

# 第13章

# 数量性状基因定位

王建康

中国农业科学院作物科学研究所

[wangjiankang@caas.cn](mailto:wangjiankang@caas.cn)

<http://www.isbreeding.net>

# 数量性状基因定位（或QTL作图）

- 随着分子标记技术的发展，目前人们已经可以像研究质量性状基因那样研究数量性状基因，也可以把数量性状基因（quantitative trait gene or locus，简称QTL）定位在染色体上，并估计单个座位上基因的遗传效应。寻找QTL在染色体上的位置并估计其遗传效应的过程，称为QTL作图或定位（QTL mapping）。
- 自Lander和Botstein（1989）提出区间作图方法以来，QTL定位已经成为数量遗传学的研究重点，也成为动植物数量性状遗传分析的主要方法。根据定位结果对数量性状基因进行精细定位、图位克隆，利用定位到的标记对数量性状进行辅助选择等，都有成功的例子。

# 本章的主要内容

- § 13.1 QTL作图群体和作图原理
- § 13.2 简单区间作图方法
- § 13.3 具有背景控制的QTL作图方法
- § 13.4 集成软件QTL IciMapping简介

# § 13.1 QTL作图群体和作图原理

- § 13.1.1 QTL作图的遗传群体和数据类型
- § 13.1.2 QTL作图的基本原理

# QTL作图的遗传群体

- 可以这么说，任何存在遗传变异的群体基本都可用于QTL定位研究。QTL作图群体包括双亲或多亲本杂交产生的各种分离后代，也包括自然群体和具有一定亲缘关系的家系群体。
- 按照基因型是否纯合，可以把它们分为暂时群体和永久群体两类。暂时群体中，个体的基因型是杂合的，自交繁殖后基因型将发生变化。双亲杂交产生的 $F_2$ 和回交都可被看作暂时群体。

# 永久遗传研究群体

- 永久群体中，每个个体具有纯合的基因型，自交繁殖后基因型不再发生变化。常见的永久群体有重组近交家系（RIL）、加倍单倍体（DH）和染色体片段置换系（CSSL）等。
- RIL群体是由 $F_2$ 个体连续自交，直至纯合而得到的家系群体，产生RIL群体的常用方法是一粒传。
- DH群体是由 $F_1$ 、 $F_2$ 或其它自交世代的配子直接加倍纯合得到的群体。
- CSSL群体是通过 $F_1$ 多代回交，并对供体亲本（也可以同时对背景亲本）的染色体进行选择而得到的。

# 永久群体的表型鉴定优势

- 永久群体中，每个家系都可以通过自交繁殖永久地保存下去，家系内不再发生基因型分离，故称永久群体。当然，能够进行无性繁殖的群体，其实也可以被看作是永久群体。
- 利用永久群体，可以在不同年份（季节）、不同地点下，开展有重复的数量性状表型鉴定试验。因此能够比较准确地获得个体或家系的基因型值，比较准确地定位QTL，并分析QTL表达的稳定性及研究QTL和环境之间的互作效应。

# 自然群体与关联分析

- 对动物和人类的遗传研究来说，可供利用的只是自然条件下的随机交配群体。利用自然群体中的剩余连锁不平衡开展QTL作图研究，又称关联分析（association mapping）。
- 关联分析的主要问题是连锁不平衡的程度很低（见 § 1.5.2），需要对全基因组大量标记（数万甚至数十万个）进行筛选，才有可能找到与目标基因连锁的标记。因此，关联分析又称全基因组关联研究（genome-wide association study，简称GWAS）。关联分析的另一个问题是难以排除群体混合造成的标记与性状假关联现象（见 § 1.5.5）。
- 为了与自然群体的关联分析相区分，基于双亲或多亲杂交群体、动物或人类核心家系的基因定位方法，有时又称为连锁作图（linkage mapping）。



# 初级群体与次级群体

- 前面的作图群体分类，基于自交繁殖过程中群体的遗传结构是否发生改变。也有研究把作图群体分为初级群体和次级群体。
- 初级群体（primary population）一般是由两个亲本杂种 $F_1$ 代衍生的未经选择的遗传群体，其中的各种基因型期望频率是已知的。
- 次级群体（secondary population）经人为选择而产生，如经过多代回交、供体亲本和背景亲本定向选择产生的近等基因系或染色体片段置换系群体，也可以是某一RIL或DH家系与亲本再杂交后形成的群体。次级群体一般用于QTL的精细定位和克隆研究。

# 作图群体的基因型和表型鉴定

- 开展QTL作图，首先需要创建一个或数个作图群体，然后对群体中每个分离个体或分离纯合家系进行分子标记的基因型检测，同时通过多环境试验对所关心的数量性状进行表型鉴定。
- 在获得作图群体的基因型和表型数据之后，就可以利用专业计算机软件开展QTL定位研究。如果定位方法需要连锁图谱，则需要首先根据群体的基因型数据构建。

# 基因型数据的编码

- 为方便起见，标记基因型一般用代码表示。以DH群体为例，两种亲本的带型分别用2和0编码，缺失带型用-1表示。如是暂时群体，则用1表示杂合型。
- 当然，双亲后代中的三种可能基因型和缺失，也可以用A（或AA）、B（或BB）、H（或AB）、X（或XX或\*）等其它编码方式表示。使用不同的编码方式，不会影响QTL作图结果。但要注意，不同QTL作图软件允许的编码方式可能存在差异，只有采用软件接收的、正确的编码方式，才能得到预期的分析结果。
- 表型数据一般用数字表示，对于一些分级调查性状，也要将其转换成数字才能进行QTL作图。表型数据也允许缺失，不同作图软件对缺失表型的表示方式也会有差异。

# 一个大麦DH作图群体

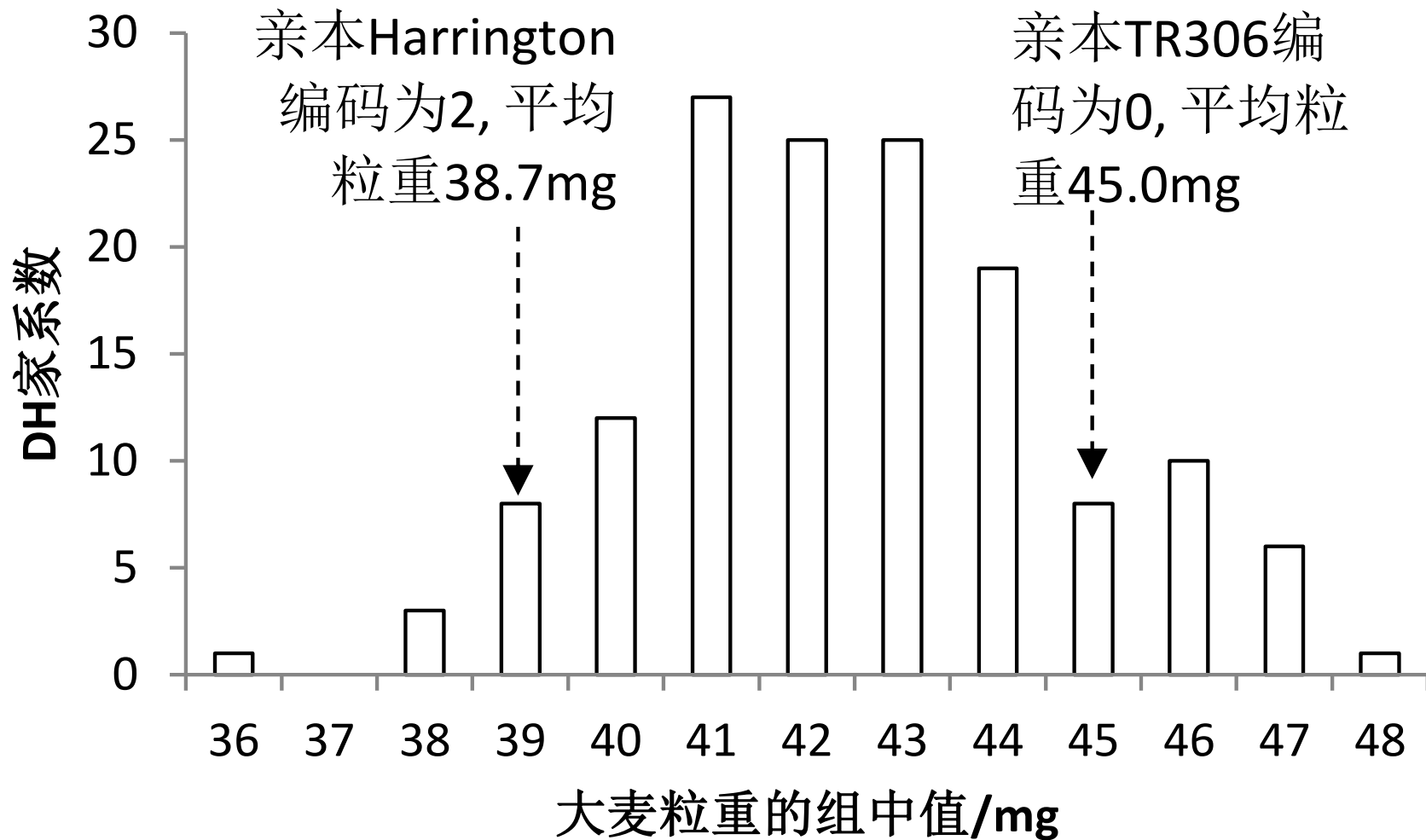
- 纯合亲本 ‘Harrington’ 和 ‘TR306’ 衍生的145个大麦 (*Hordeum vulgare* L.) DH家系构成的群体，是国际上的一个知名QTL作图群体。
- 利用该群体127个标记的基因型数据，已构建了均匀覆盖大麦7条染色体（用1H至7H表示）的遗传连锁图谱。
- 1992~1993年，在17个地点共25个环境条件下，评价了多种数量性状的表现 (Tinker et al., 1996)。这一章里，自始至终利用这个群体的127个标记数据和平均粒重作为实例。

