Analysis of mathematical approaches to the assessment of population processes on the basis of cattle blood groups characteristics

V E Gerasimova¹, L A Gerasimova¹, I Yu Eremina² and A I Kuklina¹

¹Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarskii Rabochii Av., Krasnoyarsk, 660037, Russian Federation
²Krasnoyarsk State Agrarian University, 90, Mira Av., Krasnoyarsk, 660049, Russian Federation

E-mail: lyu-gerasimova@yandex.ru

Abstract. The immunogenetic characteristics of cattle blood groups are their high variability and heterogeneity. Population processes that occur under the influence of selection pressure can be estimated on the basis of the analysis of the antigen-frequency characteristics of blood groups. Using specific herds as an example, statistical indicators were calculated, and the choice of a statistical method for identifying micropopulation processes was rationalized. A two-sample Student’s t-test for paired samples on the basis of immunogenetic indicators helped to establish differences between populations and identify processes in blood group systems.

Features of the immunogenetic characteristics of cattle blood groups are their high variability, heterogeneity and correlation with economically useful traits. On the one hand, these are qualitative indicators because 65 antigens of blood groups that can be present in every animal in almost any combination are analyzed. On the other hand, each antigen in the studied group of animals is observed with a certain typical frequency which is a quantitative characteristic. The frequency of occurrence of blood group antigens can be a very specific indicative means of describing the sample, that is, the studied group of animals. The population of animals is usually taken as a sample. As for cattle, the number of livestock of a particular farm which can also be attributed to the category of populations is taken as a sample. Thus, antigen-frequency indicators help to evaluate the population processes occurring under the pressure of natural and artificial selection in the selection process. They can be used both to characterize intrapopulation (intragroup) processes and to describe interpopulation (large groups, breeds) differences [1–4].

The purpose of the paper is to evaluate population processes that occur under the influence of selection pressure on the basis of the analysis of immunogenetic characteristics of cattle blood groups. In the course of the research the following problems were solved: 1) rationale of the choice of a statistical method for identifying micropopulation processes on the basis of cattle antigen-frequency characteristics; 2) calculation of statistical indicators on the example of specific herds; 3) biological interpretation of the obtained statistical indicators relative to ongoing population processes using real data as an example.
Population data for 3 periods was compared. The basic herd (initial data) was compared to the herd already changed during selection with an interval of animal birth in 5 years and 3 years, which are conventionally designated as periods 1, 2 and 3. Therefore, the time interval between the population data for periods from 1 to 3 is 8-10 years. In terms of the breeding work specifics, we assume that the population of period 1 will have some possibly not very significant differences compared with period 2 and significant changes with respect to period 3. Moreover, the characteristics of the population of periods 2 and 3 will be very close. The number of the analyzed livestock was 214, 435 and 170 animals for periods 1, 2 and 3, respectively.

The correlation coefficient [5-8] is usually used over the entire complex of antigens to compare changes in the state of the population as a whole. The correlation coefficient between the states of the population of all three periods is close to 1, which indicates the presence of a high degree of direct correlation between the series of frequencies in different time periods, i.e., on the basis of antigen frequency of one year we can predict the frequency of this antigen in another studied period. Moreover, the relationship between the frequencies of periods 2 and 3 is expected to be closer than the relationship between the data of the same periods with frequencies for period 1. The results are presented in table 1.

| Studied periods | 2   | 3   |
|-----------------|-----|-----|
| 1               | 0.85| 0.82|
| 2               | -   | 0.90|

Table 1. The values of the correlation coefficient between different periods.

Thus, the correlation coefficient did not reveal significant differences in antigen-frequency characteristics of the herd in the interval of almost 10 years, although there is a tendency to decrease the level of correlation with an increase in the time gap between the populations of the studied periods.

It is logical to assume that, taking into account the constantly conducted breeding process during which new genotypes are introduced into the population, there should be changes in the composition of antigens. Indeed, the values of antigen frequencies vary significantly. Nevertheless, the basic gene pool of the population also retains its influence. Therefore, it becomes necessary to identify and evaluate the speed and scale of these changes. Moreover, taking into account the generality of the underlying gene pool, samples in different time periods are considered as dependent populations.

