Molecular Biology, Vol. 37, No. 2, 2003, pp. 205–209. Translated from Molekulyarnaya Biologiya, Vol. 37, No. 2, 2003, pp. 234–239.
Original Russian Text Copyright © 2003 by Khitrinskaya, Stepanov, V. Puzyrev, Spiridonov, K. Puzyrev, Maksimova, Nogovitsyna.

UDC 575.174:599.9

Genetic Peculiarity of the Yakut Population as Inferred from Autosomal Loci

I. Yu. Khitrinskaya[1], V. A. Stepanov[1], V. P. Puzyrev[1,2], M. G. Spiridonova[1], K. V. Puzyrev[3], N. R. Maksimova[4], and A. N. Nogovitsyna[4]

[1] *Institute of Medical Genetics, Tomsk Research Center, Russian Academy of Medical Sciences, Tomsk, 634050 Russia; E-mail: alisa@img.tsu.ru*

[2] *Siberian Medical University, Tomsk, 634050 Russia*

[3] *Institute of Cardiology, Tomsk Research Center, Russian Academy of Medical Sciences, Tomsk, 634050 Russia*

[4] *Republican Hospital no. 1, National Center of Medicine, Ministry of Health of the Sakha Republic (Yakutia), Yakutsk, 677018 Russia*

Received October 28, 2002

Abstract—The autosomal gene pool of Yakuts was analyzed with a panel of polymorphic *Alu* insertions. The observed allele frequencies were typical for other Asian ethnic groups. Genetic differentiation of three Yakut populations was relatively high, 2%. East Siberian ethnic groups were shown to have a common gene pool and to experience no intense gene flow from other populations. Development of the Yakut gene pool was assumed to involve no substantial genetic effect of neighboring populations. The results fit both autochthonous and southern origin hypotheses.

Key words: *Alu* repeat, genetic polymorphism, ethnogenetics

INTRODUCTION

Ethnogenetic studies have been carried out in numerous labs throughout the world, and involve genetic reconstruction of the origin, evolution, and dissemination of modern humans and the evolutionary and demographic history of particular regional populations or race and ethnic groups. Such studies are of special importance now in view of dramatic reduction, mixing, or even complete elimination of numerous indigenous populations. Since insight into our genetic past is possible with currently available techniques, it is necessary to estimate the ethnogenetic diversity for populations of all countries and continents before the relevant information is lost irreversibly [1].

Molecular methods of evolutionary genetics take advantage of diallelic or polyallelic DNA markers and markers of maternal (mtDNA) or paternal (Y-chromosome haplotypes) lineages. A great body of data on protein systems has been accumulated [1, 2], which makes it possible to analyze the marker distribution throughout the global population, to study its regularities and mechanisms, and to infer the character of genetic demographic processes at various hierarchic levels. By the late 1990s, a vast material has been also accumulated for genetic polymorphism of mtDNA [3, 4] and various autosomal DNA markers [5, 6], allowing characterization of several genetic diversity aspects of the global population.

The populations of Siberia and the Russian Far East, which differ in origin and ethnic composition, have been poorly studied with respect to genetic polymorphism. Notwithstanding the vast ethnographic, linguistic, and anthropological information concerning the ethnogenesis of these populations, much is still unclear in the genetic relationships of Siberian and Far Eastern ethnic groups and in the contribution of various ancestral populations to their modern gene pool.

Yakuts, who call themselves Sakha, are the indigenous population (380,200 people) of Yakutia. The major groups are Amga-Lena (living between rivers Lena, lower Aldan, and Amga and on the left bank of Lena), Vilyui (living in the basin of river Vilyui), Olekma (living in the basin of river Olekma), and Northern (living in tundra in the basins of rivers Anabar, Olenek, Kolyma, Yana, and Indigirka) Yakuts. Yakuts represent the Central Asian variant of the North Asian race of Mongoloids, and, linguistically, belong to the Northeastern subgroup of the Turkic group of the Altaian family, their language including numerous Mongolian elements and Even and Russian borrowed words [7].

