* 1994; Hermand *et al.* 1995; Lee *et al.* 1995; Tournamille *et al.* 1995a; Tournamille *et al.* 1995b; Telen *et al.* 1996; Gubin *et al.* 2000; Rios *et al.* 2002).

System	Allele	Amino-acid change & SNP	Location	Forward (5' to 3')	Extension primer (Forward/Reverse) ^a	Reverse (5' to 3')	Amplicon & primer size (bp)	Conc. (μM)		
KEL	<i>KEL*02.06</i>	Pro597Leu	Exon 17	CACCCACATCCTCACCTAGG		TGAGCTTCCTGGAGGGCATG	126	0.13		
		1790C>T			(F) AGTACTGCCTGGGGGCTGCC	20	0.1			
FY	<i>FY*01/FY*02</i>	Gly42Asp	Exon 2	TCCCCCTCAACTGAGAACTC		AAGGCTGAGCCATACCAGAC	392	0.63		
		125G>A			(F) 5T -CTTCCCAGATGGAGACTATG	25	1			
		<i>FY*02M.01</i>			Arg89Cys					
		265C>T			(F) 10T -TGCTTTTCAGACCTCTCTTC	30	1			
	<i>FY*02N.01</i>		Promoter	CCAAGGCCAGTGACCCCCATA		CATGGCACCGTTTGGTTCAG	190	0.13		
		-67T>C			(F) 15T -TCATTAGTCCTTGGCTCTTA	35	0.3			
DI	<i>DI*01/DI*02</i>	Pro854Leu	Exon 19	GGTGGTAGAGGGTGCTGG		GACACAGCAGGCCAGAAC	364	0.63		
		2561C>T			(R) 20T -AGGGCAGGGCCAGGGAGGCC	40	1.2			

Table 1. continued

System	Allele	Amino-acid change & SNP	Location	Forward (5' to 3')	Extension primer (Forward/Reverse) ^a	Reverse (5' to 3')	Amplicon & primer size (bp)	Conc. (μ M)
DO	DO*01/DO*02	Asn265Asp	Exon 2	ACACAGGGGCCACCATTTCGATT		TGTGCTCAGGTTCCCAGTTG	242	0.63
		793A>G			(R) 35T -TGACCTCAACTGCAACCAGT	55	0.05	
	DO*02.-04	Gly108Val		GATCAGTACCAAGGCTGTAGC		GAATGAACATTGCTGTTCAATG	212	0.81
		323G>T			(R) 25T -TGTCTGGGGTAGAACTTTT	45	0.4	
	DO*01.-05	Thr117Ile						0.81
350C>T				(R) 30T -AAATAGCCACAGCGTGTGTG	50	0.4		
LW	LW*05/LW*07	Gln 100 Arg	Exon 1	GGGAAGTCAGTGCAGCTCAA		GCAGGTCACGAGGCAGTG	168	0.13
		299 A>G			(F) 40T -GCCGGGTTGGGTGTCTTACC	60	0.05	
IN	IN*01/IN*02	Pro46Arg	Exon 2	GCCGCTTTGCAGGTGTATTC		CCAGTTGTCATACAGGTTGC	317	0.63
		137 C>G			(F) 45T -TGGTCGCTACAGCATCTCTC	65	0.2	

^a **Bold characters refer to the 5' poly-thymidine tail**

Table 2. Occurrence and frequencies of the ABO, Rh and Kell phenotypes in five Afghan populations

		Hazara	Pashto	Tajik	Turkmen	Uzbek
		N=77	N=81	N=140	N=70	N=121
ABO	A	24 (0.31)	21 (0.26)	42 (0.3)	16 (0.23)	35 (0.29)
	AB	3 (0.04)	12 (0.15)	18 (0.13)	10 (0.14)	16 (0.13)
	B	25 (0.32)	20 (0.25)	44 (0.31)	20 (0.29)	37 (0.31)
	O	25 (0.32)	28 (0.35)	36 (0.26)	24 (0.34)	33 (0.27)
Rh	D-C+E-c+e+	2 (0.03)	-	-	1 (0.01)	-
	D-C-E-c+e+	10 (0.13)	5 (0.06)	8 (0.06)	6 (0.09)	9 (0.07)
	D-C-E+c+e+	-	-	1 (0.01)	-	-
	D+C+E-c+e+	17 (0.22)	23 (0.28)	46 (0.33)	16 (0.23)	31 (0.26)
	D+C+E+c+e+	15 (0.19)	14 (0.17)	23 (0.16)	12 (0.17)	23 (0.19)
	D+C+E-c-e+	28 (0.36)	30 (0.37)	44 (0.31)	25 (0.36)	44 (0.36)
	D+C-E+c+e+	5 (0.06)	9 (0.11)	18 (0.13)	10 (0.14)	14 (0.12)
Kell	K-k+	75 (0.97)	77 (0.95)	139 (0.99)	69 (0.99)	114 (0.94)
	K+k-	2 (0.03)	4 (0.05)	1 (0.01)	1 (0.01)	7 (0.06)

Table 3. Allele and haplotype frequencies of the ABO, Rh and Kell blood group systems in five Afghan populations compared to published data. (Cavalli-Sforza *et al.* 1994; Nanu and Thapliyal 1997; Dupire *et al.* 1999; Asadov *et al.* 2003; Periyavan *et al.* 2010; Thakral *et al.* 2010; Keramati *et al.* 2011; Radzhabov *et al.* 2011; Reid *et al.* 2012).

