



APPLICATION OF THE AUTOMATED SYSTEM-COGNITIVE ANALYSIS FOR SOLVING PROBLEMS OF GENETICS

E.V. Lutsenko, L.P. Troshin*, A.S. Zviagin, A.V. Milovanov

Kuban State Agrarian University, Russia, Krasnodar, Kalinina 13

*Corresponding author email: ltroshin@yahoo.com

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 12 November 2017
Accepted 12 December 2017
Available online 1 January 2018

ABSTRACT

This article describes an approach that provides the calculation of the amount of information in the genes about the characteristics or properties of different grape cultivars. A specific phenotypic property is considered as a noisy genetic text that includes both genetic information about the true phenotypic property (pure signal). For this purpose, automated system-cognitive analysis (ASK-analysis) and its software tools - the intelligent Eidos system are used. A numerical example based on real data is presented, confirming the effectiveness of the proposed approach.

KEYWORDS

Automated system-cognitive analysis, "Aidos" system, grapevine varieties, DNA fragment.

1. INTRODUCTION

The notion that living beings are characterized by heredity developed in ancient times [1-3]. It was noted that when the organisms multiply from generation to generation, a complex of attributes and properties inherent in a particular species (manifestation of heredity) [4-6]. Due to heredity, species are preserved over significant intervals (up to hundreds of millions of years) of time [7-9]. Genes are certain DNA fragments; they are organized into chromosomes located in the nucleus of the cell [4, 8-9]. Thus, genes carry in themselves all the information about the signs and properties of the organism [10]. However, it is still not considered the theme of the amount of information contained in specific genes on specific phenotypic characteristics and properties of the organism.

As an exception to give two quotations from 1974 and 1980, which clearly recognized the role of information in genetic [11-12]. "The absolute criterion of the living is considered to be the presence of a certain number of bits of information contained in information molecules" [11, 13]. "... the rapid growth in the amount of information has given a new quality - the ability to reproduce itself (a quality inherent only in nucleic acids, possessing an incomparable with all other compounds of our world information capacity, the quality that gave rise to a new form of existence of matter)" [13]. And yet, there is no specific calculation of the amount of information in the genes in these works.

Apparently, the formulated problem is due to the lack of access for genetic researchers to software tools that implement the theory of information and ensure the solution of problems of genetics with the application of information theory. Meanwhile, such a toolkit exists and is in full open and free access: it is an automated system-cognitive analysis (ASK-analysis) [4, 8, 13], his mathematical model - the system information theory and implementing their software tools - the intellectual system "Eidos" [13,14].

Thus, there is a contradiction between the possibility of applying information theory in genetics and the actual absence of such an application. In this regard, the goal of the work is to present to the scientific community one of the possible solutions to this problem by

calculating the amount of information in the genes about the characteristics and properties of different varieties of grapes.

2. MATERIALS AND METHODS

To solve this problem, the following steps of ASK analysis were performed [15-16]:

- 1) cognitive-target structuring of the subject area;
- 2) formalization of the subject area, i.e. the development of classification and descriptive scales and gradations and training sample;
- 3) synthesis and verification of the information model reflecting the amount of information in the genes about phenotypic traits and properties (multiparameter typing);
- 4) information about the genetic system of determination of phenotypic characteristics and properties (SWOT-analysis of phenoscience);
- 5) output information on the strength and direction of the effect of a specific gene on phenotypic signs and properties (SWOT-gene diagrams);
- 6) solving the problem of systemic identification of phenotypic signs / properties by the presence of certain genes;
- 7) quantitative determination of similarity-differences of various phenotypic features / properties according to their system of determination by genes. A particular phenotypic property (or attribute) is considered as a noisy genetic text, including both genetic information about true phenoscience (pure signal) and noise that distorts this information due to accidental environmental influences. The software tool of ASK-analysis - intelligent system "Eidos" provides noise suppression and isolation of the true signal [17-19].

In our computational research data obtained from our and colleagues' previous research was used [26, 33]. Regarding to the fact that screenshots of proposed for analysis program were published already, we want to cite previous article and anyone who interested is able to look at it [18].

3. DATA, ANALYSIS, AND RESULTS

3.1 Cognitive-Target Structuring of the Subject Area

At this stage, the expert determines in an informal way on the basis of experience and professional competence what are the factors and what are the consequences (the effect of the factors) in the modelled subject area.

In our case, it is quite obvious that it is advisable to consider the number of microsattelites of different species as the factors, and how the results of their action are different phenotypic attributes or properties of different varieties of grapes.