The problem of Yakut ethnogenesis is rather complex. Yakuts live farther to the north than any other Turkic ethnic group. None of the neighboring ethnic groups speaks similar language. There are three major hypothesis of the origin of Yakuts. One implies that Yakut ancestors migrated to the currently occupied

0026-8933/03/3702-0205$25.00 © 2003 MAIK “Nauka /Interperiodica”

area from the south, which seems to be the only idea shared by all its proponents. As the original area, they consider the Baikal region, Central Asia, Tuva, the basin of the river Amur, or even the Minusinsk Lowlands. According to the second hypothesis, Yakuts are autochthonous, originating as an ethnic group in their present area. The third one, a hypothesis of two ancestors, states that Yakuts result from the mixing of southern nomads and local ethnic groups [8].

In this work, we analyzed the genetic diversity and genetic differentiation of three Yakut populations and compared these with other Siberian ethnic groups [11–15], using eight autosomal polymorphisms. Of these, seven (loci *ACE*, *PLAT*, *APOA1*, *PV92*, *F13B*, *A25*, *D1*) result from insertion of the full-length *Alu* repeat, and one (*CD4*) results from *Alu* deletion, with only a 29-nt *Alu* fragment remaining in the locus. In this case, *Alu* was originally present in the locus, because primates (chimpanzee, gibbon, gorilla) are monomorphic and carry allele *Alu*(+) [9, 10].

Polymorphic *Alu* repeats are convenient genetic markers owing to their highly stable location, low rate of *de novo* insertion, and the absence of a precious excision mechanism. Hence *Alu* insertion into a particular locus may be considered as an independent event taking place only once. Moreover, the mechanism of *Alu* transposition allows certain identification of the original (the absence of *Alu*) and changed (*Alu* insert) alleles of a locus. In other words, the ancestral allele and the direction of mutation are always known for *Alu* polymorphism, in contrast to other diallelic polymorphic systems. Finally, the genotyping of *Alu* polymorphisms is technically simple [16, 17].

EXPERIMENTAL

We examined three population samples of unrelated Yakuts from villages Cheriktei (150 km to the northeast of Yakutsk, $N = 81$), Dyupsya ($N = 64$), and Byadi ($N = 56$).

Genomic DNA was isolated from peripheral blood lymphocytes by the standard techniques. Genotyping involved PCR with subsequent electrophoresis in 2% agarose gel. The primers and reaction conditions were as in [5, 10, 18].

Amplified fragments were visualized and gels documented with an Advanced American Biotechnology system. The results were analyzed with the Video Studio 1.0 (Ulead Systems) and Video Packer Plus 1.2p (Aura Vision & VIC Hi Tech) programs. Alleles were designated as *Alu*+ (*Alu* insertion) and *Alu*– (no insertion).

Allele frequencies, frequency errors, correspondence to the Hardy–Weinberg equilibrium, observed and expected heterozygosities, and heterozygosity errors were assessed by the standard methods [20]. Genetic distances and genetic differentiation coefficientG_{ST} were computed according to Nei [21]. A population phylogenetic tree was constructed by the neighbor-joining method, using the PHYLIP package [23] and 100 bootstrap iterations. The intensity of gene flow into the populations under study was estimated according to a published approach [24].

RESULTS AND DISCUSSION

Allele Frequencies and Genetic Diversity of the Yakut Population

The frequencies of *CD4* allele *Alu*– and alleles *Alu*+ of the other loci and the observed (H_o) and expected (H_e) heterozygosities of the three populations are shown in Table 1. All eight loci proved to be polymorphic. In the total sample, genetic diversity was high ($H_e = 0.4$–0.5) in the case of *ACE*, *PLAT*, *PV92*, and *D1*; somewhat lower ($H_e = 0.3$) in the case of *F13B*; and relatively low ($H_e = 0.05$–0.19) in the case of *APOA1*, *A25*, and *CD4*. Averaged over the eight loci, genetic diversity of the total sample was 0.31. In general, the allele frequency distributions observed in Yakuts were much the same as in other Asian ethnic groups [5, 18]. Allele frequencies were similar in the three Yakut populations. Pairwise comparisons revealed significant differences in allele frequencies of *ACE* ($\chi^2 = 7.52$, df = 1, $P < 0.01$) and *D1* ($\chi^2 = 21.5$, df = 1, $P < 0.001$) for the Byadi and Cheriktei populations, of *ACE* ($\chi^2 = 13.2$, df = 1, $P < 0.001$) for the Dyupsya and Cheriktei populations, and for *APOA1* ($\chi^2 = 10.2$, df = 1, $P < 0.01$) and *D1* ($\chi^2 = 19.9$, df = 1, $P < 0.001$) for the Dyupsya and Byadi populations. A significant departure from the Hardy–Weinberg equilibrium was observed in five cases: for *D1* in all three populations, for *APOA1* in the Byadi population, and for *CD4* in the Dyupsya population (Table 1). In each case, the cause was a lack of heterozygotes, suggesting an appreciable inbreeding level.

