

Przegląd metod statystycznych lokalizacji i estymacji efektów genów determinujących cechy ilościowe

Jan Bocianowski, Paweł Krajewski
Instytut Genetyki Roślin Polskiej Akademii Nauk,
ul. Strzeszyńska 34, 60-479 Poznań

Słowa kluczowe: regresja liniowa, regresja wielokrotna, metoda największej wiarogodności, metoda najmniejszych kwadratów, geny kontrolujące cechy ilościowe

Wstęp

Wprowadzenie różnego typu markerów do badań genetycznych i hodowlanych zrewolucjonizowało te dziedziny nauki. Markery, m.in. molekularne, stosuje się w selekcji oraz do transformacji genetycznych [16, 38]. Przed rozpoczęciem tych czynności należy jednak zlokalizować gen lub grupę genów kontrolujących daną cechę, to znaczy podać odległości sprzężeniowe szukanych genów od badanych markerów. Dokładność i poprawność wniosków zależy w dużej mierze od precyzji określenia położenia genu na chromosomie. Dlatego też bardzo ważną rolę w lokalizacji i estymacji efektów genów odgrywają metody statystyczne.

Niniejsza praca ma na celu przeglądowe spojrzenie na podstawowe metody statystyczne stosowane w mapowaniu genów kontrolujących cechy ilościowe (QTL).

Doświadczenie i obserwacje

Doświadczenie genetyczne mające na celu lokalizację genów odpowiedzialnych za cechę ilościową polega na wykonaniu kontrolowanego krzyżowania dwu homozygotycznych form rodzicielskich i obserwacji potomstwa z tego krzyżowania, tzn. obserwacji mieszańców pokolenia F_2 , mieszańców z krzyżowań wstecznych, linii wsobnych lub linii podwojonych haploidów. Istotne jest, aby jednoznacznie stwierdzić, czy obserwowany obiekt (roślina, linia) pod względem danego markera jest typu takiego

jak pierwsza z form rodzicielskich, typu takiego jak druga z form rodzicielskich czy też jest heterozygotą. Tak więc przy założeniu, że mówimy o roślinach diploidalnych, a także iż marker może posiadać jedynie dwie formy alleliczne, obserwacja genotypu markerowego jest obserwacją cechy o liczbie możliwych wartości równej dwa (w wypadku markerów dominujących lub markerów kodominujących obserwowanych w potomstwie homozygotycznym) lub trzy (w innych wypadkach). Omawiane w dalszym ciągu modele dotyczą sytuacji, w której obserwacji podlegają linie homozygotyczne (wsobne lub podwojonych haploidów).

Potomstwo, w liczbie n , jest jednocześnie obserwowane pod względem cechy ilościowej, najczęściej w doświadczeniu połowym zakładanym w układzie blokowym. Wstępna analiza wariancji dla tego doświadczenia powinna odpowiedzieć na pytanie, czy obiekty są istotnie zróżnicowane. Jeżeli tak jest, fenotyp potomstwa jest dalej opisywany przez n wartości średnich.

Założmy więc, że dany jest n -wymiarowy wektor obserwacji (średnich) fenotypowych $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ oraz q n -wymiarowych wektorów obserwacji genotypowych \mathbf{m}_l , $l = 1, 2, \dots, q$, których elementami są 1 lub -1 , w zależności od tego, po którym z rodziców linia odziedziczyła allele w danym locus markerowym (znak „T” oznacza transpozycję wektora).

Dodatkowo założmy, iż dostępna jest mapa sprzężeniowa obserwowanych markerów, wykonana na podstawie wyliczonych częstości rekombinantów. Mapa taka określa najbardziej prawdopodobne liniowe ułożenie markerów wraz z odległościami pomiędzy nimi podanymi w morganaach; sposoby konstrukcji map podają m.in. Staub i in. [29] oraz Jones i in. [10].

Celem analizy jest identyfikacja nieznanymi QTL sprzężonych z obserwowanymi markerami.

