# QUANTIFICATION METODOLGY OF GENETIC BIODIVERSITY IN PLANT AND ANIMAL POPULATIONS ON GENETIC AND INFORMATIONAL STATISTICS BASES

Gh. HRINCĂ, M. GROZA, I. CHIORESCU, Elena FECIORU

*Research and Development Station for Sheep, Popăuți-Botoşani*

## MOTIVATION

- An inventory of flora and fauna on the Terra estimates that the number of vegetable and animal species could surpass three millions. Besides the scientific interest, this genetic capital represents an inestimable agricultural, chemical and pharmaceutical potential. The intensivization of the agricultural production is achieved by specific improvement programmes, leading to morpho-productive entities with a high efficiency but which are vulnerable at morbidity factors. The supported application of these programmes, by the selection of more and more improved genotypes contributed, simultaneously with the technology optimization, to gradual increase of agricultural production. But the genetic improvement activity of plant and animals can have negative effects too. The extension of improvement activities, besides their positive and useful aspects, ousts from reproduction not only species but especially subspecies, breeds, varieties. The disappearance of subspecific genotypes limits the species variability, impoverishing it, determining severe ecological disequilibriums. These events led an will lead to the expulsion of subspecies of local populations, of “rustic” breeds and varieties with lower performances, but which are better accommodated to the environment conditions and constitutes at the same time a precious source of genetic variability.

## PURPOSE AND MEANS

- The scientific exploration, collection and preservation of the genetic resources is a pressing action of the greatest importance. In this sense, a complex and complete evaluation of the essence and dynamics of the biological world is achievable using some concepts derived from the informational statistics (*informational correlation, informational entropy, informational energy*) which do not use measurable sizes (as in the mathematical classical statistics) but they operate with qualitative sizes, frequencies, variables with a high abstraction degree. Therefore, the present theme answer to these desiderata putting at hand a very efficacious instrument concerning the evaluation of the genetic resources and of the biodiversity in plant and animal populations with the help of the genetic markers by informational statistics quantification and, alike, for the prevention of some genetic disequilibriums caused by the surrounding environmental conditions and by the selection and technological factors, for the restrengthening of the genetic patrimony with a view to for diminish the systemic entropy and to install the genetic and physiology homeostasis within the ecosystems by specific self-regulation mechanisms.

# MATERIAL AND METHOD

## Biological material

The experiment was achieved on five sheep populations belonging to the five Romanian breeds:

- Merino
- Tsigai
- Spanca
- Tsurcana
- Karakul.

Any biological material which presents polymorph structures with discontinuous variability (proteins, enzymes etc), belonging to either vegetable kingdom or animal kingdom or mineral kingdom, can be used in such studies.

## Biochemical-genetic typification of animals

- haemoglobin system – by Smithies electrophoresis method;
- transferrin system – by Ashton-Ferguson electrophoresis method;
- albumin – by Kristjansson electrophoresis method;
- amylase – by Kristjansson electrophoresis method;
- blood potassium - by Domigo-Klyne flame photometric method.

## Immunogenetic typification of animals

- blood group systems – by haemolysis test method

# INFORMATIONAL STATISTICS METHODOLOGY

1) The **genetic distance (D)** - by the ***Nei method***:

$$D = -\log_e \frac{Jx.y}{(Jx.Jy)^{1/2}} = 2\alpha t, \text{ where:}$$

-D = standardized genetic distance; -t = differentiation time;

-$\alpha$ - substitution percentage of genes per locus and per year whose value is $10^{-7}$.

-Jxy = probability of genetic identity between two populations x and y; $Jxy = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} px_{ij}\, py_{ij}$,

-Jx – probability of genetic identity within of population x: $Jx = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} (px_{ij})^2$

-Jy – probability of genetic identity within of population y: $Jy = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} (py_{ij})^2$

2) The **informational correlation** ($C_{x,y}$) – by the ***Onicescu method***

$$C_{x,y} = \sum_{i=1}^{n} f_{xi} f_{yi} \text{ ; in which}$$

$f_{xi}$ and $f_{yi}$ are the frequencies of the first, respectively, the second population.

$$R_{x,y} = \frac{\sum_{i=1}^{n} f_{xi} f_{yi}}{\sqrt{(\sum_{i=1}^{n} f_{xi}^2)(\sum_{i=1}^{n} f_{yi}^2)}}$$

The value of the informational correlation coefficient varies between *0* and *1*

$$0 \le R_{x,y} \le 1$$

Than

- the informational correlation coefficient $R_{x,y}$ can reach the value *1* only if the frequency $f_{xi}$ and its correspondent $f_{yi}$ have the value *1*, and
- the informational correlation coefficient $R_{x,y}$ takes the value *0* only if the two populations have no common characteristic, they are strange one towards the other, therefore they are in relation of estraneity.

**3)** The **informational entropy (H) -** by the **Shannon method**

$$H = -K\sum_{i=1}^{m} p_i \ln(p_i) \text{ ; where } K = 1/m$$

$$d = 1 - H$$

- The advantage of this method is that the diversity coefficient takes values between *0* and *1* and it does not require a correction to compare two structures with a different number of states (classes).
- The disadvantage is that if the frequency value of a state is *0*, the parameter *H* can not be calculated because the logarithm function is not definite in point *0*.

**4)** The **<u>informational energy</u> (e) -** by the **<u>Onicescu method</u>**

$$e = \sum_{i=1}^{m} p_i^2$$

*p* = the probability (which varies from *0* to *1*);
*m* = number of classes (states).

This indicator takes the maximum value, (*1)*, when in the system a state has maximum frequency, the other states having null frequencies, and minimum value, (*1/m)*, when all states have equal frequencies. So, the diversity coefficient can be thus defined:

***d=1-e***

▪ The advantage of this method is that it allows the diversity measurement.

▪ The disadvantage of this method is that the diversity coefficient takes values between *1* and *1/m* (which can be close to *0*, but never *0*).

To remedy this drawback in the mathematical speciality literature there was introduced the corrected informational energy indicator (***$e_c$***). This parameter takes values between *0* and *1* allowing the reduction of the system to a simple state (class) :

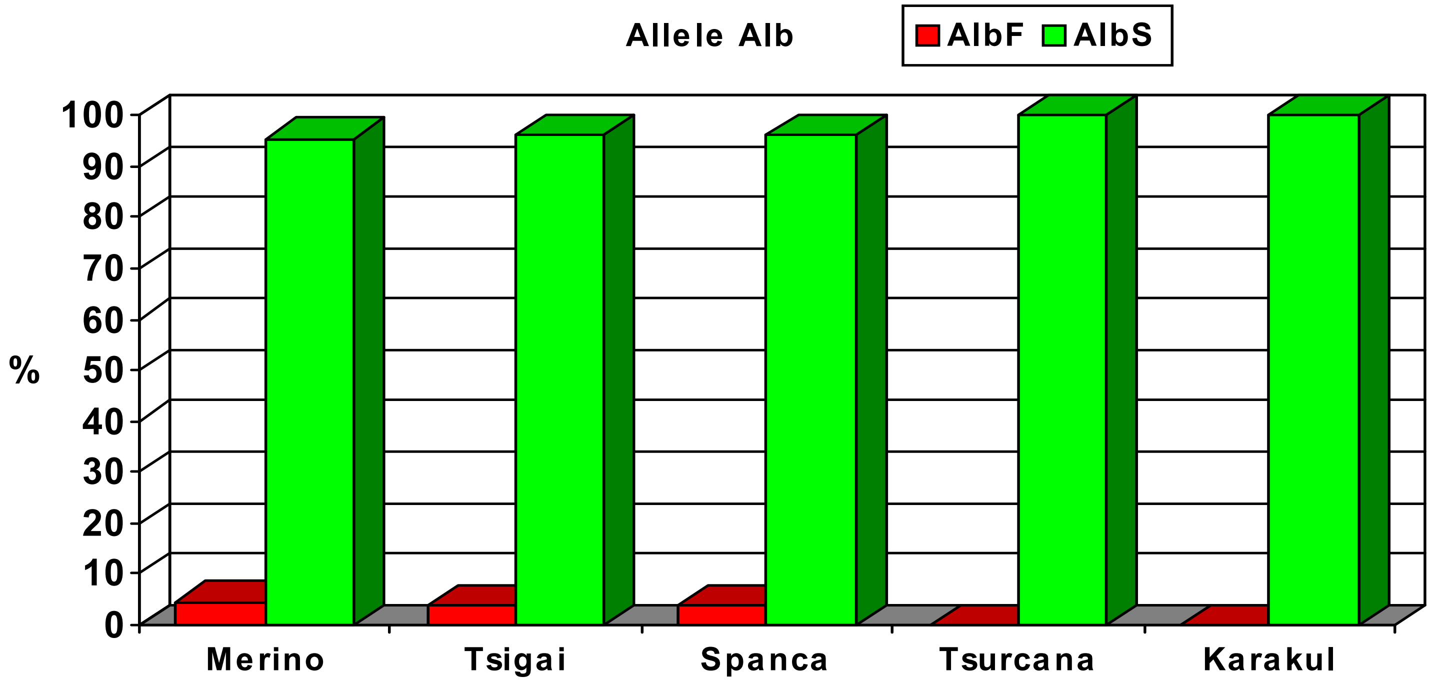
$$e_c = \frac{\left(\sum_{i=1}^{m} p_i^2 - \frac{1}{m}\right)}{\left(1 - \frac{1}{m}\right)}$$

