特集記事

QTL解析の理論

鵜飼保雄

(茨城県土浦市右舂 1994-9, 〒300-0837)

Theory of QTL analysis

Yasuo Ukai

(1994-9 Migimomi, Tsuchiura, Ibaraki 300-0837, Japan)

キーワード: QTL解析, 量的形質, 遺伝効果, マッピング, EM アルゴリズム

第1節 QTL解析の原理-F₂の区間マッピング法

量的形質は関与する遺伝子座が多く、分離集団において多数の遺伝子型が同時に分離する。さらに各遺伝子型値に環境による変動が加わり質的形質のように各個体の遺伝子型を推定することができない。従来の統計遺伝学的方法では、分離世代における形質の表現型値の平均や分散共分散に基づいて、関与する遺伝子座全体での遺伝効果の和や二乗和を推定してきた。個々の遺伝子座の遺伝効果や染色体上の位置の推定はあきらめるほかはなく、質的形質と違って解析は隔靴搔痒のもどかしさがあった。

量的遺伝子座 (Quantitative Trait Locus, QTL) の近傍に質的形質の遺伝子座 (マーカー) があれば、それとの連鎖を利用して、QTLの遺伝効果の推定が可能である。このようなQTL解析の原理はつとに1923年Saxによって提案されていたが、解析の実用化はDNAマーカー利用による詳細な連鎖地図が作れるようになってからである。ここではLanderとBotstein(1989)による区間マッピング法 (Interval Mapping) にもとづいてQTL解析の理論について要点を総説として紹介したい。ただし彼らはもっとも簡単な戻し交雑世代を例としているが、ここではQTL解析により適したF₂を対象とする (Hayashi と Ukai 1994)。なお海外ではQTL解析についての著書がいくつか (Falconer & MacKey 1989, Knapp 1994, Liu 1998) 出ているので詳細はそれらを参照されたい。

いま量的形質にk個の遺伝子座が関与しているとする。k個中のある1個の遺伝子座Qに注目する。座Qが近接する2つのマーカー (flanking markers) 座A, Bで挟まれた区間のどこかにあるとする。いまこれら3座における母親と父親の遺伝子型をそれぞれAAQQBB, aaqqbbと表す。ただし対立遺伝子の大文字小文字の区別は、優性劣性ではなく、母親と父親のどちらから由来したかを示す。F₁の遺伝子型はAQB/aqbとなる。ここで記号/はAQBとaqbが異なる相同染色体に乗っていることを示す。いまQ座はA座から測って組換え価r₁の距離にあると想定する (図1)。Q-B座, A-B座の間の組換え価をそれぞ

図1. マーカー座A,BとQTL座Qの染色体上の位置関係 (距離を組換え価で表した場合)。

ここでr₁₊₂はマーカー間距離で連鎖地図作成の過程ですでに求められている。下記の関係からr₁が与えられればr₂が求められる。

$$r_{1+2} = r_1 + r_2 - 2Cr_1r_2$$

Cは一致係数で、Haldane関数ではC=1, Kosambi関数ではC=2r₁₊₂となる。マーカーAの点 (r₁=0) から始めてマーカーBの点 (r₁=r₁₊₂) までQ座の位置を順次移して、そのつどLOD値を計算する。

れ、r₂, r₁₊₂と表す。r₁₊₂はマーカー間の組換え価で連鎖地図作成の時点ですでに分かっているので、r₁が決まればr₂はr₁の関数として求められる。QTL解析では、Q座の位置を染色体上の想定した点に決めて、その点におけるLOD(後述)値を求める。以下にQTL解析の計算手順を示す。

(1) マーカーの遺伝子型の下におけるQTL遺伝子型の条件つき確率

F₂ではA, B座のマーカーの遺伝子型は, AABb, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, aabbの9種となる。これら遺伝子型に1-9の番号をつけて示すとする。なおAB/ab (相引型) とAb/aB (相反型) は通常区別できないのでまとめてAaBbと表す。

マーカー座の遺伝子型によって、Q座の遺伝子型がQQ, Qq, qqのどれであるかの確率が異なり、その確率はマーカー座とQ座の間の組換え価の関数として表される。いまA座とQ座が同じ染色体上で近接していれば、両座の遺伝子型は相伴って次代に伝達されやすい。もし両座が同じ位置にあれば、A座の遺伝子型がAA, Aa, aaである個体は、それぞれQ座の遺伝子型が確実にQQ, Qq, qqとなる。いい換えると、この場合マーカー座の遺伝子型がAAである個体のQTL遺伝子型がQQである確率は1であり、Qq, qqである確率は0である。マーカー座とQTL座が離れるにつれてA座とQ座の間に組換えが生じる機会が増えるので、AA個体がQQである確率が1