Table 1 presents classification scales with which phenotypic features will be encoded, and in Table 2 descriptive scales, i.e. microsattelites on the 1st and 2nd spirals of DNA.

Table 1: Classification Scales

| Code | Name |
|------|-------------------------------------|
| 1 | FLOWER TYPE (151) |
| 2 | AVERAGE WEIGHT OF CLUSTER, G |
| 3 | YELD, C/HA |
| 4 | SUGARISTIC, G / 100 CUB. CM |
| 5 | ACID TITRATE, G/L |
| 6 | LEAVES DISSECTION (072) |
| 7 | LEAVES BLADDERY (075) |
| 8 | UPPER LEAFE SHAPES (094-1) |
| 9 | DOWN LEAFE SHAPES (094-2) |
| 10 | PUBESCENT OF LEAVES (084) |
| 11 | SIZE OF CLUSTER (202) |
| 12 | DENSITY OF THE CLUSTER (204) |
| 13 | BERRY SIZE (220) |
| 14 | FORM OF BERRY (223) |
| 15 | BERRY COLOR (225) |
| 16 | BERRY AROMA (236) |
| 17 | DIRECTION OF USE (603) |
| 18 | THE PERIOD OF MATURING (629) |
| 19 | DURATION OF PRODUCTION PERIOD, DAYS |
| 20 | SHOOT GROWTH, DAYS (351) |

Table 2: Descriptive scales

| Code | Name | Code | Name | Code | Name |
|------|-----------|------|-----------|------|------------|
| 1 | VRZAG47-2 | 18 | VVMD25-1 | 35 | VVIV37-2 |
| 2 | VRZAG47-1 | 19 | VVIP60-2 | 36 | VVIV37-1 |
| 3 | VVS2-2 | 20 | VVIP60-1 | 37 | VVIH54-2 |
| 4 | VVS2-1 | 21 | VVIB01-2 | 38 | VVIH54-1 |
| 5 | VVMD7-2 | 22 | VVIB01-1 | 39 | VVMD24-2 |
| 6 | VVMD7-1 | 23 | VRZAG83-2 | 40 | VVMD24-1 |
| 7 | VVMD5-2 | 24 | VRZAG83-1 | 41 | VVIV67-2 |
| 8 | VVMD5-1 | 25 | VVMD27-2 | 42 | VVIV67-1 |
| 9 | VRZAG62-2 | 26 | VVMD27-1 | 43 | VVIN73-2 |
| 10 | VRZAG62-1 | 27 | VVMD21-2 | 44 | VVIN73-1 |
| 11 | VRZAG79-2 | 28 | VVMD21-1 | 45 | VMC4F3.1-2 |
| 12 | VRZAG79-1 | 29 | VMC1B11-2 | 46 | VMC4F3.1-1 |
| 13 | VVMD28-2 | 30 | VMC1B11-1 | 47 | VVIN16-2 |
| 14 | VVMD28-1 | 31 | VVIQ52-2 | 48 | VVIN16-1 |
| 15 | VVMD32-2 | 32 | VVIQ52-1 | 49 | VVIP31-2 |
| 16 | VVMD32-1 | 33 | VRZAG67-2 | 50 | VVIP31-1 |
| 17 | VVMD25-2 | 34 | VRZAG67-1 | | |

The subsequent stages of ASK-analysis are automated in its toolkit – the intellectual system "Eidos".

descriptive scales and gradations are developed and with their help the initial data are coded, resulting in the formation of a training sample (Table 3).

3.2 Formalization of the subject domain, i.e. development of classification and descriptive scales and gradations and training sample