the diversity coefficient becomes: ***$d_c = 1 - e_c$***

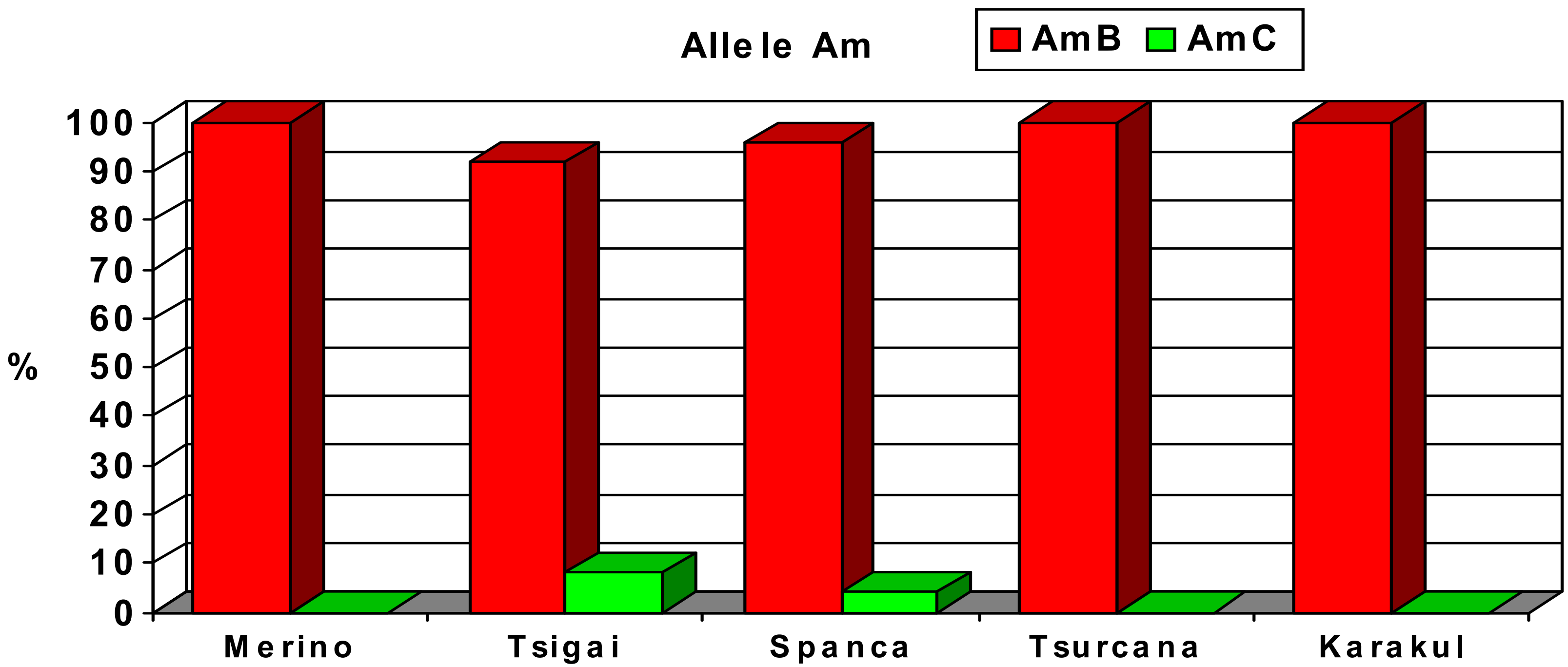
For an overall perception of the polymorphism of all the genetic systems there were calculated the *complex informational energy* (**$c_t$**) and the *complex genetic diversity* (**$d_t$**).

By extension, the *energy informational* parameter is in connection with the one of the *informational correlation*:

