SciencePG
Science Publishing Group

Algorithms for control of genetic-breeding improvement of economically valuable traits of self-pollinated plants

Mikhailenko I. M.[1], Dragavtsev V. A.[2]

[1]Scientific deputy director, head of lab of infornatic-measuring systems, Agrophysical institute of RAAS, Saint-Petersburg, 195220, Grazhdansky prospect 14, Russia

[2]Main scientist of plant ecologic physiology lab, Agrophysical institute of RAAS, Saint-Petersburg, 195220, Grazhdansky prospect 14, Russia

Email address:

ilya.mihailenko@yandex.ru (Mikhailenko I. M.), dravial@mail.ru (Dragavtsev V. A.)

Abstract: The new approaches to solving problems of selection parent's pairs (varieties) for crossing and forecast of eco-genetic portrait of the future new variety. Paper supplements the two previous publications on the formalization of the quantitative theory of eco-genetic processes [1,2].

Keywords: Eco- Genetic Portrait, Genetic- Physiological Systems, Selection, Mathematical Model of "Genotype - Environment Interaction", Management by Environmental Factors

1. Introduction

From quantitative genetics is known [3] that the phenotypic variability of any quantitative trait (it is visible by the naked eye in a segregating population) - is:

$$v_{ph} = v_g + v_e \qquad (1)$$

where: v_{ph} - the phenotypic variance, v_g - genotypic variance, v_e - environmental variance caused by variations of micro-environmental conditions for individual plants of populations.

Breeder desire to know the genotypic variance of productivity trait, because selection of the best genotypic deviations and their subsequent cloning (for example potatoes or fruit trees) leads to increase productivity, improvement of stability and quality of new clones. Now it is known the next principles for rapid estimation of genotypical variance (without raising progeny):

1. The principle of artificial leveling background of growing a segregating population [4].

2. The principle of measurement standards (comparison of phenotypic variation in wild or segregating population (v_{ph}) with the variability in clones or pure lines, the obtained from this population (v_e) [3].

3. Shrikhande principle [5,6,7].

4. The principle of background characters [8, 9, 10, 11, 12, 13], which from four principles showed oneself the most accurate and reliable. [14].

For grain self-pollinated crops can evaluate genotypic variance v_g by calculation of the variance of character's averages on the plots with different varieties in collection of any breeding center.

During the calculation of the average characteristics of the variety in the plot all the noises eliminated, and the average meaning in the plot - is genotypic value of the character, and the variance of averages of plot is v_g. However, the crops do not propagate by clones or by grafting, so to predict possible genetic improvement of crop productivity and yield the breeder should know the additive genetic variance v_a.

It is known [3] that

$$v_g = v_a + v_d + v_i \qquad (2)$$

where v_a - the additive variance (caused by a variety of additive effects of polygenes), v_d - the dominant variance caused by the effects of dominance and Vi - the variance of epistatic effects.

Until recently in quantitative genetics was only one principle for estimation of additive variance - on the correlation of "parent-offspring"- $cov_{op} = 1/2\, v_a$, or on the correlation

of relatives (half-sibs) $\text{cov}_{hs} = 1/4\, v_a$ [15].

This principle requires the change of generations, so it cannot work in the populations of F_2. Until now, breeders visually select the best phenotypes in the F_2, but not the best genotypes, and certainly not the best additive genotypes (transgressions). This selection has a very low efficiency. Litun writes: "The solvability of the modern technological schemes of selections - 0.01% [16]. Using the principle of background characters whether we brought the efficiency of selection of barley up to 15%" (that is increased the efficiency of identification of individual genotypes in 1500 times).

In 1979 was created an express method of evaluation of additive variance (without raising progeny) for quantitative traits using similyarity their responses in environmental gradient [17].

In 1998 was created an express method of evaluation for each of the seven genetic- physiological systems that contribute to the productivity of individuals and the yield per unit area phytocenosis [18, pp. 33 - 35]. This method formed the basis of mathematical models of "genotype-environment interaction" and algorithms of identification of genotype by phenotype, which were presented in [1,2].

