

GENETIC DIFFERENTIATION OF ESTONIANS

Mart VIIKMAA^a and Leiu HEAPOST^b

^a Tartu Ülikooli üld- ja molekulaarpatoloogia instituut (Institute of General and Molecular Pathology, University of Tartu), Veski 34, EE-2400 Tartu, Eesti (Estonia)

^b Tartu Ülikooli molekulaar- ja rakubioloogia instituut (Institute of Molecular and Cell Biology, University of Tartu), Veski 34, EE-2400 Tartu, Eesti (Estonia)

^b Ajaloo Instituut (Institute of History), Rütüli 6, EE-0001 Tallinn, Eesti (Estonia)

Presented by J. Martin

Received 15 March 1996, revised version received 10 October 1996, accepted 16 October 1996

Abstract. The population genetic structure of Estonians was analysed and the gene frequencies of subgroups of the Estonian population were compared with one another and with some neighbouring and other Finno-Ugric populations on the basis of seven blood group systems and the trait of PTC-tasting. A genetic heterogeneity of the Estonians is shown with the greatest genetic differences in the west–east direction in Estonia. The Estonians of the Central, East, South-West, and South-East regions are genetically the closest. The subgroups of the West Islands, the Western part of the mainland, and North-East are standing separately. On the basis of genetic distances, the whole Estonian population has the closest relation to their nearest neighbouring peoples, Russians and Latvians, but also to the Finno-Ugric peoples of Vepsians and Karelians.

Key words: population genetic features, structure, Estonians.

INTRODUCTION

The Estonian population is quite heterogeneous in many aspects. While the greatest linguistic differences appear between the northern and southern parts of Estonia (Kask, 1956), the greatest anthropological differences appear in the west–east direction. On the basis of voluminous research made by Aul (Ауль, 1964) two main anthropological types are distinguished among the Estonians: the West-Baltic type predominates in West Estonia, while the East-Baltic type occurs mainly in East Estonia, especially in the South-East region, but also in some localities of South-West Estonia.

In this report we are making an attempt to analyse the population genetic structure of the Estonians and to compare the gene frequencies of subgroups of the Estonian population with one another and with some neighbouring peoples.

MATERIAL AND METHODS

Genetic data are based on seven blood group systems (ABO, Duffy, Kell, Lewis, MN, P, and Rhesus) and the trait of phenyl thiocarbamide (PTC) tasting (altogether 24 alleles). The material was collected by L. Heapost from 39 localities in different regions of Estonia (Heapost, 1994). All the individuals examined were indigenous Estonians, whose all grandparents and parents had been born in the same locality. The local samples were joined into seven regional groups more or less according to the main dialectal areas (Murumets, 1982, 1983) (Fig. 1).

