

Y-CHROMOSOMAL HAPLOGROUPS IN BULGARIANS

Sena Karachanak, Simona Fornarino*, Viola Grugni*, Ornella Semino*, Draga Toncheva, Angel Galabov**, Boris Atanasov***

(Submitted on December 22, 2008)

Abstract

In order to reconstruct Bulgarian paternal genetic history, we analysed the Y chromosome diversity across the country. This initial effort was performed through genotyping of 13 biallelic markers in samples of 127 Bulgarian males. These markers define the main Y chromosome haplogroups and some of their subclades. We determined five predominant haplogroups, which probably reflect three different steps in the colonization of the Balkan Peninsula. The obtained pattern of the Bulgarian Y chromosome diversity was compared to that of populations previously analysed at the same phylogenetic resolution. Based on this comparison and on historical data, we estimated definitely greater Indo-European (IE) than Central Asian (Altaic, AL) contribution to the modern Bulgarian gene pool.

Key words: Y chromosome, Bulgarians, UEP, haplogroup

Introduction. The most powerful tool for recovery of the genetic record of paternal history of modern populations is the study of the Y chromosome diversity [1]. The male specific region of the Y chromosome (MSY), which comprises around 95% of the chromosome length, is the genetic system of holandric (father to son) inheritance, which escapes the reshuffling effect of the homologous recombination. Hence, the only possible source of its sequence variation is the progressive accumulation of mutational events along paternal lineages. Furthermore, the lower effective population size of the Y chromosome in comparison with the other components of the human genome further contributes to its lower sequence diversity and makes it particularly sensitive to the influence of drift and founder effect [2].

This study was supported by the “Bulgarian consortium for structural genomics and *in silico* drug design” Project of the Ministry of Education and Science Contract No DRI 5/07.02.2006, the Compagnia di San Paolo and the Italian Ministry of Education: Progetti Ricerca Interesse Nazionale 2007 (to O.S.). We would like to express our thanks to ELTA'90 Ltd. for the financial support.

The Y specific biallelic markers, mostly represented by single nucleotide polymorphisms (SNPs), are characterized by very low mutational rates (10^{-7} - 10^{-8} /bp/generation). They are considered to have occurred only once during modern human diaspora and are termed unique event polymorphisms (UEPs). They define the haplogroups (Hg-s) of the Y chromosome phylogeny. Since UEPs occurred during and after the human dispersal, haplogroups distribution displays correspondence with geography, thus providing patterns of affinity among modern populations and clues concerning past human movements.

The territory of present-day Bulgaria was one of the places from where modern humans set foot in Europe. It also played an important role in the post-glacial recolonization of the continent by low density Palaeolithic populations and in the transition from hunter-gathering to farming society. After the Black Sea Flood [3] there was an expansion of the West-Pontic populations. It is archaeologically documented as Varna necropolis culture (6th–5th millennium BC) and the Vincha culture (4th millennium BC), the carriers of which resettled all Danube River Valley.

When referring to historical times, several demographic events may have left significant imprint on the modern Bulgarian gene pool. These include Cimmerian migrations, the wide distribution of enormous Thracian populations (more than 90 tribes), and later on the settlements of Slavs [4], ancient Bulgarians and some Asian incomers.

To date, the Bulgarian genetic paternal history still has not been directly addressed and despite the importance of the region, the studies of the Bulgarian Y chromosome diversity are scarce and restricted to few haplogroups or only a few individuals [5–8].

In order to unravel Bulgarian paternal past, we genotyped a set of 13 biallelic markers in a sample of 127 Bulgarian males. These markers define the major MSY haplogroups and some of their subclades. We determined five predominant haplogroups, which probably reflect three different steps in the colonization of present-day Bulgaria. The obtained pattern of the Bulgarian Y chromosome diversity was compared to that of populations previously analysed at the same phylogenetic resolution. Based on this comparison and on historical data, we estimated definitely more greater Indo-European (IE) than Central Asian (Altaic, AL) [9] contribution to the modern Bulgarian gene pool.

Materials and methods. *Sampling.* A total of 127 blood DNA samples from healthy, unrelated Bulgarian males were analysed. Appropriate informed consent and information on paternal birthplace were obtained from each participant. Individuals were categorized according to the paternal birthplace and they were chosen in a way that the majority of the provinces of Bulgaria would be represented. DNA samples were isolated by using phenol-chloroform extraction, followed by ethanol precipitation.