# 10个DH家系在1H染色体的14个标记型和平均粒重

亲本 ‘Harrington’ 用2编码，亲本 ‘TR306’ 用0编码，-1表示缺失基因型。表型无缺失

分子标记	位置/cM	DH1	DH2	DH3	DH4	DH5	DH6	DH7	DH8	DH9	DH10
Act8A	0	0	2	-1	2	0	2	2	0	0	2
OP06	10.9	0	2	2	2	0	2	2	0	0	2
aHor2	18.5	0	2	2	2	0	2	2	0	0	2
MWG943	78.2	2	2	2	2	2	2	0	0	0	0
ABG464	91.2	2	-1	2	2	2	2	0	0	0	0
Dor3	111.2	2	0	2	0	2	2	2	0	0	0
iPgd2	114.7	2	0	2	0	2	2	2	0	0	0
cMWG733A	121.7	2	0	2	0	2	2	2	2	2	2
AtpbA	125.2	0	0	2	0	2	2	2	2	2	2
drun8	138.8	0	0	2	0	2	2	0	2	2	2
ABC261	143.7	0	0	2	0	2	2	2	2	2	2
ABG710B	150.7	0	0	2	0	0	0	2	2	2	2
Aga7	154.2	0	0	2	0	0	0	2	2	2	2
MWG912	159.9	0	0	2	-1	0	0	2	2	2	2
粒重		41.0	40.4	40.1	40.3	41.5	45.8	40.2	44.1	42.1	45.7

# 145个DH家系平均粒重的次数分布



# QTL作图的基本原理

- 不论是连锁作图还是关联分析，QTL作图利用的都是标记与控制性状基因之间的连锁不平衡。由于这种不平衡的存在，性状在不同标记基因型间才会存在差异。利用单个遗传标记，开展数量性状与标记间的连锁分析，称为单标记分析。
- Sax (1923) 最早利用单标记分析，报导了菜豆 (*Phaseolus vulgaris*) 籽粒颜色和粒重之间的连锁关系，这里的籽粒颜色为标记。单标记分析适用于标记比较少、难以构建连锁图谱的情况，目前已很少采用，但这一方法中包含的基本原理对所有作图方法都是适用的。

# 单标记分析的基本原理

- 假定标记 $M/m$ 与控制某一性状的基因 $Q/q$ 存在连锁，两个亲本的基因型是 $MMQQ$ 和 $mmqq$ ，DH群体在每个分离座位上只包含两种纯合基因型。
- 由于连锁的存在， $MM$ 标记型中 $QQ$ 的频率就会高于 $qq$ ， $mm$ 标记型中 $QQ$ 的频率就会低于 $qq$ 。
- 如果基因型 $QQ$ 比 $qq$ 有更高的表型，将群体按照标记型分成两组后，这两组标记型在表型性状上就呈现截然不同的两个分布。标记 $MM$ 的分布具有较高的均值，标记 $mm$ 的分布具有较低的均值。



# 单标记分析的基本原理

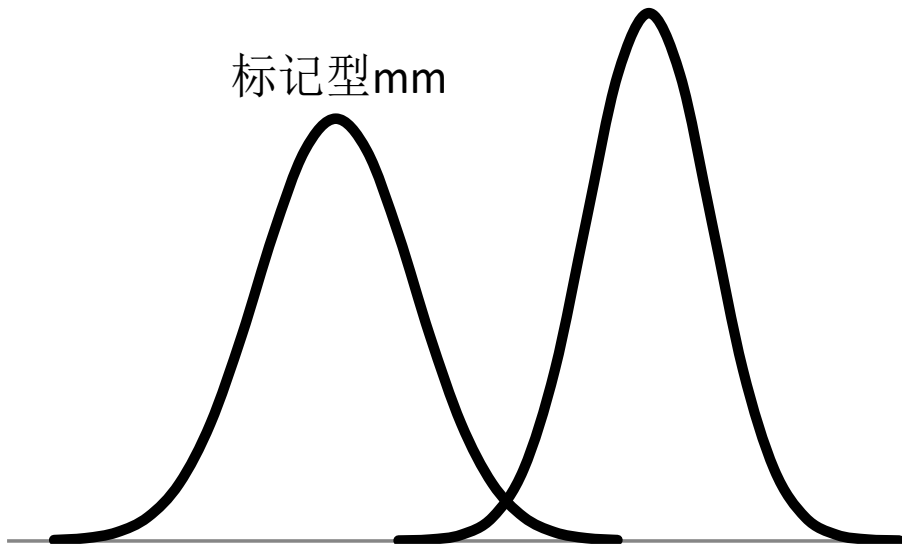
- 如果标记与控制性状的基因之间没有连锁关系，同样把群体按照标记型分成两组后， $QQ$ 和 $qq$ 的频率在这两组之间没有差异，不同标记型的分布具有相同的均值，表现出两个相似分布。
- 因此，通过检验不同标记型的性状是否服从相同的分布，就能判断该标记和QTL间是否存在连锁关系。

# 一个标记座位上2种标记型 $MM$ 和 $mm$ 的性状分布

A

标记型 $mm$

标记型 $MM$

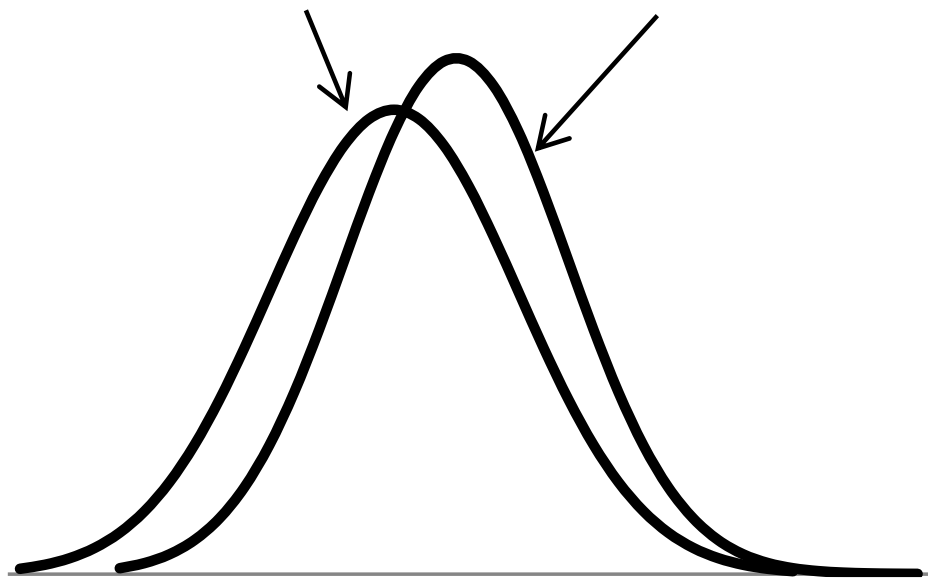


标记与性状基因存在连锁

B

标记型 $mm$

标记型 $MM$



标记与性状基因不存在连锁

# 标记和QTL两个座位上的基因型和频率

- 假定两个纯合亲本 $P_1$ 和 $P_2$ 的基因型分别为 $MMQQ$ 和 $mmqq$ 。DH群体中，标记基因型有 $MM$ 和 $mm$ 两种，QTL的基因型有 $QQ$ 和 $qq$ 两种。
- 标记与QTL结合起来共有4种基因型，即 $MMQQ$ 、 $MMqq$ 、 $mmQQ$ 和 $mmqq$ 。假定标记与QTL间的重组率为 $r$ ，4种基因型的频率分别为 $(1-r)/2$ 、 $r/2$ 、 $r/2$ 和 $(1-r)/2$ 。

# QTL的基因型值

- 标记一般是中性DNA水平的多态性，标记本身并不会产生任何表型效应。基因型 $MMQQ$ 和 $mmQQ$ 的遗传效应是由 $QQ$ 决定的；基因型 $MMqq$ 和 $mmqq$ 的遗传效应是由 $qq$ 决定的。因此，在加显性模型下，4种基因型具有两种不同的均值：

$$\mu_{MMQQ} = \mu_{mmQQ} = \mu + a$$

$$\mu_{MMqq} = \mu_{mmqq} = \mu - a$$

# 标记型的均值

- 如果一个DH家系的标记型为 $MM$ ，不能明确确定它的QTL基因型究竟是 $QQ$ 还是 $qq$ 。但是，如果把所有标记型为 $MM$ 的DH家系看作一个群体，这个群体中 $QQ$ 和 $qq$ 的频率是知道的，即 $1-r$ 和 $r$ 。标记型 $mm$ 中 $QQ$ 的频率是 $r$ ， $qq$ 的频率是 $1-r$ 。因此，

