

ZASTOSOWANIE METOD HAYMANA-JINKSA DO OCENY ZMIENNOŚCI
GENETYCZNEJ MIESZAŃCÓW Z KRZYŻOWAŃ DIALLELICZNYCH

Maria Parlińska, Wiesław Mądry

Katedra Statystyki Matematycznej i Doświadczalnictwa
SGGW - AR w Warszawie

Analiza statystyczna układów diallelicznych wg modelu Griffinga [wg 2, 3] pozwala na oczacowanie ogólnej i specyficznej zdolności kombinacyjnej oraz efektów krzyżowań przeciwnych dla form rodzicielskich. Na podstawie oceny efektów GCA i SCA w modelu stałym nie można wnioskować o charakterze działania genów badanej grupy rodziców. W modelu losowym pojawiają się wprawdzie możliwości oceny charakteru działania genów, jednakże wnioski dotyczą puli genowej całej populacji, nie zaś tej, którą wnoszą konkretne formy rodzicielskie. Bardzo często genetycy chcieliby badać możliwie wnikliwie działanie genów pochodzących od rozważanej grupy rodziców. Dużą szansę dają metody Haymana-Jinksa [2, 3, 5]. Są one często stosowane w pracach nad dziedziczeniem cech ilościowych roślin uprawnych, a ponieważ nie są one wciąż jeszcze dostatecznie spopularyzowane, zdecydowaliśmy się na kolejne podjęcie tego tematu.

W niniejszej pracy analizę struktury genetycznej wybranych linii rodzicielskich i ich potomstwa przeprowadza się na podstawie modelu Haymana-Jinksa. Dotyczy on kompletnego układu diallelicznego zawierającego mieszańce F_1 proste, przeciwne i formy rodzicielskie homozygotyczne. Kolejne założenia są następujące:

- a/ geny niealleliczne działają niezależnie;
- b/ geny segregują niezależnie wg schematu diploidalnego;
- c/ każdy gen reprezentowany jest przez dwa allele;
- d/ rodzice nie są spokrewnieni.

Ponieważ każda cecha ilościowa uwarunkowana jest poligenicznie, założmy, że badaną cechę kontroluje r genów. Weźmy pod uwagę 3 możliwe genotypy w k -tym locus ($k = 1, 2, \dots, r$) $A_k A_k$, $A_k a_k$, $a_k a_k$, których efekty zgodnie z notacją Haymana-Jinksa i Mathera są następujące:

$$a/ \text{ genotyp} - A_k A_k, A_k a_k, a_k a_k;$$

$$b/ \text{ efekt} - d_k, h_k, -d_k.$$

Założmy dalej, że w krzyżowaniu diallelicznym częstość rodziców o genotypie $A_k A_k$ jest u_k , a o genotypie $a_k a_k$ jest v_k , gdzie $u_k + v_k = 1$. Parametry statystyczne, których oceny wyznacza się na podstawie obserwacji z tablicy diallelicznej można rozbić na dwa niezależne składniki - genetyczny i środowiskowy. Wartości oczekiwane estymatorów tych składników są funkcjami parametrów genetycznych oraz wariancji błędu. Przyjmijemy za Haymanem następujące oznaczenia:

$$D = \sum d_k^2 (1-w_k^2) \quad K = 1, 2, \dots, r$$

$$H_1 = \sum h_k^2 (1-w_k^2)$$

$$H_2 = \sum h_k^2 (1-w_k^2)$$

$$F = 2 \sum d_k h_k w_k (1-w_k^2)$$

$$h = \sum h_k (1-w_k^2).$$

Dalej, przyrównując odpowiednie charakterystyki statystyczne do wartości oczekiwanych otrzymujemy układ równań, z którego wyznaczymy estymatory parametrów genetycznych D , H_1 , H_2 , F , h^2 , przyjmując jako estymator σ^2 oszacowanie wariancji błędu z tabeli analizy wariancji S_e^2 :

$$\hat{D} = \hat{V}_p - S_e^2,$$

$$\hat{F} = 2\hat{V}_p - 4\hat{W}_{PF} + \frac{4-2n}{n} S_e^2,$$

$$\hat{H}_1 = \hat{V}_p - 4\hat{W}_{PF} + \hat{V}_{HF} + \frac{2-3n}{n} S_e^2,$$

$$\hat{H}_2 = 4\hat{V}_{HF} - 4\hat{V}_F - 2S_e^2,$$

$$\hat{h}^2 = 4(\hat{\mu} - \hat{\mu}_p)^2 - \frac{4(n-1)}{n^2} S_e^2,$$

gdzie:

