

JAN BOCIANOWSKI¹

PAWEŁ KRAJEWSKI

Instytut Genetyki Roślin Polskiej Akademii Nauk w Poznaniu

Porównanie dwóch metod estymacji efektu addytywnego działania genów na podstawie linii podwojonych haploidów jęczmienia

Comparison of two methods of estimation of additive gene action effect on the basis of doubled haploid lines in barley

Praca zawiera numeryczne porównanie dwu metod estymacji efektu addytywnego działania genów. W pierwszej metodzie zakładamy, że obserwujemy jedynie fenotyp rośliny, natomiast w drugiej dodatkowe informacje otrzymujemy z obserwacji markerów molekularnych. W prezentowanej pracy analizowano dane z 90 doświadczeń przeprowadzonych na 150 liniach podwojonych haploidów jęczmienia uzyskanych z pokolenia F₁ krzyżówki Steptoe × Morex metodą *Hordeum bulbosum* oraz dane z obserwacji 223 markerów, w większości molekularnych typu RFLP, wykonanych na tych liniach, zaczerpnięte z Północnoamerykańskiego Projektu Badania Genomu Jęczmienia. W większości rozważanych doświadczeń zaobserwowano niedoszacowanie efektu addytywnego działania genów obliczonego na podstawie obserwacji fenotypowych i markerów, w stosunku do oceny jedynie na podstawie obserwacji fenotypowych. Obserwowano dużą rozpiętość wyników względnego porównania prezentowanych metod estymacji.

Słowa kluczowe: efekt addytywnego działania genów, estymacja, jęczmień, linie podwojonych haploidów, markery molekularne

The paper presents numerical comparison of two methods of estimation of the additive gene action effect. In the first method we assume that we observe only the plant phenotype, while in the second method we have additional information from the molecular markers observations. In this paper we analysed phenotypic data from 90 experiments on 150 barley doubled haploid lines and data concerning 223 molecular markers (data from the North American Barley Genome Mapping Project). In most of the considered experiments we observed that the additive effect calculated on the basis of the marker observations underestimates the total additive effect obtained from phenotypic observations only. We observed a large range of results of the relative comparison of the presented estimation methods.

Key words: barley, doubled haploid lines, effect of additive gene action, estimation, molecular markers

¹ Stypendysta Fundacji na rzecz Nauki Polskiej (FNP) na rok 2002

WSTĘP

Estymacja efektu związanego z addytywnym działaniem genów, oznaczanego zwykle symbolem $[d]$, na podstawie linii homozygotycznych (linii podwojonych haploidów, DH; linii wsobnych, RI) jest niezależna od współczynnika rozkładu genów u form rodzicielskich (Kaczmarek i in., 1984). Wykorzystując linie DH nie dało się jednak „uciec” od założenia, że wszystkie geny determinują badaną cechę fenotypową w równym stopniu (Kaczmarek i in., 1984). Założenie to w praktyce nie jest spełnione. Do oceny parametru $[d]$ nie są potrzebne dodatkowe założenia o braku sprzężeń pomiędzy loci (jak ma to miejsce w przypadku oceny innych parametrów genetycznych), gdyż estymacja tego parametru jest możliwa zarówno przy braku, jak i obecności sprzężeń genów kontrolujących badaną cechę (Surma, 1996).

Wiele nowych i znaczących informacji wniosło zastosowanie w badaniach obserwacji przeprowadzanych na markerach, m.in. molekularnych. Możliwe stało się estymowanie efektów poszczególnych genów determinujących badaną cechę ilościową (QTL), a więc rozwiązano problem założenia równego wpływu genów na fenotyp. Istnieją prace, w których oceniano efekty QTL dla linii homozygotycznych (Frova i Sari-Gorla, 1994; Sari-Gorla i in., 1997; Charmet i in., 1999; Li i in., 2001), jednakże porównania tych ocen z oceną efektu addytywnego działania genów wyliczonego na podstawie obserwacji fenotypowych, poza autorami tej pracy, dokonał jedynie Snape (1997) na przykładzie jednego doświadczenia.

