

MATHEMATICAL MODELLING IN PLANT BREEDING. II. ALGORITHMS FOR CONTROL OF GENETIC-BREEDING IMPROVEMENT OF ECONOMICALLY VALUABLE PROPERTIES IN SELF-POLLINATORS

I.M. Mikhailenko, V.A. Dragavtsev

Agrophysical Research Institute, Russian Academy of Agricultural Sciences,
14, Grazhdanskii prosp., St. Petersburg, 196220 Russia,
e-mail: ilya.mikhailenko@yandex.ru

Received January 11, 2012

Summary

The authors considered the brand new approaches to choose the breeding pairs (varieties) for crossing and prediction of eco-genetic picture of new variety. The algorithm of control of breeding process on the basis of mathematical model for «genotype—environment» interaction was proposed, which consists in successive choose the variants of mutually complementary eco-genetic character of potential parents and prediction of resultant quantitative breeding determinants. The choose of the variants is performed until a required degree of coincidence between predictable and specified breeding determinants occurs. Under the control of ecological factors such choose is supplementing by optimization of environment parameters, and it provides the maximal usage of eco-genetic potential and the increase of reliability of identification of progeny genotypes on their phenotypes in the sequel.

Keywords: mathematical modeling, plant breeding technologies, self-pollinators.

Quantitative genetics (1) describes phenotypic diversity of any quantitative trait manifested in a splitting population as:

$$v_{ph} = v_g + v_e \quad [1],$$

where v_{ph} — phenotypic variance, v_g — genotypic variance, and v_e — ecological variance caused by differences of environmental conditions for individual plants in population.

For a breeder it is important to know genotypic variance of a productive trait, because selection of best genotypic changes and further improvement of these features at vegetative propagation (eg., potatoes, fruit-berry crops) increases productivity, resistance and quality of resulting clones. Today, rapid estimate of productivity (without changes of generations) is performed in a splitting population grown in artificially leveled background conditions with the use of standards (comparison of phenotypic diversity in a wild or splitting population (v_{ph}) with diversity of clones or pure lines derived from this population (v_e) (1), as well as by Shrikhande's method (2-4) and the theory of background characters (5-10) that was found to be the most precise and reliable approach (11-13).

In grain self-pollinating crops, genotypic variance v_g can be easily assessed knowing the variance of mean values of traits manifested by cultivars grown on experimental plots in a breeding center (calculation of mean value of a trait over the plot eliminates all noises; the resulting mean corresponds to genotypic value of the trait, whose observed variance is v_g). However, cereal crops can't be propagated by cloning or grafting, so, for a breeder, any prognoses about genetic improvement of their productivity are possible only upon the data on additive genetic variance (v_a) of these crops.

It is known (1) that

$$v_g = v_a + v_d + v_i \quad [2],$$

where v_a — additive variance (caused by additively acting polygenes), v_d — genetic variance caused by domination effects, and v_i — variance provided by effects of epistasis.

Until recently, quantitative genetics could estimate v_a only by the correlation “parent – offspring” $cov_{op} = 1/2v_a$ or correlations of relatives (half-sibs) $cov_{hs} = 1/4v_a$ (12).

This approach requires the change of generations, so it doesn't work in F_2 populations. Breeders still select visually the best phenotypes from F_2 , though not the best genotypes and furthermore not the best additive genotypes (transgressions). In studies of P.P. Litun (13), the theory of background characters allowed to improve the efficiency of selection in barley up to 15% while the resolution capacity of modern technological schemes of selection is 0,01% (i.e. 1500 times more efficient identification of individual genotypes).

In 1979, it was established the method for express estimate (without the change of generations) of v_a of quantitative traits based on their similar responses in ecological gradient (14), in 1998 – the method for assessing v_a of any of the seven genetic-physiological systems that provide contributions to individual productivity and the yield from unit phytocenosis area. The second method became the basis of mathematical models “genotype-environment” and algorithms for identification of genotypes by phenotypes described in the earlier authors' works (15).

