第二章 连锁分析和遗传图谱构建

重组率是指两个标记或基因座位之间发生奇数次交换的概率，直观上反映了两个基因座位间的遗传距离．重组率的估计是遗传研究中的经典问题 (Kempthrone, 1957; Bailey, 1961; Hartl and Jones, 2005)．连锁图谱是指基因或标记在染色体上的相对位置与遗传距离，遗传距离一般以厘摩 (centi-Morgan, cM) 表示，1%的重组率对应的遗传距离定义为1cM．世界上第一张遗传连锁图谱是利用5个形态特性标记构建的果蝇X染色体 (Sturtevant, 1913)，现在的连锁图谱一般都包含成百上千个标记．建立在重组率估计之上的连锁图谱，是开展遗传研究，基因定位，精细定位和克隆的前提．本章介绍常见群体中的遗传连锁分析和图谱构建方法．

§2.1 世代转移矩阵

§2.1.1 世代转移矩阵的定义

考虑两个座位上的等位基因A-a和B-b，亲本的基因型为AABB和aabb，后代中有九种可能的基因型．给定重组率的大小，每种基因型在特定群体中有特定的理论频率(也称作期望频率)．基因型存在于群体中的理论频率是重组率估计的基础．有些群体，如回交一代，F1DH和F2，是F_1群体通过适当的交配繁殖方式产生的．由于F_1只有一种基因型，因此，容易计算经过一次回交，加倍单倍体或一次自交之后，产生出来的遗传群体中各种基因型的频率．如果一个群体是由其他群体经过多次回交和自交而产生，如BC1F2, BC2F2, F3和RIL等，基因型理论频率的推算需借助转移矩阵．

双杂合基因型AB/ab和Ab/aB在重组率估计中是不能区分的．虽然它们产生同样类似的配子，但同一种配子的频率却是不同的．在计算理论基因型频率时，要区分对待．在估计重组率时，一般仅知道两种双杂型的观察值之和．因此，需要再把这两种基因型的频率进行合并．为推导不同群体中各种基因型的频率，需考虑10种不同的基因型，称为类型1，类型2, …，类型10．两个基因座位上10种基因型的频率用行向量$\mathbf{f}^{(t)}$表示，即，

$$\mathbf{f}^{(t)} = \left[f_{AABB}^{(t)} \quad f_{AABb}^{(t)} \quad f_{AAbb}^{(t)} \quad f_{AaBB}^{(t)} \quad f_{AB/ab}^{(t)} \quad f_{Ab/aB}^{(t)} \quad f_{Aabb}^{(t)} \quad f_{aaBB}^{(t)} \quad f_{aaBB}^{(t)} \quad f_{aaBB}^{(t)} \right]$$

如果把组成群体的不同个体视为从遗传群体中抽取的一组随机样本，这组样本将服从频率为$\mathbf{f}^{(t)}$的多项分布．10种基因型包含了一个随机样本所有可能的取值．因此，行向量$\mathbf{f}^{(t)}$的元素之和为1，概率统计中称之为概率向量．为表达方便，把自交，回交，单倍

体加倍统称为交配. 交配之后, 群体进入 t+1 世代, 交配后的基因型也有 10 种可能, 但它们的频率却发生了变化, 交配后群体的频率用行向量 $\mathbf{f}^{(t+1)}$表示, 即,

$$\mathbf{f}^{(t+1)} = \left[f_{AABB}^{(t+1)} \quad f_{AABb}^{(t+1)} \quad f_{AAbb}^{(t+1)} \quad f_{AaBB}^{(t+1)} \quad f_{AB/ab}^{(t+1)} \quad f_{Ab/aB}^{(t+1)} \quad f_{Aabb}^{(t+1)} \quad f_{aaBB}^{(t+1)} \quad f_{aaBb}^{(t+1)} \quad f_{aabb}^{(t+1)} \right]$$

世代 t+1 的基因型频率仅依赖于世代 t 的基因型频率, 而与世代 t 之前的基因型频率无关. 如果把不同世代群体中, 个体的基因型看作随机变量, 这些随机变量则形成一个马尔可夫链. $\mathbf{T}$ 表示特定交配方式下, 一次交配的转移矩阵. 转移矩阵的每一行代表每种基因型产生的各种后代基因型的频率. 这个矩阵的每一行的元素之和为 1, 概率统计中称为概率转移矩阵. 这样, 一次交配发生后, 基因型的频率向量 $\mathbf{f}^{(t+1)}$就能表示为交配前的频率向量 $\mathbf{f}^{(t)}$与转移矩阵 $\mathbf{T}$ 的乘积, 即,

$$\mathbf{f}^{(t+1)} = \mathbf{f}^{(t)}\mathbf{T} \tag{2.1.1}$$

因此, 如果知道了各种交配方式的转移矩阵, 就能得到一个群体交配后, 各种基因型的理论频率. 下面首先给出与 P_1 回交一代, 与 P_2 回交一代, 自交一代和加倍单倍体一代后的转移矩阵. 用 $\mathbf{T}_{P1B}$ 表示与 P_1 回交一代的转移矩阵, $\mathbf{T}_{P2B}$ 表示与 P_2 回交一代的转移矩阵, $\mathbf{T}_S$ 表示自交一代的转移矩阵, $\mathbf{T}_D$ 表示加倍单倍体的转移矩阵.

§2.1.2 回交世代转移矩阵

首先, 以与亲本 P1 的回交为例, 来说明回交转移矩阵的计算 (公式 2.1.2 和 2.1.3). 分以下 10 种基因型来讨论.

(1) 基因型 AABB 与亲本 P1 (AABB) 回交, 后代的基因型全部为类型 1 (AABB), 因此转移矩阵 $\mathbf{T}_{P1B}$ 的第 1 行只有第 1 个元素为 1, 其他为均 0.

(2) 基因型 AABb 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 1 (AABB) 或类型 2 (AABb), 两种基因型的频率均为 $\frac{1}{2}$. 因此, 转移矩阵 $\mathbf{T}_{P1B}$ 第 2 行的前两个元素均为 $\frac{1}{2}$, 其他为均 0.

(3) 基因型 AAbb 与亲本 P1 (AABB) 回交, 后代的基因型全部为类型 2 (AABb). 因此, 转移矩阵 $\mathbf{T}_{P1B}$ 第 3 行的第 2 个元素为 1, 其他为均 0.

(4) 基因型 AaBB 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 1 (AABB) 或

类型 4 (AaBB), 两种基因型的频率均为 $\frac{1}{2}$. 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 4 行的第 1 和 4 两个元素均为 $\frac{1}{2}$, 其他为均 0.

(5) 基因型 AB/ab 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 1 (AABB), 类型 2 (AABb), 类型 4 (AaBB) 和类型 5 (AB/ab) 四种可能. 类型 1 和 5 是非交换型配子 AB 和 ab 与 P1 的配子 AB 杂交产生的基因型, 频率均为 $\frac{1}{2}(1-r)$. 类型 2 和 4 是交换型配子 Ab 和 aB 与 P1 的配子 AB 杂交产生的基因型, 频率均为 $\frac{1}{2}r$. 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 5 行的第 1, 2, 4 和 5 四个元素分别为 $\frac{1}{2}(1-r)$, $\frac{1}{2}r$, $\frac{1}{2}r$ 和 $\frac{1}{2}(1-r)$, 其他为均 0.

(6) 与基因型 AB/ab 类似, 基因型 Ab/aB 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 1 (AABB), 类型 2 (AABb), 类型 4 (AaBB) 和类型 5 (AB/ab) 四种可能. 但是, 相对于 Ab/aB 来说, 配子 AB 和 ab 是交换型, Ab 和 aB 是非交换型. 因此, 类型 1 和 5 是交换型配子 AB 和 ab 与 P1 的配子 AB 杂交产生的基因型, 频率均为 $\frac{1}{2}r$. 类型 2 和 4 是非交换型配子 Ab 和 aB 与 P1 的配子 AB 杂交产生的基因型, 频率均为 $\frac{1}{2}(1-r)$. 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 6 行的第 1, 2, 4 和 5 四个元素分别为 $\frac{1}{2}r$, $\frac{1}{2}(1-r)$, $\frac{1}{2}(1-r)$ 和 $\frac{1}{2}r$, 其他为均 0.

(7) 基因型 Aabb 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 2 (AABb) 或类型 5 (AB/ab), 两种基因型的频率均为 $\frac{1}{2}$. 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 7 行的第 2 和 5 两个元素均为 $\frac{1}{2}$, 其他为均 0.

(8) 基因型 aaBB 与亲本 P1 (AABB) 回交, 后代的基因型全部为类型 4 (AaBB). 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 8 行的第 4 个元素为 1, 其他为均 0.

(9) 基因型 aaBb 与亲本 P1 (AABB) 回交, 后代的基因型只能为类型 4 (AaBB) 或类型 5 (AB/ab), 两种基因型的频率均为 $\frac{1}{2}$. 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 第 9 行的第 4 和 5 两个元素均为 $\frac{1}{2}$, 其他为均 0.

(10) 基因型 aabb 与亲本 P1 (AABB) 回交, 后代的基因型全部为类型 5 (AB/ab). 因此, 转移矩阵 $\mathbf{T}_{\mathrm{P1B}}$ 的第 10 行只有第 5 个元素为 1, 其他为均 0.

$$\mathbf{T}_{\mathrm{P1B}}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}(1-r) & \frac{1}{2}r & 0 & \frac{1}{2}r & \frac{1}{2}(1-r) & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}r & \frac{1}{2}(1-r) & 0 & \frac{1}{2}(1-r) & \frac{1}{2}r & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.1.2}$$

与亲本 P1 回交后代中, 不可能出现类型 3 (AAbb), 类型 7 (Aabb), 类型 8 (aaBB), 类型 9 (aaBb) 和类型 10 (aabb). 因此, 转移矩阵 $\mathbf{T}_{\mathbf{P1B}}$ 的第 3, 7~10 列上的元素均为 0 (公式 2.1.2). 与亲本 P2 回交的转移矩阵 (公式 2.1.3) 的计算与 $\mathbf{T}_{\mathbf{P1B}}$ 类似, P2 回交后代中, 不可能出现类型 1~类型 4, 类型 6 和类型 8. 因此, 转移矩阵 $\mathbf{T}_{\mathbf{P2B}}$ 的第 1~4, 6 和 8 列上的元素均为 0. $\mathbf{T}_{\mathbf{P2B}}$ 第 5 列等于 $\mathbf{T}_{\mathbf{P1B}}$ 第 1 列, $\mathbf{T}_{\mathbf{P2B}}$ 第 7 列等于 $\mathbf{T}_{\mathbf{P1B}}$ 第 2 列, $\mathbf{T}_{\mathbf{P2B}}$ 第 9 列等于 $\mathbf{T}_{\mathbf{P1B}}$ 第 4 列, $\mathbf{T}_{\mathbf{P2B}}$ 第 10 列等于 $\mathbf{T}_{\mathbf{P2B}}$ 第 5 列.

$$\mathbf{T}_{\mathrm{P2B}}=\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}r & 0 & \frac{1}{2}r & \frac{1}{2}(1-r) \\ 0 & 0 & 0 & 0 & \frac{1}{2}r & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}(1-r) & \frac{1}{2}r \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1.3}$$

§2.1.2 自交世代转移矩阵

对自交转移矩阵 (公式 2.1.4), 按杂合座位的个数分以下三种情况讨论.

(1) 无杂合座位, 即两个座位上的基因型都纯合. 纯合基因型的自交后代的基因型与亲代相同, 四种纯合基因型分别对应于类型 1 (AABB), 类型 3 (AAbb), 类型 8 (aaBB) 和类型 10 (aabb). 因此, 转移矩阵 $\mathbf{T}_{\mathbf{S}}$ 第 1 行的第 1 个因素为 1, 其余因素为 0; 第 3 行的第 3 个因素为 1, 其余因素为 0; 第 8 行的第 8 个因素为 1, 其

余因素为 0; 第 10 行的第 10 个因素为 1, 其余因素为 0.

(2) 一个座位纯合, 一个座位杂合. 在杂合座位上, 自交后代的基因型按照 1:2:1 的比例分离, 即频率分别为 $\frac{1}{4}$, $\frac{1}{2}$ 和 $\frac{1}{4}$. 以类型 2 (AABb)为例, 自交后代的基因型为类型 1 (AABB), 类型 2 (AABb) 和类型 3 (AAbb), 频率分别为 $\frac{1}{4}$, $\frac{1}{2}$ 和 $\frac{1}{4}$. 因此, 转移矩阵 $\mathbf{T_S}$ 第 2 行的第 1, 2 和 3 个元素为 $\frac{1}{4}$, $\frac{1}{2}$ 和 $\frac{1}{4}$, 其余为 0. 类型 4 (AaBB), 类型 7 (Aabb) 和类型 9 (aaBb) 与类型 2 类似.

(3) 两个座位均杂合, 即类型 5 (AB/ab) 和类型 6 (Ab/aB). 所有 10 种类型都可能在自交后代中出现, 以类型 5 (AB/ab) 为例说明转移频率的计算 (公式 2.1.4). 基因型 AB/ab 将产生四种配子型, 即 AB, Ab, aB 和 ab. 相对于亲代基因型 AB/ab 来说, AB 和 ab 是非交换型, 频率均为 $\frac{1}{2}(1-r)$; Ab 和 aB 是交换型, 频率均为 $\frac{1}{2}r$. 在不存在配子选择的情况下, 基因型 AB/ab 产生同样频率的雌配子和雄配子. 自交等同于基因型 AB/ab 产生的雌配子和雄配子间的随机结合, 雌配子和雄配子间随机结合后的基因型及其频率见表 2.1.1. 对角线为四种纯合基因型及其频率, 对应于转移矩阵 $\mathbf{T_S}$ 第 5 行的第 1, 3, 8 和 10 四个元素. 对于自交后代类型 2, 可通过雌配子 AB 和雄配子 Ab 结合而产生, 也可通过雌配子 Ab 和雄配子 AB 结合而产生. 因此, 类型 2 的频率为 $\frac{1}{4}r(1-r)+\frac{1}{4}r(1-r)=\frac{1}{2}r(1-r)$, 这个频率对应于转移矩阵 $\mathbf{T_S}$ 第 5 行的第 2 个元素. 与此类似, 可以计算后代类型 4, 5, 6, 7 和 9 的转移频率. 类型 6 (AB/ab) 的自交转移概率与类型 5 (Ab/aB) 类似, 只是把类型 5 (Ab/aB) 中的 (1-*r*) 替换为 *r*, *r* 替换为 (1-*r*) 即可.