Porównanie wartości średnich

Wczesne badania dotyczące linii wsobnych [28] mające na celu ocenę efektu genu sprzężonego z markerem oparte były na testowaniu istotności różnicy średnich fenotypowych. Jeżeli przez \bar{y}_1 oraz \bar{y}_{-1} oznaczymy wartości średnie tych linii, dla których pewien obserwowany marker przyjmował wartość, odpowiednio 1 oraz -1 , to statystyka testowa

$$Z = \frac{\bar{y}_1 - \bar{y}_{-1}}{s_D} \quad (1)$$

(gdzie s_D oznacza standardowy błąd różnicy) ma rozkład t-Studenta (dla $n \leq 30$) z $n-2$ stopniami swobody lub przybliżony rozkład normalny (dla $n > 30$). Istotna różnica pomiędzy średnimi dla dwu grup linii (na ustalonym poziomie istotności) interpretowana była jako sprzężenie pomiędzy QTL i markerem [27, 31]. Efekt QTL jest wówczas

oceniony jako połowa tej różnicy. Jak łatwo zauważyć, przedstawiona wyżej statystyka Z jest równoważna statystyce F otrzymanej z jednoczynnikowej analizy wariancji przy klasyfikacji linii ze względu na genotyp markerowy [30].

Podstawową wadą tej metody, jak i przyczyną niedoszacowania efektu QTL, jest niedokładność lokalizacji działających genów. W praktyce bardzo rzadko gen kontrolujący daną cechę zajmuje dokładnie locus, w którym zlokalizowany jest marker. Najczęściej QTL umiejscowiony jest w pewnej odległości od każdego z markerów naniesionych na dostępną mapę sprzężeniową.

Zauważyć należy analogię pomiędzy opisaną powyżej analizą a pewnymi rozważaniami znanymi z klasycznej genetyki ilościowej. Wartości \bar{y}_1 i \bar{y}_{-1} utożsamiać można z wartościami średnimi tzw. grup linii ekstremalnych, służącymi w genetyce ilościowej do estymacji efektów genów działających na cechę w wypadku, gdy dostępne są tylko obserwacje cechy ilościowej dla linii homozygotycznych [11]. Linie ekstremalne to linie o największej oraz najmniejszej ekspresji badanej cechy, czyli teoretycznie te, które zawierają jedynie allele, odpowiednio, zwiększające oraz zmniejszające wartość cechy. Odpowiedniość, o której mówimy, byłaby pełna, gdyby wszystkie linie maksymalne miały genotyp jednego z rodziców, a linie minimalne – drugiego. Oczywiście w praktyce sytuacja taka nie występuje, przede wszystkim dlatego, że do linii ekstremalnych zalicza się jedynie kilka linii reprezentujących skrajne wartości empirycznego rozkładu cechy ilościowej.

Regresja liniowa

Lander i Botstein [14] zaproponowali w celu estymacji położenia i efektu QTL tzw. mapowanie przedziałowe (interval mapping). Polega ono na weryfikacji hipotezy o istnieniu QTL pomiędzy dwoma tzw. markerami flankującymi, w kolejnych pozycjach chromosomu wybranych z zadaniem krokiem, np. 1 centymorgan (cM). Dla każdej takiej pozycji obserwacje fenotypowe opisuje się modelem regresyjnym

$$y = \mu \mathbf{1} + b \mathbf{x}_l + \mathbf{e} \quad (2)$$

gdzie μ oznacza średnią ogólną, \mathbf{x}_l – wektor wartości genotypowych QTL, natomiast b oznacza efekt tego QTL. Elementy wektora \mathbf{x}_l są oczywiście nieobserwowalne, wyznacza się je za pomocą odpowiedniej funkcji wartości genotypowych dla markerów flankujących oraz odległości QTL od tych markerów. W celu zweryfikowania hipotezy zerowej

$$H_0 : b = 0 \quad (3)$$

o braku efektu QTL, przeciwko hipotezie alternatywnej

$$H_1 : b \neq 0 \quad (4)$$

mówiącej, że efekt ten jest różny od zera, Lander i Botstein [14] zastosowali metodę największej wiarygodności. Zaproponowana przez nich statystyka ma postać

