Legal Medicine 12 (2010) 208–211

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Brief Communication

A Japanese-specific allele in the *GALNT11* gene

Isao Yuasa [a,*], Kazuo Umetsu [b], Aya Matsusue [c], Hiroaki Nishimukai [d], Shinji Harihara [e], Yasuo Fukumori [f], Naruya Saitou [g], Feng Jin [h], Prasanta K. Chattopadhyay [i], Lotte Henke [j], Jürgen Henke [j]

[a] Division of Legal Medicine, Faculty of Medicine, Tottori University, Yonago 683-8503, Japan
[b] Department of Forensic Medicine, Faculty of Medicine, Yamagata University, Yamagata, Japan
[c] Department of Forensic Medicine, Fukuoka University School of Medicine, Fukuoka, Japan
[d] Division of Legal Medicine, Ehime University School of Medicine, Ehime, Japan
[e] Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan
[f] Osaka Red Cross Blood Center, Osaka, Japan
[g] Division of Population Genetics, National Institute of Genetics, Mishima, Japan
[h] Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[i] Amity Institute of Forensic Sciences, Defence Colony, New Delhi, India
[j] Institut für Blutgruppenforschung, Cologne, Germany

ARTICLE INFO

Article history:
Received 15 February 2010
Received in revised form 29 March 2010
Accepted 4 April 2010
Available online 23 May 2010

Keywords:
Ancestry-informative markers (AIMs)
Genotyping
Single nucleotide polymorphism (SNP)
Population-specific
Population study

ABSTRACT

In this study, five single nucleotide polymorphisms (SNPs) in the *ABCC4*, *FBN1*, *CEP152*, *ZNF804B*, and *GALNT11* genes were investigated to assess allele frequencies in 14 different populations by a novel pentaplex PCR method. All SNPs were polymorphic in East Asians, whereas mutant alleles were absent or rare in non-East Asians. The frequencies of a mutant allele in *FBN1* (rs140598) showed a north–south downward cline in East Asia, whereas those of a mutant allele in *ZNF804B* (rs1916830) were relatively uniform in East Asia. The highest frequencies of mutant alleles in *ABCC4* (rs3765534), *CEP152* (rs2289178), and *GALNT11* (rs3778922) were observed in Okinawa. The mutant allele in *GALNT11* was found only in Far-East Asian populations: the frequencies were about 0.153 in Okinawa, 0.076 in the main island of Japan, and 0.017–0.004 in Korea. These five East Asian- and Japanese-specific SNPs would be useful markers for forensic individualization, in particular, as ancestry-informative markers.

1. Introduction

Inference of the ancestry and population to which an individual belongs is an important aspect of forensic individualization. Information about the ancestry of the perpetrator of a crime in the absence of eyewitness evidence and of victims in disasters is invaluable. Genetic markers with a geographically restricted distribution, or with marked allele frequency differences between two regions, are effective for this purpose. Short tandem repeat (STR) polymorphisms have successfully been applied to forensic individualization, because they are the most variable types of DNA sequence in the genome. Increases and decreases in repeat number arise from replication slippage, which can occur in every human population. Regionally private STR alleles with a frequency above 0.02 have sometimes been detected on autosomal chromosomes. However, only a few alleles with a frequency above 0.13 have been observed [1–3]. In contrast, single nucleotide polymorphisms (SNPs) sometimes show significant differences in allele frequency among human populations. Such population-specific alleles are known as ancestry-informative markers (AIMs) [4,5]. The distribution patterns of allele frequencies are also useful for evolutionary elucidation of population migration. Some of these alleles have spread by positive selection and are associated with physical traits such as skin pigmentation [6]. In East Asia, genetic differentiation between northern and southern populations has been observed using classic and DNA markers [7]. Japanese from the main islands and those from Okinawa were distinguished using 140 K SNPs [8]. Japanese-specific alleles are powerful for the characterization of a forensic sample as being of Japanese origin, but only a few alleles have been identified to date. In this study, five SNPs in the *ABCC4*, *FBN1*, *CEP152*, *ZNF804B*, and *GALNT11* genes were investigated, which are known to be polymorphic in Japanese and Chinese populations, but are shown to be monomorphic in people of European descent and the Yoruba people of the international HapMap project. To assess allele frequencies for the five SNPs in 14 different populations, a novel simultaneous genotyping method was developed, based on the pentaplex PCR. Of the five mutant alleles, an allele in the *GALNT11* gene was shown to be Japanese-specific.