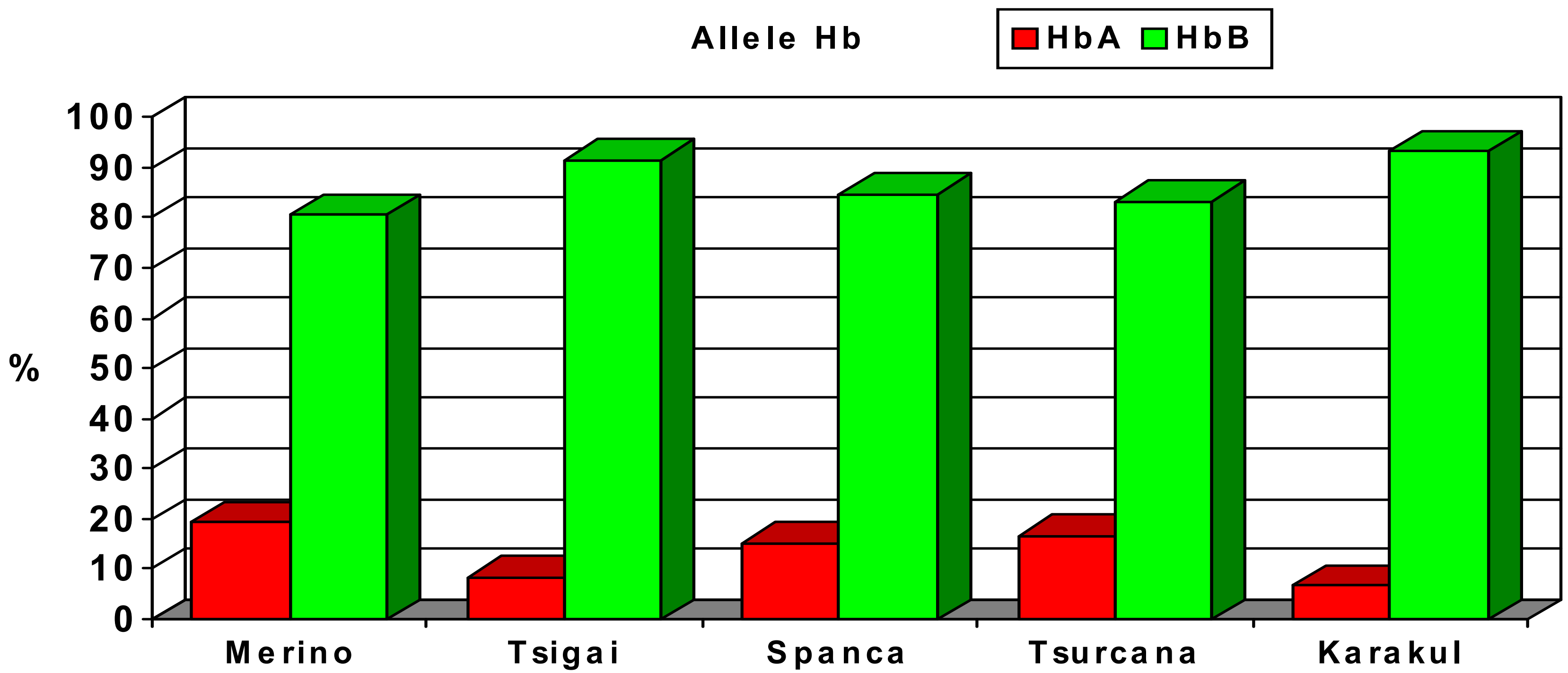
$$e = \sum_{i=1}^{n} p_i^2 = C_{xy}$$



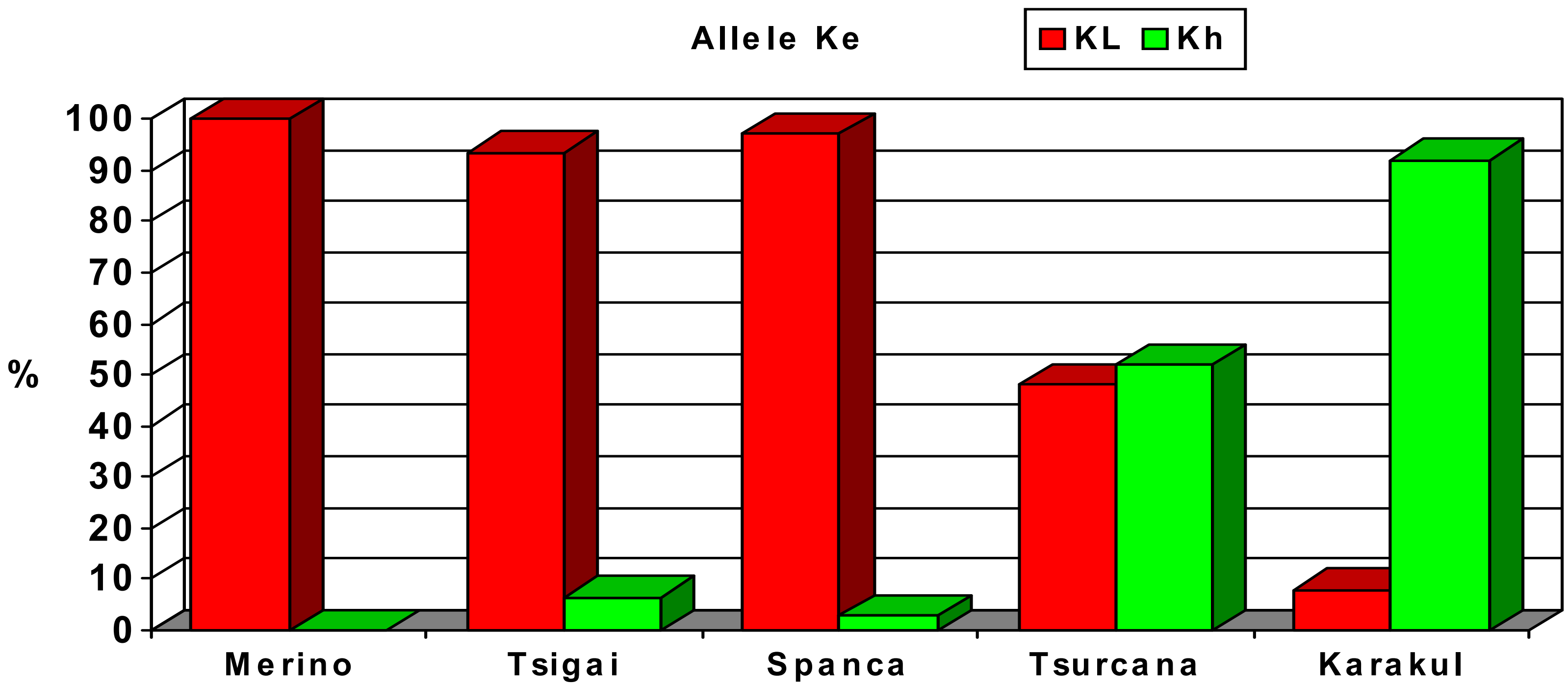

Allelic structure at the albumin locus in the Romanian sheep breeds



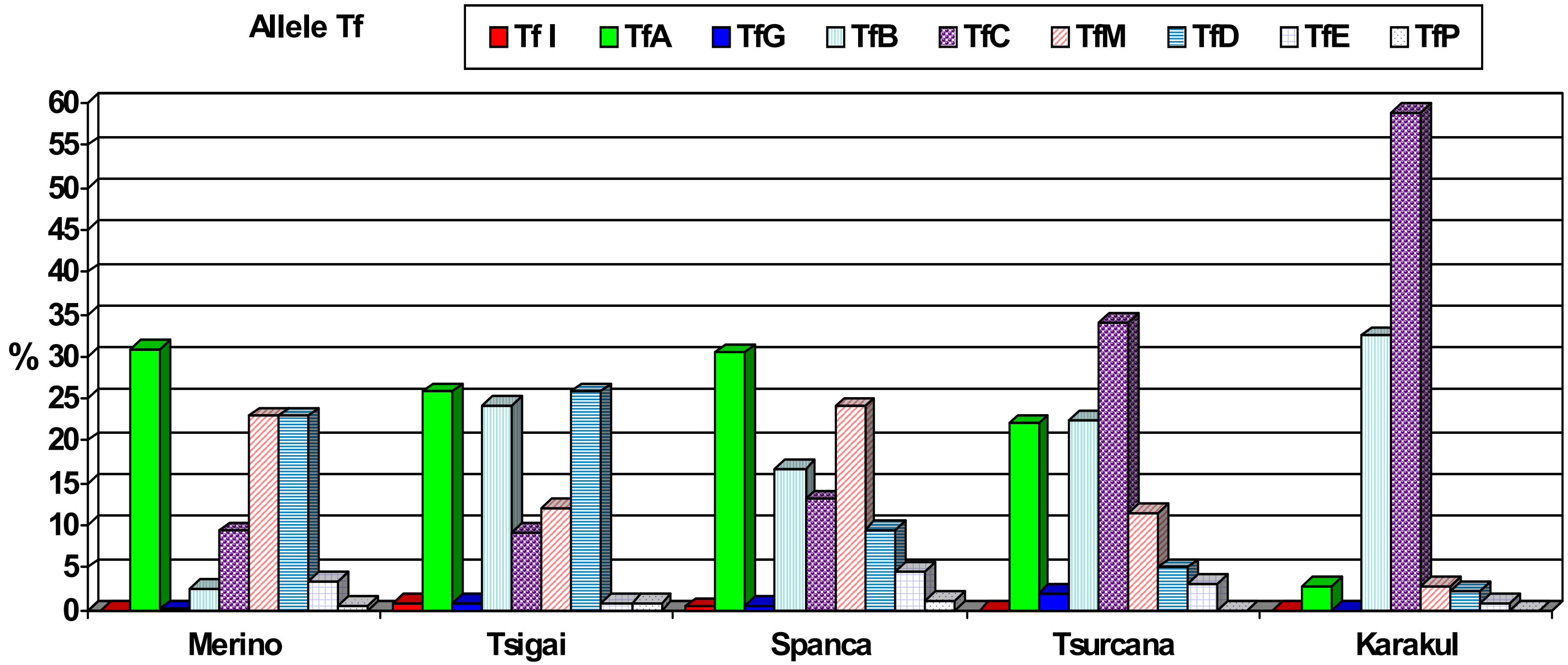

Allelic structure at the amylase locus in the Romanian sheep breeds




Allelic structure at the haemoglobin locus in the Romanian sheep breeds

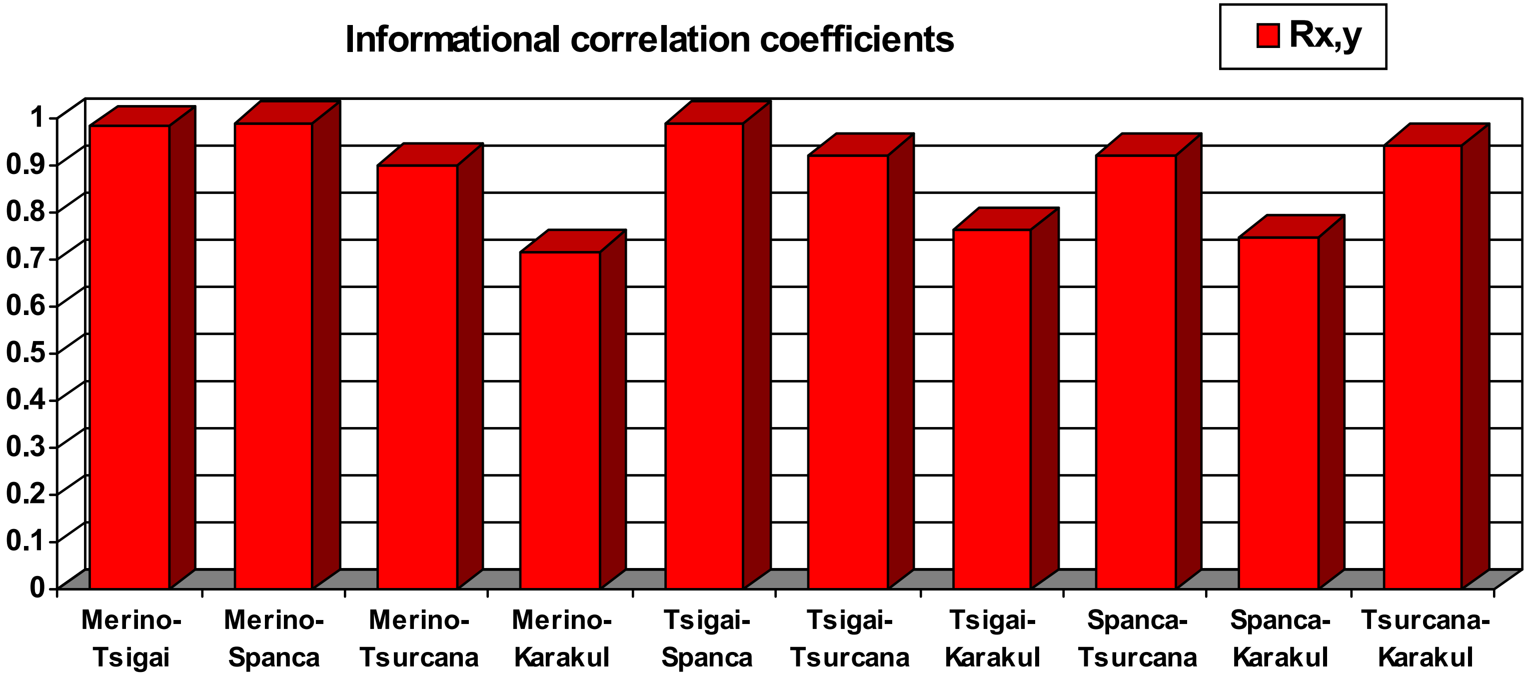



Allelic structure at the potassium locus in the Romanian sheep breeds

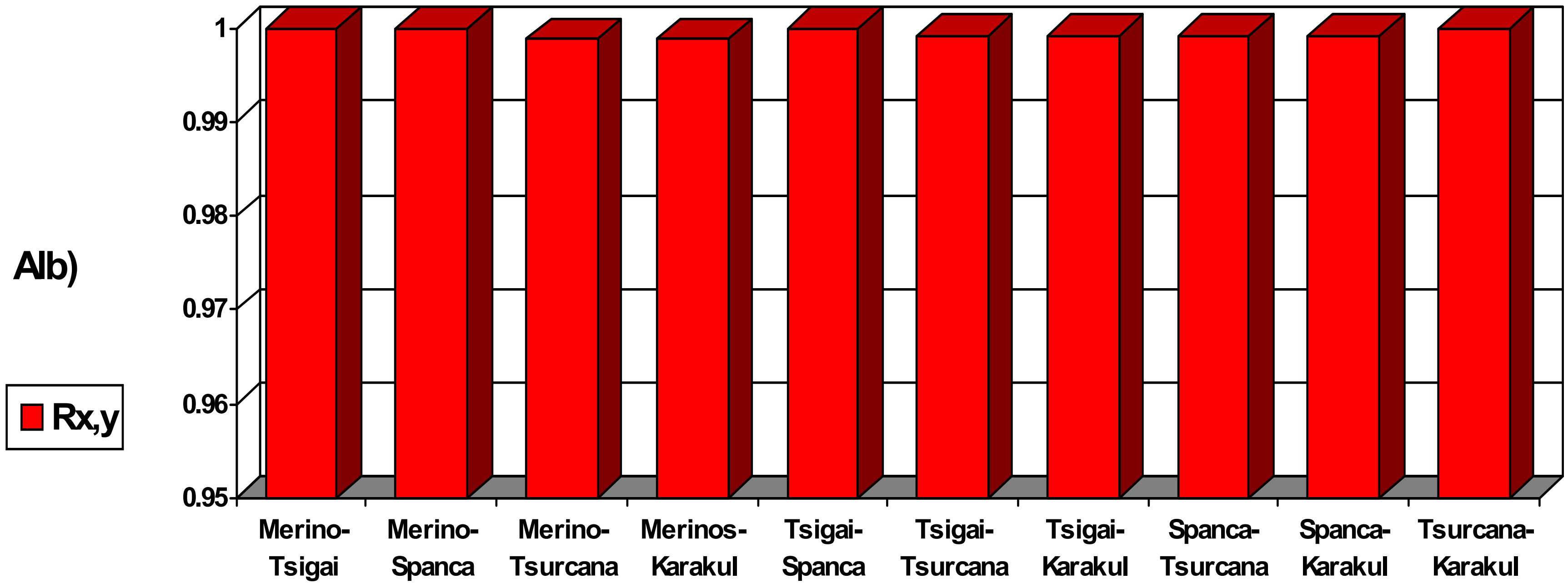



Allelic structure at the transferrin locus in the Romanian sheep breeds

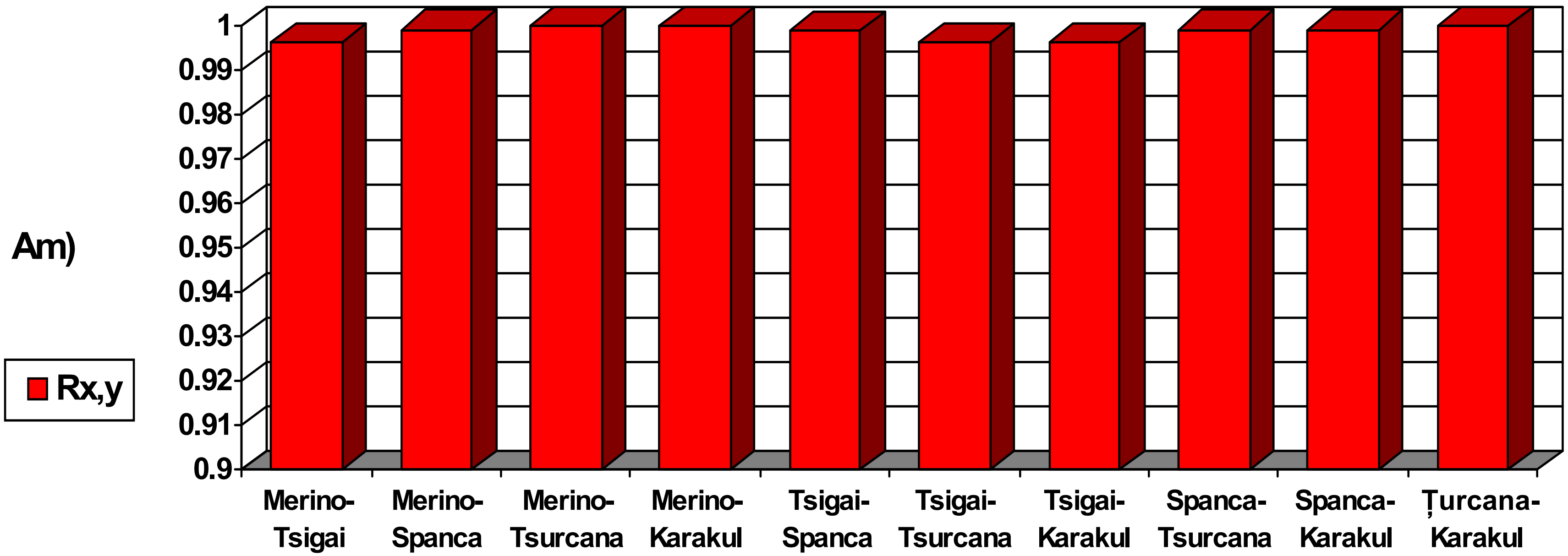
| Breed | Merino | Tsigai | Spanca | Tsurcana | Karakul | |
|---|---|---|---|---|---|---|
| Merino | | **0,01724789** | **0,00712532** | **0,10174742** | **0,33218377** | **D** |
| Tsigai | 0,9829 | | **0,01045445** | **0,08055950** | **0,26722606** | |
| Spanca | 0,9929 | 0,9896 | | **0,08088472** | **0,28634962** | |
| Tsurcana | 0,9033 | 0,9226 | 0,9223 | | **0,05763039** | |
| Karakul | 0,7173 | 0,7655 | 0,7510 | 0,9439 | | |
| | **$R_{x,y}$** | | | | | |