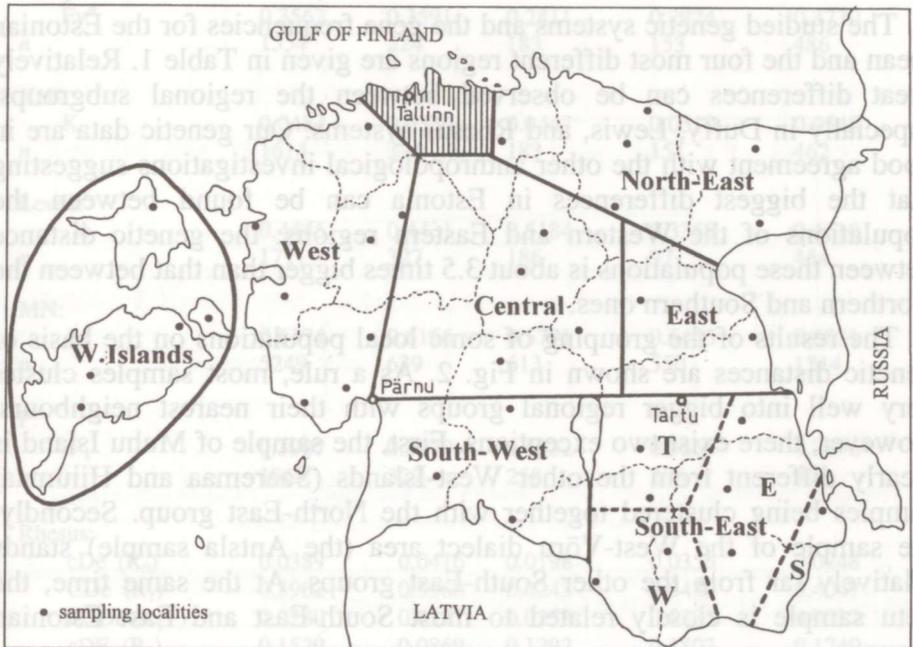


Fig. 1. Regional division of Estonia used in this study. The South-East region is divided into four local dialect areas: T = Tartu, E = East-Võru, W = West-Võru, S = Setu.

The data were analysed with tools of a population genetic program package created by M. Viikmaa in the Institute of General and Molecular Pathology of the University of Tartu. The differences of gene frequencies between the regional populations are valued by the chi-squared method. The degree of genetic diversity of the groups is determined by the method of genetic distances (Cavalli-Sforza & Edwards, 1967). The grouping of populations on the basis of these distances is made using the cluster analysis method of means. The genetic distance between two subclusters (groups of populations) was calculated as the average of distances between every pair of members from different clusters. The results of clustering are visualized in the form of graphic clustrograms.

Data for interpopulational comparisons were taken from literature: for Finns and Finnish Swedes from Nevanlinna (1973); for Karelians from Шнейдер & Тихомирова (1991); for Komis from Эрикссон & Франгс (1982); for Latvians from Kariks et al. (1966), Race et al. (1948), and Heapost (1994); for Lithuanians from Harvey et al. (1983); for Maris from Эрикссон et al. (1979); for Russians from Umnova et al. (1968), and for Vepsians from Heapost (1994).

RESULTS

The studied genetic systems and the gene frequencies for the Estonian mean and the four most different regions are given in Table 1. Relatively great differences can be observed between the regional subgroups, especially in Duffy, Lewis, and Rhesus systems. Our genetic data are in good agreement with the other anthropological investigations suggesting that the biggest differences in Estonia can be found between the populations of the Western and Eastern regions: the genetic distance between these populations is about 3.5 times bigger than that between the Northern and Southern ones.

The results of the grouping of some local populations on the basis of genetic distances are shown in Fig. 2. As a rule, most samples cluster very well into bigger regional groups with their nearest neighbours. However, there exist two exceptions. First, the sample of Muhu Island is clearly different from the other West-Islands (Saaremaa and Hiiumaa) samples being clustered together with the North-East group. Secondly, the sample of the West-Võru dialect area (the Antsla sample) stands relatively far from the other South-East groups. At the same time, the Setu groups is closely related to most South-East and East Estonian groups.

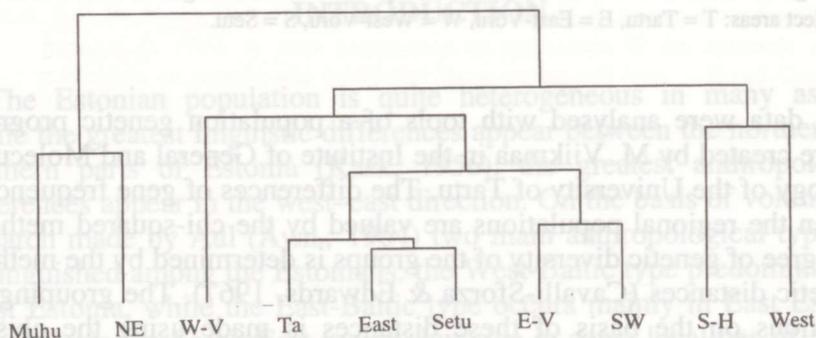


Fig. 2. Clustrogram of grouping some local Estonian samples based on the cluster analysis of Edwards' genetic distance matrix. (NE = North East, W-V = West-Võru dialect area, Ta = Tartu dialect area, E-V = East-Võru dialect area, SW = South West, S-H = Saaremaa and Hiiumaa islands, West = West Estonia.)