These tasks correspond to stages in solving the general task of controlled genetic-breeding process: 1) assessing mechanisms of transgressions and selecting parental couples from a collection of cultivars; 2) evaluating contributions of parental genetic-physiological systems to the improved trait; 3) forecasting of breeding traits of transgression genotypes; 4) crossing and obtaining F_2 population; 5) identification of individual genotypes on phenotypes and comparison of selected transgressions with forecasts.

This works is focused on solving the following tasks: selection of parental couples for a cross providing a desired result, as well as forecasting the result of such crossing. A start point in this work is determining the set of desired genetic-physiological systems manifested in F_2 (splitting) generation that would determine levels of breeding traits (BRT). Mathematical procedure of this task includes a so-called “eco-genetic portrait” described in detail previously (*Agrobiology*, 2013, 1, pp. 26-34). This is a set of positive responses, or changes of BRT, formed by contributions of the seven genetic-physiological systems in the output module of a final product. This set can be presented as a vector of these responses $\Delta\Phi^T = [\Delta\phi_1, \Delta\phi_2, \Delta\phi_3, \Delta\phi_4, \Delta\phi_5, \Delta\phi_6, \Delta\phi_7]$, or a graphical diagram (Fig. 1). In fact, such “eco-genetic portrait” is more complex, multilevel structure that includes additional changes of BRT in other modules of quantitative traits. At the same time, for a breeder, the most common target is a desired combination of BRT in yield and productivity. Such “eco-genetic portrait” is incomplete, because modules-precedes of the final one don't reflect the effect of the seven genetic-physiological systems. In particular, the effects of attraction and microdistribution systems can be observed only

at final stages of ontogenesis (formation of a final product).

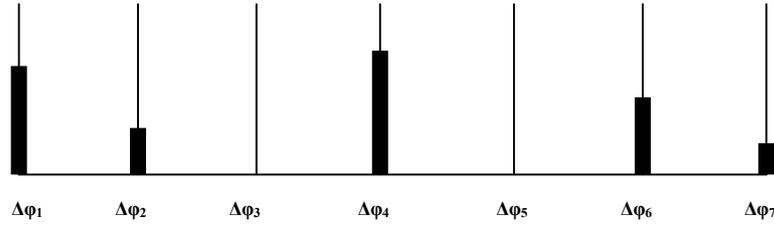


Fig. 1. Eco-genetic portrait of a cultivar (genotype) in the output module of a crop: contributions provided by genetic-physiological systems of attraction ($\Delta\varphi_1$), microdistribution ($\Delta\varphi_2$), adaptivity (resistance to climatic and chemical environmental stressors) ($\Delta\varphi_3$), polygenic immunity to diseases and pests, protective mechanisms developed by plants ($\Delta\varphi_4$), “payment” by productivity in response to limiting edaphic factors ($\Delta\varphi_5$), tolerance to thickening growth ($\Delta\varphi_6$), and diversity in duration of ontogeny periods (possibility to move a critical ontogeny stage away from the impact of stressor) ($\Delta\varphi_7$).

General scheme for solving the task of selection of an optimum parental couple. The initial informational situation of this task includes the data available for a breeder – the set of BRTs of a final product (output module) $X^*(T)$ i.e. the desired result of a cross, and the data about dynamics of all limiting factors in the region of growing the resulting hybrid $F(t)$. Along with it, a breeder has an initial base cultivar or hybrid whose BRTs need to be improved, as well as the databank of mathematical models of possible parents, which against the background of specified dynamics of ecological conditions and technology can be used to forecast a final result $\hat{X}(T)$, where T — time of the end of vegetation.