Coefficients of the genetic distances (***D***) and informational correlations (***Rx,y***) among the Romanian sheep breeds concerning the wholeness structure at the determinant loci of the biochemical-genetic systems

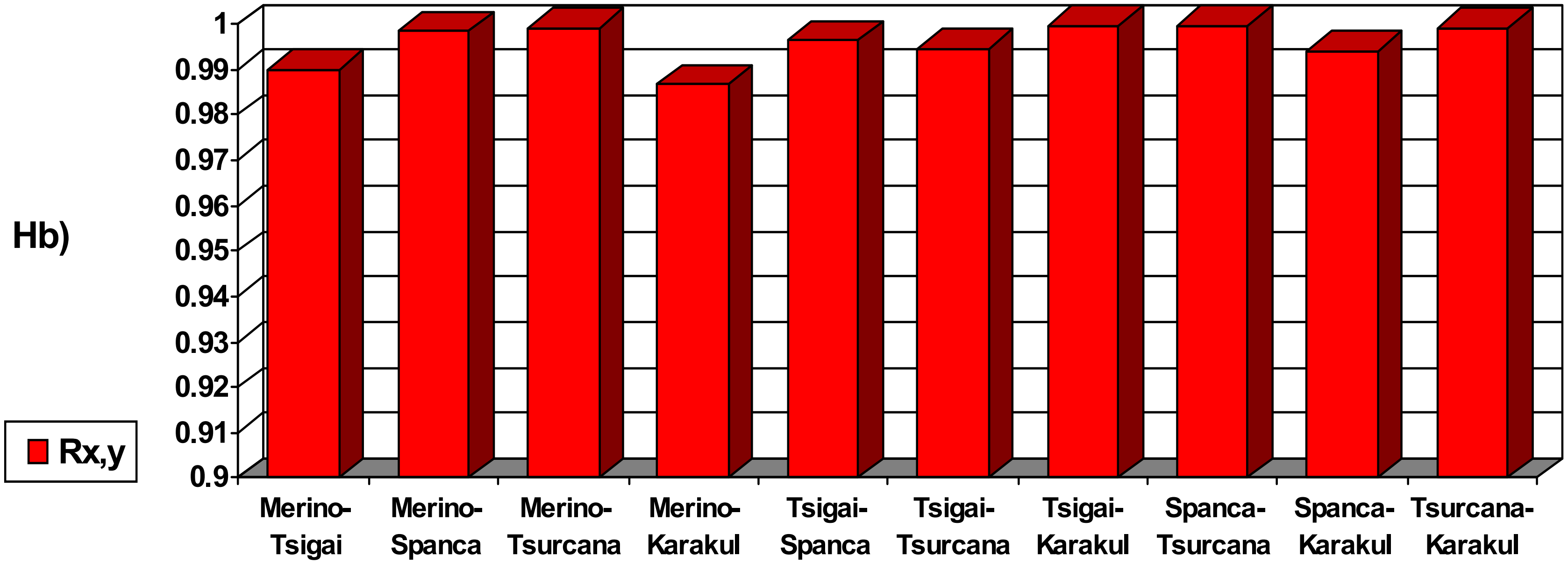



Informational correlation coefficients ($\boldsymbol{R_{x,y}}$) among the Romanian sheep breeds concerning the allelic structure at the albumin locus

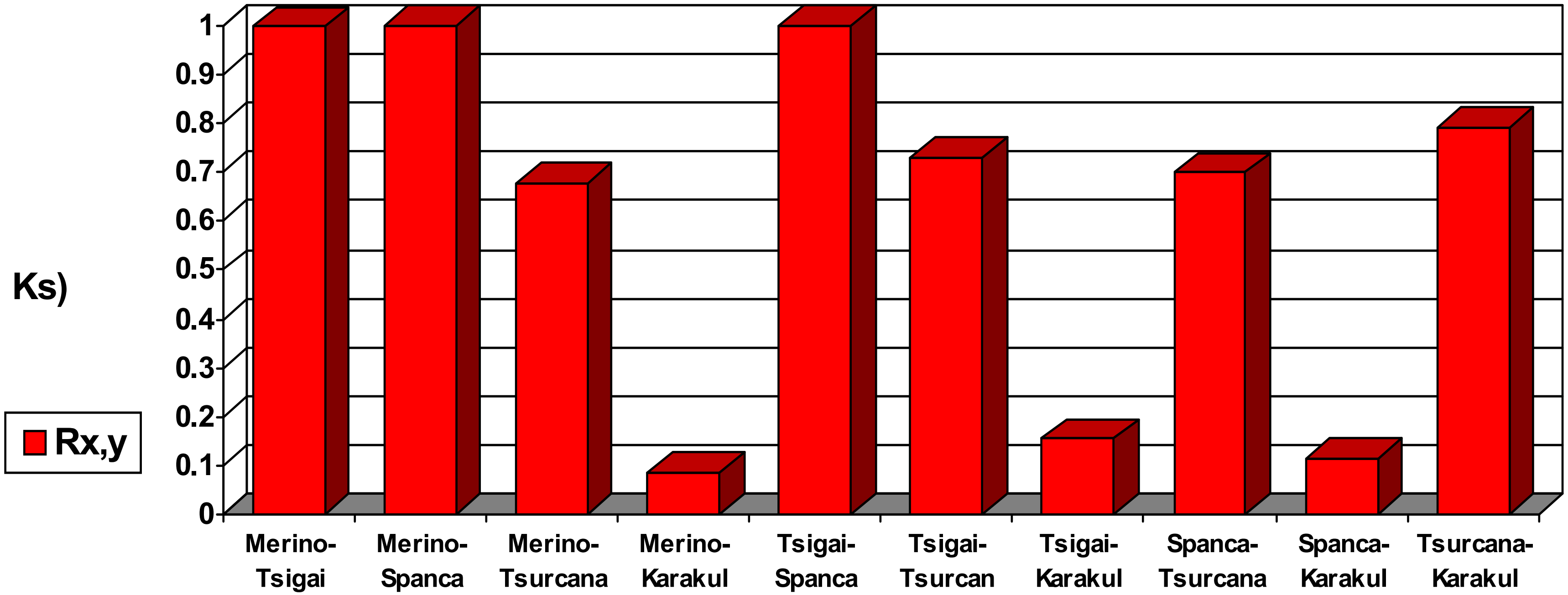



Informational correlation coefficients (***$R_{x,y}$***) among the Romanian sheep breeds concerning the allelic structure at the amylase locus

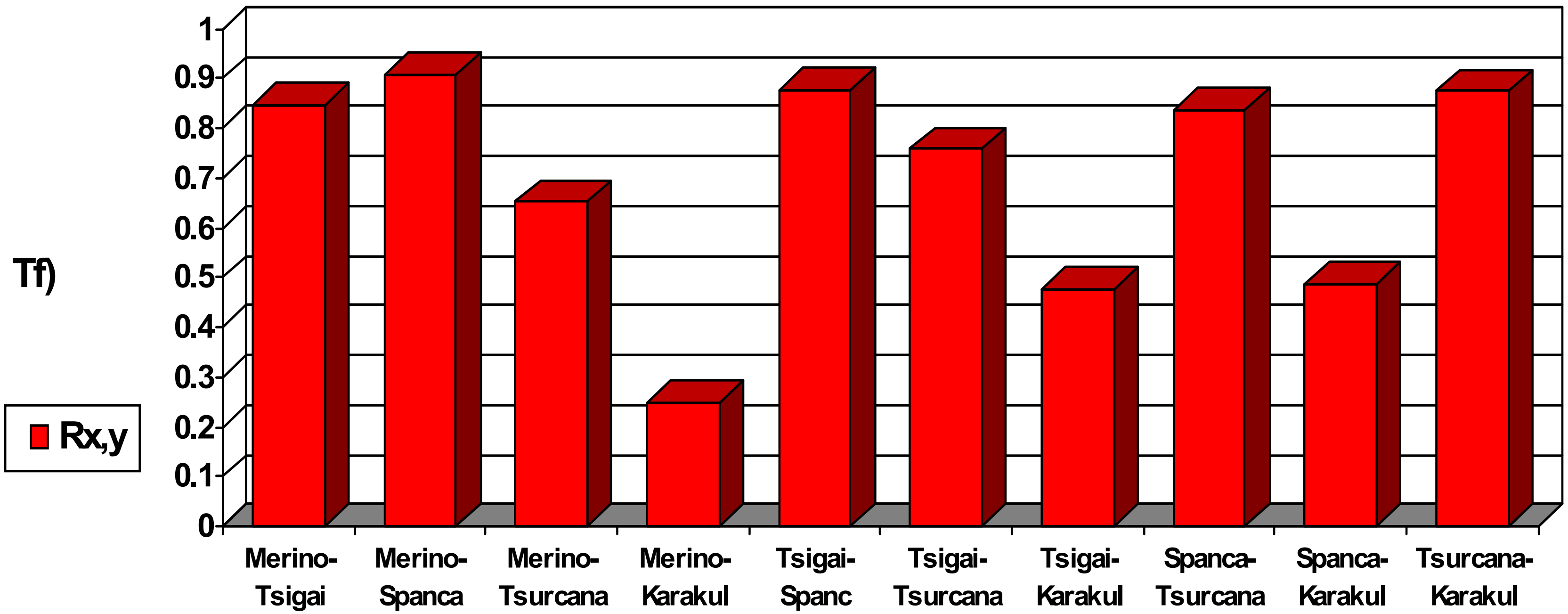
Hb)



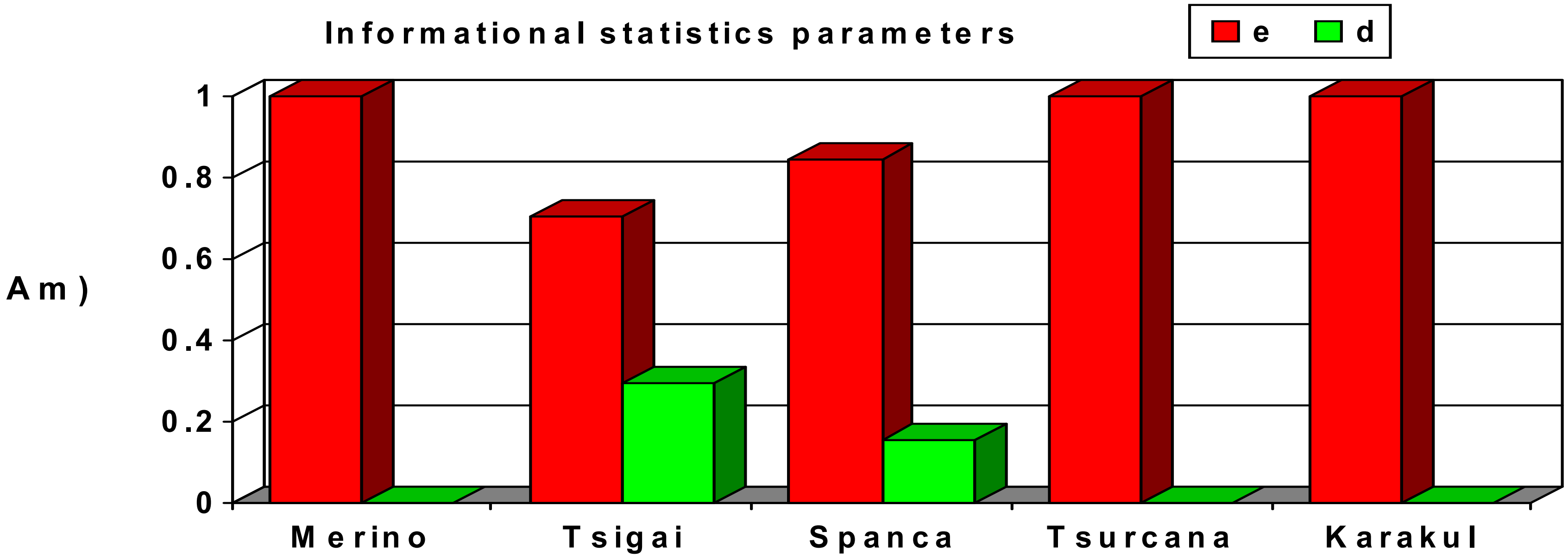

Informational correlation coefficients ($R_{x,y}$) among the Romanian sheep breeds concerning the allelic structure at the haemoglobin locus