表 1. 隣接マーカー遺伝子型と QTL 遺伝子型に対応した配偶子の遺伝子頻度

遺伝子型	Q	q
AB	q_1	q_2
Ab	q_3	q_4
aB	q_4	q_3
ab	q_2	q_1

ここで $q_1=(1-r_1-r_2+r_{12})/(1-r_{1+2})$

$q_2=r_{12}/(1-r_{1+2})$

$q_3=(r_2-r_{12})/r_{1+2}$

$q_4=(r_1-r_{12})/r_{1+2}$

$(q_1+q_2=1, q_3+q_4=1)$

表 2. QTL に近接する 2 つのマーカー座の遺伝子型が

$i(=1, 2, \dots, 9)$ であるとき QTL 遺伝子型が QQ, Qq, qq である条件付き確率 (F_2)

i: マーカー 遺伝子型	$QQ(p_{11})$	$Qq(p_{12})$	$qq(p_{13})$
1: $AABB$	q_1^2	$2 q_1 q_2$	q_2^2
2: $AABb$	$q_1 q_3$	$q_1 q_4 + q_2 q_3$	$q_2 q_4$
3: $Aabb$	q_3^2	$2 q_3 q_4$	q_4^2
4: $AaBB$	$q_1 q_4$	$q_1 q_3 + q_2 q_4$	$q_2 q_3$
5: $AaBb$	$z_1 q_1 q_2 + z_2 q_3 q_4$	$z_1 (q_1 + q_2^2) + z_2 (q_3^2 + q_4^2)$	$z_1 q_1 q_2 + z_2 q_3 q_4$
6: $Aabb$	$q_2 q_3$	$q_1 q_3 + q_2 q_4$	$q_1 q_4$
7: $aaBB$	q_4^2	$2 q_3 q_4$	q_3^2
8: $aaBb$	$q_2 q_4$	$q_1 q_4 + q_2 q_3$	$q_1 q_3$
9: $aabb$	q_2^2	$2 q_1 q_2$	q_1^2

$z_1=(1-r_{1+2})^2/\{(1-r_{1+2})^2+r_{1+2}^2\}$

$z_2=1-z_1$

$p_{11}+p_{12}+p_{13}=1 (i=1, 2, \dots, 9)$

よりしだいに小さくなり、いっぽう Qq または qq である確率は少しずつ大きくなる。マーカーと QTL が染色体の両端にあるとき、または異なる染色体上にあるときには、QTL の遺伝子型とマーカー遺伝子型は独立となる。なお単一座ではなく、近接した 2 マーカー座の遺伝子型を同時に利用すれば、QTL 遺伝子型の推定精度がやや高くなる。

F_2 個体の A, B マーカーの各遺伝子型別に、QTL 遺伝子型が QQ, Qq, qq である確率を求めるには、 F_1 個体の減数分裂で生じる配偶子におけるマーカー遺伝子型と QTL 遺伝子型の関係から考えると分かりやすい。各マーカー遺伝子型の下に Q 座の遺伝子型が Q また q となる条件付き確率は、表 1 のとおりとなる。ただし、 r_{12} は A-B 座間および Q-B 座間の両区間にも組換えが生じる

確率である。これより F_2 接合体における A, B 座の遺伝子型の下における Q 座の遺伝子型が QQ, Qq, qq となる条件付き確率が求められる (表 2)。

(2) マーカー遺伝子型および量的形質の表現型値を考慮した QTL 遺伝子型の推定

Q 座の相加効果、優性効果、残差分散を a, d, σ^2 とすると、 QQ, Qq, qq の遺伝子型値は、それぞれ $u+a, u+d, u-a$ となる。 u は定数である。表現型値はこの遺伝子型値に残差 e が加わったものとなる。ふつう残差 e は平均 0、分散 σ^2 の正規分布 $N(0, \sigma^2)$ に従うとする。 QQ, Qq, qq である個体が表現型値 y を示す確率密度は、それぞれ

$$\phi_1 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-u-a)^2}{2\sigma^2}} \quad (1a)$$

$$\phi_2 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-u-d)^2}{2\sigma^2}} \quad (1b)$$

$$\phi_3 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-u+a)^2}{2\sigma^2}} \quad (1c)$$