Niniejsza praca jest kontynuacją prac nad numerycznym porównaniem dwu metod estymacji efektu addytywnego działania genów (Bocianowski i Krajewski, 2000, 2001 b). W jednej metodzie zakładamy, że dysponujemy jedynie obserwacjami fenotypowymi, natomiast w drugiej obok obserwacji fenotypowych brane są pod uwagę również obserwacje markerów molekularnych. W pracy przedstawiono postaci estymatorów efektu addytywnego działania genów wyznaczone dwoma metodami oraz dokonano numerycznego porównania tych ocen na przykładzie 150 linii DH jęczmienia uzyskanych z pokolenia F_1 kombinacji krzyżówkowej Steptoe $\times$ Morex, będących częścią Północno-amerykańskiego Projektu Badania Genomu Jęczmienia.

ESTYMACJA EFEKTU ADDYTYWNEGO DZIAŁANIA GENÓW

Ocena na podstawie fenotypu

Dysponując jedynie obserwacjami fenotypowymi roślin, efekt związany z addytywnym działaniem genów $[d]$ można ocenić wyróżniając tzw. grupy linii maksymalnych i minimalnych (Kaczmarek i in., 1984). W skład tych grup wchodzi linie, które teoretycznie posiadają allele jedynie, odpowiednio, zwiększające i zmniejszające wartość badanej cechy. Estymator wspomnianego parametru ma wówczas postać (Kaczmarek i in., 1984):

$$[\hat{d}]_f = \frac{1}{2}(\bar{L}_{\max} - \bar{L}_{\min}), \quad (1)$$

gdzie $\bar{L}_{\max}$ oraz $\bar{L}_{\min}$ oznaczają wartości średnie dla grup linii, odpowiednio, maksymalnych oraz minimalnych. Różne metody wyboru grup linii ekstremalnych opisali Bocianowski i wsp. (1999). W niniejszej pracy, ze względu na dysponowanie jedynie średnimi dla linii, wyboru grup linii ekstremalnych dokonano metodą kwantyli (Bocianowski i in., 1999).

Ocena na podstawie fenotypu oraz markerów

Jeżeli oprócz obserwacji fenotypowych linii roślinnych dysponujemy również obserwacjami markerów molekularnych wykonanymi na tych liniach, to wówczas wybierając p markerów najlepiej wyjaśniających zmienność badanej cechy, będących najprawdopodobniej sprzężonymi z genami determinującymi cechę ilościową (QTL), obserwacje fenotypowe y dla n linii można wyrazić za pomocą modelu regresji wielokrotnej:

$$y = \mu \mathbf{1} + d_1 \mathbf{m}_1 + d_2 \mathbf{m}_2 + \dots + d_p \mathbf{m}_p + \mathbf{e},$$

gdzie y jest n -wymiarowym wektorem obserwacji fenotypowych, μ oznacza średnią ogólną, $\mathbf{m}_s$ oznacza wektor obserwacji s -tego QTL ($s = 1, 2, \dots, p$), d_s jest natomiast jego efektem, $\mathbf{e}$ jest n -wymiarowym wektorem zmiennych losowych o rozkładzie normalnym takim, że $E(e_i) = 0$, $\text{Cov}(e_i, e_j) = 0$ dla $i \neq j$, $i, j = 1, 2, \dots, n$. Parametry występujące w powyższym modelu estymować można między innymi metodą najmniejszych kwadratów (Searle, 1982). Ocenę sumarycznego efektu addytywnego działania genów dokonujemy według wzoru:

$$[\hat{d}]_g = \sum_{k=1}^p |\hat{d}_k| \quad (2)$$

W niniejszej pracy markery sprzężone z QTL wybierano poprzez selekcję krokową wstecz w trzech etapach. W pierwszym etapie dokonano selekcji wewnątrz chromosomów, w drugim wybrane markery poddano wspólnej selekcji. W trzecim etapie rozważano sytuacje, w których wybrane markery znajdowały się na chromosomie blisko siebie (tzn. w odległości mniejszej niż 5 cM). Ponieważ takie markery są sprzężone najprawdopodobniej z jednym QTL, więc do ostatecznego modelu wybierany był ten z nich, dla którego wartość statystyki testowej była największa.