At the stage of formalization of the subject area, classification and

Table 3: Fragment of raw data

| Cultivar name | FLOWER TYPE (151) | AVERAGE WEIGHT OF CLUSTER, G | YELD, C/HA | SUGARISTIC, G / 100 CUB. CM | ACID TITRATE, G/L | LEAVES DISSECTION (072) | LEAVES BLADDERY (075) | UPPER LEAFE SHAPES (094-1) | DOWN LEAFE SHAPES (094-2) | PUBESCENT OF LEAVES (084) | SIZE OF CLUSTER (202) | DENSITY OF THE CLUSTER (204) | BERRY SIZE (220) | FORM OF BERRY (223) | BERRY COLOR (225) | BERRY AROMA (236) | DIRECTION OF USE (603) | THE PERIOD OF MATURING (629) | DURATION OF PRODUCTION PERIOD, DAYS | SHOOT GROWTH, DAYS (351) | VRZAG47-2 | VRZAG47-1 | VVS2-2 | VVS2-1 | VVMD7-2 | VVMD7-1 | VVMD5-2 | |
|---------------|-------------------|------------------------------|------------|-----------------------------|-------------------|-------------------------|-----------------------|----------------------------|---------------------------|---------------------------|-----------------------|------------------------------|------------------|---------------------|-------------------|-------------------|------------------------|------------------------------|-------------------------------------|--------------------------|-----------|-----------|--------|--------|---------|---------|---------|----|
| Adzhem misket | :4 | 40,0 | 25,0 | 21,5 | 3,0 | :7 | :3 | :7 | :1 | :3 | :3 | :3 | :5 | :4 | :1 | :2 | :1 | :6 | 157 | :5 | 182 | 188 | 143 | 145 | 9 | 23 | 23 | 24 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----------------------|----|-------|-------|------|------|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|-----|-----|-----|----|----|----|
| Ag izyum | :3 | 220,0 | 120,0 | 18,5 | 8,0 | :7 | :5 | :7 | :7 | :6 | :5 | :3 | :7 | :3 | :1 | :4 | :1 | :5 | 130 | :5 | 180 | 195 | 135 | 145 | 24 | 24 | 23 |
| Akkermanskij chernyj | :3 | 95,0 | 55,0 | 20,5 | 6,5 | :3 | :5 | :3 | :3 | :4 | :3 | :7 | :5 | :2 | :7 | :1 | :4 | :6 | 152 | :7 | 182 | 195 | 133 | 143 | 23 | 24 | 24 |
| Aklyk zheltyj | :3 | 180,0 | 35,0 | 17,5 | 6,5 | :3 | :5 | :3 | :1 | :3 | :5 | :5 | :7 | :4 | :1 | :1 | :3 | :5 | 142 | :5 | 195 | 195 | 143 | 145 | 23 | 24 | 23 |
| Alen'kij | :4 | 60,0 | 30,0 | 21,9 | 10,7 | :5 | :5 | :5 | :5 | :3 | :5 | :3 | :3 | :3 | :7 | :1 | :4 | :5 | 143 | :5 | 180 | 195 | 133 | 135 | 23 | 24 | 22 |
| Alyj terskij | :3 | 225,0 | 175,0 | 18,0 | 6,7 | :7 | :9 | :7 | :7 | :5 | :5 | :3 | :5 | :3 | :6 | :1 | :4 | :7 | 146 | :7 | 182 | 195 | 135 | 143 | 24 | 24 | 23 |
| Amet adzhibram | :3 | 110,0 | 65,0 | 17,5 | 5,5 | :3 | :5 | :3 | :1 | :8 | :5 | :5 | :5 | :3 | :1 | :1 | :4 | :6 | 152 | :5 | 180 | 182 | 143 | 143 | 23 | 23 | 23 |