MSY analysis. Thirteen MSY biallelic markers (M9, M17, M35, M69, M70,

M89, M170, M201, M207, M214, M269, M304 and YAP) were analysed. M89 was typed in all samples, while the other markers were typed hierarchically according to their known phylogeny. For each UEP studied, information concerning primer sequence, amplicon size, the allelic states and nucleotide position in the amplicon are given in Table 1. Markers M35, M69, M70, M89, M170, M201, M214 and M304 were genotyped by using PCR-DHPLC (denaturing high-performance liquid chromatography) according to [10]. Markers M9, M17, M207 and M269 were examined using PCR/RFLP assay with the appropriate restriction enzymes (*Hinf*I, *Afl*III, *Dra*I, and *Tsp*509I, respectively). YAP was typed by PCR/AFLP analysis according to [11].

Haplogroups were named following the terminological conventions recommended by the Y Chromosome Consortium, as the clades excluded from a haplogroup are listed within parentheses after the name of the haplogroup [12].

Results and discussion. The informative set of 13 UEPs defines 11 main haplogroups. With the exception of Hg-s NO, all defined haplogroups were identified in the Bulgarian population. The phylogenetic relationships among the analysed markers, the nomenclature of the haplogroups and their percent frequencies in Bulgarians are shown in Fig. 1. Five haplogroups: E-M35, I-M170, J-M304, R-M17 and R-M269 (referred as: E3b, I, J, R1a and R1b in [9], respectively) comprise 95% of the total genetic variation in Bulgarians.

Haplogroup E-M35 accounts for approximately one-fifth (19.7%) of the Bulgarian Y chromosomes. Only one of the samples derived at YAP, remained unresolved at the DE haplogroups level. Both phylogeography and microsatellite variance suggest that E-M35 originated in Eastern Africa, from where it is be-

T a b l e 1
Examined markers and their characteristics

Marker	Primers sequence (5' - 3')		Amplicon size (bp)	Allelic states		Position
	Forward	Reverse		ancestral	derived	
M9	gcagcatataaaactttcagg	aaaacctaactttgctcaagc	340	C	G	068
M17	ctggtcataaactggaatc	tgaactacaaatgtgaaact	333	4G	3G	068
M35	taagcctaaagagcagtcagag	taagcctaaagagcagtcagag	513	G	C	168
M69	ggttatcatagcccactatactttg	agcaagacaaaggaataaagat	257	T	C	222
M70	ggttatcatagcccactatactttg	agcaagacaaaggaataaagat	257	A	C	045
M89	agaagcagattgatgtccact	tccagtaggatcccctca	527	C	T	347
M170	tgttcacacaaatgcgttt	ggtcttaaatgttgaaagtaattgg	405	A	C	327
M201	tatgcatttgttgagtatgtc	gttctgaatgaaagttcaaacg	144	T	G	259
M207	aggaaaaatcagaagtatccctg	caaaattaccaagaatccttg	423	A	G	079
M214	tattacaaaatgatgaaacaagcg	gaaatgccacttcactccag	460	T	C	404
M269	ctaaagatcagagtatctccctttg	aaattgtttcaatttaccag	379	C	T	358
M304	caaagtgtgggattacagg	cttctagcttcatctgattgt	496	A	G	212
YAP	caggggaagataaagaata	actgctaaaaggggatgat	150/455	Alu(-)	Alu(+)	-

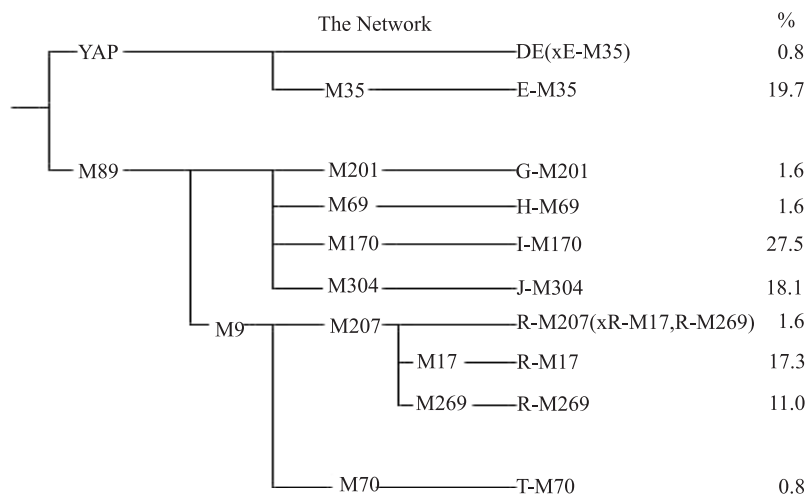


Fig. 1. Phylogenetic network of Y chromosome haplogroups observed in Bulgarians and their respective percent frequencies