MATERIAŁ

Materiał do badań stanowiło 150 linii podwojonych haploidów jęczmienia uzyskanych z pokolenia F_1 krzyżówki Steptoe $\times$ Morex metodą *Hordeum bulbosum* oraz obserwacje 223 markerów (w większości molekularnych typu RFLP) wykonanych na tych liniach (Kleinhofs i in., 1993). Dane będące częścią Północnoamerykańskiego Projektu Badania Genomu Jęczmienia (NABGMP, Romagosa i in., 1996) uzyskano z doświadczeń przeprowadzonych w 16 środowiskach: Crookston, Minnesota, 1992 (MN92); Ithaca, New York, 1992 (NY92); Guelph, Ontario, 1992 (ON92); Pullman, Washington, 1992 (WA92); Brandon, Manitoba, 1992 (MA92); Outlook, Saskatchewan, 1992 (SKo92); Goodlae,

Saskatchewan, 1992 (SKg92); Kcfr, Saskatchewan, 1992 (SKk92); Tetonia, Idaho, 1992 (ID92); Bozeman, Montana, nawadniane, 1992 (MTi92); Bonzeman, Montana, teren suchy, 1992 (MTd92); Aberdeen, Idaho, 1991 (ID91); Klamath Falls, Oregon, 1991 (OR91); Pullman, Washington, 1991 (WA91); Bozeman, Montana, nawadniane, 1991 (MTi91); Bozeman, Montana, teren suchy, 1991 (MTd91). Linie badano pod względem ośmiu cech fenotypowych: zawartości α -amylazy, AA; siły diastatycznej, DP; zawartości białka w ziarnie, GP; plonu ziarna, GY; wysokości rośliny, H; czasu kwitnienia, HD; wylegania, L; ekstraktu słodu, ME. Przed przystąpieniem do analiz obserwacje cech GP, L, ME transformowano na stopnie Blissa. Nie wszystkie cechy były obserwowane we wszystkich środowiskach. Doświadczenia w poszczególnych środowiskach traktowano jako osobne eksperymenty.

WYNIKI I DYSKUSJA

W tabeli 1 zawarto wyniki względnego porównania ocen efektu addytywnego działania genów obliczonych według wzoru (2) oraz wzoru (1), dla poszczególnych cech i środowisk, tzn. wartości współczynnika $([d]_g/[d]_f) \cdot 100$.

Tabela 1

Względne porównanie ocen efektu addytywnego działania genów obliczonych wzorami (1) i (2) dla linii podwojonych haploidów krzyżówki Steptoe $\times$ Morex
Relative comparison of the estimators of additive gene action effect calculated with the formulas (1) and (2) for the doubled haploid lines

Doświadczenie Experiment	Cecha Trait							
	AA	DP	GP	GY	H	HD	L	ME
MN92	123,20	93,36	51,30	85,37	68,36	77,68	—	81,34
NY92	—	—	—	79,26	69,34	94,23	80,09	—
ON92	—	—	—	73,55	36,35	90,21	61,51	—
WA92	67,94	90,19	50,73	47,30	94,00	74,54	—	83,83
MA92	—	—	—	47,96	111,20	69,83	125,39	—
SKo92	—	—	—	34,13	60,38	63,95	82,54	—
SKg92	—	—	—	100,15	65,41	83,14	—	—
SKk92	—	—	—	54,02	81,19	79,24	—	—
ID92	77,42	113,10	47,04	66,88	47,26	74,81	—	40,24
MTi92	105,80	69,15	66,63	40,95	94,85	82,71	82,67	56,15
MTd92	48,52	86,13	79,65	63,76	60,14	79,27	79,43	76,66
ID91	94,46	72,24	100,90	83,23	64,67	86,93	—	56,76
OR91	79,87	76,67	34,99	44,85	52,09	72,62	—	69,38
WA91	75,01	65,77	62,41	60,27	63,67	71,34	—	63,73
MTi91	36,90	111,97	133,30	103,80	90,61	80,32	—	56,59
MTd91	—	—	—	81,12	73,65	81,77	—	—

