

DOI: 10.3724/SP.J.1006.2010.00918

数量性状基因定位研究中若干常见问题的分析与解答

李慧慧 张鲁燕 王建康*

中国农业科学院作物科学研究所 / 农作物基因资源与基因改良国家重大科学工程 / 国际小麦玉米改良中心中国办事处, 北京 100081

摘要: QTL 作图是基因精细定位、克隆以及有效开展分子育种的基础, 在利用 QTL 作图开展数量性状基因定位研究的过程中经常会碰到一些问题, 与统计方法有关的一些问题包括 LOD 的统计学意义是什么? 检测 QTL 的可信度和 LOD 临界值的关系是什么? 如何评价不同的 QTL 作图方法? 提高 QTL 检测效率的途径有哪些? 与遗传参数估计有关的一些问题包括 QTL 的贡献率是如何计算出来的? 如何确定 QTL 有利等位基因的来源? 选择基因型分析的有效性如何? 复合性状是否适宜于 QTL 作图? 与作图群体及遗传图谱有关的一些问题包括 QTL 作图群体中表型数据是否要求服从正态分布? 加密标记是否可以显著提高 QTL 检测功效? 缺失分子标记对 QTL 作图有什么影响? 奇异分离标记对 QTL 作图有什么影响? 文章试图结合笔者多年研究工作对这 12 个有共性的常见问题做出分析和解答, 以供科研工作者参考。

关键词: 数量性状; QTL 作图; 完备区间作图; 似然比检验; 功效分析

Analysis and Answers to Frequently Asked Questions in Quantitative Trait Locus Mapping

LI Hui-Hui, ZHANG Lu-Yan, and WANG Jian-Kang*

Institute of Crop Sciences / National Key Facility for Crop Gene Resources and Genetic Improvement / CIMMYT China Office, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract: QTL mapping is an important step in gene fine mapping, map-based cloning, and the efficient use of gene information in molecular breeding. Questions are frequently met and asked in the application of QTL mapping in practical genetic populations. Questions related to statistical method of QTL mapping are: what does LOD score mean? What is the relationship between the reliability of detected QTL and the LOD threshold? How to evaluate different QTL mapping methods? How to improve the QTL detection power? Questions related to genetic parameter estimation are: how to calculate the phenotypic variance explained by each detected QTL? How to determine the source of favorable alleles at detected QTL? How efficient is the selective genotyping? Can composite traits be used in QTL mapping? Questions related to linkage map and mapping populations are: does the phenotype of a trait in interest have to follow a normal distribution? Does the increase in marker density greatly improve QTL mapping power? What effects will missing markers have in QTL mapping? What effects will segregation distortion have in QTL mapping? Our objective in this paper was to give an analysis and answer to each of the 12 frequently asked questions, based on our studies in past several years.

Keywords: Quantitative trait; QTL mapping; Inclusive composite interval mapping; Likelihood ratio test; Power analysis

经典数量遗传学建立在多基因假说基础之上, 把控制数量性状的基因作为一个整体, 重点研究各种遗传效应与遗传方差的分解和估计, 不区分个别基因在表型效应上的差异^[1-3]。分子标记连锁图谱的大量出现, 使我们可以像研究质量性状基因一样研究数量性状基因, 也可以把单个数量性状基因

(quantitative trait gene or locus, 简称 QTL)定位在染色体上, 并估计其遗传效应, 这一过程称为 QTL 作图或定位^[4]。QTL 作图是基因精细定位和克隆的基础, 目前已成为数量性状遗传研究的常用方法。QTL 定位结果可以帮助育种家获得目标性状的遗传信息, 借助与 QTL 连锁的分子标记在育种群体中跟踪和选

本研究由国家自然科学基金项目(30771351)资助。

* 通讯作者(Corresponding author): 王建康, E-mail: wangjk@caas.net.cn; jkwang@cgiar.org; Tel: 010-82105846

第一作者联系方式: E-mail: lihuihui@caas.net.cn; Tel: 010-82106038

Received(收稿日期): 2010-01-18; Accepted(接受日期): 2010-02-28.

择有利等位基因, 提高选择的准确性和预见性。但是, 在利用 QTL 作图开展遗传研究的过程中也经常碰到一些问题, 这些问题大致可分为有关作图统计方法、有关遗传参数估计、和有关作图群体及连锁图谱等三大类。笔者试图结合我们近些年的研究工作对具有共性的 12 个问题做出分析和解答, 供广大科研工作者在利用 QTL 作图开展遗传研究时参考。

1 QTL 作图中的统计学问题

1.1 LOD 的统计学意义是什么?

区间作图(Interval mapping, 简称 IM)^[5]、复合区间作图(composite interval mapping, 简称 CIM)^[6]和完备区间作图(Inclusive composite interval mapping, 简称 ICIM)^[4,7-9]均利用极大似然估计原理, 通过一维扫描在全基因组上逐点检测 QTL 的存在。检验的零假设(H_0)是扫描位点上不存在 QTL, 备择假设(H_A)是扫描位点上存在一个 QTL, 以似然比统计量(LRT)来判断扫描位点上是否有 QTL 存在。

$$LRT = -2\ln\left(\frac{L_0}{L_A}\right),$$

其中, \ln 是自然对数函数, L_0 是 H_0 下似然函数的极大值, L_A 是 H_A 下似然函数的极大值。似然比统计量大多数情况下有很好的统计学性质, 即渐近服从 χ^2 分布, 其自由度为两种假设下独立变量个数之差, 因此容易确定给定显著性水平下检验统计量的临界值。一些研究表明 QTL 作图时的似然比统计量可能不服从单一 χ^2 分布, 但却可以用 2 个不同自由度的混合 χ^2 分布来近似^[6,10], 因此也可以确定给定显著性水平下检验统计量的临界值。

Lander 和 Botstein 在提出 IM 的同时也提出利用人类遗传研究中常用的 LOD (likelihood of odd) 值作为检验 QTL 是否存在的标准^[5]。LOD 值被定义为极大似然函数比的常用对数。

$$LOD = \log\left(\frac{L_A}{L_0}\right) = \log(L_A) - \log(L_0)$$

因此, 若 L_A 是 L_0 的 10 倍, 则 $LOD=1$; 如果 L_A 是 L_0 的 100 倍, 则 $LOD=2$; 如果 L_A 是 L_0 的 1 000 倍, 则 $LOD=3$ 。根据 LOD 值和 LRT 值的定义, 不难得到它们有如下关系。

$$LOD = \frac{LRT}{2\ln(10)} \approx \frac{LRT}{4.6052} \text{ 或 } LRT \approx 4.6052LOD。$$

1.2 检测 QTL 的可信度和 LOD 临界值之间的关系是什么?

与其他假设检验一样, QTL 作图中的检验也可

产生 4 种结果(图 1)。当一个位点上没有 QTL, 通过测验却错误地判断有 QTL 存在, 称这种现象为假阳性; 当一个位点上有 QTL, 通过测验判断有 QTL 存在, 称这种现象为真阳性; 当一个位点上有 QTL, 通过测验却错误地判断没有 QTL 存在, 称这种现象为假阴性; 当一个位点上没有 QTL, 通过测验判断没有 QTL 存在, 称这种现象为真阴性。假阳性和假阴性是统计测验中的两类错误, 犯假阳性错误的概率(α)等于 H_0 为真时被拒绝的概率,

$$\alpha = P\{\text{拒绝 } H_0 | H_0 \text{ 为真}\} = P\{\text{假阳性}\} / [P\{\text{假阳性}\} + P\{\text{真阳性}\}]。$$

犯假阴性错误的概率(β)等于 H_0 为假时未被拒绝的概率,

$$\beta = P\{\text{未拒绝 } H_0 | H_0 \text{ 为假}\} = P\{\text{假阴性}\} / [P\{\text{假阴性}\} + P\{\text{真阳性}\}]。$$

对于特定的检验方法, 在一定的试验精确度下, 降低 α 则会提高 β , 降低 β 则会提高 α , 同时降低 α 和 β 的途径是提高试验精确度和增加样本量。 α 一般在检验前设定, 对一定的检验方法, 给定 α 后 β 也是确定的。但是除 t 测验外, 大多数统计假设检验给定 α 下的 β 难以用代数解析式表示出来。

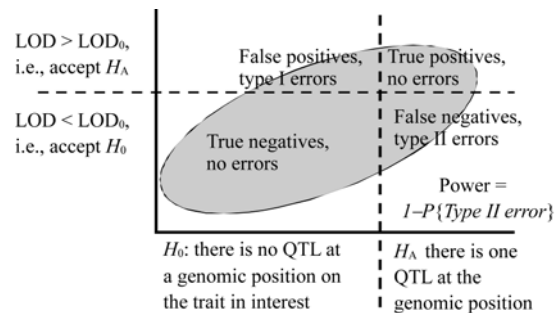


图 1 QTL 定位中假设检验的两类错误

Fig. 1 Two types of error in hypothesis test in QTL mapping
阴影部分表示检验统计量 LOD 值的分布, LOD_0 表示临界值。

The shadow area represents the distribution of the test statistic LOD score, and LOD_0 is the threshold value.