で示される。ゆえに、マーカー遺伝子型が i であるとき、Q 座の遺伝子型がたとえば $1(QQ)$ で、表現型値が y である確率は $p_{i1}\phi_1$ に比例する。マーカー遺伝子型および量的形質の表現型値という 2 つの独立した情報源を利用するとき、マーカー遺伝子型が $i (i=1, \dots, 9)$ である個体の Q 座の遺伝子型が QQ, Qq, qq である確率 z_1, z_2, z_3 は次式で与えられる ($z_1+z_2+z_3=1$)。

$$QQ: z_1 = p_{i1}\phi_1 / (p_{i1}\phi_1 + p_{i2}\phi_2 + p_{i3}\phi_3) \quad (2a)$$

$$Qq: z_2 = p_{i2}\phi_2 / (p_{i1}\phi_1 + p_{i2}\phi_2 + p_{i3}\phi_3) \quad (2b)$$

$$qq: z_3 = p_{i3}\phi_3 / (p_{i1}\phi_1 + p_{i2}\phi_2 + p_{i3}\phi_3) \quad (2c)$$

(3) 尤 度

確率 z_1, z_2, z_3 を用いると、尤度 (Likelihood, L) は

$$L \propto \prod_{i=1}^9 \prod_{j=1}^{n_i} (p_{i1}\phi_{ij1})^{z_{ij1}} (p_{i2}\phi_{ij2})^{z_{ij2}} (p_{i3}\phi_{ij3})^{z_{ij3}} \quad (3a)$$

と表せる (Hayashi と Ukai 1994)。ここで j は、マーカー遺伝子型が i 番目のタイプである個体群の j 番目 ($j=1, \dots, n_i$) の個体を示す。 $p_{i1}\phi_{ij1}, p_{i2}\phi_{ij2}, p_{i3}\phi_{ij3}$ はそれぞれ i 番目の遺伝子型をもつ個体群中の j 番目の個体における QTL 遺伝子型 QQ, Qq, qq の生起確率に比例した値である。ただし Lander と Botstein (1989) の方式では尤度は、つぎのとおりとなる。

$$L \propto \prod_{i=1}^9 \prod_{j=1}^{n_i} (p_{i1}\phi_{ij1} + p_{i2}\phi_{ij2} + p_{i3}\phi_{ij3}) \quad (3b)$$

(3a) と (3b) は QTL がマーカー上にあるとき一致する。

(4) パラメータ u , a , d , σ^2 の推定

QTL の染色体上の位置を固定したときに, QTL 解析において計算上推定しなければならない遺伝パラメータは, 定数 u , 相加効果 a , 優性効果 d , 残差分散 σ^2 の 4 つである。定数 u は推定の必要はあるが遺伝情報としては重要でない。残差分散とは, 環境分散に対象とする QTL 以外の QTL における遺伝的分離による遺伝分散が加わったものである。QTL が 1 個の場合には残差分散は環境分散に等しい。残差分散の大きさは QTL における遺伝効果や QTL 位置の推定精度に影響するが, 残差分散の推定値自体は QTL 解析においてはあまり重要でない。

尤度を最大にするようなパラメータ (ここでは u , a , d , σ^2) の値をもとめるのが, 最尤法の原理である。尤度を最大にする条件を求めるのは複雑なので, ふつう尤度の対数 (対数尤度) を最大にする値としてパラメータを推定する。対数尤度 $\log_e L$ は

$$\begin{aligned} \log_e L = & C + \sum_{i=1}^9 \sum_{j=1}^{ni} \{z_{ij1} \log p_{i1} + z_{ij2} \log p_{i2} + z_{ij3} \log p_{i3}\} \\ & - 0.5N \log (2\pi\sigma^2) - \sum_{i=1}^9 \sum_{j=1}^{ni} \left\{ z_{ij1} (y_{ij} - u - a)^2 / (2\sigma^2) \right. \\ & \left. + z_{ij2} (y_{ij} - u - d)^2 / (2\sigma^2) + z_{ij3} (y_{ij} - u + a)^2 / (2\sigma^2) \right\} \quad (4) \end{aligned}$$

となる。ただし C は定数である。

求むべきパラメータ u , a , d , σ^2 について対数尤度を偏微分して 0 に等しいとおくと, 4 つの式が得られる。それらを連立方程式として解くことになる。計算過程をなるべく省かずに以下に示す。

$$\begin{aligned} \frac{\partial \log L}{\partial u} = & \frac{1}{\sigma^2} \sum_{i=1}^9 \sum_{j=1}^{ni} \{z_{ij1} (y_{ij} - u - a) + z_{ij2} (y_{ij} - u - d) \\ & + z_{ij3} (y_{ij} - u + a)\} \\ = & \frac{1}{\sigma^2} \sum_{i=1}^9 \sum_{j=1}^{ni} \{ (z_{ij1} + z_{ij2} + z_{ij3}) y_{ij} - (z_{ij1} + z_{ij2} + z_{ij3}) u \\ & - (z_{ij1} - z_{ij3}) a - z_{ij2} d \} \\ = & \frac{1}{\sigma^2} \sum_{i=1}^9 \sum_{j=1}^{ni} (y_{ij} - u - w_{ij} a - z_{ij2} d) = 0 \quad (5) \end{aligned}$$