AA — Zawartość α -amylazy; AA — α -amylase content

DP — Siła diastyczna; DP — Diastatic power

GP — Zawartość białka w ziarnie; GP — Protein content in grain

GY — Plon ziarna; GY — Grain yield

H — Wysokość rośliny; H — Plant height

HD — Czas kwitnienia; HD — Flowering time

L — Wyleganie; L — Lodging

ME — Ekstrakt słodu; ME — Malt extract

W większości rozważanych doświadczeń (89%) zaobserwowano niedoszacowanie parametru wyliczonego z wykorzystaniem obserwacji markerowych, co jest potwierdzeniem wstępnych wyników uzyskanych w przeprowadzonych badaniach symulacyjnych (Bocianowski i Krajewski, 2001 a). Przypadki, w których ocena fenotypowa była mniejsza niż ocena dokonana według wzoru (2) wytłumaczyć można dużą liczbą wyselekcjonowanych markerów (znacznie większą niż w pozostałych przypadkach), co w konsekwencji zwiększało ocenę sumarycznego efektu addytywnego działania genów. Rozpiętość uzyskanych wyników porównywanych ocen jest znaczna i zawiera się między 34,13% a 133,30%. Najbardziej wyrównanym środowiskiem okazało się MTd91 (od 73,65% do 81,77%), natomiast cechą, dla której zaobserwowano najmniejsze zróżnicowanie był ekstrakt słoju (od 40,24% do 83,83%).

W tabeli 2 przedstawiono ocenę zmienności wyjaśnianej przez zlokalizowane QTL, tzn. odpowiedni współczynnik determinacji ($w\%$).

Tabela 2

Ocena zmienności wyjaśnianej przez zlokalizowane QTL (współczynniki determinacji, R^2)
Variability explained by localized QTL (coefficients of determination, R^2)

Doświadczenie Experiment	Cecha Trait							
	AA	DP	GP	GY	H	HD	L	ME
MN92	40,5	34,4	11,1	28,3	61,5	78,1	—	17,8
NY92	—	—	—	30,4	48,2	80,0	53,2	—
ON92	—	—	—	24,7	35,9	12,9	23,7	—
WA92	43,2	53,2	40,4	25,2	62,4	77,3	—	45,9
MA92	—	—	—	37,0	80,9	75,2	62,8	—
SKo92	—	—	—	19,7	57,5	61,4	44,8	—
SKg92	—	—	—	37,4	53,0	78,7	—	—
SKk92	—	—	—	38,4	75,1	83,4	—	—
ID92	36,6	67,4	36,5	50,1	32,4	76,2	—	26,2
MTi92	52,5	51,8	38,2	26,5	61,3	76,6	41,2	40,3
MTd92	37,5	57,1	54,9	33,8	53,3	73,9	41,9	42,5
ID91	40,9	50,9	36,5	40,5	56,8	74,5	—	33,6
OR91	52,7	42,7	22,4	13,1	26,5	47,5	—	40,0
WA91	27,6	39,6	39,8	54,9	67,6	67,3	—	31,5
MTi91	18,8	66,5	31,7	47,9	61,5	68,7	—	36,8
MTd91	—	—	—	63,7	48,4	68,9	—	—

Objaśnienia jak w tabeli 1

The key for abbreviations of traits — see in Table 1

Zaobserwowano bardzo dużą zbieżność tych wyników (a co za tym idzie: liczby i miejsc zlokalizowanych genów) z wynikami uzyskanymi dla tych samych danych metodą największej wiarygodności (Hayes i Iyarnabo, 1994). Procent zmienności cechy wyjaśnianej przez QTL zawierał się w przedziale od 11,1% (dla GP w MN92) do 80,9% (dla H w MA92). Na szczególną uwagę zasługuje cecha HP, dla której poza doświadczeniem przeprowadzonym w środowisku ON92 (gdzie $R^2 = 12,9\%$) udział zmienności wyjaśnianej przez zlokalizowane QTL przekraczał wartość 50%. W 41% rozważanych doświadczeń wartości współczynników determinacji przekraczały 50%, co jest dobrym wynikiem, jeżeli

weźmiemy pod uwagę ostre kryterium wyboru markerów sprzężonych z *loci* determinującymi rozpatrywane cechy, zastosowane w niniejszej pracy.

WNIOSKI

1. Nieszacowanie (w 89% rozważanych doświadczeń) efektu addytywnego działania genów obliczonego na podstawie obserwacji fenotypowych i genotypowych, w stosunku do oceny jedynie na podstawie obserwacji fenotypowych, jest potwierdzeniem wyników poprzednich obliczeń na danych rzeczywistych oraz badań symulacyjnych. Sytuacje, w których ocena na podstawie markerów była wyższa, wynikają ze zbyt dużej liczby wyselekcjonowanych markerów.
2. Zlokalizowane QTL są zgodne z publikowanymi wynikami uzyskanymi metodą mapowania przedziałowego opartego na estymatorach największej wiarygodności.
3. Istnieje duża rozpiętość wyników względnego porównania dwu rozważanych metod estymacji efektu addytywnego działania genów.