$$\mu_{MM} = (1-r)(\mu + a) + r(\mu - a) = \mu + (1-2r)a$$

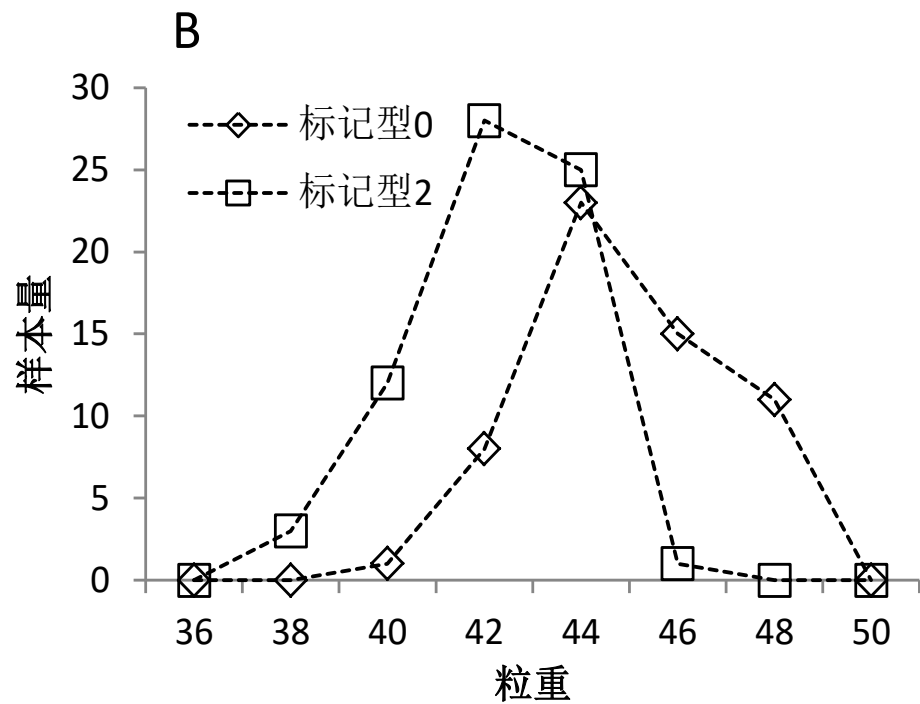
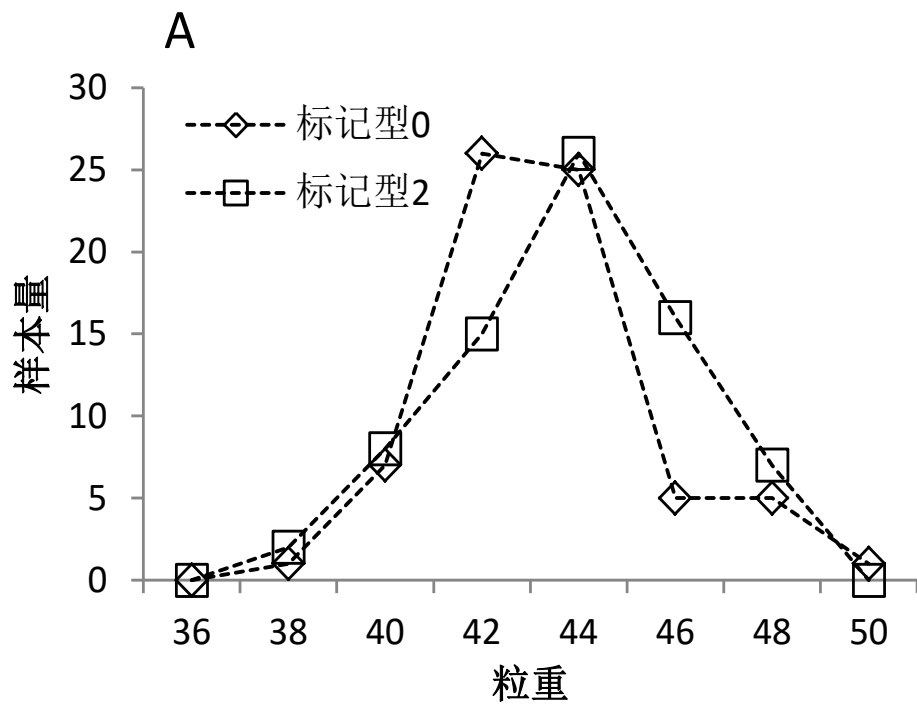
$$\mu_{mm} = r(\mu + a) + (1-r)(\mu - a) = \mu - (1-2r)a$$

$$\mu_{MM} - \mu_{mm} = 2(1-2r)a$$

# 标记型均值的差异显著性检验

- 如果标记与QTL之间不存在连锁关系，即重组率 $r=0.5$ ，标记型 $MM$ 和 $mm$ 平均数间的差异便不复存在。反过来说，如果两种标记型间存在明显差异，则说明该标记与QTL之间存在连锁。
- 利用这一原理，就可以对作图群体按照标记型进行分组，计算两种标记型的平均数与以及它们的方差，然后利用 $t$ 检验，检验两个平均数之间的差异显著性。差异显著则说明标记与QTL之间存在连锁关系；否则就说明标记与QTL之间不存在连锁关系。

# 标记Act8A (A) 和Act8B (B) 分组的 的粒重次数分布



# 两个标记座位上两种标记型粒重的 差异显著性检验

参数	标记Act8A		标记Act8B	
	标记型0	标记型2	标记型0	标记型2
样本量	70	74	58	69
自由度	69	73	57	68
均值	42.23	42.79	43.89	41.25
方差	4.45	5.32	3.53	2.79
标准差	2.11	2.31	1.88	1.67
合并方差	4.90		3.13	
t统计量	1.51 ( $P=0.1341$ )		8.37 ( $P=1.00 \times 10^{-13}$ )	



# § 13.2 简单区间作图方法

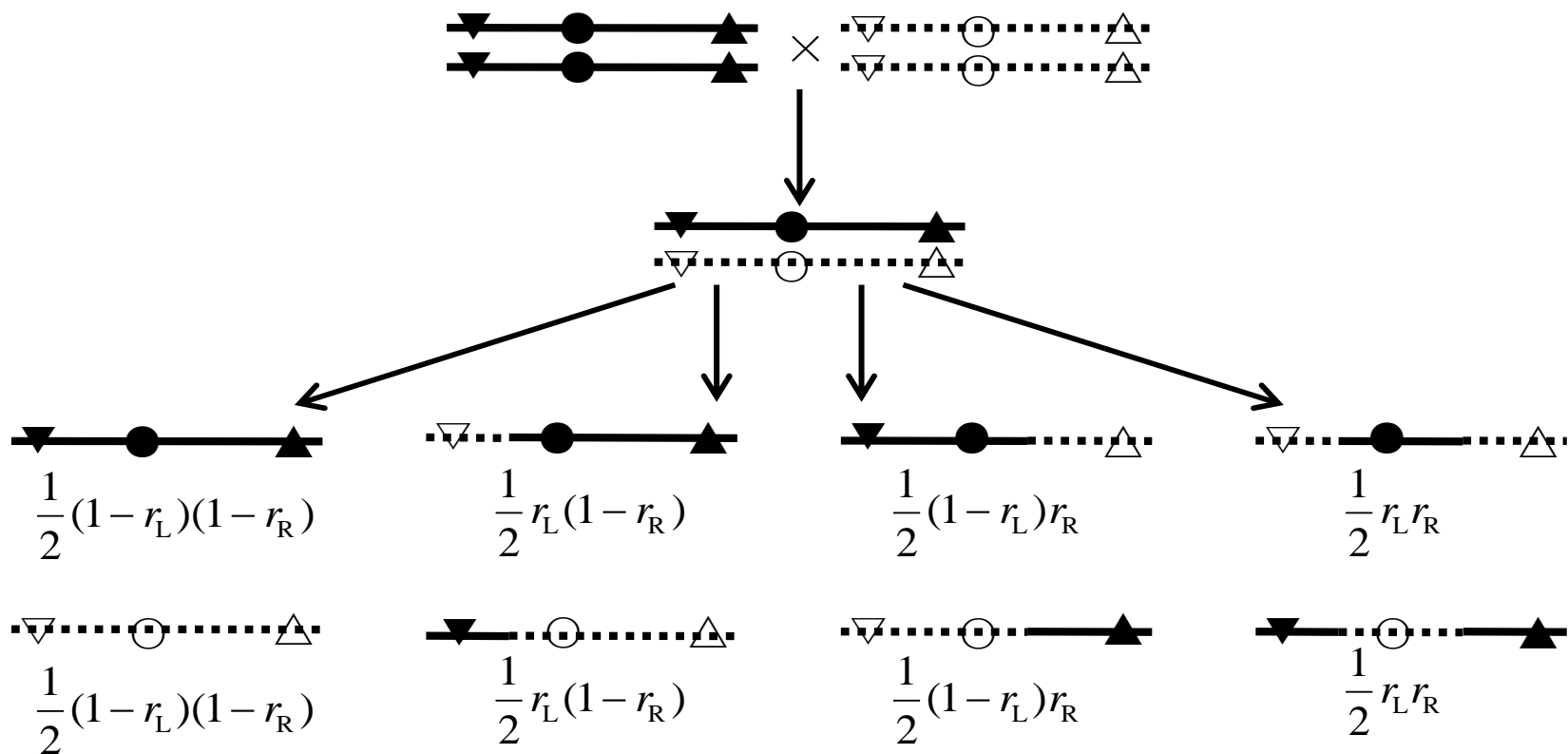
- § 13.2.1 标记区间中QTL基因型的频率
- § 13.2.2 QTL基因型均值和方差的极大似然估计
- § 13.2.3 QTL存在的假设检验与遗传参数估计
- § 13.2.4 大麦DH群体中粒重的简单区间作图
- § 13.2.5 简单区间作图方法的局限性