より

$$N\hat{u} + A\hat{a} + (N-B)\hat{d} = \sum_{i=1}^9 \sum_{j=1}^{ni} y_{ij}$$

後述の (11) を代入すると,

$$N\hat{u} + A\hat{a} + \left\{ \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} y_{ij} - \hat{u} (N-B) \right\} = \sum_{i=1}^9 \sum_{j=1}^{ni} y_{ij}$$

$$\therefore B\hat{u} + A\hat{a} = \sum_{i=1}^9 \sum_{j=1}^{ni} (1 - z_{ij2}) y_{ij} \quad (6)$$

$$\frac{\partial \log L}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^9 \sum_{j=1}^{ni} \{z_{ij1} (y_{ij} - u - a) + z_{ij3} (y_{ij} - u + a)\} = 0 \quad (7)$$

より

$$\begin{aligned} \hat{a} \sum_{i=1}^9 \sum_{j=1}^{ni} (z_{ij1} + z_{ij3}) &= \sum_{i=1}^9 \sum_{j=1}^{ni} (z_{ij1} - z_{ij3}) y_{ij} \\ - \hat{u} \sum_{i=1}^9 \sum_{j=1}^{ni} (z_{ij1} - z_{ij3}) & \\ \hat{a} B + \hat{u} A &= \sum_{i=1}^9 \sum_{j=1}^{ni} w_{ij} y_{ij} \quad (8a) \end{aligned}$$

$$\therefore \hat{a} = \left(\sum_{i=1}^9 \sum_{j=1}^{ni} w_{ij} y_{ij} - A\hat{u} \right) / B \quad (8b)$$

2x(8a) - 1x (6) より

$$\begin{aligned} (A^2 - B^2) \hat{u} &= A \sum_{i=1}^9 \sum_{j=1}^{ni} w_{ij} y_{ij} - B \sum_{i=1}^9 \sum_{j=1}^{ni} (1 - z_{ij2}) y_{ij} \\ \therefore \hat{u} &= \left\{ A \sum_{i=1}^9 \sum_{j=1}^{ni} w_{ij} y_{ij} - B \sum_{i=1}^9 \sum_{j=1}^{ni} (1 - z_{ij2}) y_{ij} \right\} / (A^2 - B^2) \quad (9) \end{aligned}$$

$$\frac{\partial \log L}{\partial d} = \frac{1}{\sigma^2} \{z_{ij2} (y_{ij} - u - d)\} = 0 \quad (10)$$

$$\begin{aligned} (\hat{u} + \hat{d}) \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} &= \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} y_{ij} \\ \hat{u} + \hat{d} &= \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} y_{ij} / \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} \\ \therefore \hat{d} &= \sum_{i=1}^9 \sum_{j=1}^{ni} z_{ij2} y_{ij} / (N - B) - \hat{u} \quad (11) \end{aligned}$$

$$\begin{aligned} \frac{\partial \log L}{\partial \sigma^2} = & -\frac{N}{2} \frac{1}{\sigma^2} - \sum_{i=1}^9 \sum_{j=1}^{ni} \left\{ z_{ij1} (y_{ij} - u - a)^2 / 2 \right. \\ & \left. + z_{ij2} (y_{ij} - u - d)^2 / 2 + z_{ij3} (y_{ij} - u + a)^2 / 2 \right\} \left[-\frac{1}{(\sigma^2)^2} \right] \\ = & 0 \quad (12) \end{aligned}$$

より

$$\begin{aligned} \hat{\sigma}^2 = & \frac{1}{N} \sum_{i=1}^9 \sum_{j=1}^{ni} \left\{ z_{ij1} (y_{ij} - u - a)^2 + z_{ij2} (y_{ij} - u - d)^2 \right. \\ & \left. + z_{ij3} (y_{ij} - u + a)^2 \right\} \quad (13) \end{aligned}$$