* Corresponding author. Tel.: +81 859 38 6123; fax: +81 859 38 6120.
E-mail address: yuasai@med.tottori-u.ac.jp (I. Yuasa).

1344-6223/$ - see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.legalmed.2010.04.001

2. Materials and methods

2.1. DNA samples

DNA samples were obtained from 1541 unrelated individuals living in various areas of Eurasia and South America and were used for a population study. Sub-Saharan African (Nigerian and Ghanaian) and Turkish samples were collected from people residing in Germany (Table 1). Most of these samples were from a set used in previous studies [9–12], where the geographical locations of the investigated populations were also shown. This study was approved by the Ethical Committee at the Faculty of Medicine, Tottori University.

2.2. Genotyping procedure

Five SNPs, rs3765534, rs140598, rs2289178, rs1916830, and rs3778922, were investigated as shown in Table 1. These SNPs were simultaneously genotyped by pentaplex PCR based on the amplified product length polymorphism (APLP) method [9–12]. The nucleotide sequence and final concentration of each primer are shown in Table 2. The PCR cocktail consisted of 50 μl of Multiplex PCR Master Mix from a Multiplex PCR Kit (Qiagen, Hilden, Germany), 8.25 μl of 15 primers with a concentration of 100 pmol/μl, and 41.75 μl of water. PCR was performed in a volume of 8 μl containing 7.5 μl of the PCR cocktail and 0.5 μl of a solution containing 10–20 ng of genomic DNA. The cycle conditions were 95 °C for 15 min; then 30 cycles of 94 °C for 10 s, 54 °C for 10 s, 72 °C for 10 s, and a final extension step of 15 min at 72 °C. The products were separated using a polyacrylamide gel (9%T, 5%C) together with positive and negative controls and then visualized by ethidium bromide staining.

2.3. Statistical analysis

The estimation of allele and haplotype frequencies and population pairwise F_{ST}, and testing of Hardy–Weinberg equilibrium and linkage disequilibrium were carried out using the Arlequin program ver. 3.11 [13]. The statistics D' and r^2 values were measured according to Hartl and Clark [14].

3. Results and discussion

Fig. 1 shows the band patterns of the products obtained by pentaplex PCR. The nucleotide substitutions were clearly and unambiguously detected as bands of different sizes. The sizes ranged from 70 bp for *GALNT11*G* to 119 bp for *ABCC4*A*. Simultaneous genotyping by the present pentaplex PCR method is advantageous for investigating genotypes, because the amplicons at every locus are obtained by only one amplification without sampling errors. In addition, this method is technically simple and inexpensive and does not require specialized instruments.

We genotyped five SNPs in 1541 individuals from the 14 populations. The frequencies of the mutant allele at each locus are shown in Table 1. In each locus, there was no statistically significant departure from Hardy–Weinberg equilibrium except for the *CEP152* locus in Han Chinese at Huizhou ($p = 0.0356$), where the observed numbers of the homozygous types were increased, but the reason for this was unknown. The allele frequencies in our African, German, Chinese, and Japanese samples were similar to those in the SNP database. The results confirmed marked differences in frequency between East Asian and non-East Asian populations.

