

第七章

QTL作图中的其它常见问题

王建康，李慧慧，张鲁燕

中国农业科学院作物科学研究所

wangjiankang@caas.cn

<http://www.isbreeding.net>

QTL作图中的常见问题

- 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规方法. 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, 以及有关作图群体及连锁图谱等三大类. 有些问题第一~第六章已有所解答, 如LOD临界值的选择, 不同作图方法检测功效模拟和比较等. 本章对之前没有或较少涉及的一些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.

第七章 QTL作图中的其它常见问题

- § 7.1 QTL遗传方差和贡献率的计算
- § 7.2 复合性状的QTL作图
- § 7.3 加密标记对QTL检测功效的影响
- § 7.4 缺失标记的填补以及缺失对QTL作图的影响
- § 7.5 奇异分离对遗传研究的影响
- § 7.6 数量性状表型分布的非正态性

§ 7.1 QTL遗传方差和贡献率的计算

- § 7.1.1 单个QTL的遗传方差和贡献率
- § 7.1.2 连锁QTL的遗传方差和贡献率
- § 7.1.3 QTL贡献率与QTL检测功效的提高

单个QTL的加显性遗传模型

- 在一个遗传群体中, 三种QTL基因型QQ, Qq和qq的频率用 f_{QQ} , f_{Qq} 和 f_{qq} 表示. QTL作图中, 首先估计的是三种基因型的平均表现 μ_{QQ} , μ_{Qq} 和 μ_{qq} . 在此基础上, 根据加显性模型估计QTL的加显性效应 a 和 d .
- 加显性模型给出三种基因型的平均表现和加显性效应之间的线性关系, 即,

$$\mu_{QQ}=m+a; \mu_{Qq}=m+d; \mu_{qq}=m-a$$

单个QTL产生的遗传方差

- QTL的遗传方差 (用 V_Q 表示) 依赖于三种基因型的平均表现, 以及它们在一个群体中的频率, 即,

$$V_Q = f_{QQ} \times \mu_{QQ}^2 + f_{Qq} \times \mu_{Qq}^2 + f_{qq} \times \mu_{qq}^2 - (f_{QQ} \times \mu_{QQ} + f_{Qq} \times \mu_{Qq} + f_{qq} \times \mu_{qq})^2$$

- 根据加显性模型, 可以得到QTL的遗传方差与加显性效应之间的关系为,

$$V_Q = [f_{QQ} + f_{qq} - (f_{QQ} - f_{qq})^2]a^2 - 2f_{Qq}(f_{QQ} - f_{qq})ad + (f_{Qq} - f_{Qq}^2)d^2$$

- 如群体中仅存在两种QTL基因型QQ和qq, 它们的频率用 f_{QQ} 和 f_{qq} 表示. 则QTL遗传方差的表达式为,

$$V_Q = 4f_{QQ}f_{qq}a^2$$

- 对于三种QTL基因型的群体, 当基因型的频率偏离孟德尔分离比时, 遗传方差中除加性效应和显性效应的平方项外, 还包含有加性效应和显性效应的乘积项. 乘积项可正可负, 这时很难从QTL加显性效应的大小判断QTL遗传方差的高低.

单个QTL的表型贡献率

- QTL的贡献率定义为QTL的遗传方差占表型方差的比例,有时也称为QTL所能解释的表型变异或表型方差的比例 (phenotypic variance explained, PVE). QTL贡献率或PVE一般用百分数表示,其计算公式是,

$$PVE = \frac{V_Q}{V_P} \times 100\%$$

- 其中, V_Q 是公式7.1.2~4给出的QTL遗传方差, V_P 是性状的表型方差. 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, 其遗传方差和PVE不一定就高.

两个连锁QTL在无奇异分离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, Q_1 和 Q_2 间的重组率为 r .

基因型	频率	基因型值
$Q_1Q_1Q_2Q_2$	$(1-r)/2$	$\mu_{11} = m + a_1 + a_2$
$Q_1Q_1q_2q_2$	$r/2$	$\mu_{12} = m + a_1 - a_2$
$q_1q_1Q_2Q_2$	$r/2$	$\mu_{21} = m - a_1 + a_2$
$q_1q_1q_2q_2$	$(1-r)/2$	$\mu_{22} = m - a_1 - a_2$

连锁QTL的遗传方差

- QTL间的连锁或不平衡, 可能造成多个QTL的PVE之和超过100%的情形, 甚至可能造成单个PVE超过100%的情形. 这里, 用两个连锁QTL的情形说明PVE计算中可能存在的一些现象, 不考虑其他遗传因素. 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 分别为两个QTL的加性效应, 两个QTL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见表7.1.1. 两个QTL各自的遗传方差分别为

$$V_{Q_1} = a_1^2 \quad V_{Q_2} = a_2^2$$

- 但是, DH群体中总遗传方差并非两个QTL遗传方差之和, 而是,

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

两个独立遗传QTL的遗传方差和PVE

- 从等式 (7.1.6) 不难看出, 只有在 $r=0.5$ 的情况下才有

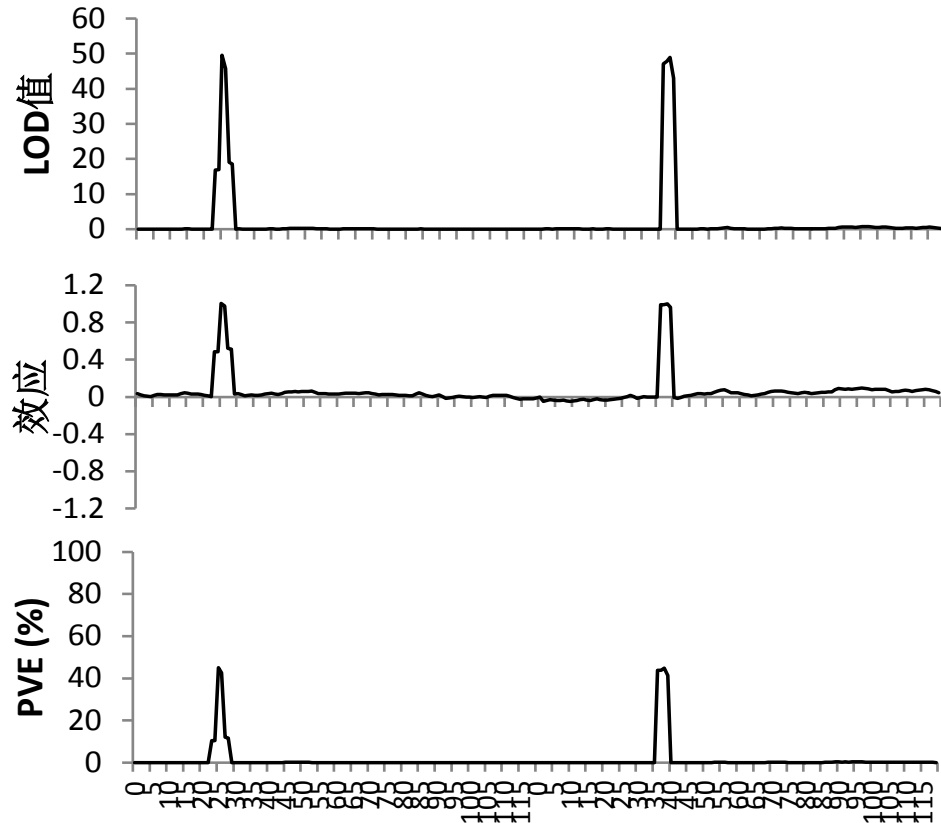
$$V_G = V_{Q_1} + V_{Q_2}$$

- 例如, 当 $a_1=1.0$, $a_2=1.0$, 误差方差=0.4时, 两个QTL各自的遗传方差为1. 当 $r=0.5$ 时, 总的遗传方差=2, 表型方差=2.4, 遗传力为0.833. 两个QTL的理论PVE均为41.7%, 两个QTL的PVE之和等于遗传力, 即总遗传方差占表型方差的比例.
- 在独立遗传的情况下, QTL效应正负号的改变, 不会影响QTL的方差和总的遗传方差, 也不影响QTL的PVE的大小.

一个包含200个DH家系模拟群体的QTL作图

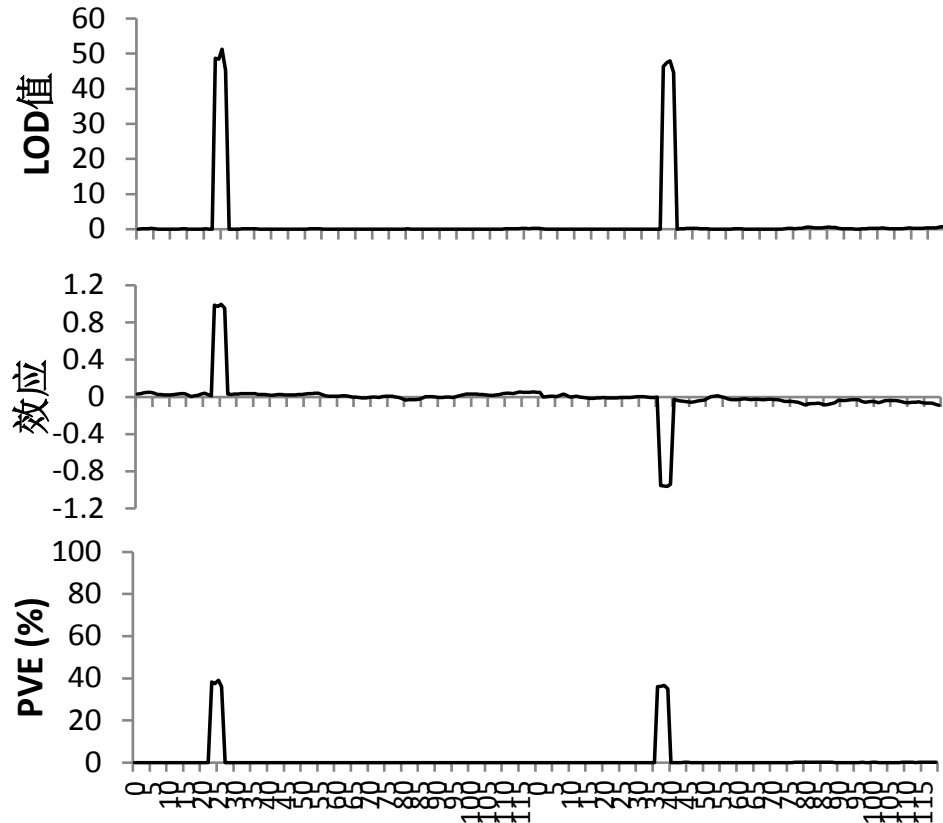
两条染色体的长度均为120cM, 各分布一个QTL. QTL的效应和随机误差方差见图例说明. 标记间的距离为2cM.

两个独立遗传QTL, 加性效应均为1, 误差方差为0.4



每隔1 cM扫描一次

两个独立遗传QTL, 加性效应分别1和-1, 误差方差为0.4



每隔1 cM扫描一次

两个连锁遗传QTL的遗传方差和PVE

- 当两个QTL间存在连锁, 并且 a_1 和 a_2 的效应方向相同 (即处于相引连锁状态) 时, 群体中总的遗传方差会高于单个QTL遗传方差之和, 即

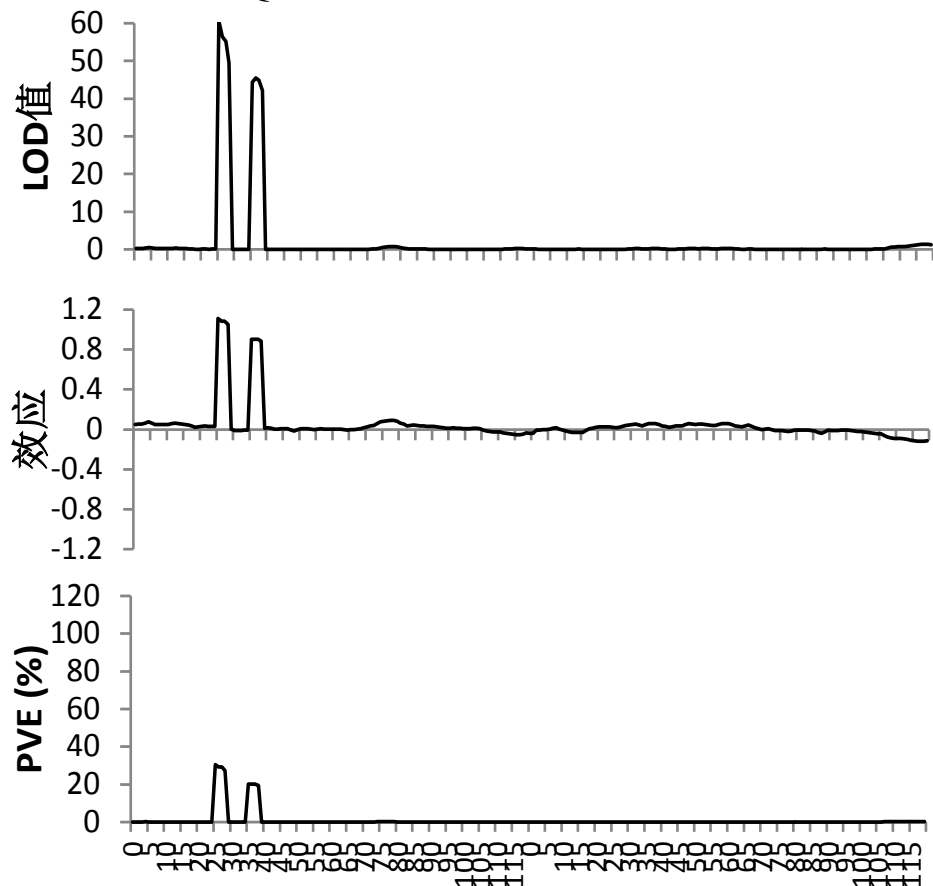
$$V_G > V_{Q_1} + V_{Q_2}$$

- 这样就会造成两个QTL的PVE远低于单个QTL的PVE之和的情况. 例如, 当 $a_1=1.0$, $a_2=1.0$, 误差方差=0.4时, 两个QTL各自的遗传方差为1. 当 $r=0.1$ 时, 总的遗传方差=3.6, 表型方差=4. 这时, 每个QTL的理论PVE为25%, 总的遗传方差占表型方差的比例到达90%, 远高于两个QTL的PVE之和.

一个包含200个DH家系模拟群体的QTL作图

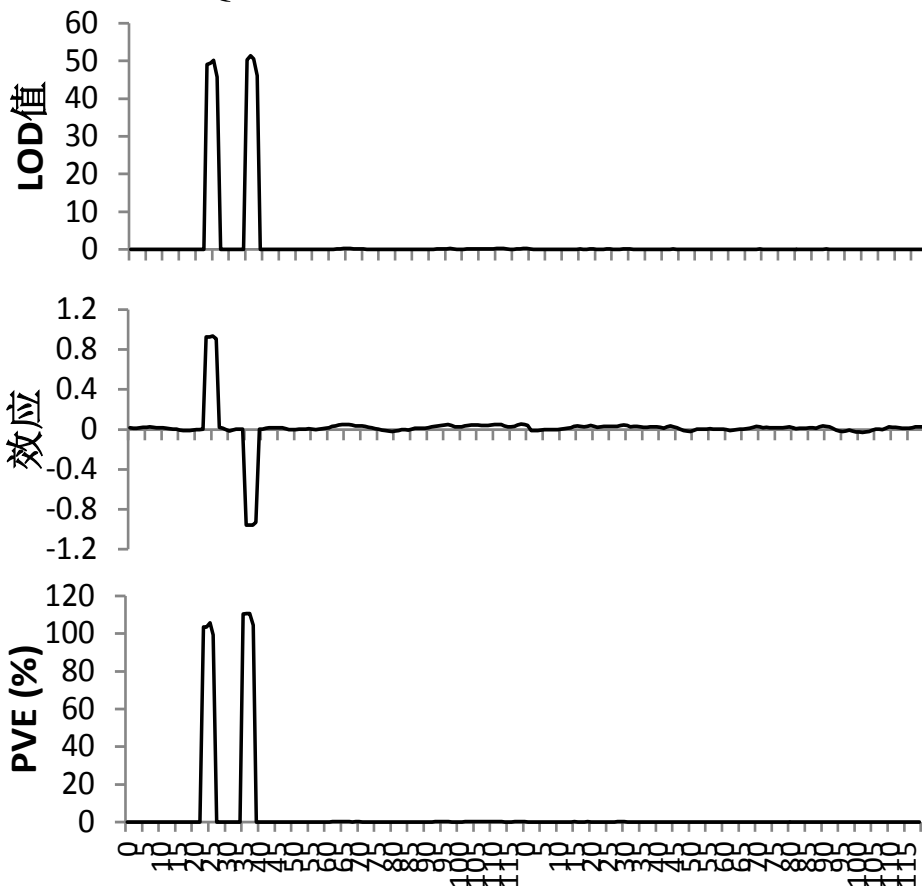
两条染色体的长度均为120cM, 两个连锁QTL分布第一条染色体上, 之间的重组率为0.1. QTL的效应和随机误差方差见图例说明. 标记间的距离为2cM.