表 2.1.1 杂合基因型 AB/ab 产生的配子型和自交后代基因型的频率

雌配子型及其频率	雄配子型及其频率			
	AB, $\frac{1}{2}(1-r)$	Ab, $\frac{1}{2}r$	aB, $\frac{1}{2}r$	ab, $\frac{1}{2}(1-r)$
AB, $\frac{1}{2}(1-r)$	类型 1: AABB $\frac{1}{4}(1-r)^2$	类型 2: AABb $\frac{1}{4}r(1-r)$	类型 4: AaBB $\frac{1}{4}r(1-r)$	类型 6: AB/ab $\frac{1}{4}(1-r)^2$
Ab, $\frac{1}{2}r$	类型 2: AABb $\frac{1}{4}r(1-r)$	类型 3: AAbb $\frac{1}{4}r^2$	类型 5: Ab/aB $\frac{1}{4}r^2$	类型 7: Aabb $\frac{1}{4}r(1-r)$
aB, $\frac{1}{2}r$	类型 4: AaBB $\frac{1}{4}r(1-r)$	类型 5: Ab/aB $\frac{1}{4}r^2$	类型 8: aaBB $\frac{1}{4}r^2$	类型 9: aaBb $\frac{1}{4}r(1-r)$
ab, $\frac{1}{2}(1-r)$	类型 6: AB/ab $\frac{1}{4}(1-r)^2$	类型 7: Aabb $\frac{1}{4}r(1-r)$	类型 9: aaBb $\frac{1}{4}r(1-r)$	类型 10: aabb $\frac{1}{4}(1-r)^2$

$$\mathbf{T}_{\mathrm{S}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 \\ \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{2}r^2 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1.4}$$

§2.1.3 加倍单倍体世代转移矩阵

加倍单倍体的情况比较简单，后代基因型的频率等于配子的频率 (公式 2.1.5). 与自交类似，按杂合座位数分以下三种情况讨论.

(1) 无杂合座位，即两个座位上的基因型都纯合. 纯合基因型只产生一种配子，加倍之后的基因型与亲代的基因型相同，四种纯合基因型分别对应于类型 1 (AABB), 类型 3 (AAbb), 类型 8 (aaBB) 和类型 10 (aabb). 因此，转移矩阵 $\mathbf{T_D}$ 第 1 行的第 1 个因素为 1，其余因素为 0；第 3 行的第 3 个因素为 1，其余因素为 0；第 8 行的第 8 个因素为 1，其余因素为 0；第 10 行的第 10 个因素为 1，其余因素为 0.

(2) 一个座位纯合，一个座位杂合. 配子按照 1:1 的比例分离，即频率分别为 $\frac{1}{2}$ 和 $\frac{1}{2}$. 以类型 2 (AABb)为例，配子型 AB 和 Ab 的频率为 $\frac{1}{2}$ 和 $\frac{1}{2}$，加倍后分别为类型 1 (AABB) 和类型 3 (AAbb)，频率仍为 $\frac{1}{2}$ 和 $\frac{1}{2}$. 因此，转移矩阵 $\mathbf{T_D}$ 第 2 行的第 1 和 3 个元素为 $\frac{1}{2}$ 和 $\frac{1}{2}$，其余为 0. 类型 4 (AaBB)，类型 7 (Aabb) 和类型 9 (aaBb) 与类型 2 类似.

(3) 两个座位均杂合，即类型 5 (AB/ab) 和类型 6 (Ab/aB). 四种纯合类型都可能出现，以类型 5 为例说明转移频率的计算. 基因型 AB/ab 产生四种配子型，即 AB, Ab, aB 和 ab，频率分别为 $\frac{1}{2}(1-r)$，$\frac{1}{2}r$，$\frac{1}{2}r$ 和 $\frac{1}{2}(1-r)$. 这四种配子加倍后为类型 1 (AABB), 类型 3 (AAbb), 类型 8 (aaBB) 和类型 10 (aabb)，频率仍分别为 $\frac{1}{2}(1-r)$，$\frac{1}{2}r$，$\frac{1}{2}r$ 和 $\frac{1}{2}(1-r)$. 因此，$\mathbf{T_D}$ 第 5 行的第 1, 3, 8 和 10 四个元素分别为 $\frac{1}{2}(1-r)$，$\frac{1}{2}r$，$\frac{1}{2}r$ 和 $\frac{1}{2}(1-r)$. 类型 6 (AB/ab) 的加倍单倍体转移概率与类型 5 (Ab/aB) 类似，只是把类型 5 (Ab/aB) 中的 (1-*r*) 替换为 *r*, *r* 替换为 (1-*r*) 即可.

$$\mathbf{T}_{\mathrm{D}}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2}(1-r) & 0 & \frac{1}{2}r & 0 & 0 & 0 & 0 & \frac{1}{2}r & 0 & \frac{1}{2}(1-r) \\ \frac{1}{2}r & 0 & \frac{1}{2}(1-r) & 0 & 0 & 0 & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}r \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1.5}$$

§2.1.4 连续自交的世代转移矩阵

连续自交是产生重组近交家系经常采用的交配方式. 连续自交的转移矩阵与加倍单倍体类似, 区别之处在于重组率, 连续自交过程中会产生更多的交换机会. 对连续自交转移矩阵 (公式 2.1.6), 也按杂合座位的个数分以下三种情况讨论.

$$\mathbf{T}_{\mathrm{R}}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2}(1-R) & 0 & \frac{1}{2}R & 0 & 0 & 0 & 0 & \frac{1}{2}R & 0 & \frac{1}{2}(1-R) \\ \frac{1}{2}R & 0 & \frac{1}{2}(1-R) & 0 & 0 & 0 & 0 & \frac{1}{2}(1-R) & 0 & \frac{1}{2}R \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1.6}$$

(1) 无杂合座位, 即两个座位上的基因型都纯合. 纯合基因型的连续自交后代的基因型与亲代相同, 四种纯合基因型分别对应于类型 1 (AABB), 类型 3 (AAbb), 类型 8 (aaBB) 和类型 10 (aabb). 因此, 转移矩阵 $\mathbf{T_R}$ 第 1 行的第 1 个因素为 1, 其余元素为 0; 第 3 行的第 3 个元素为 1, 其余元素为 0; 第 8 行的第 8 个元素为 1, 其余元素为 0; 第 10 行的第 10 个元素为 1, 其余元素为 0.

(2) 一个座位纯合, 一个座位杂合. 每自交一次, 杂合基因型的频率下降一半. 连续自交多代后, 杂合基因型的频率接近于 0, 两种纯合基因型的频率分别为 $\frac{1}{2}$ 和 $\frac{1}{2}$. 以类型 2 (AABb)为例, 连续自交后代的基因型为类型 1 (AABB) 和类型 3 (AAbb), 频率分别 $\frac{1}{2}$ 和 $\frac{1}{2}$. 因此, 转移矩阵 $\mathbf{T_R}$ 第 2 行的第 1 和 3 个元素为 $\frac{1}{2}$ 和 $\frac{1}{2}$, 其余为 0. 类型 4 (AaBB), 类型 7 (Aabb) 和类型 9 (aaBb) 与类型 2 类似.

(3) 两个座位均杂合, 即类型 5 (AB/ab) 和类型 6 (Ab/aB). 在连续自交过程中, 杂合基因型的频率逐渐下降到 0, 四种纯合基因型的频率之和逐渐接近于 1. 现以类型 5 为例说明连续自交转移频率的计算 (表 2.1.6). 基因型 AB/ab 连续自交无穷多代后, 群体中也只有四种纯合基因型 AABB, AAbb, aaBB 和 aabb. 相对于基因型 AB/ab 来说, AAbb 和 aaBB 是交换型, 频率之和用 *R* 表示, AABB 和 aabb 是非交换型, 频率之和用 1-*R* 表示. 由于等位基因 A, a, B 和 b 的期望频率均为 $\frac{1}{2}$, AAbb 和 aaBB 有相等的频率, 即频率均为 $\frac{1}{2}R$. AABB 和 aabb 有相等的频率, 即

频率均为$\frac{1}{2}(1-R)$. 因此, 四种纯合基因型 AABB, AAbb, aaBB 和 aabb 的频率分别为$\frac{1}{2}(1-R)$, $\frac{1}{2}R$, $\frac{1}{2}R$ 和$\frac{1}{2}(1-R)$, 转移矩阵第 5 行的第 1, 3, 8 和 10 个元素分别为$\frac{1}{2}(1-R)$, $\frac{1}{2}R$, $\frac{1}{2}R$ 和$\frac{1}{2}(1-R)$, 其他元素为 0. R 表示连续自交的累计重组率, 利用矩阵的谱分解和马尔可夫链的性质等方面的知识, 可以证明 R 与一次交换的重组率 r 的关系是,

$$R=\frac{2r}{1+2r} \quad 或 \quad r=\frac{R}{2(1-R)} \tag{2.1.7}$$

§2.1.6 基因型理论频率的矩阵表示

利用 2.1.2~2.1.6 这五种转移矩阵, 图 1.1.1 的 20 种双亲群体中, 各种基因型的理论频率都可以用杂种 F_1 的频率与转移矩阵的乘积来表示. F_1 的基因型为 AB/ab, 频率为 1, 其余类型的频率为 0, 即,

$$\mathbf{f}^{(0)}=\begin{bmatrix}0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\end{bmatrix} \tag{2.1.8}$$

表 2.1.2 给出 20 种双亲遗传群体中, 基因型理论频率与 F_1 的频率向量 $\mathbf{f}^{(0)}$和各种转移矩阵 (公式 2.1.2~2.1.6) 的关系. 利用这些关系就能推导出这些群体中基因型理论频率, 进而用于重组率的极大似然估计.

表 2.1.2 双亲遗传研究群体中, 基因型理论频率与杂种 F_1 的频率 $\mathbf{f}^{(0)}$和转移矩阵的关系

群体编号	群体名称	基因型理论频率的表达式
1	P1BC1F1	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}$
2	P2BC1F1	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}$
3	F1DH	$\mathbf{f}^{(0)}\times\mathbf{T}_{D}$
4	F1RIL	$\mathbf{f}^{(0)}\times\mathbf{T}_{R}$
5	P1BC1RIL	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{R}$
6	P2BC1RIL	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{R}$
7	F2	$\mathbf{f}^{(0)}\times\mathbf{T}_{S}$
8	F3	$\mathbf{f}^{(0)}\times\mathbf{T}_{S}\times\mathbf{T}_{S}$
9	P1BC2F1	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{P1B}$
10	P2BC2F1	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{P2B}$
11	P1BC2RIL,	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{P1B}\times\mathbf{T}_{R}$
12	P2BC2RIL,	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{P2B}\times\mathbf{T}_{R}$
13	P1BC1F2	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{S}$
14	P2BC1F2	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{S}$

15	P1BC2F2	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{P1B}\times\mathbf{T}_{S}$
16	P2BC2F2	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{P2B}\times\mathbf{T}_{S}$
17	P1BC1DH	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{D}$
18	P2BC1DH	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{D}$
19	P1BC2DH	$\mathbf{f}^{(0)}\times\mathbf{T}_{P1B}\times\mathbf{T}_{P1B}\times\mathbf{T}_{D}$
20	P2BC2DH	$\mathbf{f}^{(0)}\times\mathbf{T}_{P2B}\times\mathbf{T}_{P2B}\times\mathbf{T}_{D}$

§2.2 两个座位上各种基因型的理论频率

§2.2.1 十种基因型的理论频率

根据表 2.1.2 中的表达式，可以计算图 1.1.1 中各种双亲群体的 10 种基因型的理论频率，结果列于表 2.2.1. 表 2.2.1 中的理论频率是利用各种双亲遗传群体估计重组率的基础 (Nelson, 2011; Sun et al., 2012).

表 2.2.1 双亲群体中, 两个基因座位上十种可能基因型的理论频率 (空白表示频率为 0, $R=\frac{2r}{1+2r}$)

群体名称	AABB	AABb	AAbb	AaBB	AB/ab	Ab/aB	Aabb	aaBB	aaBb	aabb
P1BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$					
P2BC1F1					$\frac{1}{2}(1-r)$		$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F1DH	$\frac{1}{2}(1-r)$		$\frac{1}{2}r$					$\frac{1}{2}r$		$\frac{1}{2}(1-r)$
F1RIL	$\frac{1}{2}(1-R)$		$\frac{1}{2}R$					$\frac{1}{2}r$		$\frac{1}{2}(1-R)$
P1BC1RIL	$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$					$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}(1-r)(1-R)$
P2BC1RIL	$\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$					$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$		$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$
F2	$\frac{1}{4}(1-r)^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(1-r)^2$	$\frac{1}{2}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}(1-r)^2$
F3	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r^4+\frac{1}{4}(1-r)^4$	$\frac{1}{2}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$					
P2BC2F1					$\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC2RIL	$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$					$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}(1-r)^2(1-R)$
P2BC2RIL	$\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$					$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$		$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$
P1BC1F2	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)^3$	$\frac{1}{4}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{8}(1-r)^3$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)^3$	$\frac{1}{4}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{8}(1-r)^4$	$\frac{1}{8}r^2(1-r)^2$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{16}(1-r)^4$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{8}(1-r)^4$	$\frac{1}{8}r^2(1-r)^2$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$
P1BC1DH	$\frac{1}{2}+\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$					$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}(1-r)^2$
P2BC1DH	$\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$					$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC2DH	$\frac{3}{4}+\frac{1}{8}(1-r)^3$		$\frac{1}{8}-\frac{1}{8}(1-r)^3$					$\frac{1}{8}-\frac{1}{8}(1-r)^3$		$\frac{1}{8}(1-r)^3$
P2BC2DH	$\frac{1}{8}(1-r)^3$		$\frac{1}{8}-\frac{1}{8}(1-r)^3$					$\frac{1}{8}-\frac{1}{8}(1-r)^3$		$\frac{3}{4}+\frac{1}{8}(1-r)^3$

§2.2.2 永久群体中四种纯合基因型的理论频率

为方便起见, 把 10 种永久群体中, 各种基因型的理论频率列于表 2.2.2. 从中可以看出, F1DH 中理论频率有最简单的表达形式, 重组型 AAbb 和 aaBB 占的比例为 r, 亲本型 AABB 和 aabb 占的比例为 1-r. 因此, 重组型占总 DH 家系的比例就是重组率的估计. F1RIL 中, 重组型 AAbb 和 aaBB 占的比例为 R, 亲本型 AABB 和 aabb 占的比例为 1-R. 因此, 重组型占总 RIL 家系的比例可作为累积重组率 R 的估计, 进而根据公式 (2.1.6) 计算一次减数分裂时的重组率. 其他 8 种永久群体的期望频率中, P1BC1RIL 和 P2BC1RIL 有共同项 $(1-r)(1-R)$, P1BC2RIL 和 P2BC2RIL 有共同项 $(1-r)^2(1-R)$, P1BC1DH 和 P2BC1DH 有共同项 $(1-r)^2$, P1BC2DH 和 P2BC2DH 有共同项 $(1-r)^3$ (表 2.2.2). 进行重组率估计时, 可先估计出这些共同项, 然后估计一次交换的重组率. 这样, 可以避免使用一些较复杂的迭代算法.