表 1 前 5 列给出 LOD 值 1.0~5.0 时对应的 LRT 值以及 3 种自由度下一次检验中犯假阳性错误的概率, 有时也叫做显著性概率; 后 4 列给出显著性概率 0.1~0.0001 时对应的 LOD 临界值。以自由度 2 为例, $LOD=1.30$ 对应的显著性概率为 0.05, 如果只做一次假设检验, 采用这样的 LOD 临界值就能保证假阳性的概率不超过 0.05, 即鉴定出的 QTL 是假阳性的概率低于 0.05, 或者说鉴定出的 QTL 为真的概率超过 0.95。基于区间测验的 QTL 作图, 一般在基因组内所有染色体上按一定步长逐点检验 QTL 的存在,

表 1 3 种自由度下不同 LOD 值对应的 α 和不同 α 下对应的 LOD 临界值
 Table 1 Corresponding probabilities of type I error to LOD and the corresponding LOD threshold to the significance probabilities under three degrees of freedom

LOD	LRT	α			α	LOD		
		$df=1$	$df=2$	$df=3$		$df=1$	$df=2$	$df=3$
1.00	4.61	0.031786	0.099759	0.202686	0.1	0.59	1.00	1.36
2.00	9.21	0.002407	0.010002	0.026625	0.05	0.83	1.30	1.70
3.00	13.82	0.000201	0.000998	0.003161	0.01	1.44	2.00	2.46
4.00	18.42	0.000018	0.000100	0.000360	0.001	2.35	3.00	3.53
5.00	23.03	0.000002	0.000010	0.000040	0.0001	3.29	4.00	4.58

而这些检验又不是完全独立的, 因此也很难确定一次检验的显著性水平所对应的全局 α 。例如, 如果每次检验均采用临界值 $LOD=3$, 当自由度为 2 时每个扫描位置对应的 α 近似为 0.001, 对于 QTL 作图需要知道的是全基因组检验后的 α 。由于 QTL 作图的复杂性, 如涉及到多次非独立假设检验, 不同物种有不同大小的基因组, 不同作图群体有不同的标记数, 零假设下似然比检验统计量服从什么样的渐近分布尚无定论, 因此难以准确确定 LOD 临界值对应的一次检验和全局 α , 但一些非参数统计方法已用于给定全局 α 后 LOD 临界值的确定^[11-12]。一般认为采用 2~3 的 LOD 临界值可以把全局 α 控制在 0.05 以内, 在显性 QTL 和互作 QTL 作图中, 似然比统计量有较大的自由度, 还可适当考虑采用较高的临界值, 如 3~4。但通过理论或模拟, 进一步明确 QTL 作图中似然比检验统计量的渐近分布还是必要的, 笔者推测渐近分布可能与染色体条数、每条染色体长度、标记密度和 QTL 遗传效应类型等因素有关。

与所有假设检验一样, 采用较高的 LOD 临界值会更好控制假 QTL 的发生, 同时遗传效应较小的真 QTL 却不易被检测出来。如何平衡两类错误的概率, 这不仅仅是统计学上的问题, 还要考虑到具体的研究目标。如果 QTL 作图只是初步确定基因在染色体上的位置, 然后根据作图结果构建其他次级群体对检测到的 QTL 精细定位、甚至图位克隆, 然后开展转基因工作, 则几乎不允许假 QTL 的发生。此时要适当提高检验 QTL 时的 LOD 临界值, 保证后续研究中 QTL 的可靠性。另一方面, 如果研究目标是把 QTL 作图结果用于标记辅助选择聚合育种, 这时只有尽可能多地检测出控制育种目标性状的 QTL, 才能保证对所有控制育种性状的基因的选择, 因此有必要适当降低检验 QTL 时的 LOD 临界值, 使遗传效应较小的 QTL 也有机会被检测出来。此时即使有一些假 QTL 的存在, 也不至于造成很大的损失。

1.3 如何评价不同的 QTL 作图方法?

图 2 给出模拟的加倍单倍体(DH)群体中 IM、CIM 和 ICIM 三种方法的 LOD 值和估计遗传效应的曲线。遗传模型中包含 7 个 QTL, 第 1 染色体上 1 个, 第 2 和第 4 染色体上分别有 2 个相引相连锁 QTL, 第 3 染色体上有 2 个互斥相连锁 QTL, 所有 QTL 遗传效应的绝对值均为 1, 效应的方向如图 2 箭头所示。从图中我们可以很直观地看到不同 QTL 作图方法找到的 QTL 不尽相同, IM 无法准确定位连锁 QTL, CIM 无法准确定位互斥连锁的 QTL。统计检验的功效定义为 $1-\beta$, 即发现真 QTL 的概率, 常用来比较不同统计方法的有效性。QTL 作图包含着复杂的统计假设检验, 难以从理论上推导出不同 QTL 作图方法的统计功效。一般采用模拟方法比较不同方法 QTL 检测功效和假阳性的大小, 功效高同时又具有较低假阳性在统计上就是较好的方法。

我们用独立遗传模型说明功效的计算(表 2), 假定基因组包含 10 条染色体, 每条长 160 cM 且均匀分布 17 个标记, 8 个 QTL(IQ1~IQ8)分别位于前 8 条染色体, 另外 2 条为空白染色体。为简单起见, 我们用 IQ2、IQ5、IQ6 和 IQ7 说明功效的计算, 4 个 QTL 分别解释 2%、5%、10% 和 20% 的表型变异(即 PVE)。假定表型方差为 1.0, 这样加性遗传效应等于 PVE 的平方根(表 2)。模拟 4 个 DH 群体, 群体大小均为 200, 将每个群体中 ICIM 和 IM 检测到的所有超过 LOD 临界值 2.5 的 QTL 列于表 3, 用于统计每个 QTL 检测功效以及假阳性。以第一个模拟群体为例, ICIM 检测到 5 个 QTL, 其中 3 个分别位于第 5、6 和 7 条染色体的 51.10、60.00 和 70.00 cM 处, 分别落在了 IQ5、IQ6 和 IQ7 的 10 cM 置信区间内(表 3), 因此当置信区间为 10 cM 时把 IQ5、IQ6 和 IQ7 的统计功效加 1; 第 2 条染色体上检测到一个 QTL, 与 IQ2 的距离为 6.90 cM, 不在 IQ2 的 10 cM 置信区间内, 第 7 条染色体上 40 cM 处检测到一个 QTL, 既

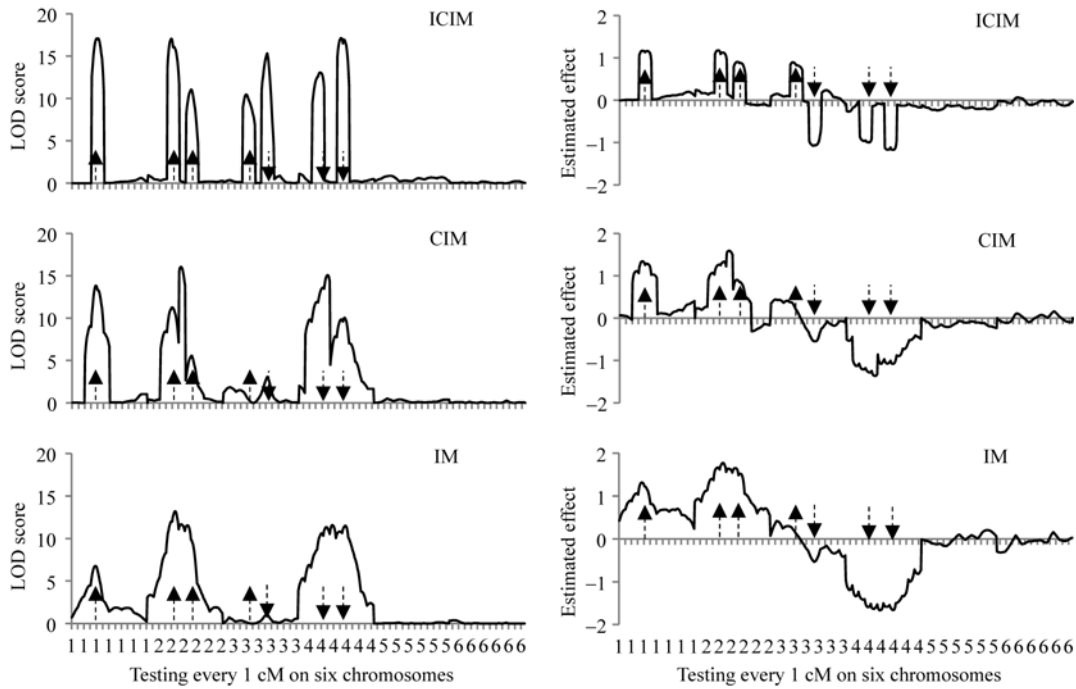


图 2 200 个加倍单倍体家系群体中区间作图、复合区间作图和完备区间作图的比较

Fig. 2 Comparison of interval mapping, composite interval mapping, and inclusive composite interval mapping in a simulated population with 200 doubled haploid lines

ICIM : 完备区间作图; CIM : 复合区间作图; IM : 区间作图。假定一个基因组包含 6 条染色体, 每条长度为 120 cM, 等距离分布 13 个标记, 第 1 条染色体上 35 cM 处存在 1 个 QTL, 第 2、3、4 条染色体上都在 35 cM 和 68 cM 处存在 2 个 QTL, 图中箭头表示 QTL 的近似位置, 向上箭头表示 QTL 的加性效应为正, 向下箭头表示 QTL 的加性效应为负, 效应的绝对值均为 1。

A genome with 6 chromosomes was assumed, each of 120 cM and evenly distributed with 13 markers. One QTL was located at 35 cM on chromosome 1, and two QTLs were located at 35 cM and 68 cM on chromosomes 2, 3, and 4. Arrows point to the approximate QTL positions. Upward arrows indicate the QTLs have positive effects, while downward arrows indicate the QTLs have negative effects. The absolute genetic effect is 1 for all QTLs.