In order to establish the presence or absence of differences between the antigen frequencies in one population in different periods, it is necessary to check the homogeneity of these characteristics, that is, to establish two sets of antigen frequencies from one general population or from different ones. Since research is conducted on antigens in one animal population, the samples are frequency dependent. Each antigen is associated with the frequency of occurrence in the population in each of the studied periods, which determines the paired data series. In this case, it is advisable to use the two-sample Student’s t-test for dependent (paired) samples [5], [7-10], which allows identifying whether two rows of antigens are homogeneous or different.

To calculate the empirical value of the t-test, the difference between the frequencies of each antigen in the series is calculated:

$$Z = \frac{1}{k}\sum_{i=1}^{k}z_i, \quad z_i = p_{1i} - p_{2i},$$

where $z_i$ is the frequency difference of the $i$-th antigen in different periods, $p_{1i}$ is the frequency of the $i$-th antigen in period 1, $p_{2i}$ is the frequency of the $i$-th antigen in period 2, $k$ is the number of antigens in the frequency series.

Next, the average and the difference between
where $\bar{Z}$ is the average value of the difference of antigen frequencies, $k$ is the number of antigens in the frequency series, $s_x$ is the standard deviation of the differences.

$\bar{Z}$ is similarly calculated for periods 2-3 and 1-3.

We consider the null hypothesis of the absence of differences (homogeneity) in two series of frequencies, i.e., the mean value of the frequency difference is equal to zero: $H_0: \bar{Z} = 0$; and the alternative hypothesis of the difference in the two series of frequencies, i.e., the fact that the average value of the frequency difference is not equal to zero: $H_0: \bar{Z} \neq 0$.

The empirical value of the two-sample Student’s $t$-test for dependent (paired) samples:

$$t = \frac{\bar{Z}}{s_x/\sqrt{k}}$$

These statistics have distribution $t (k-1)$. The calculated value $t$ is compared with the tabular $t_{cr}$ for $(k-1)$ number of degrees of freedom and the set level of significance [5], [7-10].

If the module of the criterion calculated value is less than the tabular $|t| < t_{cr}$, then the null hypothesis is accepted and the conclusion is made that the two series of antigen frequencies are homogeneous. If the module of the criterion calculated value is greater than the tabular $|t| > t_{cr}$, then the null hypothesis is rejected and an alternative hypothesis about the difference in two series of frequencies is accepted, it is concluded that two series of antigen frequencies are different.

Presence or absence of differences between the frequencies of all considered antigens in different periods was established in the paper.

Table 2. The use of Student’s $t$-test for comparing populations over the entire set of antigens.

| Studied periods | 2 | 3 |
|-----------------|---|---|
| 1               | Presence of differences | Presence of differences |
| 2               | - | Absence of differences |

Table 2 clearly shows that over the entire set of blood group antigens the two-sample Student’s $t$-test revealed differences between the immunogenetic characteristics of the population between periods 1-2 and 1-3, which corresponds to the initial assumption.

In addition, the research had to establish the presence or absence of differences between the antigen frequencies in classes (according to the genetic systems of blood groups). Cattle have several multivariate blood group systems. In the studied population there have been distinguished 5 such systems of 2, 6, 7, 10, and 34 antigens. All studied antigens were divided into classes in accordance with blood group systems, and then for each class of antigens we considered the question of homogeneity (or difference) between the series of antigen frequencies in different periods, i.e., we compared the series of antigens of one blood group system obtained in a population in different time periods. No statistically significant differences were found for systems represented by no more than 10 antigens. Significant differences were determined only in the most multivariate system including 34 antigens (EAB-system), reliability levels being 95 and 99% (table 3).