表 2.2.2 永久群体中可识别的四种基因型的理论频率 ($R=\frac{2r}{1+2r}$)

群体名称	AABB	AAbb	aaBB	aabb
F1DH	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F1RIL	$\frac{1}{2}(1-R)$	$\frac{1}{2}R$	$\frac{1}{2}R$	$\frac{1}{2}(1-R)$
P1BC1RIL	$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$	$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$	$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$	$\frac{1}{4}(1-r)(1-R)$
P2BC1RIL	$\frac{1}{4}(1-r)(1-R)$	$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$	$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$	$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$
P1BC2RIL	$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$	$\frac{1}{8}(1-r)^2(1-R)$
P2BC2RIL	$\frac{1}{8}(1-r)^2(1-R)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$	$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$
P1BC1DH	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$
P2BC1DH	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC2DH	$\frac{3}{4}+\frac{1}{8}(1-r)^3$	$\frac{1}{8}-\frac{1}{8}(1-r)^3$	$\frac{1}{8}-\frac{1}{8}(1-r)^3$	$\frac{1}{8}(1-r)^3$
P2BC2DH	$\frac{1}{8}(1-r)^3$	$\frac{1}{8}-\frac{1}{8}(1-r)^3$	$\frac{1}{8}-\frac{1}{8}(1-r)^3$	$\frac{3}{4}+\frac{1}{8}(1-r)^3$

§2.2.3 两个共显性标记在暂时群体中基因型的理论频率

实际遗传群体中, 两种双杂合类型 AB/ab 和 Ab/aB 是无法区分的. 将二者合并, 用基因型 AaBb 表示, 其理论频率是 AB/ab 和 Ab/aB 的频率之和. 这样就得到, 当两个座位上的标记均为共显性时, 可识别的九种基因型的理论频率 (表 2.2.3). 利用这些理论频率, 就能构造一组样本观测值的极大似然函数, 从而估计两个共显性标记间的重组率.

2.2.3 等位基因 A 和 a 是共显性, B 和 b 是共显性时, 暂时群体中可识别的九种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
P1BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$				
P2BC1F1					$\frac{1}{2}(1-r)$	$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F2	$\frac{1}{4}(1-r)^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(1-2r+2r^2)$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}(1-r)^2$
F3	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}(1-2r+2r^2)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$				
P2BC2F1					$\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC1F2	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)(1-2r+2r^2)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{8}(1-r)^3$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)(1-2r+2r^2)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{8}(1-r)^2(1-2r+2r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{16}(1-r)^4$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{8}(1-r)^2(1-2r+2r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$

§2.2.4 一个共显性和一个显性标记在暂时群体中基因型的理论频率

如果标记基因 A 和 a 是共显性，标记基因 B 对 b 表现为显性，即标记基因型 BB 和 Bb 无法区分. 群体中能识别的标记类型只有六种，即 (1) AAB_ (包含 AABB 和 AABb 两种基因型); (2) AAbb; (3) AaB_ (包含 AaBB 和 AaBb 两种基因型); (4) Aabb; (5) aaB_ (包含 aaBB 和 aaBb 两种基因型); (6) aabb. 这六种可识别基因型的理论频率列于表 2.2.4. 容易看出, P1BC1F1 和 P1BC2F1 群体中，无法估计共显性标记和显性标记之间的重组率.

表 2.2.4 等位基因 A 和 a 是共显性, B 对 b 为显性时，暂时群体中可识别的 6 种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB+AABb (或 AAB_)	AAbb	AaBB+ AaBb (或 AaB_)	Aabb	aaBB+ aaBb (或 aaB_)	aabb
P1BC1F1	$\frac{1}{2}$		$\frac{1}{2}$			
P2BC1F1			$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F2	$\frac{1}{4}(1-r^2)$	$\frac{1}{4}r^2$	$\frac{1}{2}(1-r+r^2)$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r(2-r)$	$\frac{1}{4}(1-r)^2$
F3	$\frac{1}{4}(\frac{3}{2}-r-r^2+2r^3-r^4)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}(\frac{1}{2}-r+2r^2-2r^3+r^4)$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	$\frac{3}{4}$		$\frac{1}{4}$			
P2BC2F1			$\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC1F2	$\frac{1}{2}+\frac{1}{8}(1-r)^2(1+r)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}(1-r)(1-r+r^2)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{8}(1-r)^2(1+r)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{3}{4}+\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{8}-\frac{1}{8}r(1-r)^3$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$

§2.2.5 一个共显性和一个隐性标记在暂时群体中基因型的理论频率

如果标记基因 A 和 a 是共显性，标记基因 B 对 b 表现为隐性，即标记基因型 Bb 和 bb 无法区分. 群体中能识别的标记类型只有六种，即 (1) AABB; (2) AA_b (包含 AABb 和 AAbb 两种基因型); (3) AaBB; (4) Aa_b (包含 AaBb 和 Aabb 两种基因型); (5) aaBB; (6) aa_b (包含 aaBB 和 aaBb 两种基因型). 这六种可识别基因型的理论频率列于表 2.2.5. 容易看出, P2BC1F1 和 P2BC2F1 群体中，无法估计共显性标记和隐性标记之间的重组率.

表 2.2.5 等位基因 A 和 a 是共显性, B 对 b 为隐性时，群体中可识别的六种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB	AABb+ AAbb (或 AA_b)	AaBB	AaBb+ Aabb (或 Aa_b)	aaBB	aaBb+ aabb (或 aa_b)
P1BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$		
P2BC1F1				$\frac{1}{2}$		$\frac{1}{2}$
F2	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(2-r)$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(1-r+r^2)$	$\frac{1}{4}r^2$	$\frac{1}{4}(1-r^2)$
F3	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{2}(\frac{1}{2}-r+2r^2-2r^3+r^4)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{4}(\frac{3}{2}-r-r^2+2r^3-r^4)$
P1BC2F1	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$		
P2BC2F1				$\frac{1}{4}$		$\frac{3}{4}$
P1BC1F2	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)(1-r+r^2)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{8}(1-r)^2(1+r)$
P2BC1F2	$\frac{1}{8}(1-r)^3$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{4}r+\frac{1}{4}(1-r)(1-r+r^2)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{2}+\frac{1}{8}(1-r)^2(1+r)$
P1BC2F2	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}(1-r)^3(1+r)$
P2BC2F2	$\frac{1}{16}(1-r)^4$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{8}-\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{3}{4}+\frac{1}{16}(1-r)^3(1+r)$

§2.2.6 两个显性标记在暂时群体中基因型的理论频率

如果标记基因 A 对 a 是显性，标记基因 B 对 b 表现为显性，即标记基因型 AA 和 Aa 无法区分，标记基因型 Bb 和 bb 无法区分. 群体中能识别的标记类型只有四种，即 (1) A_B_ (包含 AABB, AABb, AaBB 和 AaBb 四种基因型); (2) A_bb (包含 AAbb 和 Aabb 两种基因型); (3) aaB_ (包含 aaBB 和 aaBb 两种基因型); (4) aabb. 这四种可识别基因型的理论频率列于表 2.2.6. 容易看出, P1BC1F1 和 P1BC2F1 群体中，无法估计显性标记之间的重组率.

表 2.2.6 等位基因 A 对 a 是显性, B 对 b 为显性时，群体中可识别的四种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB+AABb+AaBB+AaBb (或 A_B_)	AAbb+ Aabb (或 A_bb)	aaBB+ aaBb (或 aaB_)	aabb
P1BC1F1	1			
P2BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F2	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(2-r)$	$\frac{1}{4}r(2-r)$	$\frac{1}{4}(1-r)^2$
F3	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	1			
P2BC2F1	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC1F2	$\frac{3}{4}+\frac{1}{8}(1-r)^3$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^3$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{7}{8}+\frac{1}{16}(1-r)^4$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$

§2.2.7 一个显性和一个隐性标记在暂时群体中基因型的理论频率

如果标记基因 A 对 a 是显性，标记基因 B 对 b 表现为隐性，即标记基因型 AA 和 Aa 无法区分，标记基因型 Bb 和 bb 无法区分. 群体中能识别的标记类型只有四种，即 (1) A_BB (包含 AABB 和 AaBB 两种基因型); (2) A_ _b (包含 AABb, AAbb, AaBb 和 Aabb 四种基因型); (3) aaBB; (4) aa_b (包含 aaBb 和 aabb 两种基因型). 这四种可识别基因型的理论频率列于表 2.2.7. 容易看出, P1BC1F1, P2BC1F1, P1BC2F1 和 P2BC2F1 四种群体中，无法估计显性标记和隐性标记之间的重组率.

表 2.2.7 等位基因 A 对 a 是显性, B 对 b 为隐性时，群体中识别的四种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB+ AaBB (或 A_BB)	AABb+AAbb+ AaBb+ Aabb (或 A_ _b)	aaBB	aaBb+ aabb (或 aa_b)
P1BC1F1	$\frac{1}{2}$	$\frac{1}{2}$		
P2BC1F1		$\frac{1}{2}$		$\frac{1}{2}$
F2	$\frac{1}{4}(1-r^2)$	$\frac{1}{2}+\frac{1}{4}r^2$	$\frac{1}{4}r^2$	$\frac{1}{4}(1-r^2)$
F3	$\frac{1}{4}(\frac{3}{2}-r-r^2+2r^3-r^4)$	$\frac{1}{4}(1+r+r^2-2r^3+r^4)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{4}(\frac{3}{2}-r-r^2+2r^3-r^4)$
P1BC2F1	$\frac{3}{4}$	$\frac{1}{4}$		
P2BC2F1		$\frac{1}{4}$		$\frac{3}{4}$
P1BC1F2	$\frac{1}{2}+\frac{1}{8}(1-r)^2(1+r)$	$\frac{3}{8}r+\frac{1}{8}(1-r)(2+r^2)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{8}(1-r)^2(1+r)$
P2BC1F2	$\frac{1}{8}(1-r)^2(1+r)$	$\frac{3}{8}r+\frac{1}{8}(1-r)(2+r^2)$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{2}+\frac{1}{8}(1-r)^2(1+r)$
P1BC2F2	$\frac{3}{4}+\frac{1}{16}(1-r)^3(1+r)$	$\frac{3}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}(1-r)^3(1+r)$
P2BC2F2	$\frac{1}{16}(1-r)^3(1+r)$	$\frac{3}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{1}{16}-\frac{1}{16}(1-r)^3(1+r)$	$\frac{3}{4}+\frac{1}{16}(1-r)^3(1+r)$

§2.2.8 两个隐性标记在暂时群体中基因型的理论频率

如果标记基因 A 对 a 是隐性，标记基因 B 对 b 表现为隐性，即标记基因型 Aa 和 aa 无法区分，标记基因型 Bb 和 bb 无法区分. 群体中能识别的标记类型只有四种，即 (1) AABB; (2) AA_b (包含 AABb 和 AAbb 两种基因型); (3) _aBB (包含 AaBB 和 aaBB 两种基因型); (4) _a_b (包含 AaBb, Aabb, aaBb 和 aabb 四种基因型). 这四种可识别基因型的理论频率列于表 2.2.8. 容易看出, P2BC1F1 和 P2BC2F1 两种群体中，无法估计隐性标记之间的重组率.

表 2.2.8 等位基因 A 对 a 是隐性, B 对 b 为隐性时，群体中识别的四种基因型的理论频率 (空白表示频率为 0)

群体名称	AABB	AABb+ AAbb (或 AA_b)	AaBB+ aaBB (或_aBB)	AaBb+Aabb+aaBb+ aabb (或_a_b)
P1BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
P2BC1F1				**1**
F2	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(2-r)$	$\frac{1}{4}r(2-r)$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
F3	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)(2-r+r^2)$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$
P2BC2F1				**1**
P1BC1F2	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{3}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{8}(1-r)^3$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{1}{8}r+\frac{1}{8}r(1-r)(2-r)$	$\frac{3}{4}+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{3}{16}-\frac{1}{8}(1-r)^2-\frac{1}{16}(1-r)^4$	$\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{16}(1-r)^4$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{1}{16}-\frac{1}{16}(1-r)^4$	$\frac{7}{8}+\frac{1}{16}(1-r)^4$

§2.3 两个标记/基因座位间重组率的估算

§2.3.1 DH 群体中重组率的极大似然估计

利用杂种 F_1 植株上的配子培养的 DH 群体, 具有最简单的遗传结构, 所谓遗传结构就是一个遗传群体中的基因和基因型频率. 我们首先以 DH 群体为例, 介绍重组率的极大似然估计的基本原理. 假定亲本 P_1 和 P_2 的标记基因型分别为 AABB 和 aabb, 两个标记间的重组率为 r. 杂种 F_1 的基因型为 AB/ab, 在 F_1 将产生基因型为 AB, Ab, aB 和 ab 的四种配子类型. AB 和 ab 称为亲本配子型, Ab 和 aB 称为交换配子型. 根据遗传学的交换原理, 亲本型的频率等于 1-r, 交换型的频率为 r. F_1 群体中, 每个等位基因的频率均为 0.5, AB 和 ab 出现的频率相同, Ab 和 aB 出现的频率相同. 因此, 四种配子类型 AB, Ab, aB 和 ab 的频率分别为 $\frac{1}{2}$(1-r), $\frac{1}{2}r$, $\frac{1}{2}r$ 和 $\frac{1}{2}$(1-r). 同时, 这些频率也就是 DH 群体中四种基因型 AABB, AAbb, aaBB 和 aabb 的频率. 表 2.3.1 中, n_1 和 n_4 为亲本基因型的 DH 家系数, n_2 和 n_3 为重组基因型的 DH 家系数, 总的观测个体数为 $n = n_1 + n_2 + n_3 + n_4$.

表 2.3.1 DH 群体中的期望基因型频率和观测值

基因型	AABB	AAbb	aaBB	aabb
基因型编码	(2, 2)	(2, 0)	(0, 2)	(0, 0)
期望或理论频率	$f_1 = \frac{1}{2}(1-r)$	$f_2 = \frac{1}{2}r$	$f_3 = \frac{1}{2}r$	$f_4 = \frac{1}{2}(1-r)$
观测样本量	n_1	n_2	n_3	n_4
Act8A 和 OP06 的样本量	64	8	7	61

以图 1.2.4 的大麦 DH 群体为例. 对标记 Act8A 和 OP06 来说, AABB 代表亲本 Harrington 的标记型, 编码为 (2, 2). aabb 代表亲本 TR306 的标记型, 编码为 (0, 0). 两种重组基因型的编码为 (2, 0) 和 (0, 2). 四种标记型的观测值分别为 64, 8, 7 和 61, 总样本量 n=140 (表 2.3.1). 家系 3 的标记 Act8A 缺失, 家系 55, 85, 105, 120 的标记 TR306 缺失, 因此这里的样本量小于图 1.2.4 的家系数 145. 采用极大似然方法估计重组率的基本步骤为如下.

(1) 建立重组率 r 的似然函数. 表 2.3.1 中的观测次数 n_1, n_2, n_3 和 n_4 服从频率为 f_1, f_2, f_3 和 f_4 的多项分布. 因此似然函数为,

$$L(r)=\frac{n!}{n_1!n_2!n_3!n_4!}[\tfrac{1}{2}(1-r)]^{n_1}(\tfrac{1}{2}r)^{n_2}(\tfrac{1}{2}r)^{n_3}[\tfrac{1}{2}(1-r)]^{n_4}$$

$$=C(1-r)^{n_1+n_4}r^{n_2+n_3} \tag{2.3.1}$$

其中 $C=\frac{n!}{n_1!n_2!n_3!n_4!}(\frac{1}{2})^{n_1+n_2+n_3+n_4}$ 为不依赖于重组率 *r* 的常数.