Genetic Differentiation of the Yakut Populations

We estimated genetic differentiation G_{ST}, which characterizes the contribution of interpopulation variation to the total genetic diversity of a group of populations. Total genetic diversity H_T, diversity resulting from an individual variation within (H_S) and among (D_{ST}) populations, and genetic differentiation G_{ST} were computed for each locus and for the total locus set (Table 2). Total G_{ST} was estimated at 2.0%; i.e., the interpopulation diversity substantially contributed to the total genetic diversity of the Yakut population. In Siberia, G_{ST} is maximal (2.1%) in Altaians; about 1% in Tuvinians, Buryats, and Evens; and lower in other ethnic groups examined [12]. For comparison, G_{ST} estimated with Y-chromosome haplotypes is 15.9% in Yakuts [14]. The difference between two G_{ST} estimates seems natural in view of partrilocality of this

Table 1. Genotype frequency distribution and genetic diversity parameters of the three Yakut populations

Population	*N*	Genotype			Frequency of *Alu*+ (*Alu*– for *CD4*)	χ^{2}*	H_o	H_e
		+/+	+/–	–/–				
ACE								
Dyupsya	128	20	34	10	0.5781 ± 0.0437	0.508	0.5313	0.4878
Byadi	112	22	26	8	0.6250 ± 0.1457	0.005	0.4643	0.4688
Cheriktei	162	43	31	7	0.7772 ± 0.0352	0.172	0.3827	0.4012
PLAT								
Dyupsya	128	18	31	15	0.5235 ± 0.0441	0.054	0.4844	0.4989
Byadi	112	19	26	11	0.5714 ± 0.0468	0.152	0.4643	0.4898
Cheriktei	160	23	43	14	0.5562 ± 0.0393	0.6305	0.5375	0.4937
PV92								
Dyupsya	126	32	24	7	0.6984 ± 0.0409	0.577	0.3810	0.4213
Byadi	108	33	19	2	0.7870 ± 0.0394	0.133	0.3519	0.3352
Cheriktei	158	39	32	8	0.6962 ± 0.0366	0.142	0.4051	0.4230
APOA1								
Dyupsya	124	56	6	0	0.9516 ± 0.0193	0.160	0.0963	0.0921
Byadi	112	41	10	5	0.8214 ± 0.0362	8.575$^{\$}$	0.1786	0.2934
Cheriktei	162	66	13	2	0.8951 ± 0.1049	1.7181	0.1605	0.1879
F13B								
Dyupsya	128	46	15	3	0.8359 ± 0.0327	1.355	0.2344	0.2743
Byadi	112	37	17	2	0.8125 ± 0.0369	0.001	0.3036	0.3047
Cheriktei	162	52	27	2	0.8086 ± 0.0309	0.481	0.3333	0.3095
A25								
Dyupsya	128	0	9	55	0.0703 ± 0.0226	0.366	0.1406	0.1307
Byadi	112	0	3	53	0.0268 ± 0.0153	0.042	0.0536	0.0521
Cheriktei	162	0	7	74	0.0432 ± 0.0160	0.165	0.0864	0.0827
CD4								
Dyupsya	120	57	2	1	0.0333 ± 0.0164	13.98$^{\$}$	0.0333	0.0644
Byadi	112	53	3	0	0.0268 ± 0.0153	0.042	0.0536	0.0521
Cheriktei	162	77	4	0	0.9753 ± 0.0122	0.052	0.0494	0.0482
D1								
Dyupsya	118	26	13	20	0.5508 ± 0.0458	18.15$^{\&}$	0.2203	0.4948
Byadi	112	15	5	35	0.3304 ± 0.0444	35.57$^{\#}$	0.0893	0.4424
Cheriktei	116	31	12	15	0.6379 ± 0.0446	17.68$^{\#}$	0.2069	0.4620

* A departure from the Hardy–Weinberg equilibrium was significant at ($^{\#}$) $P < 0.001$, ($^{\$}$) $P < 0.01$, or ($^{\&}$) $P < 0.05$.

ethnic group and a small effective size of the Y chromosome pool as compared with the autosome pool. These factors enhance the effect of gene drift on the male gene pool and consequently lead to a greater interpopulation variation.

