

## **X Chromosome Linkage Disequilibrium Study in East Asian Genetic Isolates**

*Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, B. Munkhbat  
Department of Genetic Information, T. Katoh  
Biological Information Research Center, National Institute of  
Advanced Industrial Science and Technology, Tokyo S. Mano  
Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, T. Ikuta,  
Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, K. Tounai,  
Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, H. Ando,  
Biological Information Research Center, National Institute of  
Advanced Industrial Science and Technology, T. Imanishi  
Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, H. Inoko  
Department of Genetic Information, National Institute of Health,  
National Medical University of Mongolia, G. Tamiya  
National Institute of Health,  
National Medical University of Mongolia, N. Munkhtuvshin*

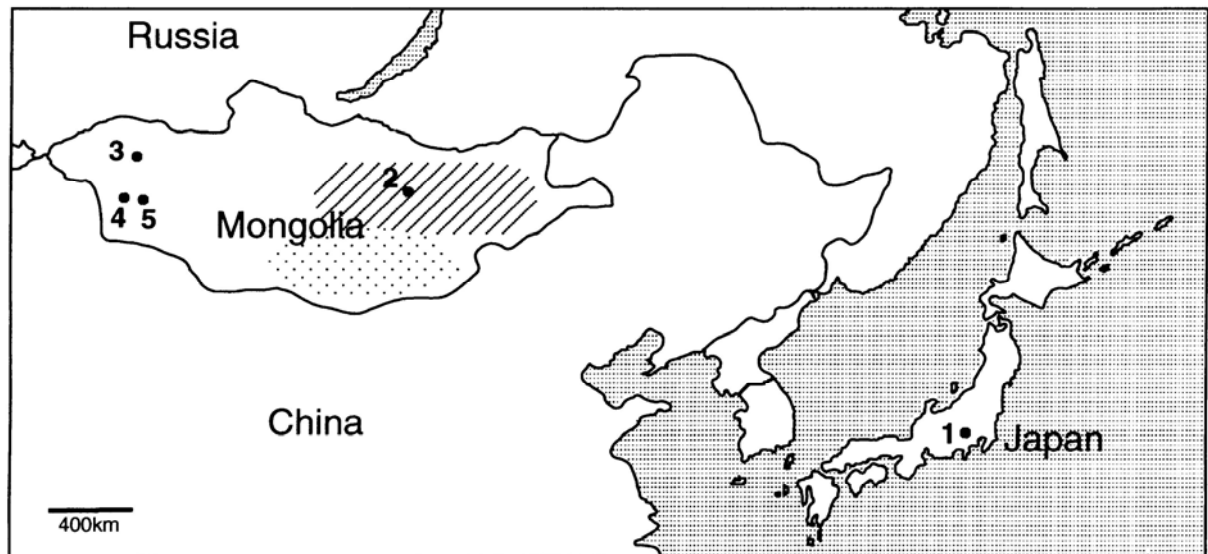
The background linkage disequilibrium (LD) in genetic isolates is of great interest in human genetics. Although many empirical studies have evaluated the background LD in European isolates, such as the Finnish and Sardinians, few data from other regions, such as Asia, have been reported. To evaluate the extent of background LD in East Asian genetic isolates, we analyzed the X chromosome in the Japanese population and in four Mongolian populations (Khalkh, Khoton, Uriankhai, and Zakhchin), the demographic histories of which are quite different from one another. Fisher's exact test revealed that the Japanese and Khalkh, which are the expanded populations, had the same or a relatively higher level of LD than did the Finnish, European American, and Sardinian populations. In contrast, the Khoton, Uriankhai, and Zakhchin populations, which have kept their population size constant, had a higher background LD. These results were consistent with previous genetic anthropological studies in European isolates and indicate that the Japanese and Khalkh populations could be utilized in the fine mapping of both complex and monogenic diseases, whereas the Khoton, Uriankhai, and Zakhchin populations could play an important role in the initial mapping of complex disease genes.

There are many debates among geneticists, concerning the utility of genetically isolated populations for the mapping of disease genes (Wright et al. 1999; Shifman and Darvasi 2001). One type of genetic isolate - for example, the populations of Finland, Iceland, and Sardinia, which have undergone distinct population isolation and subsequent expansion - has modestly higher levels of background linkage disequilibrium (LD) on many chromosomal regions than do outbred populations (Dunning et al. 2000; Eaves et al. 2000; Taillon-Miller et al.