两个连锁QTL, 加性效应均为1, 误差方差为0.4



每隔1 cM扫描一次

两个连锁QTL, 加性效应分别1和-1, 误差方差为0.4



每隔1 cM扫描一次

两个QTL产生的四种纯合基因型的频率, 平均表现, 以及边际频率和边际平均数

	Q_2Q_2	q_2q_2	行边际频率	行边际平均数
Q_1Q_1	f_{11}, μ_{11}	f_{12}, μ_{12}	$f_{1\cdot} = f_{11} + f_{12}$	$f_{1\cdot} \times \mu_{11} + f_{12} \times \mu_{12}$
q_1q_1	f_{21}, μ_{21}	f_{22}, μ_{22}	$f_{2\cdot} = f_{21} + f_{22}$	$f_{21} \times \mu_{21} + f_{22} \times \mu_{22}$
列边际频率	$f_{\cdot 1} = f_{11} + f_{21}$	$f_{\cdot 2} = f_{12} + f_{22}$		
列边际平均数	$f_{\cdot 1} \times \mu_{11} +$ $f_{2\cdot} \times \mu_{21}$	$f_{\cdot 1} \times \mu_{12} +$ $f_{2\cdot} \times \mu_{22}$		

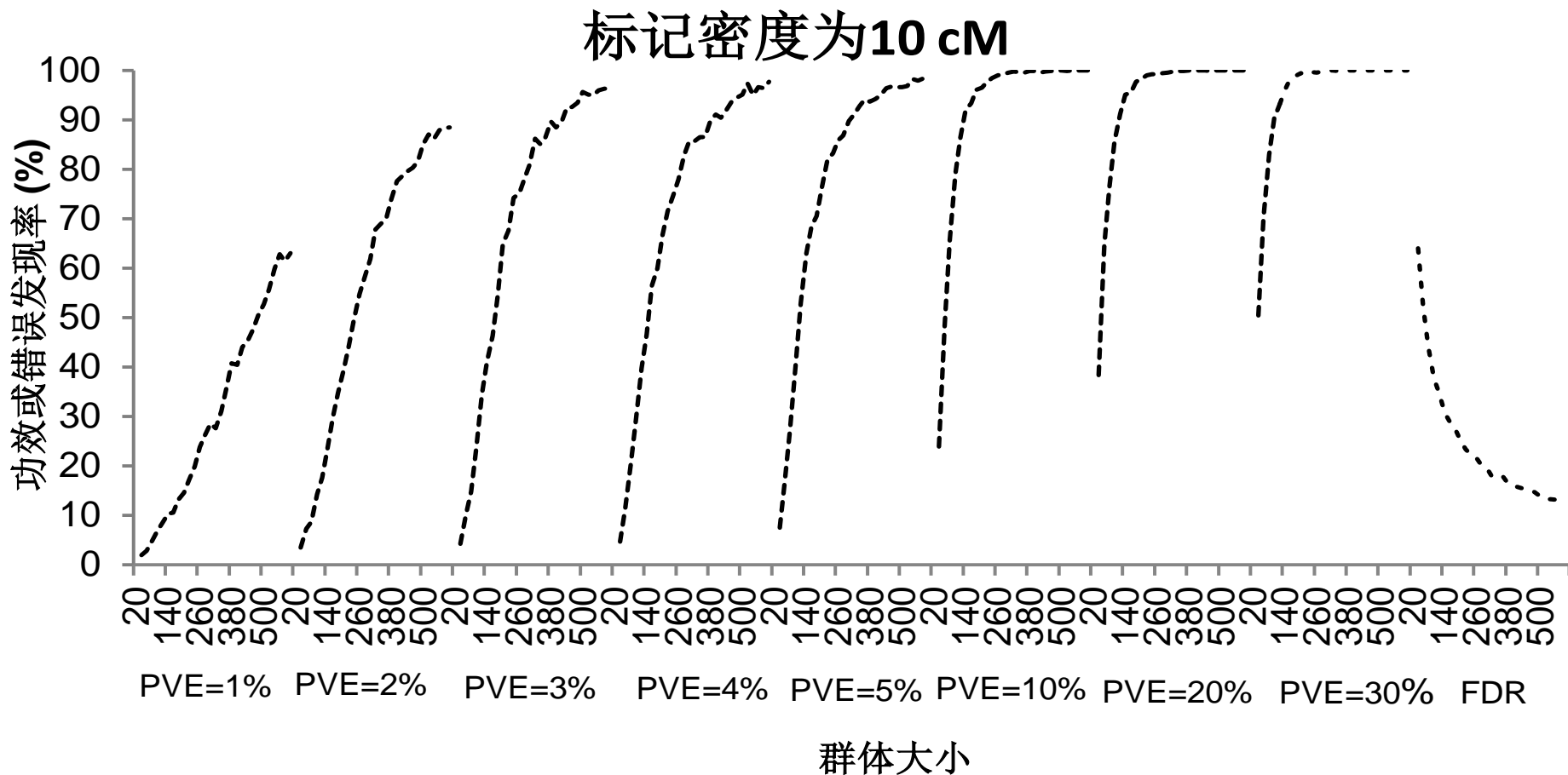
方差和PVE不可加的原因

- 除遗传连锁外, 两个座位上的基因型频率与单个座位上基因型频率间的不平衡, 以及两个QTL间的上位型互作, 都会造成两个QTL的遗传方差不等于单个QTL方差之和的现象.
- 表7.1.2给出某群体中, 两个QTL产生的四种纯合基因型的频率以及它们的平均表现, 四种频率之和等于1. 行边际频率代表的是群体中Q1座位上两种基因型的频率, 列边际频率代表的是群体中Q2座位上两种基因型的频率. 根据行边际平均数得到的方差代表的是群体中Q1座位的遗传方差, 根据列边际平均数得到的方差代表的是群体中Q2座位的遗传方差. 根据表7.1.2中四种基因型的频率和它们的平均表现, 得到的方差代表总的遗传方差.
- 仅在特定的情况下, 表7.1.2计算出的总遗传方差才等于两个边际方差之和. 特定情况就是不存在连锁, 不存在奇异分离.

方差和PVE不可加的原因

- 当两个QTL存在连锁时, 在无奇异分离的DH中, 每个座位上两种基因型的频率均为0.5. 但是, 两个座位上基因型的频率并不等于0.25, 表7.1.1看到的不平衡是由遗传连锁引起的.
- 但是, 遗传研究群体在产生过程中的选择和漂变等因素也能引起两个座位的不平衡. 基因间的上位型互作, 也会造成总的遗传方差不等于单个座位遗传方差之和. 有些群体中, 还可能同时存在基因型频率上的不平衡和上位型互作.
- 这些因素究竟是降低还是增加总遗传方差, 一般也难以确定. 因此, QTL作图中要尽可能避免对单个QTL遗传方差或PVE的简单相加.

不同群体大小和PVE对QTL检测功效的影响



群体大小对QTL检测功效的影响

- 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. 减小表型测定时的误差可以提高性状的遗传力, 同时也提高了单个QTL解释表型变异的比例, 进而提高QTL检测功效.
- 以ICIM为例, 增大RIL群体对提高QTL检测功效是明显的 (图7.1.3). 对大小为100, 200和400的RIL群体来说, 把PVE=4%的QTL定位到10 cM支撑区间内的概率分别为29%, 67%和91%, 而把PVE=10%的QTL定位到10 cM支撑区间内的概率分别为79%, 97%和100%.

QTL贡献率对检测功效的影响

- 减小表型误差则间接提高单个QTL的贡献率. 如果通过降低表型误差把QTL对表型的贡献率由4%提高到5%, 对大小为100, 200和400的RIL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的大, 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, 还要在多地点/多年份进行表型鉴定.

提高PVE和QTL检测功效的途径

- 降低群体的遗传变异也可间接提高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等群体, 都是通过这种途径提高遗传分析可靠性的.
- 例如, 假定某群体中三个独立遗传QTL的遗传方差分别为0.1, 0.2和0.3, 误差方差为0.4. 三个QTL分别解释**10%**, **20%**和**30%**的表型变异. 在这三个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, 三个QTL近等基因系群体的表型方差分别为0.5, 0.6和0.7. 三个QTL解释的表型变异则增加到**20%**, **33%**和**43%**. 因此, 在这三个近等基因系群体中, 通过控制QTL的个数, 间接提高了QTL的表型变异解释比例, 开展遗传研究和QTL定位将更加有效.

QTL IciMapping 4.1版的矫正PVE

- 在QTL作图完成之后，利用所有检测到QTL的侧连标记，建立表型对标记的回归模型，将回归模型的决定系数 R^2 视为所有QTL的遗传方差，对每个QTL的遗传方差进行矫正，利用矫正的QTL遗传方差除以表型方差来重新估计PVE。
- 矫正后的PVE具有了可加性！

§ 7.2 复合性状的QTL作图

- § 7.2.1 复合性状及其在遗传研究和育种中的应用
- § 7.2.2 一个玉米RIL群体中构成性状和复合性状的QTL作图
- § 7.2.3 复合性状的基因效应和遗传方差
- § 7.2.4 复合性状QTL作图的功效分析
- § 7.2.5 复合性状的遗传力

构成性状和复合性状

- QTL作图遗传研究中, 多数情况下的表型数据均直接来源于一些特定环境条件, 如大田或温室中的测量值, 有时为寻找环境间稳定表达的QTL, 也利用多年多点测量值的线性估计进行遗传研究.
- 在某些QTL作图研究中, 用于作图的表型数据则来自两个或更多个数量性状测量值的简单代数运算, 例如和性状, 差性状, 积性状及商性状等.
- 为方便起见, 将表型数据直接来源于测量值的数量性状称为构成性状, 将那些表型值来自于其他数量性状测量值代数运算的性状称为复合性状. 作为一种指数性状, 复合性状在遗传研究和育种中广为使用.

玉米的雌雄开花间隔期

- 玉米的雌雄开花间隔期 (anthesis-silking interval, ASI) 是对产量和抗旱有较大影响的一个复合性状. ASI的表型数据来自玉米雄花开花期 (male flowering time, MFLW) 和雌花开花期 (female flowering time, FFLW) 之间的差值. 在玉米的抗旱研究中, 育种家常把ASI作为一个有效的抗旱指标进行研究, 并借以选择抗旱性这一复杂性状.

玉米雌雄开花间隔期的QTL作图

- Ribaut等 (1996, 1997) 利用142个分子标记, 对一个大小为234的玉米F₂群体, 在良好灌溉和水分胁迫等条件下对开花期等性状进行QTL分析. 发现了4个QTL同时控制MFLW和FFLW, 1个QTL同时控制ASI和MFLW, 以及4个QTL同时控制ASI和FFLW. 在良好灌溉和水分胁迫的条件下, 共发现2个ASI独有的QTL, 其中1个QTL位于第2条染色体上, 可以解释11.4%的表型变异. 另外1个位于第6条染色体上, 可以解释13.0%的表型变异. 奇怪的是, 这2个QTL仅在ASI的QTL作图中被检测到, 而在MFLW和FFLW的QTL作图中未被检测到.
- Sari-Gorla等 (1999) 利用153个分子标记, 对另外一个大小为142的玉米RIL群体进行了QTL分析. 在良好灌溉条件下, 发现5个QTL影响MFLW, 0个QTL影响FFLW, 以及7个QTL影响ASI. 其中, 位于第9条染色体上影响ASI的QTL并没有被MFLW或FFLW发现. 在水分胁迫条件下, 发现了4个影响MFLW的QTL, 2个影响FFLW的QTL和2个影响ASI的QTL. 同样, 位于第5条染色体上影响ASI的QTL也没有被MFLW或FFLW发现.

水稻粒形长宽比及其QTL作图

- 水稻的长宽比 (也称粒型) 也是一个比较重要的复合性状. 长宽比根据水稻粒长和粒宽的比值进行计算, 是一个重要的水稻品质指标.
- Redona和Mackill (1998) 利用116个分子标记对一个大小为204的水稻F₂群体的粒长, 粒宽和粒型进行QTL分析, 发现了7个影响粒长的QTL, 4个影响粒宽的QTL, 以及3个影响粒型的QTL. 其中, 在第3和第7条染色体上发现了3个控制粒型的QTL, 与控制粒长和粒宽的QTL位于相近的染色体位置上.
- Tan等 (2000) 对来自于一个水稻骨干品种的F_{2:3}和RIL群体进行了QTL分析, 发现无论使用精米还是糙米, 控制粒长, 粒宽和粒型的主效QTL是一致的. 但是, 对于一些微效的QTL, 粒长和粒宽的QTL作图结果不尽相同.

水稻粒形的长宽比

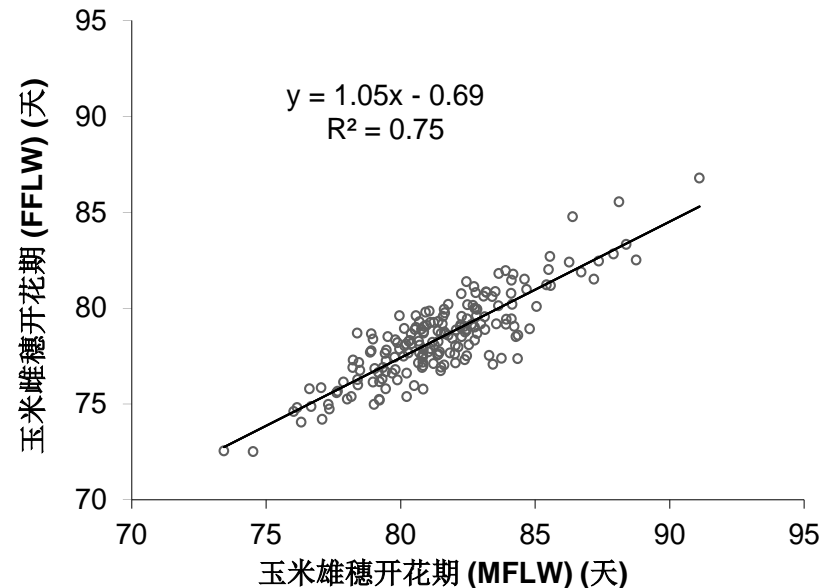
- Li等 (2004) 对一个大小为308的水稻BC₃F₁群体的粒长, 粒宽和粒型进行了QTL分析. 发现2个控制粒长的QTL, 分别位于第3和第10条染色体上. 1个控制粒宽的QTL, 位于第12条染色体上. 2个控制粒型的QTL, 其位置与控制粒长的2个QTL相近.
- Rabiei等 (2004) 对一个大小为192的水稻F₂群体的粒长, 粒宽和粒型进行了QTL分析. 共发现18个QTL, 5个控制粒长, 7个控制粒宽, 6个控制粒型. 在这18个QTL中, 有1个解释15.0%表型变异的粒型QTL, 既没有被粒长发现, 也没有被粒宽发现.