These tasks are the stages of the general problem - control genetic-breeding process. In this paper we consider the following problem: the selection of the parental couples to ensure the desired result of crossing, and the prediction result of crossing parental pairs (Fig. 1).
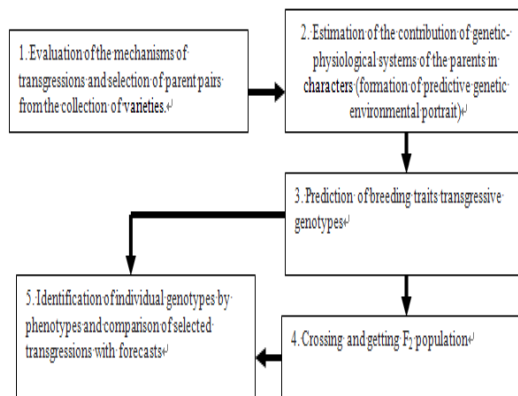

Fig. 1. *Block diagram of the interaction of management tasks by genetic-breeding process.*

Main body
Eco-genetic portraits

The starting point for these tasks is to set the desired combination of breeding characters (BC), which must be obtained in the second (segregated) progeny of the results of cross-breeding. Information and algorithmic foundation of these problems is the concept of "eco-genetic portrait", formulated in [2]. It represents a combination of positive feedback, or shifts BC derived from the contributions of 7 genetic-physiological systems in a module of final products. This combination can be represented as a vector of these responses $\Delta F^T = [\Delta\varphi_1, \Delta\varphi_2, \Delta\varphi_3, \Delta\varphi_4, \Delta\varphi_5, \Delta\varphi_6, \Delta\varphi_7]$, or a
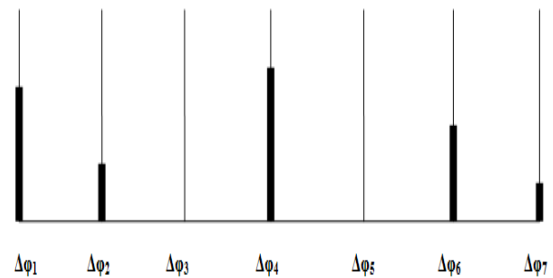
graphical diagram (Fig. 2).

Fig. 2. *Eco-genetic portrait of the variety (genotype) on the output module of crop.*

In fact, such eco-genetic portrait is more complex, has several levels, and includes joint ventures and changes in other modules of quantitative traits. However, most breeders efforts aimed at getting the desired combination of BC is the levels of productivity and yield. In addition, this portrait is incomplete, as in the other modules, prior to the end, not shown the effect of a 7-genetic-physiological systems. In particular the systems of attractions and micro - distribution appear only in the last stages of ontogenesis in the final product.

$\Delta\varphi_1$-contribution of attractions, $\Delta\varphi_2$ - contribution of the micro-distribution, $\Delta\varphi_3$ - contribution systems adaptation (resilience to climate protection and chemical stressors), $\Delta\varphi_4$ - polygenic contribution of immunity to the agents of plant diseases, the development of plant defense mechanisms, $\Delta\varphi_5$-contribution system of "payment" limiting the productivity of soil nutrition elements, $\Delta\varphi_6$ - contribution of tolerance to density, $\Delta\varphi_7$ - contribution variability periods of ontogenesis (the possibility of a selection "withdrawal" critical phase of ontogenesis from hits of the stressor).

The general scheme of solving task selection **the optimal parental pair**

We start from demonstration of the initial situation in which the problem is solved. At the disposal of the breeder has a set BC of the final product (output module) - X*(T), which must be obtained from a cross, and there is data on the dynamics of the limiting factors for the place of cultivation of hybrid - F (t). In addition, the breeder has the original base variety or hybrid, the BC of which must be improved, as well as a database of mathematical models of potential parents, by which for a given environmental conditions and using technology can be predicted outcome - $\widehat{X}(T)$, where T - the time since the end of vegetation.