表 2 8 个独立遗传 QTL 在基因组中的位置、加性遗传效应、解释表型变异的百分数以及 10 cM 和 20 cM 置信区间
Table 2 Eight independent QTLs and their chromosomal positions, additive effects and explained phenotypic variation (%), along with the 10 cM and 20 cM confidence intervals (CI)

QTL	染色体 Chromosome	位置 Position (cM)	加性效应 Additive	贡献率 PVE (%)	置信区间 Confidence interval (CI)	
					CI=10 cM	CI=20 cM
IQ1	1	25	0.1000	1	(20 cM, 30 cM)	(15 cM, 35 cM)
IQ2	2	32	0.1414	2	(27 cM, 37 cM)	(22 cM, 42 cM)
IQ3	3	39	0.1732	3	(34 cM, 44 cM)	(29 cM, 49 cM)
IQ4	4	46	0.2000	4	(41 cM, 51 cM)	(36 cM, 56 cM)
IQ5	5	53	0.2236	5	(48 cM, 58 cM)	(43 cM, 63 cM)
IQ6	6	60	0.3162	10	(55 cM, 65 cM)	(50 cM, 70 cM)
IQ7	7	67	0.4472	20	(62 cM, 72 cM)	(57 cM, 77 cM)
IQ8	8	74	0.5477	30	(69 cM, 79 cM)	(64 cM, 84 cM)

不在 IQ7 的 10 cM 置信区间内也不在其 20 cM 置信区间内, 因此判定为假阳性。同理, IM 发现 4 个 QTL, 其中两个分别位于第 6 和第 7 染色体的 60.00 cM 和 70.00 cM 处, 分别落在 IQ6 和 IQ7 的 10 cM 置信区间内, 这 2 个 QTL 对应的功效加 1, 另外 2 个不在这 4 个 QTL 的 10 cM 置信区间内, 因此被判定为假阳性。

按照上述算法, 对于 4 次模拟来说, 在 LOD 临

界值 2.5 下当置信区间为 10 cM 时(表 3), ICIM 检测 IQ2、IQ5、IQ6 和 IQ7 的功效分别为 2、1、3 和 2, 即 IQ2 在 4 次模拟中被检测到 2 次, IQ5 被检测到 1 次, IQ6 被检测到 3 次, IQ7 被检测到 2 次; IM 检测 4 个独立 QTL 的功效分别为 1、0、3 和 2。ICIM 检测到的 16 个 QTL 中有 8 个假阳性, IM 检测到的 13 个 QTL 中有 7 个假阳性。当置信区间为 20 cM 时(表 3), ICIM

表 3 利用 4 个模拟 DH 群体计算完备区间作图和区间作图两种方法的检测功效和假阳性(QTL 置信区间长度分别是 10 cM 或 20 cM)
Table 3 Calculation of power and false QTL in four simulated DH populations for ICIM and IM (the confidence interval of each predefined QTL is 10 or 20 cM)

群体 Pop.	完备区间作图 Inclusive composite interval mapping						
	染色体 Chromosome	位置 Position	加性效应 Additive	LOD	贡献率 PVE (%)	CI=10 cM	CI=20 cM
1	2	25.10	0.19	2.56	3.48	False QTL	IQ2
	5	51.10	0.29	6.05	8.14	IQ5	IQ5
	6	60.00	0.30	6.72	8.86	IQ6	IQ6
	7	40.00	0.20	2.94	3.71	False QTL	False QTL
	7	70.00	0.42	11.87	16.64	IQ7	IQ7
2	2	30.50	0.27	5.35	7.78	IQ2	IQ2
	5	45.00	0.27	5.25	7.94	False QTL	IQ5
	6	59.10	0.26	4.94	7.50	IQ6	IQ6
	7	59.40	0.38	9.84	15.61	False QTL	IQ7
3	2	30.00	0.21	2.50	3.96	IQ2	IQ2
	6	55.40	0.29	4.47	7.81	IQ6	IQ6
	7	70.00	0.28	4.42	7.14	IQ7	IQ7
	7	90.00	0.25	3.39	5.41	False QTL	False QTL
4	6	69.30	0.22	2.65	4.99	False QTL	IQ6
	7	60.00	0.33	5.87	11.34	False QTL	IQ7
	10	38.90	-0.23	2.92	5.70	False QTL	False QTL
群体 Pop.	区间作图 Interval mapping						
	染色体 Chromosome	位置 Position	加性效应 Additive	LOD	贡献率 PVE (%)	CI=10 cM	CI=20 cM
1	5	47.10	0.26	2.71	6.74	False QTL	IQ5
	6	6.70	0.25	2.51	6.15	False QTL	False QTL
	6	60.00	0.30	3.99	8.77	IQ6	IQ6
	7	70.00	0.52	13.39	26.53	IQ7	IQ7
2	2	30.00	0.30	4.38	9.59	IQ2	IQ2
	6	60.00	0.28	3.81	8.39	IQ6	IQ6
	7	63.90	0.43	8.90	19.44	IQ7	IQ7
3	6	56.20	0.31	3.66	8.81	IQ6	IQ6
	7	72.60	0.48	9.55	21.01	False QTL	IQ7
	7	86.30	0.47	8.94	19.93	False QTL	False QTL
	7	108.00	0.38	5.77	13.04	False QTL	False QTL
4	7	60.00	0.31	4.57	9.99	False QTL	IQ7
	10	38.70	-0.26	3.04	7.18	False QTL	False QTL

检测 IQ2、IQ5、IQ6 和 IQ7 的功效分别为 3、2、4 和 4, IM 的检测功效分别为 1、1、3 和 4; ICIM 检测到的 16 个 QTL 中有 3 个假阳性, IM 检测到的 13 个 QTL 中有 4 个假阳性。因此通过 100 甚至 1 000 次的模拟, 就能得到准确的检测功效和假阳性率, 从而判断不同 QTL 作图方法的有效性。

1.4 提高 QTL 检测功效的途径有哪些?

统计上提高假设检验功效的途径主要是增加样本量和减小试验误差, 对 QTL 作图来说就是增大作

图群体、减小表型测定时的误差(即提高性状的遗传力)。以 ICIM 为例, 增大群体对提高 QTL 检测功效的作用是明显的(图 3)。对大小为 100、200 和 400 的群体来说, 把 PVE = 4% 的 QTL 定位到 10 cM 置信区间内的概率分别为 29%、67% 和 91%; 而把 PVE = 10% 的 QTL 定位到 10 cM 置信区间内的概率分别为 79%、97% 和 100%。

减小表型误差则间接提高单个 QTL 的贡献率^[13], 如果通过降低表型误差把 QTL 对表型的贡献率由

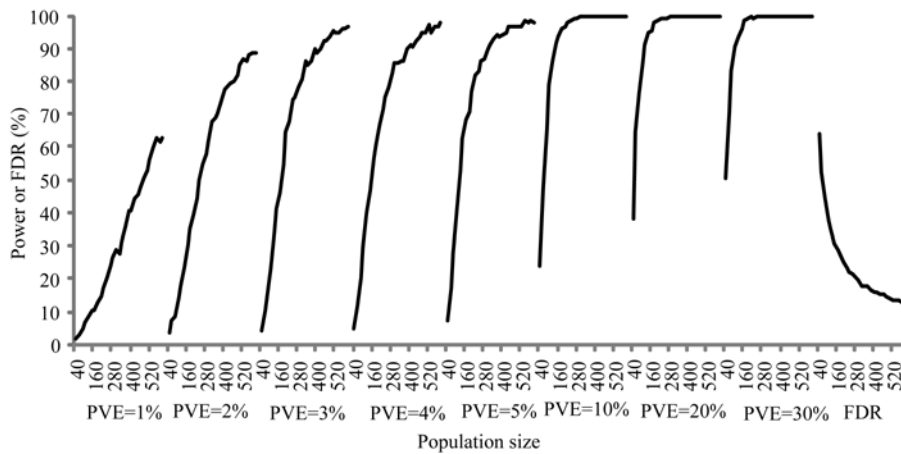


图 3 QTL 的检测功效与群体大小的关系

Fig. 3 Relationship between QTL detection power and mapping population size

4% 提高到 5%，对大小为 100、200、400 的群体来说，检测功效则分别由 29%、67% 和 91% 提高到 44%、77% 和 94%。因此，QTL 作图研究中作图群体在资源允许的条件下要尽可能地大，同时表型鉴定时要尽量减小随机误差。当然对于环境影响大并具有较大基因型和环境互作的性状，还要在多地/多年份进行表型鉴定。