\hat{V}_p - wariancja rodziców;

\hat{W}_{PF} - kowariancja między rodzicami i średnimi ich potomstwa;

- \hat{V}_{HF} - wariancja średnich dla półrodzin;
 \hat{V}_F - wariancja średnich dla rodzin;
 $\hat{\mu}$ - średnia potomstwa;
 $\hat{\mu}_p$ - średnia rodziców;
 n - liczba linii.

Analiza genetyczna możliwa jest tylko wtedy, gdy spełnione są założenia przyjęte w modelu w odniesieniu do wszystkich linii. Kryterium sprawdzającym jest ocena jednorodności różnic $W_i - V_i$ ($i = 1, 2, \dots, n$), gdzie: W_i - jest kowariancją między i -tą rodziną i wszystkimi rodzicami oraz V_i - wariancją i -tej rodziny. Można wykazać, że: $W_i - V_i = 0,25(\hat{D} - \hat{H}_1)$. Ponieważ $0,25(\hat{D} - \hat{H}_1)$ jest stałe dla danej tablicy niezależnie od i , więc różnica ocen $\hat{W}_i - \hat{V}_i$ powinna być dla każdego i taka sama. Oznaczając przez V zmienną rzeczywistą niezależną przyjmującą wartości będące ocenami V_1, V_2, \dots, V_n , a przez W zmienną losową zależną od wartościach W_1, W_2, \dots, W_n , dla oszacowania liniowej zależności W od V wyznaczamy równanie regresji. Weryfikacja hipotezy $H_0: \beta = 1 / \beta$ - współczynnik regresji/ jest jednocześnie weryfikacją hipotezy o jednorodności różnic między kowariancją rodzin i form rodzicielskich oraz wariancją form rodzicielskich. Odrzucenie tej hipotezy oznacza, że założenia przyjętego modelu genetycznego nie są spełnione przez wszystkie Linie. W takim przypadku należy zakończyć analizę genetyczną mieszaneńców z krzyżowań diallelicznych i poprzestać na analizie wariancji, która daje estymatory efektów ogólnej i specyficznej zdolności kombinacyjnej oraz estymatory efektów krzyżowania odwrotnego, lub wyeliminować Linie nie spełniające założeń modelu. Eliminację można rozpocząć od tych linii, dla których $W_i - V_i$ jest maksymalna. Dla zredukowanego zbioru konieczne jest ponowne przeprowadzenie analizy wariancji. Umożliwia ona [1, 2, 4] weryfikację hipotez związanych ze zmiennością genetyczną badanych form rodzicielskich. Wyrażenie wartości oczekiwanych średnich kwadratów jako funkcji parametrów genetycznych i rozkład sumy kwadratów dla specyficznej zdolności kombinacyjnej na 3 niezależne składniki pozwala na wnioskowanie o zmienności związanej z addytywnym działaniem genów, dominowaniem i względną częstością genów dominujących i recesywnych warunkujących badaną cechę. Wnioskowanie należy rozpocząć od weryfikacji hipotezy $H_{01}: H_2 = 0$ /brak dominacji/. Przy istnieniu dominacji kolejnym krokiem jest testowanie hipotezy $H_{02}: h^2 = 0$ /brak dominacji jednokierunkowej/. Następną hipotezą

jest $H_{03}: H_1 - H_2 = 0$ /brak dominacji w każdym locus oraz równość częstości alleli zwiększających i zmniejszających wartość cechy/. Ostatnia hipoteza $H_{04}: D - F + H_1 - H_2 = 0$ /brak zmienności addytywnej/ zależy od wyników testowania hipotezy H_{04} i H_{03} . W przypadku gdy odrzucamy hipotezę H_{04} i H_{03} pozostaje w mocy hipoteza H_{02} , lub nie odrzucamy H_{04} a odrzucamy H_{03} , wykazujemy istotną zmienność addytywną.