复合性状特异的QTL?!

- 从以上介绍的玉米开花间期和水稻粒型这两个性状的遗传研究来看, 复合性状的QTL作图和构成性状的QTL作图往往得到不同的结果, 有时甚至出现一些复合性状仅有的QTL (Ribaut et al. 1996; Sari-Gorla et al. 1999; Rabiei et al., 2004). 这些QTL仅在复合性状中检测到, 而在构成性状中都没有检测到.
- 下面利用一个玉米RIL群体中构成性状和复合性状的QTL作图, 进一步说明复合性状QTL定位中的这一现象. 在此实例中, 四种复合性状并非都具有生物学上的含义, 仅用这些复合性状说明定位到的QTL与构成性状定位到的QTL之间的差异.

一个玉米RIL群体中雄穗开花期和雌穗开花期表型数据之间的相关性

- 该玉米RIL群体包括187个RIL家系, 使用756个多态性标记进行基因型鉴定. 建立的遗传图谱总长为1380.8cM, 覆盖玉米的10条染色体, 标记之间的平均间距是1.85cM.
- 构成性状I是玉米雌花的开花时间 (也称吐丝期), 构成性状II是玉米雄花的开花时间 (也称散粉期). 构成性状I的最小值是73.44, 均值是81.47, 最大值是91.11. 构成性状II的最小值是72.50, 均值是78.40, 最大值是86.78.
- 这两个构成性状之间的相关系数为0.75, 说明两个构成性状之间存在高度的正相关.



两个构成性状（吐丝期和散粉期）的作图结果

性状	QTL编号	染色体	位置 (cM)	LOD值	加性效应	PVE (%)
构成性状I (吐丝期)	qZ1	1	87	7.56	0.7496	8.05
	qZ3	2	52	17.33	1.2165	20.47
	qZ4	2	77	5.04	0.6483	5.10
	qZ6	3	57	11.73	0.9558	12.99
	qZ7	4	61	3.08	-0.4703	3.15
	qZ10	7	46	4.73	-0.5835	4.78
	qZ11	9	42	12.61	0.9959	14.23
构成性状II (散粉期)	qZ2	1	129	2.57	0.3561	2.68
	qZ4	2	77	9.49	0.7615	10.24
	qZ5	2	108	5.71	0.5704	5.62
	qZ8	5	70	3.09	0.3875	2.93
	qZ9	6	27	2.87	-0.3693	2.88
	qZ10	7	49	4.77	-0.4757	4.73
	qZ11	9	40	11.01	0.7396	11.56

和性状与差性状的作图结果

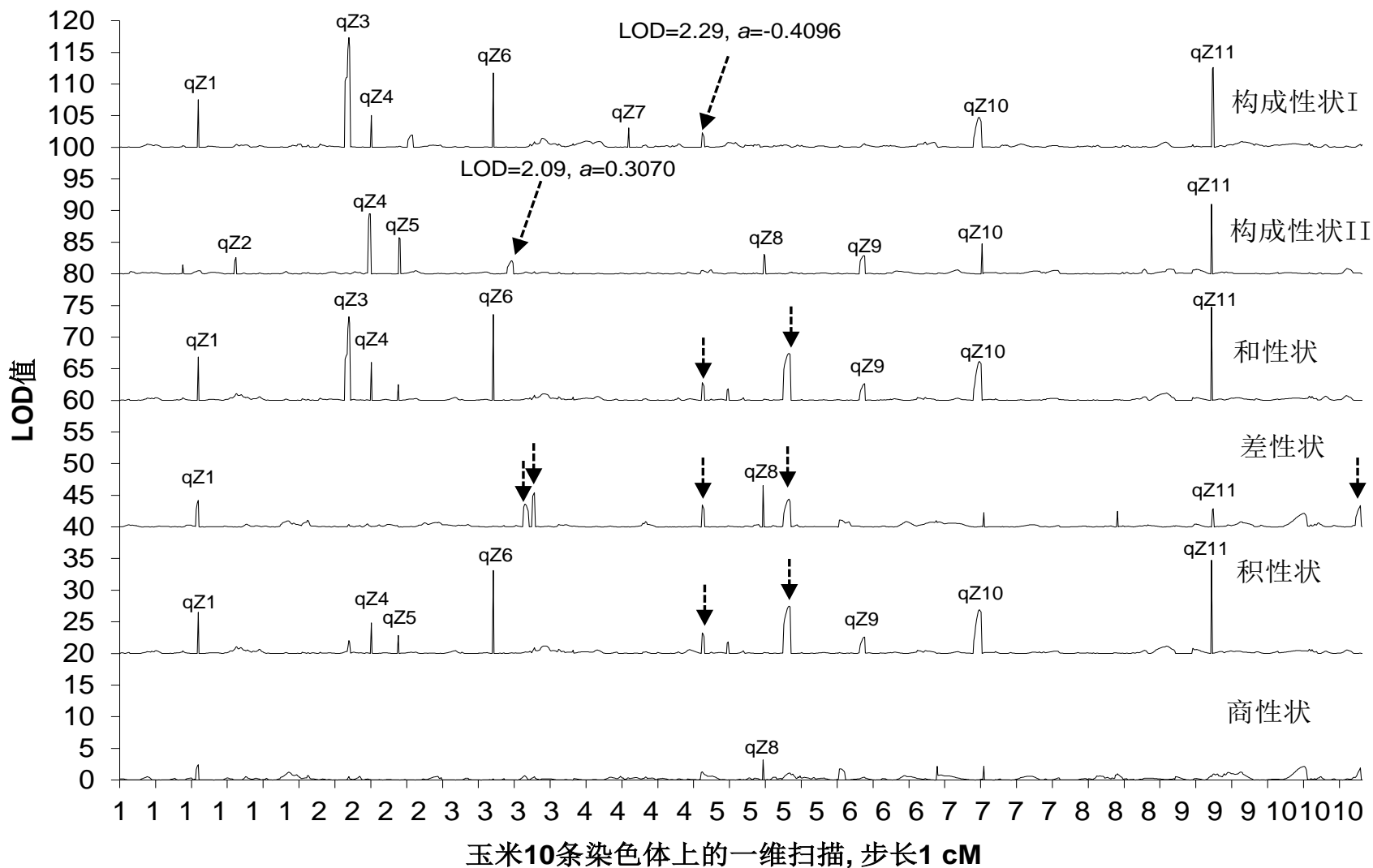
性状	QTL编号	染色体	位置 (cM)	LOD值	加性效应	PVE (%)
和	qZ1	1	87	6.88	1.1885	6.53
	qZ3	2	52	13.24	1.7202	13.21
	qZ4	2	77	6.02	1.1830	5.49
	qZ6	3	57	13.6	1.7287	13.72
	qZ9	6	28	2.63	-0.7126	2.35
	qZ10	7	46	6.15	-1.1128	5.62
	qZ11	9	40	14.73	1.7997	14.98
		5	1	2.80	-0.7590	2.55
		5	98	7.42	1.2351	7.00
差	qZ1	1	87	4.21	0.3319	6.12
	qZ8	5	69	6.53	-0.4414	10.08
	qZ11	9	42	2.89	0.2735	4.16
		3	93	3.57	-0.3077	5.28
		3	103	5.39	0.3782	7.98
		5	1	3.47	-0.3084	5.06
		5	98	4.39	0.3435	6.50
		10	91	3.40	0.2953	4.87

积性状与商性状的作图结果

性状	QTL编号	染色体	位置 (cM)	LOD值	加性效应	PVE (%)
积	qZ1	1	87	6.49	94.2285	6.37
	qZ4	2	77	4.85	86.4043	4.54
	qZ5	2	107	2.84	69.1877	2.75
	qZ6	3	57	13.12	138.6902	13.70
	qZ9	6	28	2.58	-57.8760	2.40
	qZ10	7	46	6.92	-97.2126	6.65
	qZ11	9	40	14.74	147.5234	15.62
		5	1	3.24	-67.0759	3.09
		5	98	7.46	101.4883	7.33
商	qZ8	5	69	3.16	-0.0042	5.78

构成性状和复合性状步长为1cM的一维扫描结果

两个构成性状中检测到的11个QTL按照染色体上的顺序依次用qZ1~qZ11表示。
构成性状LOD曲线的箭头指向未达到显著性水平的两个峰, 复合性状LOD曲线的箭头指向达到显著性水平, 但构成性状的相近位置上不存在显著的峰。



复合性状的基因效应和遗传方差

- 采用理论推导和模拟相结合的方法, 说明复合性状的一些遗传特性, 以及用于QTL作图可能存在的问题. 假定有四个QTL, Q_1 和 Q_2 控制构成性状I, 它们的加性效应分别用 a_1 和 a_2 表示. Q_3 和 Q_4 控制构成性状II, 它们的加性效应分别用 a_3 和 a_4 表示. 构成性状I的均值用 m_1 表示, 构成性状II的均值用 m_2 表示, 作图群体为双亲衍生的重组近交家系.
- 在RIL群体中, 构成性状的QTL基因型共有16种. 但是, 每个构成性状的基因型只有四种. 构成性状I的四种基因型值用 G_{11} , G_{12} , G_{13} 和 G_{14} 表示, 构成性状II的四种基因型值用 G_{21} , G_{22} , G_{23} 和 G_{24} 表示. 在加性遗传模型下, 构成性状的基因型值与遗传效应的关系是,

$$G_{11}=m_1+a_1+a_2; G_{12}=m_1+a_1-a_2; G_{13}=m_1-a_1+a_2; G_{14}=m_1-a_1-a_2;$$

$$G_{21}=m_2+a_3+a_4; G_{22}=m_2+a_3-a_4; G_{23}=m_2-a_3+a_4; G_{24}=m_2-a_3-a_4$$

包含4个QTL的遗传模型中复合性状的基因型值与构成性状基因型值之间的关系

编号	Q ₁	Q ₂	Q ₃	Q ₄	性状I	性状II	和性状	差性状	积性状	商性状
1	Q ₁ Q ₁	Q ₂ Q ₂	Q ₃ Q ₃	Q ₄ Q ₄	G ₁₁	G ₂₁	G ₁₁ +G ₂₁	G ₁₁ -G ₂₁	G ₁₁ ×G ₂₁	G ₁₁ /G ₂₁
2	Q ₁ Q ₁	Q ₂ Q ₂	Q ₃ Q ₃	q ₄ q ₄	G ₁₁	G ₂₂	G ₁₁ +G ₂₂	G ₁₁ -G ₂₂	G ₁₁ ×G ₂₂	G ₁₁ /G ₂₂
3	Q ₁ Q ₁	Q ₂ Q ₂	q ₃ q ₃	Q ₄ Q ₄	G ₁₁	G ₂₃	G ₁₁ +G ₂₃	G ₁₁ -G ₂₃	G ₁₁ ×G ₂₃	G ₁₁ /G ₂₃
4	Q ₁ Q ₁	Q ₂ Q ₂	q ₃ q ₃	q ₄ q ₄	G ₁₁	G ₂₄	G ₁₁ +G ₂₄	G ₁₁ -G ₂₄	G ₁₁ ×G ₂₄	G ₁₁ /G ₂₄
5	Q ₁ Q ₁	q ₂ q ₂	Q ₃ Q ₃	Q ₄ Q ₄	G ₁₂	G ₂₁	G ₁₂ +G ₂₁	G ₁₂ -G ₂₁	G ₁₂ ×G ₂₁	G ₁₂ /G ₂₁
6	Q ₁ Q ₁	q ₂ q ₂	Q ₃ Q ₃	q ₄ q ₄	G ₁₂	G ₂₂	G ₁₂ +G ₂₂	G ₁₂ -G ₂₂	G ₁₂ ×G ₂₂	G ₁₂ /G ₂₂
7	Q ₁ Q ₁	q ₂ q ₂	q ₃ q ₃	Q ₄ Q ₄	G ₁₂	G ₂₃	G ₁₂ +G ₂₃	G ₁₂ -G ₂₃	G ₁₂ ×G ₂₃	G ₁₂ /G ₂₃
8	Q ₁ Q ₁	q ₂ q ₂	q ₃ q ₃	q ₄ q ₄	G ₁₂	G ₂₄	G ₁₂ +G ₂₄	G ₁₂ -G ₂₄	G ₁₂ ×G ₂₄	G ₁₂ /G ₂₄
9	q ₁ q ₁	Q ₂ Q ₂	Q ₃ Q ₃	Q ₄ Q ₄	G ₁₃	G ₂₁	G ₁₃ +G ₂₁	G ₁₃ -G ₂₁	G ₁₃ ×G ₂₁	G ₁₃ /G ₂₁
10	q ₁ q ₁	Q ₂ Q ₂	Q ₃ Q ₃	q ₄ q ₄	G ₁₃	G ₂₂	G ₁₃ +G ₂₂	G ₁₃ -G ₂₂	G ₁₃ ×G ₂₂	G ₁₃ /G ₂₂
11	q ₁ q ₁	Q ₂ Q ₂	q ₃ q ₃	Q ₄ Q ₄	G ₁₃	G ₂₃	G ₁₃ +G ₂₃	G ₁₃ -G ₂₃	G ₁₃ ×G ₂₃	G ₁₃ /G ₂₃
12	q ₁ q ₁	Q ₂ Q ₂	q ₃ q ₃	q ₄ q ₄	G ₁₃	G ₂₄	G ₁₃ +G ₂₄	G ₁₃ -G ₂₄	G ₁₃ ×G ₂₄	G ₁₃ /G ₂₄
13	q ₁ q ₁	q ₂ q ₂	Q ₃ Q ₃	Q ₄ Q ₄	G ₁₄	G ₂₁	G ₁₄ +G ₂₁	G ₁₄ -G ₂₁	G ₁₄ ×G ₂₁	G ₁₄ /G ₂₁
14	q ₁ q ₁	q ₂ q ₂	Q ₃ Q ₃	q ₄ q ₄	G ₁₄	G ₂₂	G ₁₄ +G ₂₂	G ₁₄ -G ₂₂	G ₁₄ ×G ₂₂	G ₁₄ /G ₂₂
15	q ₁ q ₁	q ₂ q ₂	q ₃ q ₃	Q ₄ Q ₄	G ₁₄	G ₂₃	G ₁₄ +G ₂₃	G ₁₄ -G ₂₃	G ₁₄ ×G ₂₃	G ₁₄ /G ₂₃
16	q ₁ q ₁	q ₂ q ₂	q ₃ q ₃	q ₄ q ₄	G ₁₄	G ₂₄	G ₁₄ +G ₂₄	G ₁₄ -G ₂₄	G ₁₄ ×G ₂₄	G ₁₄ /G ₂₄

四个QTL所影响的构成性状以及在染色体上的三种分布模型

QTL	影响性状	QTL分布模型A		QTL分布模型B		QTL分布模型C	
		染色体	位置(cM)	染色体	位置(cM)	染色体	位置(cM)
Q ₁	构成性状I	1	18.0	1	18.0	1	18.0
Q ₂	构成性状I	2	28.0	1	53.0	2	28.0
Q ₃	构成性状II	3	53.0	2	28.0	1	53.0
Q ₄	构成性状II	4	63.0	2	63.0	2	63.0

四个QTL的三种效应模型

遗传参数	QTL效应模型A	QTL效应模型B	QTL效应模型C
Q_1 对构成性状I的加性效应 (a_1)	1	1	0
Q_2 对构成性状I的加性效应 (a_2)	1	1	0
Q_3 对构成性状II的加性效应 (a_3)	1	1	0
Q_4 对构成性状II的加性效应 (a_4)	1	1	0
Q_1 和 Q_2 对构成性状I的互作效应 (aa_{12})	0	1	1
Q_3 和 Q_4 对构成性状II的互作效应 (aa_{34})	0	1	1