ただしここで

$$w_{ij} = z_{ij1} - z_{ij3} \quad (14a)$$

$$A = \sum_i \sum_j^{ni} (z_{ij1} - z_{ij3}) = \sum_i \sum_j^{ni} w_{ij} \quad (14b)$$

$$B = \sum_i \sum_j^{ni} (z_{ij1} + z_{ij3}) \quad (14c)$$

$$N = \sum_i \sum_j^{ni} (z_{ij1} + z_{ij2} + z_{ij3}) = \sum_i \sum_j^{ni} (1) \quad (14d)$$

(5) 反復計算と収束

遺伝効果の推定に最尤法 (maximum likelihood) が使われる点では、QTL 解析は組換え価の推定と同じであるが、後者では原則として各遺伝子型の分離頻度が観察値として直接得られるのに対して、前者では各個体が QTL についてどのような遺伝子型をもっているかが分からない。すなわち分離頻度が欠測値になっている。そこで分離頻度をなんらかの方法で推定しなければならない。その計算には EM アルゴリズムという方法が使われる。EM アルゴリズムは上記の頻度の期待値を求める過程 (E ステップ) と尤度を最大にするパラメータ値を求める過程 (M ステップ) からなる。E は Expectation, M は Maximization の略である。

相加効果, 優性効果, 残差分散が既知であるとしたときに期待される QTL 遺伝子型の分離頻度を求める過程を E ステップとする。そうして得られた分離頻度に基づいて, パラメータの最尤推定値すなわち尤度を最大にする値 (すなわち相加効果, 優性効果, 残差分散) を求める過程が M ステップである。しかし実際にはパラメータは推定すべき対象であり既知ではないので, 適当な初期値を代入してまず 1 回目の E ステップと M ステップの計算を行う。2 回目の計算サイクルに入り, 得られたパラメータ値を使って再び E ステップで分離頻度の期待値を求める。それを用いてパラメータの最尤推定値を求める。以下このように E ステップと M ステップの 2 過程からなる計算をくり返すと, パラメータの推定値は次第にある一定値に収束する。サイクルが 1 回進むことによるパラメータの推定値の差があらかじめ決めた閾値より小さくなったら, 解が収束したとして計算の反復を終了し, 収束時におけるパラメータの値を記録しておく。初期値としては, 相加効果, 優性効果については 0, 定数は F_2 の平均, 残差分散は F_2 の分散を用いるのが反復計算における解の収束を速める上でよい。なお, 反復計算ではより簡単な次の関係を用いることができる。

$$\hat{u} + \hat{a} = \sum_i \sum_j^{ni} z_{ij1} y_{ij} / \sum_i \sum_j^{ni} z_{ij1} \quad (15a)$$

$$\hat{u} + \hat{d} = \sum_i \sum_j^{ni} z_{ij2} y_{ij} / \sum_i \sum_j^{ni} z_{ij2} \quad (15b)$$

$$\hat{u} - \hat{a} = \sum_i \sum_j^{ni} z_{ij3} y_{ij} / \sum_i \sum_j^{ni} z_{ij3} \quad (15c)$$

(6) L O D

$\hat{u}, \hat{a}, \hat{d}, \hat{\sigma}_0^2$ が収束したとき, その解を用いて対数尤度 $\log L_1$ を求める。

これと遺伝効果がまったくない場合を帰無仮説として, $H_0: \hat{u} = u_0, \hat{a} = 0, \hat{d} = 0, \hat{\sigma}_0^2$ とおいたときの対数尤度 $\log L_0$ を求める。すなわち

$$\log L_0 = C + \sum_{i=1}^9 \sum_{j=1}^{ni} (z_{ij1} \log p_{i1} + z_{ij2} \log p_{i2} + z_{ij3} \log p_{i3} - 0.5N \log (2\pi\sigma_0^2) - \sum_{i=1}^9 \sum_{j=1}^{ni} (y_{ij} - u_0)^2 / (2\sigma_0^2)) \quad (16)$$

これより対数を常用対数に変換して

$$LOD = \log_{10} L - \log_{10} L_0 \quad (17)$$

を求める。LOD は logarithm of odds (対数オッズ比) の略である。なお右辺第 1 項および第 2 項に含まれる $C \log_{10} e$ は相殺される。以上までの手順をフローチャートで示す (図 2)。

QTL の位置を染色体 (連鎖群) の端からある幅 (たとえば 1cM) ずつ移動させながらこのような反復計算を行い, 各点での LOD 値を求める。LOD が極大値を示した点を QTL の位置と推定する (図 3)。また推定された QTL 位置におけるパラメータの収束値を, その QTL についてのパラメータの推定値とする。