Gene frequencies of the polymorphic systems used in all investigated Estonians and in the four most different regions

| System and alleles | Estonia | West Islands | West Estonia | North-East Estonia | South-East Estonia |
|--|---------|--------------|--------------|--------------------|--------------------|
| ABO: | | | | | |
| A ₁ | 0.2009 | 0.2126 | 0.1851 | 0.2008 | 0.2072 |
| A ₂ | 0.0391 | 0.0433 | 0.0317 | 0.0489 | 0.0382 |
| B | 0.1606 | 0.1443 | 0.1737 | 0.1267 | 0.1829 |
| 0 | 0.5993 | 0.5998 | 0.6096 | 0.6235 | 0.5716 |
| n | 2722 | 650 | 456 | 330 | 573 |
| Duffy: | | | | | |
| Fy ^a | 0.3562 | 0.3591 | 0.2811 | 0.2924 | 0.3730 |
| n | 1554 | 224 | 183 | 153 | 486 |
| Kell: | | | | | |
| K | 0.0483 | 0.0330 | 0.0447 | 0.0510 | 0.0548 |
| n | 1614 | 324 | 183 | 151 | 462 |
| Lewis: | | | | | |
| le | 0.4445 | 0.4431 | 0.5184 | 0.3368 | 0.4610 |
| n | 1711 | 327 | 186 | 97 | 454 |
| MN: | | | | | |
| M | 0.6174 | 0.6166 | 0.5791 | 0.6445 | 0.6231 |
| n | 5249 | 639 | 613 | 550 | 1214 |
| P: | | | | | |
| P ₁ | 0.3907 | 0.3970 | 0.4532 | 0.3369 | 0.3896 |
| n | 1969 | 327 | 258 | 206 | 590 |
| Rhesus: | | | | | |
| cDe (R ₀) | 0.0389 | 0.0416 | 0.0198 | 0.0356 | 0.0248 |
| CDe (R ₁) | 0.3902 | 0.3960 | 0.4243 | 0.3479 | 0.4047 |
| C ^W De (R ₁ ^W) | 0.0280 | 0.0227 | 0.0253 | 0.0335 | 0.0332 |
| cDE (R ₂) | 0.1529 | 0.0869 | 0.1292 | 0.1803 | 0.1740 |
| CDE (R _Z) | 0.0208 | 0.0206 | 0.0046 | 0.0470 | 0.0253 |
| cde (r) | 0.3287 | 0.3805 | 0.3350 | 0.3249 | 0.3177 |
| Cde (r') | 0.0271 | 0.0243 | 0.0497 | 0.0309 | 0.0127 |
| cdE (r'') | 0.0134 | 0.0274 | 0.0121 | 0.0000 | 0.0077 |
| n | 2039 | 330 | 257 | 209 | 558 |
| PTC-tasting: | | | | | |
| t | 0.5007 | 0.6003 | 0.4976 | 0.4900 | 0.4780 |
| n | 2796 | 332 | 571 | 366 | 685 |

As the next step, we looked at the clustering of the regional populations with one another. The regions used here are the same as the main regions of dialects. The genetic differences between all the regions are statistically significant, with the exception of the East group, which is very similar to the southern regions (Table 2). As it is demonstrated by the clustrogram of genetic distances (Fig. 3), the Central, South-West, East, and South-East regions are very close. The North-East, West Islands, and West-Estonia are standing separately.