各种性状中QTL的遗传效应, 遗传方差和广义遗传力

		构成性状I	构成性状II	和性状	差性状	积性状	商性状
遗传效应	平均数	25	20	45	5	500	1.2563
	A ₁	1	0	1	1	20	0.0503
	A ₂	1	0	1	1	20	0.0503
	A ₃	0	1	1	-1	25	-0.0631
	A ₄	0	1	1	-1	25	-0.0631
	A ₁₂	0	0	0	0	0	0
	A ₁₃	0	0	0	0	1	-0.0025
	A ₁₄	0	0	0	0	1	-0.0025
	A ₂₃	0	0	0	0	1	-0.0025
	A ₂₄	0	0	0	0	1	-0.0025
	A ₃₄	0	0	0	0	0	0.0063
	A ₁₂₃	0	0	0	0	0	0
	A ₁₂₄	0	0	0	0	0	0
	A ₁₃₄	0	0	0	0	0	0.0003
	A ₂₃₄	0	0	0	0	0	0.0003
A ₁₂₃₄	0	0	0	0	0	0	
遗传方差	分布模型A	2.000	2.000	4.000	4.000	2054.000	0.013
	分布模型B	2.993	2.993	5.986	5.986	3072.000	0.020
	分布模型C	2.000	2.000	5.986	2.014	3047.171	0.007
广义遗传力	分布模型A	0.300	0.300	0.300	0.300		
	分布模型B	0.391	0.391	0.391	0.391		
	分布模型C	0.300	0.300	0.391	0.177		

QTL效应模型A的QTL检测功效

参数	QTL	构成性状 I	构成性状 II	和性状	差性状	积性状	商性状
功效 (%)	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 (%)		21.63	22.98	27.42	28.05	28.07	29.68
位置估计 (cM)	Q ₁	18.54		18.55	18.62	18.36	18.45
	Q ₂	28.46		28.49	28.38	28.44	28.52
	Q ₃		52.65	52.68	52.61	52.75	52.65
	Q ₄		62.85	62.83	62.63	62.88	62.58
加性效应估计	Q ₁	1.00		1.10	1.11	23.32	0.06
	Q ₂	1.01		1.09	1.11	23.42	0.06
	Q ₃		1.00	1.11	-1.11	26.46	-0.07
	Q ₄		1.00	1.10	-1.12	26.61	-0.07

构成性状和复合性状的功效分析

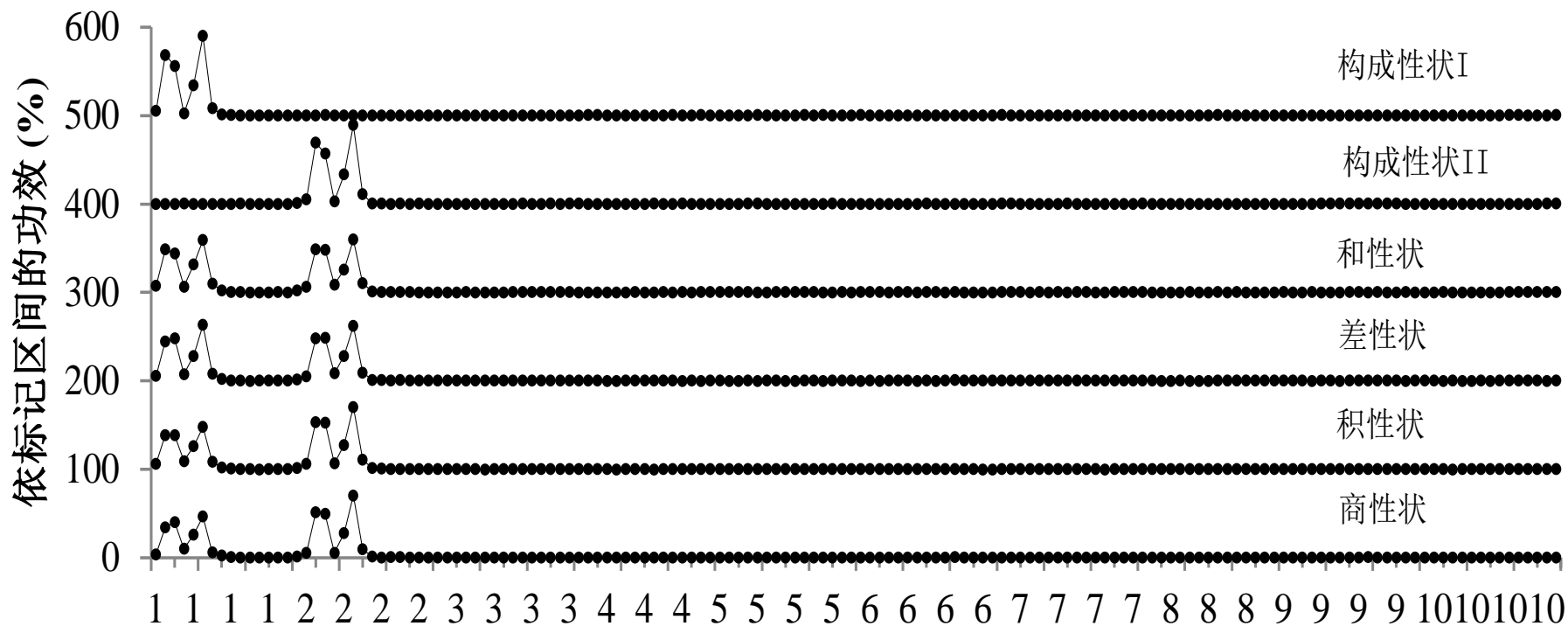
- QTL检测功效的模拟研究表明, 与构成性状相比, 复合性状的检测功效有不同程度的降低, 而且错误发现率会升高. QTL效应模型A中, 四个QTL的加性遗传效应相等. 在三种QTL分布模型下, 它们在构成性状中的检测功效都超过90%. 在三种分布模型中, 和与差这两个复合性状对不同的QTL也有着近似相等的检测功效. 但是, 这些检测功效远低于构成性状的检测功效.
- 分布模型A和B中, 和与差这两个复合性状的QTL检测功效减少了25%左右. 分布模型C中, 和与差这两个复合性状的QTL检测功效减少了40%以上. 在三种分布模型中, 积与商这两个复合性状QTL检测功效也比构成性状的功效有很大的下降. 比较起来, Q_1 和 Q_2 的检测功效比 Q_3 和 Q_4 下降得更多.

构成性状和复合性状的功效分析

- 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, QTL的效应和位置估计有更大的误差. 当连锁距离大于或等于35cM时, 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.

QTL效应模型A在位置模型B下按标记区间的功效分析

为了便于分析, 分别对积性状, 差性状, 和性状, 构成性状II和构成性状I的功效值加上100, 200, 300, 400和500.



10条染色体上的标记区间 (每个圆点对应于一个标记区间, 数字为染色体编号)

复合性状中独有的QTL没有重复性

- 图7.2.3给出按染色体上的标记区间统计的功效, 容易看出所有检测到的QTL在染色体上的分布情况.
- 从效应模型A在位置模型B下检测到QTL的分布来看, 构成性状I检测到的QTL分布在Q1和Q2附近, 构成性状II检测到的QTL分布在Q3和Q4附近. 四种复合性状检测到的QTL集中在四个预设的QTL附近, 其他标记区间上也偶尔检测到QTL, 但频率极低. 在不存在构成性状QTL的染色体区域上, 不存在很高的检测功效.
- 这也从另一个方面说明, 复合性状中独有的QTL没有重复性. 也因此可以推测, 复合性状中独有的QTL很可能是假阳性QTL.

复合性状开展QTL作图存在的问题

- 在真实的玉米RIL群体中, 根据观测到的QTL分布和各种效应进行模拟 (Wang et al., 2012). 结果发现, 对于复合性状和构成性状都能检测到的QTL, 复合性状的检测功效与构成性状的检测功效相似. 对于构成性状检测到, 但是复合性状没有检测到的QTL, 复合性状的检测功效就会下降.
- 同时, 差性状和商性状的FDR高于和性状和积性状. 究其原因, 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.
- 因此, 遗传研究中使用复合性状进行QTL作图的意义可能不大.

复合性状在育种中的作用

- 在遗传和育种中是否应该使用复合性状, 应从二者研究目标的差异方面来考虑. 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, 就更易于研究单个QTL的遗传, 实现遗传研究的目标. 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. 与多性状选择指数类似, 复合性状的使用可以同时选择到影响多个性状的多个有利等位基因, 以达到提高育种效率的目的.
- 为了实现遗传研究的目标, 需要把一个复杂性状或复杂模型简单化, 以有利于发现控制性状的基因及这些基因的遗传规律. 育种的目的是对多个性状的综合改良, 需要对这些性状同时进行选择, 育种群体中最好包含所有控制性状的有利基因和有利基因组合. 这时, 复合性状的使用恰恰有利于实现多性状和多基因聚合的育种目标.

模拟RIL群体中构成性状和复合性状的广义遗传力

效应模型	分布模型	构成性状I	构成性状II	和性状	差性状	积性状	商性状
A: 仅有加性	A	0.302	0.301	0.301	0.301	0.300	0.278
	B	0.366	0.364	0.367	0.362	0.365	0.332
	C	0.303	0.302	0.367	0.224	0.364	0.208
B: 加性和上位型互作	A	0.394	0.392	0.394	0.391	0.392	0.334
	B	0.435	0.433	0.436	0.431	0.435	0.378
	C	0.395	0.393	0.451	0.323	0.457	0.276
C: 仅有上位型互作	A	0.178	0.177	0.178	0.176	0.177	0.162
	B	0.161	0.161	0.162	0.160	0.161	0.154
	C	0.178	0.177	0.194	0.160	0.193	0.148
一个真实的玉米RIL群体		0.597	0.600	0.698	0.397	0.699	0.392

从构成性状的定位结果推测复合性状的遗传基础

- 既然复合性状不利于进行QTL研究, 那么, 如何获得复合性状的遗传信息? 实际群体中, 其实可从构成性状的QTL定位结果推测控制复合性状的QTL. 如一个QTL同时控制两个构成性状, 遗传效应的方向是一致的, 可以肯定该QTL也控制和性状. 如一个QTL同时控制两个构成性状, 遗传效应的方向是相反的, 可以肯定该QTL也控制差性状.
- 很难说, 控制和性状的QTL一定控制积性状, 控制差性状的QTL一定控制商性状. 在研究积性状与商性状的QTL时, 可能要考虑对积与商进行对数变换, 然后按和与差的情况处理. 但如何做更合适, 还有待进一步研究.

谨慎使用复合性状开展遗传研究

- 综上所述, 控制复合性状的QTL至少控制一个构成性状, 不太可能出现不控制任一构成性状却控制复合性状的QTL. 因此, 实际群体中检测到的只控制复合性状的QTL有以下两种可能. 一是构成性状中未超过给定临近值的微效QTL, 二是表型在数学运算过程中产生出的假阳性.
- 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 (包括二阶和更高阶的上位型互作) 和连锁关系, QTL作图功效比构成性状的功效有明显下降. 复合性状引起遗传结构复杂程度的增加, 可能是造成QTL作图中检测功效降低, 假阳性提高的主要原因. 因此, 遗传研究中使用复合性状开展QTL作图的意义不是很大.
- 育种研究与遗传研究有着不同的目的, 复合性状在育种中的使用可以同时选择到影响多个性状的有利等位基因或基因组合, 以达到提高育种效率的目的. QTL作图研究中应谨慎使用复合性状, 但这并不是要排除在育种中使用复合性状, 或者利用复合性状进行选择.

§ 7.3 加密标记对QTL检测功效的影响

- § 7.3.1 加密标记对单个QTL检测的影响
- § 7.3.2 加密标记对连锁QTL检测的影响

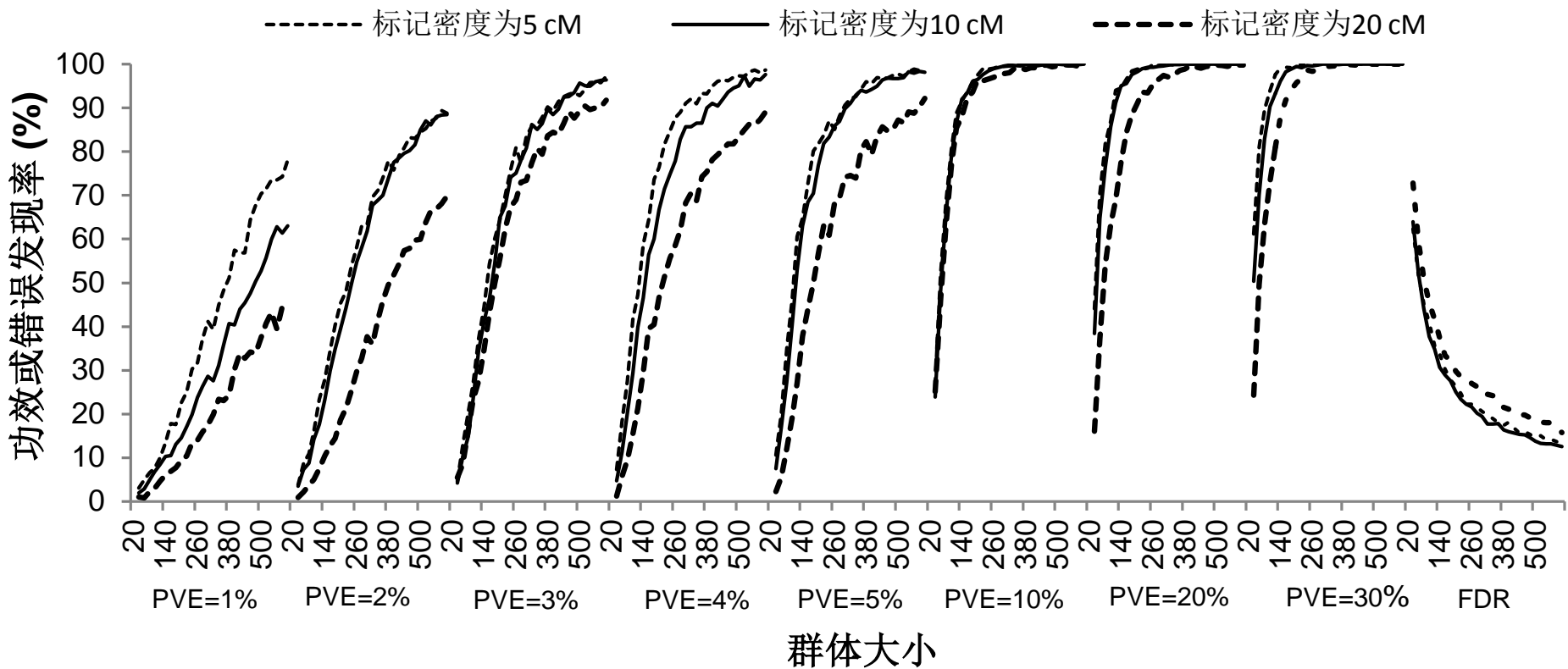
提高QTL检测功效的途径

- 统计学上提高假设检验的功效有两个主要途径, 即增加样本量和降低误差方差. 从实际应用的角度来看, 在一个已建成的遗传群体中, 一般难以再增加群体中的个体或家系数数量. 重复试验可以有效降低基因型平均表现中的误差方差, 提高基因型平均表现的广义遗传力, 进而提高QTL的检测功效.
- 在基因型鉴定方面, 过去一般只用几十个RFLP标记或数百个SSR标记, 开展一个遗传群体的基因型鉴定工作. 随着分子标记技术的发展, 现在已经能用上万甚至更多的SNP标记开展基因型鉴定. 因此, 人们自然关心是否可以通过加密标记的方式来更准确地定位QTL. 在关联分析中, 越多的标记, 越有利于检测群体中的剩余连锁不平衡. 加密标记会提高关联分析的QTL检测效率. 这一节介绍加密标记对连锁分析QTL检测的影响.