# 三个连锁座位重组率的关系

- 假定两个纯合亲本在两个标记座位上存在多态性，一个座位上的两种标记型分别用 $A$ 和 $a$ 表示，另一个座位上的两种标记型用 $B$ 和 $b$ 表示，两个标记之间存在一个QTL，等位基因分别用 $Q$ 和 $q$ 表示。
- 左侧标记 $A/a$ 与QTL间的重组率用 $r_L$ 表示，QTL与右侧标记 $B/b$ 的重组率用 $r_R$ 表示，侧连标记 $A/a$ 与 $B/b$ 之间的重组率用 $r$ 表示。当交换独立时，即不存在干涉现象，三个重组率之间的关系为：

$$r = r_L + r_R - 2r_L r_R$$

# 三个连锁座位上杂种F<sub>1</sub>产生配子示意图



无交换

$$(1-r_L)(1-r_R)$$

一次交换发生在左侧标记与QTL间

$$r_L(1-r_R)$$

一次交换发生在QTL与右侧标记间

$$(1-r_L)r_R$$

两次交换, 一次在左侧标记与QTL间, 一次在QTL与右侧标记间

$$r_Lr_R$$

# DH群体中两个相邻标记与它们之间QTL 共三个连锁座位上的8种基因型频率

标记型		频率	QTL 基因型的频率	
左侧标记	右侧标记		$QQ$	$qq$
$AA$	$BB$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r_L-r_R+r_Lr_R)$	$\frac{1}{2}r_Lr_R$
$AA$	$bb$	$\frac{1}{2}r$	$\frac{1}{2}(1-r_L)r_R$	$\frac{1}{2}r_L(1-r_R)$
$aa$	$BB$	$\frac{1}{2}r$	$\frac{1}{2}r_L(1-r_R)$	$\frac{1}{2}(1-r_L)r_R$
$aa$	$bb$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r_Lr_R$	$\frac{1}{2}(1-r_L-r_R+r_Lr_R)$

# 区间作图中的一维扫描

- 区间作图通过染色体上逐点扫描来检测QTL。当扫描到一个染色体的特定位置时，根据连锁图谱，就知道这个位置的左右两侧标记。于是就可以利用这两个侧连标记，对群体进行分组。
- DH群体中，4组标记型的样本量分别用 $n_1 \sim n_4$ 表示，总样本量为 $n$ 。下面的表中，每行的两种QTL基因型频率之和等于标记型的频率。QTL基因型频率除以相应的标记型频率，就得到每种标记型下QTL基因型的条件频率或条件概率，用符号 $\pi$ 表示，两个数字的下标用于区分4种标记类型和两种QTL基因型。

# DH群体中两个相邻标记上四种标记型的QTL基因型条件频率

标记型		样本量	QTL 基因型的条件频率	
左侧标记	右侧标记		QQ	qq
<i>AA</i>	<i>BB</i>	$n_1$	$\pi_{11} = \frac{1 - r_L - r_R + r_L r_R}{1 - r}$	$\pi_{12} = \frac{r_L r_R}{1 - r}$
<i>AA</i>	<i>bb</i>	$n_2$	$\pi_{21} = \frac{(1 - r_L) r_R}{r}$	$\pi_{22} = \frac{r_L (1 - r_R)}{r}$
<i>aa</i>	<i>BB</i>	$n_3$	$\pi_{31} = \frac{r_L (1 - r_R)}{r}$	$\pi_{32} = \frac{(1 - r_L) r_R}{r}$
<i>aa</i>	<i>bb</i>	$n_4$	$\pi_{41} = \frac{r_L r_R}{1 - r}$	$\pi_{42} = \frac{1 - r_L - r_R + r_L r_R}{1 - r}$

# DH群体中QTL基因型的表型分布

- 如也能按照QTL基因型对DH家系进行分类，两种基因型 $QQ$ 和 $qq$ 服从相同的分布，则说明这个座位上的不同基因型不会产生表型差异，也就不是控制性状的基因。
- 如果 $QQ$ 和 $qq$ 服从不同的分布，用下面的公式表示它们服从的分布，说明这个座位上的不同基因型产生了有差异的表型，是一个控制性状的座位。

$$QQ \sim N(\mu_1, \sigma^2) \quad qq \sim N(\mu_2, \sigma^2)$$

# DH群体中表型观察值的分布

- 用 $k=1\sim 4$ 表示表13.4的4种标记型， $j=1, 2, \dots, n_k$ 表示标记型 $k$ 中的家系。标记型 $k$ 中，DH家系的性状观测值用 $Y_{kj}$ 表示，可以看作是两个QTL基因型按照比例 $\pi_{k1}$ 和 $\pi_{k2}$ 组成的混合分布，即：

$$Y_{kj} \sim \pi_{k1}N(\mu_1, \sigma^2) + \pi_{k2}N(\mu_2, \sigma^2)$$

$$k=1\sim 4, j=1, 2 \dots, n_k$$



# 表型观察值的样本似然函数

- 用 $f(Y|\mu, \sigma^2)$ 表示任意正态分布 $N(\mu, \sigma^2)$ 的概率密度函数（具体表达式见公式6.32），所有DH家系表型数据 $Y_{kj}$ 的联合概率密度函数或似然函数为：

$$L(Y | \mu_1, \mu_2, \sigma^2) = \prod_{k,j} [\pi_{k1} f(Y_{kj} | \mu_1, \sigma^2) + \pi_{k2} f(Y_{kj} | \mu_2, \sigma^2)]$$

# 参数极大似然估计的EM算法

- 类似于公式13.8的似然函数，都难以直接求解，常常要用到EM迭代算法。利用EM算法时，首先要对似然函数13.8中的待估参数指定一组初始值。这样就相当于知道了两个QTL基因型的分布，然后计算每个DH家系基因型属于 $QQ$ 和 $qq$ 的概率，这种概率又称为后验概率（posterior probability）。

$$w_{kj1} = \frac{\pi_{k1} f(Y_{kj} | \mu_1, \sigma^2)}{\pi_{k1} f(Y_{kj} | \mu_1, \sigma^2) + \pi_{k2} f(Y_{kj} | \mu_2, \sigma^2)}$$

$$w_{kj2} = \frac{\pi_{k2} f(Y_{kj} | \mu_2, \sigma^2)}{\pi_{k1} f(Y_{kj} | \mu_1, \sigma^2) + \pi_{k2} f(Y_{kj} | \mu_2, \sigma^2)}$$

# 参数极大似然估计的EM算法

- 从群体的角度来讲，每种标记型下，QTL基因型的频率完全取决于重组率（表13.4）。但从公式13.9不难看出，每个DH家系的后验QTL基因型概率，除与重组率有关外，还与家系自身的性状观测值有关。
- 对于每个DH家系，都可以按照公式13.9计算它的基因型是 $QQ$ 或是 $qq$ 的后验概率。接下来，把每个DH家系按照公式13.9计算出的后验概率形象地分成两份，分别对应于两种基因型 $QQ$ 和 $qq$ ，这两份样本服从的分布分别对应于 $QQ$ 和 $qq$ 的分布。

# 参数极大似然估计的EM算法

- 在此基础上，下面的公式给出一个新的似然函数及其对数函数。

$$L(\mu_1, \mu_2, \sigma^2 | Y) = \prod_{k,j} [f(Y_{kj} | \mu_1, \sigma^2)]^{w_{kj1}} [f(Y_{kj} | \mu_2, \sigma^2)]^{w_{kj2}}$$

$$\ln L = \sum_{k,j} [w_{kj1} \ln f(Y_{kj} | \mu_1, \sigma^2) + w_{kj2} \ln f(Y_{kj} | \mu_2, \sigma^2)]$$

# 参数极大似然估计的EM算法

- 对对数似然函数求导数并令导数等于0，就得到新的均值估计公式和方差估计公式。

$$\mu_1 = \frac{\sum_{k,j} w_{kj1} Y_{kj}}{\sum_{k,j} w_{kj1}} \quad \mu_2 = \frac{\sum_{k,j} w_{kj2} Y_{kj}}{\sum_{k,j} w_{kj2}}$$

$$\sigma^2 = \frac{1}{n} \sum_{k,j} [w_{kj1} (Y_{kj} - \mu_1)^2 + w_{kj2} (Y_{kj} - \mu_2)^2]$$

# EM算法中的两个步骤

- 将公式13.12和公式13.13得到的估计值作为下一轮的起始值，并不断重复这一过程，直到相邻两次迭代间的估计值达到一定的精度为止；也可以用两次迭代间，似然函数差异小于一定的精度作为停止迭代的标准。
- 公式13.9代表的过程，相当于待估参数已知的情况下，计算每个DH家系分属于未知基因型 $QQ$ 和 $qq$ 的期望概率，称为EM算法的期望步骤（expectation step）。
- 公式13.10~公式13.13代表的过程，相当于知道每个DH家系QTL基因型的情况下，通过求解似然函数公式13.10或公式13.11的最大化，计算两种分布均值和方差的极大似然估计，称为EM算法的最大化步骤（maximization step）。
- 迭代结束时的估计值，就是待估参数的极大似然估计。

# QTL存在的假设检验与遗传参数估计

- 如果基因型 $QQ$ 和 $qq$ 的均值 $\mu_1$ 和 $\mu_2$ 之间存在显著差异, 则说明不同QTL基因型产生了具有显著差异的表型效应, 这时就说这个位置是一个QTL。
- 反之, 则说明不同QTL基因型不会产生具有显著差异的表型效应, 这时就说这个位置不是一个QTL。
- 检验QTL存在与否的零假设和备择假设为:

$$H_0 : \mu_1 = \mu_2 \quad H_A : \mu_1 \neq \mu_2$$

# 零假设下参数的极大似然估计

- 假设 $H_0$ 下，所有DH的表型服从同一个正态分布。样本似然函数为：

$$Y_{kj} \sim N(\mu_0, \sigma_0^2) \quad k=1 \sim 4, j=1, 2 \dots, n_k$$

$$L(Y \mid \mu_0, \sigma_0^2) = \prod_{k,j} f(Y_{kj} \mid \mu_0, \sigma_0^2)$$



# 零假设下参数的极大似然估计

- 对数似然函数为：

$$\ln L(Y | \mu_0, \sigma_0^2) = \sum_{k,j} \ln f(Y_{kj} | \mu_0, \sigma_0^2)$$

- 对对数似然函数公式13.17求导数，并令导数等于0，就得到待估参数的极大似然估计。

$$\hat{\mu}_0 = \frac{1}{n} \sum_{j,k} Y_{jk} \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum_{j,k} (Y_{jk} - \hat{\mu}_0)^2$$