The base cultivar is the result of preliminary selection with known mathematical model “genotype-environment” and eco-genetic portrait $\Delta\Phi_0^T = [\Delta\varphi_{10}, \Delta\varphi_{20}, \Delta\varphi_{30}, \Delta\varphi_{40}, \Delta\varphi_{50}, \Delta\varphi_{60}, \Delta\varphi_{70}]$ for optimum growing conditions. Having this model and eco-genetic portrait, a breeder can predict a final result and evaluate contributions of all the seven genetic-physiological systems. That’s why modeling growth and development of a cultivar allows to assess negative changes or sharp declines in all specified BRTs knowing negative variations of ecological factors:

$$-\Delta\hat{X}_k(T)$$

and simultaneously determine a predicted eco-genetic portrait for specified growing conditions:

$$\Delta\hat{\Phi}_k(T).$$

Then, using the principle of express estimate of the additive action of genetic-physiological systems (14) it is possible to add missing elements to eco-genetic portrait and obtain a predicted eco-genetic portrait of best transgressions F_2 :

$$\Delta\hat{\Phi}_{k+1}(T),$$

Introducing it into the mathematical model “genotype-environment” allows forecasting a final result for the specified BRTs:

$$\hat{X}_k(T).$$

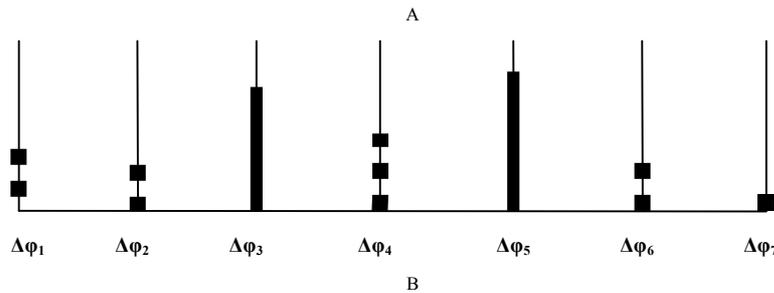
Comparing this result with desired values of $X^*(T)$, a breeder can make a decision to terminate this variant of crossing, or start a new possible variant. Successive calculations for all available variants of crosses is the way to choose the optimum variant ensuring the most close manifestation of desired BRT $X^*(T)$.

A general scheme for selection of parental couples was observed with respect to the module of a final product. However, this task is often complicated by need in considering the influence of genetic-physiological systems on condition of some intermediate modules, eg. the module: (“number of grains per plant (NGP)” · “kernel weight (KW)”= “weight of grains per plant (WGP)”). The output result expressed by this module defines productivity of an individual.

Algorithm for selection of the optimum parental couple. In this work, the subject of such algorithm is cereal self-pollinating crop with target traits of grain productivity and the module (NGP·KW=WGP).

Figure 1 presents eco-genetic portrait of an initial genotype (cultivar) that must be improved for stress resistance and “payment” by productivity to edaphic lim-factors; gaps on corresponding bars reflect zero responses of these genetic-physiological systems ($\Delta\varphi_3 = 0$, $\Delta\varphi_5 = 0$).

A breeder observes a databank of possible parents to select the variant whose eco-genetic portrait includes complete bars of the abovementioned traits (Fig. 2, A).



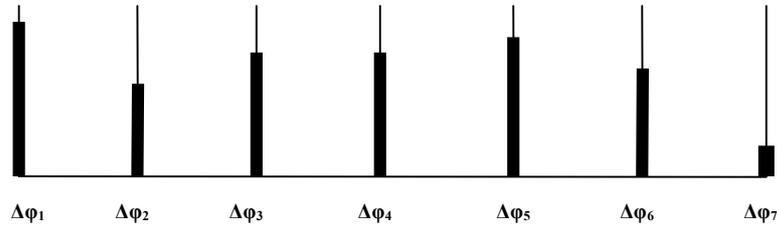


Fig. 2. Eco-genetic portrait of a cultivar (genotype) with desired selective traits of adaptivity and “payment” by productivity to edaphic lim-factor (bars with gaps – contributions of other genetic-physiological systems) (A) and predicted eco-genetic portrait of a cultivar (B), derived by its combination with initial cultivar (Fig. 1) under the condition of additive action of genetic-physiological systems.