Chi-squared values of pairwise comparison of gene frequency distributions in main regional populations of Estonians. Degree of freedoms = 15; statistically significant differences with $p < 0.05$ are shown by asterisk

| | West | South-West | Central | South-East | East | North-East |
|--------------|-------|------------|---------|------------|-------|------------|
| West Islands | 69.5* | 47.7* | 61.2* | 82.5* | 58.5* | 76.4* |
| West | | 39.1* | 55.5* | 64.0* | 52.5* | 82.9* |
| South-West | | | 26.1* | 42.8* | 23.2 | 49.3* |
| Central | | | | 36.5* | 43.1* | 41.6* |
| South-East | | | | | 19.6 | 58.7* |
| East | | | | | | 52.4* |

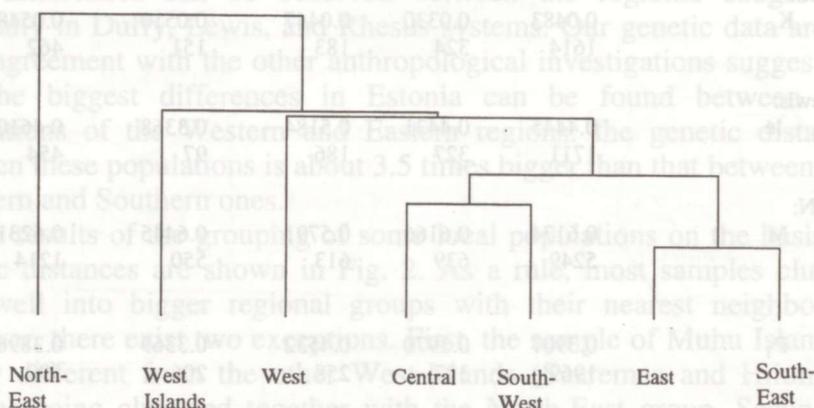


Fig. 3. Clustrogram of grouping of Estonian main regional populations based on cluster analysis using the genetic distance matrix of Cavalli-Sforza & Edwards (1967).

Next, we compared the Estonian regional populations with the neighbouring and some Finno-Ugric peoples (Fig. 4). The populations are grouped in two main clusters. The Estonian samples are joined together in one cluster with the two Finno-Ugric populations of Vepsians and Karelians. The other cluster comprises the Finno-Ugric populations of Komis, Maris (both standing separately), and Finns, as well as Finnish Swedes, Lithuanians, Latvians, and Russians. Thus, the Finno-Ugric linguistic group is clearly not a genetic unit.

The genetic distances of the whole Estonian population to other peoples (Table 3) suggest that Estonians are most closely related to Russians and Latvians, to their nearest neighbours. In addition, Vepsians and Karelians are very close to Estonians, Finns come only after them. On the demonstrated clustrogram (Fig. 4) the Russians and Latvians fall into the second cluster because they have some more closely related populations than Estonians – Lithuanians, Finns, and Finnish Swedes.

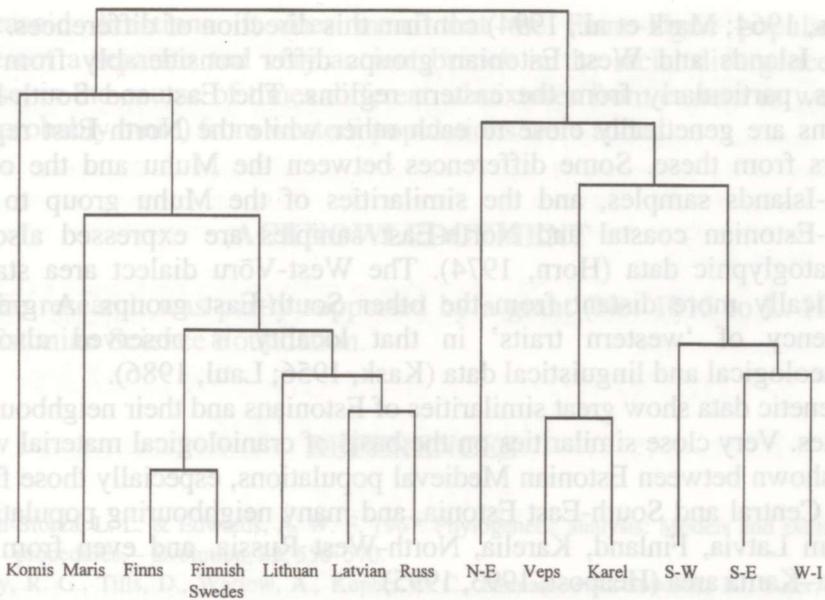


Fig. 4. Clustrogram of grouping of Estonian regional populations with some neighbouring and other Finno-Ugric peoples based on the cluster analysis using the genetic distance matrix of Cavalli-Sforza & Edwards (1967). (Lithuan = Lithuanians, Russ = Russians, Veps = Vepsians, Karel = Karelians, N-E = North-East Estonia, S-W = South-West Estonia, S-E = South-East Estonia, W-I = West Estonia with West Islands.)