(7) QTL の有無の判定と位置の信頼区間

QTL が存在すると判断できる LOD の閾値は, マーカー数, マーカー間距離に依存する (表 3)。 F_2 ではマーカーが多い場合でも, 5% 水準で $LOD=3$ を閾値とすれば十分であろう (Rebai ら 1994)。QTL が存在する場合には, LOD 値は供試個体数に比例する。また QTL がマーカー上にある場合には, 1 個体当たりの LOD の期待値は広義の遺伝率の簡単な関数として表される (図 4) (鶴飼 1998)。

LOD 極大値の点の両側で極大値より 1 だけ低い LOD 値を示す染色体上位置 (L_1, L_2) を求め, 区間 L_1-L_2 を support interval とよび, QTL 位置の信頼区間 (の近似値, やや過小評価) として使う (Johan と Ooijen 1992)。シミュレーションの結果によると, 推定精度は QTL の遺伝効果については一般に高いが, QTL 位置については連鎖地図作成におけるマーカー位置ほどの場合ほど高くない。またマーカー数が多くマーカー区間が狭いほど QTL 位置の推定精度は高くなるが, それにも限度があり, マーカーを無限に増やしても QTL 位置の信頼区間は 10cM 以下にはならない。それ以上の精度を必要とする場合には, 準同質系統, 戻し交雑近交系など生物学的な手法を加味し

図2. LOD 値の計算のフローチャート。ただし Hayashi と Ukai(1994)の手順による。近接した2マーカー A, B の区間において、想定されたQTL座の位置を順次ずらしながら、その各点でLOD値を求める。A-B区間が終わったら、次の区間に移る。同じ操作を染色体末端のマーカー区間までくりかえし、LOD値が極大値を示す染色体上の位置を求め、そこをQTLの位置と推定する。その極大値点におけるパラメータ(相加効果、優性効果、残差分散)の値を推定値として採用する。

てQTL解析を行う必要がある(Darvasi ら 1993)。

第2節 F_2 以外の世代のQTL解析

倍加半数体(DH)および戻し交雑世代(BC_1)については、 F_2 における表2の代わりにそれぞれ表4と表5を用いる。その他の計算は基本的には F_2 と同じである。組換え近交系の場合には、最も簡単な自殖の場合でも確率のマトリクスは 27×27 以上の大きさとなり、またマトリクス要素を直接書き下すことが難しく、世代行列の積とし

図3. 染色体上位置に対してプロットしたLODスコア。

染色体上位置は地図距離(cM)を単位として示される。通常LODが極大値を示す点がQTLの位置と推定される。並行な点線はLOD=3としたQTL検出の閾値。パラメータ推定にEMアルゴリズムを採用しているため、マーカー以外の点では、LOD値が一般に過大となる。

図4. 1個体当たりのLOD期待値と広義の遺伝率の関係。

表3. F_2 におけるQTLの有無判定のための5%水準LOD閾値(長さ100cMの1本の染色体につき)
(Rebaiらの表より計算)

マーカー間距離 (cM)	1.0	2.0	2.5	5.0	10.0	20.0	50.0	100.0
マーカー数	101	51	41	21	11	6	3	2
シミュレーション		2.90	2.87	2.48	2.38	2.20	1.96	1.78
Davies近似	3.08	2.89	2.80	2.54	2.31	2.11	1.87	1.83

表 4. QTL に近接する 2 つのマーカー座の遺伝子型が $i(i=1, 2, \dots, 9)$ であるとき QTL 遺伝子型が QQ, Qq, qq である条件付き確率 (倍加半数体 DH)

i: マーカー 遺伝子型	QQ(p_{i1})	qq(p_{i3})
1: AABBB	q_1	q_2
3: AABbb	q_3	q_4
7: aaBBB	q_4	q_3
9: aabbb	q_2	q_1

表 5. QTL に近接する 2 つのマーカー座の遺伝子型が $i(i=1, 2, \dots, 9)$ であるとき QTL 遺伝子型が QQ, Qq, qq である条件付き確率 (戻し交雑世代 BC₁)

i: マーカー 遺伝子型	QQ(p_{i1})	Qq(p_{i2})
1: AABBB	q_1	q_2
2: AABbb	q_3	q_4
4: AaBBB	q_4	q_3
5: AaBbb	q_2	q_1

て計算しなければならない。QTL 解析にはどのような世代を使うのがよいかは、実験の目的や生物種により異なり、一概に言えない。統計遺伝学的な面からだけで判断すれば、優性効果の推定を必要とする場合には共優性マーカーを用いた F₂ 世代が望ましい。DH では優性効果は推定できない。また BC では定数、相加効果、優性効果を分けて推定できない。優性効果の推定が必要でない場合には、DH や遺伝的に固定した F₂ 由来の組換え近交系が薦められる。BC 由来の組換え近交系では、遺伝効果の推定精度が低く不安定である。