# QTL存在的似然比检验

- 把公式13.18的极大似然估计值代入公式13.16，就得到 $H_0$ 的似然函数极大值，用  $\max L(H_0)$  表示；备择假设 $H_A$ 的似然函数极大值用  $\max L(H_A)$  表示。
- 备择假设对待估参数没有施加任何限制条件，零假设需要对待估参数施加一定的约束条件。因此，备择假设的极大似然函数，不会小于零假设的极大似然函数。
- 可以想象，如果备择假设为真，这时带有约束条件零假设的似然函数就会远低于真的备择假设；如果零假设为真，这时即使没有约束条件，似然函数也不会比零假设高得太多。

# QTL存在的似然比检验

- 因此，两种假设下极大似然函数的比值，提供了一种对零假设的检验方法，这就是统计上用途很广的似然比检验（likelihood ratio test, 简称LRT）。检验统计量及其大样本分布由下面的公式给出。

$$LRT = -2 \ln \frac{\max L(H_0)}{\max L(H_A)} \sim \chi^2(df)$$

# 似然比检验的自由度

- 似然比检验有广泛的适用范围。当样本量足够大时，LRT统计量服从 $\chi^2$ 分布，自由度等于两种假设的独立待估参数个数之间的差异。
- 检验公式13.14的零假设包含两个待估参数，备择假设包含三个待估参数。因此，区间作图LRT的 $\chi^2$ 分布自由度 $df=1$ 。
- 这样，利用公式13.19的统计量，就可以对 $QQ$ 和 $qq$ 的均值差异进行显著性检验。

# LOD检验统计量

- 公式13.19的统计量是一个自然对数。实际中，人们更习惯于以10为底的常用对数，即下面公式给出的LOD统计量。

$$\text{LOD} = \log_{10} \left( \frac{\max L(H_A)}{\max L(H_0)} \right)$$

# LOD和LRT检验统计量的换算

- 如果 $H_A$ 和 $H_0$ 的极大似然函数比值为10，则LOD值等于1；如果比值为100，则LOD值等于2等等。
- LRT和LOD之间相差一个常数比值，下面的公式给出两者的换算关系。

$$\text{LOD} = \frac{\text{LRT}}{2 \ln 10} \approx 0.217 \text{ LRT} \quad \text{LRT} \approx 4.605 \text{ LOD}$$

- LOD统计量并不满足 $\chi^2$ 分布。如想知道它的显著性概率，仍需要用上面的公式把LOD转换为LRT，然后根据分布公式13.19计算显著性概率值。

# QTL遗传效应的估计

- 如前所述，任意一个扫描位置上，都能通过EM迭代算法，得到QTL基因型均值 $\mu_1$ 和 $\mu_2$ 的极大似然估计。它们与性状平均数 $\mu$ 和QTL加性效应 $a$ 之间的关系，以及由此得到性状平均数 $\mu$ 和QTL加性效应 $a$ 的估计公式为：

$$\hat{\mu}_1 = \mu + a \quad \hat{\mu}_2 = \mu - a$$

$$\hat{\mu} = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) \quad \hat{a} = \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)$$

# QTL贡献率的计算

- 加性效应的单位与表型相同。为了便于比较不同性状的QTL，需要定义QTL的表型贡献率，有时也称QTL解释表型变异的大小（phenotypic variance explained, PVE）。
- 一个QTL的PVE定义为这个座位上的遗传方差占表型方差的百分数， $V_G$ 是QTL的遗传方差， $V_P$ 是数量性状的表型方差，等于零假设下的方差估计值，PVE的计算公式是：

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



# QTL遗传方差的计算

- 如不考虑奇异分离，QTL在DH群体中的遗传方差等于 $a^2$ ；下面的公式给出两种基因型频率不等时遗传方差的计算方法，基因型 $QQ$ 和 $qq$ 频率的计算要用到公式13.9的后验概率。

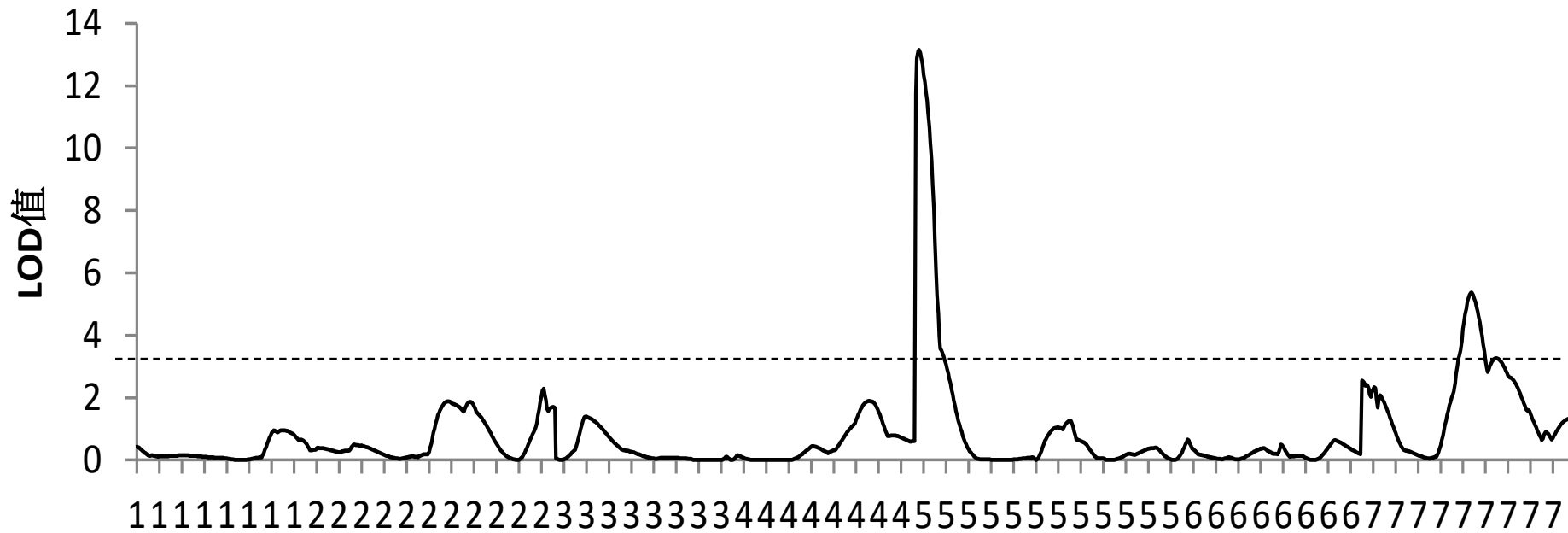
$$V_G = 4 f_{QQ} f_{qq} a^2$$

$$\text{其中 } f_{QQ} = \frac{1}{n} \sum_{j,k} w_{jk1} \quad f_{qq} = \frac{1}{n} \sum_{j,k} w_{jk2}$$

# QTL遗传效应和贡献率的不一致性

- 从QTL的遗传方差计算公式13.25可以看到，QTL的贡献率除依赖于它的遗传效应外，还依赖于基因型的频率。
- 在不存在奇异分离的群体中，效应较大的QTL，也会具有较高的表型贡献率。
- 如果作图群体中存在较严重的奇异分离，效应较大的QTL，有时也可能具有较低的表型贡献率。

# 大麦全基因组七条染色体粒重一维扫描的LOD曲线



大麦7条染色体的一维扫描, 步长1cM

# 简单区间作图鉴定出的粒重QTL

染色体	位置/cM	最近左侧标记	最近右侧标记	LOD值	贡献率/%	加性效应
5	3	ABA306B	Act88	13.15	34.55	-1.31
7	0	dRpg1	iPgd1A	2.55	7.79	-0.62
7	98	VAtp57A	MWG571D	5.36	15.77	-0.89

# 有利等位基因来源的判断方法

- 加性效应既衡量QTL对性状的贡献，又能从效应的正负号判断有利等位基因的亲本来源。
- 在双亲遗传群体中，每个鉴定出的QTL都携带有两个等位基因。有些性状在育种中要求越高越好，有些性状要求越低越好，有些性状要求最好在一定的范围内，太低或太高都不好。
- 因此，等位基因的有利或不利是根据所关心的性状和特定的育种目标而言。

# 有利等位基因来源的判断方法

- 该DH群体中，亲本 ‘Harrington’ 和 ‘TR306’ 的平均粒重分别为38.7mg和45.0mg。QTL作图时，用2表示 ‘Harrington’ 的标记型，0表示 ‘TR306’ 的标记型。育种对粒重的要求一般是越高越好。
- 因此，如果加性效应为正说明 ‘Harrington’ 携带的等位基因能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。