lieved to have arrived in Europe through the Near East/Anatolia. The great majority of Balkan E-M35 Y chromosomes belongs to the E-M78 subclade, which reached the Balkans anytime after 17 000 years ago and expanded to Europe in Neolithic times [13]. During this period a high civilization flourished around the Pontic Basin [4]. One of the cultures associated with this civilization is the Varna necropolis culture [14]. About 8500 years ago, after the raising of the water level of the Black Sea, part of the West-Pontic farmers spread the “Linear Pottery” (which reached present-day France via Central Europe and dispersed beyond the Carpathians) and the Vincha culture (which spread along the Danube River Valley) [4]. Based on these data, we can consider the E-M35 component of the modern Bulgarian gene pool as “prehistorically domestic” one.

Haplogroup I accounts for 27.5% of the total sample, thus being the most commonly represented Y chromosome lineage in the Bulgarian gene pool. This haplogroup is widespread over Europe and West Asia: Anatolia, Lebanon, Mideast Arab, Armenia, Georgia up to Persia [9]. It was supposed that it represents the paternal genetic component of pre-Neolithic Europeans. Its initial spread in Europe may be linked to the diffusion of the largely pan-European Gravettian technology [15]. Haplogroup I marks the post Last Glacial Maximum recolonization of Europe from isolated refuge areas. Its two major subclades reach their highest frequencies in two distinct regions of Europe: Scandinavia (Germanous) and the Balkans (Cimmero-Sarmatians) [16]. Because this haplogroup covers the Iranian Plato up to region of Punjab, we hypothesize that the main migration stream was directed East-West and that it represents the wide community of Below-Caspian IE populations, which gave origin to many tribes and nations like Cimmerians (which migrated up to Ireland), old Tocharians (which migrated up to the Tarim River Valley), the Celtic tribes, Sarmatian peoples (to whom ancient Bulgarians

belong), Persian and Baluchi tribes. This link binds present Bulgarians with their oldest great-grandfathers.

Haplogroup J comprises 18.1% of the total Bulgarian sample. Since long time, this haplogroup has been considered to represent a signature of Middle Eastern migrations, which included the Neolithic demic diffusion associated with the spread of agriculture [17,18]. One of the haplogroup J subclades (J-M12) displays frequency distribution similar to that of the aforementioned E-M35 subclade. This is probably, a consequence of the same microevolutionary event, namely, an expansion from South-Eastern to continental Europe following the Danube waterways during the Neolithic [14]. From archaeological point of view, this population can be associated with Vincha farmers. However, this haplogroup is widely distributed on the territory of Anatolia, Mideast Arab, Morocco, Sudan and Ethiopia, Persia, Georgia, Armenia, Uzbek-Kazakhstan-Kirgizstan up to the Tarim River Valley, too. Thus this has to be a megapopulation in that time. We hypothesized that it can be only the Thracian one – from West Balkan (Dardanians, Illyrians, Pelagonians, Macedonians, etc.), North Balkan (Bessians, Gateans, Dacians, Misians, Medians, Thracoiranian Cimmerians, Seenians, etc.), South Balkan (Brigians, Peonians, Triballians, Odrysians, Pelasgians, etc.), over all Anatolia (Vithinians, Frisians, Trojans) and at the East Black Sea coast (East Medians, Frigo-Armenians, Parthians, Massagetes, etc.), covering all North Mesopotamia, Lebanon, Palestine and North Egypt (Mitanni, Phillistinians, South-Pelasgians and many others). For Thracians Herodotus had written that they are “...the greatest and most populous on Earth (after the Indians)”. Besides this, Herodotus accounted that the Greeks absorbed an earlier non-Greek population (Pelasgians) and that the Greek history is based on their traditions [5]. In summary, we can put Thracian assignation to the second most prevalent MSY haplogroup in Bulgarians.

Haplogroup R in Bulgarians is mainly represented by its two clades R-M17 and R-M269, which occur at 17.3% and 11.0% frequencies, respectively. Their understanding is of great importance for the Bulgarian history.