Some mutations in the *FBN1* gene are associated with Marfan syndrome and aortic aneurysm, but *FBN1*G* (p.1148A) is innocent for these diseases [15]. Buryats and Mongolians showed the highest frequencies for this allele among the 14 populations investigated in this study. The frequencies were higher in the northern parts of East Asia than in the southern parts. There was a significant correlation between allele frequencies and the degree of latitude in 9 East Asian populations ($r = 0.935$, $p < 0.001$), showing a north–south downward cline in East Asia. ZNF804B is one of the zinc finger proteins, and its function has not well been characterized. *ZNF804B*A* showed a relatively uniform distribution in East Asia, with frequencies ranging from 0.35 in Okinawa to 0.49 in Tottori. Multiple drug resistance protein 4 (MRP4), the product of the *ABCC4* gene, mediates some form of xenobiotic transport and/or drug resistance. *ABCC4*A* (p.757 K) exhibits a 10% reduction in the level of protein expression and 20% higher transporter activities in comparison with the wild type [16]. *ABCC4*A* was mainly found in East Asian populations, and its frequency was somewhat higher in the northern parts of East Asia than in the southern parts. Interestingly, the highest frequencies were observed in Japan, in

Table 1
Frequencies of the mutant alleles in 14 populations with data from the HapMap project.

No.	Population	n	ABCC4 3q32 rs3765534 exon 18 c.2269 GAG>AAG Glu757Lys	FBN1 15q21.1 rs140598 exon 28 c.3442 CCC>GCC Pro1148Ala	CEP152 15q21.1 rs2289178 exon 18 c.2378 AGC>ATC Ser793Ile	ZNF804B 7q21.13 rs1916830 exon 4 c.743 TGT>TAT Cys248Tyr	GALNT11 7q34-q36 rs3778922 exon 5 c.589 GAT>TAT Asp197Tyr
1	African	68	0	0.0074	0	0.0074	0
2	German	92	0.0109	0	0	0.0054	0
3	Turk	109	0.0092	0.0092	0.0046	0.0413	0
4	Indian	107	0.0607	0.0140	0.0093	0.0935	0
5	Buryat	108	0.0833	0.4028	0.1343	0.3889	0
6	Mongolian	120	0.1292	0.3667	0.0875	0.4292	0
7	Korean (Seoul)	140	0.0893	0.2893	0.1143	0.4786	0.0036
8	Korean (Kwangju)	144	0.0903	0.2917	0.1215	0.4444	0.0174
9	Japanese (Tottori)	131	0.1412	0.2863	0.2901	0.4885	0.0763
10	Japanese (Okinawa)	88	0.2330	0.2727	0.3636	0.3523	0.1534
11	Han (Wuxi)	119	0.0420	0.2437	0.1261	0.4202	0
12	Han (Huizhou)	111	0.0541	0.2297	0.1982	0.4685	0
13	Thai	111	0.0450	0.2027	0.2162	0.3964	0
14	Columbian	93	0.0215	0.0430	0.1935	0.4140	0
	African (YRI)	53–60	0	0.009	0	0.025	0
	Caucasian (CEU)	53–60	0	0	0	0.017	0
	Japanese (JPT)	42–45	0.2	0.393	0.4	0.489	0.114
	Han Chinese (CHB)	42–45	0.067	0.214	0.133	0.433	0

Table 2
PCR primers for simultaneous typing of 5 SNPs and their final concentrations.

SNP	Primer name	Sequence (5′→3′)*	Final concentration (μM)
rs3765534	ABCC4-FA2	tatattATGGAGGAGGAAATGTtACCA	0.45
	ABCC4-FG	TGGAGGAGGAAATGTAtCCG	0.6
	ABCC4-R2	TCATTGTAAGCACCTCAGCC	0.6
rs140598	FBN1-F	aCTCTCCTATGCCGAGGTG	0.5
	FBN1-RG	tatatTACACGCGGAGATGTTaGC	0.5
	FBN1-RC	TACACGCGGAGATGTTtGG	0.5
rs2289178	CEP152-F	ttAAGGAGTGGCAGTCTAAGCTG	0.2
	CEP152-RT	tataTGGTTACTTGGTCAGaTTGGA	0.2
	CEP152-RG	TGGTTACTTGGTCAGTTaGGC	0.2
rs1916830	ZNF804B-FA	atatACACAGAAGAAACCCAaGATTA	0.9
	ZNF804B-FG	ACACAGAAGAAACCCATGtTTG	0.6
	ZNF804B-R	tCTGCAGCACTTGCACTTATC	0.6
rs3778922	GALNT11-FT	attaAATTGTGGGACTTTATTTTCtGATT	1.2
	GALNT11-FG	AATTGTGGGACTTTATTTTCAGtTG	0.6
	GALNT11-R	CAGGGAGGTATTTTTGGACAT	0.6

* Lowercase letters indicate noncomplementary nucleotide.