$$LOD = \log \frac{L_1}{L_0} \quad (5)$$

gdzie L_1 oraz L_0 oznaczają maksima funkcji wiarygodności wyznaczone przy prawdziwości hipotez, odpowiednio, alternatywnej oraz zerowej. Jako wartość progową (krytyczną) LOD_{KR} Lander i Botstein [14] zaproponowali liczbę z przedziału $\langle 2; 3 \rangle$, wybraną w zależności od wielkości genomu i gęstości rozmieszczenia markerów. Sami zaś w praktyce istotność efektów QTL oceniali przy wartości progowej 2,4. Odpowiada to, w przybliżeniu, poziomowi istotności równemu 0,001.

Problem wyboru wartości progowej LOD_{KR} okazał się dość istotny w badaniach praktycznych. Różni autorzy przyjmowali różne wartości: od 2 [12] do 3 [19]. Pewne rozwiązanie tego problemu zaproponowali Churchill i Doerge [3], sugerując użycie testów permutacyjnych do wyznaczania LOD_{KR} .

Opisywana tu metoda regresji znalazła szerokie zastosowanie w lokalizacji QTL i estymacji efektów w badaniach empirycznych (m.in. [5, 9, 21] oraz liczne prace poświęcone różnym gatunkom roślin); spośród prac autorów polskich na uwagę zasługują pozycje Masojć i Milczarski [17], Masojć i in. [18] oraz Wolko i in. [36]. Metoda ta jest podstawą działania programu MAPMAKER/QTL, najpopularniejszego programu komputerowego służącego do lokalizacji QTL.

Wadą metody największej wiarygodności jest niewykrywanie istotnych QTL w pewnych sytuacjach. Na przykład, gdy pomiędzy markerami flankującymi znajdują się dwa loci (lub więcej) determinujące daną cechę, użycie metody największej wiarygodności spowoduje wykrycie tylko jednego z nich – tego, dla którego wartość statystyki testowej będzie największa.

Regresja wielokrotna

Reiter i in. [23] zastosowali w badaniach nad kukurydzą model wielokrotnej regresji liniowej do lokalizacji i estymacji efektów genów odpowiedzialnych za odporność na stres. Markery o istotnych efektach regresyjnych (ocenianych indywidualnie) umieszczane były w modelu jako zmienne niezależne (objaśniające). Przy stosunkowo niewielkiej liczbie obserwacji genotypowych (77 markerów typu RFLP) niedoszacowanie efektów QTL tłumaczono tym, że loci z dużymi efektami fenotypowymi mogły znajdować się w rejonach genomu, które nie były pokryte markerami, a co za tym idzie, nie mogły być wykryte.

Dążąc do poprawy oszacowania parametrów, zwiększano liczbę obserwowanych markerów, udoskonalając zarazem metody estymacji. Zaproponowano model regresji wielokrotnej

$$\mathbf{y} = \mu \mathbf{1} + \sum_{i=1}^p d_{l_i} \mathbf{m}_{l_i} + \mathbf{e} \quad (6)$$

gdzie $l_i \in \{1, 2, \dots, q\}$, $i = 1, 2, \dots, p$. Występujące w tym modelu markery \mathbf{m}_{l_i} utożsamiane są z QTL determinującymi badaną cechę, d_{l_i} oznaczają efekty poszczególnych QTL. Rozważanie determinowania badanej cechy przez wiele QTL pozwala na wyeliminowanie markerów, które testowane indywidualnie wykazują wpływ na cechę, lecz w analizach wielokrotnych nie charakteryzują się żadnym wpływem na wartość fenotypową. Wybór markerów dokonywany jest najczęściej poprzez selekcję krokową w przód lub wstecz. Model (6) można przedstawić w postaci macierzowej jako