In the case of additive action of genetic-physiological systems, parental eco-genetic portraits can be combined to form the predicted eco-genetic portrait of a new future variety (Fig. 2, B).

Introducing the responses of genetic-physiological systems as components of the forecasted eco-genetic portrait into the output module of the mathematical model “genotype-environment” (15) leads to:

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix}_j &= \begin{bmatrix} a_{11}(\varphi_3) & a_{12}(\varphi_2) & a_{13}(\varphi_1) \\ a_{21}(\varphi_2) & a_{33}(\varphi_3) & 0 \\ a_{31}(\varphi_1) & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}_j + \begin{bmatrix} b_1(\varphi_5) \\ b_2 \\ b_3 \end{bmatrix} [u(t)] + \\ &+ \begin{bmatrix} 0 & c_{12}(\varphi_3) & c_{13}(\varphi_3) \\ 0 & 0 & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & d_{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{34} & 0 & d_{36} & 0 \end{bmatrix} \begin{bmatrix} \Delta\varphi_1 \\ \Delta\varphi_2 \\ \Delta\varphi_3 \\ \Delta\varphi_4 \\ \Delta\varphi_5 \\ \Delta\varphi_6 \\ \Delta\varphi_7 \end{bmatrix} + \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \xi_3(t) \end{bmatrix}, \quad [3], \\ &t \in [t_0(\varphi_7); T(\varphi_7)] \end{aligned}$$

with the following notations: x_{1i} — grain weight per ear in i^{th} individual; x_{2i} — weight of chaff per ear; x_{3i} — weight of straw per ear; u — supply (control) of nitrogen nutrition; f_1 — luminous efficiency factor; f_2 — temperature factor of productivity; f_3 — moisture as a productivity factor; $\varphi_1 \dots \varphi_7$ — influence of genetic and physiological systems; ξ_1, ξ_2, ξ_3 — random disturbances reflecting informational uncertainty of the model; $a_{kj}, b_{ki}, c_{kj}, d_{kj}$ — dynamic parameters of the model.

This model can be represented in compact vector-matrix form (15):

$$\begin{aligned} \dot{X}_i &= A(\varphi_1, \varphi_2, \varphi_3)X(t) + b(\varphi_5)u(t) + C(\varphi_3)F(t) + D * [\varphi_4(t)\varphi_6(t)] + \xi(t), \\ &t \in [t_0(\varphi_7); T(\varphi_7)] \end{aligned} \quad [4],$$

in which all variables and parameters are combined in corresponding vectors and matrices.

BRTs obtained as result of modeling $X_j(T)$ then should be compared with desired $X^*(T)$ values, which necessitates introducing the parameter “termination of a procedure”:

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

where δ — specified threshold value exceeding which leads to continuation of the procedure by selecting a new parental couple from the databank as the new variant of crossing.

If a breeder has no available data about eco-genetic portraits of possible parents, then it can be used the statistical variant of parent’s models “ecological perturbation – response of genetic-physiological systems” described in the first authors’ report:

$$\Delta\varphi = W^T \Delta\tilde{E} \quad [6].$$

Role of controlled ecological factors in breeding. The idea of controlled genetic-breeding process is based on the theory of eco-genetic organization of quantitative traits and modeling of “genotype-environment” interaction. According to these bases, ecological factors provide a major contribution to formation of BRTs in generations, because expression of any productive trait is the result of interaction between genotype and environment. That’s why a control of ecological factors is one of the important conditions for obtaining reliable results of breeding a new variety in a phytotron. Along with it, the possibility to control ecological factors can greatly improve the efficiency of identification of genotypes on phenotypes and thereby speed up the creation of new varieties.