降低表型变异也可间接提高 PVE，从而提高 QTL 的检测功效。遗传研究中近等基因系和染色体片段置换系^[14-15]都是通过这种途径提高遗传分析可靠性的。例如，假定某群体中 3 个独立遗传 QTL 的遗传方差分别为 0.1、0.2 和 0.3，误差方差为 0.4，因此 3 个 QTL 分别解释 10%、20% 和 30% 的表型变异，在这 3 个 QTL 的近等基因系中假定 QTL 的遗传效应和误差方差保持不变，那么 3 个 QTL 近等基因系群体的表型方差分别为 0.5、0.6 和 0.7，3 个 QTL 解释的表型变异则增加到 20%、33% 和 43%，因此在这 3 个近等基因系群体中 QTL 的定位研究将更加有效。

2 QTL 作图中的遗传学问题

2.1 QTL 的贡献率是如何计算出来的？

一个 QTL 所解释表型变异(PVE，又称为 QTL 的贡献率)的计算公式是 $PVE = V_G/V_P \times 100\%$ ，其中 V_G 是 QTL 的遗传方差， V_P 是表型方差。在一个无奇异分离的群体中，QTL 的遗传方差只依赖于 QTL 的遗传效应，效应大的 QTL 同时也具有较高的 PVE；如果存在奇异分离，QTL 的遗传方差除依赖于 QTL 的遗传效应外，还依赖于基因型频率，这时效应大的 QTL，其 PVE 不一定高。以 F_2 群体为例，假定 QTL 三种基因型 QQ、Qq 和 qq 的频率为 f_2 、 f_1 和 f_0 ，

QTL 的加显性效应为 a 和 d ，

$$V_G = [f_2 + f_0 - (f_2 - f_0)^2] a^2 - 2f_1(f_2 - f_0) ad + (f_1 - f_1^2) d^2 \quad [1]$$

例如，一个 QTL 的加显性效应为 1，即完全显性，当 $f_2=0.25$ 、 $f_1=0.5$ 和 $f_0=0.25$ 即不存在奇异分离时， $V_G=0.75$ ；当 $f_2=0.4$ 、 $f_1=0.5$ 和 $f_0=0.1$ 时， $V_G=0.36$ ；当 $f_2=0.5$ 、 $f_1=0.5$ 和 $f_0=0.0$ 时， $V_G=0$ ，这时 QTL 就检测不出来了。大多数作图软件中，QTL 的遗传方差都是根据 3 种 QTL 基因型的观测频率计算出，因此 QTL 作图结果中，出现遗传效应大而 PVE 偏低的情形也是正常的。

QTL 间的互斥连锁甚至可能造成单个 QTL 的 PVE 或两个 QTL 的 PVE 之和超过 100% 的情形。假定亲本的基因型为 $Q_1Q_1Q_2Q_2$ 和 $q_1q_1q_2q_2$ ， a_1 和 a_2 分别为 2 个 QTL 的加性效应，2 个位点间的重组率为 r 。在双亲衍生的 DH 群体中，4 种基因型的频率和基因型值如表 4。RIL 群体中，只需将 r 用 $R = 2r/(1+2r)$ 代替。2 个 QTL 各自的遗传方差分别为 $V_{G_1} = a_1^2$ 和 $V_{G_2} = a_2^2$ ，总遗传方差

$$\begin{aligned} V_G &= \frac{1}{2}(1-r)(a_1 + a_2)^2 + \frac{1}{2}(a_1 - a_2)^2 + \frac{1}{2}(a_1 - a_2)^2 \\ &\quad + \frac{1}{2}(1-r)(a_1 + a_2)^2 \\ &= a_1^2 + a_2^2 + 2(1-2r)a_1a_2 \end{aligned}$$

因此，只有在 $r=0.5$ 的情况下，才有 $V_G = V_{G_1} + V_{G_2}$ 。当 2 个 QTL 间存在连锁并且 a_1 和 a_2 的效应方向相反时(即处于互斥连锁状态)， $V_G < V_{G_1} + V_{G_2}$ ，造成 2 个 QTL 的 PVE 之和甚至单个 QTL 的 PVE 大于 100% 的情况。例如，不考虑随机

表 4 两个连锁 QTL 在 DH 群体中 4 种基因型的频率和基因型值, a_1 代表 Q_1-q_1 的加性遗传效应, a_2 代表 Q_2-q_2 的加性遗传效应
Table 4 Frequencies and genotypic values of four genotypes at two linked QTLs in a DH mapping population, where a_1 is the additive effect of Q_1-q_1 , and a_2 is the additive effect of Q_2-q_2

基因型 Genotype	频率 Frequency	基因型值 Genotypic value
$Q_1Q_1Q_2Q_2$	$\frac{1}{2}(1-r)$	$a_1 + a_2$
$Q_1Q_1q_2q_2$	$\frac{1}{2}r$	$a_1 - a_2$
$q_1q_1Q_2Q_2$	$\frac{1}{2}r$	$-a_1 + a_2$
$q_1q_1q_2q_2$	$\frac{1}{2}(1-r)$	$-a_1 - a_2$

环境误差和其他遗传变异, 当 $a_1=1.0$, $a_2=-1.0$, $r=0.1$, $V_e=0.4$ 时, $V_{G_1}=V_{G_2}=1$, $V_G=0.4$, 表型方差 $V_P=0.8$, 故两个 QTL 理论 PVE 均为 125%。在这样的遗传模型下, 用 ICIM 仍然可以较准确地定位到两个紧密连锁的 QTL, 效应的估计也是近似无偏的(图 4)。因此当 QTL 间存在连锁时, 2 个 QTL 所解释的遗传方差不等于它们遗传方差之和, 即方差不具有可加性, 因此由方差计算出的 PVE 之间也不具可加性。

2.2 如何判断 QTL 有利等位基因的来源?

确定每个 QTL 上有利等位基因的来源是把作图结果应用于分子育种的前提^[15]。假设一个 QTL 上有

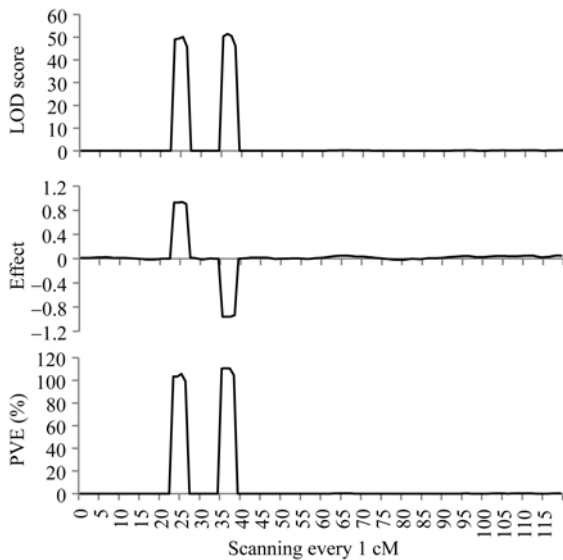


图 4 一个模拟的包含 200 个加倍单倍体家系作图群体
Fig. 4 QTL mapping in a simulated population of 200 doubled haploid lines

一条长度为 120 cM 的染色体上有一对重组率为 0.1 的互斥连锁 QTL, 遗传效应分布为 1.0 和 -1.0, 随机误差方差假定为 0.4, 标记间的距离为 2 cM。

Two QTLs are linked in the repulsion phase on one chromosome of 120 cM. Genetic effects of the two QTLs are 1.0 and -1.0, respectively. Variance of random error is 0.2, and distance between markers is 2 cM.

两个等位基因 Q 和 q, 分别代表两个亲本中的等位基因, 这样就有 QQ、Qq 和 qq 三种可能的基因型。我们用 \bar{P}_1 表示 $P_1(QQ)$ 的平均表现, \bar{P}_2 表示 $P_2(qq)$ 的平均表现, \bar{F}_1 表示 $F_1(Qq)$ 的平均表现, 并且 qq 代表较小的亲本基因型, QQ 表示较大的亲本基因型。m 表示中亲值, 即两个亲本的平均值, 记为 $m=1/2(\bar{P}_1+\bar{P}_2)$ 。亲本与中亲值的距离为 a, 或者说 $a=1/2(\bar{P}_1-\bar{P}_2)$, 即 a 为大值亲本(QQ)与小值亲本(qq)之差的一半, 是正值。 F_1 与中亲值 m 的离差为 d, 其值可正可负(图 5)。亲本与中亲值之差 a 称为加性效应, F_1 与中亲值之差 d 称为显性效应, 这就是经典数量遗传中的加显性遗传模型^[3]。Q 对 q 并不一定具有显性, 只表示取较大值或有较大的效应, 所以可以是无显性, 可以是部分显性, 也可以是完全显性或超显性^[3]。

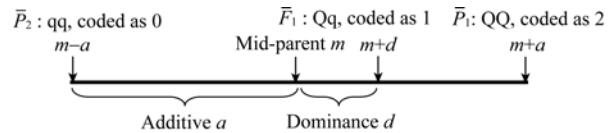


图 5 一个位点的数量性状加显性遗传模型
Fig. 5 One-locus additive and dominance genetic model of quantitative traits