| Asma | :3 | 400,0 | 95,0 | 17,1 | 7,6 | :1 | :5 | :5 | :3 | :1 | :7 | :7 | :7 | :7 | :7 | :1 | :1 | :8 | 160 | :7 | 186 | 195 | 149 | 151 | 24 | 24 | 23 |
| Bayat kapy | :3 | 312,0 | 148,0 | 17,9 | 10 | :5 | :5 | :5 | :5 | :9 | :7 | :9 | :5 | :3 | :6 | :8 | :4 | :5 | 141 | :7 | 182 | 182 | 135 | 145 | 23 | 24 | 23 |
| Bogos zerva | :3 | 135,0 | 60,0 | 19,0 | 5,5 | :5 | :4 | :7 | :5 | :8 | :5 | :7 | :3 | :3 | :1 | :1 | :4 | :5 | 151 | :5 | 182 | 182 | 145 | 149 | 23 | 23 | 24 |
| Bor kara | :3 | 286,0 | 150,0 | 16,0 | 5,5 | :7 | :5 | :7 | :7 | :7 | :7 | :5 | :5 | :3 | :6 | :1 | :4 | :7 | 144 | :7 | 176 | 186 | 141 | 143 | 23 | 25 | 23 |
| Boryu izyum | :3 | 225,0 | 170,0 | 14,5 | 4,5 | :5 | :1 | :5 | :5 | :3 | :5 | :5 | :5 | :3 | :1 | :1 | :1 | :5 | 143 | :5 | 190 | 195 | 135 | 137 | 23 | 24 | 23 |
| Bruskovaten'kij | :4 | 100,0 | 17,5 | 22,1 | 8,5 | :5 | :5 | :5 | :3 | :1 | :3 | :3 | :3 | :3 | :7 | :1 | :4 | :5 | 140 | :5 | 180 | 182 | 129 | 143 | 23 | 24 | 23 |
| Budaj shuli | :3 | 325,0 | 275,0 | 17,5 | 7,4 | :7 | :5 | :7 | :7 | :7 | :7 | :5 | :7 | :7 | :1 | :4 | :3 | :7 | 147 | :6 | 180 | 186 | 141 | 143 | 23 | 24 | 23 |
| Bulanyj belyj | :3 | 125,0 | 80,0 | 20,2 | 7,2 | :5 | :5 | :5 | :3 | :1 | :5 | :5 | :3 | :2 | :1 | :1 | :3 | :5 | 140 | :7 | 180 | 184 | 133 | 145 | 23 | 23 | 24 |
| CHaush chaban | :3 | 250,0 | 100,0 | 17,0 | 5,0 | :7 | :5 | :7 | :7 | :2 | :6 | :5 | :6 | :3 | :1 | :4 | :1 | :6 | 150 | :7 | 180 | 190 | 135 | 143 | 24 | 24 | 23 |
| Chol ber | :4 | 277,0 | 151,0 | 15,0 | 6,9 | :3 | :4 | :3 | :3 | :1 | :7 | :5 | :7 | :3 | :6 | :1 | :3 | :7 | 150 | :5 | 180 | 182 | 133 | 135 | 23 | 24 | 23 |
| Chorkuca rozoveh | :3 | 240,0 | 125,0 | 17,0 | 7,0 | :7 | :4 | :7 | :7 | :8 | :5 | :5 | :3 | :3 | :2 | :4 | :4 | :6 | 152 | :7 | 182 | 195 | 133 | 135 | 23 | 23 | 23 |
| Demir kara | :3 | 165,0 | 65,0 | 20,5 | 7,2 | :7 | :4 | :7 | :3 | :3 | :5 | :7 | :5 | :2 | :7 | :7 | :4 | :5 | 151 | :5 | 180 | 188 | 135 | 145 | 23 | 25 | 23 |
| Dokur | :3 | 304,0 | 70,0 | 16,0 | 5,5 | :7 | :1 | :7 | :7 | :1 | :7 | :5 | :6 | :4 | :1 | :1 | :1 | :7 | 150 | :5 | 186 | 186 | 135 | 151 | 24 | 24 | 23 |
| Dubut | :3 | 325,0 | 150,0 | 17,4 | 6,0 | :3 | :4 | :5 | :3 | :8 | :7 | :5 | :7 | :3 | :1 | :4 | :1 | :7 | 143 | :7 | 182 | 195 | 135 | 145 | 23 | 25 | 23 |
| Dzhagar | :3 | 308,0 | 249,0 | 17,0 | 6,2 | :7 | :4 | :7 | :7 | :3 | :7 | :5 | :7 | :3 | :2 | :1 | :3 | :7 | 148 | :7 | 182 | 186 | 135 | 143 | 24 | 25 | 23 |
| Dzhevat kara | :3 | 172,0 | 120,0 | 19,2 | 4,2 | :5 | :7 | :5 | :3 | :8 | :5 | :5 | :5 | :3 | :6 | :1 | :4 | :6 | 153 | :5 | 180 | 184 | 143 | 145 | 23 | 24 | 23 |