# 有利等位基因来源的判断方法

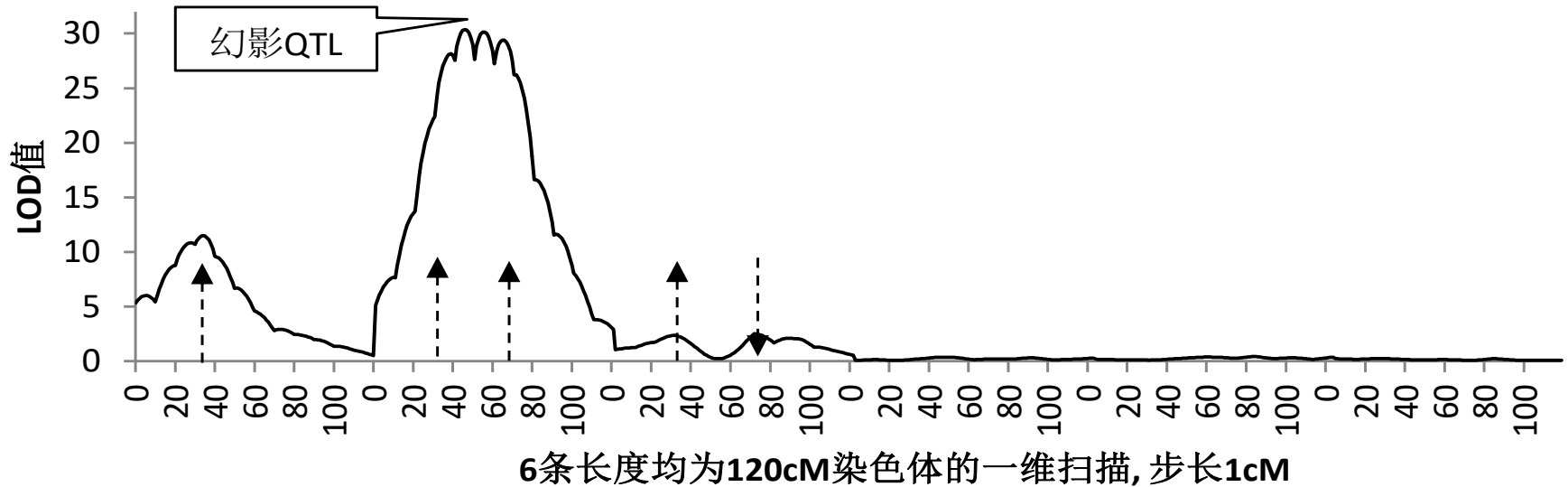
- 反之，如果某个QTL加性效应为负值，则说明‘Harrington’携带的等位基因起到降低粒重的作用，‘TR306’携带的等位基因则起到增加粒重的作用。
- 从表13.5的结果来看，三个被检测到的控制粒重的QTL均具有负的加性效应，这三个QTL上提高粒重的等位基因，均来源于编码为0的亲本，即粒重较高的亲本‘TR306’。

# 简单区间作图方法的局限性

- Lander和Botstein（1989）提出区间作图方法时，曾假定每条染色体上的QTL不超过一个。
- 如果存在两个连锁QTL，并且QTL的遗传效应有相同的方向（又称为相引连锁），区间作图会在两个QTL的中间出现一个峰，这种现象称为幻影QTL。
- 如果存在两个遗传效应方向相反的连锁QTL（又称为互斥连锁），区间作图在整条染色体上的LOD会很低，可能检测不出任何QTL的存在。



# 一个模拟群体中，简单区间作图的LOD曲线



- 共考虑6条染色体，第一条存在一个QTL，加性效应为1；第二条存在两个QTL，加性效应均为1；第三条存在两个QTL，加性效应为1和-1；其它三条无QTL。图中箭头对应的X轴，为QTL的真实位置，方向表示加性效应的正负

# 幻影QTL产生的原因

- 究其原因，简单区间作图在某一标记区的一个扫描位置上进行参数估计和假设检验时，没有对区间之外的QTL进行任何控制。
- 当存在QTL连锁时，简单区间作图不能有效定位QTL，更谈不上准确估计遗传效应了。即使QTL位于不同的染色体上，假设检验中的分布也会有较大的方差，对效应相对较小的QTL的检验功效较低。
- 同时，QTL的影响范围很大，在没有QTL的区间上也会出现检验统计量超过临界值的情况；每个位置上的检验统计量只利用两个标记的信息，利用的信息量较少。针对简单区间作图的这些问题，人们随后提出了具有背景控制的QTL作图方法。

# § 13.3 具有背景控制的QTL作图方法

- § 13.3.1 单个QTL的标记回归模型
- § 13.3.2 多个加性QTL的标记回归模型
- § 13.3.3 完备区间作图的背景控制
- § 13.3.4 大麦DH群体中粒重的完备区间作图

# 简单区间作图的主要问题

- 简单区间作图没有对背景遗传变异施加任何控制，背景遗传变异与随机误差方差一起组成参数估计和假设检验中的样本方差。
- 统计学上，参数估计的精确度和假设检验的功效，一方面受分布参数真实大小的影响，另一方面也受随机样本所服从分布方差的影响。样本方差越大，参数估计的精确度和假设检验的功效越低。
- 这一节仍以最简单的DH群体为例，介绍完备区间作图（Inclusive composite interval mapping，简称ICIM），以说明如何通过背景控制，使得样本方差仅来自随机误差，进而提高参数估计的精确度和QTL检测的功效。

# 单个QTL的基因型值

- 在只有一个QTL的加性遗传模型下， $Q$ 和 $q$ 表示该座位上2个等位基因，两种纯合QTL基因型 $QQ$ 和 $qq$ 的基因型值 $G$ 统一用下面的公式表示：

$$G = \mu + aw$$

- 其中， $\mu$ 代表两种纯合基因型 $QQ$ 和 $qq$ 的中亲值， $a$ 为加性效应， $w$ 是QTL基因型的指示变量， $w=1$ 代表基因型 $QQ$ 、 $w=-1$ 代表基因型 $qq$ 。

# 区间标记中的QTL基因型频率

- QTL作图前，个体的QTL基因型是未知的，遗传参数 $a$ 有待估计，难以对公式13.26进行估计。但是，个体的标记型是已知的。由于标记和QTL之间存在连锁，标记型提供了QTL基因型的信息。因此，需要寻求QTL基因型与标记基因型间的关系。
- 假定两个多态性标记之间存在一个QTL，两个亲本的基因型分别为 $AAQQBB$ 和 $aaqqbb$ 。DH群体中有4种不同的标记型，每种标记型下，两种QTL基因型的频率可通过QTL与两个标记间的重组率来估计。类似遗传模型公式13.26中QTL基因型的指示变量，如下表定义两个侧连标记的指示变量，每种标记下的QTL基因型频率及重组率的含义均与表13.4相同。

# DH群体中两个相邻标记座位上4种 标记型内QTL基因型的期望频率

标记型	标记变量		QTL 基因型的频率		$w$ 的条件期望 $E(w x_L, x_R)$
	$x_L$	$x_R$	QQ ( $w=1$ )	qq ( $w=-1$ )	
<i>AABB</i>	1	1	$\frac{1 - r_L - r_R + r_L r_R}{1 - r}$	$\frac{r_L r_R}{1 - r}$	$1 - \frac{2r_L r_R}{1 - r} \hat{=} f_1$
<i>AAbb</i>	1	-1	$\frac{(1 - r_L)r_R}{r}$	$\frac{r_L(1 - r_R)}{r}$	$\frac{r_R - r_L}{r} \hat{=} f_2$
<i>aaBB</i>	-1	1	$\frac{r_L(1 - r_R)}{r}$	$\frac{(1 - r_L)r_R}{r}$	$\frac{r_L - r_R}{r} = -f_2$
<i>aabb</i>	-1	-1	$\frac{r_L r_R}{1 - r}$	$\frac{1 - r_L - r_R + r_L r_R}{1 - r}$	$-1 + \frac{2r_L r_R}{1 - r} = -f_1$

# QTL基因型期望频率的计算

- 根据上表中指示变量 $w$ 的取值和各种标记型下的条件概率，计算各种标记型下QTL指示变量 $w$ 的条件期望，列于上表最后一列。
- 以标记型 $AABB$ 为例，并利用三个重组率之间的关系，QTL指示变量 $w$ 期望的计算方法如下。

$$E(w | x_L = 1, x_R = 1) = \sum w \times \Pr\{w | x_L = 1, x_R = 1\} = 1 - \frac{2r_L r_R}{1-r} \hat{=} f_1$$



# QTL基因型期望频率的一般表示

- 如用 $f_1$ 和 $f_2$ 分别表示 $AABB$ 和 $AAbb$ 中 $w$ 的期望，那么， $aaBB$ 和 $aabb$ 中， $w$ 期望分别等于 $-f_2$ 和 $-f_1$ 。根据表13.6最后一列的4个期望，定义两个新的参数，稍后给出这两个参数的具体含义。不难验证，QTL基因型指示变量 $w$ 在4种标记型下的期望值，可以通过 $\lambda_L$ 和 $\lambda_R$ 表示为标记指示变量 $x_L$ 和 $x_R$ 的线性函数。

$$\lambda_L = \frac{1}{2}(f_1 + f_2) = \frac{r - r_L + r_R - 2r_L r_R}{2r(1-r)} \quad \lambda_R = \frac{1}{2}(f_1 - f_2) = \frac{r + r_L - r_R - 2r_L r_R}{2r(1-r)}$$

$$E(w | x_L, x_R) = \lambda_L x_L + \lambda_R x_R$$

# QTL基因型值对标记的期望

- 单基因加性效应模型中，基因型值 $G$ 对标记型的条件期望就能用下面的公式表示出来。将公式中标记指示变量的系数视为标记的效应，分别用 $A_L$ 和 $A_R$ 表示，那么基因型值 $G$ 对标记型的条件期望，就可以用标记效应表示。

$$E(G | x_L, x_R) = \mu + aE(w | x_L, x_R) = \mu + a\lambda_L x_L + a\lambda_R x_R$$

$$E(G | x_L, x_R) = \mu + A_L x_L + A_R x_R$$

$$\text{其中, } A_L = a\lambda_L \quad A_R = a\lambda_R$$

# 标记效应中的QTL位置和效应信息

- 可以看到，公式13.27乘以QTL加性效应后，就是左侧标记的加性效应；公式13.28乘以QTL加性效应后，就是右侧标记的加性效应。
- 从上述的推导过程可以看出，线性模型公式13.31的系数 $A_L$ 和 $A_R$ ，既包含QTL的位置信息，又包含QTL加性效应的信息。
- 反过来，如果能够估计公式13.31中标记的系数，就能推断QTL在标记区间上的相对位置，并估计其加性遗传效应。