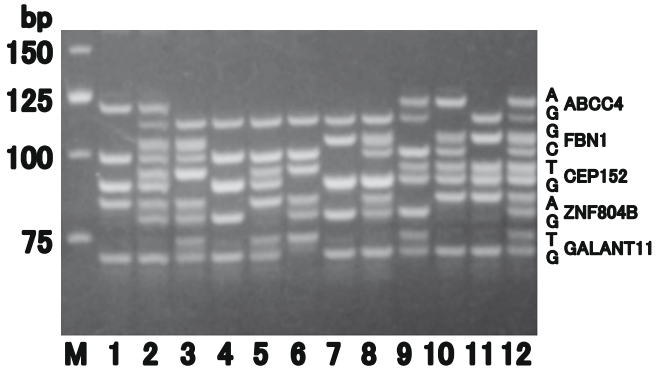

Fig. 1. Simultaneous genotyping of the five SNPs by pentaplex PCR based on the APLP method. Lane M: 25-bp ladder.

particular, in Okinawa where a high frequency of 0.23 was observed. The association between the prevalence of this allele in Okinawa and the expression level or functional activities of the MRP4 variant is unknown. CEP152, a centrosomal protein of 152 kDa, is essential for cilia formation [17]. GALNT11, one of the UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferases, is expressed only in kidney [18]. Although the significance of amino acid substitutions in these proteins is unknown, *CEP152*T* and *GALNT11*T* also showed the highest frequencies in Okinawa. *GALNT11*T* was found only in Far-East Asian populations. The frequencies were about 0.153 in Okinawa, 0.076 in Tottori, and 0.017–0.004 in Korea. Thus, this allele was nearly restricted to Japanese.

The F_{ST} values, obtained by comparisons between German and Japanese (Tottori) populations, ranged from 0.437 for *ZNF804B* to 0.061 for *GALNT11*. *FBN1* and *CEP152* are adjacent to each other on chromosome 15q21.1, and the two polymorphisms were found to be in significant linkage disequilibrium in five populations, including Buryat, Korean (Kwangju), Japanese (Tottori and Okinawa), and Han (Huizhou) populations. However, the low r^2 values suggested that linkages between them were relatively weak. Although both *ZNF804B* and *GALNT11* are located on chromosome 7q, significant linkage disequilibrium was not observed between them in any populations. The haplotype frequencies and D' and r^2 values are shown in Appendix A. The frequencies of people having at least one of five mutant alleles ranged from 0.029 in Africans to 0.971 in Japanese (Okinawa) (Appendix A).