The results obtained for individual loci showed that the contribution of differences in allele frequencies to the interpopulation diversity was greatest in the case of *D1* (G_{ST} = 6.4%); lower in the case of *PV92* (2.0%), *ACE* (1.5%), and *A25* (1.6%); and only 0.1–0.8% in the case of the other four loci (Table 2).

Genetic Relationships among Populations

Cluster analysis was used to elucidate the genetic relationships among the three Yakut populations and among these and other Siberian populations. A phylogenetic tree was based on the above results and our previous data. As an outgroup, we used a hypothetical ancestral population, taking the frequency of the original allele as zero for each of the eight loci. As already

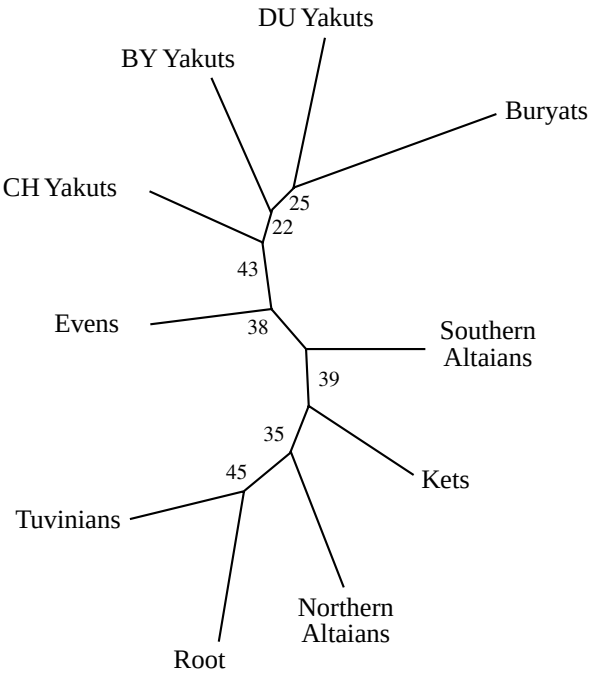

Fig. 1. Dendrogram of genetic relationships between several Siberian ethnic groups. Numbers of bootstrap iterations supporting particular branches are indicated. Here and in Fig. 2: Yakut populations are designated as CH, Cheriktei; BY, Byadi; and DU, Dyupsya.

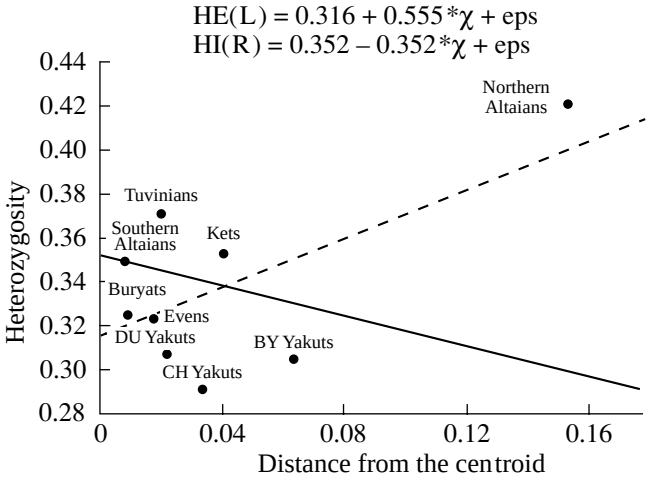

Fig. 2. Distance to the allele frequency centroid vs. heterozygosity regression plot based on the allele frequencies of the eight autosomal loci. HI (R), regression of expected heterozygosity predicted according to Harpending and Ward [24] (solid line); HE (L), regression of actual expected heterozygosity (dotted line).

mentioned, it is possible to reconstruct the ancestral state of *Alu* polymorphisms for modern human populations owing to the unidirectional character of mutation: the original allele is always *Alu*– and the more recent one, *Alu*+. In the case of *CD4*, *Alu*+ is original. To obtain more reliable results, we used 100 bootstrap iterations of the initial data. A consensus dendrogram of phylogenetic relationships among Siberian populations is shown in Fig. 1. The bootstrap values and, consequently, the dendrogram were nonsignificant.