	Hazara	Pashto	Tajik	Turkmen	Uzbek	Western Europe	Sub-Sahara	West Asia	South Asia	East Asia
ABO										
<i>ABO*AO1</i>	0.20	0.23	0.24	0.20	0.24	0.27	0.17	0.24	0.18	0.20
<i>ABO*B01</i>	0.21	0.22	0.25	0.24	0.25	0.07	0.13	0.19	0.23	0.20
<i>ABO*O.01.01</i>	0.60	0.56	0.50	0.56	0.51	0.66	0.70	0.56	0.58	0.59
RH										
<i>DCe</i>	0.51	0.53	0.54	0.52	0.54	0.42	0.17	0.44	0.72	0.69
<i>DcE</i>	0.16	0.18	0.13	0.17	0.19	0.14	0.11	0.13	0.09	0.20
<i>Dce</i>	-	0.04	0.03	-	0.03	0.04	0.44	0.09	0.04	0.05
<i>DCE</i>	-	0.01	0.02	-	-	-	-	0.02	0.03	0.01
<i>dCe</i>	0.05	-	-	0.03	-	0.02	0.02	0.01	0.03	0.01
<i>dcE</i>	-	-	0.01	-	-	0.01	-	0.01	-	-
<i>dce</i>	0.29	0.24	0.26	0.28	0.25	0.37	0.26	0.28	0.10	0.03
<i>dCE</i>	-	-	-	-	-	-	-	0.01	-	-

Table 3. continued

	Hazara	Pashto	Tajik	Turkmen	Uzbek	Western Europe	Sub-Sahara	West Asia	South Asia	East Asia
Kell										
<i>KEL*01.1</i>	0.97	0.95	0.99	0.99	0.94	0.91	0.98	0.98	0.96	0.99
<i>KEL*02</i>	0.03	0.05	0.01	0.01	0.06	0.09	0.02	0.02	0.04	0.01

Table 4. Frequencies of 15 biallelic polymorphisms in five Afghan and eight Pakistani populations.

System	Allele	Afghanistan					Pakistan							
		Hazara n=62	Pashto n=71	Tajik n=123	Turkmen n=67	Uzbek n=108	Balochi n=21	Brahui n=19	Burusho n=24	Hazara n=20	Kalash n=24	Makrani n=17	Pathan n=22	Sindhi n=21
MNS	<i>GYPB*03</i>	0.23	0.36	0.30	0.22	0.30	0.40	0.45	0.21	0.28	0.19	0.35	0.39	0.38
	<i>GYPB*04</i>	0.77	0.64	0.70	0.78	0.70	0.60	0.55	0.79	0.73	0.81	0.65	0.61	0.62
KEL	<i>KEL*01.1</i>	0.02	0.04	0.00	0.01	0.04	0.00	0.03	0.02	0.03	0.00	0.00	0.00	0.00
	<i>KEL*02</i>	0.98	0.96	1.00	0.99	0.96	1.00	0.97	0.98	0.98	1.00	1.00	1.00	1.00
	<i>KEL*02.06</i>	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FY	<i>FY*01</i>	0.64	0.52	0.54	0.47	0.62	0.48	0.58	0.58	0.73	0.54	0.47	0.59	0.69
	<i>FY*02</i>	0.36	0.48	0.46	0.53	0.38	0.52	0.42	0.42	0.28	0.46	0.53	0.41	0.31
	<i>FY*02M.01</i>	0.00	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
	<i>FY*02N.01</i>	0.01	0.00	0.01	0.03	0.00	0.12	0.03	0.00	0.00	0.00	0.18	0.00	0.00
DI	<i>DI*01</i>	0.00	0.01	0.01	0.03	0.00	0.02	0.00	0.00	0.03	0.00	0.00	0.00	0.02
	<i>DI*02</i>	1.00	0.99	0.99	0.97	1.00	0.98	1.00	1.00	0.98	1.00	1.00	1.00	0.98
DO	<i>DO*01</i>	0.39	0.37	0.37	0.38	0.36	0.29	0.45	0.31	0.28	0.42	0.62	0.45	0.29
	<i>DO*02</i>	0.61	0.63	0.63	0.62	0.64	0.71	0.55	0.69	0.73	0.58	0.38	0.55	0.71
	<i>DO*02.-04</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>DO*01.-05</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Continued.

System	Allele	Afghanistan					Pakistan								
		Hazara n=62	Pashto n=71	Tajik n=123	Turkmen n=67	Uzbek n=108	Balochi n=21	Brahui n=19	Burusho n=24	Hazara n=20	Kalash n=24	Makrani n=17	Pathan n=22	Sindhi n=21	
IN	<i>IN*01</i>	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	
	<i>IN*02</i>	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	
LW	<i>LW*05</i>	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	<i>LW*07</i>	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
CO	<i>CO*01.01</i>	0.99	1.00	1.00	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	
	<i>CO*02</i>	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	
JK	<i>JK*01</i>	0.52	0.51	0.51	0.60	0.51	0.55	0.50	0.48	0.60	0.73	0.71	0.50	0.50	
	<i>JK*02</i>	0.48	0.49	0.49	0.40	0.49	0.45	0.50	0.52	0.40	0.27	0.29	0.50	0.50	
YT	<i>YT*01</i>	0.90	0.91	0.90	0.88	0.94	0.90	1.00	0.92	0.90	0.83	0.91	0.84	0.86	
	<i>YT*02</i>	0.10	0.09	0.10	0.12	0.06	0.10	0.00	0.08	0.10	0.17	0.09	0.16	0.14	