Combination of models of the output module [4] and the state of genetic-physiological systems (15):

$$\begin{aligned} \dot{X} &= A(\Phi)X(t) + bu(t) + C\Pi(t) + D\Phi(t) + \xi(t), \\ \dot{\Phi} &= A_\varphi \Phi + B\Pi(t) \end{aligned} \quad [7]$$

is the system with controlled ecological factors where $\Pi(t)$ — vectors of parameters of controlled ecological factors.

Then the task of obtaining the generation with desired BRTs is expressed as:

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

which means the search for an optimum parental couple for crossing and optimization of ecological factors during a breeding process.

It's notably that in the case of controlled ecological factors, contributions of individual genetic-physiological systems are not analyzed separately – they are just intermediate variables of searching procedure that ends by formation of a forecasted eco-genetic portrait with corresponding predicted levels of BRTs.

The described algorithm can be completely realized virtually, while a real cross enables only one variant after which the only genotype with desired traits can be selected from F₂ population (with a purpose of improving a zoned variety), which significantly accelerates breeding work.

So, the authors have proposed the algorithm for controlled breeding process based on mathematical models of “genotype-environment” interaction, which includes a successive selection of the variants of eco-genetic portraits of potential parents and prediction of values of resulting quantitative breeding traits. Selection of the variants is performed until finding a desired degree of coincidence between predicted and desired breeding traits. In the case of controlled ecological factors, such selection is complemented with optimization of environmental conditions, which provides the most efficient utilization of parental eco-genetic potential and more reliable identification of genotypes by phenotypes in offspring.

REFERENCES

1. Falconer D.S. *Introduction to quantitative genetics*. Oliver-Boyd, Edinburg, 1961.
2. Shrikhande V.J. Some considerations in designing experiments on coconut trees. *J. Indian Soc. Agric. Stat.*, 1957, 9: 82-91.
3. Sakai K.I., Hatakeyama S. Estimation of genetic parameters in forest trees without raising progeny. *Silvae Genetica*, 1963, 12(5): 152-160.
4. Sakai K.I., Mukaide H. Estimation of genetic, environmental and competition variances in standing forests. *Silvae Genetica*, 1967, 16(5-6): 159-168.
5. Dragavtsev V.A. *Uspekhi Sovremennoi Biologii*, 1961, 52(3/6): 347-361.
6. Dragavtsev V.A. *Vestnik AN KazSSR*, 1963, 10(223): 33-42.
7. Dragavtsev V.A. *Botanicheskii Zhurnal*, 1966, 51(7): 939-946.
8. Dragavtsev V.A., Ostrikova V.M. *Genetika*, 1972, 8(4): 33-37.
9. Dragavtsev V.A., Pogozhev I.B., Sokolova T.A. *Modeli ekosistem i metody opredeleniya ikh parametrov* [Models of Ecosystems and Methods of Determining Their Parameters: Compilation of Sci. Works]. Novosibirsk, 1981: 109-116.
10. Dragavtsev V.A. *Agrarnaya Rossiya*, 2008, 4: 2-10.
11. Gronin V.V. *Ispol'zovanie kachestvennykh i morfometricheskikh priznakov dlya obespecheniya otlichimosti roditel'skikh linii i gibridov podsolnechnika. Avtoreferat kandidatskoi dissertatsii* [The Use of Quantitative and Morphometric Traits for Distinguishing Sunflower Parental Lines and Hybrids, Extended Abstract of Cand. Sci. Dissertation]. Krasnodar, 2007.
12. Mather Wharton B. *Principles of quantitative genetics*. Burgess Publishing, USA, 1964.
13. Litun P.P. *Materialy 4-go Vsesoyuznogo s"ezda VOGiS* [Proc. 4th Congress of Vavilov Society of Geneticist and Breeders]. Kishinev, 1982: 89-91.
14. Dragavtsev V.A., Aver'yanova A.F. *Genetika*, 1979, 15(3): 518-526.
15. Mikhailenko I.M., Dragavtsev V.A. *Sel'skokhozyaistvennaya Biologiya* [Agricultural Biology], 2010, 3: 26-35.