LITERATURA

- Bocianowski J., Krajewski P. 2000. Estymacja efektów genetycznych na podstawie fenotypu i genotypu markerowego. *Colloquium Biometryczne* 30: 270 — 280.
- Bocianowski J., Krajewski P. 2001 a. Badania symulacyjne na temat porównania metody fenotypowej i genotypowej estymacji efektu addytywnego działania genów. XIV Zjazd Polskiego Towarzystwa Genetycznego „Genetyka w służbie człowieka”, Streszczenia: 200.
- Bocianowski J., Krajewski P. 2001 b. Numerical comparison of classical and marker-based methods of QTL effects estimation. In: *Quantitative genetics and breeding methods: the way ahead. Proceedings of the Eleventh Meeting of the EUCARPIA Section Biometrics in Plant Breeding*. A. Gallais, C. Dillmann and I. Goldringer (eds.). Institut National de la Recherche Agronomique, Paris, France, Les Colloques 96: 35 — 40.
- Bocianowski J., Krajewski P., Kaczmarek Z. 1999. Porównanie metod wyboru ekstremalnych linii podwojonych haploidów w estymacji parametrów genetycznych. *Colloquium Biometryczne* 29: 193 — 202.
- Charmet G., Robert N., Perretant M. R., Gay G., Sourdille P., Groos C., Bernard S., Bernard M. 1999. Marker-assisted recurrent selection for cumulating additive and interactive QTL's in recombinant inbred lines. *Theor. Appl. Genet.* 99: 1143 — 1148.
- Frova C., Sari-Gorla M. 1994. Quantitative trait *loci* (QTL) for pollen thermotolerance detected in maize. *Mol. Gen. Genet.* 245: 424 — 430.
- Hayes P. M., Iyarnabo O. 1994. Summary of QTL effects in the Steptoe × Morex population. *Barley Genetics Newsletter* 23: 98 — 143.
- Kaczmarek Z., Surma M., Adamski T. 1984. Parametry genetyczne — ich interpretacja i sposoby wyznaczania. *Listy Biometryczne — Biometrical Letters* 21: 3 — 20.
- Kleinohs A., Kilian A., Saghai Maroof M. A., Biyashev R. M., Hayes P., Chen F. Q., Lapitan N., Fenwick A., Blake T. K., Kanazin V., Ananiev E., Dahleen L., Kudrna D., Bollinger J., Knapp S. J., Liu B., Sorrells M., Heun M., Franckowiak J. D., Hoffman D., Skadsen R., Steffenson B. J. 1993. A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. *Theor. Appl. Genet.* 86: 705 — 712.
- Li Z.-K., Luo L. J., Mei H. W., Wang D. L., Shu Q. Y., Tabien R., Zhong D. B., Ying C. S., Stansel J. W., Khush G. S., Paterson A. H. 2001. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* 158: 1737 — 1753.

- Romagosa I., Ullrich S. E., Han F., Hayes P. M. 1996. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor. Appl. Genet.* 93: 30 — 37.
- Sari-Gorla M., Krajewski P., Binelli G., Frova C., Taramino G., Villa M. 1997. Genetic dissection of herbicide tolerance in maize by molecular markers. *Molecular Breeding* 3: 481 — 493.
- Searle S. P. 1982. *Matrix algebra useful for statistics*. John Wiley and Sons, New York.
- Snape J. W. 1997. Application of doubled haploid lines in plant breeding and genetical research: current issues and approaches. In: *Advances in Biometrical Genetics, Proceedings of the Tenth Meeting of the EUCARPIA Section Biometrics in Plant Breeding*, P. Krajewski and Z. Kaczmarek (eds.). Institute of Plant Genetics, Poznań, 35 — 46.
- Surma M. 1996. *Biometryczno-genetyczna analiza cech ilościowych mieszańców i linii podwojonych haploidów jęczmienia jarego*. IGR PAN, Poznań.