加密标记对单个QTL检测的影响

- 模拟三种标记密度, 即5, 10和20cM, 10条染色体, 长度均为160cM. 因此, 三种标记密度需要总的标记数分别为330, 170和90. QTL具有不同的效应, PVE设置了1%, 2%, 3%, 4%, 5%, 10%, 20%和30%八种水平. 这些QTL位于8条不同的染色体的22cM处, 即不考虑QTL之间的连锁.
- 模拟群体为RIL, 群体大小从20到500, 间隔为20. 功效统计的支撑区间长度为10cM, 即功效统计时仅考虑区间17~27cM上检测到的QTL, 该区间之外的QTL统计为假阳性. 每个QTL效应, 每个标记密度, 每个群体大小, 模拟次数为1000 (Li et al., 2010).

QTL检测功效与标记密度和群体大小的关系



加密标记对单个QTL检测的影响

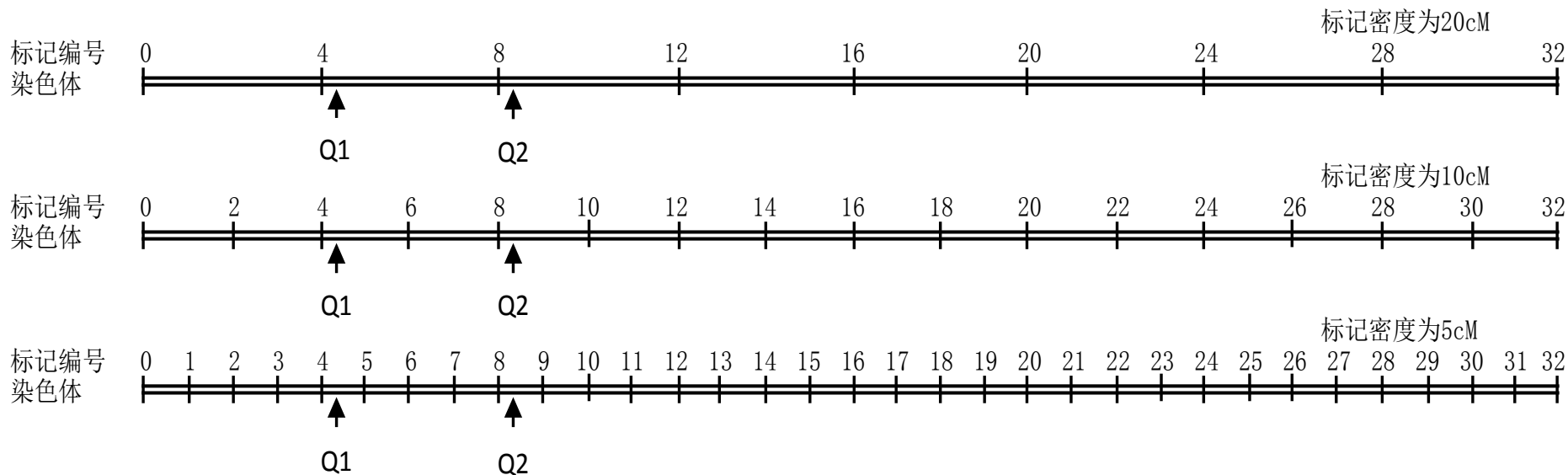
- 当标记数由密度为10cM时的170个减少到密度为20cM时的90个时, 对于PVE小于10%的QTL, 即使群体大小为500, 检测功效也有明显的下降. 对于PVE大于10%的QTL, 当群体增大到200之后, 其检测功效接近于100%.
- 但对于较小的群体, 检测功效也有明显的下降. 当标记数由密度为10cM时的170增加到密度为5cM时的330时, 只有PVE=1%的QTL的检测功效才有明显的变化. 其他QTL检测功效的增加不明显. 因此, 对于一般有100-200个体的作图群体, QTL连锁作图中每隔10cM左右就需要有一个分子标记. 增加标记有增加检测功效的作用, 但更有利于提高效应较小QTL的检测功效.
- 在关注加密标记提高检测功效的同时, 也应该注意增加标记在大群体中会产生更大的检测功效的增加. 同时, 增加标记使得假阳性QTL有增加的趋势. 因此, 只有大遗传群体才能充分体现加密标记对QTL作图的贡献.

加密标记对连锁QTL检测的影响

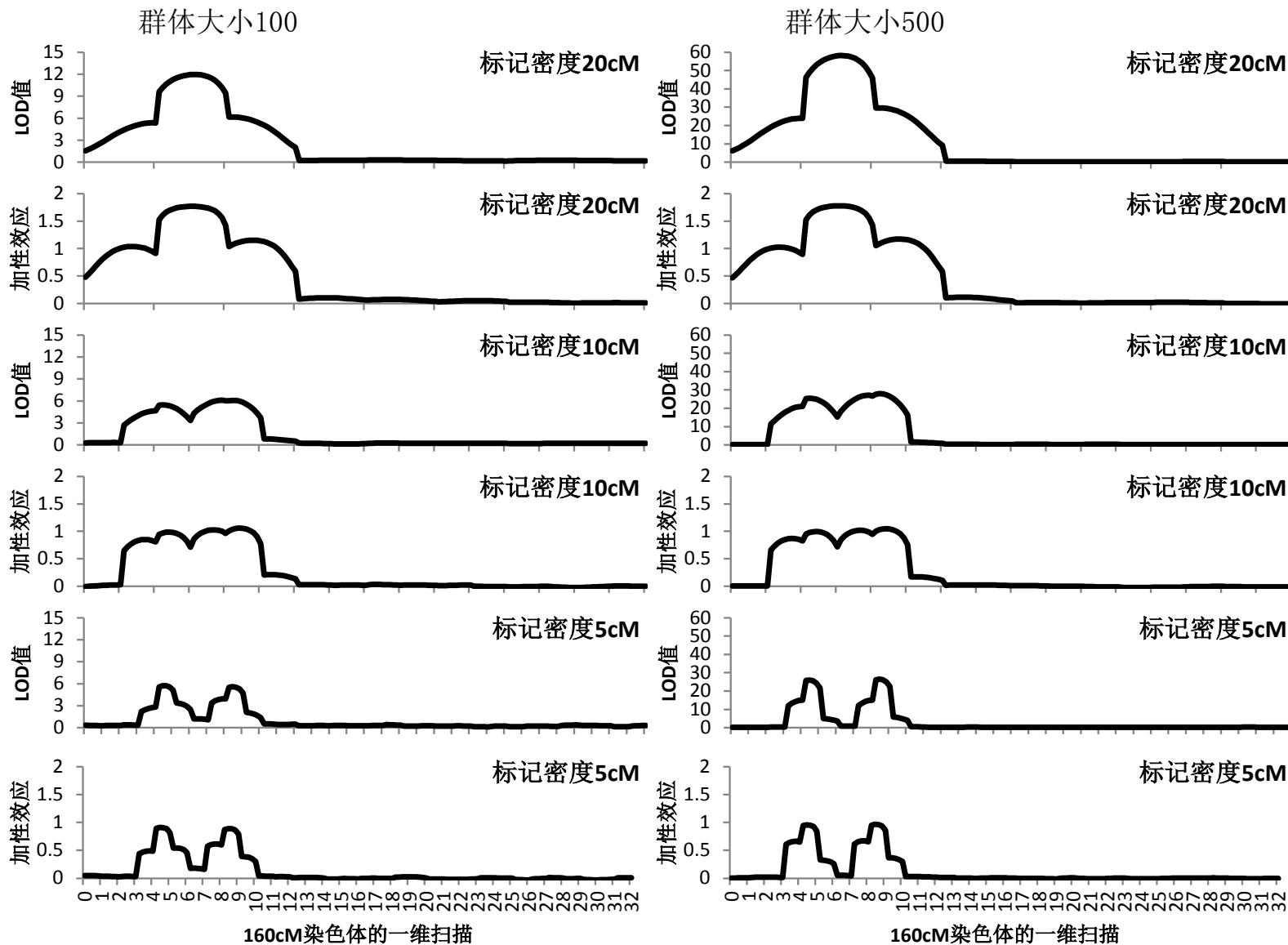
- QTL之间的紧密连锁, 极大地增加了遗传研究的难度. 但这却是遗传研究中不得不面对的问题. 在第四章建立表型对标记的线性模型的过程中, 已经知道一个QTL位置和效应的信息, 在理论上完全包含在与其连锁最紧密的左右两个标记变量的系数中. 从而在一个扫描区间上, 可以通过矫正表型的方法, 使得矫正值中只包含当前区间上QTL的信息. 以达到控制区间以外QTL影响的目的.
- 如果一个标记变量同时受两个QTL的影响, 当扫描区间含有这个标记时, 矫正值中将同时包含这两个QTL的信息. 这时, 难以把这两个QTL区分开的. 因此, ICIM作图方法一般要求连锁QTL之间要有一些空白区间 (Whittaker et al., 1996; Li et al., 2007). 所谓空白区间, 就是这些标记区间上不存在任何控制性状的QTL. 很明显, 增加标记可能使得原来没有空白区间的连锁QTL之间出现空白区间, 从而达到区分连锁QTL的目的.

距离为20cM的两个连锁QTL与三种标记密度在染色体上的示意图

染色体长度为160cM, 两个QTL位于22cM和42cM处



三种标记密度下100次模拟群体中连锁QTL的完备区间的平均作图结果



标记密度与连锁QTL的检测

- 当标记密度为20cM时, 即使群体大小为500, ICIM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效应之和的“幻影”QTL.
- 当标记密度为10cM时, 有一部分模拟群体检测到两个QTL. 但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. 当标记密度为5cM时, 大多数模拟群体中都能正确检测到预设的两个QTL. 如连锁QTL之间不存在空白区间, 更极端的情况如一个标记区间上存在两个QTL, QTL作图方法是难以将二者区分开的.
- LOD曲线反映的是两个QTL累加的效果. 这种情况下, 仅靠增加群体规模来区分连锁QTL是徒劳无益. 唯一的方法是增加标记密度, 使得连锁QTL之间至少存在两个空白区间. 只有这样, 才有可能将连锁QTL区分开来. 因此, 连锁QTL的区分需要更密的分子标记和更大的群体规模.

§ 7.4 缺失标记的填补以及缺失对QTL作图的影响

- § 7.4.1 缺失标记的填补
- § 7.4.2 一个水稻F₂群体中的株高QTL
- § 7.4.3 缺失标记对QTL检测功效的影响

缺失标记的填补

- 在知道标记间的连锁关系后, 可以根据与缺失标记连锁的无缺失标记的标记型, 来推测缺失标记型的概率. 然后, 根据缺失标记型的条件概率, 对缺失的标记型进行填补. 下面介绍 F_2 群体中缺失标记型的填补方法, 其他群体的填补方法与此类似 (Zhang et al., 2010).
- 假设所有标记都已经通过连锁分析排好顺序, M 和 m 是某个体中缺失标记处的两个标记等位基因. 从染色体或连锁群的第一个标记开始, 逐个填补缺失标记型. 可以认为, 当前标记之前的所有标记都无缺失, 或者已经对缺失值进行了填补. 缺失标记可分为如下三种情况.

缺失标记没有任何不缺失的连锁标记

- 在这种情况下, 除了知道三种标记型MM, Mm和mm的频率满足1:2:1的孟德尔分离比外, 没有任何可用的连锁信息. 这时, 只能通过抽取随机数的方法补缺失.
- 具体地说, 从均匀分布 $U(0, 1)$ 中抽取一个随机数 x . 如果 $x > 0.75$, 缺失标记基因型补为MM; 如果 $x < 0.25$, 补为mm; 否则, 补为Mm. 即以0.25的概率将缺失标记补为MM, 以0.5的概率补为Mm, 以0.25的概率补为mm.

缺失标记仅在一侧能找到连锁的未缺失标记

- 此时, 可利用缺失标记基因型在未缺失标记 (取距离最近的未缺失标记) 基因型下的条件概率 (表7.4.1) 来填补缺失. 以无缺失标记基因型AA为例. 从均匀分布 $U(0, 1)$ 中抽取一个随机数 x . 若 $x < (1-r)^2$, 则缺失标记基因型补为MM; 若 $(1-r)^2 < x < (1-r)^2 + 2r(1-r)$, 补为Mm; 否则, 补为mm.
- 很明显, 如果重组率很小, 会远大于0.25, 把缺失标记填补为MM的概率远大于0.25.

缺失标记M与一个未缺失标记A连锁时 (重组率用 r 表示), 缺失标记基因型的条件概率

无缺失标记的基因型	缺失标记的基因型		
	MM	Mm	mm
AA	$(1-r)^2$	$2r(1-r)$	r^2
Aa	$r(1-r)$	$1-2r+2r^2$	$r(1-r)$
aa	r^2	$2r(1-r)$	$(1-r)^2$

缺失标记在其左右两侧均能找到未缺失的连锁标记

- 此时, 利用缺失标记基因型在两个相邻的未缺失标记 (左右两侧分别取距离最近的那个未缺失标记) 基因型下的条件概率 (表7.4.2) 填补缺失. 以无缺失标记基因型AABB为例. 从均匀分布 $U(0, 1)$ 中抽取一个随机数 x . 若 $x < (1-r_1)^2(1-r_2)^2/(1-r)^2$, 则缺失标记基因型补为MM; 若 $(1-r_1)^2(1-r_2)^2/(1-r)^2 < x < (1-r_1)^2(1-r_2)^2/(1-r)^2 + 2r_1 r_2 (1-r_1)(1-r_2)/(1-r)^2$, 补为Mm; 否则, 补为mm.
- 很明显, 如果两个未缺失标记之间的重组率很小, 会远大于0.25, 把缺失标记填补为MM的概率远大于0.25.