Table 3. Identification of intrapopulation changes on the basis of antigenic differences in the EAB-system of blood groups.

| Indices          | t-test values | Presence of differences |
|------------------|---------------|-------------------------|
| Reliability level 95% | 2.03          | -                       |
Table 3 shows the pronounced interpopulation differences in the EAB system between periods 1 and 3. Changes in the genetic structure between the most distant in time states of the population are clearly visible. Populations close to each other in time frames showed differences only with reliability level of 95%. Thus, we can note the initial stages of genetic divergence. The divergence only intensifies over time, which is explained by the action of directed selection. Obviously, antigens of the multifactorial EAB-system of blood groups are subjected to the highest selection pressure during the selection process, that is, they are probably more actively correlated with economically useful traits.

In addition, the research found the presence or absence of differences between the frequencies within the sorted groups of antigens. In each of the three populations antigens are ranked according to the frequency of occurrence. All the studied antigens were divided into classes: rarely occurring and eliminating, often occurring and medium ones with occurrence thresholds of 0-15%, over 50% and 16-50%, respectively.

Due to the fact that in each of the studied periods the division of antigens into these groups is different, we consider the comparison of the antigen group frequencies separately for each period. That is, first, for each of the studied periods, the division of all antigens into groups is made and this division is considered to be basic. Further, the question of homogeneity or difference in the frequency series of each antigen group with the corresponding frequency series of the same antigens in another (non-basic) period is considered during the basic division of antigens into groups. It should be noted that different numbers of antigens fall into the frequency classes in different periods and they differ in their qualitative composition.

In the analysis, when period 1 was taken as the basic one and frequency classes were compared in pairs with the same indicators of periods 2 and 3, we obtained statistically significant differences with both periods in the group of rare antigens, which is also significant. Rarely occurring and eliminated antigens are a group washed away by selection, i.e. a part of the gene pool that is not supported by selection and environmental factors.

It is logical that in base period 2 there are no differences with the frequency classes of periods 1 and 3. Comparing period 3 with the data of periods 1 and 2, significant differences were noted in the class of medium-encountered antigens only with period 1. In all other cases differences are not defined.

The two-sample Student’s t-test is most informative for assessing changes in the genetic structure of populations at the initial stages of their development. The correlation coefficient does not capture the initial levels of divergence of population traits. Population processes are most clearly manifested by the antigen-frequency characteristics of the EAB-system of blood groups. Antigen classes in frequency of occurrence exhibit characteristic distinguishing features only between distant periods. A class of rare antigens can be considered as indicative of identifying cumulation of differences.

Reference
[1] Bukarov N and Silkin S 2011 Genetic monitoring in dairy cattle breeding using marker blood groups Molochnoe I myasnoe skotovodstvo 7 14-6
[2] Tihonov V N and Bobovich V E 2004 Monitoring of microevolution and breed formation of pigs on the basis of molecular immunogenetic analysis Sel’skohozyaistvennaya biologiya 2 10-27
[3] Eremina I Yu, Gerasimova L A and Lushhenko A E 2016 Retrospective analysis of phylogenesis in the formation of uterine subpopulation of Holstein dairy cattle in Krasnoyarsk region (Vestnik Omskogo gosudarstvennogo agrarnogo universiteta ) 2 (22) pp 100-8
[4] Nikitin S V et al 2009 Dynamics of heterozygosity and its connection with adaptability in the population of domestic pigs Genetika. 45 pp 1401-10
[5] Aivazyan S A and Mhitaryan V. S 2001 Probability theory and applied statistics (Moskva: Yuniti-Dana) vol 1 p 656
[6] Vencel’ E S and Ovcharov L A 1969 Probability theory (Moskva) p 576
[7] Gmurman V E., Rukovodstvo 1979 Guide to solving problems in probability theory and mathematical statistics (Moskva: Vy’ssh. shkola) p 400
[8] Kremer N Sh 2010 Probability theory and mathematical statistics: Ucheb. Posobie dlya studentov vuzov (Moskva: Vy’ssh. shkola) p 551
[9] Ruban A I 2004 Data analysis methods (Krasnoyarsk: IPTs KGTU) p 319
[10] Khizhnyak S V and Muchkina E Y 2005 Mathematical methods in biology and ecology (Krasnoyarsk: Gos. agrarn. universitet ) p 54