第 3 節 区間マッピング法以外の QTL 解析法

分散分析法: DNA 多型マーカーを用いて QTL を検出する簡便な方法として、分散分析法がある (Soller ら, 1976)。すなわちマーカーの遺伝子型により個体を群別し、量的形質の表現型値について一元配置分散分析を行い、群間差を群内差で検定する。この方法は詳細な連鎖地図がすでに作成されていてマーカー間距離が 10cM 以下にできる場合には、QTL の有無を通常の統計ソフトで検定できる便利さがある。連鎖群上に密に並んだマーカーを用意できない場合には使うべきではない。マーカー間距離が小さい場合でもなお、つぎの難点がある。1) QTL 検出の感度が遺伝効果だけでなく、マーカーと QTL の間の地図距離に依存する。マーカー近傍に遺伝効果の低い QTL が存在する場合とマーカーと離れて遺伝効果の大きい QTL がある場合とが区別しにくい。そのため重要な QTL を見逃したり、その遺伝効果を過小評価するおそれがある。2) QTL の位置を正確に決定できない。3)

遺伝効果についてはなんの情報も与えない。これは致命的な欠点である。

重回帰分析法: Haley と Knott (1992) はマーカー遺伝子型から通常の統計手法である重回帰分析法によって QTL の位置と遺伝効果を推定する方法を提案した。この方法にはつぎの利点がある。1) 単一 QTL だけでなく、複数の QTL の場合に容易に拡張できる。2) 関与する全 QTL の分離を同時に考慮したモデルの下に残差変動を減少できる。3) 環境変動が正規分布に従わない場合 (対数正規分布、指数分布など) にも適用できる。また量的形質の表現型が連続分布でなく、計数データ (離散分布) や順序データである場合にも適用できる。4) エピスタシスを含む遺伝モデルへ容易に拡張できる。5) 遺伝子型 $\times$ 環境相互作用や実験計画法における反復ブロックを考慮することにより、残差を減少できる。6) いくつかの異なる世代のデータを統合して解析できる。7) 区間マッピング法より計算が簡便で迅速である。

複合区間マッピング法 (Composite Interval Mapping): Zeng (1994) による方法で、対象とする QTL 以外の効果を予め重回帰分析法で除き、残差について区間マッピング法を適用するものである。複数の QTL が近接している場合に、通常の区間マッピング法では LOD のピークがひとつに重なってしまい、QTL 間の識別が難しく、また QTL の位置や遺伝効果の推定の精度が低くなる。それに対してこの方法では、個々の QTL を識別しやすい。区間マッピング法にくらべて LOD 値は、相加効果が同符号の QTL の場合には小さくなるが、逆の場合には大きくなる。また対象とする QTL と連鎖していない QTL の効果を重回帰分析法で除くと、対象 QTL の解析における残差変動が小さくなり、それだけ LOD 値が高くなり QTL が検出しやすくなる。これを半複合区間マッピング法という。

第 4 節 QTL 解析についての留意点

- 1) 遺伝効果推定: QTL 解析において質的形質や DNA マーカーのマッピングと同じ扱いで QTL の連鎖群上の位置だけを求めている論文が少なくないが、それではせっかく QTL 解析を行ったメリットが半減する。QTL 解析では、各 QTL の相加効果、優性効果、優性度、環境分散の対する遺伝分散の比、エピスタシス、多面発現などについて推定値を求め、十分な考察を行うことが重要である。
- 2) 環境変動: 量的形質は、それを計測した環境条件によって変化する。結果の発表においては、栽培条件 (地名、時期、施肥、土壌)、気象条件、計測法などを実験結果に必ず付記する必要がある。環境条件による変動を確実に調べてはじめて QTL 解析が完成するといえる。
- 3) 全遺伝分散: QTL 解析とともに、従来の統計遺伝学的解析を同時に行い、遺伝分散、相加分散、優性分散を求め、QTL 解析から得られた QTL 別の遺伝効果から

期待される遺伝分散, 相加分散, 優性分散と比較することにより, 検出し残した QTL の有無や QTL 間エピスタシスの有無について推察することが望ましい. QTL 効果と表現型分散をくらべることは, 広義の遺伝率が高い形質でないかぎりあまり意味がない.