There are three different hypotheses regarding the origins of East Asian populations: (i) South East Asian origin, (ii) North Asian origin, and (iii) a combination of northern and southern origins [7]. This suggests at least two main centers of human expansion and migration in East Asia. East Asian-specific alleles, *rs17822931*A* at *ABCC11* and *rs3827760*C* at *EDAR*, which are associated with ear wax type and hair thickness, respectively, are more frequent in the northern parts of East Asia, where they are nearly fixed [19,20]. *OCA2*481Thr*, which exhibits a 70% of the function of the wild type allele in terms of melanogenesis, was also found almost exclusively in the northern parts of East Asia [10]. A north–south downward cline of *FBN1*G* frequency observed in East Asia reflects the migration and expansion of ancient North East Asians. Some mutants in the *HERC1* gene have frequencies as high as about 0.9 in East Asians, but only about 0.1 in sub-Saharan African and German populations [12]. Similarly, *ZNF804B*A* was distributed uniformly in East Asia, with a relatively high F_{ST} value between Japanese in Tottori and Germans. *ABCC4*A*, *CEP152*T*, and *GALNT11*T* showed the highest frequencies in Okinawa. It is suggested that *ABCC4*A* and *CEP152*T* spread into the Japanese archipelago before the Neolithic Yayoi period (300 BC–300 AD). According to a dual structure model [21], modern Japanese are the result of an intermixture between the upper Paleolithic native population of Japan (Jomon people) and migrants from northeast Asia who passed through the Korean peninsula during the Neolithic Yayoi period. A representative Jomon allele, YAP+(M1) on the Y chromosome, is observed at a frequency of 0.556 in Okinawa, whereas it is observed at frequencies of 0.257–0.385 in the main islands of Japan. This allele is absent or low in other East Asians except for some populations including Tibetans [22,23]. The high frequencies of *ABCC4*A* and *CEP152*T* in Okinawa may reflect a smaller degree of migration of Yayoi ancestors to Okinawa. It is also possible that these two alleles have been maintained at high frequencies in Japanese populations partly through genetic drift during and after migration from the Asian continent to Japan.

Only a few Okinawa-specific alleles have been identified. *AHSG*5* in the alpha2-HS-glycoprotein gene is polymorphic in Okinawa with a frequency of 0.026, whereas it is quite rare or absent in other populations [24]. Some Japanese-specific alleles have been found. The fusion gene (se^{fus}) at the ABO-secretor locus (*FUT2*) was found in the Japanese populations with relatively high frequencies ranging from 0.048 to 0.079 in the main islands of Japan to 0.055 in Okinawa. This allele was found to be rare in a Korean population (0.006) and absent in two Chinese populations. This mutation was suggested to have arisen in the Japanese archipelago [25]. Recently, we also found a Japanese-specific allele in the complement factor I (*CFI*) gene. *CFI*As* was detected with a frequency of 0.03 in

two main island Japanese populations, whereas it was found to be rare in a Japanese population in Okinawa (0.006) and among Korean populations (0–0.007) [11]. The frequency of *GALNT11*T* in Japanese populations was found to be 0.076–0.153, whereas it was found to be 0.004–0.017 in Korean populations and absent in other populations. The frequency in the Japanese populations was about tenfold higher than that in the Korean populations. Evidently, *GALNT11*T* is a unique allele showing a high frequency in Okinawa. These findings suggest that *GALNT11*T* emerged in Japanese living in Okinawa after the divergence of the Japanese–Korean group from other populations and then very recently entered into the Korean populations. Conversely, *GALNT11*T* as well as the fusion gene (se^{fus}) may have arisen in the Korean peninsula. This new mutation could have spread to the Japanese islands through a phenomenon called allelic surfing [26]. However, *GALNT11*T* could show an irregular distribution in Asia in a fashion similar to that of the YAP+(M1) allele [22,23]. Further studies are needed to investigate the frequencies of *GALNT11*T* in other populations, including Central Asians.

In conclusion, genetic markers with a geographically restricted distribution, or with marked allele frequency differences between two regions, are useful to infer the ancestry of an individual. The distribution of the five SNPs investigated in this study was restricted to East Asia. One of them, *GALNT11*T*, was observed only in Far-East Asians, and the highest frequency was observed in Japanese living in Okinawa. These SNPs would be useful AIMs for forensic individualization, and also help to decipher the migration patterns of East Asians.

Acknowledgement

This study was supported in part by a Grant-in-Aid for Scientific Research (20590678 to IY) from the Japan Society for the Promotion of Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.legalmed.2010.04.001.