Po odrzuceniu hipotezy H_{01} wskazane jest zbadanie stopnia dominacji zdefiniowanego w przypadku tablicy diallelicznej jako $\sqrt{\frac{\hat{H}_1}{\hat{D}}}$. Potwierdzenie wniosku z testowania hipotezy H_{03} można uzyskać obliczając wartość ilorazu $\frac{\hat{H}_2}{4\hat{H}_1}$, który przy założeniu równości efektów dominowania szacuje średnią wartość uv . Jeżeli odrzucimy hipotezę H_{02} i przyjmiemy założenie o równości efektów dominowania, to obliczając wartość ilorazu $\frac{\hat{h}^2}{\hat{H}_2}$ można określić dolną granicę czynników efektywnych kontrolujących daną cechę. Korzystając z parametrów genetycznych Haymana można wyznaczyć także współczynniki odziedziczalności w sensie węższym [1, 7]:

$$\hat{H}_{ns}^2 = \frac{\hat{D} + \hat{H}_1 - \hat{H}_2 - \hat{F}}{\hat{D} + \hat{H}_1 - \hat{H}_2 - \hat{F} + 2S_e^2}$$

oraz szerszym:

$$\hat{H}_{bs}^2 = \frac{\hat{D} + \hat{H}_1 - \frac{1}{2}\hat{H}_2 - \hat{F}}{\hat{D} + \hat{H}_1 - \frac{1}{2}\hat{H}_2 - \hat{F} + 2S_e^2}.$$

Analiza graficzna V_R , W_R pozwala na dalszą dokładną interpretację genetyczną. W tym celu wyznacza się parabolę „limitującą” o równaniu $W_R^2 = \hat{V}_P V_R$.

Należy pamiętać, że jeśli badane linie rodzicielskie są próbą losową z populacji, to otrzymane wyniki można interpretować w odniesieniu do całej populacji. Jeżeli jednak z powodu niejednorodności chociaż jedna linia jest odrzucona, to dalsze wnioskowanie dotyczy już tylko analizowanych linii, a nie całej populacji.

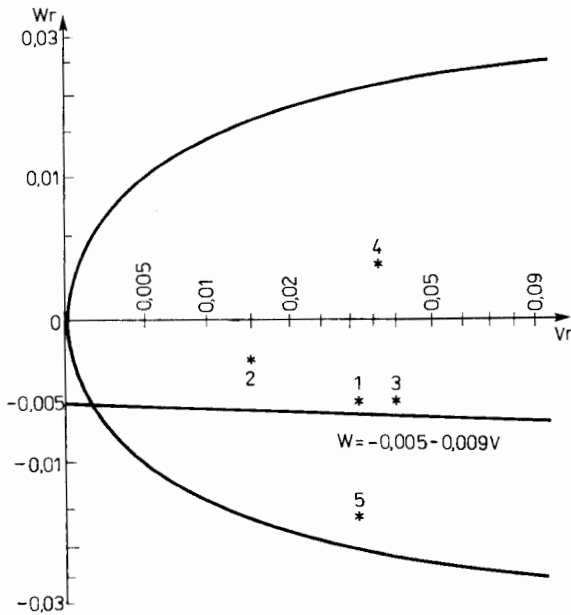
Do analizy genetycznej użyto danych dotyczących plonu ziarna /w dag z 1 rośliny/ potomstwa otrzymanego z krzyżowania w pełnym układzie diallelicznym [6] przy $n=6$ odmianach pszenicy ozimej.

Po utworzeniu tablicy diallelicznej przeprowadzono testowe - nie istotności efektów krzyżowania odwrotnego. Odrzucenie tej hipotezy pozwoliło na przejście do wyznaczenia \hat{W}_i i \hat{V}_i oraz testowania jednorodności linii rodzicielskich. Ponieważ stwierdzono niejednorodność, więc odrzucono linię /Scout/, która tę niejednorodność powodowała. Dla zredukowanego zbioru linii ponownie przeprowadzono test jednorodności. Wartości \hat{W}_i i \hat{V}_i dla badanych linii są następujące:

linie	\hat{V}_i	\hat{W}_i	$\hat{W}_i - \hat{V}_i$
Sturdy	0,032	-0,005	-0,037
Omecha	0,015	-0,002	-0,017
WS1003	0,04	-0,005	-0,045
WWRN	0,037	0,003	-0,033
Norin	0,032	-0,016	-0,048

Jednorodność badano testem t-Studenta, weryfikując hipotezę $H_0: \beta = 1$. Wobec wartości funkcji testującej $t_{emp} = -2,412$ przy wartości krytycznej $t_{0,05;4} = 2,776$ hipoteza pozostaje w mocy, tzn. uznajemy jednorodność badanych linii.