(2) 建立对数似然函数. 对似然函数 (公式 2.3.1) 直接求解有时很困难, 这时, 往往对似然函数求自然对数, 即,

$$\ln L(r)=\ln C+(n_1+n_4)\ln(1-r)+(n_2+n_3)\ln(r) \tag{2.3.2}$$

(3) 求对数似然函数 (公式 2.3.2) 对重组率 *r* 的一阶和二阶导数.

$$[\ln L(r)]'\hat{=}\frac{d\ln L}{dr}=-\frac{n_1+n_4}{1-r}+\frac{n_2+n_3}{r} \tag{2.3.3}$$

$$[\ln L(r)]''=\frac{d^2\ln L}{d^2r}=-\frac{n_1+n_4}{(1-r)^2}-\frac{n_2+n_3}{r^2} \tag{2.3.4}$$

(4) 求解重组率 *r* 的极大似然估计. 令一阶导数 (公式 2.3.3) 等于 0, 得到重组率估计的极大似然估计为,

$$\hat{r}=\frac{n_2+n_3}{n_1+n_2+n_3+n_4}=\frac{n_2+n_3}{n} \tag{2.3.5}$$

(5) 求重组率估计值的方差. 极大似然估计的方差一般从 Fisher 信息量获得, Fisher 信息量 *I* 等于对数似然函数二阶导数的相反数, 一般可作为估计量方差的估计. 因此,

$$I=-[\ln L(r)]''|_{r=\hat{r}}=[-\frac{n_1+n_4}{(1-r)^2}-\frac{n_2+n_3}{r^2}]|_{r=\hat{r}}=\frac{n}{\hat{r}(1-\hat{r})} \tag{2.3.6}$$

$$V_{\hat{r}}=\frac{1}{I}=\frac{\hat{r}(1-\hat{r})}{n} \tag{2.3.7}$$

(6) 重组率显著性的似然比检验. 显著性检验的零假设是 H_0: r=0.5, 即两个基因座位间不存在连锁关系. 备择假设是 H_A: r<0.5, 即两个基因位点间存在连锁关系. 似然比统计量 (likelihood ratio test, LRT) 定义为备择假设和零假设两种情形下, 极大似然函数比值的自然对数的 2 倍. LRT 统计量在大样本的情况下, 近似服从于卡方分布, 卡方分布的自由度等于两种假设下独立参数个数间的差异, 此时为 1. 即,

$$\max L(H_0) = L(r = 0.5) = C(\tfrac{1}{2})^n,$$

$$\max L(H_a) = L(r = \hat{r}) = C(1-\hat{r})^{n_1+n_4}(\hat{r})^{n_2+n_3},$$

$$LRT = -2\ln\frac{\max L(H_0)}{\max L(H_A)} = -2\ln\frac{(\frac{1}{2})^n}{(1-\hat{r})^{n_1+n_4}(\hat{r})^{n_2+n_3}}$$

$$= 2(n_1+n_4)\ln[2(1-\hat{r})] + 2(n_2+n_3)\ln(2\hat{r}) \sim \chi^2(1) \qquad (2.3.8)$$

对大麦DH中的标记Act8A和OP06来说 (表2.3.1), $\hat{r} = 0.1071$, $SE(\hat{r}) = 0.0261$. 似然比统计量 $LRT = 98.44$ (P=2.88×10^{-23}), 说明它们之间存在极显著的遗传连锁关系.

§2.3.2 重组率极大似然估计的一般形式

假定某个遗传群体中的可能基因型类型为 k 种, 每种基因型的理论频率为 f_i (i=1, 2, …, k). 观察群体中 n 个个体的基因型, 每种类型基因型的观察次数为 n_i (i=1, 2, …, k).

(1) 建立似然函数.

$$L(r) = \frac{n!}{n_1!n_2!\cdots n_k!}(f_1)^{n_1}(f_2)^{n_2}\cdots(f_k)^{n_k} \qquad (2.3.9)$$

(2) 建立对数似然函数. 对似然函数 (公式 2.3.8) 直接求解有时很困难, 为便于微分, 往往对似然函数 (公式 2.3.8) 作对数变换. 得到的对数似然函数为,

$$\ln L(r) = \ln C + n_1 \ln f_1 + n_2 \ln f_2 + \cdots + n_k \ln f_k \qquad (2.3.10)$$

其中 $C=\frac{n!}{n_1!n_2!\cdots n_k!}$ 为常数项，与待估计的重组率 r 无关.

(3) 求对数似然函数 (公式 2.3.2) 对重组率 r 的一阶和二阶导数.

$$[\ln L(r)]' \triangleq \frac{d\ln L(r)}{dr}=\sum_{i=1}^{k} n_i \frac{d(\ln f_i)}{dr}=\sum_{i=1}^{k}\frac{n_i}{f_i}(\frac{df_i}{dr}) \tag{2.3.11}$$

$$[\ln L(r)]''=\frac{d^2\ln L}{dr^2}=-\sum_{i=1}^{k}\frac{n_i}{f_i^{\,2}}(\frac{df_i}{dr})^2+\sum_{i=1}^{k}\frac{n_i}{f_i}(\frac{d^2 f_i}{dr^2}) \tag{2.3.12}$$

(4) 求解重组率 r 的极大似然估计. 有些群体中，令一阶导数 (公式 2.3.10) 等于 0 (称为似然方程)，可以直接计算出重组率，如 DH, RIL, BC1F1 等. 还有些群体，难以对似然方程直接求解，这时需采用迭代算法. 当一个函数的一阶和二阶导数有明显的表达式时，Newton 迭代算法 (也称 Newton-Raphson 算法) 是通用的求解方法. 首先选定一个重组率的起始值 $r^{(0)}$，利用下面的公式计算一个新的重组率 $r^{(1)}$，

$$r^{(1)}=r^{(0)}-\frac{[\ln L(r)]'|_{r=r^{(0)}}}{[\ln L(r)]''|_{r=r^{(0)}}} \tag{2.3.13}$$

重复这一过程，当两次迭代间重组率之差的绝对值小于事先设定的允许误差 ε 时，则停止迭代. 并把最后一次的迭代值，作为的重组率的极大似然估计值. 允许误差 ε 可取 10^{-4} 或更小的数字.

(5) 求重组率估计值的方差. 极大似然估计的方差一般从 Fisher 信息量获得. Fisher 信息量 I 等于对数似然函数二阶导数期望值的相反数，因此，

$$I=-\frac{d^2\ln L}{dr^2}|_{r=\hat{r}},\ V_{\hat{r}}=\frac{1}{I} \tag{2.3.14}$$

(6) 重组率显著性的似然比检验. 显著性检验的零假设是 H_0: r=0.5，即两个基因位点间不存在连锁关系. 备择假设是 H_0: r<0.5，即两个基因位点间存在连锁关系. 似然比统计量 LRT 定义为，备择假设和零假设两种情形下极大似然函数比值的自然对数的 2 倍. LRT 统计量在大样本的情况下，近似服从于卡方分布，卡方分布的自由度等于两种假设下独立参数个

数间的差异. 在重组率的检验中, 两种假设下独立参数个数的差异为 1. 因此,

$$LRT = -2\ln\frac{\max L(H_0)}{\max L(H_A)} = -2\ln\frac{L(r=0.5)}{L(r=\hat{r})}$$

$$= -2[\ln L(r=0.5) - \ln L(r=\hat{r})] \sim \chi^2(1) \qquad (2.3.15)$$

EM 算法也曾应用于 F_2 群体中不同类型标记间重组率的计算 (见§2.3.5). 但是, 对于 F3, BC1F2, BC2F1, BC2F2 等群体, EM 算法难以实现. Newton 迭代可作为重组率极大似然估计的通用算法, 适宜于所有群体和所有标记类型. 它的另外一个优点是迭代结束时, 可同时得到重组率的估计值, 以及重组率估计值的方差.

以表 2.3.1 中的数据为例, 图 2.3.1 给出对数似然函数曲线, 其中不包含常数项 ln*C*. 可以看出, 对数似然函数在 0.08~0.14 处有一个极大值点. 图 2.3.2 给出相应的一阶 (左) 和二阶导数 (右) 曲线. 一阶导数对较小的重组率为正值, 随着重组率增大逐渐下降, 在 0.08 和 0.14 与 *x*-轴有一个交点. 由于二阶导数为负值, 因此, 这个交点是对数似然函数的一个极大值点. 二阶导数在重组率的取值范围内一直为负值, 但随着重组率增大而逐渐上升. 表明对数似然函数在区间 (0, 0.5) 上是一个凸函数, 有唯一极大值点. 对于初始值 0.01, 经过 8 次迭代收敛到极大值点, 得到重组率的估计值为 0.1071 (表 2.3.3), 与直接计算的估计值相同. 极大值点的二阶导数值-1463.47 可用于估计重组率极大似然估计的方差和标准差, 即,

$$V_{\hat{r}} = -\frac{1}{-1463.37} = 6.83\times10^{-6},\ SE_{\hat{r}} = \sqrt{V_{\hat{r}}} = 0.0261$$

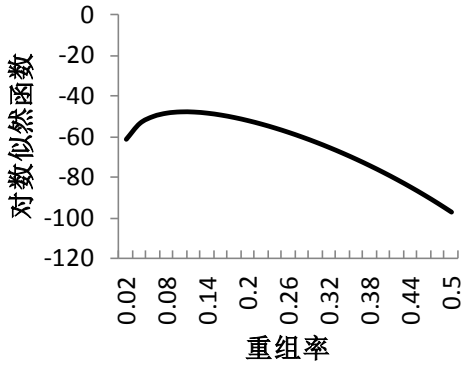

图 2.3.1 一个 DH 群体中重组率的对数似然函数曲线 (不包含常数项 ln*C*)

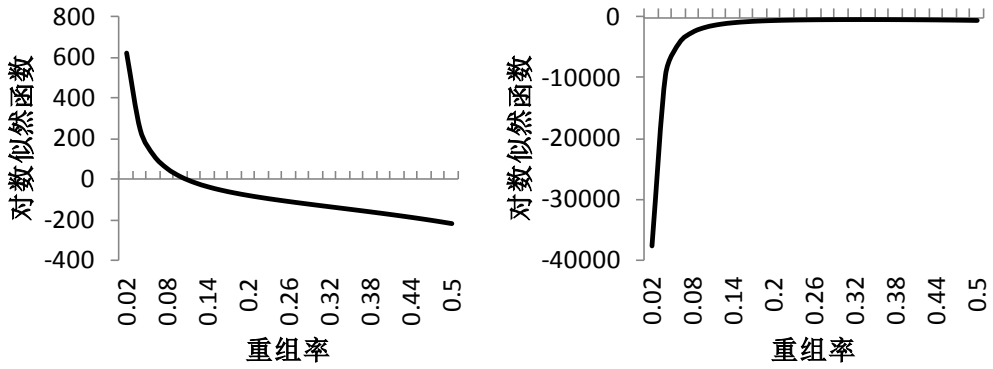

图 2.3.2 一个 DH 群体中重组率对数似然函数的一阶 (左) 和二阶导数 (右) 曲线

表 2.3.3 DH 群体中重组率估计的 Newton 迭代算法

迭代次数	1	2	3	4	5	6	7	8
重组率	0.0100	0.0196	0.0351	0.0593	0.0866	0.1035	0.1070	0.1071
$\ln L(r)$	-70.33	-61.75	-54.70	-50.01	-48.02	-47.68	-47.67	-47.67
$[\ln L(r)]'$	1373.74	655.83	297.38	119.90	36.40	5.49	0.16	0.00
$[\ln L(r)]''$	-1.50E5	-4.10E4	-1.23E4	-4401.17	-2150.79	-1555.66	-1466.11	-1463.47

§2.3.3 F_2 群体中一个共显性座位和一个显性座位间的重组率估计

以 F_2 群体中一个共显性座位和一个显性座位间的重组率估计为例. 群体中有六种可能的基因型, 每种基因型的理论频率为 f_i (i=1, 2, …, 6), 理论频率与重组率 r 的关系见表 2.2.4. 每种基因型的观察次数为 n_i (i=1, 2, …, k), 总样本量为 n. F_2 群体的似然函数为,

$$L(r) = C(1-r^2)^{n_1}(r^2)^{n_2}(1-r+r^2)^{n_3}[r(1-r)]^{n_4}[r(2-r)]^{n_5}(1-r)^{2n_6}$$

$$= C(r)^{2n_2+n_4+n_5}(1+r)^{n_1}(1-r)^{n_1+n_4+2n_6}(2-r)^{n_5}(1-r+r^2)^{n_3} \qquad (2.3.16)$$

其中 $C = \dfrac{n!}{n_1!n_2!n_3!n_4!n_5!n_6!}(\frac{1}{4})^{n_1+n_2+n_5+n_6}(\frac{1}{2})^{n_3+n_4}$, 与待估计的重组率 r 无关.

对数似然函数为,

$$\ln L(r) = \ln C + (2n_2 + n_4 + n_5)\ln(r) + n_1 \ln(1+r)$$

$$+ (n_1 + n_4 + 2n_6)\ln(1-r) + n_5 \ln(2-r) + n_3 \ln(1-r+r^2) \qquad (2.3.17)$$

对数似然函数的一阶和二阶导数为,

$$[\ln L(r)]' = \frac{d \ln L(r)}{dr} = \frac{2n_2 + n_4 + n_5}{r} + \frac{n_1}{1+r} - \frac{n_1 + n_4 + 2n_6}{1-r}$$

$$+ \frac{n_5}{2-r} - \frac{n_3(1-2r)}{1-r+r^2} \qquad (2.3.18)$$

$$[\ln L(r)]'' = \frac{d^2 \ln L(r)}{dr^2} = -\frac{2n_2 + n_4 + n_5}{r^2} - \frac{n_1}{(1+r)^2} + \frac{n_1 + n_4 + 2n_6}{(1-r)^2}$$

$$+ \frac{n_5}{(2-r)^2} + \frac{n_3(1+2r-2r^2)}{(1-r+r^2)^2} \qquad (2.3.19)$$

令一阶导数等于 0 得到的是重组率的一元六次方程, 难以直接求解, 只能采用迭代算法. 以一个抗病小麦亲本 P_1 和一个感病亲本 P_2 的 F_2 群体为例 (表 2.3.3), 抗病单株中观察到三种分子标记带型的植株数分别为 572, 1161 和 14, 感病单株中观察到三种带型的植株数分别为 3, 22 和 569. 表 1.2.1 的适合性检验表明, 抗性为单基因控制, 抗病表现为显性, 分子标记为共显性. 选取重组率的初始值 0.001, 经过 9 次迭代, 重组率收敛到 0.0179, 对数似然值到达最大值-201.11. 这时, 对数似然的一阶导数接近于 0 (表 2.3.3), 最终的二阶导数-1.29×10^5 可用于重组率估计值的方差和标准差的估计, 即,

$$V_{\hat{r}} = -\frac{1}{-1.29 \times 10^5} = 7.76 \times 10^{-6}, \quad SE_{\hat{r}} = \sqrt{V_{\hat{r}}} = 0.0028$$

表 2.3.2 抗病和感病亲本间杂交产生的小麦 F_2 群体中, 一个共显性标记和单基因抗性性状的观测个体数. 抗病相对于感病表现为显性, 标记等位基因用 A 和 a 表示, 抗病性等位基因用 B 和 b 表示

标记基因型	抗病性	
	抗病, 用 B_表示	感病, 用 bb 表示

抗病亲本标记型，用 AA 表示	n_1=572	n_2=3
杂合标记型，用 Aa 表示	n_3=1161	n_4=22
感病亲本标记型，用 aa 表示	n_5=14	n_6=569

表 2.3.3 共显性标记和显性抗病基因间重组率估计的 Newton 迭代算法

迭代次数	1	2	3	4	5	6	7	8	9
重组率	0.0010	0.0019	0.0037	0.0066	0.0108	0.0151	0.0175	0.0179	0.0179
$\ln L$	-282.75	-257.02	-234.28	-216.51	-205.69	-201.67	-201.12	-201.11	-201.11
$\frac{d(\ln L)}{dr}$	39670.85	19268.08	9081.59	4018.63	1548.44	430.81	50.88	-0.26	0.0071
$\frac{d^2(\ln L)}{dr^2}$	-4.12×10^7	-1.11×10^7	-3.10×10^6	-9.5×10^5	-3.58×10^5	-1.81×10^5	-1.35×10^5	-1.29×10^5	-1.29×10^5

§2.3.4 Newton 迭代算法中初始值的选取

Newton 迭代算法的收敛性和收敛速度与初始值的选取有关，当初始值接近真实值时，Newton 迭代算法的收敛速度很快；当初始值超过真实值太远时，Newton 迭代算法可能不收敛. 对于重组率来说，尽管不同群体中似然函数有不同的表达形式，但是对数似然函数有效取值范围在 0~0.5 之间，并且有类似于图 2.3.1 的形状，对数似然函数的一阶和二阶导数有类似于图 2.3.2 的形状. 因此，重组率有唯一的极大似然估计值. 图 2.3.3 给出 Newton 迭代算法的几何解释. 对于初始值 $r^{(0)}$，过点 $(r^{(0)},\ \ln L'(r=r^{(0)}))$ 做函数 $\ln L'(r)$ 的切线，切线与 x-轴的交点就是迭代公式 (2.3.12) 给出的更新重组率.