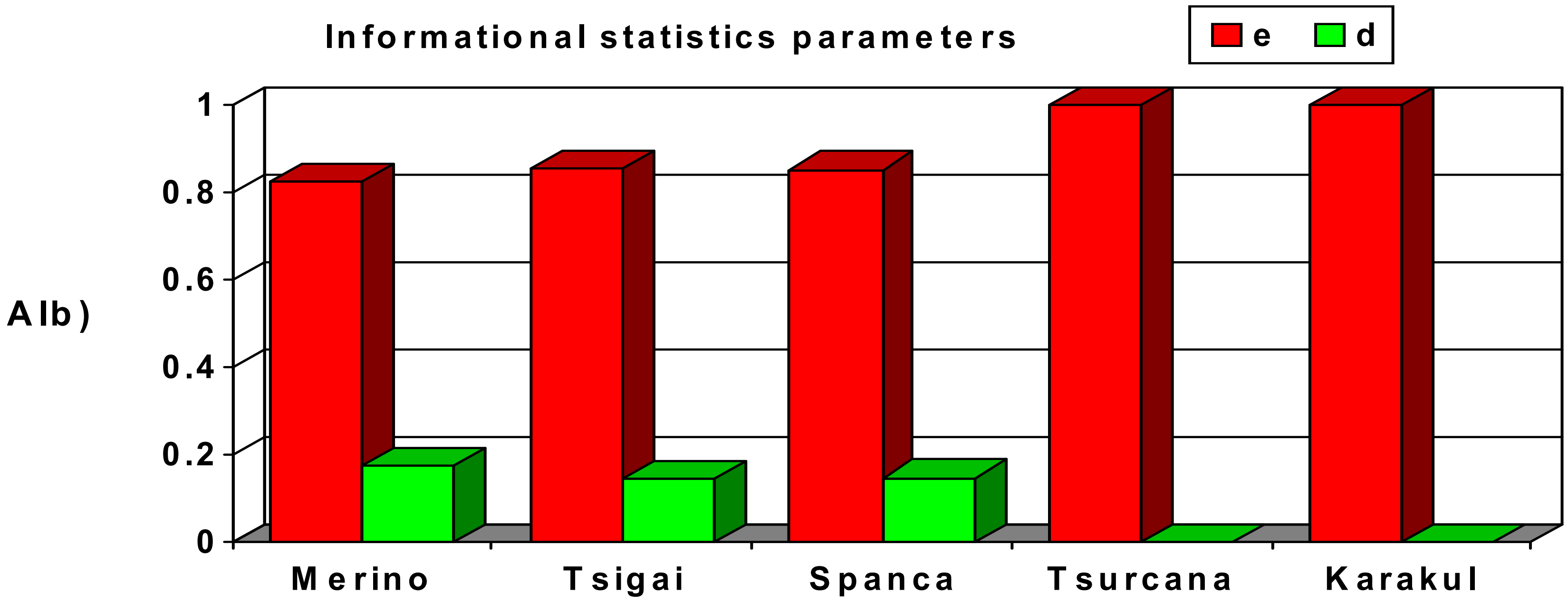

Informational correlation coefficients ($R_{x,y}$) among the Romanian sheep breeds concerning the allelic structure at the potassium locus



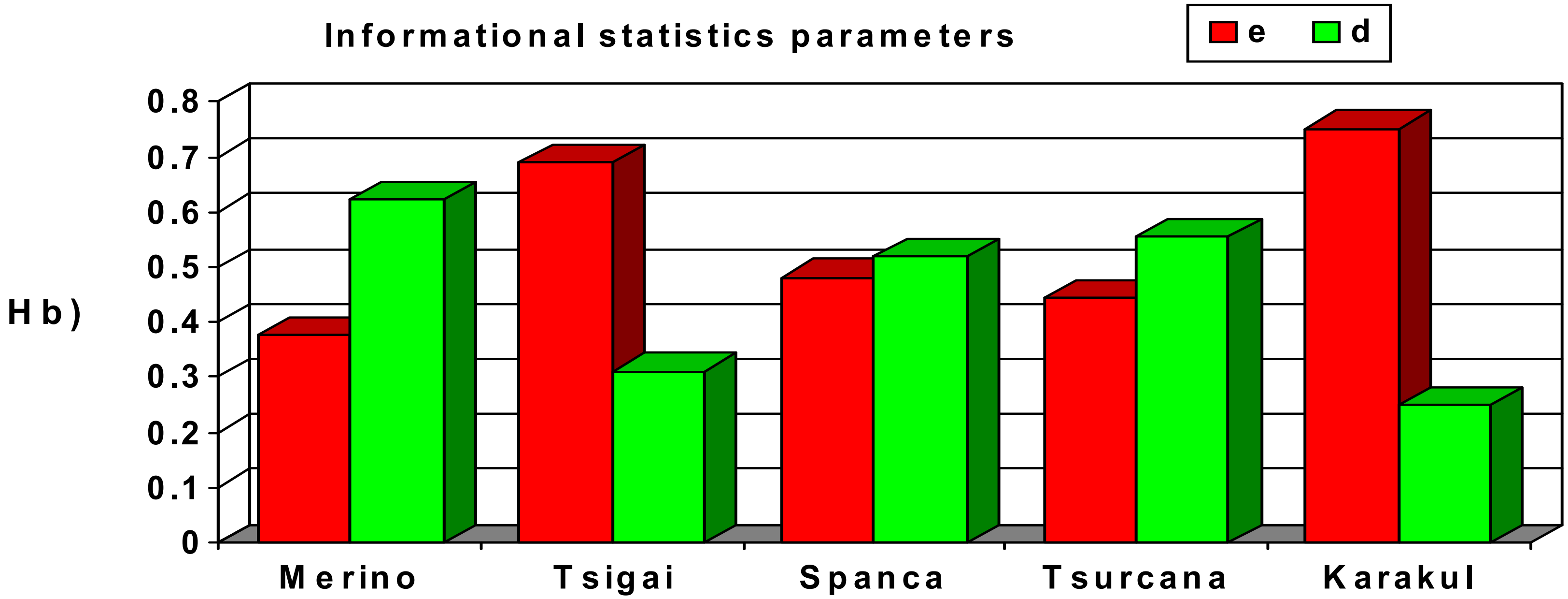

Informational correlation coefficients ($R_{x,y}$) among the Romanian sheep breeds concerning the allelic structure at the transferrin locus



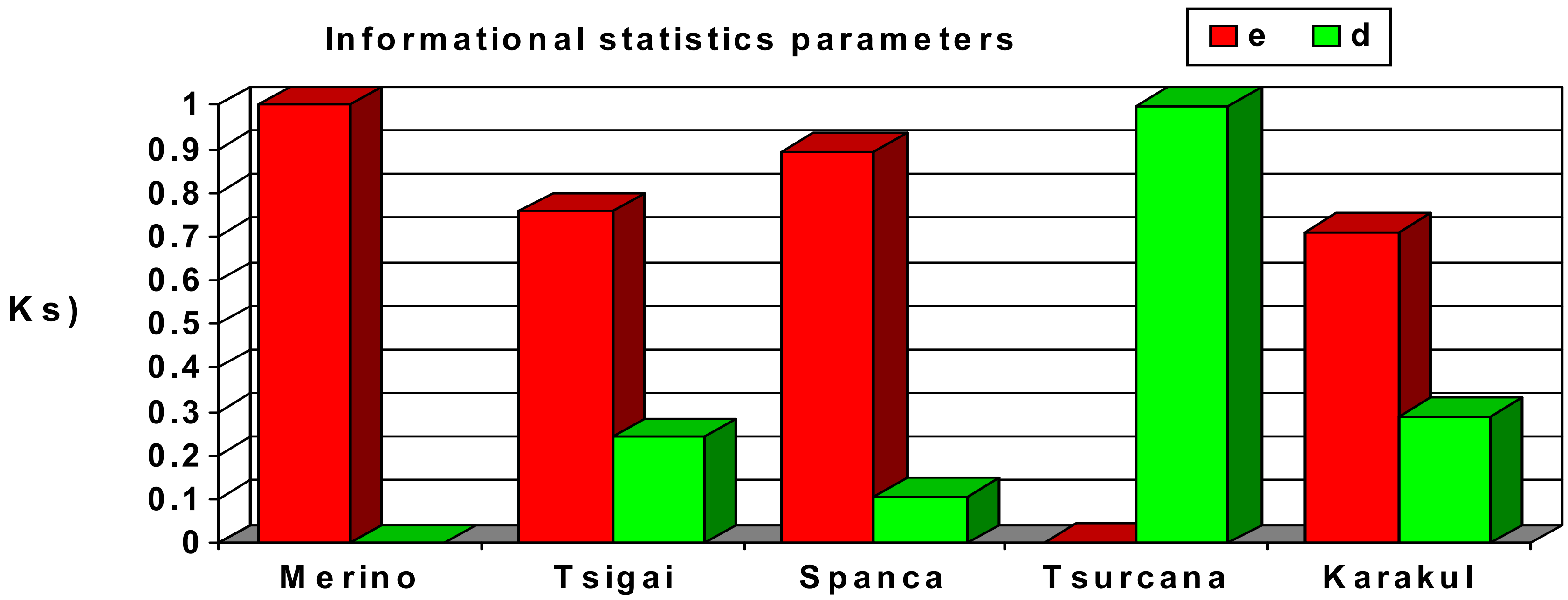

Coefficients of informational energy and (**e**) of genetic diversity (**d**) in the amylase system in the Romanian sheep breeds