References

[1] Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, et al. A private allele ubiquitous in the Americas. Biol Lett 2007;3:218–23.
[2] Phillips C, Rodriguez A, Mosquera-Miguel A, Fondevila M, Porras-Hurtado L, Rondon F, et al. D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. Forens Sci Int Genet 2009;3:7–13.
[3] Yuasa I, Irizawa Y, Nishimukai H, Fukumori Y, Umetsu, K, Nakayashiki N, et al. A hypervariable STR polymorphism in the complement factor I (CFI) gene: asian-specific alleles. Int J Legal Med doi:10.1007/s00414-009-0369-0.
[4] Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 1998;63:1839–51.
[5] Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 2003;112:387–99.
[6] Sturm RA. Molecular genetics of human pigmentation diversity. Hum Mol Genet 2009;18:R9–R17.
[7] Zhang F, Su B, Zhang Y, Jin L. Genetic studies of human diversity in East Asia. Phil Trans R Soc B 2007;362:987–95.
[8] Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 2008;83:445–56.
[9] Yuasa I, Umetsu K, Harihara S, Kido A, Miyoshi A, Saitou N, et al. Distribution of two Asian-related coding SNPs in the MC1R and OCA2 genes. Biochem Genet 2007;45:535–42.
[10] Yuasa I, Umetsu K, Harihara S, Miyoshi A, Saitou N, Park KS, et al. OCA2*481Thr, a hypofunctional allele in pigmentation, is characteristic of northeastern Asian populations. J Hum Genet 2007;52:690–3.
[11] Yuasa I, Nakagawa M, Umetsu K, Harihara S, Matsusue A, Nishimukai H, et al. Molecular basis of complement factor I (CFI) polymorphism: one of two polymorphic suballeles responsible for CFI A is Japanese-specific. J Hum Genet 2008;53:1016–21.
[12] Yuasa I, Umetsu K, Nishimukai H, Fukumori Y, Harihara S, Saitou N, et al. HERC1 polymorphisms: population-specific variations in haplotype composition. Cell Biochem Funct 2009;27:402–5.
[13] Excoffier L, Laval G, Schneider S. Arlequin ver 3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 2005;1:47–50.
[14] Hartl DL, Clark AG. Principles of population genetics. 4th ed. Sunderland MA: Sinauer Associates; 2007. p. 45–92.
[15] Schrijver I, Liu W, Francke U. The pathogenicity of the Pro1148Ala substitution in the FBN1 gene: causing or predisposing to Marfan syndrome and aortic aneurysm, or clinically innocent? Hum Genet 1997;99:607–11.
[16] Janke D, Mehralivand S, Strand D, Godtel-Armbrust U, Habermeier A, Gradhand U, et al. 6-Mercaptopurine and 9-(2-phosphonyl-methoxyethyl) adenine (PMEA) transport altered by two missense mutations in the drug transporter gene *ABCC4*. Hum Mutat 2008;29:659–69.
[17] Blachon S, Gopalakrishnan J, Omori Y, Polyanovsky A, Church A, Nicastro D, et al. Drosophila *asterless* and vertebrate Cep152 are orthologs essential for centriole duplication. Genetics 2008;180:2081–94.
[18] Schwientek T, Bennett EP, Flores C, Thacker J, Hollmann M, Reis CA, et al. Functional conservation of subfamilies of putative UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyl- transferases in Drosophila, *Caenorhabditis elegans*, and mammals. One subfamily composed of *l(2)35Aa* is essential in Drosophila. J Biol Chem 2002;277:22623–38.
[19] Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, et al. A SNP in the *ABCC11* gene is the determinant of human earwax type. Nat Genet 2006;38:324–30.
[20] Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, et al. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 2008;17:835–43.
[21] Hanihara K. Dual structure model for the population history of Japanese. Jpn Rev 1991;2:1–33.
[22] Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, et al. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J Hum Genet 2006;51:47–58.
[23] Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, et al. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. BMC Biol 2008;6:45.
[24] Yuasa I, Taira T, Suenaga K, Ito K, Okada K. Determination of alpha 2HS-glycoprotein phenotypes by isoelectric focusing and immunoblotting: polymorphic occurrence of HSGA*5 in Okinawa. Hum Genet 1985;70:32–4.
[25] Liu YH, Koda Y, Soejima M, Pang H, Wang BJ, Kim DS, et al. The fusion gene at the ABO-secretor locus (FUT2): absence in Chinese populations. J Hum Genet 1999;44:181–4.
[26] Excoffier L, Ray N. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 2008;23:347–51.