Haplogroup R-M17 (R1a) is widespread in Central Asian Turkic-speaking populations (from East to West: Altaics, Uyghurs, Kirgisiens, Selkups, Uzbekistanis), in eastern European (Kazan Tatars, Bashkirs), in Finno-Ugric (Finns/Suomi, Estonians, Hungarians), in Slavic (Russians, Ukrainians, Poles) and in Germanic (Germans, Swedes, Norwegians) speaking populations and has also been found less frequently in populations of the Caucasus and the Middle East, in India, and in Sino-Tibetans of Northern China. To date, in literature there is no yet a clear consensus about the place and time of origin of this haplogroup. From one side, its spread has been attributed to expansion(s) from a Ukrainian glacial refuge or to infiltrations from Southern Russia [15]. Alternatively, an Asian source or a deeper Palaeolithic time depth for the occurrence of M17 has been suggested [19]. We hypothesize that it represents the East-West “Great Migration of the Nations”

(culminating with the Hun invasion), which transferred lot of Altaic peoples to the West. It is known that in the 6th century a big group of Chachkent/Tashkent Bulgarians have been settled in the Rhodope Mountains; that many Central Asian peoples migrated together with Asparuch's incoming to the Balkan Peninsula; that the Bulgarian history is full with invasions of Pecheneges, Kumans and other Turkic speaking populations [20]. Thus the obtained R-M17 frequency in Bulgarians is an under-expected result.

Haplogroup R-M269 (R1b) frequency in Europe shows a major West-to-East decline: as it increases from the Middle East to North-Western Ireland. It represents the diversification of the Upper Palaeolithic precursor (R-M45) arrived in Europe from Western Asia and spread with the Aurignacian culture [15,19]. The opposite frequency distribution of lineages internal to this haplogroup identified with the complex 49a,f/*TaqI* system, possibly reflects re peopling of Europe from Iberia and Asia Minor during the Late Upper Palaeolithic and Holocene. The 49a,f and microsatellite haplotypes distribution associated with the R-M269 in the Balkan Peninsula suggested its possible arrival from two different source populations during the recolonization of Europe [17]. The R-M269 frequency peak observed in Basques is due to genetic drift of lineages which have arrived from the Caucasus (probably with the Cim merian migrations). We can assume that the high R-M269 frequency in Wales, Scotland and Ireland is a consequence of the same event. Historical data point that after Attila drove back, a huge contingent of Alans and other Caucasians, resettled South France and the Iberian Peninsula. They probably left their genetic imprint in Hun, Germanic and Proto-Bulgarian tribes. Therefore, the observed frequency of this marker agrees with the expectations.

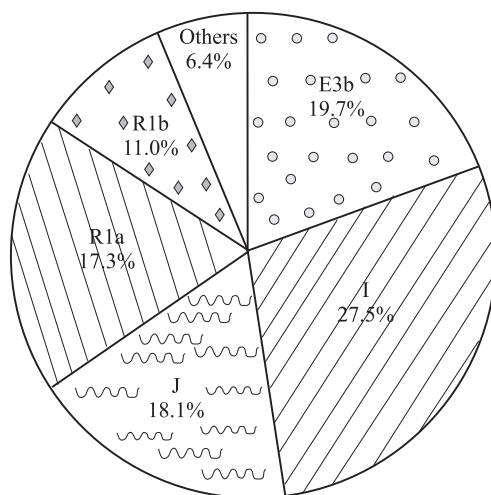


Fig. 2. Distribution of Y chromosome haplogroups in Bulgarians. Haplogroup designations correspond to those in [9]

The frequencies of the other major haplogroups observed in Bulgarians (G-M201 – shows affinity with the south Caucasus populations, H-M69 – associated to the Roma and T-M70 – only sporadically observed in Europe) are lower than 2%.

Speaking of the Bulgarian Y chromosome diversity, analysed at the level of main binary haplogroups and some of their subclades, we can draw the following conclusions: (1) 95% of the Y chromosome diversity is represented by five haplogroups: E-M35, I-M170, J-M304, R-M17 and R-M269 (Fig. 2), which probably reflect three different periods of the colonization of the Balkan Peninsula; (2) the greatest part of the present-day Bulgarian gene pool is typical for European populations, and (3) the ratio between European (I-M170 and R-M269) and Altaic (R-M17) components of the Bulgarian gene pool is greater than two (2.3).