The basic variety was obtained as a result of previous acts of breeding, and for it, except the mathematical model of "genotype-environment", at the disposal of the breeder has its eco-genetic portrait - $\Delta F0T = [\Delta\varphi10, \Delta\varphi20, \Delta\varphi30, \Delta\varphi40, \Delta\varphi50, \Delta\varphi60, \Delta\varphi70]$ obtained for optimal growing conditions. Availability of models and eco-genetic portrait give possibility to predict the final result and estimate the contribution of all genetic-physiological systems. Therefore, modelling its growth and development, we can at a negative effects of environmental factors to assess negative

changes or failures in all the required joint venture BC $-\Delta\hat{X}_k(T)$ and in addition to forecast eco-genetic portrait for the given growth conditions $\Delta\hat{\Phi}_k(T)$. Then, using the principle of estimation of the additive genetic-physiological systems [18], we fill the "holes" in the eco-genetic portrait, and we have forecast eco-genetic portrait of top transgressions in generation F2 - $\Delta\hat{\Phi}_{k+1}(T)$, entering this information into a mathematical model of the "genotype-environment" possible to predict the expected results for a given BC- $\hat{X}_k(T)$ .By comparing it with the required values X*(T), we can decide to stop the process and go to the next possible crossings. Consistently going through all the available options for the crossing, we'll do the final selection on the option that provides the greatest proximity to the required values BC - X*(T).

We reviewed the general scheme of selection of parental pairs for the final module of products. However, often the problem is complicated by the need to select a substantial impact of the genetic-physiological systems, and the state of some of intermediate modules, such as the module, "the number of grains per plant," multiplied by the "weight of one grain," as well "grain weight per plant". Result expressed on this module determines the productivity of individual.

Algorithm for the selection of optimal parental pair

Let us consider in more detail the algorithm of optimal selection of parental pairs for grain self-pollinated crops in considering module grain productivity. Assume that in Fig. 2 shows the eco-genetic portrait of the initial genotypes (variety), which we want to improve resistance to stress and to "pay" by productivity the lim-factor of soil nutrition. As can be seen in the eco-genetic portrait of these indicators are the "holes" because the responses appropriate genetic-physiological systems are zero - $\Delta\varphi_3 = 0$, $\Delta\varphi_5 = 0$.

From the breeder's data bank of potential parents we choose the option for eco-genetic portrait which these "holes" are filled (see Fig. 3).
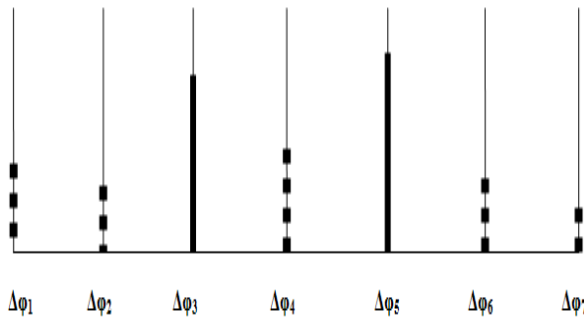

Fig. 3. *Eco-genetic portrait of the variety (genotype) with the required BC for adaptability and "payment" of lim-factor of soil nutrition (dotted line shows the contributions of other genetic- physiological systems).*

In case if the genetic-physiological systems are additive, we have the right to combine eco-genetic portraits of parents, and get the forecast eco-genetic portrait of the future new variety:
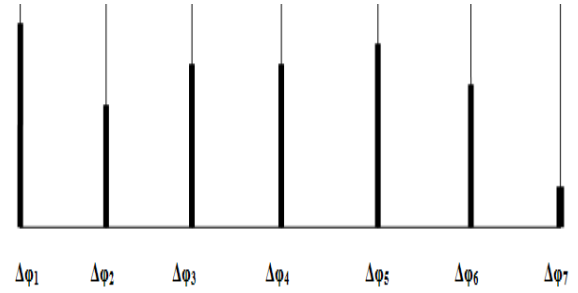