两个未缺失标记连锁时, 缺失标记基因型的条件概率

无缺失标记的基因型	边缘频率	缺失标记的基因型		
		MM	Mm	mm
AABB	$(1-r)^2$	$(1-r_1)^2(1-r_2)^2$	$2r_1r_2(1-r_1)(1-r_2)$	$r_1^2r_2^2$
AABb	$r(1-r)$			
AAbb	r^2			
AaBB	$r(1-r)$			
AaBb	$(1-2r+2r^2)$			
Aabb	$r(1-r)$			
aaBB	r^2			
aaBb	$r(1-r)$			
aabb	$(1-r)^2$			

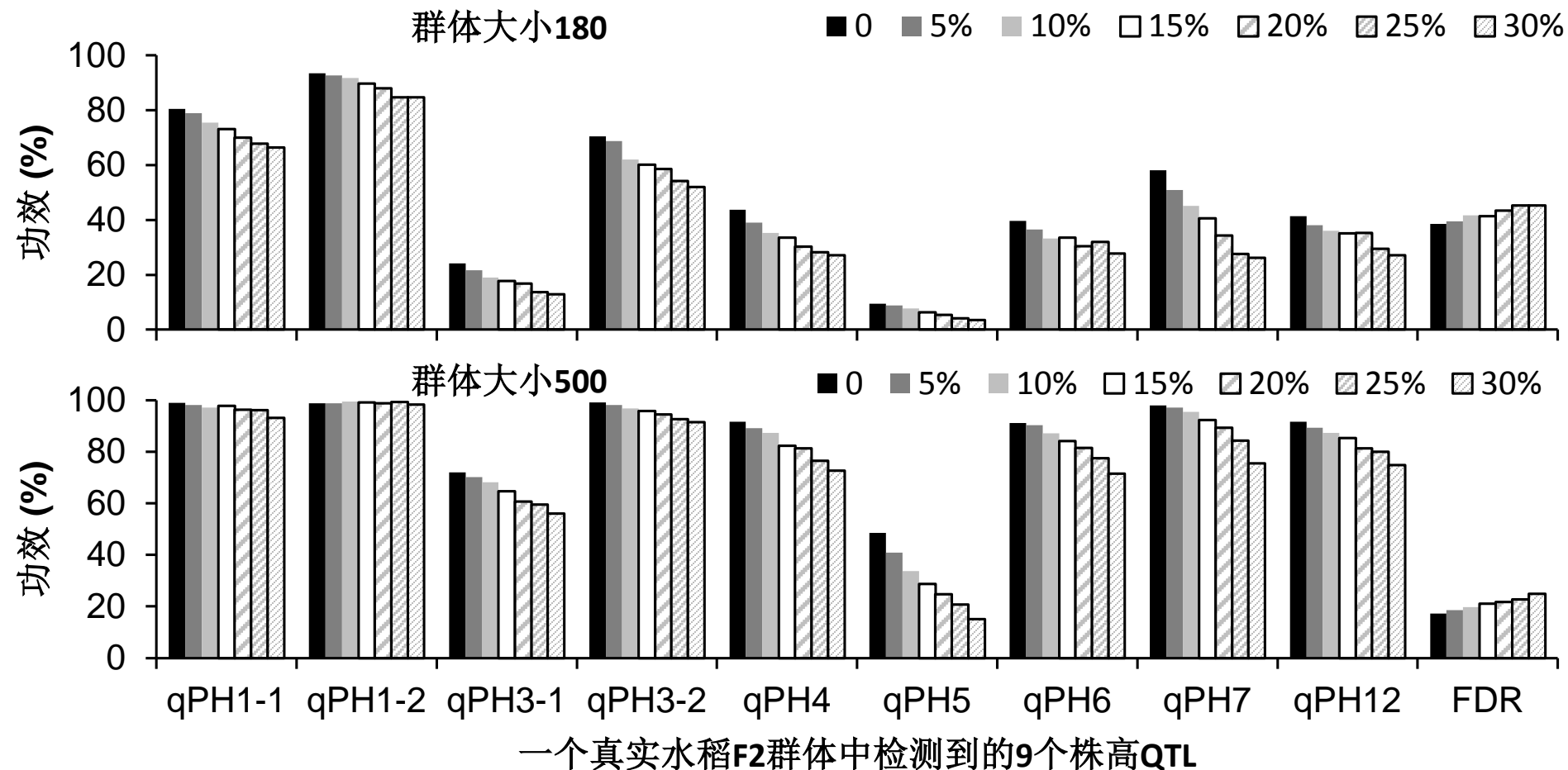
一个水稻F₂群体

- 一个水稻F₂群体中包含180个个体 (叶少平等, 2005; Zhang et al., 2010), 杂交组合的两个亲本是籼稻品种PA64s和粳稻品种Nipponbare. 2002年Nipponbare已实现完全测序, PA64s在同一年也实现了部分测序.
- 利用137个多态性SSR标记构建连锁图谱, 12条染色体上分布着6~12个标记, 遗传图谱的总长度为2046.2cM, 平均标记间距离为17.1cM. 该F₂群体中共有24,660个标记点 (即180 × 137), 其中5,131个是PA64s标记型, 6,175个是Nipponbare标记型, 11,114个是杂合型. 所有标记基因型数据中, 有2,240个缺失标记型, 占有所有标记数据的9.08%.
- 亲本PA64s里有一个主要的矮秆基因, 其株高是74.4 cm. 亲本Nipponbare的株高是98.3 cm.

一个水稻F₂群体的株高QTL

QTL编号	标记区间	左标记距离 (cM)	加性效应	显性效应	LOD 值	PVE (%)	显性度
qPH1-1	RM246—RP2	12.0	-0.57	-7.98	8.04	12.03	13.96
qPH1-2	RP82—RP3	19.5	-8.59	0.59	15.54	25.57	-0.07
qPH3-1	RM523—RM251	16.9	4.35	-4.86	6.51	13.30	-1.12
qPH3-2	RP242—RM520	11.4	-4.69	-1.00	5.04	6.84	0.21
qPH4	RP67—OSR15	13.7	-3.56	-2.09	4.61	5.53	0.59
qPH5	RM159—RP299	13.0	-0.44	-4.48	3.13	3.86	10.24
qPH6	RP199—RM276	6.2	-0.79	-5.05	3.17	4.96	6.36
qPH7	RM82—RM180	7.0	0.26	6.48	5.27	7.56	25.24
qPH12	RM19—RM247	2.4	-1.66	3.93	3.98	5.44	-2.36

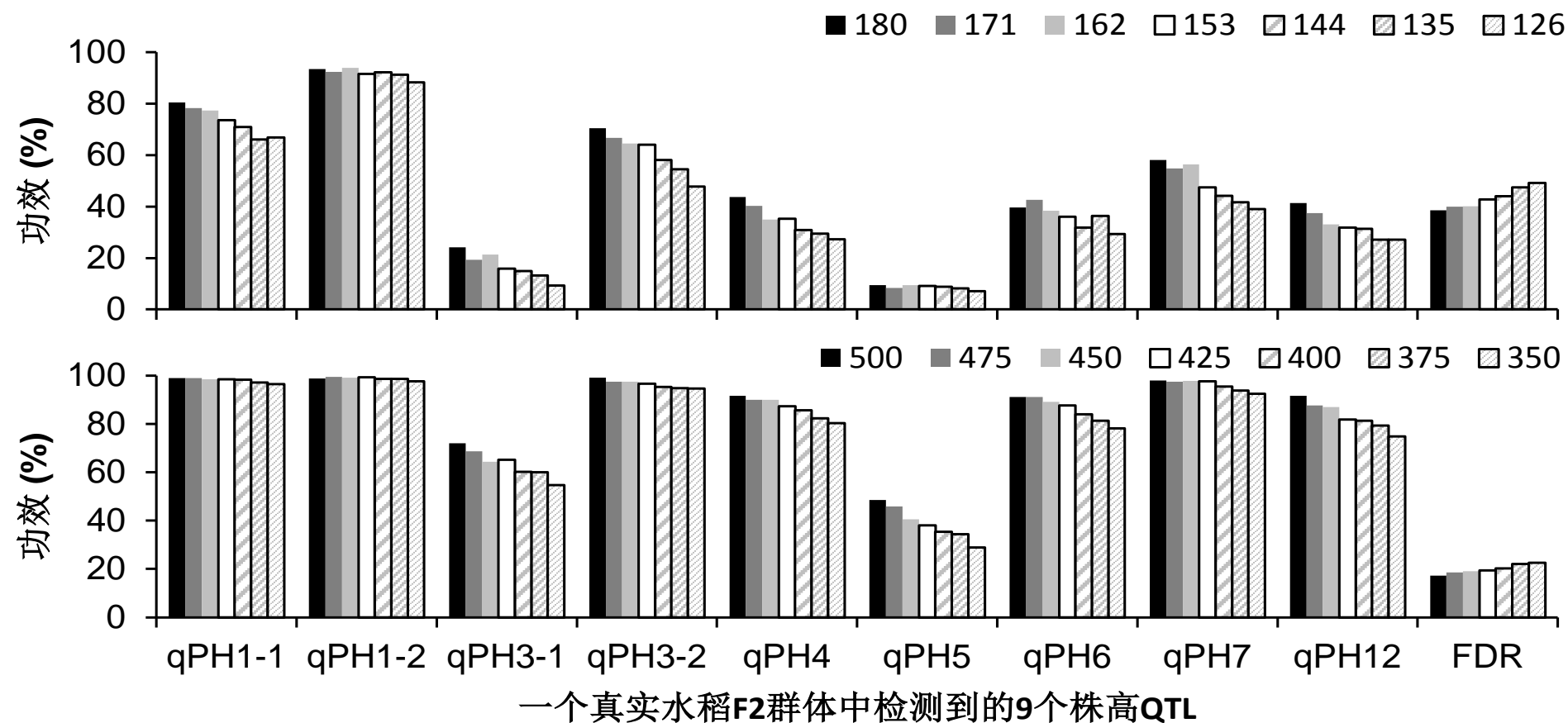
F₂群体中缺失标记对QTL检测功效的影响



无缺失标记的不同F₂群体大小对QTL检测功效的影响

上图的群体大小从180开始, 然后逐渐减少5%, 10%, 15%, 20%, 25%和30%.

下图的群体大小从500开始, 然后逐渐减少5%, 10%, 15%, 20%, 25%和30%.



缺失标记对QTL检测功效的影响

- 当模拟群体大小为180 (与真实群体一致) 时, 随着缺失率的增加, 所有QTL的检测功效会逐步降低, 假阳性QTL的比例呈逐渐上升的趋势.
- 当模拟群体大小为500时, 随着缺失率的增加, 效应较大的QTL的检测功效变化不明显, 效应较小QTL的检测功效有明显的降低. 与模拟群体大小180类似, 假阳性QTL的比例呈逐渐上升的趋势. 因此, 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.
- 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.

缺失标记对QTL检测功效的影响

- 利用水稻 F_2 群体构建的连锁图谱, 以及表7.4.3列出的QTL建立遗传模型, 模拟无缺失标记的不同大小 F_2 群体.
- 当模拟群体大小从180 (与真实群体一致) 开始, 然后逐渐减少5%, 10%, 15%, 20%, 25%和30%的情况下, 除PVE最高的 $qPH1-2$ 外, 所有QTL的检测功效逐步降低, 假阳性QTL的比例呈逐渐上升的趋势.
- 当模拟群体大小从500开始, 然后逐渐减少5%, 10%, 15%, 20%, 25%和30%的情况下, 效应较大的QTL的检测功效变化不明显, 效应较小QTL的检测功效有明显的降低. 与模拟群体大小从180开始的情况类似, 假阳性QTL的比例呈逐渐上升的趋势. 因此, 群体大小对效应较小的QTL的影响较大.

缺失标记对QTL检测功效的影响

- 比较图7.4.1和图7.4.2的检测功效和假阳性发现率可以发现, 一定比例的标记缺失对QTL作图结果的影响相当于一定比例个体缺失对QTL作图的影响. **也就是说, 一个群体大小为 n , 标记型缺失率为 p 的群体的作图功效, 与大小为 $n(1-p)$, 无标记型缺失群体的作图功效大致相同.**
- 因此, 尽管大多数QTL作图方法和软件可以处理带有缺失标记的数据, 在遗传群体的基因型鉴定的过程中, 还是要尽可能避免缺失数据的发生, 从而得到更好的遗传研究结果. 缺失标记的填补只是一种技术手段. 如果不利用其他额外信息, 就填补本身来说, 是不会增加QTL检测功效的.
- 当然, 如果采用对缺失的标记型重新进行基因型分析的方式进行填补, 这时相当于利用额外信息对缺失标记进行填补, 其结果是降低标记数据的缺失率. 采用这种方式填补缺失标记, 可以增加QTL的检测功效, 提高遗传分析的准确性.

§ 7.5 奇异分离对遗传研究的影响

- § 7.5.1 一个水稻F₂群体中的奇异分离标记
- § 7.5.2 奇异分离在三种基因型群体中对QTL作图的影响
- § 7.5.3 奇异分离影响的距离
- § 7.5.4 奇异分离在两种基因型群体中对QTL作图的影响

奇异分离

- 奇异分离几乎存在于所有的遗传研究群体中, 是一个广泛研究的遗传学问题 (Hedrick and Muona, 1990; Lorieux et al., 1995; Hackett and Broadfoot, 2003; Luo et al., 2005; Zhu et al., 2007; Xu, 2008).
- 在 § 7.4 的水稻F₂群体中, 对每个标记做1:2:1的孟德尔分离比适合性检验, 在0.01的显著性水平下, 存在9个显著偏离孟德尔分离比的标记 (表7.5.1). 这些标记分布在水稻第2, 3, 5, 8, 10, 11和12染色体上 (Zhang et al., 2010). 显著性检验中, $-\log_{10}(P)$ (P 为显著性水平概率) 的最大值是20.02, 对应于第12条染色体上的标记RP129. RM304和RP129周围的一些标记也是显著的, 但这很可能是由连锁造成的, 因此没有在表7.5.1中给出.

一个水稻F₂群体中的奇异分离标记

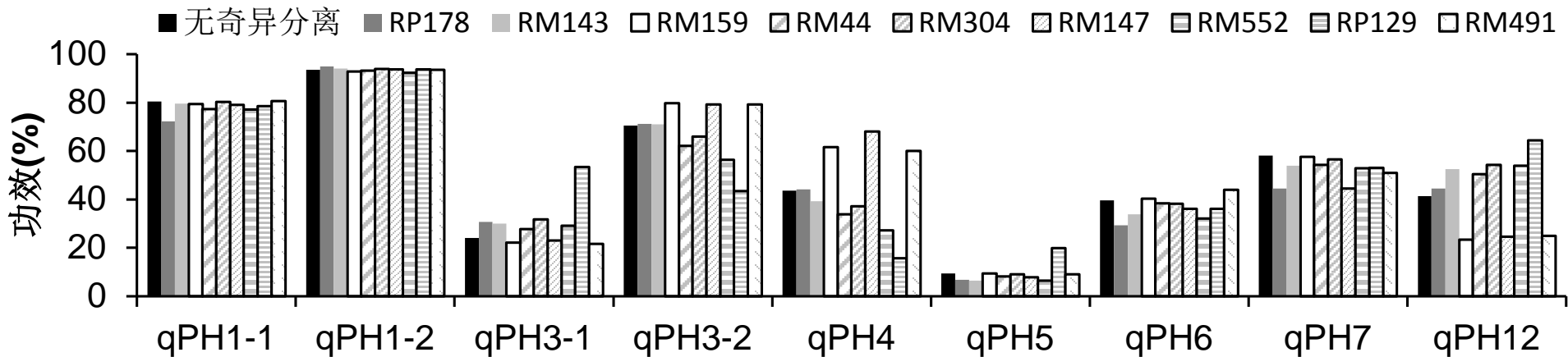
标记名称	染色体	样本量			M:m	-log ₁₀ (P)	适合度		
		MM	Mm	mm			MM	Mm	mm
RP178	2	72	33	27	2.03	13.83	1.00	0.23	0.38
RM143	3	85	64	27	1.98	11.14	1.00	0.38	0.32
RM159	5	32	33	79	0.51	15.84	0.41	0.21	1.00
RM44	8	64	83	22	1.66	4.54	1.00	0.65	0.34
RM304	10	75	78	25	1.78	6.69	1.00	0.52	0.33
RM147	10	39	27	78	0.57	16.80	0.50	0.17	1.00
RM552	11	57	110	13	1.65	6.60	1.00	0.96	0.23
RP129	12	92	87	1	3.04	20.02	1.00	0.47	0.01
RM491	12	27	28	60	0.55	10.69	0.45	0.23	1.00

适合度及其估计

- 适合度是群体遗传学中研究选择效应的重要参数. 适合度是指某基因型繁殖后代的相对能力, 取值在0和1之间. 假定两种基因型AA和aa的个体各有100个, 有10个基因型AA的个体和90个aa基因型的个体成活下来. 那么, aa的繁殖成活率是AA的0.9倍. 这时, 认为AA的适合度为1, aa的适合度为0.9.
- 对于一个F₂群体中的三种基因型AA, Aa和aa, 如果适合度均为1, 三种基因型的频率将分别为0.25, 0.5和0.25. 否则的话, 三种基因型将偏离1:2:1的孟德尔分离比. 因此, 可以通过基因型的相对频率, 定义三种基因型的适合度. 假定M来自亲本PA64s, m来自亲本Nipponbare. 用 n_{MM} , n_{Mm} 和 n_{mm} 表示三种标记型MM, Mm和mm的样本量, 用 n_{MAX} 表示 n_{MM} , $0.5 * n_{Mm}$ 和 n_{mm} 三者中的最大值. 那么, 三种标记型的适合度分别为 n_{MM}/n_{MAX} , $0.5 * n_{Mm}/n_{MAX}$ 和 n_{mm}/n_{MAX} .