Table 2. Genetic differentiation of the three Yakut populations

Locus	H_T	H_S	G_{ST}
ACE	0.4408	0.4343	0.0149
PLAT	0.4949	0.4941	0.0017
PV92	0.4013	0.3932	0.0204
APOA1	0.1917	0.1911	0.0035
F13B	0.2973	0.2962	0.0038
A25	0.0900	0.0885	0.0166
CD4	0.0544	0.0549	0.0085
D1	0.4987	0.4664	0.0647
Total	0.3087	0.3023	0.0205

Yet it is possible to isolate two clusters: the East Siberian one includes Evens, Buryats, and Yakuts, and the Altai-Sayan one, Tuvinians and Altaians. Kets are between two Altaian groups on the dendrogram, which is consistent with a hypothesis of their Altai-Sayan origin. The division of Altaians into the two groups possibly reflects their specific ethnogenesis. The position of Yakuts suggests their close relatedness to Evens and Buryats, suggesting that the Yakut gene pool is mostly of a non-Turkic origin. A Turkic language was probably acquired as a result of social domination of a Turkic-speaking elite. Similar data have been obtained with Y-chromosome haplogroups [14]. Basing on the data on polymorphic protein systems, Dubrova *et al.* [25] have observed that Yakuts are closest to northwestern Buryats. Anthropologically, the Central Asian type prevails and some features of the Baikal type are detectable in Yakuts. Notwithstanding the substantial linguistic differences, East Siberian ethnic groups, which speak languages of the Mongol (Buryats), Tungus-Manchurian (Evens), or Turkic (Yakuts) groups of the Altaian family, show close genetic relationships by autosomal gene pool and Y-chromosome lineages [14].

Analysis of the Gene Flow

Historically, the ethnic groups examined vary in intensity of ethnogenetic events, which first and foremost include migration. Their modern populations also differ in extent of gene exchange with adjacent ethnic groups. To estimate the relative intensity of gene flow in Siberian populations, we used an approach developed by Harpending and Ward [24] on

the basis of Wright's island model [26, 27]. The results are presented in Fig. 2. A theoretic dependence between the distance from the centroid and the heterozygosity is shown with a solid line. Gene flow is lower or higher than expected in populations located below and above the line, respectively. In general, the East Siberian populations are characterized by the absence of intense gene flow from other ethnic groups, and seem to be more genetically isolated as compared with Tuvinians, Altaians, and Kets, who experience a substantial gene flow.

No intense gene flow was observed for the three Yakut populations, like for the other East Siberian ethnic groups. This suggests that the Yakut gene pool was formed on the local basis and was only slightly genetically affected by the adjacent populations. Our results fit both autochthonous and southern origin hypotheses of the Yakut origin, if the interaction of southern migrants with the original population of the modern Yakut area mostly involved displacement of local non-Turkic groups by migrants, rather than assimilation or mixing.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-04-49166, 00-04-48506, 01-04-63076, 00-15-9786, 02-04-06041), the Wener-Gren Foundation (grant no. 6801), and the foundation Databases on Human, Animal, Plant, and Microbial Gene Pools of the Federal program Frontiers in Science and Technology, 2002–2004.