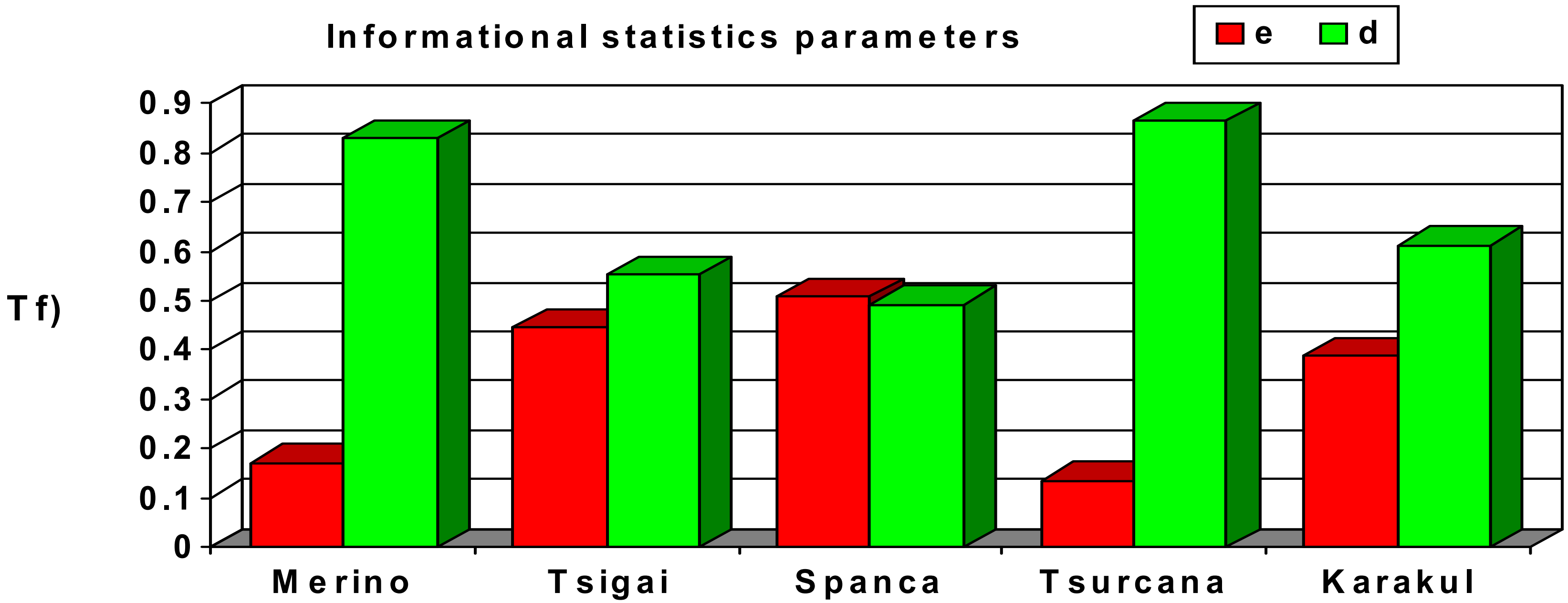

Coefficients of informational energy and (**e**) of genetic diversity (**d**) in the albumin system in the Romanian sheep breeds



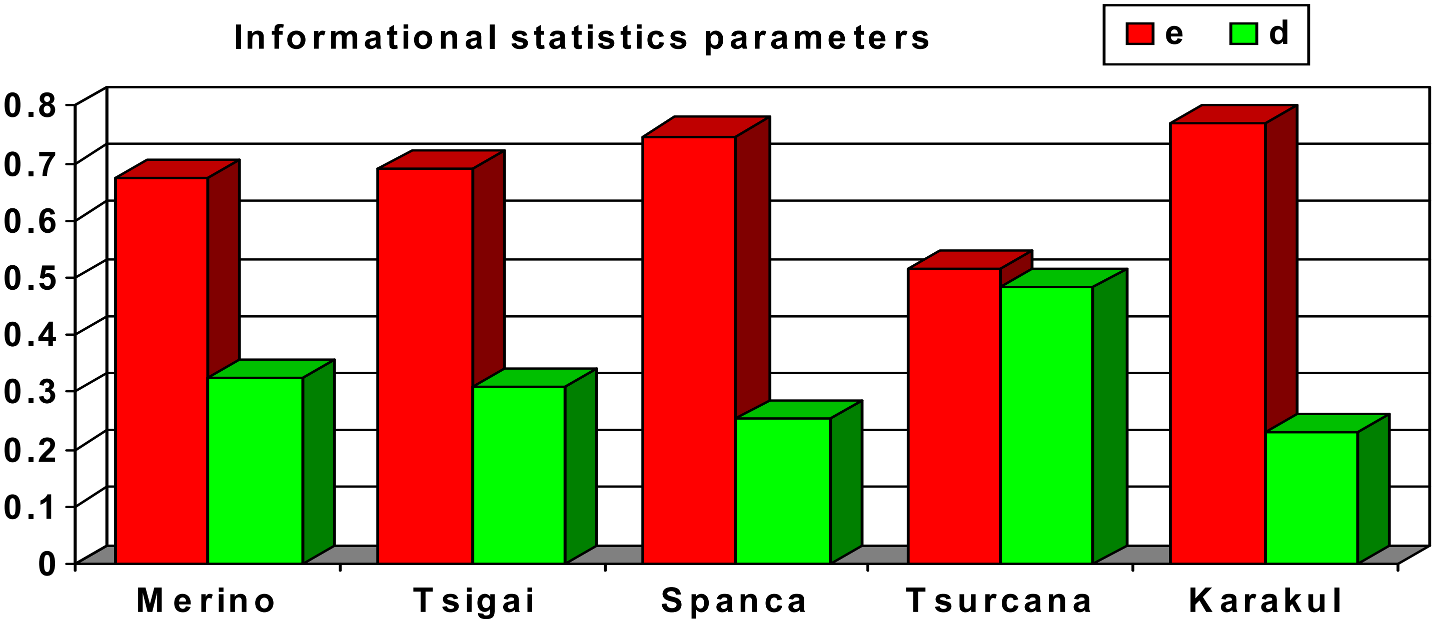

Coefficients of informational energy and (**e**) of genetic diversity (**d**) in the haemoglobin system in the Romanian sheep breeds




Coefficients of informational energy and (**e**) of genetic diversity (**d**) in the potassium system in the Romanian sheep breeds

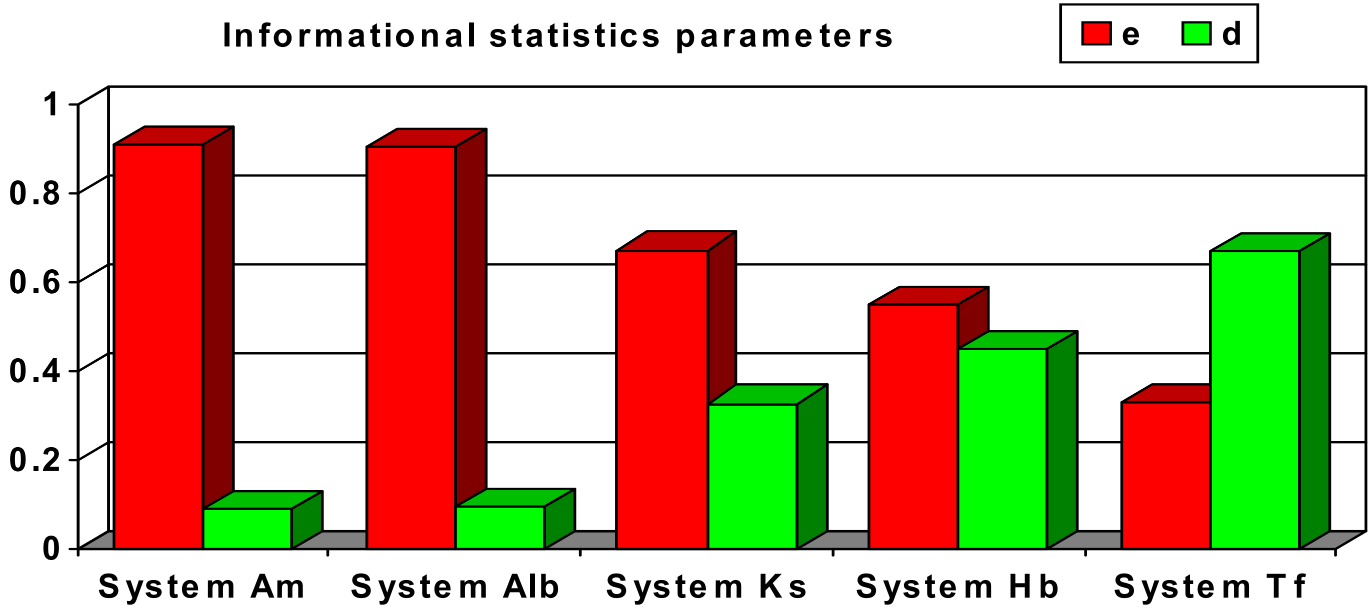



Coefficients of informational energy (**e**) and of genetic diversity (**d**) in the transferrin system in the Romanian sheep breeds



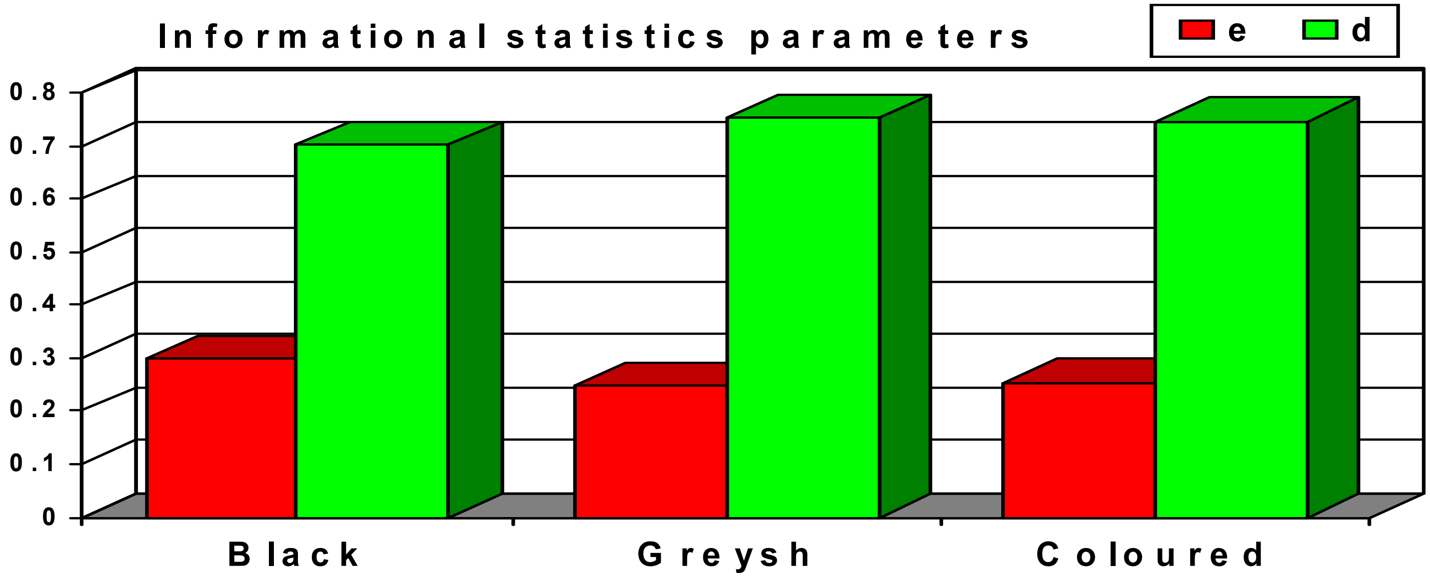

$e^{Karakul} > e^{Spanca} > e^{Tsigai} > e^{Merino} > e^{Tsurcana}$; $\Rightarrow$ $d^{Tsurcana} > d^{Merino} > d^{Tsigai} > d^{Spanca} > d^{Karakul}$