QTL 作图中常用 2、1 和 0 表示 3 种标记和 QTL 的基因型, 以亲本 Harrington 和 TR306 衍生的大麦 DH 群体^[4,8,16]为例。亲本 Harrington 和 TR306 的平均粒重分别为 38.7 mg 和 45.0 mg, QTL 作图时分别用 2 表示 Harrington 的标记型、0 表示 TR306 的标记型。因此如果加性效应为正, 说明 Harrington 携带的等位基因起到增加粒重的作用, TR306 携带的等位基因则起到降低粒重的作用; 反之, 如果某个 QTL 加性效应为负值, 说明 Harrington 携带的等位基因起到降低粒重的作用, TR306 携带的等位基因则起到增加粒重的作用。从表 5 的结果来看, 9 个被检测到的控制粒重的 QTL 中有 6 个具有负的加性效应, 说明这 6 个 QTL 上提高粒重的等位基因来源于粒重较高的亲本 TR306, 但亲本 Harrington 在 $qKWT2-1$ 、 $qKWT2-3$ 和 $qKWT3-2$ 这 3 个 QTL 上携带着提高粒重的等位基因, 这些基因解释了群体中的超亲分离现象。育种中高粒重一般来说是理想性状, 因此在利用粒重 QTL 作图结果开展单标记或区间标记辅助选择时, 对 $qKWT2-1$ 、 $qKWT2-3$ 和 $qKWT3-2$ 应该选择亲本 Harrington 的标记类型, 其

表 5 大麦 DH 群体中检测到的粒重 QTL
Table 5 QTL of kernel weight identified in the barley DH population

QTL	ICIM (PIN=0.01, POUT=0.02)						选择基因型分析
	位置 Position (cM)	左侧连标记 Left flanking marker (cM)	右侧连标记 Right flanking marker (cM)	加性效应 Additive	LOD	贡献率 PVE (%)	LOD
<i>qKWT2-1</i>	83	74.30	83.40	0.39	4.16	3.04	0.53
<i>qKWT2-2</i>	139	130.90	142.10	-0.46	5.28	4.23	2.44
<i>qKWT2-3</i>	201	195.40	201.70	0.45	5.60	4.20	4.36
<i>qKWT3-1</i>	0	0.00	0.00	-0.35	3.35	2.41	0.02
<i>qKWT3-2</i>	22	17.20	25.20	0.57	8.00	6.50	2.97
<i>qKWT4</i>	125	124.80	140.10	-0.31	2.73	1.95	3.55
<i>qKWT5</i>	5	3.80	7.00	-1.38	32.19	38.27	19.82
<i>qKWT7-1</i>	4	3.40	7.60	-0.56	7.81	6.38	3.55
<i>qKWT7-2</i>	95	92.60	97.70	-0.94	18.86	17.51	9.41

最后一列为选择基因型分析得到的 LOD 值, 所在位置为 ICIM 鉴定出 QTL 的最近标记。选择基因型分析中, 标记频率的差异由表型最低和最高的 25% 的个体进行计算和检验, LOD 临界值为 2.5。粗体代表选择基因型分析中得到的 LOD 值超过临界值 2.5, 即与 ICIM 检测结果相同。

LOD score of selective genotyping given in the last column is at the nearest marker of the corresponding QTL identified by ICIM. In selective genotyping, the selected proportion was 25% for each tail. LOD score from selective genotyping is in bold when it is higher than LOD threshold 2.5.

他 QTL 应该选择亲本 TR306 的标记类型, 这样才能选择到所有增加粒重的等位基因。

2.3 选择基因型分析的有效性如何?

QTL 定位和基因精细定位时通常利用全群体分析方法, 但是全群体分析需要测定群体中所有个体的表现型和基因型, 要检测到微效的 QTL 需要大量的样本才能达到相应的功效, 花费较高。相比较, 选择基因型分析(selective genotyping)一般利用分离群体的双尾或单尾极端表型个体, 分析标记与 QTL 的连锁关系, 检验标记基因型在两极端类型内的分离比例是否偏离孟德尔规律, 是一种花费较低的方法,

常常被用于检测控制质量性状的主效 QTL^[17-20]。在对数量性状遗传分析时, 我们是否仍然可以采用选择基因型分析方法呢? 通过模拟分析大量遗传模型下多种因素对选择基因型作图结果的影响, 并利用真实群体比较选择基因型分析与全群体作图结果的差异^[20], 我们得到以下结论(表 5 和图 6)。

选择基因型分析的检测功效一般都高于 IM 的检测功效, 但是低于 ICIM 的检测功效。增大群体能迅速提高选择基因型分析的检测功效, 在不同群体大小下, 选择比例介于 0.15~0.35 时检测功效较高。而对于连锁 QTL, 增大标记密度有助于检测连锁

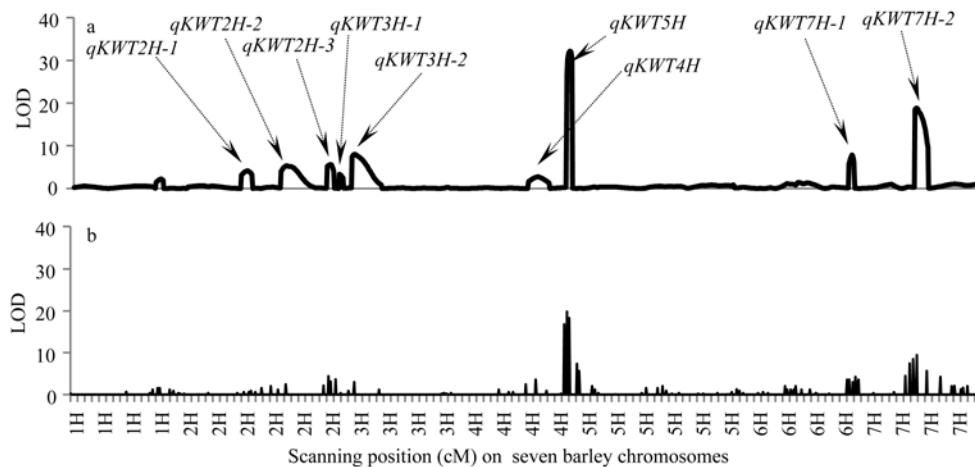


图 6 大麦 DH 群体完备区间作图(a)和选择基因型分析作图结果(b)

Fig. 6 Mapping results for a barley DH population under inclusive composite interval mapping (a) and selective genotyping (b)

QTL 的准确性。选择基因型分析对加性 QTL 的检测功效不受它们之间互作效应的影响。换句话说, 两个互作的加性 QTL 检测功效与这两个 QTL 只有加性没有上位效应时的检测功效相当。选择基因型分析无法检测到只存在上位性而没有加性效应的 QTL。利用真实群体(包括 RIL、DH、BC 和 F_2)比较选择基因型分析与全群体 ICIM 的结果, 发现对于只存在加性效应的群体, 两种方法的作图结果一致性较好, 而对于 F_2 群体, 虽然显性效应存在, 但是利用选择基因型分析也只能检测到加性 QTL, 并且双向选择的作图效果远远优于单向选择的。

2.4 复合性状是否适宜于 QTL 作图?

植物育种中, 有不少性状可以看作是由两个或两个以上性状的代数运算得到的, 如玉米雌雄开花间期等于雌穗开花期减去雄穗开花期, 水稻籽粒的长宽比等于粒长除以粒宽。为便于叙述, 我们把通过单独测量得到性状值的性状称为构成性状, 由若干构成性状经过代数运算得到的性状称为复合性状。不少研究者利用构成性状和复合性状同时进行作图, 有时会发现一些复合性状独有的 QTL, 但对复合性状 QTL 作图的理论根据和有效性缺乏了解。其实复合性状的使用使遗传模型变得更为复杂, 为说明这一点, 假定分别有 2 个独立遗传 QTL 控制 2 个构成性状(表 6), 加性遗传效应为 1, 性状 I 和 II 的群体平均数分别为 25 和 20。可以证明 4 个 QTL 在和性状上的遗传效应均为 1、在差性状上的遗传效应分别为 1、1、-1 和 -1, 在积和商性状上的遗传效应不仅与 QTL 在构成性状中的效应有关, 而且还依赖于群体平均数(表 6)。进一步还能证明除加性效应外, 4 个 QTL 在积和商性状上还存在互作效应(表 6), 因此复合性状的遗传效应也比构成性状更复杂^[21]。

模拟研究也表明, 对于独立的 QTL 而言, 复合性状的遗传力与构成性状基本一致, 而对于连锁的 QTL, 积和商性状的遗传力有所下降(表 7)。对于不同的遗传模型, 与构成性状相比, 复合性状的检测

功效始终会有不同程度的降低, 而且错误发现率会升高; 选择较大的作图群体, 复合性状检测功效下降的程度会减小, 同时错误发现率也会略有减小, 但错误发现率仍然高于构成性状(表 7)。

QTL 间连锁的存在, 不利于复合性状作图, 检测功效会大大降低, 对 QTL 的效应和位置的估计不准确; 当连锁距离大于或等于 35 cM 时, 对复合性状效应和位置的估计基本无偏, 但检测功效仍然低于构成性状。构成性状间的正相关会极大地降低差与商性状的检测功效, 而构成性状间的负相关会极大地降低和与积性状的检测功效, 但不管是正相关或是负相关, 构成性状的 QTL 检测功效都高于复合性状的检测功效。复合性状独有 QTL 的遗传机制尚不明确, 模拟实验显示检测到的复合性状独有 QTL 可能是由构成性状中的微效 QTL 引起的。