Table 3

Genetic distances by Cavalli-Sforza & Edwards (1967) between Estonians and some neighbouring and other Finno-Ugric peoples

| | Vepsians | Karelians | Komis | Maris | Finns | Finn. Swedes | Latvians | Lithuanians | Russians |
|----------------|----------|-----------|--------|--------|--------|--------------|----------|-------------|----------|
| Estonians | 0.0069 | 0.0074 | 0.0167 | 0.0189 | 0.0108 | 0.0115 | 0.0064 | 0.0114 | 0.0065 |
| Vepsians | | 0.0029 | 0.0168 | 0.0128 | 0.0089 | 0.0075 | 0.0085 | 0.0072 | 0.0066 |
| Karelians | | | 0.0120 | 0.0104 | 0.0096 | 0.0095 | 0.0073 | 0.0094 | 0.0059 |
| Komis | | | | 0.0082 | 0.0106 | 0.0103 | 0.0120 | 0.0133 | 0.0081 |
| Maris | | | | | 0.0080 | 0.0100 | 0.0079 | 0.0085 | 0.0058 |
| Finns | | | | | | 0.0017 | 0.0041 | 0.0055 | 0.0043 |
| Finnish Swedes | | | | | | | 0.0063 | 0.0042 | 0.0051 |
| Latvians | | | | | | | | 0.0045 | 0.0032 |
| Lithuanians | | | | | | | | | 0.0033 |

DISCUSSION

According to our genetic data there exist relatively great differences between the local population samples in Estonia. The differences in gene frequencies between West and East Estonia are more impressive than those between North and South Estonia. Other anthropological data

(Ауль, 1964; Mark et al., 1994) confirm this direction of differences. The West Islands and West Estonian groups differ considerably from the others, particularly from the eastern regions. The East and South-East regions are genetically close to each other while the North-East region differs from these. Some differences between the Muhu and the other West-Islands samples, and the similarities of the Muhu group to the West-Estonian coastal and North-East samples are expressed also in dermatoglyphic data (Horn, 1974). The West-Võru dialect area stands genetically more distant from the other South-East groups. A greater frequency of 'western traits' in that locality is observed also in archaeological and linguistical data (Kask, 1956; Laul, 1986).

Genetic data show great similarities of Estonians and their neighbouring peoples. Very close similarities on the basis of craniological material were also shown between Estonian Medieval populations, especially those from East, Central and South-East Estonia, and many neighbouring populations – from Latvia, Finland, Karelia, North-West Russia, and even from the Volga–Kama area (Heapost, 1993, 1995).

Finally, let us comment on the problem of the so-called western and eastern traits in Finno-Ugric populations. In literature, we can find data suggesting the existence of eastern traits (correctly, eastern frequencies of traits) in these populations. That phenomenon is usually explained with Mongoloid admixture. However, there are evidences of the simultaneous presence of western (and sometimes even super-western) frequencies of traits in Finno-Ugric populations. The best example of that kind of data is found in the traits of odontological complexes (Зубов, 1982; Sarap, 1994). Our data of gene frequencies in Estonians also illustrate the tendencies in two opposite directions (Table 4). We are of the opinion that the Mongoloid admixture cannot explain the simultaneous presence of clear western frequencies. We suppose that these antagonistic frequencies of different traits are tracks of the original genetic structure of the Finno-Ugric ancestor population which was not clearly differentiated in Mongoloid–