Coefficients of ***complex informational energy*** *(*$\mathbf{e_c}$*) and of* ***complex genetic diversity*** *(*$\mathbf{d_c}$*)* on the whole biochemical-genetic systems in the Romanian sheep breeds



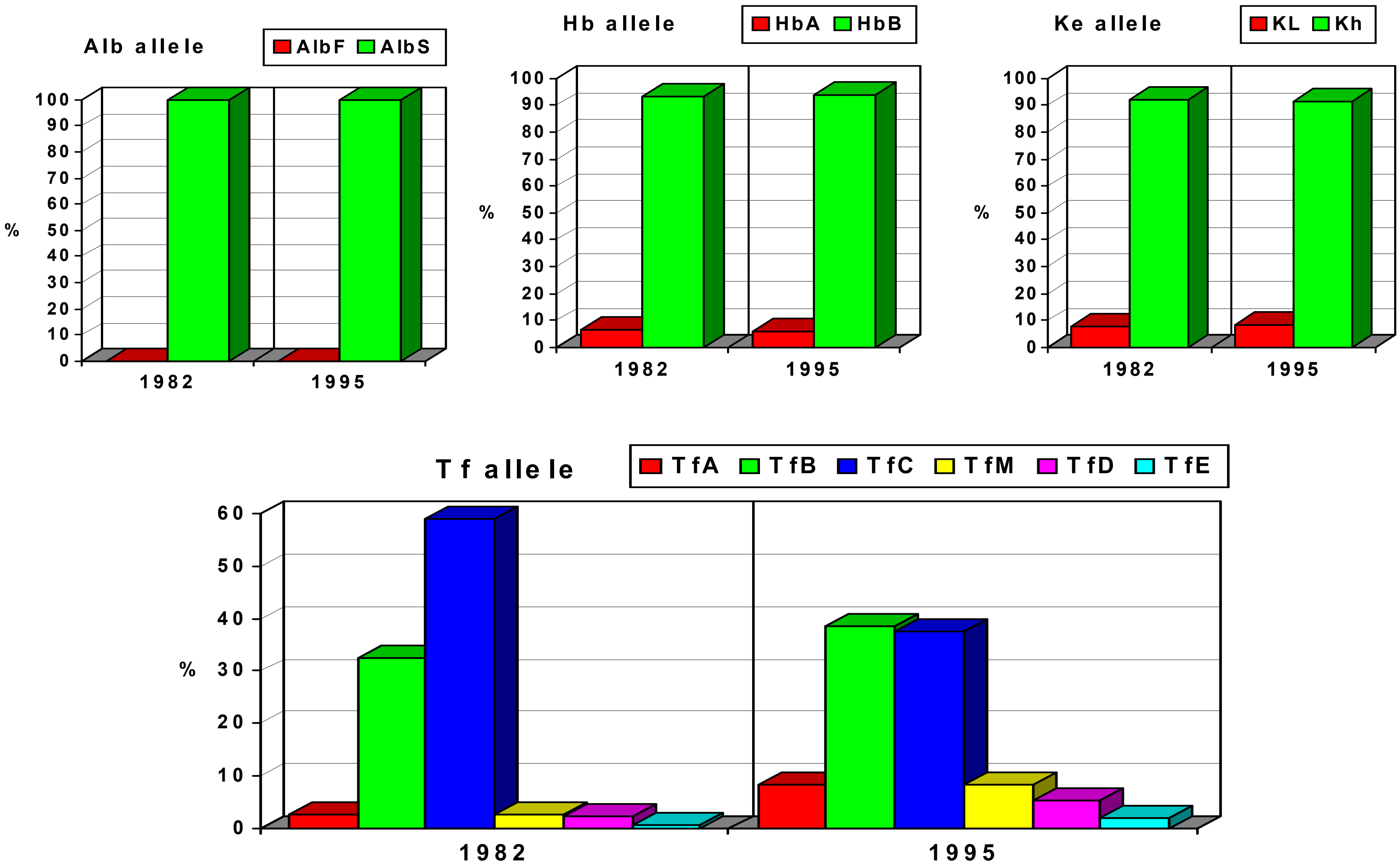

$e^{Am} > e^{Alb} > e^{Ks} > e^{Hb} > e^{Tf}; \quad \Rightarrow \quad d^{Tf} > d^{Hb} > d^{Ks} > d^{Alb} > d^{Am}$

Coefficients of ***complex informational energy*** ($\mathbf{e_c}$) and of ***complex genetic diversity*** ($\mathbf{d_c}$) on the whole Romanian sheep breeds in the biochemical-genetic systems

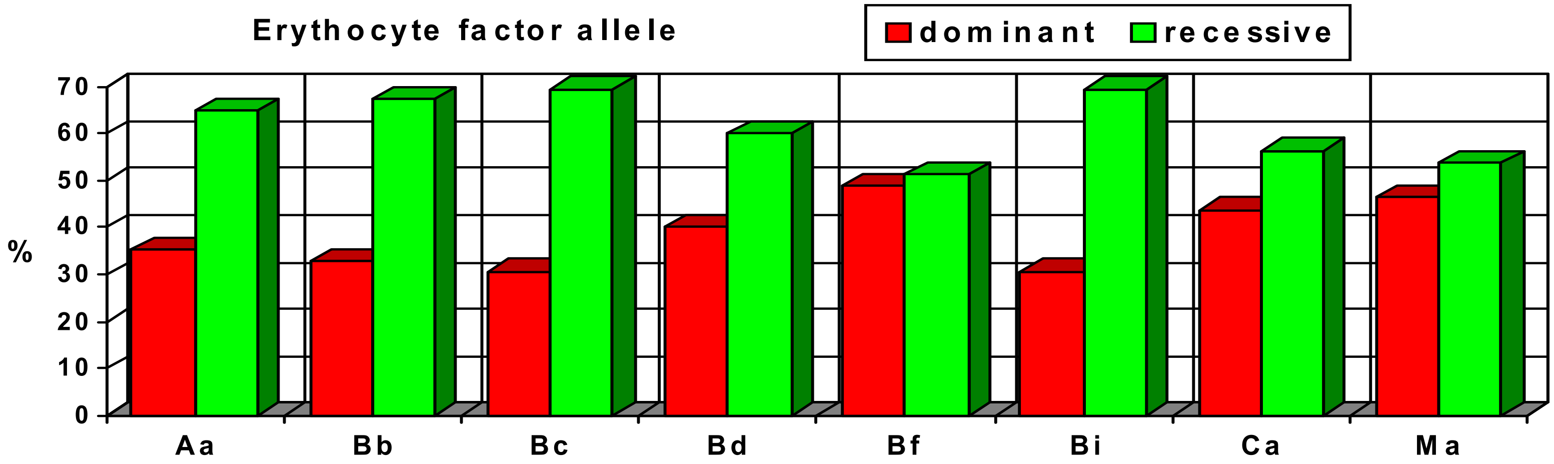


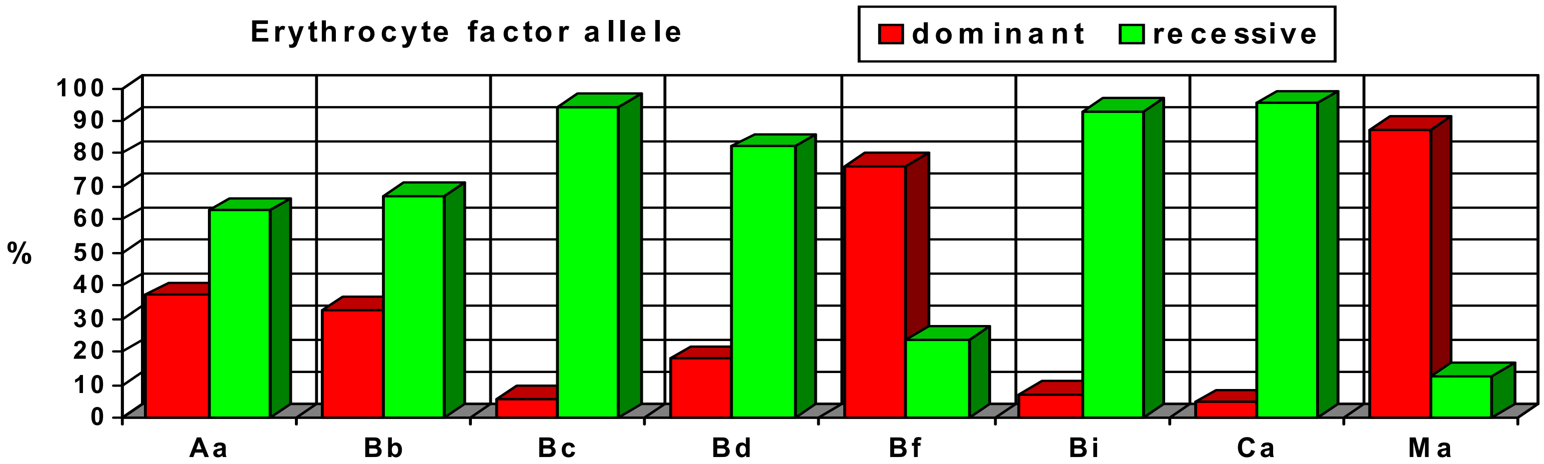

$$e^{\text{Black variety}} > e^{\text{Coloured varieties}} > e^{\text{Greysh variety}}; \Rightarrow d^{\text{Greysh variety}} > d^{\text{Coloured varieties}} > d^{\text{Black variety}}$$

Coefficients of ***complex informational energy*** *(*$\mathbf{e_c}$*) and of* ***complex genetic diversity*** *(*$\mathbf{d_c}$*)* on the whole biochemical-genetic and immunogenetic systems in the colour varieties of the Botosani Karakul breed