与构成性状相比, 复合性状受较多 QTL 控制、QTL 具有更复杂的遗传效应和连锁关系、QTL 作图功效比构成性状下的功效明显下降, 因此使用复合性状作图的意义不大。复合性状在遗传和育种中的使用应充分考虑二者有着不同的研究目标, 遗传研究在于尽可能多地了解控制目标性状基因的遗传规律, 构成性状受较少 QTL 控制、具有较简单的遗传模型, 因此易于研究单个 QTL 的遗传; 育种的目的是要同时选择尽可能多的有利基因和基因组合。与多性状选择指数^[3]类似, 复合性状的选择可以同时选择影响多个性状的有利等位基因, 适度提高育种效率。因此, QTL 作图研究中应谨慎使用复合性状, 但这并不是排除育种中使用复合性状进行选择。

3 QTL 作图中有关作图群体及遗传图谱的问题

3.1 QTL 作图群体中的表型数据是否要求服从正态分布?

一些 QTL 应用研究论文中, 误以为表型数据的正态分布是 QTL 作图的前提。其实大部分作图方法

表 6 构成性状和复合性状的理论遗传效应(假定各有 2 个 QTL 控制 2 个构成性状)
Table 6 Theoretical QTL effects of two individual traits and four composite traits (assuming each individual trait is controlled by two QTL)

QTL	染色体 Chromosome	位置 Position (cM)	加性效应 Additive					
			性状 I Trait I	性状 II Trait II	和性状 Addition	差性状 Subtraction	积性状 Multiplication	商性状 Division
Q ₁	1	18	1.0		1.0	1.0	20	0.0503
Q ₂	2	28	1.0		1.0	1.0	20	0.0503
Q ₃	3	53		1.0	1.0	-1.0	25	-0.0631
Q ₄	4	63		1.0	1.0	-1.0	25	-0.0631
群体均值 Population mean			25	20	45	5	500	1.2563

表 7 利用构成性状和复合性状的 QTL 检测功效
Table 7 QTL detection power of individual and composite traits

参数 Parameter	QTL	性状 I Trait I	性状 II Trait II	和性状 Addition	差性状 Subtraction	积性状 Multiplication	商性状 Division
功效 Power (%)	Q ₁	95.10		69.60	69.30	55.20	50.50
	Q ₂	94.80		69.80	70.40	54.10	50.90
	Q ₃		92.50	67.20	65.30	76.90	75.20
	Q ₄		94.50	68.40	65.40	77.80	75.20
错误发现率 False discovery rate (%)	Q ₁	21.63	22.98	27.42	28.05	28.07	29.68
	Q ₂	18.54		18.55	18.62	18.36	18.45
	Q ₃	28.46		28.49	28.38	28.44	28.52
	Q ₄		52.65	52.68	52.61	52.75	52.65
位置估计 Estimated position (cM)	Q ₁		62.85	62.83	62.63	62.88	62.58
	Q ₂	1.00		1.10	1.11	23.32	0.06
	Q ₃	1.01		1.09	1.11	23.42	0.06
	Q ₄		1.00	1.11	-1.11	26.46	-0.07
加性效应估计 Estimated additive effect	Q ₁		1.00	1.10	-1.12	26.61	-0.07
	Q ₂		1.01	1.09	1.11	23.42	0.06
	Q ₃			1.11	-1.11	26.46	-0.07
	Q ₄			1.10	-1.12	26.61	-0.07

只是要求表型数据的随机误差项服从正态分布, 数量性状只有在多基因假说下才真正符合正态分布, 表型数据的非正态并不影响 QTL 作图。一般地, 个体的表型是群体平均数 μ 、遗传型值 G 和随机误差效应 ε 之和, 即统计模型 $P = \mu + G + \varepsilon$ ^[3]。加显性模型

下包含 q 个 QTL 的遗传模型为 $G = \sum_{j=1}^q [a_j w_j + d_j v_j]$,

其中 w_j 和 v_j 是第 j 个 QTL 基因型的指示变量^[4,9], 因此 G 一般服从多项分布, 而 ε 则要求服从正态分布, 并且是大多数统计方法建模的前提。在多基因假说下, G 服从正态分布, 因此 P 也服从正态分布。当 QTL 的个数较少、并存在少数遗传效应较大的 QTL 时, P 将不服从正态分布。

举例说明, 假定长度为 160 cM 的染色体 25 cM 处有一 QTL 解释 80% 的表型变异, 遗传效应为 1.0, 群体平均数为 10。图 7 为一个模拟作图群体中性状的次数分布, 作图群体包含 200 个 DH 家系, 表型数据明显不服从正态分布。ICIM 一维扫描结果(图 7)表明 LOD 在 27 cM 处达到峰值 88.44, 对应位置的遗传效应估计值为 0.9989, 解释 81.12% 的表型变异。

3.2 加密标记是否可以显著提高 QTL 检测功效?

问题 1.2 中列出了统计学上提高假设检验功效的 2 个主要途径, 即提高样本量和降低表型误差。从实际应用角度来看, 对一个已建成的群体一般难以再增加个体数, 人们有时候关心是否可以通过加密标记来更准确地定位 QTL。在图 3 模拟实验的基础上, 我们又模拟了两种标记密度(MD), 即 5 cM 和

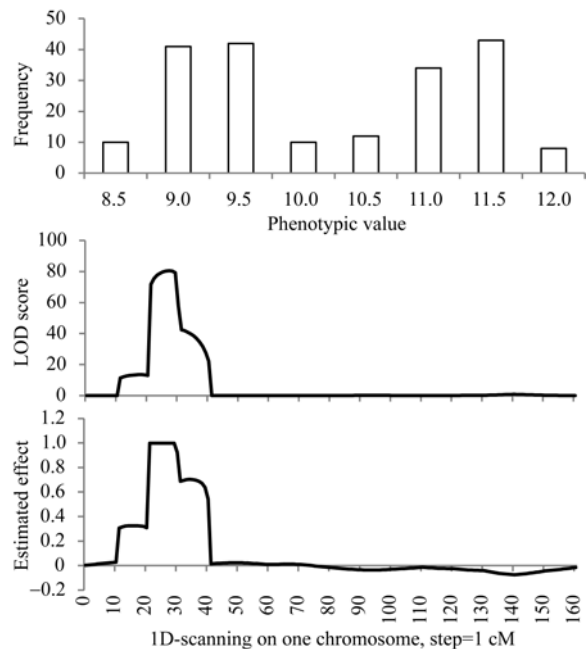


图 7 一个表型非正态的模拟作图群体及其 QTL 作图结果
Fig. 7 QTL mapping in a simulated population with 200 doubled haploid lines that do not follow a normal distribution
染色体 25 cM 处有一 QTL 解释 80% 的表型变异, 遗传效应为 1.0, 作图群体包含 200 个加倍单倍体家系。

One QTL is simulated at 25 cM on the chromosome, explaining 80% of the phenotypic variation.

20 cM (图 8)。MD 为 5、10 和 20 cM 时需要总的标记数分别为 330、170 和 90, 当标记数由密度为 10 cM 时的 170 减少到密度为 20 cM 时的 90 个时, 对于 PVE 小于 10% 的 QTL, 即使群体大小为 600, 10 cM 置信区间检测功效也有明显的下降; 对于 PVE

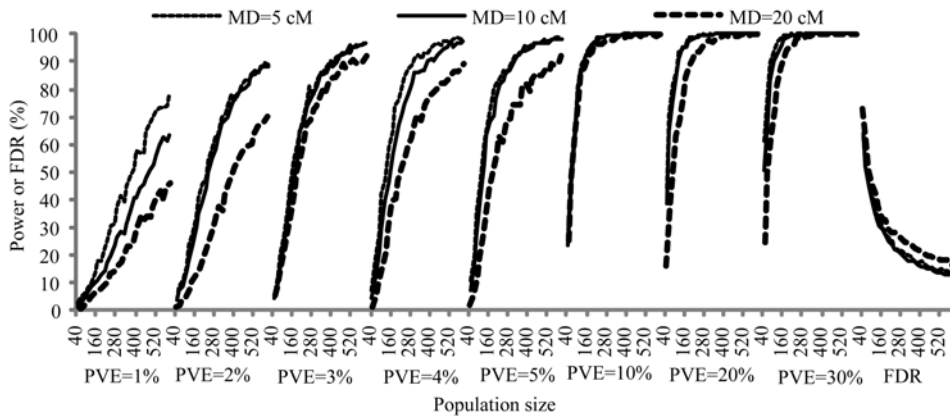


图 8 QTL 检测功效与标记密度和群体大小的关系

Fig. 8 Relationship of QTL detection power with marker density and mapping population size

大于 10% 的 QTL, 随着群体的增大, 检测功效接近于 100%, 但对于较小的群体, 检测功效也有明显的下降。当标记数由密度为 10 cM 时的 170 增加到密度为 5 cM 时的 330 时, 只有 PVE=1% 的 QTL 的检测功效才有明显的变化, 其他 QTL 检测功效的增加不明显。对于一般有 100~200 个体的作图群体, QTL 连锁作图中每隔 10 cM 左右有一个分子标记可能就足够了^[13]。