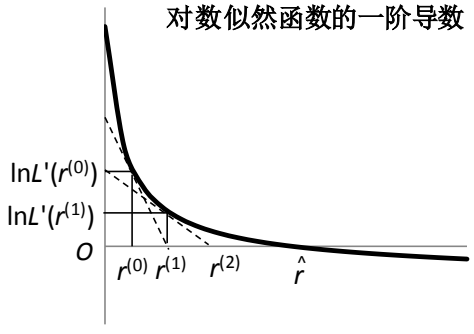

图 2.3.3 Newton 迭代计算重组率极大似然估计示意图

由图 2.3.3 可以看出，对于小于极大似然估计 $\hat{r}$ 的初始值 $r^{(0)}$，Newton 迭代能很快收敛到 $\hat{r}$. 当要计算的 $\hat{r}$ 很小时，选取较大的正数如 0.2 作为初始值时，Newton 迭代算法可能收敛

不到 $\hat{r}$. 这时应该逐渐减小初始值, 如选取 0.2 的一半, 即 0.1, 作为新的初始值进行迭代. 研究表明 (Sun et al., 2012), 选取较小的一个正数作为初始值, 如 $r^{(0)}$=0.01 或 0.001, Newton 迭代算法在绝大多情况下都能收敛到极大似然估计 $\hat{r}$, 对于较大的 $\hat{r}$, 只不过迭代次数多些而已.

§2.3.5 F_2 群体中重组率估计的 EM 算法

§2.3.2 节提过, EM 算法也可用于一些群体中重组率的计算. 现以两个共显性标记为例, 说明 F_2 群体中重组率估计的 EM 算法. 表 2.3.4 给出一个大豆 F_2 群体中, 两个共显性标记九种基因型的观测值. 第 3 列的理论频率来自表 2.2.4, 总样本量用 n 表示. 给定重组率的一个起始值, EM 算法计算群体中重组单倍型的比例, 并作为重组率的估计值. 以此作为新的重组率起始值, 重复上述过程, 直到两次迭代间重组率的差值小于预设的标准为止. 因此, EM 算法的重点是计算每种基因型中重组单倍型的概率. 产生基因型 AABB 和 aabb 的两个单倍型均为亲本型, 因此, 这两种基因型中重组单倍型的频率为 0. 产生基因型 AAbb 和 aaBB 的两个单倍型均为重组型, 因此, 这两种基因型中重组单倍型的频率为 1. 产生基因型 AABb, AaBB, Aabb 和 aaBb 的一个单倍型为亲本型，一个为重组型, 因此, 重组单倍型的频率为 $\frac{1}{2}$.

双杂合 AaBb 的情况要复杂一些, 它包含了 AB/ab 和 Ab/aB 两种可能, 它们的理论频率分别为 $\frac{1}{2}(1-r)^2$ 和 $\frac{1}{2}r^2$ (表 2.2.1), 频率之和为 $\frac{1}{2}(1-2r+2r^2)$ (表 2.2.3). 形成 AB/ab 的两个单倍型都是亲本型, 形成 Ab/aB 的两个单倍型都是重组型. 因此, 基因型 AaBb 中重组配子型的频率为 $\frac{\frac{1}{2}r^2}{\frac{1}{2}(1-2r+2r^2)}=\frac{r^2}{1-2r+r^2}$. 这样就得到表 2.3.5 第 4 列各种基因型的重组单倍型的频率.

EM 算法的基本原理如下: 给定一个重组率的初始值, 根据表 2.3.5 最后一列计算每种基因型下重组配子型的概率. 每种基因型的重组配子型概率乘以观测值即为重组体的个数, 所有重组体占总观测值的比例作为新的重组率估计值. EM 算法包含期望和极大化两个基本步骤, 即 E-步骤和 M-步骤 (Dempster et al., 1977; McLachlan, 1988).

E-步骤: 根据重组率的初始值计算各种标记基因型属于重组型的期望概率. 给定初始重组率 r, 一般可以让初始重组率 r＝0.25. 根据表 2.3.5 最后一列, 计算各种标记基因型的重

组频率 p_i, i 表示不同的标记基因型.

M-步骤: 在 E-步骤得到的各种基因型重组概率的基础上, 重新计算重组率的极大似然估计. 根据标记基因型属于重组基因型的概率重新计算重组率 r,

$$r' = \frac{1}{n} \sum_{i=1,2,\cdots,9} n_i p_i \tag{2.3.20}$$

利用公式 (2.3.20) 计算出的重组率作为新的起始值, 重复上述过程, 直到指定的精度为止. 例如, 当两次迭代间重组率差值的绝对值 $|r'-r|$ 小于 10^{-4}, 则停止迭代.

表 2.3.4 两个共显性标记间重组率估计的 EM 迭代算法

可分辨的基因型	观测值	理论频率	重组单倍型的频率
AABB	n_1=10	$f_1 = \frac{1}{4}(1-r)^2$	$p_1 = 0$
AABb	n_2=2	$f_2 = \frac{1}{2}r(1-r)$	$p_2 = \frac{1}{2}$
AAbb	n_2=1	$f_3 = \frac{1}{4}r^2$	$p_3 = 1$
AaBB	n_3=1	$f_4 = \frac{1}{2}r(1-r)$	$p_4 = \frac{1}{2}$
AaBb	n_4=21	$f_5 = \frac{1}{2}(1-2r+2r^2)$	$p_5 = \frac{r^2}{1-2r+2r^2}$
Aabb	n_4=3	$f_6 = \frac{1}{2}r(1-r)$	$p_6 = \frac{1}{2}$
aaBB	n_5=0	$f_7 = \frac{1}{4}r^2$	$p_7 = 1$
aaBb	n_6=1	$f_8 = \frac{1}{2}r(1-r)$	$p_8 = \frac{1}{2}$
aabb	n_6=17	$f_9 = \frac{1}{4}(1-r)^2$	$p_9 = 0$

对于表 2.3.4 中九种基因型的观测值, 表 2.3.5 给出 0.01, 0.25 和 0.5 三种初始值的迭代结果. 可以看出, EM 算法经过六次迭代, 得到重组率的估计值为 0.0834. 说明该算法有很快的收敛性, 收敛性和收敛到的极值点不依赖于初始值, 同时不用计算似然函数的一阶和二阶导数. F_2 群体中, 当标记不是共显性时, 也能利用 EM 算法. 但对有些群体, 如 F3, BC1F2, BC2F1, BC2F2 等, 由于存在多次减数分裂过程, E-步骤重组型的期望频率难以计算. 因此, EM 算法在有些群体中难以实现. 另外, 如果想要通过 Fisher 信息量获得重组率估计值的方差, 仍然要计算二阶导数.

表 2.3.5 三种重组率初始值下, EM 算法六次迭代的结果

重组率初始值	迭代次数					
	1	2	3	4	5	6
0.01	0.0804	0.0832	0.0834	0.0834	0.0834	0.0834
0.25	0.1179	0.0869	0.0837	0.0835	0.0834	0.0834
0.5	0.2679	0.1246	0.0878	0.0838	0.0835	0.0834

§2.3.6 奇异分离对重组率估计的影响

奇异分离一般是由于不同基因型有不同的适合度 (用 w 表示) 造成的. 假定两种基因型 AA 和 aa 各 100 个个体, AA 个体的繁殖成活率为 1, aa 个体为 0.9. 那么, 我们就说 aa 相对于 AA 的适合度 0.9. 所以, 适合度是指某基因型间能繁殖成活后代的相对能力, 其值在 0 和 1 之间. 当基因型的个数多于两个时, 繁殖成活率最高的基因型的适合度设为 1, 其他基因型的适合度为各自的繁殖成活率与最高繁殖成活率的比值. 1-w 在群体遗传学中称为选择系数, 用 s 表示. 奇异分离现象几乎存在于所有的遗传群体, 一个位点上的奇异分离会引起连锁标记或基因出现奇异, 从而导致基因型偏离孟德尔分离比, 基因型偏离孟德尔分离比会影响群体的遗传方差, 从而影响基因定位的功效. 但是, 奇异分离对重组率估计的影响却很小, 在此我们以最简单的 DH 群体为例说明这一现象.

假定 AA 和 aa 的适合度分别为 1 和 1-s, BB 和 bb 的适合度均为 1, 即 bb 相对于 BB 的选择系数为 0, 选择后的频率列于表 2.3.6 最后一列. 用 r' 表示存在奇异分离时的重组率, 仍定义为重组型基因型的比例. 那么,

$$r' = \frac{\frac{1}{2}r + \frac{1}{2}r(1-s)}{\frac{1}{2}(2-s)} = r \tag{2.3.21}$$

表 2.3.7 基因型 bb 相对于 BB 的选择系数为 s 时基因型频率的计算

基因型	无奇异分离的理论频率	选择系数	选择后的频率
AABB	$\frac{1}{2}(1-r)$	s	$\frac{1}{2}(1-r)(1-s)$
AAbb	$\frac{1}{2}r$	1	$\frac{1}{2}r$
aaBB	$\frac{1}{2}r$	s	$\frac{1}{2}r(1-s)$
aabb	$\frac{1}{2}(1-r)$	1	$\frac{1}{2}(1-r)$
总和	1		$\frac{1}{2}(2-s)$

利用表 2.3.1 中的数据, 表 2.3.8 给出基因型 aa 的不同选择系数下重组率的估计. 可以看出, 即使在 aa 的选择系数为 1 的情形下, 重组率的估计值仍然很接近无选择的情形. 因此,

在大多数情况下, 可以忽略奇异分离对重组率估计的影响. 连锁图谱构建建立在重组率的基础之上, 因此奇异分离对连锁图谱构建的影响也是可以忽略的.

表 2.3.8 基因型 aa 相对于 AA 的选择系数 *s* 取不同值时重组率的估计值

标记 Act8A	标记 OP06	*s*=0	*s*=0.5	*s*=0.75	*s*=1
AA	BB	64	64	64	64
AA	bb	8	8	8	8
aa	BB	7	4	2	0
aa	bb	61	31	15	0
重组基因型个数		15	12	10	8
观测值之和		140	107	89	72
重组率估计值		0.1071	0.1122	0.1124	0.1111

§2.4 不同遗传群体重组率估计的比较研究

利用表 2.2.1~2.2.8 中的理论频率和§2.3 中的算法, 就能利用不同的遗传群体估计两个基因座位间的重组率. 不同群体包含的重组信息不尽相同, 因此, 对重组率估计的准确度也有差异 (Sun et al., 2012). 利用 Fisher 信息量, 可以衡量不同群体包含重组率的信息量高低. 在此, 结合模拟方法, 给出不同群体中检验连锁的 LOD (likelihood of Odd) 统计量的大小, 估计重组率与真实值的离差, 重组率估计的标准差, 以及检测连锁的最小样本量等结果. 在回交群体中, 只考虑 P1 是轮回亲本的情形, 即比较 F2, F3, RIL, DH, P1BC1F1, P1BC1F2, P1BC1RIL, P1BC1DH, P1BC2F1, P1BC2F2, P1BC2RIL, P1BC2DH 这 12 种群体对重组率估计的准确度. 为了书写方便, 群体名称中略去“P1”. 根据等位基因频率, 这 12 个群体可分为三类. 在 F1-衍生群体 (包含 F2, F3, RIL 和 DH) 中, P1 等位基因的频率为 0.5. 在 BC1F1-衍生群体 (包含 BC1F1, BC1F2, BC1RIL 和 BC1DH) 中, P1 等位基因的频率为 0.75. 在 BC2F1-衍生群体 (包含 BC2F1, BC2F2, BC2RIL 和 BC2DH) 中, P1 等位基因的频率为 0.875.

§2.4.1 不同遗传群体中检验连锁的 LOD 统计量

重组率估计中的 LOD 值, 是用来检测两个标记位点间是否连锁的统计量. 通常当 LOD≥3 时, 认为两个标记位点是相互连锁的. 因此, LOD 值越大, 两个标记间连锁的可能性越大, 即检测到这两个标记连锁的可能性越大. 图 2.4.1 给出真实重组率为 0.05 和 0.20 时, 不同大小的双亲群体中 1000 次模拟得到的平均 LOD 值. 在暂时群体中, 只考虑共显性标记的情况. 可以看到, 无论哪种群体类型, LOD 值都随群体大小的增加而增大. 因此, 遗传群体

越大，越有利于连锁的检测. 从图2.4.1上和下的两种真实重组率可以看出，真实重组率越小，检验连锁的 LOD 值越高，说明采用 LOD 检验连锁是合理的.

从图 2.4.1 还可以看到，虽然各种遗传群体的 LOD 值随着重组率和群体大小的变化保持相似的变化趋势，但不同群体之间存在较大差异. 四种 F1-衍生群体 (即 F2, F3, DH, RIL) 的 LOD 值，分别高于回交一次的四种相应群体 (即 BC1F1, BC1F2, BC1DH, BC1RIL). 回交一代的四种相关群体的 LOD 值，又分别高于回交两次的四种相应群体 (即 BC2F1, BC2F2, BC2DH, BC2RIL). 因此，当群体中等位基因的频率越偏离 0.5，就越难检测到基因座位间的连锁. 回交对基因频率产生重大影响，对一个座位来说，每回交一次，非轮回亲本等位基因的频率就下降 50%. 因此，实际遗传连锁研究中很少利用回交两次以上的群体.