The names of grape varieties are given in the first column of the table. The columns from the 2nd to 21-th classification scale given by a gradation which describe the severity of symptoms and phenotypic properties. Moreover, in the names of the scales, and in the names of their gradations, a generally accepted by international system for coding phenotypic attributes and properties are used (highlighted in a light blue background) [20]. In columns from 22-th to 71-th given here microsatellite genes. In the Table 1 columns with text and numeric data types are shown [13].

Microsatellites are selected for study for two reasons:

- firstly, because we had data about them;
- secondly, because it is traditionally acknowledged that they are not associated with phenotypic traits and properties and have a scientific actuality [18].

In the internal calculator of the "Eidos" system, we set 5 gradations in numerical classification and descriptive scales. As a result of the program interface, we obtain classification scales, gradations, descriptive scales and a training sample or eventological database. The training sample is the raw data (Table 3), coded using classification and descriptive scales and

gradations. After formalizing the domain, we start mode 3.5, which provides the synthesis and verification of models with parameters, resulting in private models: ABS, PRC#, INF#. Models ABS, PRC #, INF #, differ from each other by particular knowledge criteria [18, 21] (Table 4).

Table 4: Particular criteria of knowledge currently used in ASC analysis and the "Eidos-X ++" system

| The name of the knowledge model and the particular criterion | Expression for a particular criterion | |
|--|--|---|
| | through the relative frequencies | through the absolute frequencies |
| INF1 , a particular criterion: the amount of knowledge according to A. Kharkevich, the first version of the calculation of relative frequencies: N_j - the total number of attributes for the j -th class. The relative frequency of the fact that if an object of the j th class has a feature, then this is the i -th attribute | $I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$ |
| INF2 , a particular criterion: the amount of knowledge according to A. Kharkevich, the second version of the calculation of relative frequencies: N_j - the total number of objects in the j -th class. The relative frequency of the fact that if an object of the j -th class is presented, then it will have an i -th attribute. | $I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$ |
| INF3 , a particular criterion: Chi-square: the differences between the actual and theoretically expected absolute frequencies. | --- | $I_{ij} = N_{ij} - \frac{N_i N_j}{N}$ |
| INF4 , a particular criterion: ROI - Return On Investment, the first version of the calculation of relative frequencies: N_j - the total number of signs for the j -th class. | $I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$ | $I_{ij} = \frac{N_{ij}N}{N_i N_j} - 1$ |
| INF5 , a particular criterion: ROI - Return On Investment, 2nd variant of calculation of relative frequencies: N_j - total number of objects in j -th class. | $I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$ | $I_{ij} = \frac{N_{ij}N}{N_i N_j} - 1$ |
| INF6 , a particular criterion: the difference between the conditional and unconditional relative frequencies, the first variant of calculating the relative frequencies: N_j - the total number of attributes for the j -th class. | $I_{ij} = P_{ij} - P_i$ | $I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$ |
| INF7 , a particular criterion: the difference between the conditional and unconditional relative frequencies, the second variant of calculating the relative frequencies: N_j - the total number of objects in the j -th class. | $I_{ij} = P_{ij} - P_i$ | $I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$ |

Notation:

i - previous parameter value;

j - future parameter value;

N_{ij} - the number of meetings of the j -th value of the future parameter at the i -th value of the past parameter;

M - total number of values of all past parameters;

W - total number of values of all future parameters;

N_i - the number of occurrences of the i -th value of the past parameter across the entire sample;

N_j - the number of meetings of the j -th value of the future parameter for the whole sample;

N - the number of meetings of the j -th value of the future parameter at the i -th value of the past parameter for the whole sample;

I_{ij} - a particular criterion of knowledge: the amount of knowledge in the fact of observing the i -th value of the past parameter that the object will go into a state corresponding to the j -th value of the future parameter;

Ψ - normalization factor (E.V. Lutsenko, 2002), which converts the number in the formula A. Harkevicha information in bits and providing for her compliance with the principle of conformity with the formula of R. Hartley;

P_i - unconditional relative frequency of meeting the i -th value of the past parameter in the training sample;

P_{ij} – conditional relative meeting frequency of the i -th value of the past parameter at the j -th value of the future parameter.

All these metrization methods using 7 particular knowledge criteria are implemented in the system-cognitive analysis and the intellectual system "Eidos" and provide a comparison of the gradations of all types of scales of numerical values that have the meaning of the amount of information in the gradation about the belonging of the object to the class. Therefore, it is correct to use integral criteria, including operations of multiplication and summation, to process numeric values corresponding to gradations of scales. This allows for the uniform and comparable processing of empirical data obtained with the help of any type of scale, using all mathematical operations.

The frequency distributions of similarity levels for true and false positive and negative solutions in the most reliable INF3 model are shown in the 4.1.3.11 screen form. From this form, we see that even at levels of similarity above 18%, the entire number of true-positive solutions far exceeds the number of false-positive solutions (which are not observed at similarity levels), and true-negative solutions are generally observed at all levels of difference, i.e. false-negative decisions there is very little. To evaluate the reliability of models, the proposed by Prof. E.V. Lutsenko metric, similar to the well-known F-criterion of Van Riesbergen [17].

3.3 Solving the problem of systemic identification of phenotypic traits and properties by the presence of certain genes

The purpose of this task is to use the knowledge of the strength and direction of the influence of various genome structures (the degree of DNA genes expression) reflected in the models on the degree of expression of various phenotypic traits and properties by the description of the genome structures to determine what phenotypic characteristics and properties will be possessed by grapes with such a genome.

In ASK-analysis and the "Eidos" system this problem is solved in two ways:

- 1) considered that the object of the more relates to a certain class, the more information about belonging to it is contained in its characteristics;
- 2) considered that the object of the more relates to a certain class, the more its profile in terms of the degree of expression of characteristics is similar to the profile of the class, which reflects the degree of characterization of these characteristics for the class.