2000; Zavattari et al. 2000; Angius et al. 2001). In contrast, another type of isolate - for example, the Scandinavian Saami isolate, which is from a small population that remained constant in size after isolation - shows very strong and long-range background LD (Laan and Pддbo 1997; Lonjou et al. 1999). These types of genetic isolates have led to the identification of disease genes (Hдstbacka et al. 1992; de la Chapelle and Wright 1998), and, therefore, the value of genetic isolates has increasingly been appreciated during recent attempts at whole-genome LD mapping and association mapping (Wright et al. 1999). However, most studies have mainly focused on particular populations from European regions, such as those from Finland and Sardinia. On the other hand, few data from Asian regions have been reported. It is necessary to investigate the properties of genetic isolates in East Asia and to compare these properties to those of genetic isolates in Europe.

We focused on five populations from two regions of East Asia: Japan and Mongolia (fig. 1). All of the census data are available at the authors' Web site. The demographic history of the Japanese population has been described elsewhere (Koyama 1978; Hayami 2001; Statistics Bureau/Statistical Research and Training Institute 2001). The population size in Japan ~2,000 years ago was estimated to be ~600,000 (Koyama 1978). After several gradual expansions, the modern Japanese population consists of

**Figure 1. Geographic map around East Asia**



Sampling was performed on the following populations: Japanese in Kanagawa and Tokyo (1); Khalkh in Ulaanbaatar (2); Khoton in Tarialan, Uvs aimag (3); Uriankhai in Munkhkhairkhan and Duut, Khovd aimag (4); and Zakhchin in Mankhan and Zereg, Khovd aimag (5). Tarialan, Munkhkhairkhan, and Mankhan are bordered by the steep mountain district formed by part of the Altai Mountains. The hatched and dotted areas indicate the regions of the vast steppe and the Gobi desert, respectively. The Japanese population has the historical and demographic properties of an island population, which would make them a genetic isolate, like the Finnish (Graham and Thompson 1998). The Mongolian populations experienced a historical event in the 12th century. Their country, the "Mongol Empire," expanded from the eastern steppe to the Near East, and then reduced to its modern size. In this century, although there were immigrations from

Turkish and central Asian regions, these immigrants have settled in the mountainous district in western Mongolia (Nyambuu 1992; Badamkhatan 1996).

~120 million people (Statistics Bureau/Statistical Research and Training Institute 2001). Throughout the history of Japan, there seems to be no obvious recent admixture. The Mongolian population consists of ~2.4 million people, which are divided into ~20 ethnic groups (Tumen 1992). Among them, the Khalkh is the largest population, with ~1.8 million people. The Khalkh group is considered to be the direct descendant of the core Mongol tribes, which have inhabited the present-day climatically mild geographical territory of eastern Mongolia (Badamkhatan 1987; Nyambuu 1992). In contrast, the Khoton, Uriankhai, and Zakhchin are regarded as the young, isolated subpopulations of western Mongolia, where the Altai Mountains are located. The Khoton are an extremely small population, with ~6,000 people living in northwestern Mongolia (Figure 1). They are a Turkish descendant population that has not experienced significant admixture and that migrated into Mongolia in the 12th century (Nyambuu 1992). The Uriankhai and Zakhchin populations consist of ~23,000 and ~25,000 people, respectively. They are independent descendants of mixed populations formed by western tribes of Mongolian and Turkish origin (Nyambuu 1992; Badamkhatan 1996). These subpopulations, especially the Khoton, have been separated by severe geographic factors and have become isolates.

**Table 1. Sequence Variation of HVR1 in mtDNA**

Population	Sample size	No. of Haplotypes	Nucleotide Diversity	Tajima's D	P Value <sup>1</sup>
Japanese	182	123	0.018±0.001	-1.925	<0.05
Khalkh	146	106	0.018±0.001	-1.979	<0.05
Khoton	74	23	0.012±0.001	-1.491	>0.10
Uriankhai	58	38	0.017±0.001	-1.619	>0.05
Zakhchin	59	43	0.020±0.001	-1.608	>0.05
European American	50	44	0.014±0.001	-2.241	<0.05
Finnish <sup>2</sup>	50	35	0.011±0.001	-1.820	<0.05
Saami <sup>2</sup>	25	11	0.009±0.001	-0.986	>0.10
Western Pygmy <sup>3</sup>	17	8	0.023±0.003	1.219	>0.10

<sup>1</sup> – Signification values (P<0.05) are given in boldface italics.