- 4) 形質構成要素: 収量, 草丈, 早晩性などの QTL 解析では, その構成要素に基づいて解析をすることも重要である. オオムギを例にとれば, 収量については収量構成要素, 草丈については節数, 節間長, 細胞数, 細胞長, 早晩性については播性, 日長感応性, 狭義の早晩性などと関連づけることが, 個々の QTL の役割を理解するのに役立つ. それにより 1 回の QTL 解析において, 同時に分離する QTL の数を減らすことができれば, 解析精度があがることも期待できる.
- 5) 遺伝子型×環境交互作用: 年次や場所による変動の大きい形質では, 複数の年次, 場所で同じ実験を行い, その総合的解析から, 年次または場所による QTL 解析結果の差異を調べ, また QTL 別に遺伝子型 x 環境交互作用を調べることが重要である (Sari-Gorla ら 1997).
- 6) 計測精度: 量的形質の計測法をいままで以上に精密で客観的なものにする必要がある. 従来の統計遺伝学的手法と異なり, 0.1 程度の低い遺伝率の形質でも充分 QTL 解析が可能であるので, 形質の計測法もそれにみあう精度を要求される. 目視による観察や, スコアによる観察値の表現は出来るだけ避ける.
- 7) 遺伝率: QTL 解析にさきだち形質の遺伝率ができるだけ高くなるように, 実験上の工夫が必要である. それには, 実験の精度を高める, 形質発現の環境条件を均一にして環境変動を抑える, 実験区の反復を増やす, 個体単位ではなく系統または集団単位で値を求める, などが有効である.
- 8) 形質間相関: 晩生が早生よりも多収である場合には, 早晩性に関与する QTL が収量の QTL としても検出されてくることになる. 解析にさきだち, あらかじめ形質間相関が予想される場合には, 目的外の形質については QTL における分離がないような交配組合せを用いて解析するか, 統計学的に目的外形質値に対する目的形質の回帰を求め, その残差について QTL 解析することが考えられる.
- 9) 遺伝子座: 交配親の間で分離していない QTL については, 当然ながら QTL 解析においてもなんの情報も得られない. 目的とする形質に関与する座を出来るだけ残りなく検出するには, 多くの異なる交配組合せについ

て QTL 解析を行う必要がある.

- 10) 致死因子: 遠縁交雑などで分離世代において致死因子などの分離比を乱す要因が含まれると, 近傍の QTL の検出感度が下がり, 相加効果, 優性効果などの遺伝効果の推定値が低くなる. QTL 解析の結果には, 分離比を乱すような要因の有無を併記することが望ましい.
- 11) データベース: QTL 解析結果は, 作物別に原データを含めて結果をデータベース化しておくことが, 今後国内外の研究者の結果を統合して, 量的形質のマッピングや遺伝効果の推定を行う上できわめて重要である.

付記: F_2 , BC, DH, 組換え近交系など各種の世代についての QTL 解析のためのパソコン用プログラム MAPL97 を web 上の下記 URL からダウンロードできる. 17 インチモニターと DOS/V 系マシンの計算環境があれば, Win95/98 上で利用できる.

URL: <http://peach.ab.a.u-tokyo.ac.jp/~ukai/>

引用文献

- Darvasi, A., A. Weinreb, V. Minke, J.I. Weller and M. Soller (1993) *Genetics* 134:943–951.
- Falconer, D.S. and T.F.C. MacKey (1996) *Introduction to Quantitative Genetics*. 4th Edition. Longman.
- Hayashi, T. and Y. Ukai (1994) *Theor. Appl. Genet.* 87:1021–1027.
- Haley, C.S. and Knott, S.A. (1992) *Heredity* 69:315–324.
- Knapp, S.J. (1994) In: *DNA-based Markers in Plants*. ed. by Phillips R.L. and I.K. Vasil. Kluwer.
- Lander, E.S. and Botstein, D. (1989) *Genetics* 121: 185–199.
- Liu, B.H. (1998) *Statistical Genomics*. CRC Press.
- Rebai, A., B. Goffinet and B. Mangin (1994) *Genetics* 138:235–240.
- Sari-Gorla, M., Calinski, T., Z. Kaczmarek and P. Krajewski (1997) *Heredity* 78:146–157.
- Sax, K. (1923) *Genetics* 8:552–560.
- Soller, M., Brody, T. and Genizi, M.A. (1976) *Theor. Appl. Genet.* 47:35–39.
- 鶴飼保雄 (1998) 育種学雑誌 48 (別冊 2) (印刷中)
- van Ooijen J.W. (1992) *Theor. Appl. Genet.* 84:803–811.
- Zeng, Z. (1994) *Genetics* 136:1457–1468.