REFERENCES

- [1] UNDERHILL P. A. *Cold Spring Harb. Symp. Quant. Biol.*, **68**, 2003, 487–493.
- [2] JOBLING M. A., C. TYLER-SMITH. *Nat. Rev. Genet.*, **4**, 2003, 598–612.
- [3] RYAN W., W. PITMAN. *Noah's Flood*, New York, Simon & Schuster Publ., 1998.
- [4] MALLORY J. P. *In the Search of the Indo-Europeans*, New York, Thames & Hudson Publ., 1889/2001.
- [5] ROSSER Z. H., T. ZERJAL, M. E. HURLES, M. ADOJAAN, D. ALAVANTIC, A. AMORIM, W. AMOS, M. ARMENTEROS, E. ARROYO, G. BARBUJANI, G. BECKMAN, L. BECKMAN, J. BERTRANPETIT, E. BOSCH, D. G. BRADLEY, G. BREDE, G. COOPER, H. B. CÔRTE-REAL, P. DE KNIJFF, R. DECORTE, Y. E. DUBROVA, O. EVGRAFOV, A. GILISSEN, S. GLISIC, M. GÖLGE, E. W. HILL, A. JEZIOROWSKA, L. KALAYDJIEVA, M. KAYSER, T. KIVISILD, S. A. KRAVCHENKO, A. KRUMINA, V. KUCINSKAS, J. LAVINHA, L. A. LIVSHITS, P. MALASPINA, S. MARIA, K. McELREAVEY, T. A. MEITINGER, A. V. MIKELSAAR, R. J. MITCHELL, K. NAFA, J. NICHOLSON, S. NØRBY, A. PANDYA, J. PARIK, P. C. PATSALIS, L. PEREIRA, B. PETERLIN, G. PIELBERG, M. J. PRATA, C. PREVIERÉ, L. ROEWER, S. ROOTSI, D. C. RUBINSZTEIN, J. SAILLARD, F. R. SANTOS, G. STEFANESCU, B. C. SYKES, A. TOLUN, R. VILLEMS, C. TYLER-SMITH, M. A. JOBLING. *Am. J. Hum. Genet.*, **67**, 2000, 1376–81.
- [6] CRUCIANI F., R. LA FRATTA, P. SANTOLAMAZZA, D. SELBITTO, R. PASCONE, P. MORAL, E. WATSON, V. GUIDA, E. B. COLOMB, B. ZAHAROVA, J. LAVINHA, G. VONA, R. AMAN, F. CALI, N. AKAR, M. RICHARDS, A. TORRONI, A. NOVELLETTO, R. SCOZZARI. *Am. J. Hum. Genet.*, **74**, 2004, 1014–102.
- [7] DI GIACOMO F., F. LUCA, L. O. POPA, N. AKAR, N. CIAVARELLA, F. CUCCI, L. DI STASI, L. GAVRILA, M. G. KERIMOVA, D. KOVATCHEV, A. I. KOZLOV, A. LOUTRADIS, V. MANDARINO, C. MAMMI, E. N. MICHALODIMITRAKIS, G. PAOLI, K. I. PAPPAS, G. PEDICINI, L. TERRENATO, S. TOFANELLI, P. MALASPINA, A. NOVELLETTO. *Hum. Genet.*, **115**, 2004, 357–371.
- [8] MALASPINA P., A. I. KOZLOV, F. CRUCIANI, P. SANTOLAMAZZA, N. AKAR, D. KOVATCHEV, M. G. KERIMOVA, J. PARIK, R. VILLEMS, R. SCOZZARI, A. NOVELLETTO. In: *Ancient interactions: east and west in Eurasia* (eds K. Boyle, C. Renfrew, M. Levine). McDonald Institute Monographs, 2003, 309–313.