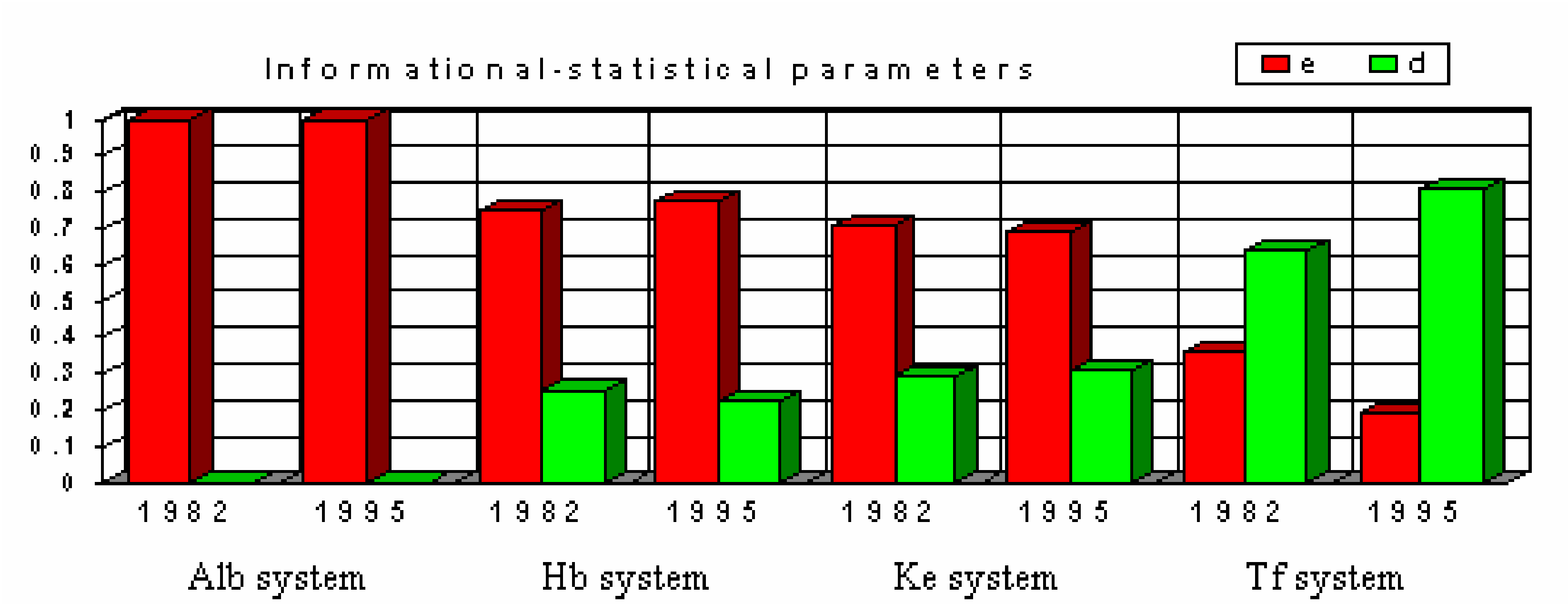



Allelic structure at the determinant loci of the biochemical-genetic systems in the Botoşani Karakul breed in two stages of its evolution



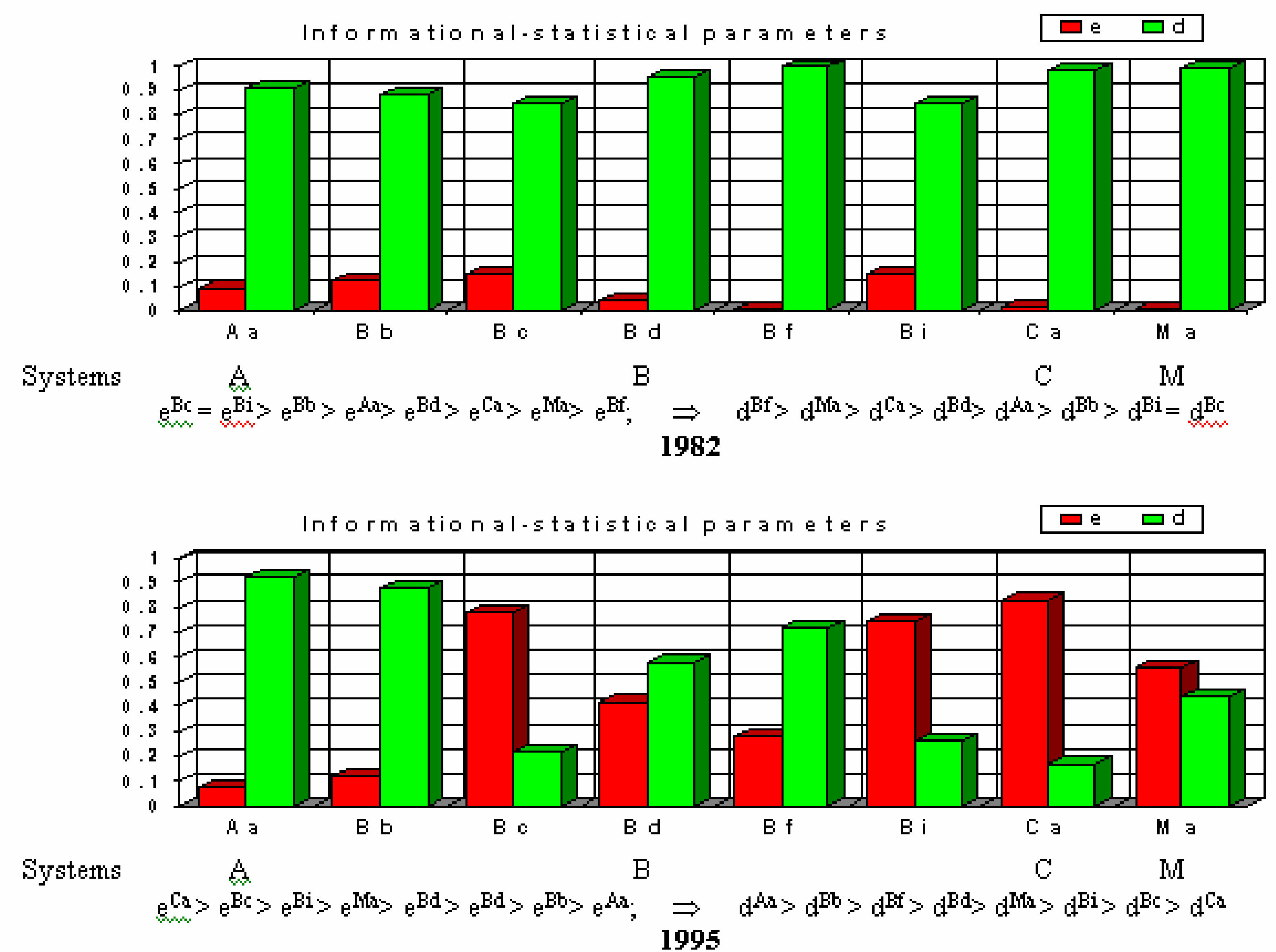




Allelic structure at the determinant loci of the immunogenetic systems in the Botoşani Karakul breed in two stages of its evolution

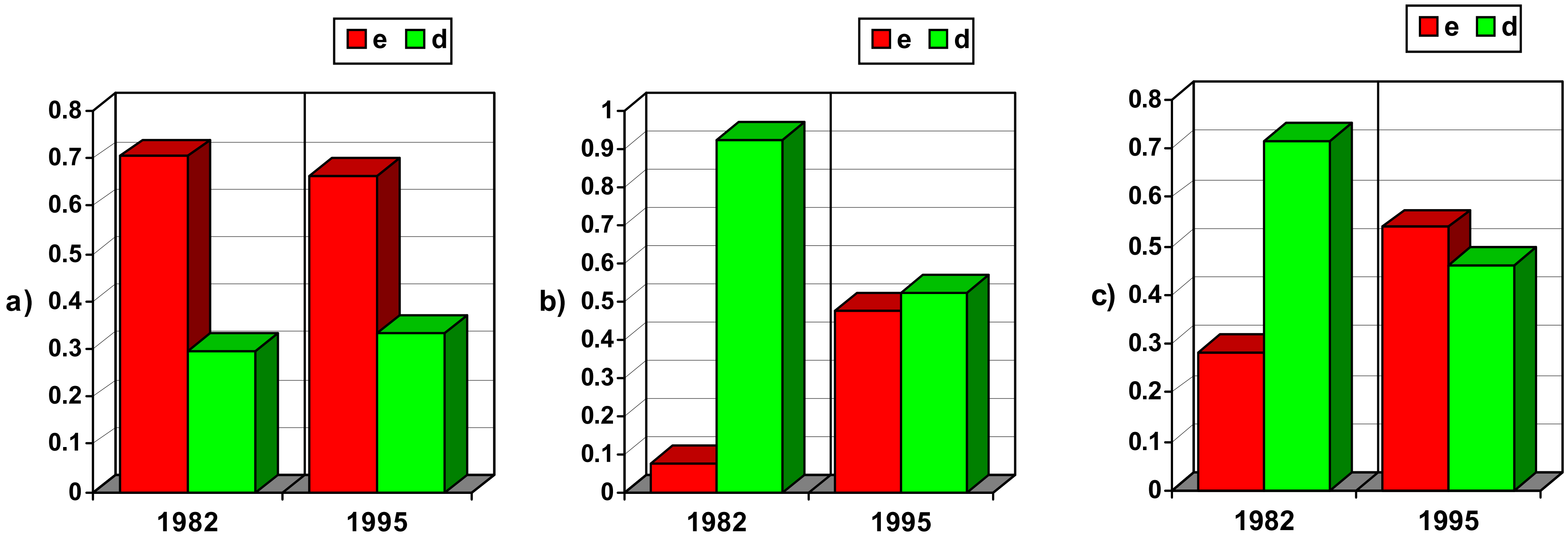



$$e^{Alb} > e^{Hb} > e^{Ks} > e^{Tf}; \quad \Rightarrow \quad d^{Tf} > d^{Ks} > d^{Hb} > d^{Alb}$$

Coefficients of ***complex informational energy*** *(*$\mathbf{e}_c$*) and of* ***complex genetic diversity*** *(*$\mathbf{d}_c$*)* on the biochemical-genetic systems in the Botoşani Karakul breed in two stages of its evolution




Coefficients of ***complex informational energy*** *(*$\mathbf{e_c}$*) and of* ***complex genetic diversity*** *(*$\mathbf{d_c}$*)* on the immunogenetic systems in the Botoşani Karakul breed in two stages of its evolution




Coefficients of ***complex informational energy*** (**e**$_c$) *and of* ***complex genetic diversity*** (**d**$_c$) on the on the whole biochemical-genetic systems ***(a),*** on the whole immunogenetic systems ***(b)*** and on the whole biochemical-genetic+immunogenetic systems ***(c)*** in the Botoşani Karakul breed in two stages of its evolution

## CONCLUSIONS

- 1.Certain parameters of informational statistics (*genetic distance, informational correlation, informational entropy, informational energy*) can be successfully used to definite the status of one biological system (*ecosystem, species, breed, variety etc*.) concerning *the biodiversity quantification* within them and *redimensioning the phylogeny or genealogy relations* among these entities, as well as *the dynamics* of these ecosystems in different stages of their evolution ; the calculus algorithm of this methodology uses *qualitative features with discontinuous variability*.
- 3.The authenticity of the size of the informational statistics parameters is proportional with the number of the polymorph systems taken in the calculus algorithm.
- 4.The biodiversity degree of different biological systems, as well as the intensity of relations among them shows the common or different particularities of the biotope conditions and of the improvement processes used in the breeding technologies that influence these biosystems .
- 5.The measurement of the informational-statistical parameters on the biochemical-genetic bases is useful in the process of preservation and improvement of the environment bioresources and of the genetic patrimony of the existing breeds or varieties, but also in the process of setting up of some new breeds or varieties, offering information on the compatibility of the cross-breeding systems among breeds or varieties which are genetically more or less far-off.