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (7)$$

gdzie $\mathbf{X} = [\mathbf{1}, \mathbf{m}_{l_1}, \mathbf{m}_{l_2}, \dots, \mathbf{m}_{l_p}]$, a $\boldsymbol{\beta} = [\mu, d_{l_1}, d_{l_2}, \dots, d_{l_p}]^T$. Przy spełnieniu założeń rozważanych przez Landera i Botsteina [14] dotyczących wektora \mathbf{y} oraz macierzy obserwacji markerowych \mathbf{X} do estymacji wektora nieznanymi parametrów $\boldsymbol{\beta}$ zastosować można metodę największej wiarygodności, jak czyni chociażby Zeng [40]. Dla powyższego modelu zastosować można również metodę najmniejszych kwadratów. Wówczas estymator wektora $\boldsymbol{\beta}$ ma postać [26]:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (8)$$

Weryfikacji hipotezy o istotności składowych wektora $\boldsymbol{\beta}$ dokonujemy za pomocą statystyki

$$F = \frac{MS_R}{MS_E} \quad (9)$$

gdzie MS_R i MS_E oznaczają średnie kwadraty dla, odpowiednio, modelu i błędu. Statystyka (9) ma rozkład F -Snedecora z p i $n - p - 1$ stopniami swobody.

Sprzężenie markera z QTL, jak już wspomniano, bardzo rzadko jest całkowite. W celu poprawy estymacji efektów QTL Zeng [39] zaproponował więc tzw. złożone mapowanie przedziałowe (composite interval mapping), łączące mapowanie przedziałowe z regresją wielokrotną. W modelu (6) pojawiają się dodatkowo zmienne związane z QTL zlokalizowanymi pomiędzy obserwowanymi markerami. Otrzymujemy wówczas model

$$\mathbf{y} = \mu \mathbf{1} + \sum_{i=1}^p \mathbf{d}_{l_i} \mathbf{m}_{l_i} + \sum_{j=1}^t \mathbf{d}_j \mathbf{x}_j + \mathbf{e} \quad (10)$$

gdzie t oznacza liczbę QTL zlokalizowanych pomiędzy markerami. Elementy każdego z wektorów \mathbf{x}_j wyznacza się, podobnie jak w wypadku (2), jako funkcję wartości genotypowych markerów flankujących oraz odległości QTL od nich. Wszystkie parametry występujące w modelu (10) można estymować metodą najmniejszych kwadratów, podobnie jak w wypadku modelu (6). Model regresji wielokrotnej zastosowano w takich programach komputerowych służących do lokalizacji QTL, jak m.in. MQTL, QTL Cartographer.

Zaprezentowane powyżej metody lokalizacji i estymacji efektów genów determinujących cechę ilościową ukazują kształtowanie się procesu opracowywania danych doświadczalnych wraz ze wzrostem znaczenia obserwacji genotypowych markerów. W pracy ograniczono się do ogólnego scharakteryzowania trzech metod: opartej na porównaniu średnich wartości fenotypowych, wykorzystującej regresję liniową oraz opartej na regresji wielokrotnej. Ogromny wkład w proces opracowywania tych metod wniosły badania symulacyjne (m.in. [6, 20, 35]).

Z wymienionych w pracy metod dużą popularnością cieszy się obecnie, pomimo pewnych wad, metoda zaproponowana przez Landera i Botsteina [14], oparta na regresji liniowej i zasadzie największej wiarygodności. Ma to bezpośredni związek z dostępnością programu komputerowego MAPMAKER/QTL bazującego na tej właśnie metodzie. Niedostatkami metody (oraz wspomnianego programu) jest trudność w przystosowaniu analizy statystycznej do konkretnej sytuacji eksperymentalnej, np. doświadczenia blokowego lub doświadczenia przeprowadzonego w wielu środowiskach. Lokalizacja QTL dokonywana jest więc na podstawie średnich dla linii.

Wydaje się, że lepszą, bardziej precyzyjną metodą jest złożone mapowanie przedziałowe [39]. Do wykrycia i oceny efektów QTL zastosować w tym wypadku można, przyjmując odpowiednie założenia, zarówno metodę największej wiarygodności, jak i najmniejszych kwadratów. Używając tej ostatniej, obliczenia można wykonać nie tylko za pomocą jednego ze specjalistycznych programów wspomnianych poprzednio, ale także używając ogólnych pakietów statystycznych, stosując standardowe procedury regresji wielokrotnej.