IM 和 ICIM 相比, ICIM 可以更好地检测连锁 QTL (图 2)。当然连锁程度对检测功效有较大影响, 如果 2 个 QTL 分布在同一个标记区间内, 可能用所有的方法都难以检测到。ICIM 要求连锁 QTL 间至少有一个空白区间^[7], 即分隔 QTL, 因此更密的标记更有可能把连锁 QTL 变成分隔 QTL, 也更容易被检测出来。假定在长度为 160 cM 的一条染色体上 22 cM 和 42 cM 处分布 2 个等效相引相 QTL, 分别解释 31% 的表型变异, 图 9 给出群体大小为 100、300 和 500 时 100 个群体的 ICIM 平均 LOD 曲线图。当 MD 为 20 cM 时, 2 个 QTL 位于 2 个相邻的标记区间, 之间没有空白区间。ICIM 作图结果表明即使群体大小为 500, 也无法将二者区分开, 而是在两个 QTL 之间发现一个效应近似等于两个 QTL 之和的“幻影”QTL。当 MD 为 10 cM 时, 两个 QTL 被一个空白区间(即区间[30 cM, 40 cM])分隔, 这时有一部分模拟群体检测到 2 个 QTL, 大多数模拟群体还是只定位到一个遗传效应较大的“幻影”QTL。当 MD 为 5cM 时, 两个 QTL 被 3 个空白区间(即区间[25 cM, 30 cM]、[30 cM, 35 cM]和[35 cM, 40 cM])分隔, 当群体大小为 300 和 500 时大多数模拟群体中都能正确地检测到 2 个 QTL, 当群体大小为 100 时大多数模拟群体都不能检测到 2 个 QTL。

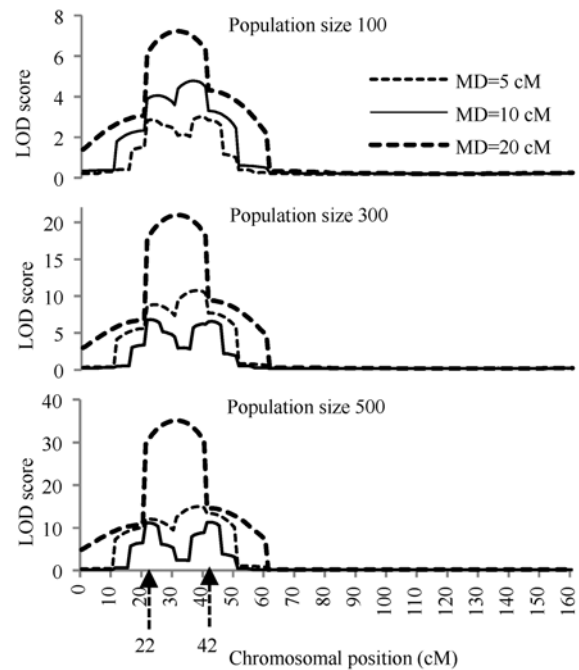


图 9 3 种群体大小中连锁 QTL 的完备区间作图结果

Fig. 9 Inclusive composite interval mapping of linked QTL in populations of three sizes

因此, 标记的加密尽管可以把紧密连锁的 QTL 分隔开, 但仍需足够大的群体, 才能提供更多 QTL 的重组基因型, 从而把 2 个连锁的 QTL 分离出来。在低于 200 的作图群体中, 标记增加到一定程度, 如每隔 5~10 cM 有一个标记, 再增加更多的标记也不会明显提高 QTL 的检测功效。

3.3 缺失分子标记对 QTL 作图有什么影响?

对于缺失标记, 作图软件 QTL IciMapping (可从网站 <http://www.isbreeding.net> 获得)中利用连锁信息以一定概率水平把缺失标记基因型插值为 P_1 、 P_2 或 F_1 型, 从而避免 QTL 作图中样本量的减少。在一

个包含 180 个单株的水稻 F_2 群体中, ICIM 定位到 9 个影响株高的 QTL^[9]。根据这些 QTL 建立遗传模型, 模拟不同程度标记的缺失^[22], 研究缺失标记对株高 QTL 检测功效的影响(图 10)。随着缺失率的增加, QTL 作图功效会逐步降低, 同时错误发现率逐步提高。对于效应较小的 QTL 和较小的群体, 缺失标记对 QTL 作图的影响较大; 对于效应较大的 QTL 和较大的群体, 缺失标记对 QTL 作图的影响可以忽略(图 10)。尽管作图功效会受到缺失标记的影响, 但在缺失标记条件下, 被检测到的 QTL 位置和效应估计与无缺失条件下基本一致, 都是渐近无偏的, 估计值的方差也基本相同。通过模拟不同大小的群体还可以发现, 一定比例的标记缺失对 QTL 作图结果的影响相当于对应比例个体缺失对 QTL 作图的影响(图 10), 也就是说, 一个群体大小为 n 、缺失率为 p 的群体的作图功效与大小为 $n(1-p)$ 、无缺失群体的作图功效大致相同。

3.4 奇异分离标记对 QTL 作图有什么影响?

奇异分离(SD, segregation distortion)会引起周围的标记出现奇异, 而奇异分离标记(SDM)也会导致连锁的 QTL 出现 SD。当 SDM 与 QTL 不连锁时, 对 QTL 定位基本没有影响。但是当 SDM 与 QTL 紧密连锁时, 会对 QTL 作图产生影响, 尤其当 QTL 效应较小、作图群体也较小时, 会产生较大的影响, 有时会提高与其连锁的 QTL 的检测功效, 有时会降低, 这取决于 QTL 的效应和 SD 类型(图 11)。以 F_2 群体为例, 检测 QTL 实际上就是检测 3 种 QTL 基因型

QQ、Qq 和 qq 之间的差异。若 SD 会使 QTL 的遗传方差高于无奇异分离时的遗传方差(见公式[1]), 则会提高作图功效; 反之, 则会降低作图功效。然而, 无论功效提高还是降低, QTL 的位置和效应估计与无 SD 情况下基本一致, 都是渐近无偏的, 估计值的方差也基本相同^[22]。

推广到其他具有 2 种等比例基因型的作图群体, 如 RIL 和 DH, 任何 SD 都会引起连锁 QTL 的 2 种基因型偏离 1:1 的分离比, 这种偏离对检测 2 种 QTL 基因型间的差异会产生不利影响, 从而减低 QTL 的检测功效。如果 2 种基因型的理论分离比为 3:1, 如回交群体衍生的重组近交家系, 若 SDM 导致 2 种基因型更接近 1:1, 此时 SD 便可提高与 SDM 连锁 QTL 的检测功效(图 12)。若 SDM 导致 2 种基因型更加偏离 3:1, 如 4:1, 这样的 SD 无疑要降低与 SDM 连锁 QTL 的检测功效, 在图 12 方差比值的曲线中可以看出如果奇异分离导致 QQ 频率在(0.25, 0.75)中变化, 奇异分离时 QQ 具有较大的遗传方差, 这样的奇异分离将提高 QTL 的检测功效。

SD 一般是由不同配子或合子生活力的差异引起的, 如产生 RIL 的一粒传方法就是为了尽可能避免在遗传研究群体中 SD 的发生。但在实际数据中, 即使是在通过一粒传法产生的作图群体中 SD 也不可避免。在实际应用中, 对遗传效应较大的 QTL 和较大的作图群体, 少量的 SDM 对 QTL 作图的影响是有限的, 这种影响也会随着与 QTL 间的连锁距离的增大而消失^[22]。

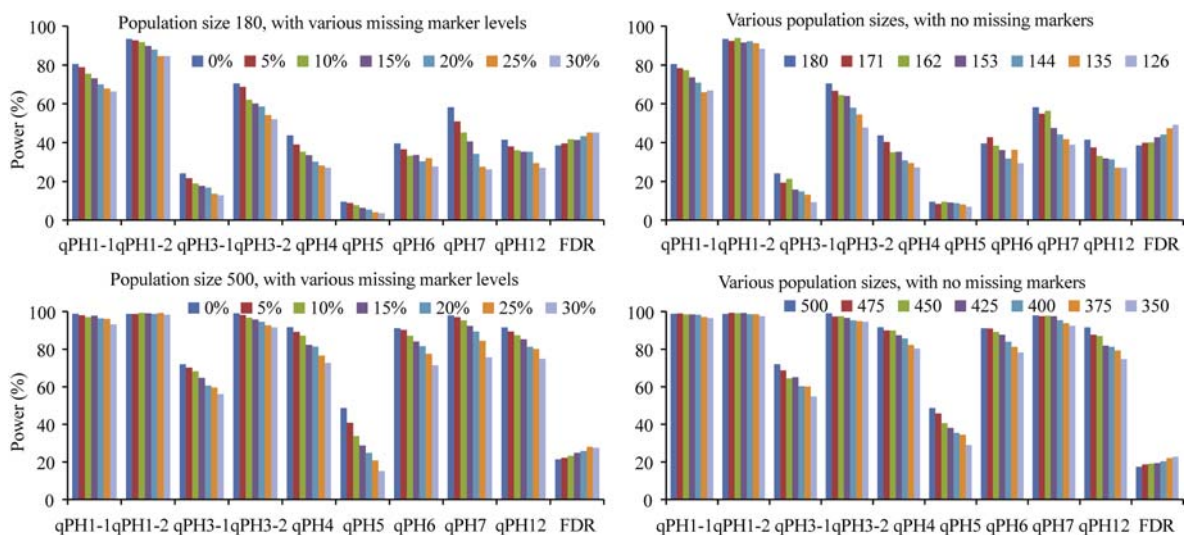


图 10 F_2 群体中缺失标记对 QTL 检测功效的影响
Fig. 10 Effect of missing markers on QTL detection in F_2 population