Table 4

'Western' and 'eastern' gene frequencies in the Estonians (in comparison with other East and North European peoples)

| Frequencies | 'Western' type | | 'Eastern' type | |
|-------------|-----------------|-----------|-------------------|-----------|
| | Allele | Frequency | Allele | Frequency |
| Higher | K | 0.05 | B | 0.16 |
| | Lu ^a | 0.05 | CDE | 0.02 |
| | MS | 0.32 | Hypolactasia | 0.53 |
| | S | 0.34 | | |
| | Hp ¹ | 0.41 | | |
| Lower | CDe | 0.39 | A ₂ | 0.04 |
| | Fy ^a | 0.36 | P ₁ | 0.39 |
| | | | cde | 0.33 |
| | | | t | 0.50 |
| | | | Le (a+) phenotype | 0.13 |

Caucasoid directions. It does mean that the Finno-Ugric populations represent a separate and very ancient branch in the racial divergence. Of course, in the course of times different admixtures from eastern as well as (but probably more) from western populations were added.

ACKNOWLEDGEMENT

The research was partly supported by a grant (No. 1015 to L. H.) of the Estonian Science Foundation.

REFERENCES

- Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967. Phylogenetic analysis: Models and estimation procedures. – *Evolution*, **32**, 550–570.
- Harvey, R. G., Tills, D., Warlow, A., Kopec, A. C., Domaniewska-Sobczak, K., Suter, D. & Lord, J. M. 1983. Genetic affinities of the Balts. A study of blood groups, serum proteins and enzymes of Lithuanians in the United Kingdom. – *Man (N.S.)*, **18**, 535–552.
- Heapost, L. 1993. Makita kalmistu antropoloogiline aines. – In: Muinasaja teadus 2. Vadjäpärased kalmes Eestis 9.–16. sajandil. Eesti TA Ajaloo Instituut, Tallinn, 233–248.
- Heapost, L., 1994. Populatsioonigeneetilised tunnused eestlastel. – In: K. Mark, L. Heapost, G. Sarap. Eestlaste antropoloogia seoses etnogeneesi küsimustega. TA Kirjastus, Tallinn, 110–196.
- Heapost, L. 1995. On craniology of South-East Estonian population in XI–XVIII cc. – In: Papers on Anthropology VI. Tartu Ülikooli Kirjastuse trükikoda, Tartu, 57–69.
- Horn, A. 1974. Sõrmemustrite põhitüüpide esinemissagedusest eestlastel. – *Antropoloogia-alaseid töid*, II. TRÜ Toimetised, **330**, 67–90.
- Kariks, J., Bradley, M. A. & Walsch, R. J. 1966. The blood groups of Latvians resident in Australia. – *Vox Sanguinis*, **11**, 699–704.
- Kask, A. 1956. Eesti murrete kujunemisest ja rühmitumisest. – In: Eesti rahva etnilisest ajaloost. Eesti Riiklik Kirjastus, Tallinn, 24–40.
- Laul, S. 1986. Kas lõuna-estli murdealade kujunemine on arheoloogiliselt jälgitav. – In: Emakeele Seltsi Aastaraamat 30. Eesti Raamat, Tallinn, 28–38.
- Mark, K., Heapost, L. & Sarap, G. 1994. Eestlaste antropoloogia seoses etnogeneesi küsimustega. TA Kirjastus, Tallinn.
- Murumets, S. 1982. Eesti keeleala murdelisest liigendusest “Väikese murdesõnastiku” põhjal. I. – *Keel ja Kirjandus*, **1**, 11–17.
- Murumets, S. 1983. Eesti keeleala murdelisest liigendusest “Väikese murdesõnastiku” põhjal. II. – *Keel ja Kirjandus*, **11**, 615–623.
- Nevanlinna, H. R. 1973. Suomen väestörakenne. Geneettinen ja genealoginen tutkimus. Vammalan Kirjapaino OY, Vammala.
- Race, R. R., Sanger, R., Lawler, S. D. & Keetch, D. V. 1948. Blood groups of Latvians A₁A₂B₀, MN and Rh. – *Ann. Eugenics*, **14**, 134–138.
- Sarap, G. 1994. Eestlaste odontoloogilise kompleksi omapära. – In: K. Mark, L. Heapost, G. Sarap. Eestlaste antropoloogia seoses etnogeneesi küsimustega. TA Kirjastus, Tallinn, 197–240.
- Umnova, M. A., Prokop, O., Piskunov, T. M., Samusova, G. S., Ishalovskaya, T. A. & Prozorovskaya, G. P. 1968. Blood group distribution among the Moscow population. 7th Int. Congr. Anthropol. Ethnol. Sci., Moscow, 496–500.
- Аулю Ю. 1964. Антропология эстонцев. – TRÜ Toimetised, 158.