Analiza wariancji diallelicznej tablicy badanej cechy upoważnia do stwierdzenia genetycznej zmienności pomiędzy formami rodzicielskimi. Efekt dominowania jest istotny $/H_{01}/$, a podział tej zmienności daje informacje o wielokierunkowym dominowaniu $/H_{02}/$, asymetrii działania genów, oraz że indywidualne interakcje dominujących efektów mieszańców diallelicznej tablicy są istotne $/H_{03}/$. Różnice wynikające z krzyżowania odwrotnego, udział jakiegokolwiek linii występującej w mieszańcu jako forma mateczna lub ojcowska, jak również różnice między poszczególnymi mieszańcami są istotne. Na podstawie wartości $\hat{H}_1/\hat{D}=2,17$ można wnioskować, że w dziedziczeniu tej cechy w badanych liniach przeważa naddominowanie, natomiast $\sqrt{\hat{H}_1/\hat{D}} = 1,47$ oszacowuje średni stopień dominowania w każdym locus. Estymator $uv = \hat{H}_2/4\hat{H}_1 = 0,23$ określa średnią wartość uv . Stosunek ten mierzy względny rozdział tylko tych alleli, które ujawniają dominowanie. Współczynniki odziedziczalności wynoszą odpowiednio: w sensie węższym $\hat{H}_{ns}^2 = 48\%$ i w sensie szerszym $\hat{H}_{bs}^2 = 87\%$.



Graficzna analiza W_r, V_r obrazuje obecność naddominowania, ponieważ prosta regresji przecina oś OW poniżej układu. Oznacza to, że plon jest uwarunkowany naddominowaniem.

LITERATURA

1. Dobek A., Kaczmarek Z., Kielczewska H., Łuczkiwicz T.: Podstawy i założenia analizy statystycznej krzyżówek diallelicznych *Materiały VII Col. Agr.*, 146-168, 1978.
2. Hayman B.J.: The theory and analysis of diallel-crosses, *Genetics* 39, 789-809, 1954.
3. Jinks J.L.: The analysis of continuous variation in a diallel cross of *Nicotiana rusticalis* varieties, *Genetics* 39, 767-788, 1954.
4. Królikowski Z.: Dziedziczenie cech ilościowych na podstawie diallelicznych krzyżowań, *Hod. Rośl. Aklim.* 19, 3, 187-229, 1975.
5. Mather K., Jinks J.L.: *Biometrical genetics*, Chapman, 1971.
6. Mądry W., Ubysz-Borucka L.: Biometryczna analiza zdolności kombinacyjnej w niekompletnym układzie diallelicznym. Model stały dla cech ilościowych, *Roczn. Nauk Roln.* 1, 105, 1982.
7. Parlińska M., Ubysz-Borucka L.: Some multivariate estimations of genetic parameters, *Mitteilungsblatt Österreichische Statistische Gesellschaft*, 197-202, 1984.

M. Parlińska, W. Mađry

APPLICATION OF STATISTICAL ANALYSIS HAYMAN-JINKS'
FOR ESTIMATION OF VARIABILITY OF DIALLEL CROSSINGS

S u m m a r y

The analysis of genetics structure of n parent lines presented in this paper, is based on the Hayman-Jinks' and Mather's model, in which it was assumed that there was no episthesis. The analysis allows to determine the effects of additivity and dominance and provides the information about the distribution of genes in parental forms. The final section is an example for 6 lines of winter wheat.

М. Парлиньска, В. Мондры

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЕЙМАНА-ДЖИНКСА ДЛЯ ОЦЕНКИ
ГЕНЕТИЧЕСКОЙ ИЗМЕНЧИВОСТИ ГИБРИДОВ ПОЛУЧЕННЫХ
В ДИАЛЛЕЛЬНОМ СКРЕЩИВАНИИ

Р е з ю м е

В статье представлен аналитический анализ n родительских линий базирующий на модели Геймана-Джинкса и Матера, в котором принято отсутствие эпистаза. Этот анализ позволяет определить аддитивные эффекты и доминантность и информирует о расположении генов в родительских формах. В последней части дается пример для шести линий озимой пшеницы.