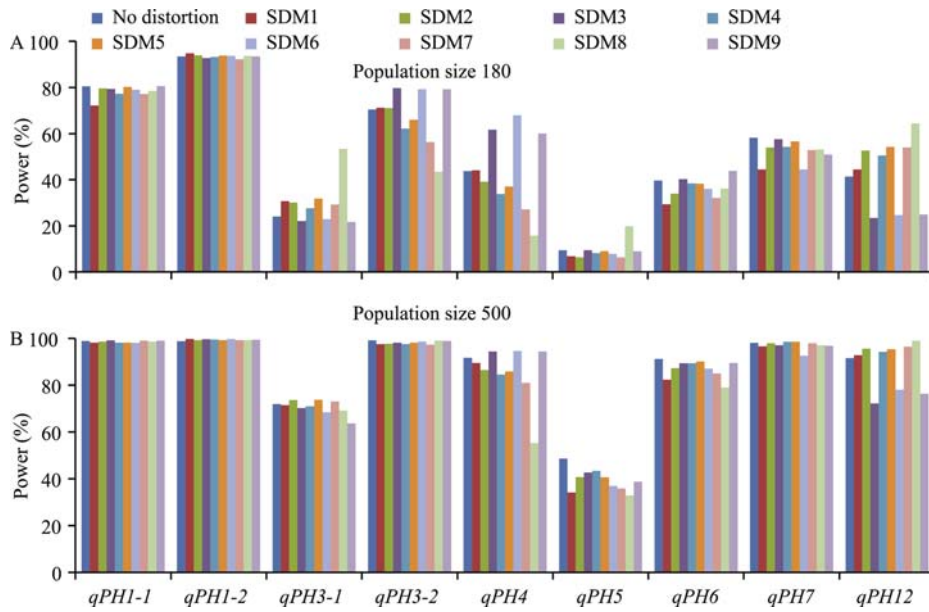


图 11 F₂ 群体中奇异分离标记对 QTL 检测功效的影响
 Fig. 11 Effect of segregation distortion markers on QTL detection in F₂ population

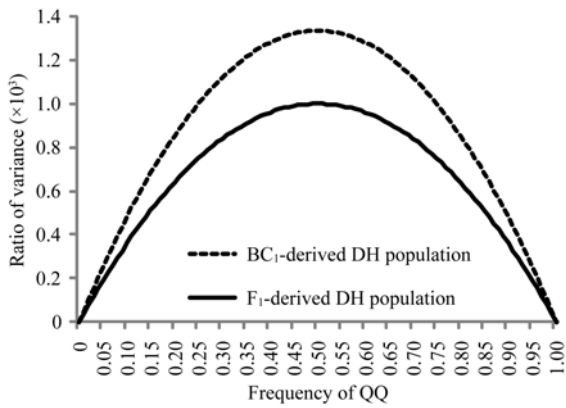


图 12 BC₁ 和 F₁ 产生的 DH 作图群体中奇异分离时 QTL 的遗传方差与无奇异分离时方差的比值
 Fig. 12 Ratio of genetic variance of QTL under segregation distortion and non-distortion in BC₁ and F₁ derived doubled haploids populations

4 结论

LOD 是 2 种假设检验下极大似然函数比值的常用对数, 较高的 LOD 临界值将更好地控制假 QTL 的发生, 同时遗传效应较小的真 QTL 就不易被检测出来。如何平衡两类错误的概率, 不仅是统计学问题, 还要考虑到具体的研究目标。QTL 作图包含着复杂的统计假设检验, 一般采用模拟方法估计 QTL 检测功效和假阳性的大小, 功效高同时又具有较低假阳性在统计上就是较好的方法。增加群体大小、减小表型误差、创造近等基因系和染色体片段置换系、大群体中适当增加标记密度都是提高 QTL 检测

功效的有效途径。

QTL 的遗传方差占表型方差的比例即 QTL 的贡献率, 遗传方差受 QTL 的遗传效应和 QTL 基因型的频率共同影响, 不同 QTL 的贡献率一般不具有可加性, 互斥连锁有可能导致超过 100% 的贡献率。有利等位基因的判断依赖于 QTL 作图时如何编码 2 个亲本, 正加性效应 QTL 意味着编码为 2 的亲本携带的等位基因能够提高性状值, 反之编码为 0 的亲本携带的等位基因能提高性状值。选择基因型分析的检测功效一般都高于 IM 的检测功效, 但低于 ICIM 的检测功效。与构成性状相比, 复合性状 QTL 作图功效比构成性状下的功效明显下降, 因此使用复合性状进行 QTL 作图的意义不大, 但并不排除育种中使用复合性状进行选择。

表型正态分布不是该性状是否适合 QTL 作图的必要条件, 但一般要求表型数据的随机误差服从正态分布。高密度分子标记为分离紧密连锁 QTL 提供了可能性, 但高密度标记必须辅以大作图群体, 才能提高 QTL 作图的准确度。缺失标记对 QTL 作图的影响等同于一个缩小了的作图群体。奇异分离对作图的影响可以用 QTL 遗传方差的变化来衡量, 有时奇异分离也可能会增大 QTL 的方差进而提高检测功效, 而且这种影响会随着连锁距离的增大而减弱。缺失标记和奇异分离对 QTL 位置和效应估计的影响不大, 对于效应较大的 QTL 和较大的群体, 缺失标记和奇异分离对 QTL 作图的影响可以忽略。

References

- [1] Lynch M, Walsh B. Genetic and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates, 1998
- [2] Doerge R W. Mapping and analysis of quantitative trait loci in experiment populations. *Nat Rev Genet*, 2002, 3: 43–52
- [3] Zhai H-Q(翟虎渠), Wang J-K(王建康). Applied Quantitative Genetics (应用数量遗传). Beijing: China Agricultural Science and Technology Press, 2007 (in Chinese)
- [4] Wang J-K(王建康). Inclusive composite interval mapping of quantitative trait genes. *Acta Agron Sin* (作物学报), 2009, 35(2): 239–245 (in Chinese with English abstract)
- [5] Lander E S, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 1989, 121: 185–199
- [6] Zeng Z B. Precision mapping of quantitative trait loci. *Genetics*, 1994, 136: 1457–1468
- [7] Li H H, Ye G Y, Wang J K. A modified algorithm for the improvement of composite interval mapping. *Genetics*, 2007, 175: 361–374
- [8] Li H H, Ribaut J M, Li Z L, Wang J K. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet*, 2008, 116: 243–260
- [9] Zhang L Y, Li H H, Li Z L, Wang J K. Interactions between markers can be caused by the dominance effect of QTL. *Genetics*, 2008, 180: 1177–1190
- [10] Goffinet P, Loisel B, Lawrent B. Testing in normal mixture models when the proportions are known. *Biometrika*, 1992, 79: 842–846
- [11] Churchill G A, Doerge R W. Empirical threshold values for quantitative trait mapping. *Genetics*, 1994, 138: 963–971
- [12] Chen L, Storey J D. Relaxed significance criteria for linkage analysis. *Genetics*, 2006, 173: 2371–2381
- [13] Li H, Hearne S, Bänziger M, Li Z, Wang J. Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity*, 2010 (in press)
- [14] Wang J K, Wan X Y, Crossa J, Crouch J, Weng J F, Zhai H Q, Wan J M. QTL mapping of grain length in rice (*Oryza sativa* L.) using chromosome segment substitution lines. *Genet Res*, 2006, 88: 93–104
- [15] Wang J K, Wan X Y, Li H H, Pfeiffer W M, Crouch J, Wan J M. Application of identified QTL-marker associations in rice quality improvement through a design breeding approach. *Theor Appl Genet*, 2007, 115: 87–100
- [16] Tinker N A, Mather D E, Rosnagel B G, Kasha K J, Kleinhofs A, Hayes P M, Falk D E, Ferguson T, Shugar L P, Legge W G, Irvine R B, Choo T M, Briggs K G, Ullrich S E, Franckowiak J D, Blake T K, Graf R J, Dofing S M, Saghai-Marooof M A, Scoles G J, Hoffman D, Dahleen L S, Kilian A, Chen F, Biyashev R M, Kudrna D A, Steffenson B J. Regions of the genome that affect agronomic performance in two-row barley. *Crop Sci*, 1996, 36: 1053–1062
- [17] Zhang L P, Lin G Y, Niño-Liu D, Foolad M R. Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. *Mol Breed*, 2003, 12: 3–19
- [18] Wingbermuehle W J, Gustus C, Smith K P. Exploiting selective genotyping to study genetic diversity of resistance to Fusarium head blight in barley. *Theor Appl Genet*, 2004, 109: 1160–1168
- [19] Docherty S J, Butcher L M, Schalkwyk L C, Plomin R. Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genome wide association studies. *BMC Genomics*, 2007, 8: 214
- [20] Sun Y P, Wang J K, Crouch J H, Xu Y B. Efficiency of selective genotyping for genetic analysis and crop improvement of complex traits. *Mol Breed*, 2010 (in press)
- [21] Wang Y, Li H, Zhang L, Lü W, Wang J. On the use of composite traits in quantitative trait locus mapping. *Genetics*, 2010 (in press)
- [22] Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, Li P, Li Z, Wang J. Effects of missing markers and segregation distortion on QTL mapping in F₂ populations. *Theor Appl Genet*, 2010 (in press)