The first method corresponds to the application of the integral criterion: "The sum of knowledge", and the second – "Resonance of knowledge" (explanations are given below).

Consider the integral knowledge criteria currently used in the ASK analysis and the "Eidos-X ++" system for model verification and identification and prediction problems, and then we present some output forms with results of recognition.

3.4 The integral criterion is the "Sum of Knowledge"

The 1st integral criterion "Sum of knowledge" is the total amount of knowledge contained in a system of factors of different nature that characterize the control object itself, control factors and the environment, the transition of the object to future target or undesirable states.

The integral criterion is an additive function of the particular criteria of knowledge and has the form:

$$I_j = (\bar{I}_{ij}, \bar{L}_i). \quad (1)$$

The expression in parentheses denotes the scalar product. In the coordinate form this expression has the form:

$$I_j = \sum_{i=1}^M I_{ij} L_i, \quad (2)$$

where: M – the number of gradations of descriptive scales (attributes);

\bar{I}_{ij} = $\{I_{ij}\}$ – state vector of the j -th class;

\bar{L}_i = $\{L_i\}$ – the state vector of the recognized object, including all kinds of factors characterizing the object itself, control actions and the environment (array-locator), i.e.:

$$L_i = \begin{cases} 1, & \text{if } i\text{-th factor works,} \\ n, & \text{where } n > 0, \text{ if } i\text{-th factor works with } n\text{-intensity;} \\ 0, & \text{if } i\text{-th factor doesn't works.} \end{cases} \quad (3)$$

In the current version of the Eidos-X ++ system, the values of the coordinates of the state vector of the recognized object were assumed to be equal to either 0, if there is no attribute, or n if it is present for an object with an intensity of n , i.e. is represented n times (for example, the letter "o" in the word "milk" is presented 3 times, and the letter "m" – once).

3.5 The integral criterion is the "Resonance of knowledge"

The 2-nd integral criterion "Semantic resonance of knowledge" is a normalized total amount of knowledge contained in a system of factors of different nature that characterize the control object itself, control factors and the environment, the transition of the object to future target or undesirable states.

The integral criterion is an additive function of the particular criteria of knowledge and has the form:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j)(L_i - \bar{L}), \quad (4)$$

where:

M – the number of gradations of descriptive scales (attributes);

\bar{I}_j – average informativity for the vector of the class;

\bar{L} – mean over the object vector;

σ_j – mean-of-square deviation of the particular criteria of knowledge of the vector of the class;

σ_l – mean-of-square deviation along the vector of the recognized object.

The above expression for the integral criterion "Semantic Resonance of Knowledge" is obtained directly from the expression for the "Sum of Knowledge" criterion after replacing the coordinates of the multiplied vectors with their standardized values:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}. \quad (5)$$

Its name is the integral criterion of similarity "Semantic resonance of knowledge" was received because in its mathematical form is the correlation of two vectors: the state of the j-th class and the state of the recognized object.

3.6 Automated SWOT-analysis of the system of determination of phenotypic characteristics and properties of the genome structure

On the screen modes of the Eidos Mode 4.4.8 the user can select any phenotypic property or feature and model in which to display the system of determination of this phenotypic property / attribute by the genome structure and the degree of expression of the various DNA genes (23, 24, 25).

3.7 Automated SWOT-analysis of the strength and direction of the effect of a specific gene on the phenotypic properties/attributes

In the screen form user can select any phenotypic property and include a filter on it. Then the screen forms will reflect the effect of a given degree of expression of a given gene on this phenotypic property: Thus, using automated cognitive SWOT analysis of genes and phenotypic properties/attributes, we can obtain information about which genes a given phenotypic property/characteristic is determined and vice versa, what phenotypic attributes/properties are linked with a given gene. Possession of this knowledge gives hope for the possibility of obtaining an answer to the question of what a genome should be, so that the variety possesses the desired phenotypic properties.

However, there is at least one problem here that is that some phenotypic properties/attributes are linked with the presence or severity of certain genes, while others are either completely absent or weakly expressed. It is clear that simultaneously the gene cannot both be present and absent, be both strong and weakly expressed. It follows that some phenotypic properties/signs cannot be simultaneously obtained in one class, while others are quite possible. For example, a berry cannot be sweet, and sour, and dark and light. Perhaps this is the justification for the existence of many varieties. To find out exactly which properties are compatible in one class and which are not, it is necessary to compare different phenotypic properties according to their gene determination system.