Fig. 4. *Predicted eco-genetic portrait of variety.*

We introduce responses of genetic-physiological systems. They are components of the predicted eco-genetic portrait in the output module of the mathematical model of "genotype-environment" [2]:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} a_{11}(\varphi_3) & a_{12}(\varphi_2) & a_{13}(\varphi_1) \\ a_{21}(\varphi_2) & a_{33}(\varphi_3) & 0 \\ a_{31}(\varphi_1) & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_1(\varphi_5) \\ b_2 \\ b_3 \end{bmatrix} [u(t)] +$$

$$+ \begin{bmatrix} 0 & c_{12}(\varphi_3) & c_{13}(\varphi_3) \\ 0 & 0 & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix} + \begin{bmatrix} d_{11} & d_{12} \\ 0 & 0 \\ d_{31} & d_{32} \end{bmatrix} \begin{bmatrix} \varphi_4(t) \\ \varphi_6(t) \end{bmatrix} + \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \xi_3(t) \end{bmatrix}, \quad (3)$$

$$t \in [t_0(\varphi_7); T(\varphi_7)]$$

where the following notation: x_1 - grain weight per ear i-th individual, x_2 - mass of chaff in the ear, x_3 - straw weight, u- control of nitrogen nutrition; f_1 - luminous efficiency factor, f_2 - temperature factor productivity, f_3 - moisture, as a factor of productivity; $\varphi_1 ... \varphi_7$ - influence of genetic-physiological systems; ξ_1, ξ_2, ξ_3 - random disturbances, reflecting the informational uncertainty of the model; a_{kj}, b_k, c_kj, d_{kj} - the dynamic parameters of the model.

Model (3) is more convenient to consider a more compact vector-matrix form

$$\dot{X} = A(\Phi)X(t) + bu(t) + CF(t) + D\Phi(t) + \xi(t), (4)$$

where all variables and parameters are combined in the respective vectors and matrices.

The resulting model BC - $X_j(T)$, we will compare with the required values X*(T), then we need a stopping criterion procedures

$$KR_J = [X^*(t) - X_j(t)]^T [X^*(t) - X_j(t)] \leq \delta, \quad (5)$$

where δ-specified threshold number it's above leads to the need to continue process of selection of the parental pair by selecting from the database following options for the crossing.

In the case when breeder has no of eco-genetic portraits of potential parents, use a static version of the model "ecological disturbance - the reaction of genetic-physiological systems" of the parents, as indicated in our previous paper [2]

$$\Delta\varphi = W^T \Delta\tilde{E}. \qquad (6)$$

Management of environmental factors in process of breeding

The idea of controlling the genetic-breeding process is based on the theory of eco-genetic organization of quantitative traits and simulation of the system interaction of "genotype-environment." According to them it follows, that environmental factors often make a significant contribution to the joint venture BC of generations, as the level of productivity of any character is the result of the interaction of "genotype-environment". Therefore, in terms of the breeding process in the phytotrons, the most important factor in increasing the reliability is management by environmental factors of breeding. This allows to make the breeding process from a passive procedure of consistent comparison of parental pairs, in active controlled process of quick search the best options while making maximum use of their genetic potential. Furthermore, the presence of control by environmental factors will significantly improve the identification of genotypes to phenotypes, and thus to speed up the release of new varieties.

Combine model output module (4) with the state model of genetic-physiological systems (3):

$$\dot{X} = A(\Phi)X(t) + bu(t) + C\Pi(t) + D\Phi(t) + \xi(t),$$

$$\dot{\Phi} = A_\varphi \Phi + B\Pi(t), \qquad (7)$$

and to get managed the system of environmental factors, where P(t) - vector of parameters control environmental factors.

Now the task of achieving the desired BC in a selectable generation looks as follows

$$KR^* =$$
$$\arg\min_{j,\Pi(t)} [X^*(t) - X_j(t)]^T [X^*(t) - X_j(t)] \qquad (8)$$

and consists in finding the best pair for crossing and optimization of environmental effects in the breeding process.

Note that in the presence of control by environmental factors of breeding, response of the genetic-physiological systems are not analyzed separately, but are only intermediate variables search procedure, which is formed when stopping forecast eco-genetic portrait and the corresponding forecast values BC.

The algorithm is implemented entirely on a virtual level, but nature is carried only a single option for the crossing, which resulted from a population of F2 then select a single genotype that meets the given requirements (in case of the improvements of commercial variety). In this realized a significant acceleration of the breeding process.