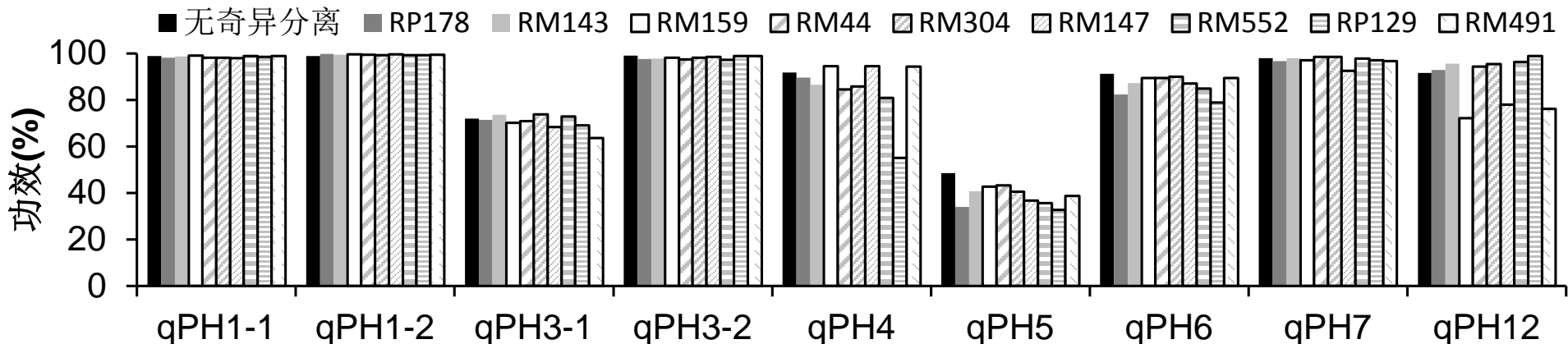
F₂群体中存在奇异分离时的QTL检测功效

群体大小180



一个真实水稻F₂群体中检测到的9个株高QTL

群体大小500



一个真实水稻F₂群体中检测到的9个株高QTL

有、无奇异分离的方差比值

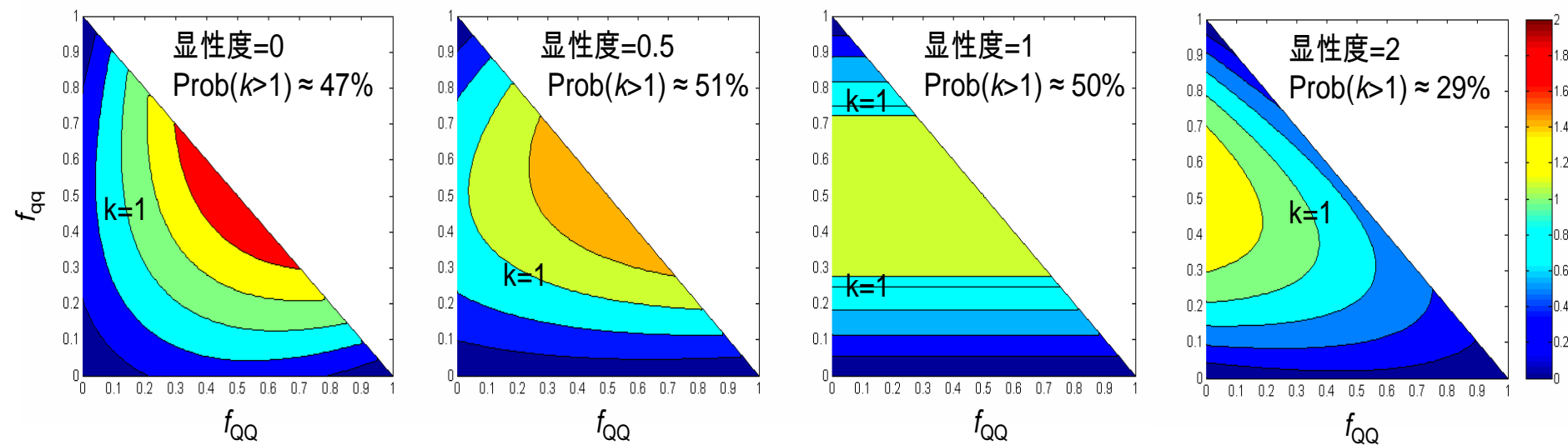
- 将QTL的显性度用 $s=d/a$ 表示. 从等式 (7.5.1) 和 (7.5.2) 可以得到有无奇异分离情况下QTL的两个遗传方差的比值为,

$$k = \frac{\sigma_{SD}^2}{\sigma^2} = \frac{4[f_{QQ} + f_{qq} - (f_{QQ} - f_{qq})^2] - 8f_{Qq}(f_{QQ} - f_{qq})s + 4(f_{Qq} - f_{Qq}^2)s^2}{2 + s^2}$$

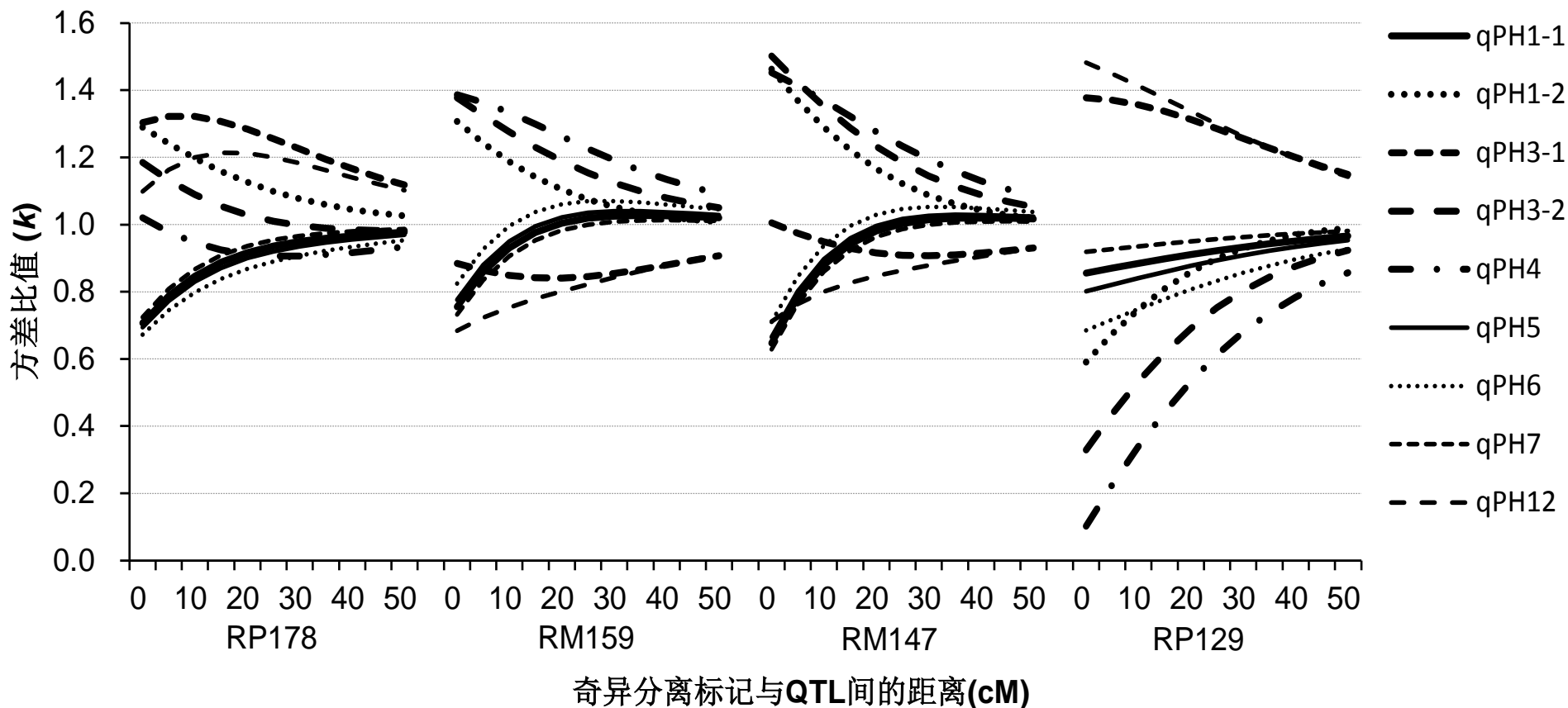
- 统计学上, 一个决定因素的不同水平之间的方差越大, 这些水平间的差异越易于检测. 不同水平间较大的差异会造成高的显著性概率, 因此也会有较高的检测功效. QTL作图也是如此.
- 如果一个QTL的遗传方差很大, 这个QTL就会产生很高的LOD值, 也就容易被检测到. 等式 (7.5.3) 给出了有无奇异分离QTL遗传方差的相对变化, 可用于定量解释奇异分离对QTL作图的影响. 从直观上看, 若 k 大于1, 奇异分离会引起QTL遗传方差的增加, 因此有益于QTL的检测. 若 k 小于1, 奇异分离会引起QTL遗传方差的降低, 因此不利于QTL的检测. 若 k 等于1, 奇异分离不会影响到QTL的检测 (Zhang et al., 2010).

F₂群体中不同基因型频率的遗传方差与无奇 异分离时方差的比值

从左到右代表四种显性度, 即无显性, 部分显性, 完全显性和超显性.



F₂群体中奇异分离标记与QTL之间连锁 距离对QTL遗传方差比值的关系



奇异分离影响的距离

- 根据表7.5.2, 可以计算不同连锁距离时QTL基因型的频率. 进而计算不同连锁距离下遗传方差比值 k , 并由此推测一个奇异分离所能影响的距离.
- 以表7.5.1中的四种奇异分离标记为例, 图7.5.3给出QTL遗传方差比值随标记与QTL间遗传距离的变化曲线. 可以看出, 当奇异分离标记与QTL紧密连锁时, 有些QTL的 k 值远低于1, 有些QTL的 k 值远高于1. 随着遗传距离的增加, 所有QTL的方差比值都趋向于1. 说明奇异分离对遗传分析的影响变得越来越小.
- 从图7.5.3可以看出, 当标记与QTL间遗传距离超过20cM时, k 值将在0.8到1.2的范围内. 如果这个范围内的 k 值可以忽略奇异分离影响的话, 可以认为, 奇异分离的影响距离不超过20cM (Zhang et al., 2010).

奇异分离在两种基因型群体中对QTL作图的影响

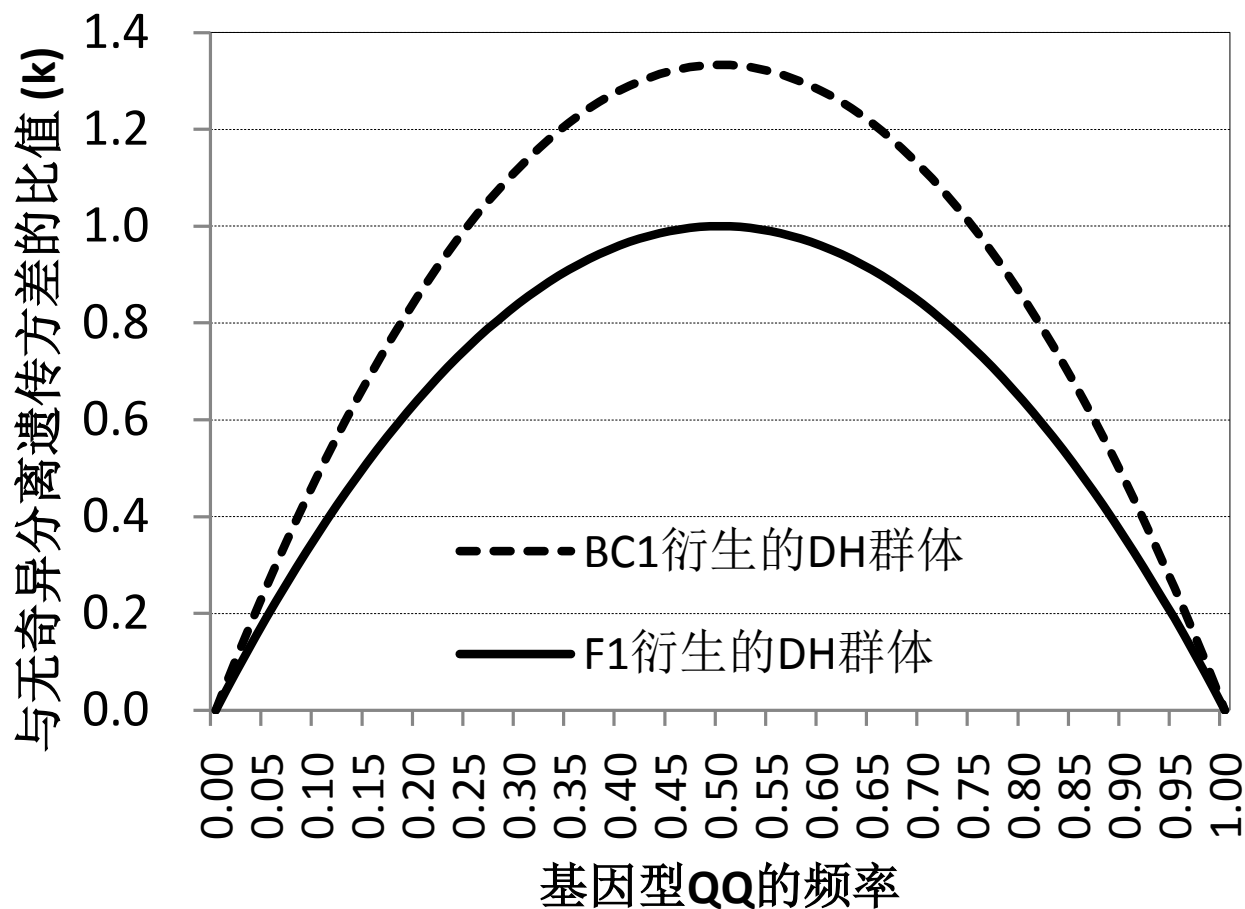
- 对于仅有两种基因型的群体, 奇异分离对QTL作图的影响显得简单些 (李慧慧等, 2010). 假定真实群体中, 两种QTL基因型的频率分别用 p 和 $1-p$ 表示. 无奇异分离群体中QTL基因型频率分别用 f 和 $1-f$ 表示. 那么, QTL在两种情况下遗传方差的比值为,

$$k = \frac{\sigma_{SD}^2}{\sigma^2} = \frac{p(1-p)}{f(1-f)} = \frac{1-(1-2p)^2}{1-(1-2f)^2}$$

奇异分离在两种基因型群体中对QTL作图的影响

- 在 F_1 衍生的RIL或DH群体中, 两种基因型的期望频率比为1:1. 这时, $f=0.5$, $k=4p(1-p)$. 因此, 只有在 $p=0.5$ 的情况下才有 $k=1$. 任何奇异分离都会使 k 小于1, 从而降低QTL的作图功效.
- 如果两种基因型的期望比不为1:1, 例如回交衍生的重组近交家系或回交衍生的加倍单倍体群体中, 无奇异分离时两种基因型频率满足3:1的分离比. 这时, 如果奇异分离能引起两种基因型偏向于频率1:1的方向, 就会增加QTL的遗传方差. 因此, 方差的比值 k 就会大于1, 奇异分离会有利于QTL的检测.

BC₁和F₁产生的DH作图群体中, 奇异分离QTL的遗传方差与无奇异分离时的遗传方差的比值



§ 7.6 数量性状表型分布的非正态性

- § 7.6.1 数量性状的表型模型与表型分布
- § 7.6.2 非正态表型分布的QTL作图

随机误差效应的正态性假定

- 大部分QTL作图方法只是要求表型数据中的随机误差项服从正态分布, 并没有要求表型数据满足正态分布. 数量性状只有在多基因假说下才真正符合正态分布, 表型数据的非正态性并不影响QTL作图.
- 误差理论的基石是误差效应服从均值为0的正态分布. 这一假定也是大多数统计理论的基石.
- 一个遗传群体中的表型不一定服从正态分布.