3.8 Quantitative determination of similarity and differences of various phenotypic properties according to their system of determination by genes. Compatible and incompatible phenotypic properties

This problem is solved in mode 4.2.2 of the "Eidos" system, in which class similarity matrices are first calculated, and then they are simply visualized in the form of cognitive diagrams or, in addition, cognitive cluster analysis (Lutsenko, 2011). The results of cluster analysis of classes are shown on the tabular screen form of mode 4.2.2.2 and in the graphical form of the cognitive diagram. The diagram shows a construct with two clusters at the poles. At one pole, a cluster centred in the class is code: 12 "The yield is very high" and classes 98, 81, 66 and 2, and at the other pole we see a cluster with phenotypic signs: 37, 9, 92, 93, 10, 16.

Classes that are in the same cluster are determined by similar systems of genes and are attainable simultaneously in one class. Classes belonging to opposite clusters located at the poles of the construct are not attainable simultaneously in one class. For example, before plant breeders correctly set the task to obtain a high-yielding grape variety with the colouring of berries 225 (code: 81) and berries size 220 (code: 66), and a very massive bunch (code: 7), while this variety will have a long ripening period of 629 (code: 98). But to demand from breeders that these berries be sweet (code: 16) is incorrect, because this requires a completely different genome than predetermining the previously listed phenotypic properties. If we want to get a sweet variety, then, apparently, we have to reconcile with its small or at best average yield. Note that usually only experts in this subject area have such knowledge, and in this paper it is shown how they can be obtained with the help of modern automated intellectual technologies (ASK-analysis and the Eidos system) directly on the basis of empirical data.

4. DISCUSSION

Traditionally it is believed that certain genes cause the corresponding phenotypic properties and these properties cannot be determined by other genes (7). But our research shows that, apparently, it is not so. This is evidenced by the fact of the similarity between genes in their effect on the phenotype, because of what they combine into clusters, and the opposite clusters form poles of constructs. In one cluster, there are genes that have a similar effect on the phenotypic properties, and on the poles of the construct – the opposite.

To obtain the corresponding output forms, we first form a matrix of similarity of genes by their effect, and then display it graphically. And from these output forms it is clear which genes have a similar, and which – opposite effect on the phenotype. If there was a one-to-one correspondence between genes and phenotypic characteristics/properties, there would not be a similarity between them and the difference in their effect on the phenotype, i.e. their system would be orthonormal. An algorithm for deciding the structure of the genome with the aim of creating a new variety with a given system of phenotypic properties/attributes.

The authors propose the following algorithm for deciding the structure of the genome in order to create a new variety with a given system of phenotypic properties/characteristics, which essentially coincides with the logic of the presentation in this paper (assuming that all stages of ASC analysis before the creation and verification of models have been successfully performed):

Step-1: 4.7. Quantitative determination of similarity and differences of various phenotypic properties according to their system of determination by genes. Choose the phenotypic properties compatible in one variety.

Step-2: 4.5. Carry out an automated SWOT-analysis of the determination system of each of the phenotypic properties chosen at the 1-st step of the genome structure. Select genes that probably determine the selected phenotypic properties.

Step-3: Exclude repeats of the genes selected at the 2-nd step.

Step-4: 4.4. Solve the problem of system identification of the variety according to the genes chosen at the 3-rd step. This variety and take as a basis for selection.

To perform the 4th step, we create another model in the same way as described above, but as classes we choose the names of the varieties, and as the signs – the genes. Authors has created this model. It shows one of the forms that reflects the results of a cluster-constructive analysis of varietal comparisons by their genome in this model. Cognitive functions are offered by prof. E.V. Lutsenko generalization of the classical concept of functional dependence on the basis of information theory in system fuzzy interval mathematics (14). The visualization module of cognitive functions was developed by Prof. E. V. Lutsenko and the developer of intelligent systems from Belarus D. K. Bandyk (25). In total, in each of the 7 knowledge models, similar 1000 cognitive functions are obtained, because in each knowledge model there are 20 classification scales and 50 descriptive scales, i.e. 1000 submatrices. These cognitive functions reflect the dependence of the degree of expression of various phenotypic properties/attributes of grapes on the degree of expression of certain genes.

5. CONCLUSION

The results obtained in the work have a world scientific novelty, i.e. previously were not received anywhere in the world. The technology of synthesis and application of an intelligent measuring system described in the article can be applied not only in genetic studies in ampelography, but also in other subject areas. Of course, this work is far from complete and only demonstrated the possibility of applying the theory of information and cognitive technologies in research in the field of genetics. In the

future, the authors plan to conduct a more extensive study on a much larger volume of initial data: more varieties of grapes, more phenotypic properties/attributes and more genes, not only microsatellite, but also directly influencing the phenotypic properties/attributes. The materials of this article can be used in conducting laboratory work on disciplines related to intellectual technologies, knowledge representation and artificial intelligence systems, as well as econometrics, biometrics, ecology, pedagogy, psychology, medicine, criminalistics, entomology, etc.