Na tle przedstawionych metod lokalizacji QTL oraz wynikających z literatury potrzeb badań genetycznych powstaje pytanie o kierunki rozwoju odpowiedniej metodologii statystycznej. Jednym z takich kierunków jest niewątpliwie uwzględnienie problemu interakcji QTL \times środowisko [1, 4, 8, 25, 33]. Drugim wydaje się wzięcie pod uwagę interakcji pomiędzy determinującymi cechę genami [22, 24, 33, 37]. Estymacja efektów związanych z interakcją pomiędzy QTL jest najczęściej pomijana celowo, gdyż branie pod uwagę tych efektów powoduje szybkie zwiększenie liczby parametrów w modelu [7, 32]. Wreszcie, interesujące wydają się prace nad wykorzystaniem w lokalizacji QTL modeli wielowymiarowych, pomocnych szczególnie w badaniach nad plejotropowością i sprzężeniem QTL determinujących różne cechy fenotypowe [2, 13, 15, 34].

Literatura

- [1] Aastveit A.H., Aastveit K. 1993. Effects of genotype-environment interactions on genetic correlations. *Theoretical and Applied Genetics* 86: 1007–1013.

- [2] Caliński T., Kaczmarek Z., Krajewski P., Frova C., Sari-Gorla M. 2000. A multivariate approach to the problem of QTL localization. *Heredity* 84: 303–310.
- [3] Churchill G.A., Doerge R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
- [4] Hayes P.M., Liu B.H., Knapp S.J., Chen F., Jones B., Blake T., Franckowiak J., Rasmusson D., Sorrells M., Ullrich S.E., Wesenberg D., Kleinjans A. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. *Theoretical and Applied Genetics* 87: 392–401.
- [5] Heun M., Kennedy A.E., Anderson J.A., Lapitan N.L.V., Sorrellis M.E., Tanksley S.D. 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). *Genome* 34: 437–447.
- [6] Jansen R.C. 1996. A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* 142: 305–311.
- [7] Jansen R.C., Stam P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447–1455.
- [8] Jansen R.C., van Ooijen J.W., Stam P., Lister C., Dean C. 1995. Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. *Theoretical and Applied Genetics* 91: 33–37.
- [9] Jensen J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. *Theoretical and Applied Genetics* 78: 613–618.
- [10] Jones N., Ougham H., Thomas H. 1997. Markers and mapping: we are all geneticists now. *New Phytologist* 137: 165–177.
- [11] Kaczmarek Z., Surma M., Adamski T. 1984. Parametry genetyczne – ich interpretacja i sposoby wyznaczania. *Listy Biometryczne – Biometrical Letters* 21: 3–20.
- [12] Kjær B., Jensen J., Giese H. 1995. Quantitative trait loci for heading date and straw characters in barley. *Genome* 38: 1098–1104.
- [13] Korol A.B., Ronin Y.I., Nevo E., Hayes P.M. 1998. Multi-interval mapping of correlated trait complexes. *Heredity* 80: 273–284.
- [14] Lander E.S., Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- [15] Mangin B., Thoquet P., Grimsley N. 1998. Pleiotropic QTL analysis. *Biometrics* 54: 88–99.
- [16] Marczewski W. 1995. Markery molekularne w genetyce i hodowli roślin. *Postępy Biochemii* 41(4): 237–243.
- [17] Masojć P., Milczarski P. 1999. Wykorzystanie mapy molekularnej genomu żyta i analizy QTL do identyfikacji genów warunkujących wczesność kłoszenia. *Biuletyn IHAR* 211: 205–210.
- [18] Masojć P., Milczarski P., Banek-Tabor A. 2001. Mapowanie interwałowe (IM) genów warunkujących odporność żyta (*Secale cereale* L.) na porastanie. XIV Zjazd Polskiego Towarzystwa Genetycznego „Genetyka w służbie człowieka”, Streszczenia: 66–67.
- [19] Menancio-Hautea D., Fatokun C.A., Kumar L., Danesh D., Young N.D. 1993. Comparative genome analysis of mungbean (*Vigna radiata* L. WILCZEK) and cowpea (*V. unguiculata* L. WALPERS) using RFLP mapping data. *Theoretical and Applied Genetics* 86: 797–810.