<sup>2</sup> – Date are from Sajantila et al. (1995)

<sup>3</sup> – Date are Vigilant et al. (1991)

We collected samples from 182 healthy unrelated Japanese individuals from the Kanto region of the central part of the mainland, which includes Kanagawa and Tokyo (fig. 1). From Mongolia, we collected 146 Khalkh from eastern Mongolia, as well as 74 Khoton, 58 Uriankhai, and 59 Zakhchin from western Mongolia (Figure 1); all individuals were unrelated or at least second cousins. After the mtDNA analysis, we chose 100 Japanese, 83 Khalkh, 40 Khoton, 55 Uriankhai, and 59 Zakhchin male samples for the X-chromosome analysis, including the 49 European Americans (Human variation panel, Caucasian, HD100CAU) as a control population.

**Table 2. Gene Diversity and Haplotype Mismatch Based on Seven Microsatellite Markers**

Population	Sample size	No. of Haplotypes	No. of Alleles (Diversity), Based on Marker							Mean±SD Gene Diversity	Mean Haplotype Mismatch	Variance of Haplotype Mismatch
			DXS983	DXS986	DXS809 2	DXS808 2	DXS122 5	DXS803 7	DXS99 5			
Japanese	100	94	5 (0.45)	14 (0.90)	16 (0.91)	8 (0.70)	7 (0.69)	3 (0.06)	3 (0.51)	0.60±0.33	4.21	1.46
Khalkh	83	80	5 (0.43)	12 (0.87)	13 (0.90)	9 (0.80)	9 (0.74)	6 (0.41)	4 (0.53)	0.67±0.37	4.67	1.78
Khoton	40	32	6 (0.66)	11 (0.89)	11 (0.88)	7 (0.84)	9 (0.86)	3 (0.53)	4 (0.55)	0.74±0.41	5.21	2.21
Uriankhai	55	50	3 (0.35)	14 (0.91)	13 (0.90)	10 (0.84)	9 (0.82)	6 (0.61)	6 (0.64)	0.72±0.40	5.07	1.51
Zakhchin	59	54	6 (0.49)	14 (0.91)	14 (0.91)	8 (0.81)	10 (0.81)	5 (0.46)	5 (0.56)	0.71±0.39	4.95	1.43
European American	49	49	8 (0.76)	7 (0.69)	10 (0.81)	8 (0.78)	10 (0.77)	6 (0.68)	6 (0.59)	0.73±0.40	5.09	1.54
Finnish	80	75	6 (0.70)	11 (0.80)	12 (0.85)	9 (0.75)	9 (0.73)	8 (0.71)	6 (0.63)	0.74	5.15	1.38
Saami <sup>1</sup>	54	32	4 (0.58)	10 (0.80)	8 (0.83)	6 (0.73)	6 (0.76)	5 (0.44)	3 (0.44)	0.67	4.59	3.08

on X<sub>q</sub>13

<sup>1</sup>- Data are from Laan and Paabo (1997). SD's were unavailable

To verify the demographic history of our five tested populations, we first tested Tajima's *D* statistic, using the nucleotide sequence variations of hypervariable region 1 (HVR1) in mtDNA; this statistic can identify a past expansion in population size (Tajima 1989a, 1989b). As is illustrated in table 1, this test showed that the values were significantly negative in the Japanese and Khalkh ( $P < 0.05$ ), as well as in the Finnish and European Americans, indicating that these two East Asian populations had expanded in size in the past. In contrast, the *D* statistics did not significantly deviate from zero in the Khoton ( $P > 0.10$ ), Uriankhai, and Zakhchin ( $P > 0.05$ ) or in the Saami, indicating that these three young subpopulations have kept their sizes small and constant over time.