2. Conclusions

The algorithm controls by the breeding process based on the mathematical model of "genotype-environment" interaction, is a logical choice of variants of complementary eco-genetic portraits of potential parents and the resulting quantitative prediction of breeding characters. At the same time the choice of options is carried out to the required degree of agreement between predicted and experimental breeding characters. In the presence of control environmental factors such an option is supplemented by optimizing parameters of the environment, which ensures the maximum use of eco-genetic potential of the parents.

References

[1] Mikhailenko I.M., Dragavtsev V.A. Basic principles of mathematical modelling of the "genotype-environment" interaction, // Agricultural Biology, 2010, № 3, p. 24-31 (rus).

[2] Mikhailenko I.M., Dragavtsev V.A. Mathematical modelling in plant breeding. I. Theoretical basis of genotypes identification on their phenotypes during selection in segregating generations. //Agricultural biology, 2013, №1, p. 26-34, (rus).

[3] Falconer D.S. Introduction to Quantitative Genetics, 1961, Oliver-Boyd, Edinburg.

[4] Matveev, N.D. On one of the ways to improve the initial stages of breeding process / / Genetica, 1966, № 4, p. 158-165. (rus).

[5] Shrikhande V.J. Some considerations in designing experiments on coconut trees. / / J. Indian Soc. Agric. Stat., 1957, v. 9, p. 82-91.

[6] Sakai K.I., Hatakeyama S., Estimation of genetic parameters in forest trees without raising progeny, / / Silvae Genetica, 1963, v. 12, No 5, p. 152-160.

[7] Sakai K.I., Mukaide H. Estimation of genetic, environmental and competitional variances in standing forests. / / Silvae Genetica, 1967, v. 16, No 5-6, p. 159-168.

[8] Dragavtsev V.A. Philosophical questions of the theory of morphogenesis and theory of the information / / Successes of modern biology, 1961, vol 52, no. 3 (6), p. 347-361 (rus).

[9] Dragavtsev V.A. Phenogenetic analysis of the variability in plant populations, / / Vestnik AN Kaz. SSR, 1963, № 10 (223), p. 33-42 (rus).

[10] Dragavtsev V.A. The method of estimation of the role of heredity and environment in the development traits of woody plants, which does not require the change of generations, / / Botanical Journal, 1966, v. 51, № 7, p. 939-946 (rus).

[11] Dragavtsev V.A. Ostrikova V.M. Search for background characters express evaluation of genetic variation in plant populations / / Genetica, 1972, Vol 8, № 4, p. 33-37 (rus).

[12] Dragavtsev V.A. Pogozhev I.B., Sokolova T.A. Quantitative estimates of the critical values of genotypic features of the plants with the environmental distribution of the deviation from the phenotypes / / Proc. "Models of ecosystems and methods of determining their parameters." Calculate. Center of the Academy of Sciences, Novosibirsk, 1981, p. 190-116 (rus).

[13] Dragavtsev V.A. Foundation for future high-tech of breeding for improving of polygenic economically important characters of plants. / / Journal "Agrarian Russia", 2008, № 4, p. 2-10. (rus).

[14] Gronin V.V. Using quantitative and morphometric characters for distinctness parental lines and hybrids. / / Thesises of diss. candidate. biol. sciences (breeding and seeds production), Krasnodar, Institute of Rice, 2007 (rus).

[15] Mather Wharton B. Principles of Quantitative Genetics, 1964, Burgess Publishing, USA.

[16] Litun P.P. Resolution of the current breeding schemes selection, Report to the 4th All-Union Congress of VOGiS, February 2,1982, Kishinev (rus).

[17] Dragavtsev V.A. Averyanova A.F., On the correlation between the level of the additive variance and the degree of reaction similyarity of quantitative traits of wheat, / / Genetica, 1979, v. 15, № 3, p. 518-526 (rus).

[18] Dragavtsev V.A., Eco-genetic screening of the gene pool and methods of creating varieties of agricultural plants on productivity, resistance and quality. / / Guidelines (new approaches), VIR, St. Petersburg, 1998, 50 p. (pp. 33-35) (rus).