- [20] Moreno-Gonzalez J. 1992. Estimates of marker-associated QTL effects in Monte Carlo backcross generations using multiple regression. *Theoretical and Applied Genetics* 85: 423–434.
- [21] Paterson A.H., Damon S., Hewitt J.D., Zamir D., Rabinowitch H.D., Lincoln S.E., Lander E.S., Tanksley S.D. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127: 181–197.
- [22] Pressoir G., Albar L., Ahmadi N., Rimbault I., Lorieux M., Fargette D., Ghesquiere A. 1998. Genetic basis and mapping of the resistance to rice yellow mottle virus. II. Evidence of a complementary epistasis between two QTLs. *Theoretical and Applied Genetics* 97: 1155–1161.
- [23] Reiter R.S., Coors J.G., Sussman M.R., Gabelman W.H. 1991. Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. *Theoretical and Applied Genetics* 82: 561–568.
- [24] Roper R.J., Griffith J.S., Lyttle C.R., Doerge R.W., McNabb A.W., Broadbent R.E., Teuscher C. 1999. Interacting quantitative trait loci control phenotypic variation in murine estradiol-regulated responses. *Endocrinology* 140(2): 556–561.
- [25] Sari-Gorla M., Caliński T., Kaczmarek Z., Krajewski P. 1997. Detection of QTL \times environment interaction in maize by a least squares interval mapping method. *Heredity* 78: 146–157.
- [26] Searle S. R. 1982. Matrix algebra useful for statistics. John Wiley & sons, New York: 438 ss.
- [27] Simpson S.P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. *Theoretical and Applied Genetics* 77: 815–819.
- [28] Soller M., Brody T., Genizi A. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* 47: 35–39.
- [29] Staub J.E., Serquen F.C, Gupta M. 1996. Genetic markers, map construction, and their application in plant breeding. *HortScience* 31(5): 729–740.
- [30] Stuber C.W., Lincoln S.E., Wolff D.W., Helentjaris T., Lander E.S. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132: 823–839.
- [31] Tanksley S.D., Medina-Filho H., Rick C.M. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49: 11–25.
- [32] van der Schaar W., Alonso-Blanco C., Leon-Kloosterzei K.M., Jansen R.C., van Ooijen J.W., Koornneef M. 1997. QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. *Heredity* 79: 190–200.
- [33] Wang D.L., Zhu J., Li Z.K., Paterson A.H. 1999. Mapping QTLs with epistatic effects and QTL \times environment interactions by mixed linear model approaches. *Theoretical and Applied Genetics* 99: 1255–1264.
- [34] Weller J.I., Wiggans G.R., van Raden P.M., Ron M. 1996. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theoretical and Applied Genetics* 92: 998–1002.
- [35] Whittaker J.C., Curnow R.N., Haley C.S., Thompson R. 1995. Using marker-maps in marker-assisted selection. *Genetical Research Cambridge* 66: 255–265.

- [36] Wolko B., Irzykowska L., Święcicki W.K., Gawłowska M. 2001. Lokalizacja loci cech ilościowych (QTL) na mapie chromosomów *Pisum sativum* L. XIV Zjazd Polskiego Towarzystwa Genetycznego „Genetyka w służbie człowieka”, Streszczenia: 77–78.
- [37] Wu P., Liao C.Y., Hu B., Yi K.K., Jin W.Z., Ni J.J., He C. 2000. QTLs and epistasis for aluminum tolerance in rice (*Oryza sativa* L.) at different seedling stages. *Theoretical and Applied Genetics* 100: 1295–1303.
- [38] Young N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* 5: 505–510.
- [39] Zeng Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the USA* 90: 10972–10976.
- [40] Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.

Statistical methods in localization and estimation of the effects of quantitative trait loci

Key words: linear regression, multiple regression, maximum likelihood method, least squares method, quantitative trait locus

Summary

The paper presents three methods of localization and estimation of the effects of quantitative trait loci: the method based on the comparison of mean values for marker genotypes, the interval mapping method which uses the linear regression and the composite interval mapping method based on the multiple regression.