REFERENCES

- [1] Gehring, C., Flores-Rentería, D., Sthultz, C.M., Leonard, T.M., Flores-Rentería, L., Whipple, A.V., Whitham, T.G. 2014. Plant genetics and interspecific competitive interactions determine ectomycorrhizal fungal community responses to climate change. *Molecular ecology*, 23 (6), 1379–1391.
- [2] Ronald, P.C. 2014. Lab to farm: applying research on plant genetics and genomics to crop improvement. *PLoS biology*, 12 (6), e1001878.
- [3] Waldron, D. 2015. Plant genetics: A clever TrAP circumvents host defence. *Nature Reviews Genetics*, 16 (12), 686.
- [4] Lee, I. 2014. A showcase of future plant biology: moving towards next-generation plant genetics assisted by genome sequencing and systems biology. *Genome biology*, 15 (5), 305.
- [5] Leyser, O., Sedwick, C., Leyser, O. 2014. The beauty of plant genetics. Interviewed by Caitlin Sedwick.
- [6] Lokody, I. 2014. Plant genomics: The wild side of soybeans. *Nature Reviews Genetics*, 15 (11), 704.
- [7] Andersen, R. 2013. *Governing agrobiodiversity: plant genetics and developing countries*, Ashgate Publishing, Ltd.
- [8] Schweitzer, J.A., Madritch, M.D., Felker-Quinn, E., Bailey, J.K. 2013. From genes to ecosystems: plant genetics as a link between above- and belowground processes. *Soil Ecology and Ecosystem Services*, 82–97.
- [9] Prunet, N., Meyerowitz, E.M. 2016. Genetics and plant development. *Comptes rendus biologies*, 339 (7), 240–246.
- [10] Lutsenko, E.V. 2012. Investigation of the influence of subsystems of different levels of the hierarchy on the emergent properties of the system as a whole with the use of ASK-analysis and the intellectual system "Eidos" (microstructure of the system as a factor in controlling its macroproperties). *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU)*, 01 (075), 638 – 680.
- [11] Lehninger, A. 1974. *Biochemistry*. Mir, Moscow. (in Russ.).
- [12] Khilchevskaya, R.I. 1980. The role of asymmetry-symmetry of matter in the processes of the origin of life on the Earth. In Zh-L VHO im. D.I. Mendeleev, XXV4, Publishing House. "Chemistry", 418–424.
- [13] Lutsenko, E.V. 2014. Quantitative automated SWOT- and PEST-analysis by means of ASK-analysis and intellectual system "Eidos-X ++". *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU) (Electronic resource)*, 07 (101), 1367–1409.
- [14] Orlov, A.I., Lutsenko, E.V. 2014. *System fuzzy interval mathematics*, Monograph (scientific publication). Krasnodar: KubSAU, 600.
- [15] Lutsenko, E.V. 2002. *Automated system-cognitive analysis in the management of active objects (system theory of information and its application in the study of economic, socio-psychological, technological and organizational-technical systems): Monograph (scientific publication)*, Krasnodar: KubSAU, 605.
- [16] Duval, R. 2006. A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics*, 61 (1), 103–131.
- [17] Lutsenko, E.V. 2017. Open scalable interactive intellectual on-line environment for teaching and research based on ASC-analysis and the Eidos system. *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU) (Electronic resource)*, 06 (130), 1 – 55.
- [18] Lutsenko, E.V. 2016. Synthesis of adaptive intelligent measuring systems using ASC-analysis and the Eidos system and system identification in econometrics, biometrics, ecology, pedagogy, psychology and medicine. *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU)*, 02 (116), 1 – 60.
- [19] Lutsenko, E.V. 2017. Problems and perspectives of the theory and methodology of scientific knowledge and automated system-cognitive analysis as an automated method of scientific cognition, providing meaningful phenomenological modeling. *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU) (Electronic resource)*, 03 (127), 1 – 60.
- [20] Troshin, L.P., Magradze, D.N. 2013. *Ampelographic screening of the grape gene pool (Educational visual aid)*. Krasnodar, 120.
- [21] Lutsenko, E.V. 2013. Metrization of measuring scales of various types and joint comparable quantitative processing of heterogeneous factors in system-cognitive analysis and the "Eidos" system". *Polytematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal KubSAU)*, 08 (092), 859 – 883.