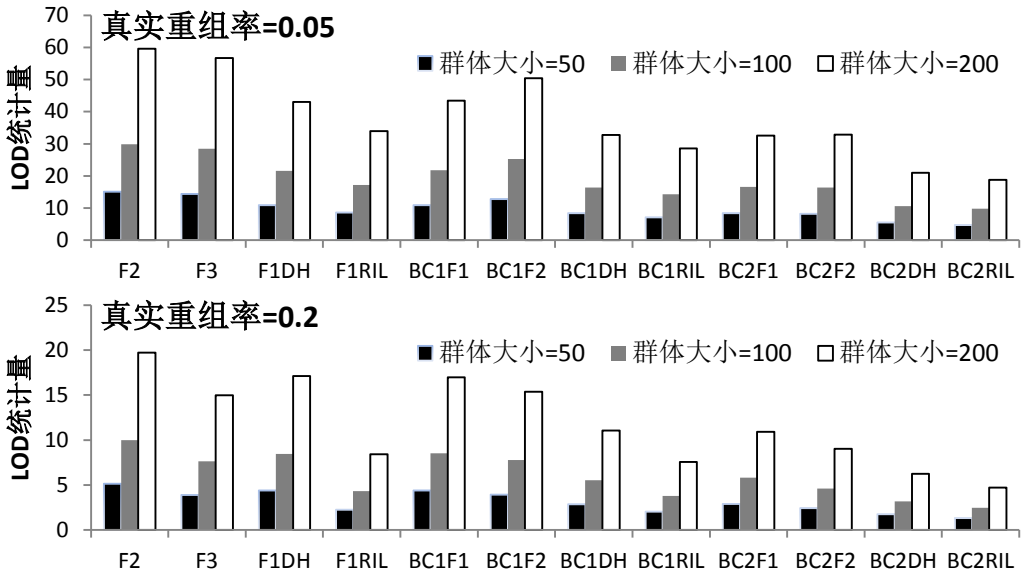

图 2.4.1 真实重组率为 0.05 (上图) 和 0.20 (下图) 时，不同大小的双亲群体中 1000 次模拟得到的平均 *LOD* 值. 只考虑共显性标记的情况.

等位基因频率相等的群体，也不一定意味着它们有相同的检测连锁的功效. 对于基因频率为 0.5 的四种群体来说中，按照 LOD 值从大到小的顺序是 F2, F3, DH, RIL. 对共显性标记来说, F2 和 F3 群体有九种可以识别的基因型. 因此，提供了最多的关于重组和交换的信息, LOD 统计量也最高. DH 和 RIL 群体仅包含四种可识别的基因型，在同样的群体大小下，它们的 LOD 值低于 F2 和 F3 群体. F2 和 F3 间的差异是由于基因型频率的不同引起的，DH 和 RIL 群体间的差异同样也由于基因型频率的不同造成的. 与 F2 相比较, F3 在自交过程中，多了一次重组的机会. 因此, F3 中的累积重组率要大于 F2 的重组率，这意味着两个连锁座位之间的关联或连锁不平衡程度降低了. 因此，引起 LOD 值的下降. 与此类似，DH 群体中只经

历一次重组，RIL 群体在重复自交的过程中有多次重组机会，累积重组率 *R* 大于一次交换的重组率 *r*. 因此, DH 群体的 LOD 值高于 RIL 群体.

对于轮回亲本等位基因频率为 0.75 的四种群体来说中, LOD 值按照从大到小的顺序是 BC1F2, BC1F1, BC1DH, BC1RIL. BC1F2 比 BC1F1 有较高的 LOD 值可根据可识别基因型的个数得以解释. BC1F1 只包含四种基因型，自交之后的 BC1F2 与 F2 类似，包含九种可识别的基因型. 因此，提供了更多的重组和交换信息. 与 DH 和 RIL 类似，BC1DH 和 BC1RIL 群体中 LOD 值的降低，可以从群体中包含了较少的基因型个数得以解释. 对于轮回亲本等位基因频率为 0.875 的四种群体来说中，LOD 值按照从大到小的顺序也是 BC2F2，BC2F1, BC2DH, BC2RIL，只不过群体间的差异变得更小些.

§2.4.2 不同遗传群体中重组率估计的准确度

不同遗传群体中，重组率估计的准确度与图 2.4.1 中观察到的 LOD 有相似的结果. 表现为 LOD 值越大，重组率估计得越精确，重组率估计值的方差和标准差也越小 (图 2.4.2). 真实重组率为 0.3 时，图 2.4.2 上给出三种大小的群体中 1000 次模拟得到的重组率估计值与真实值的离差，图 2.4.2 下给出重组率估计值的标准差. 在暂时群体中，只考虑共显性标记的情况. 可以看到，无论哪种群体类型，估计值与真实值的离差，以及估计值的标准差随群体大小的增加而减小. 因此，遗传群体越大，重组率估计得越准确. 回交次数越多，离差和标准差也越大.

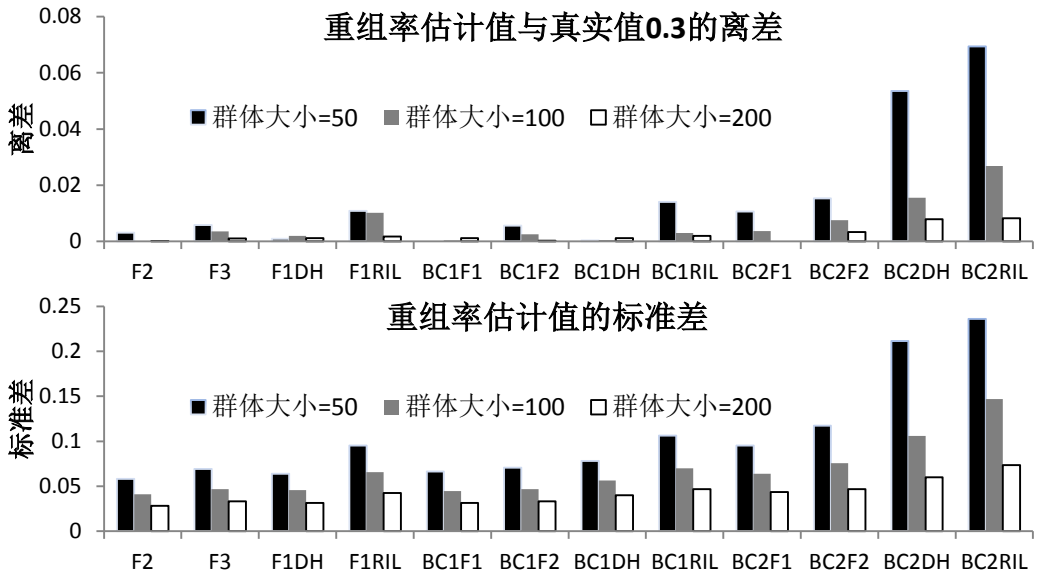

图 2.4.2 真实重组率为 0.3 时，不同大小的双亲群体中 1000 次模拟得到的重组率估计值与真实值的离差 (上图) 以及重组率估计值的标准差 (下图). 只考虑共显性标记的情况.

图 2.4.1 和图 2.4.2 中的暂时群体，我们假定两个标记表现为共显性. 显然，这些结果不能简单推及其他五种标记类型 (表 2.2.4~表 2.2.8). 如果两标记均为显性，这时 F2 群体中可识别的标记只有四种，重组率估计的准确性不一定高于同样大小的 DH 或 RIL 群体. 回交群体还会出现重组率无法估计的情况. 如果重组率无法估计，就更无法谈及重组率估计的准确性. 非共显性标记的情形，还将在§2.4.3 中考虑.

§2.4.3 不同遗传群体检测到显著连锁所需的样本量

检测两个标记是否连锁，要足够大的群体来保证: (1) 连锁紧密时至少出现一个重组体; (2) 连锁不太紧密时两标记间的连锁 LOD 值不小于 3. 这样，对于紧密的连锁，不会把它估计为 0; 对于不太紧密的连锁，也不会把它判断为独立遗传，即重组率与 0.5 无显著差异. 为方便读者，我们把不同真实重组率在 95%概率水平下，至少出现一个重组体所需的最低群体大小列于表 2.4.1，保证 LOD 统计量不小于 3 的最低群体大小列于表 2.4.2. 两个标记连锁越紧密，LOD 值越高，检测标记间连锁越容易. 但连锁越紧密，发生重组的可能性越小，重组率越容易被估计为 0. 因此，需要较大的群体，才能保证出现至少一个重组体 (表 2.4.1). 两个标记连锁越松散，LOD 值越低，检测标记间连锁越困难. 因此，也需要较大的群体，才能保证不把松散的连锁判断为独立遗传 (表 2.4.2). 实际研究中，应该选取表 2.4.1 和表 2.4.2 相应位置上的两个数字中的较大值作为最低群体大小的判断依据.

标记的显隐性不影响永久群体中重组率的估计，但对暂时群体却有较大影响. 表 2.4.1 和表 2.4.2 也给出非共显性标记检测连锁的最低群体大小. 以 F_2 群体和真实重组率 0.01 为例，如两个标记为共显性，平均每 150 个 F_2 个体中，就能观测到重组体的出现. 如一个标记仍为共显性，另一个标记为显性，则需要近 300 个 F_2 个体中，才能观测到重组体的出现. 如一个标记为显性，另一个标记为隐性，则需要在数十万个 F_2 个体中，才能观测到重组体的出现. 因此，显性标记和隐性标记可能代表了重组率估计最差的情形. 这从另一个角度说明，如果选择暂时群体开展遗传研究，要尽可能采用共显性标记开展基因型的鉴定工作.

表 2.4.1 在 95%概率水平下至少出现一个重组体所需的群体大小

群体*	*r*=0.01	*r*=0.02	*r*=0.03	*r*=0.05	*r*=0.1	*r*=0.2	*r*=0.3
F_2 (C, C)	150	75	50	30	15	8	5

F_2 (C, D)	299	149	99	60	31	16	11
F_2 (C, R)	299	149	99	60	31	16	11
F_2 (D, D)	299	149	99	61	31	16	11
F_2 (D, R)	149786	29956	13616	4754	1197	299	132
F_2 (R, R)	299	149	99	61	31	16	11
F_3 (C, C)	121	61	41	25	13	7	5
F_3 (C, D)	199	99	67	41	21	11	8
F_3 (C, R)	199	99	67	41	21	11	8
F_3 (D, D)	213	99	67	41	21	11	8
F_3 (D, R)	998	598	373	229	110	52	34
F_3 (R, R)	213	99	67	41	21	11	8
DH	299	149	99	59	29	14	9
RIL	152	77	52	32	17	9	7
BC_1F_1 (C, C)	299	149	99	59	29	14	9
BC_1F_1 (C, R)	299	149	99	59	29	14	9
BC_1F_1 (R, R)	299	149	99	59	29	14	9
BC_1F_2 (C, C)	172	86	58	35	18	9	7
BC_1F_2 (C, D)	427	199	135	82	43	24	17
BC_1F_2 (C, R)	249	119	80	48	24	12	8
BC_1F_2 (D, D)	373	213	135	82	43	24	17
BC_1F_2 (D, R)	2995	998	748	427	213	102	66
BC_1F_2 (R, R)	249	124	82	49	24	12	8
BC_1DH	300	150	100	60	31	16	11
BC_1RIL	203	103	70	43	23	13	10
BC_2F_1 (C, C)	300	150	100	60	31	16	11
BC_2F_1 (C, R)	300	150	100	60	31	16	11
BC_2F_1 (R, R)	300	150	100	60	31	16	11
BC_2F_2 (C, C)	242	122	82	50	27	15	11
BC_2F_2 (C, D)	748	332	213	129	70	39	31
BC_2F_2 (C, R)	299	157	99	61	32	17	12
BC_2F_2 (D, D)	748	299	213	124	70	39	31
BC_2F_2 (D, R)	2995	1497	748	498	249	124	85
BC_2F_2 (R, R)	299	149	99	61	32	17	12
BC_2DH	403	203	136	83	43	24	17
BC_2RIL	305	156	106	66	36	21	16

*群体名称后给出两个标记座位 A/a 和 B/b 的显隐性. (C, C) 表示等位基因 A 和 a 为共显性, B 和 b 也为共显性; (C, D) 表示等位基因 A 和 a 为共显性, B 对 b 为显性; (C, R) 表示等位基因 A 和 a 为共显性, B 对 b 为隐性; (D, D) 表示等位基因 A 对 a 为显性, B 对 b 为显性; (D, R) 表示等位基因 A 对 a 为显性, B 对 b 为隐性; (R, R) 表示等位基因 A 对 a 为隐性, B 对 b 为隐性.

表 2.4.2 在 95%概率水平下, 检验连锁的 *LOD* 统计量≥3 所需的群体大小

群体*	*r*=0.01	*r*=0.02	*r*=0.03	*r*=0.05	*r*=0.1	*r*=0.2	*r*=0.3
F_2 (C, C)	8	9	9	11	15	31	78
F_2 (C, D)	14	15	16	19	26	51	123

F_2 (C, R)	14	15	16	19	26	51	123
F_2 (D, D)	14	15	16	19	27	56	147
F_2 (D, R)	82	83	83	86	96	138	262
F_2 (R, R)	14	15	16	19	27	56	147
F_3 (C, C)	8	9	9	11	17	41	121
F_3 (C, D)	12	13	15	17	26	58	162
F_3 (C, R)	12	13	15	17	26	58	162
F_3 (D, D)	12	14	15	17	26	62	179
F_3 (D, R)	31	33	35	39	52	100	246
F_3 (R, R)	12	14	15	17	26	62	179
DH	11	12	13	14	19	36	84
RIL	12	14	15	18	29	73	219
BC_1F_1 (C, C)	11	12	13	14	19	36	84
BC_1F_1 (C, R)	11	12	13	14	19	36	84
BC_1F_1 (R, R)	11	12	13	14	19	36	84
BC_1F_2 (C, C)	9	10	11	12	18	40	107
BC_1F_2 (C, D)	21	23	25	29	42	90	236
BC_1F_2 (C, R)	12	13	14	16	23	49	125
BC_1F_2 (D, D)	21	23	25	29	44	101	289
BC_1F_2 (D, R)	54	57	59	65	84	150	343
BC_1F_2 (R, R)	12	13	14	16	23	49	128
BC_1DH	14	15	16	19	27	56	147
BC_1RIL	15	16	18	22	34	83	238
BC_2F_1 (C, C)	14	15	16	19	27	56	147
BC_2F_1 (C, R)	14	15	16	19	27	56	147
BC_2F_1 (R, R)	14	15	16	19	27	56	147
BC_2F_2 (C, C)	13	15	16	19	29	68	199
BC_2F_2 (C, D)	34	37	41	48	72	166	469
BC_2F_2 (C, R)	17	18	20	23	34	78	218
BC_2F_2 (D, D)	34	38	41	49	76	193	606
BC_2F_2 (D, R)	66	70	75	84	114	229	585
BC_2F_2 (R, R)	17	18	20	23	34	79	220
BC_2DH	21	23	25	29	44	101	289
BC_2RIL	22	24	27	33	52	133	406

*群体名称后给出两个标记位点 A/a 和 B/b 的显隐性. (C, C) 表示等位基因 A 和 a 为共显性, B 和 b 也为共显性; (C, D) 表示等位基因 A 和 a 为共显性, B 对 b 为显性; (C, R) 表示等位基因 A 和 a 为共显性, B 对 b 为隐性; (D, D) 表示等位基因 A 对 a 为显性, B 对 b 为显性; (D, R) 表示等位基因 A 对 a 为显性, B 对 b 为隐性; (R, R) 表示等位基因 A 对 a 为隐性, B 对 b 为隐性.