**Table 3. Pairwise LD Based on Fisher's Exact Test Using Seven Microsatellite Markers on X<sub>q</sub>13**

Locus Pair	Distance (cM)	Uncorrected P Value In									
		Japanese (n=100)	Japanese (n=80)	Khalkh (n=83)	Khoton (n=40)	Uriankhai (n=35)	Zakhchin (n=59)	European American (n=49)	Finnish (n=80)	Saami (n=54)	Sardinian (n=73)
DXS995-DXS983	4.0	0.480	0.291	1.000	0.003	0.097	0.927	0.030	0.508	0.012	0.394
DXS995-DXS986	2.5	0.068	0.051	0.326	0.011	0.482	0.402	0.328	0.729	0.000	0.482
DXS995-DXS8092	2.4	0.120	0.055	0.076	0.047	0.185	0.625	0.755	0.115	0.104	0.829
DXS995-DXS8082	2.3	0.034	0.033	0.095	0.013	0.579	0.021	0.662	0.128	0.000	0.430
DXS1225-DXS983	2.0	0.227	0.151	0.165	0.000	0.773	0.449	0.243	0.630	0.000	0.169
DXS995-DXS8037	2.0	0.517	0.310	0.072	0.070	0.563	0.907	0.182	0.874	0.124	0.650
DXS8037-DXS983	2.0	0.515	0.342	0.449	0.000	0.019	0.521	0.604	0.683	0.300	0.036
DXS995-DXS1225	2.0	0.910	0.388	0.062	0.002	0.654	0.469	0.282	0.154	0.001	0.355
DXS8082-DXS983	1.7	0.064	0.096	0.556	0.002	0.922	0.970	0.448	0.565	0.000	0.142
DXS8092-DXS983	1.6	0.383	0.065	0.318	0.051	0.154	0.796	0.458	0.314	0.000	0.876
DXS986-DXS983	1.5	0.276	0.071	0.407	0.011	0.835	0.210	0.530	0.829	0.000	0.825
DXS8037-DXS986	0.5	0.139	0.179	0.150	0.035	0.046	0.190	0.929	0.620	0.000	0.302
DXS1225-	0.5	0.003	0.002	0.000	0.000	0.019	0.098	0.233	0.393	0.000	0.166

DXS986											
DXS1225- DXS8092	0.4	0.002	0.004	0.105	0.000	0.047	0.211	0.583	0.283	0.000	0.921
DXS8037- DXS8092	0.4	0.426	0.294	0.000	0.086	0.025	0.000	0.723	0.180	0.000	0.285
DXS8037- DXS8082	0.3	0.373	0.353	0.008	0.000	0.000	0.001	0.219	0.238	0.012	0.630
DXS1225- DXS8082	0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DXS8082- DXS986	0.2	0.000	0.000	0.001	0.000	0.001	0.004	0.524	0.92	0.000	0.322
DXS8082- DXS986	0.1	0.612	0.003	0.161	0.000	0.021	0.004	0.502	0.331	0.000	0.125
DXS8082- DXS8092	0.1	0.051	0.014	0.910	0.000	0.017	0.300	0.115	0.044	0.000	0.319
DXS8037- DXS1225	0.0	0.372	0.301	0.000	0.000	0.000	0.000	0.198	0.836	0.091	0.710

To compare these results with those from the European isolates studied elsewhere (Laan and PaEaEbo 1997; Zavattari et al. 2000), we analyzed the extent of background LD among the East Asian males through use of seven microsatellite markers on Xq13 (DXS983, DXS8092, DXS8037, DXS8082, DXS1225, DXS986, and DXS995). Detailed data for number of alleles in the populations are listed in table 2, and the distribution of haplotypes among populations is detailed in an onlineonly table. Fisher's exact test was used to detect significant LD between every pair of the microsatellite markers. All of the uncorrected *P* values are listed in table 3. The results showed that the Japanese population exhibited a relatively higher level of background LD than did the Finnish, European American, and Sardinian populations. Before correction, the Japanese had 5 of 21 pairs with significant LD ( $P < .05$ ), and the other 3 pairs had suggestive LD ( $.05 \leq P < .10$ ). The rapid expansion of the Japanese population to a size of 120 million occurred at a rate of 108%, which is slightly higher than that of the Finnish population (106%). Such a rapid expansion might reduce the LD level in the Japanese, compared with the Finnish. This inconsistency could be explained by population structure differences (Kittles et al. 1998; Jorde et al. 2001).