数量性状的表型模型与表型分布

- 个体在特定环境下的表型是个体遗传效应值 G 和随机误差效应 ε 之和. 这就是纯系理论的主要内容, 用一个线性统计模型表示为: $P=G+\varepsilon$, 其中, 随机误差效应 ε 服从均值为0和一个特定方差 σ_ε^2 的正态分布, 即 $\varepsilon\sim N(0, \sigma_\varepsilon^2)$.
- 从上面的表型模型可以看出, 对于一个特定的基因型来说, 其表型重复观测值服从正态分布. 这一要求也是绝大多数统计分析方法的前提.
- 在一个遗传群体中, 不同基因型有着不尽相同的基因型值, 基因型值之间差异是开展遗传研究的前提. 不同基因型的表型 P 服从的分布不仅与随机误差效应 ε 的分布有关, 而且还依赖于不同基因型值 G 构成的分布.

数量性状的基因型值模型

- 在加显性模型下, 包含 q 个QTL的遗传模型为,

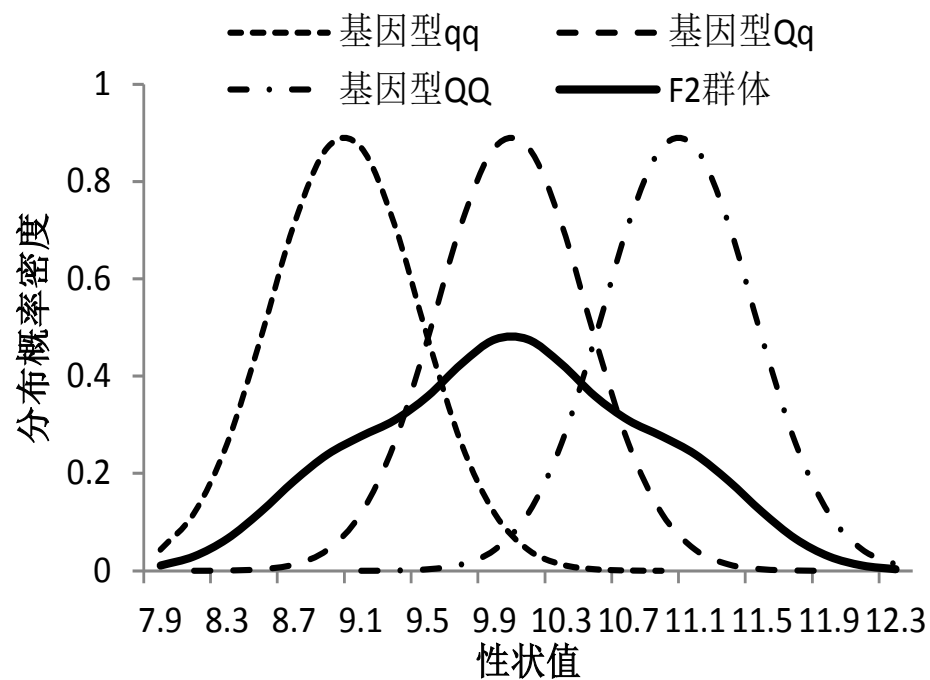
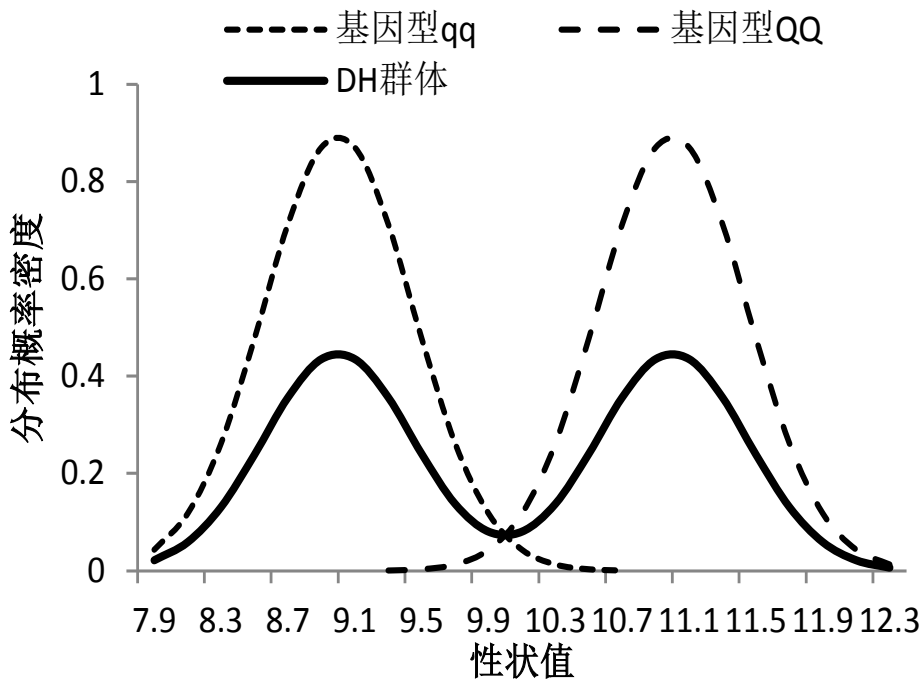
$$G = m + \sum_{j=1}^q [a_j w_j + d_j v_j]$$

- 其中, m 是所有可能纯合基因型的平均表现, w_j 和 v_j 是第 j 个QTL基因型的指示变量, 对于基因型QQ取1和0, 对于基因型Qq取0和1, 对于基因型qq取-1和0. 不同的遗传群体, 有着不同的基因频率和基因型频率. 因此, 模型 (7.6.2) 中的 G 有不同的取值, 这些不同的取值还有着不同的频率.

数量性状的表型分布

- 在多基因假说下, 群体的基因型值 G 是一个近似正态分布, 再加上具有正态分布的随机效应 ε 的修饰, 群体的表型才会呈现出正态分布.
- 在多基因假说不成立的情况下, 如QTL的个数较少, 并存在少数遗传效应较大的QTL, 模型 (7.6.2) 中的 G 不服从正态分布, 甚至与正态分布相去甚远. 这时, 即使有正态分布随机效应 ε 的修饰, 模型 (7.6.1) 中的 P 也不会服从正态分布.

一个QTL的不同基因型在DH (左) 和 F_2 (右) 群体中的分布

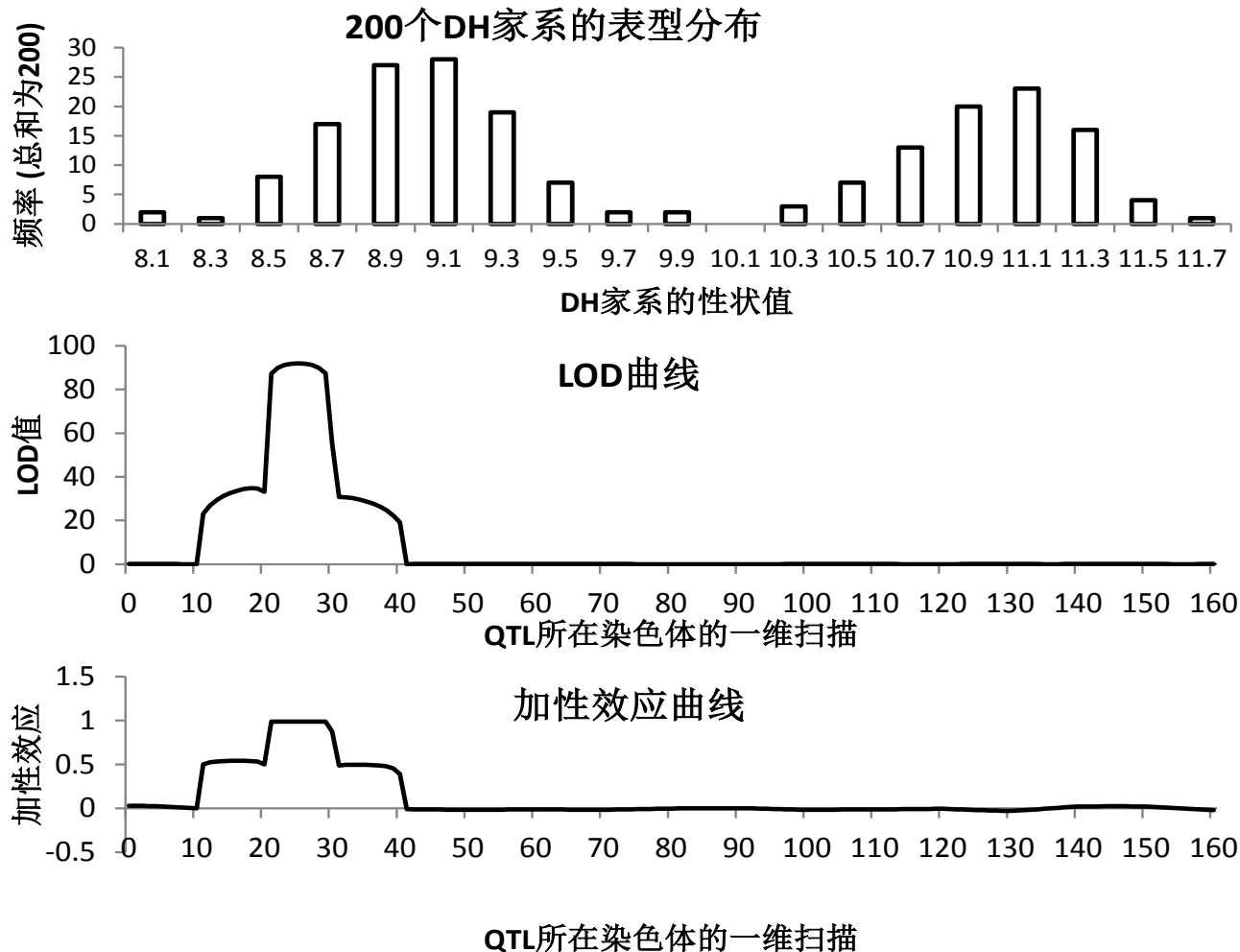


非正态表型分布DH群体的QTL作图

- 利用QTL IciMapping软件, 产生一个包含200个DH家系的作图群体. 表型数据分布的范围为8.1~11.7, 但有更多DH家系的表型分布位于9和11附近, 表型明显不服从正态分布.
- 表型分布的均值为10.04, 方差为1.15, 接近于理论值10和1.2. ICIM一维扫描结果表明LOD在25cM处达到峰值92.01, 对应位置的遗传效应估计值为0.9867, 解释81.12%的表型变异, 接近于真实加性效应值1和广义遗传力83.3%.

一个表型非正态的模拟DH群体中性状的分布 (上), QTL作图LOD曲线 (中), 以及QTL作图的加性效应曲线 (下)

QTL存在于25 cM处, 加性效应为1, 显性效应为0, 随机误差方差为0.2, 作图群体中包含200个加倍单倍体家系

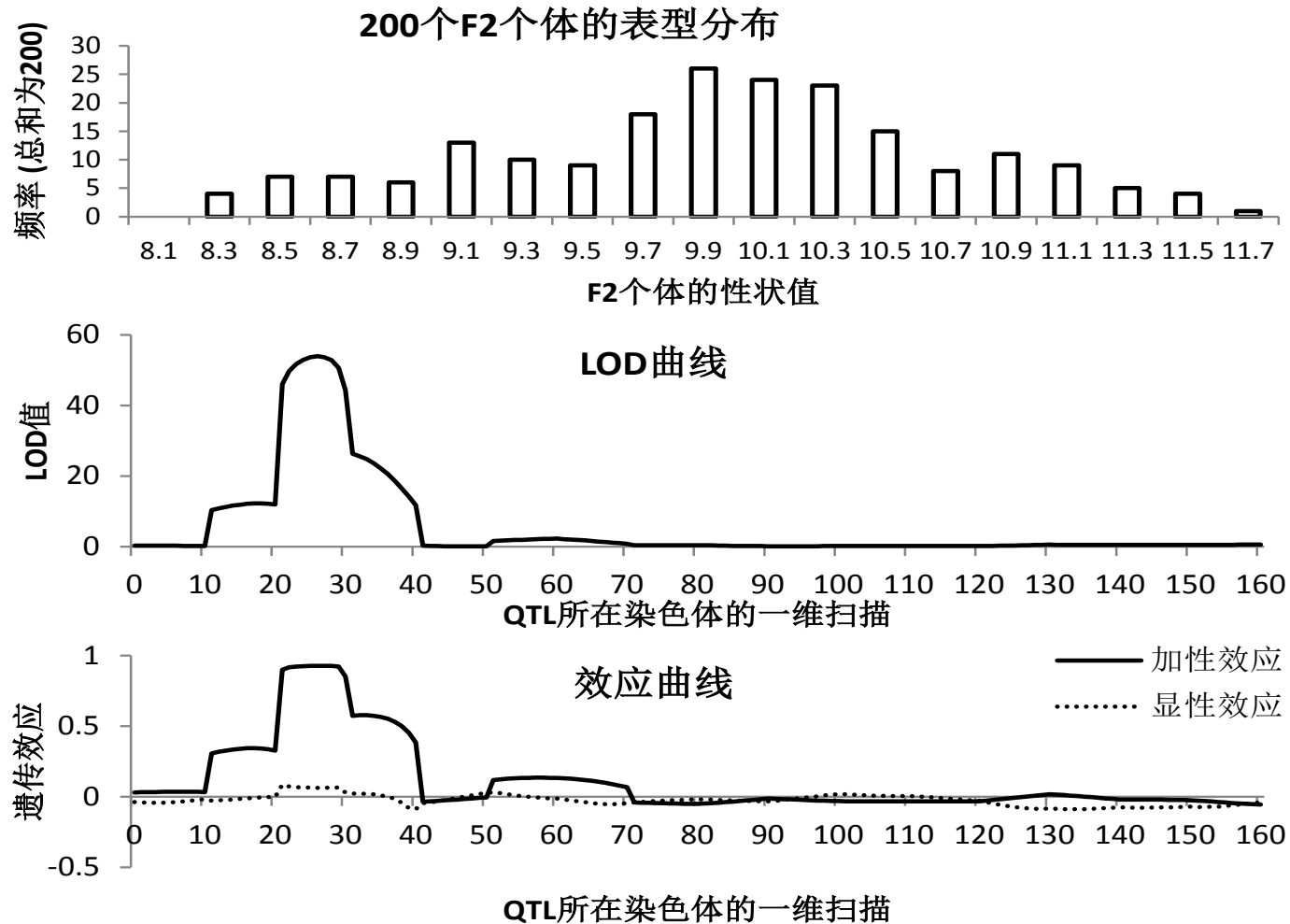


非正态表型分布F₂群体的QTL作图

- 利用QTL IciMapping软件产生一个包含200个个体的F₂作图群体. 表型数据分布为8.3~11.7, 但有更多个体的表型分布位于10附近. 表型分布有两个较长的尾部, 明显不服从正态分布.
- 表型分布的均值为10.03, 方差为0.73, 接近于理论值10和0.7. ICIM一维扫描结果表明LOD在26cM处达到峰值53.91, 对应位置的加显遗传效应估计值分别为0.9269和0.0632, 接近于真实加显性效应值1和0. 该QTL解释68.46%的表型变异, 接近于广义遗传力71.4%.

一个表型非正态的模拟F2群体中性状的分布 (上), QTL作图LOD曲线 (中), 以及QTL作图的加性效应曲线 (下)

QTL存在于25 cM处, 加性效应为1, 显性效应为0, 随机误差方差为0.2, 作图群体中包含200个个体.



误差效应的正态性检验

- 从前面的两个模拟群体不难看出, 数量性状表型分布是否为正态性不是QTL作图的决定因素. QTL作图也不以表型正态分布为先决条件, 表型分布的非正态性不影响QTL位置和效应的估计.
- 但是, 与大多数统计分析方法一样, 随机误差效应一般要求满足正态分布. 测量误差在大多数情况下都满足或近似服从正态分布, 这一点已被大量的观测数据所证实.
- 对于QTL作图群体来说, 如果没有重复观测数据, 检验误差效应分布的正态性不太可能. 但是, 如果有重复观测数据, 可利用观测值与重复平均数的离差作为随机效应的估计值, 从而对随机效应服从的分布进行检验.