REFERENCES

1. Cavalli-Sforza L.L., Menozzi P., Piazza A. 1994. *The history and geography of human genes*. Princeton: Prinseton Univ. Press.
2. *Genofond i genogeografiya narodonaseleniya* (Human Gene Diversity and Gene Geography). St. Petersburg: Nauka. 2000. vol. 1.
3. Wallace D.C. 1995. Mitochondrial DNA variation in human evolution, degenerative disease and aging. *Am. J. Hum. Genet*. **57**, 201–223.
4. Jorde L.B., Bamshad M., Rogers A.R. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *BioEssays*. **20**, 126–136.
5. Stoneking M., Fontius J.J., Clifford S.L., Soodyall H. *et al*. 1997. *Alu* insertion polymorphism and human evolution: evidence for a larger population size in Africa. *Genome Res*. **7**, 1061–1071.
6. Bowcock A.M., Ruiz-Linares A., Tomfohrde J. *et al*. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. **386**, 455–457.
7. *Narody Rossii: entsiklopediya* (Ethnic Groups of Russia: An Encyclopedia). 1994. Tishkov V.A., Ed. Moscow: Bol'shaya Rossiiskaya Entsiklopediya.
8. Gogolev A.I. 2000. *Istoriya Yakutii (Obzor istoricheskikh sobytii do nachala XX veka)* (History of Yakutia (A review of historical events up to the early 20th century)). Yakutsk: Yakutsk. Univ.
9. Tishkoff S.A., Dietzsch E., Speed W. *et al*. 1996. Global patterns of linkage equilibrium at the CD4 lokus and modern human origins. *Science*. **271**, 1380–1387.
10. Majumder P.P., Roy B., Banrjee S. *et al*. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur. J. Hum. Genet*. **7**, 435–446.
11. Stepanov V.A., Puzyrev V.P., Spiridonova M.G., Khitrinskaya I.Yu. 1999. Analysis of *Alu* polymorphisms in urban and rural Russian populations of Siberia. *Genetika*. **35**, 1138–1143.
12. Stepanov V.A., Khitrinskaya I.Yu., Puzyrev V.P. 2000. Genetic differentiation of the Siberian indigenous population, as revealed by *Alu* polymorphisms. *Genetika cheloveka i patologiya* (Human Genetics and Pathology). Puzyrev V.P., Ed. Tomsk: STT. **5**, 98–107.
13. Stepanov V.A., Khitrinskaya I.Yu., Puzyrev V.P. 2001. Genetic differentiation of the Tuva population with respect to the *Alu* insertions. *Genetika*. **37**, 563–569.
14. Stepanov V.A., Puzyrev V.P., Spiridonova M.G., Kharkov V.N., Soltobaeva J.O. 2002. Y-chromosome diversity in population of Altaic language family. in *Thesis of HUGO Meeting*. Edinburgh.
15. Khitrinskaya I.Yu., Stepanov V.A., Puzyrev V.P. 2001. Analysis of *Alu* polymorphisms in Buryat populations. *Genetika*. **37**, 1553–1558.
16. Novick G.E., Batzer M.A., Deininger P.L., Herrera R.J. 1996. The mobile genetic element *Alu* in the human genome. *Bioscience*. **46**, 32–41.
17. Batzer M.A., Deininger P.L. 2002. *Alu* repeats and human genomic diversity. *Nature Rev. Genetics*. **3**, 370–379.
18. Batzer M.A., Arcot S.S., Phinney J.M. *et al*. 1996. Genetic variation of recent *Alu*-insertions in human populations. *J. Mol. Evol*. **42**, 22–29.
19. Arcot S.S., Fontius J.J., Deininger P.L., Batzer M.A. 1995. Identification and analysis of a "young" polymorphic *Alu* element. *Biochim. Biophys. Acta*. **1263**, 99–102.
20. Zhivotovsky L.A. 1991. *Populyatsionnaya biometriya* (Population Biometrics). Moscow: Nauka.
21. Nei M. 1987. *Molecular evolutionary genetics*. N.Y.: Columbia Univ. Press.
22. Saitou N., Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406–425.
23. Felsenstein J. 1993. PHYLIP, version 3.5. Univ. of Washington, Seattle.
24. Harpending H.C., Ward R.H. 1982. Chemical systematics and human populations. in: *Biochemical Aspects of Evolutionary Biology*. Nitecki M.H. Ed. Chicago: Univ. of Chicago Press.
25. Dubrova E.Yu., Bogatyreva L.V., Pushkina E.I. 1992. Variation of polymorphic gene markers in Buryat and Russian newborns of Ulan-Ude. *Genetika*. **28**, 153–158.
26. Wright S. 1943. Isolation by distance. *Genetics*. **28**, 139–156.
27. Wright S. 1951. The genetical structure of populations. *Ann. Eugenics*. **15**, 323–354.