Our Japanese samples were collected in the Kanto region, which is inhabited by the Japanese homogeneous majority. Our Japanese samples may show stronger LD, reflecting that genetic homogeneity. It should be noted that the Japanese sample size (100) was somewhat larger than that of the Finnish (80) and Sardinians (73). To avoid an overestimation due to the effect of sample size on LD, the equilibration of sample size by the randomization test described by Varilo et al. (2000) was applied to the Fisher's exact test. The sample size equilibration in the Japanese showed that similar LD results were obtained between the original and adjusted Japanese population sizes in table 3, indicating that the larger sample size is not responsible for the strong LD in the Japanese.

The Khalkh showed a relatively higher level of LD than the Japanese. The largest genetic distance for the significant LD was 2.3 cM, and that for the suggestive LD was 2.5 cM. Three pairs still showed the significant LD even after correction, with a largest distance of 0.4 cM (DXS1225-DXS8092). Of the three Mongolian

subpopulations, the Khoton had the largest number of pairs, almost all of which showed significant LD even after the correction, with an extent similar in magnitude to that in the Saami population. Before the correction, 18 of 21 pairs showed significant LD, with the remaining pairs showing suggestive LD. After the correction, 13 pairs remained significant for the largest distance reached, 4.0 cM (DXS995-DXS983). In contrast, the two remaining subpopulations, the Uriankhai and Zakhchin, had four pairs showing significant LD after correction, with largest distances of 0.3 cM (DXS8037-DXS8082) and 0.4 cM (DXS8037-DXS8092), respectively.

The strongest statistical significance and largest distance for LD in this study was observed in the Khoton. The background LD in the Khoton could be explained by strong isolation and constant population size. On the other hand, it should be noted that the microsatellite allelic diversity was generally greater in the Khoton than in the other Asian populations and that the average haplotype mismatch distance in the Khoton was also high, with a large variance (table 2). Moreover, the *F<sub>ST</sub>* values obtained using SNPs on mtDNA and microsatellites on the X chromosome showed the close relationship between the Khoton and either the Europeans and Asians (see Appendix 4 at the authors' Web site). Therefore, the long-range LD in the Khoton may be the result of admixture between the Asian and European populations. However, many data from historical, ethnological, and anthropological studies previously performed in Mongolia suggest that the Khoton population was of Turkish origin and that there was no obvious admixture between the Khoton and other Mongolian populations (Batsuuri 1977; Nyambuu 1992). It is worth mentioning that, in contrast to the high allelic diversity of microsatellites, the mtDNA nucleotide diversity in the Khoton was the smallest of the values among the Asian populations (table 1). Taken together with the similar propensity for these diversities in the Saami, the inconsistency could be explained by the difference between the marker types. In addition, the average pairwise mismatch of haplotypes in the Khoton was not a bimodal distribution, which is the distribution that would be predicted for an admixed population (see online-only figure). The Khoton population was known to have the closest genetic relationship with the Turkish Kyrgyz (Batsuuri 1978). The bilaterally close relationship in the Khoton that is implied by the *F<sub>ST</sub>* value may reflect the original character of an ancestral population, such as the Kyrgyz, rather than admixture in the Khoton. Moreover, although the Uriankhai and Zakhchin are obviously admixed populations with constant sizes, the extent of their LD was smaller than that of the Khoton. These facts suggest that the strong LD with long distance in the Khoton was contributed mainly by the isolation and constant size rather than by the admixture, even if there was some minor admixture.

Our analysis is, to our knowledge, the first report on genetic isolates in East Asia. Although our results were restricted to the X chromosome and mtDNA, they were consistent with demographic, historical, ethnological, and anthropological studies of these populations. Studies of the Khoton, Uriankhai, and Zakhchin populations could play an important role in initial gene mapping of complex diseases, and studies of the Japanese and Khalkh populations could be applied to the fine mapping of either complex or monogenic diseases (Yu et al. 1996; Graham and Thompson 1998).