- [9] <http://www.scs.uiuc.edu/~mcdonald/WorldHaplogroupsMaps.pdf>
- [10] UNDERHILL P. A., L. JIN, A. A. LIN, S. Q. MEHDI, T. JENKINS, D. VOLLRATH, R. DAVIS, L. L. CAVALLI-SFORZA, P. J. OEFNER. *Genome Res.*, **7**, 1997, 996–1005.
- [11] HAMMER M. F., S. HORAI. *Am. J. Hum. Genet.*, **56**, 1995, 951–62.
- [12] Y Chromosome Consortium (YCC), *Genome Res.*, **12**, 2002, 339–348.
- [13] CRUCIANI F., R. LA FRATTA, B. TROMBETTA, P. SANTOLAMAZZA, D. SELLITO, E. B. COLOMB, J. DUGOUJON, F. CRIVELLARO, T. BENINCASA, R. PASCONI, P. MORAL, E. WATSON, B. MELEGH, G. BARBUJANI, S. FUSELLI, G. VONA, B. ZAGRADISHNIK, G. ASSUM, R. BRDICKA, A. I. KOZLOV, G. D. EFREMOV, A. COPPA, A. NOVELLETO, R. SCOZZARI. *Mol. Biol. Evol.*, **24**, 2007, 1300–1311.
- [14] RENFREW C. *Antiquity*, **52**, 1978, 197–203.
- [15] SEMINO O., G. PASSARINO, P. J. OEFNER, A. A. LIN, S. ARBUZOVA, L. E. BECKMAN, G. DE BENEDICTIS, P. FRANCALACCI, A. KOVATSI, S. LIMBORSKA, M. MARCIKI, A. MIKA, B. MIKA, D. PRIMORAC, S. A. SANTACHIARA-BENERECETTI, L. L. CAVALLI-SFORZA, P. A. UNDERHILL. *Science*, **290**, 2000, 1155–1159.
- [16] ROOTSI S., C. MAGRI, T. KIVISILD, G. BENUZZI, H. HELP, M. BERMISHEVA, I. KUTUEV, L. BARAC, M. PERICIC, O. BALANOVSKY, A. PSHENICHNOV, D. DION, M. GROBEI, L. A. ZHIVOTOVSKY, V. BATTAGLIA, A. ACHILLI, N. AL-ZAHERY, J. PARIK, R. KING, C. CINNIUGLU, E. KHUSNUTDINOVA, P. RUDAN, E. BALANOVSKA, W. SCHEFFRAHN, M. SIMONESCU, A. BREHM, R. GONCALVES, A. ROSA, J. MOISAN, A. CHAVENTRE, V. FERAK, S. FUREDI, P. J. OEFNER, P. SHEN, L. BECKMAN, I. MIKEREZI, R. TERZIC, D. PRIMORAC, A. CAMBON-THOMSEN, A. KRUMINA, A. TORRONI, P. A. UNDERHILL, A. S. SANTACHIARA-BENERECETTI, R. VILLEMS, O. SEMINO. *Am. J. Hum. Genet.*, **75**, 2004, 128–137.
- [17] CINNIUGLU C., R. KING, T. KIVISILD, E. KALFOGLU, S. ATASOY, G. L. CAVALLERI, A. S. LILLIE, C. C. ROSEMAN, A. A. LIN, K. PRINCE, P. J. OEFNER, P. SHEN, O. SEMINO, L. L. CAVALLI-SFORZA, P. A. UNDERHILL. *Hum. Genet.*, **114**, 2004, 127–148.
- [18] SEMINO O., C. MAGRI, G. BENUZZI, A. A. LIN, N. AL-ZAHERY, V. BATTAGLIA, L. MACCIONI, C. TRIANTAPHYLIDIS, P. SHEN, P. J. OEFNER, L. A. ZHIVOTOVSKY, R. KING, A. TORRONI, L. L. CAVALLI-SFORZA, P. A. UNDERHILL, A. S. SANTACHIARA-BENERECETTI. *Am. J. Hum. Genet.*, **74**, 2004, 1023–1034.
- [19] WELLS S. R., N. YULDASHEVA, R. RUZIBAKIEV, P. A. UNDERHILL, I. EVSEEVA, J. BLUE-SMITH, L. JIN, B. SU, R. PITCHAPPAN, S. SHANMUGALAKSHMI, K. BALAKRISHNANG, M. READH, N. M. PEARSONI, T. ZERJALJ, M. T. WEBSTERK, I. ZHOLOSHVILIL, E. JAMARJASHVILIL, S. GAMBAROV, B. NIKBINN, A. DOSTIEVO, O. AKNAZAROV, P. ZALLOUAQ, I. TSOYR, M. KITAEVS, M. MIRRAKHIMOV, A. CHARIEVT, W. F. BODMERA. *Proc. Natl. Acad. Sci. USA*, **98**, 2001, 10244–10249.
- [20] DOBREV P. *The Golden Core of the Bulgarian Antiquity*, TANGRA Publ., Sofia, 2005.

Department of Medical Genetics
Medical University
1134 Sofia, Bulgaria

**Department of Genetics and Microbiology*
University of Pavia
27 100 Pavia, Italy

***Institute of Microbiology*
Bulgarian Academy of Sciences
1113 Sofia, Bulgaria

****Institute of Organic Chemistry*
Bulgarian Academy of Sciences
1113 Sofia, Bulgaria
e-mail: boris@orgchm.bas.bg