# 多个加性QTL的标记回归模型

- 为简单起见，假定 $m$ 个QTL在两个纯合亲本 $P_1$ 和 $P_2$ 中分离，分布在由 $m+1$ 个标记分隔的 $m$ 个区间内。
- 假定 $P_1$ 的基因型为 $Q_1Q_1Q_2Q_2\cdots Q_mQ_m$ ， $P_2$ 的基因型为 $q_1q_1q_2q_2\cdots q_mq_m$ 。
- 对DH群体中的一个家系， $\mathbf{X}=(x_1, x_2, \dots, x_m, x_{m+1})$ 代表标记指示变量，亲本 $P_1$ 标记型用1表示，亲本 $P_2$ 标记型用-1表示； $\mathbf{W}=(w_1, w_2, \dots, w_m)$ 代表QTL指示变量，亲本 $P_1$ 基因型用1表示，亲本 $P_2$ 基因型用-1表示。

# 多个加性QTL的标记回归模型

- QTL的加性效应用 $a_1, a_2, \dots, a_m$ 表示。假设QTL的效应是可加的，DH家系的基因型值 $G$ 可表示为：

$$G = \mu + \sum_{j=1}^m a_j w_j$$

- 为区分不同QTL产生的标记效应，将公式13.29另记为：

$$E(w_j | \mathbf{X}) = \lambda_{j(L)} x_j + \lambda_{j(R)} x_{j+1}$$

# 多个加性QTL的标记回归模型

- 因此，在标记基因型已知的条件下，基因型值 $G$ 的期望可表示为标记变量的线性函数。

$$E(G | \mathbf{X}) = \mu + \sum_{j=1}^m a_j (\lambda_{j(L)} x_j + \lambda_{j(R)} x_{j+1}) \hat{=} \beta_0 + \sum_{j=1}^{m+1} \beta_j x_j$$

其中，  $\beta_0 = \mu$

$$\beta_1 = \lambda_{1(L)} a_1$$

$$\beta_j = \lambda_{(j-1)(R)} a_{j-1} + \lambda_{j(L)} a_j \quad (j=2, \dots, m)$$

$$\beta_{m+1} = \lambda_{m(R)} a_m$$

# 多个加性QTL的标记回归模型

- 从标记系数的构成来看，第 $j$ 个标记的系数仅受标记区间  $(j-1, j)$  和  $(j, j+1)$  上QTL加性效应的影响。
- 如果与  $(j, j+1)$  相邻的左右两个区间上没有QTL，区间  $(j, j+1)$  两端标记的回归系数 和 仅包含该区间内QTL位置和加性效应的信息。
- 这正是ICIM加性作图的理论依据，也是其它回归作图方法的理论依据。

# 表型对标记的回归模型

- 假定一个群体中有 $n$ 个DH家系，目标性状的表型和 $m+1$ 个已排序标记的基因型都是已知的。基于遗传模型公式13.34，可以得到用于QTL加性作图的线性回归模型公式：

$$Y_i = E(G | \mathbf{X}) + \varepsilon_i = \beta_0 + \sum_{j=1}^{m+1} \beta_j x_{ij} + \varepsilon_i$$

- 其中， $i=1, 2, \dots, n$ ， $n$ 为家系数目； $Y_i$ 是DH群体中第 $i$ 个家系的表型值； $\beta_0$ 是线性回归模型的常数项； $\beta_j$ 是表型对第 $j$ 个标记变量的偏回归系数； $x_{ij}$ 是第 $j$ 个标记在第 $i$ 个家系中的标记指示变量，亲本 $P_1$ 标记型用1表示，亲本 $P_2$ 标记型用-1； $\varepsilon_i$ 是残差项，假定服从均值为0、方差为 $\sigma_\varepsilon^2$ 的正态分布。



# 完备区间作图的背景控制

- 线性模型公式13.35是ICIM方法实现背景控制的理论基础。
- 从推导过程可以看出，在QTL的效应满足可加性的假定下，表型对标记的偏回归系数只依赖于两个相邻标记所标定区间上的QTL，而不受其他区间上QTL的影响。
- ICIM的基本思想是用全基因组上所有标记信息构建回归模型公式13.35，利用矫正的表型值进行区间作图，以达到控制背景遗传变异的目的。

# 完备区间作图的背景控制

- 考虑到QTL个数通常远低于标记个数，采用逐步回归策略选择公式13.35的重要标记变量，未中选标记的偏回归系数设为0。公式13.35中的参数只估计一次。如当前扫描的标记区间为  $(k, k+1)$ ，利用下面的公式对观测值进行矫正。

$$\Delta Y_i = Y_i - \sum_{j \neq k, k+1} \hat{\beta}_j x_{ij}, \quad i=1, 2, \dots, n, \quad n \text{ 为群体大小}$$

- 如果样本量足够大，而且当前区间  $(k, k+1)$  的相邻区间上不存在QTL时，两个侧连标记回归系数的估计值仅包含区间  $(k, k+1)$  上QTL位置和加性效应的信息。

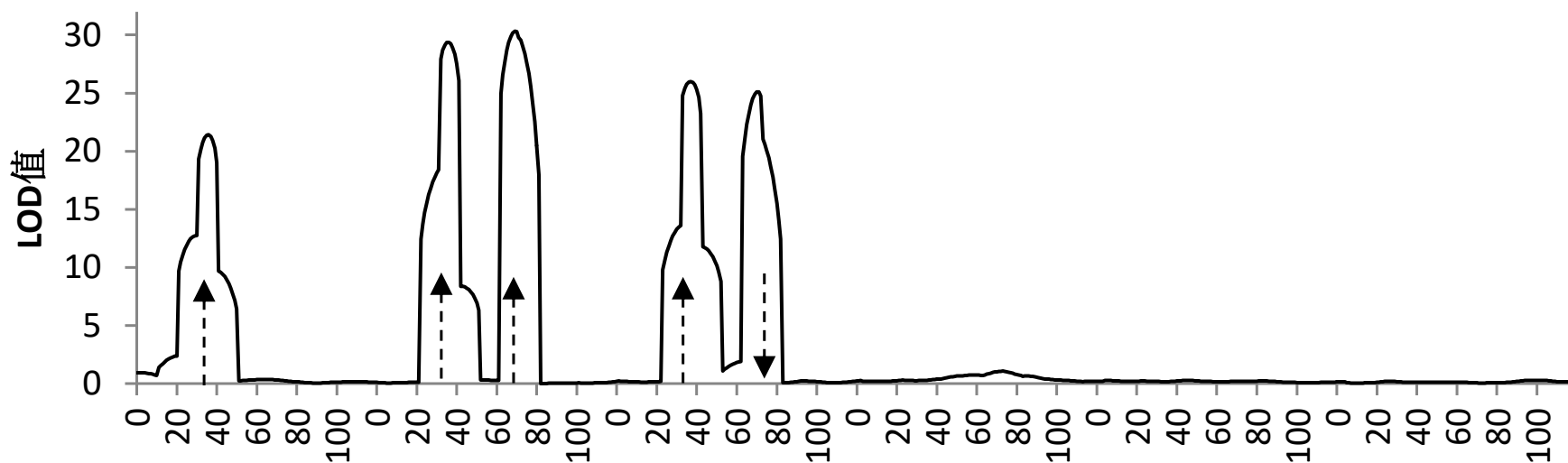
# 完备区间作图的背景控制

- 因此，在随后的区间作图中，表型矫正值就不会遗漏当前扫描区间  $(k, k+1)$  上QTL位置和效应的信息。
- 同时，矫正值中扣除了当前区间之外标记的回归系数，相等于控制了其他区间和染色体上QTL的效应。因此，随后利用矫正值进行区间作图时，也就排除了当前区间之外所有QTL的影响。
- 只有当扫描位置移动到下一个区间时，才需要重新计算表型矫正值。

# 完备区间作图的两步骤策略

- ICIM方法包含以下两个步骤。第一步，利用所有标记的信息，通过逐步回归选择重要的标记变量并估计其效应。第二步，利用逐步回归得到的线性模型矫正表型数据，利用矫正后的数据进行区间作图。
- 下图给出一个模拟群体的完备区间作图LOD曲线，所用的遗传模型和群体大小与图13.6完全相同。可以看到，不管是第二染色体上的两个相引连锁、还是第三染色体上的两个互斥连锁QTL，ICIM都在真实位置附近产生很高的峰，远离QTL的地方、没有QTL的染色体，ICIM产生的LOD值接近于0。

# 一个模拟群体中，完备区间作图的LOD曲线



6条长度均为120cM染色体的一维扫描, 步长1cM

- 共考虑6条染色体，第一条存在一个QTL，加性效应为1；第二条存在两个QTL，加性效应均为1；第三条存在两个QTL，加性效应为1和-1；其它三条无QTL。图中箭头对应的X轴，为QTL的真实位置，方向表示加性效应的正负

# 完备区间作图的优点

- 大量模拟研究和实际数据都表明，ICIM是一个行之有效的QTL定位方法。ICIM有较低的抽样误差，较高的作图效率。存在QTL的区域，ICIM能够产生出显著高的LOD值；没有QTL的区域，ICIM的LOD值接近于0。
- ICIM对作图参数有着很好的稳健性，同时也容易被推广到上位性作图，以及多亲本杂交和两个无性系杂交产生的遗传研究群体。在上位性作图时，不仅可以检测到有加性效应QTL间的互作，而且还可以检测到没有明显加性效应的QTL之间的互作。