§2.5 作图函数和遗传图谱构建

连锁图谱是指基因或标记在染色体上的相对位置与遗传距离. 通过连锁图谱可以大致了解基因和标记之间的相对位置, 了解哪些基因更靠近着丝粒, 哪些更靠近端粒等. 连

锁图谱的构建是很多遗传研究的基础，使用的标记越多，遗传连锁图谱的分辨率就越高. 但是标记数目增加之后，也给标记的分群和排序带来难度. 因此，高密度连锁图谱的构建方法也一直是遗传学研究的一个热点问题 (Haldane, 1919; Buetow and Chakravarti, 1987; Lander and Green, 1987; Lander et al., 1987; Weeks and Lange, 1987; Hackett and Broadfoot, 2003; Mester et al., 2003. van Os et al., 2005; Mollinari et al., 2009).

§2.5.1 遗传干涉和干涉系数

对于三个连锁的基因座 M_1, M_2 和 M_3，根据§2.3 的内容，可以估计三个成对座位间的重组率. 用 r_{12}, r_{23} 和 r_{13} 表示标记区间 M_1-M_2, M_2-M_3 和 M_1-M_3 上的重组率. 根据这三个重组率的估计值，就能够判断这三个基因座在染色体上的相对位置. 例如，如果 r_{13} 的估计值大于 r_{12} 和 r_{23}，三个基因座排列顺序可能为 M_1-M_2-M_3. 假定连锁图上三个座位的排列顺序为 M_1-M_2-M_3，标记区间 M_1-M_2 和 M_2-M_3 上不存在干涉时，即交换独立发生，三个重组率的关系为，

$$(1-r_{13})=(1-r_{12})(1-r_{23})+r_{12}r_{23} \quad \text{或}$$

$$r_{13}=r_{12}(1-r_{23})+(1-r_{12})r_{23}=r_{12}+r_{23}-2r_{12}r_{23} \tag{2.5.1}$$

对于完全干涉，即区间 M_1-M_2 (或 M_2-M_3) 上的交换将完全阻止区间 M_2-M_3 (或 M_1-M_2) 上交换的发生，这时，

$$r_{13}=r_{12}+r_{23} \tag{2.5.2}$$

一般情况下，用 δ 表示干涉系数，则有，

$$r_{13}=r_{12}+r_{23}-2(1-\delta)r_{12}r_{23} \tag{2.5.3}$$

容易看出，当 δ=0 时，等式 (2.5.3) 与 (2.5.1) 相同. 因此，δ=0 表示两个区间上的交换是独立的. 当 δ=1 时，等式 (2.5.3) 与 (2.5.2) 相同. 因此，δ=1 表示两个区间上的交换是完全干涉. 这时，一个区间上的交换完全阻止另外一个区间上的交换的发生. 如果三个连锁的基因座的顺序为 M_1-M_2-M_3，干涉系数可利用三个重组率进行估计，即，

$$\delta = 1 - \frac{r_{12} + r_{23} - r_{13}}{2r_{12}r_{23}} \tag{2.5.4}$$

表 2.5.1 列出图 1.2.4 的大麦 DH 群体中, 1H 染色体上 14 个标记的成对重组率估计值. 以前三个标记为例, Act8A 与 OP06 之间重组率的估计值为 0.107, OP06 与 aHor2 之间为 0.076, Act8A 与 aHor2 之间为 0.111. 从这三个估计值可以看出标记 OP06 应该排序在 Act8A 与 aHor2 之间. 根据公式 (2.5.4) 得到干涉系数 $\delta = -3.422$, 说明区间 Act8A - OP06 与区间 Act8A - aHor2 可能存在负干涉, 即双交换的频率大于无干涉时的频率 $r_{12}r_{23}$. 再以第 5-7 个标记为例, ABG464 与 Dor3 之间重组率的估计值为 0.184, Dor3 与 iPgd2 之间为 0.036, ABG464 与 iPgd2 之间为 0.214. 从这三个估计值可以看出标记 Dor3 应该排序在 ABG464 与 iPgd2 之间. 根据公式 (2.5.4) 得到干涉系数 $\delta = 0.617$, 说明区间 ABG464 - Dor3 与区间 Dor3 - iPgd2 可能存在正干涉, 即双交换的频率小于无干涉时的频率 $r_{12}r_{23}$.

表 2.5.1 大麦 DH 群体中 1H 染色体上 14 个标记的成对重组率估计值

	Act8A	OP06	aHor2	MWG943	ABG464	Dor3	iPgd2	cMWG733A	AtpbA	drun8	ABC261	ABG710B	Aga7
OP06	0.107												
aHor2	0.111	0.076											
MWG943	0.419	0.429	0.419										
ABG464	0.475	0.485	0.458	0.128									
Dor3	0.457	0.460	0.459	0.308	0.184								
iPgd2	0.438	0.468	0.419	0.321	0.214	0.036							
cMWG733A	0.451	0.482	0.448	0.370	0.283	0.101	0.070						
AtpbA	0.437	0.482	0.455	0.390	0.304	0.122	0.105	0.036					
drun8	0.500	0.532	0.529	0.467	0.436	0.262	0.241	0.175	0.133				
ABC261	0.483	0.507	0.511	0.441	0.410	0.236	0.222	0.155	0.113	0.049			
ABG710B	0.493	0.525	0.530	0.496	0.475	0.317	0.294	0.227	0.184	0.105	0.070		
Aga7	0.479	0.504	0.515	0.504	0.500	0.355	0.331	0.266	0.224	0.145	0.111	0.035	
MWG912	0.464	0.489	0.481	0.504	0.529	0.400	0.376	0.317	0.273	0.192	0.171	0.094	0.057

§2.5.2 作图函数

由于遗传干涉的存在, 重组率一般不满足可加性. 而距离一般是可加的, 对于遗传图谱来说, 希望图谱上的距离也满足可加性. 设连锁图上有排列顺序为 M_1-M_2-M_3 的 3 个座位, M_1 与 M_3 之间的图距用 m_{13} 表示, M_1 与 M_2 之间的图距用 m_{12} 表示, M_2 与 M_3 之间的图距用 m_{23} 表示. 根据距离的可加性,

$$m_{13} = m_{12} + m_{23} \tag{2.5.5}$$

公式式 (2.5.5) 中 m 是两个位点间的遗传距离, 称为图距. 图矩的单位为摩尔根 (用 M 表示)或厘摩 (用 cM 表示), 1M=100cM. 图距 m 是交换率 r 的函数, 即 $m = f(r)$, 称 f 为作图函数. 交换率 r=0.01 的两个位点间的图距大约为 1cM. 在连锁作图研究中, 有不同的作图函数, 可以把重组率转换为图距, 这里介绍常用的三种作图函数.

(1) Morgan 作图函数. 由 Morgan 在 1928 年和 Sturtevant (1931) 提出, 它将重组率的百分数作为图距, 即 m=100×r, 单位为 cM. 对于紧邻的两个区间, 可以采用求和的办法计算图距. 例如顺序排列的 3 个位点 M_1-M_2-M_3, M_1-M_2 间的重组率为 0.02, 即图距为 2cM; M_2-M_3 间的重组率为 0.01, 即图距为 1cM. 根据 Morgan 作图函数, M_1-M_3 间的图距为 3cM. Morgan 作图函数没有考虑大标记区间中存在多重交换的可能, 且假定干涉系数 $\delta = 1$. 事实上, 一个较长的染色体区间上可能存在双交换甚至多次交换, 使得重组率不具有线性可加性的. 因此, Morgan 作图函数不能应用于比较长的染色体区段.

(2) Haldane 作图函数. 对于顺序排列的 3 个位点 M_1-M_2-M_3, 在没有干涉的情况下, 即假定 M_1-M_2 间的交换和 M_2-M_3 间的交换独立发生, 并考虑到一个区间可以发生多次交换, Haldane (1919) 给出下面的作图函数,

$$m = f(r) = -\tfrac{1}{2}\ln(1-2r), \text{ 或 } r = \tfrac{1}{2}(1-e^{-2m}) \tag{2.5.6}$$

其中, m 的单位为 M. 实际中, m 常用 cM 为单位, 这时,

$$m = f(r) = -50\ln(1-2r), \text{ 或 } r = \tfrac{1}{2}(1-e^{-m/50}) \tag{2.5.7}$$

(3) Kosambi 作图函数. Kosambi (1944) 考虑到遗传干涉的存在，提出干涉系数应是重组率的函数. 即, 染色体区间越短, 干涉的程度越大; 染色体区间越长, 干涉系数越小. 由此建立的作图函数为,

$$m = \frac{1}{4}\ln\frac{1+2r}{1-2r}, \text{ 或 } r = \frac{1}{2}\frac{e^{4m}-1}{e^{4m}+1} \tag{2.5.8}$$

其中, m 的单位为 M. 当 m 以 cM 为单位时,

$$m = 25\ln\frac{1+2r}{1-2r}, \text{ 或 } r = \frac{1}{2}\frac{e^{m/25}-1}{e^{m/25}+1} \tag{2.5.9}$$

上述三种作图函数，Haldane 和 Kosambi 作图函数用得较多. 对于给定的重组率，Haldane 作图函数给出的图距最大，Morgan 函数给出的图距最小 (图 2.5.1). 当重组率 $r < 0.1$ 时，三种作图函数得到非常相近的图距.

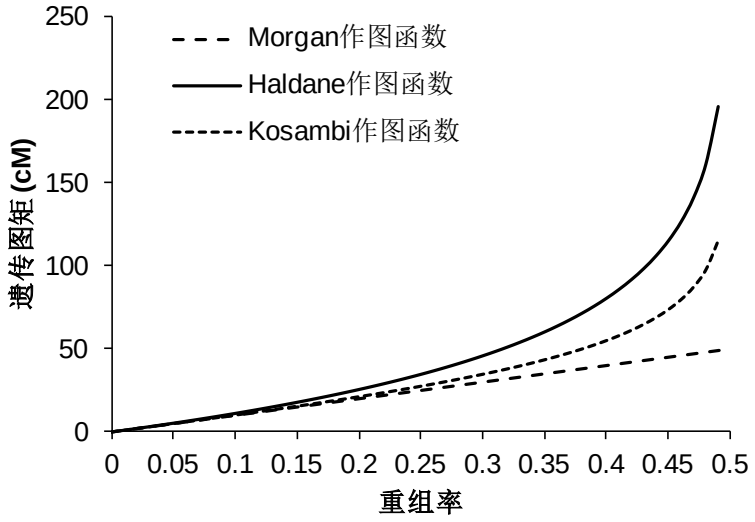

图 2.5.1 三种作图函数的比较

§2.5.3 标记分群算法

建立连锁图谱的第一步是将来自不同染色体的标记进行分群. 理想的情况是，有多少条染色体，就把标记分成多少个群，一个标记群代表一条染色体. 分群时采用的标准可以是检测连锁的 LOD 统计量，也可以是重组率的估计值，还可以是根据重组率转换成的图距. 现以 LOD 分群标准为例，说明分群的过程. 设定一个 LOD 临界值，n 个待分群标记用集合的形式表示为 $G_0=\{M_1, M_2, \ldots, M_n\}$. 分群后标记用 k 个非空集合 $G_1, \ldots, G_k$ 表示. 分 $k=0$ 和 $k>0$ 两种情形讨论.

情形 1: $k=0$，即当前没有任何标记群.

(1.1) 在 G_0 中，确定一对优先分群标记 M_{j1} 和 M_{j2} (即连锁最紧密的两个标记)，满足:

$$D_{j_1 j_2} = \mathrm{Max}\{LOD(M_{i_1}, M_{i_2}); i_1, i_2 = 1,2,\cdots,n, i_1 \neq i_2\} \tag{2.5.10}$$

(1.2) 如果 $D_{j_1 j_2}$ 大于指定的 LOD 临界值，则生成第一个群 G_1，将 M_{j1} 和 M_{j2} 分入 G_1 中；否则，生成 2 个群 G_1 和 G_2，将 M_{j1} 和 M_{j2} 分别分入 G_1 和 G_2 中.

(1.3) 将 M_{j1} 和 M_{j2} 从 G_0 中删除.

情形 2: $k>0$，即已经产生一些标记群，适用于有锚定标记的分群.

(2.1) 在 G_0 中，确定一个优先分群标记 M_j，方法如下：对 G_0 中的任意 $M_{j'}$，计算

$$C_{j'} = \mathrm{Max}\{LOD(\mathrm{M}_{j'}, \mathrm{G}_{xy}); x = 1,2,\cdots k, y = 1,2,\cdots n_x\} \tag{2.5.11}$$

其中，G_{xy} 表示第 x 个标记群 G_x 中的第 y 个标记；优先分群标记 M_j 是具有最大 $C_{j'}$ 的标记，即 $C_j = \mathrm{Max}\{C_{j'}; j' = 1,2,\cdots,n\}$.

(2.2) 确定 M_j 优先分进的群 G_i，方法如下：对任意 $G_{i'}$，计算

$$D_{i'} = \mathrm{Max}\{LOD(\mathrm{M}_j, \mathrm{G}_{i'y}); y = 1,2,\cdots n_{i'}\} \tag{2.5.12}$$

优先分进的群 G_i 是具有最大 $D_{i'}$ 的群，即 $D_i = \mathrm{Max}\{C_{i'}; i' = 1,2,\cdots,k\}$.

(2.3) 确定 M_j 是否应该分入 G_i 中，方法如下：如果 D_i>指定的 LOD 临界值，则把 M_j 分进 G_i 中；否则，生成 1 个新群 G_{k+1}，并把 M_j 分进新群 G_{k+1} 中.

(2.4) 将 M_j 从 G_0 中删除.

(2.5) 如果 G_0=ø，则分群完成；否则重复上述过程.

最后得到的 G_1, G_2, … 就是对这 n 个标记的分群. 如选择重组率或图距作为分群标准，公式 (2.5.10) (2.5.11) 和 (2.5.12) 中的最大化改为最小化，判断标准改为小于即可.

以图 1.2.4 大麦 DH 群体中 1H 染色体上 14 个标记为例，表 2.5.2 上三角给出检验连锁关系的成对 LOD 值，下三角为成对标记间的图距. 图距根据表 2.5.1 的重组率转换而来，利用 Haldane 作图函数. 为整齐起见，并避免标记分群和排序时出现缺失数据，当重组率的估计值等于或大于 0.5 时，图距用 1000 表示. 设 LOD 临界值为 3. 在没有任何锚定标记时，这

14 个标记分为两个群，前三个在一个群，后面的 11 个在另外一个群. 如认为第 1 和 14 个标记在同一个群内，即把第 1 和 14 个标记锚定在一个群内，则这 14 个标记在 LOD 临界值为 3 时就分成一个群.