## References

1. Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB, Fraumene C, Bebbere D, Forabosco P, Pirastu M (2001) *Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits*. Hum Genet 109:198–209
2. Badamkhatan S (ed) (1987) *Ethnography of Mongolia*. Vol 1. Mongolian Academy of Sciences, Ulaanbaatar, Mongolia
3. ——— (1996) *Ethnography of Mongolia*. Vols 2 and 3. Mongolian Academy of Sciences, Ulaanbaatar, Mongolia
4. Batsuuri J (1977) *Anthropological characteristics of the ethnic group "Khoton."* Proceedings of the Institute of General and Experimental Biology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia 12:111–127
5. ——— (1978) *Genetic structure of the ethnic group "Khoton."* Proceedings of the Institute of General and Experimental Biology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia 13:115–129
6. de la Chapelle A, Wright FA (1998) *Linkage disequilibrium mapping in isolated populations: the example of Finland revisited*. Proc Natl Acad Sci USA 95:12416–12423
7. Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, Xu CF, Dawson E, Rhodes S, Ueda S, Lai E, Luben RN, Van Rensburg EJ, Mannermaa A, Kataja V, Rennart G, Dunham I, Purvis I, Easton D, Ponder BA (2000) *The extent of linkage disequilibrium in four populations with distinct demographic histories*. Am J Hum Genet 67:1544–1554
8. Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) *The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes*. Nat Genet 25:320–323
9. Graham J, Thompson EA (1998) *Disequilibrium likelihoods for fine-scale mapping of a rare allele*. Am J Hum Genet 63: 1517–1530
10. Hdstbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) *Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland*. Nat Genet 2:204–211
11. Hayami A (2001) *The historical demography of pre-modern Japan*. University of Tokyo Press, Tokyo
12. Jorde LB, Watkins WS, Bamshad MJ (2001) *Population genomics: a bridge from evolutionary history to genetic medicine*. Hum Mol Genet 10:2199–2207
13. Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long J (1998) *Dual origins of Finns revealed by Y chromosome haplotype variation*. Am J Hum Genet 62:1171–1179
14. Koyama S (1978) *Jomon subsistence and population*. Senri Ethnol Stud 2:1–65
15. Laan M, Pддбо S (1997) *Demographic history and linkage disequilibrium in human populations*. Nat Genet 17:435–438
16. Lonjou C, Collins A, Morton NE (1999) *Allelic association between marker loci*. Proc Natl Acad Sci USA 96:1621– 1626

17. Ludbrook J (1998) *Multiple comparison procedures updated*. Clin Exp Pharmacol Physiol 25:1032–1037
18. Nyambuu K (1992) *Introduction to the ethnography of Mongolia*. Mongolian State Press, Ulaanbaatar, Mongolia
19. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Godde-Dahl T, Issel-Tarver L, Di Rienzo A, PaÈaÈbo S. (1995) *Genes and languages in Europe: an analysis of mitochondrial lineages*. Genome Research 5:42–52
20. Shifman S, Darvasi A (2001) *The value of isolated populations*. Nat Genet 28:309–310
21. Statistics Bureau/Statistical Research and Training Institute (ed) (2001) *Japan statistical yearbook 2002*. Japan Statistical Association, Tokyo
22. Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) *Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28*. Nat Genet 25: 324–328
23. Tajima F (1989a) *The effect of change in population size on DNA polymorphism*. Genetics 123:597–601
24. ——— (1989b) *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism*. Genetics 123: 585–595
25. Tumen D (1992) *Anthropology of the contemporary population of Mongolian People's Republic*. PhD thesis, MV Lomonosov Moscow State University, Moscow
26. Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L (2000) *Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo*. Eur J Hum Genet 8:604–612
27. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) *African populations and the evolution of human mitochondrial DNA*. Science 253:1503–1507
28. Wright AF, Carothers AD, Pirastu M (1999) *Population choice in mapping genes for complex diseases*. Nat Genet 23: 397–404
29. Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD (1996) *Positional cloning of the Werner's syndrome gene*. Science 272:258–262
30. Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Toddo JA, Cucca F (2000) *Major factors influencing linkage disequilibrium in distinct populations: demography, chromosome recombination frequency and selection*. Hum Mol Genet 9:2947–2957