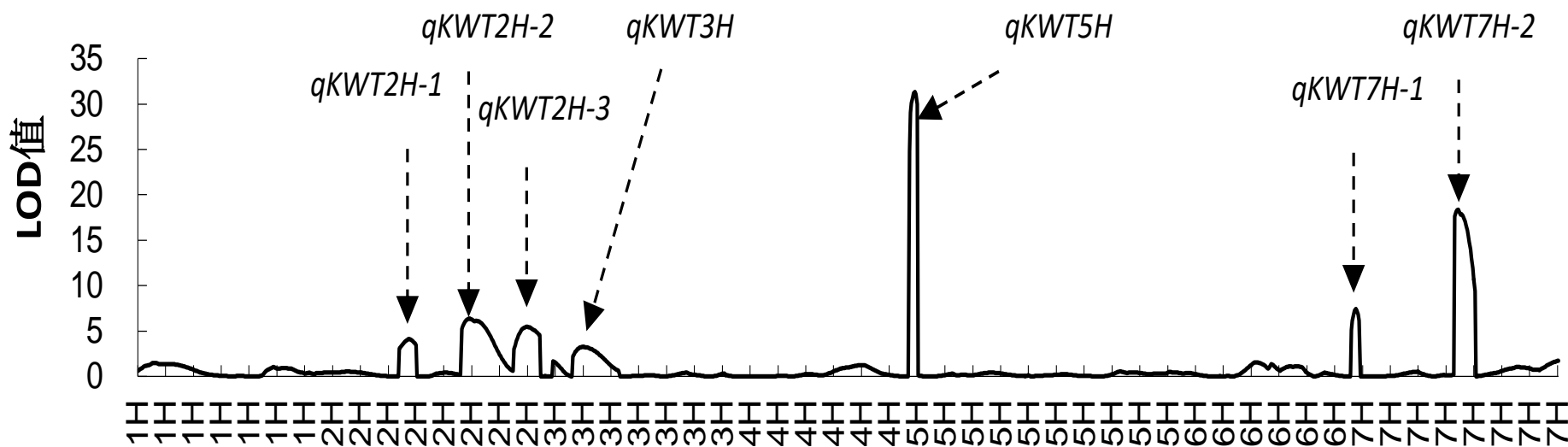
# 完备区间作图对连锁QTL的检测

- 与简单区间作图相比，ICIM放弃了单条染色体至多存在一个QTL的假定。
- 但这并不是说，不管多么近的连锁QTL都可以分辨出来。ICIM区分连锁QTL是有条件的，要求连锁QTL之间至少存在一个空白区间，即“分隔QTL” (isolated QTL)；同时也要求有足够大的作图群体。

# 大麦145个DH家系粒重性状的完备区间作图

逐步回归中变量进入模型的概率水平为0.001

变量退出模型的概率水平为0.002





# 完备区间作图检测到的粒重QTL

QTL名称	染色体上的位置 (cM)	左侧最邻标记	右侧最邻标记	LOD值	加性效应	贡献率/%
qKWT2H-1	84	Pox	BCD351B	4.10	0.43	3.63
qKWT2H-2	138	ABC620	MWG882	6.36	-0.56	6.21
qKWT2H-3	190	BCD453B	ABG317	5.45	0.53	5.60
qKWT3H	27	Ugp2	Ugp1	3.23	0.38	2.99
qKWT5H	5	Act8B	MWG502	31.33	-1.47	43.92
qKWT7H-1	4	iPgd1A	BCD129	7.41	-0.59	7.07
qKWT7H-2	96	MWG626	VAtp57A	18.35	-1.01	20.53

# 粒重QTL上增效等位基因的分布

- 位于5H染色体5cM上的 $qKWT5H$ 和位于7H染色体96cM上的 $qKWT7H$ ，是加性遗传效应最大的2个QTL，分别解释表型变异的43.92%和20.53%。
- 4个效应较大QTL的加性遗传效应均为负值，说明在这些座位上，提高粒重的等位基因来源于编码为0的亲本，即粒重较高的亲本‘TR306’。
- 亲本‘Harrington’的粒重低于亲本‘TR306’，但在QTL座位 $qKWT2H-1$ 、 $qKWT2H-3$ 和 $qKWT3H$ 上仍携带有提高粒重的等位基因。
- 增效基因在两个亲本中的分散分布，很好地解释了表型分布图13.1中观察到的超亲分离现象。

# 表型回归的过拟合和拟合不足问题

- 利用ICIM，首先要对表型对标记的回归模型进行估计。一般群体中QTL的个数都会远低于标记的个数。当标记数目较大，或者标记数远超过群体大小时，回归模型包含很多变量。尽管我们知道大多数变量的系数都是0，但要确定哪些效应是0、哪些不是0，并不是一件容易的事。
- 逐步回归的变量选择过程中有两个重要参数，一个是变量进入模型的显著性概率，用PIN表示；另一个是变量离开模型的显著性概率，用POUT表示。PIN决定哪些变量需要进入回归模型，POUT决定在新变量加入模型后哪些变量需要离开回归模型。不同的PIN和POUT会选择到不同的标记变量，因此QTL作图也会产生不同的结果。

# 表型的过拟合和拟合不足问题

- 如果回归模型未能拟合大部分或全部的遗传变异，ICIM第二步区间作图时的背景控制就会不完全，从而影响QTL的检测功效，就可能检测不出来效应较小的QTL。
- 如果回归模型出现过拟合，ICIM第二步区间作图时就可能会产生假阳性QTL。
- 正确使用ICIM，既要避免表型对标记变量的不完全拟合而产生的假阴性问题，又要避免过拟合而产生的假阳性问题。

# 过拟合和拟合不足问题的避免

- 实际数据中，通过对比回归模型的决定系数和性状的广义遗传力，可以大致判定模型的拟合程度。回归模型的决定系数定义为回归平方和占总平方和的比例，可以视为回归模型解释表型变异的比例。
- 如果决定系数远高于性状的广义遗传力，则表明可能存在过拟合问题。如果决定系数低于性状的广义遗传力，则表明可能存在拟合不足问题。
- 另外，作图结果中是否存在紧密连锁QTL，也可以用来判断是否存在过拟合。如果在检测到的QTL中存在紧密连锁的现象，甚至检测到紧密连锁的互斥QTL，都表明可能有过拟合问题。
- 当存在过拟合现象时，可通过降低标记进入回归模型的概率，逐步改善过拟合问题。

## § 13.4 集成软件QTL IciMapping简介

# QTL IciMapping的特点

- 构建高密度的遗传连锁图谱、定位重要性状的QTL是目前动植物遗传研究的重要内容。这些研究工作，需要借助专门的计算机软件才能完成。
- QTL IciMapping是中国农业科学院作物科学研究所数量遗传课题组，通过10年努力研制的一个集成软件，是国际上首例能够同时进行连锁图谱构建和数量性状基因定位的遗传分析工具。
- 软件采用最新的编程技术，把不同功能包装在一个工程中，用户可以随时查看已经完成的各项操作，浏览已完成操作产生的各种结果。软件具有友好的用户交互界面，可以满足大多数遗传研究群体的表型数据和基因型数据分析，并根据遗传研究的发展和用户的需要不断地更新完善。

# QTL IciMapping 4.1版的9大功能

- 1、多环境表型数据方差分析的AOV功能；
- 2、删除冗余标记的BIN功能；
- 3、构建连锁图谱的MAP功能；
- 4、整合连锁图谱的CMP功能；
- 5、定位奇异分离座位的SDL功能；
- 6、双亲衍生群体QTL作图的BIP功能；
- 7、多环境表型鉴定数据QTL分析的MET功能；
- 8、染色体片段置换系QTL作图的CSL功能；
- 9、巢式关联分析群体QTL作图的NAM功能。



# QTL IciMapping 4.1版的两个工具

- 1、绘制已构建遗传连锁图谱的MapShow工具；
  - 2、两点重组率估计的2pointREC工具。
- 利用QTL IciMapping软件集成的各种功能，研究工作者现在可以很方便地开展单环境/多环境表型鉴定数据的方差分析和遗传力估计、分析基因型鉴定数据的冗余性、利用分子标记构建遗传连锁图谱、鉴定奇异分离座位、通过一维扫描定位加性和显性QTL、通过二维扫描定位上位性互作QTL、以及开展QTL与环境互作分析等多方面的遗传研究。

# 输入文件的格式

- QTL IciMapping软件除接收按照一定规则整理的纯文本数据文件外，还能读入不同版本的Excel文件。
- 用户在使用Excel文件导入数据时，只需参照软件中的例子文件，或者干脆使用这些例子作模板，把自己群体的一般性信息如群体类型、群体大小、标记个数、性状个数或环境个数等存放在一个工作表中，把基因型数据、表型数据、标记锚定信息等存放于其他不同的工作表中，软件就可以根据用户选定的功能，读取不同Excel工作表中的数据。

# 结果的浏览和保存

- 同时，用户在完成一个功能的分析后，软件会输出大量的分析结果。有些结果以文件的形式存放在用户的工程下，有些结果是各种各样的图形。
- 用户可以在工程下浏览这些文件，将文件保存到其它地方或保存为其它格式。用户也可以在工程下浏览图形输出，对图形进行编辑并保存成具有不同分辨率的图片文件。

# 不同功能之间的衔接

- 为了使用更方便，软件在研制过程中还注重不同功能之间的衔接。用户在使用某些功能时，软件会根据用户可能开展的后续分析，利用一个功能得到的分析结果，自动地创建出后续功能的输入文件。
- 例如，用户在使用删除冗余标记的BIN功能后，软件会自动生成连锁图谱构建功能MAP的输入文件，这样，用户就可以很方便地利用去除冗余后的标记来构建图谱。
- 用户在使用MAP功能后，软件会自动生成奇异分离作图SDL功能的输入文件，用户可以立即利用构建出的连锁图谱，来定位奇异分离座位；同时还会自动生成各种QTL作图功能的输入文件，用户只需指定性状的个数、添加这些性状的表型数据，然后就能顺利开展各种QTL定位分析。