- Зубов А. А. 1982. Географическая изменчивость одонтологических комплексов финно-угорских народов. – In: Финно-угорский сборник (антропология, археология, этнография). Наука, Москва, 134–148.
- Шнейдер Ю. В. & Тихомирова Е. В. 1991. Генетический полиморфизм в популяции карелов. – Генетика, 27, 1460–1466.
- Эриксон А. В., Золотарева И. М., Козинцев А. Г., Шевченко А. В., Эскола М. Р., Кирьяринта М., Партанен К., Фельман Дж. 1979. Генетические исследования марийцев (черемисов). – In: Новые исследования по антропологии марийцев. Наука, Москва, 7–39.
- Эриксон А. В. & Франтс Р. Р. 1982. Исследования групп крови у коми-зырян в СССР. – In: Финно-угорский сборник (антропология, археология, этнография). Наука, Москва, 191–206.

EESTLASTE GENEETILISEST MITMEKESISUSEST

Mart VIIKMAA, Leiu HEAPOST

On analüüsitud eestlaste populatsioonigeneetilist struktuuri. Eestlaste lokaalseid ja regionaalseid rühmi on võrreldud üksteisega ning mõne naaber-, sealhulgas soome-ugri rahvaga, kasutades selleks seitset polümorfset vererühmasüsteemi ja PTC (fenüülitiokarbamiidi) maitsetundmise tunnust (kokku 24 alleeli). On tutvustatud vastavaid geneetilisi süsteeme ja alleelisagedusi eestlastel keskmiselt ja neljas kõige omanäolisemas regioonis.

Eesti lokaalsete rühmade vahel esineb suhteliselt suuri lahknevusi. Seejuures on geneetilised erinevused Lääne- ja Ida-Eesti vahel oluliselt tugevamini väljendunud kui Põhja- ja Lõuna-Eesti vahel.

Enamik lokaalseid valimeid (uuritud 39-st) rühmitub geneetiliste distantside klasteranalüüsil oma lähimate naabritega suuremateks regionaalseteks rühmadeks kahe erandiga. Muhu rühm ei seostu mitte teiste saarte ja Läänemaa omadega, vaid Kirde-Eesti regiooni rühmadega. Võru murdeala lääneosa (Antsla rühm) erineb märgatavalt teistest Kagu-Eesti valimeist. Samal ajal on Setu rühmad väga tihedalt seotud Kagu- ja Ida-Eesti regiooni omadega.

Kesk-, Ida-, Kagu- ja Edela-Eesti regionaalsed populatsioonid on omavahel väga lähedased. Kirde-Eesti, Lääne-Eesti saarte ja mandriosa populatsioonid on omavahel ja ülejäänud eestlaste rühmadest võrdlemisi erinevad.

Geneetiliste distantside alusel on eestlastega kõige sarnasemad lähimad naabrid (venelased, lätlased) ja mõned soome-ugri rahvad (vepslased, karjalased). Seega ei lange geneetiline ja keeleline ühtekuuluvus kokku.

On diskuteeritud ka nn. mongoliidsete tunnuste üle soome-ugri rahvastel ning leitud, et mongoliidne lisand ei võimalda seletada samaaegseid läänelikke tunnuste- ja geenisagedusi. Oletatavasti on mitmete tunnuste antagonistlikud sagedused soome-ugri alpopulatsiooni originaalse geneetilise struktuuri jäljed.