表 2.5.2 大麦 DH 群体中 1H 染色体上 14 个标记的成对 LOD 值 (上三角) 和成对图距 (下三角). 下三角中的图距根据表 2.5.1 重组率利用 Haldane 作图函数转换而来. 当重组率的估计值等于或大于 0.5 时，图距用 1000 表示

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
M1		21.4	20.2	0.8	0.1	0.2	0.5	0.3	0.5	0.0	0.0	0.0	0.1	0.2
M2	10.9		24.4	0.6	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M3	11.3	7.6		0.7	0.2	0.2	0.8	0.3	0.2	0.0	0.0	0.0	0.0	0.0
M4	60.8	64.1	60.6		18.0	4.4	3.9	2.0	1.4	0.1	0.4	0.0	0.0	0.0
M5	91.4	105.1	78.2	13.1		12.8	10.6	5.9	4.7	0.5	1.0	0.1	0.0	0.0
M6	77.7	79.4	78.6	36.0	19.3		33.1	22.1	19.4	7.2	8.9	4.2	2.6	1.2
M7	67.7	85.3	60.8	38.1	22.9	3.6		27.3	22.2	8.8	10.2	5.4	3.7	1.9
M8	74.0	100.0	72.5	47.6	32.0	10.2	7.0		33.1	14.3	16.2	9.6	7.1	4.2
M9	67.3	100.0	76.5	52.2	35.3	12.5	10.6	3.6		18.7	21.0	13.2	10.0	6.4
M10	1000	1000	1000	84.6	66.9	29.1	26.3	18.3	13.6		31.2	22.2	17.6	12.5
M11	100.7	1000	1000	69.3	57.9	25.6	23.9	16.0	11.5	4.9		27.0	21.5	14.3
M12	123.7	1000	1000	139.9	91.4	37.3	33.7	24.5	19.4	10.6	7.1		33.6	23.1
M13	96.3	1000	1000	1000	1000	44.3	39.8	29.6	24.1	14.9	11.3	3.5		29.1
M14	82.4	112.6	98.9	1000	1000	54.9	48.9	37.3	30.7	20.2	17.9	9.5	5.7	

§2.5.4 标记排序算法

已知 n 个城市之间的直线距离，有一个旅行商需要遍访这 n 个城市，并且每个城市只能访问一次，最后返回出发城市，这就是组合数学中的旅行商问题 (traveling salesman problem, TSP). 求解 TSP 问题，就是要选择一条路程最短的旅行路线. 数学上已经证明, TSP 是运筹学，图论和组合优化中的一个 NP 难题 (non-deterministic poly-nominal time hard, NP-hard). 当城市数较大时，不存在全局最优解的精确算法，所有的解都是近似最优解. 但可喜的是，目前已研究出多种有效求解 TSP 问题的近似算法 (Lin and Kernighan, 1973; Laporte, 1992).

连锁图谱构建过程中，排序的目的是寻求图距最短的一个标记顺序. 当一个群中有 n 个标记时，所有可能的排序有 $\frac{1}{2}n!$ 种. 当 n=50 时，$\frac{1}{2}n! = 1.52 \times 10^{64}$. 因此，要比较所有可能的顺序几乎是不可能的. 高通量分子标记可以构建超高密度遗传连锁图谱，但同时对连锁图谱构建算法提出巨大的挑战. 一些传统的方法如顺序排列法 (Buetow and Chakravarti,

1987),重组计数排序法 (van Os et al., 2005), 单向生长算法 (Tan and Fu, 2006) 等, 存在时间复杂度过高, 排序准确度差等问题. 连锁图谱构建与 TSP 问题求解之间存在极大的相似性. 成对标记间的重组率或图距可看作 TSP 问题中两两城市间的路程. 但两者之间又有一定区别, 遗传距离的估计受群体类型, 群体大小, 标记缺失等诸多因素的影响, 估计值有一定误差. 而 TSP 中的物理距离一般没有误差, 或者误差很小. 标记排序的近似算法已有很多, 这里着重介绍基于旅行商问题的近似算法.

步骤 1: 构造一个起始序列

构造算法也有很多, 这里介绍最近邻居算法, 也称为贪婪算法. 算法从距离最短的两个标记开始, 然后在待排标记中, 依次加入与已排顺序具有最短距离的标记. 实际计算中, 可以从任意一个标记开始, 构造出不同的顺序, 然后选择具有最短距离的一个, 作为起始序列. 假定群 G 中有 *n* 个标记, 用集合的形式表示为 G={M_1, M_2, …, M_n}表示. 构造算法如下:

(1.1) 对于 G 中的任一标记 M_i, 将 M_i 作为起始序列的起始标记, 同时 M_i 也是序列的终止标记, 并将 M_i 从中 G 中删除, 删除后的标记群用 G_0 表示.

(1.2) 在 G_0 中寻找与已有序列的终止标记具有最短距离的标记 M_j, 作为新的终止标记, 同时并将 M_j 从中 G_0 中删除.

(1.3) 如果 G_0=ø, 则从 (1.1) 循环, 直到 G 中最后一个标记; 否则从 (1.2) 循环.

(1.4) 从上述的 *n* 个序列中, 选择最短的一个.

步骤 2: 序列改进的 Two-opt 算法

把步骤 1 构造出的序列首尾相连, 形成一个类似 TSP 问题的回路. 将回路从任意两个位置上断开, 颠换后对接, 如颠换对接后有更短的图距, 则把对接后的回路作为新的回路继续改进, 直到回路不再变短为止. 假定从标记 X 与 X+1, 以及 Y 与 Y+1 之间, 将回路断为两段. 将标记 X 与 Y+1 对接, X+1 与 Y 对接, 形成一个新的回路 (图 2.5.2). 如果新的回路与之前的回路相比有较短的路程, 则在新回路的基础上重复 Two-opt 改进算法. 如果新的回路与之前的回路相比没有较短的路程, 则在旧回路的其他位置上重复 Two-opt 改进算法. 对最终得到的回路, 把首尾相连的两个标记重新分开, 或从最长的区间上将回路断开, 就得到我们

想要的连锁图.

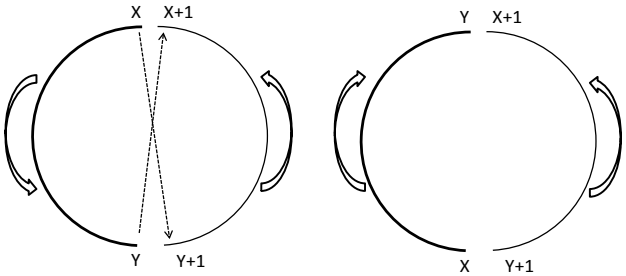

图 2.5.2 Two-opt 改进算法示意图. 左图为交换前的回路, 右图为交换后的回路.

步骤 3: 序列的进一步改进

对于只包含数十个标记的连锁群, 通过步骤 1 和 2 一般就能得到最优的标记顺序. 当标记更多时, 步骤 1 和 2 得到的标记顺序还有进一步改进的必要. 具体过程如下:

(3.1) 选定一个窗口大小 w, w 一般在 5~10 个标记之间. 窗口太小, 改进效果不明显; 窗口太大, 则耗时较长. 假定一个连锁群体上有 n 个标记, 一般认为 $n>>w$.

(3.2) 对 i=1, 2, …, n-w 做循环. 对 w 个标记 M_i, M_{i+1}, …, M_{i+w} 的所有 $w!$ 个可能排列中, 寻找最短的一种排列顺序. 例如 w=5 时, $w!$=120; w=8 时, $w!$=40320.

利用表 2.5.2 中的成对距离矩阵, 对大麦 DH 群体中 1H 染色体上 14 个标记进行排序. 得到的结果见表 2.5.3, 平均间距 12.34cM, 标记 aHor2 与 MWG943 有最大的间距, 达 60.59cM. 一些区间上存在明显的干涉.

表 2.5.3 大麦 DH 群体中 1H 染色体上 14 个标记的排序

标记编号	标记名称	与下一个标记的间距 (cM)	染色体上位置 (cM)	两个相邻区间上的干涉系数
1	Act8A	10.88	0	0
2	OP06	7.64	10.88	0
3	aHor2	60.59	18.52	0.1738
4	MWG943	13.07	79.11	0.9282
5	ABG464	19.28	92.18	0.6166
6	Dor3	3.55	111.46	0.0581
7	iPgd2	7.04	115.01	0.9
8	cMWG733A	3.56	122.05	1
9	AtpbA	13.61	125.61	0
10	drun8	4.88	139.22	0

11	ABC261	7.09	144.10	1
12	ABG710B	3.50	151.19	1
13	Aga7	5.70	154.69	
14	MWG912		160.39	

§2.6 随机交配群体的连锁分析

§2.6.1 随机交配与连锁不平衡

对一个基因座位来说，随机交配一代后，群体就达到Hardy-Weinberg平衡．处于Hardy-Weinberg平衡的群体，基因型的频率可以从基因频率推出．用f_A和f_a表示一个座位上两个等位基因A和a的频率，则三种基因型AA，Aa和aa的频率对应于二项式 $(f_A+f_a)^2$ 的展开项．对于另外一个座位，用f_B和f_b表示两个等位基因B和b的频率，随机交配群体中三种基因型BB，Bb和bb的频率对应于二项式 $(f_B+f_b)^2$ 的展开项．如果九种基因型中，存在一种或多种基因型，其频率不等于单个座位上基因型频率的乘积，则说明这两个座位之间存在不平衡．否则，两个座位间不存在不平衡．平衡时，基因型AABB的频率等于基因型AA的频率和BB的频率的乘积，基因型AABb的频率等于基因型AA的频率和Bb的频率的乘积等．即，九种基因型的频率对应于公式 (2.6.1) 的展开项．

$$(f_A^2+2f_Af_a+f_a^2)\times(f_B^2+2f_Bf_b+f_b^2) \tag{2.6.1}$$

随机交配群体中，基因型的数目较多．从基因型的频率度量不平衡度较为复杂．一般从配子型的频率是否等于基因频率的乘积，来度量座位间是否存在不平衡．如果两个座位间不存在连锁，四种配子型AB，Ab，aB和ab的频率等于两个座位上等位基因频率的乘积．即，对应于多项式 $(f_A+f_a)(f_B+f_b)$ 的展开项．因此，四种配子型偏离平衡的程度可度量为，

$$D_{AB}=f_{AB}-f_Af_B,\ D_{Ab}=f_{Ab}-f_Af_b,$$
$$D_{aB}=f_{aB}-f_af_B,\quad D_{ab}=f_{ab}-f_af_b, \tag{2.6.2}$$

其中，f_{AB}，f_{Ab}，f_{aB}和f_{ab}表示四种配子型的频率．四个等位基因的频率可以用四种配子型频率表示为，

$$f_{\mathrm{A}} = f_{\mathrm{AB}} + f_{\mathrm{Aa}}, \quad f_{\mathrm{a}} = f_{\mathrm{aB}} + f_{\mathrm{ab}},$$

$$f_{\mathrm{B}} = f_{\mathrm{AB}} + f_{\mathrm{aB}}, \quad f_{\mathrm{b}} = f_{\mathrm{Ab}} + f_{\mathrm{ab}} \tag{2.6.3}$$

将公式 (2.6.3) 代入公式 (2.6.2) 可以得到,

$$D_{\mathrm{AB}} = D_{\mathrm{ab}} = f_{\mathrm{AB}} f_{\mathrm{ab}} - f_{\mathrm{Ab}} f_{\mathrm{aB}}, \quad D_{\mathrm{Ab}} = D_{\mathrm{aB}} = -(f_{\mathrm{AB}} f_{\mathrm{ab}} - f_{\mathrm{Ab}} f_{\mathrm{aB}}) \tag{2.6.4}$$

把公式 (2.6.4) 中, 配子AB和ab的频率乘积与配子Ab和aB的频率乘积之间的差值定义为连锁不平衡度, 用*D*表示. 即,

$$D = f_{\mathrm{AB}} f_{\mathrm{ab}} - f_{\mathrm{Ab}} f_{\mathrm{aB}} \tag{2.6.5}$$

这样定义的不平衡称为配子型不平衡或连锁不平衡. 但要注意, 有时不平衡并非一定是由连锁造成的, 选择也可以造成独立遗传的两个位点间的不平衡, 具有不同结构的群体按一定比例混合也可以造成不平衡.

在定义了配子型连锁不平衡后, 四种配子型的频率可以用等位基因频率和连锁不平衡度表示为,

$$f_{\mathrm{AB}} = f_{\mathrm{A}} f_{\mathrm{B}} + D, \quad f_{\mathrm{Ab}} = f_{\mathrm{A}} f_{\mathrm{b}} - D,$$

$$f_{\mathrm{aB}} = f_{\mathrm{a}} f_{\mathrm{B}} - D, \quad f_{\mathrm{ab}} = f_{\mathrm{a}} f_{\mathrm{b}} + D \tag{2.6.6}$$

随机交配可以打破连锁不平衡, 用 D_1 表示随机交配一代的连锁不平衡, D_t 表示t (t>0)代随机交配后群体的连锁不平衡度, r为两个位点间的重组率. 则有,

$$D_t = D_1(1-r)^{t-1} \tag{2.6.7}$$

根据公式 (2.6.4) 和公式 (2.6.6), 就能得到随机交配 t 代后, 四种配子型的频率, 即,

$$f_{\mathrm{AB}} = f_{\mathrm{A}} f_{\mathrm{B}} + D_t, \quad f_{\mathrm{Ab}} = f_{\mathrm{A}} f_{\mathrm{b}} - D_t,$$

$$f_{\mathrm{aB}} = f_{\mathrm{a}} f_{\mathrm{B}} - D_t, \quad f_{\mathrm{ab}} = f_{\mathrm{a}} f_{\mathrm{b}} + D_t \tag{2.6.8}$$

现以杂种F_1作为起始群体为例，说明随机交配后配子型理论频率的计算. 设两个亲本P_1和P_2的基因型分别为AABB和aabb，位点A和B间的重组率为r，四个等位基因A，a，B和b的频率均为$\frac{1}{2}$. F_1产生的四种配子型AB，Ab，aB和ab的频率分别为$\frac{1}{2}(1-r)$，$\frac{1}{2}r$，$\frac{1}{2}r$ 和$\frac{1}{2}(1-r)$. 如果产生配子就视为新世代的开始，根据公式 (2.6.5) 得到随机交配一代群体中的配子型连锁不平衡度为，

$$D_1 = \tfrac{1}{2}(1-r)\times\tfrac{1}{2}(1-r) - \tfrac{1}{2}r\times\tfrac{1}{2}r = \tfrac{1}{4}(1-2r) \tag{2.6.9}$$

可以验证四种配子AB, Ab, aB和ab的频率分别为$\frac{1}{4}+D_1$，$\frac{1}{4}-D_1$，$\frac{1}{4}-D_1$ 和$\frac{1}{4}+D_1$，它们之间随机结合产生的基因型与表2.2.1的F_2群体有相同的频率. 因此，有时也把F_2群体视为等位基因频率均为$\frac{1}{2}$ 的随机交配群体. 从F_1开始，随机交配t代后的连锁不平衡度为，

$$D_t = \tfrac{1}{4}(1-2r)(1-r)^{t-1} \tag{2.6.10}$$

图2.6.1给出随机交配过程中，连锁不平衡度的变化曲线. 对于较松散的连锁 (如$r \geq 0.2$)，经过几代随机交配，连锁不平衡度趋于0. 在这样的群体中，两个基因座位间即使存在遗传上的连锁，也难以检测. 对于经历很多代随机交配的群体来说，即使两个基因座位存在紧密的连锁，但由于群体的连锁不平衡度很低，紧密连锁也难以被发现.

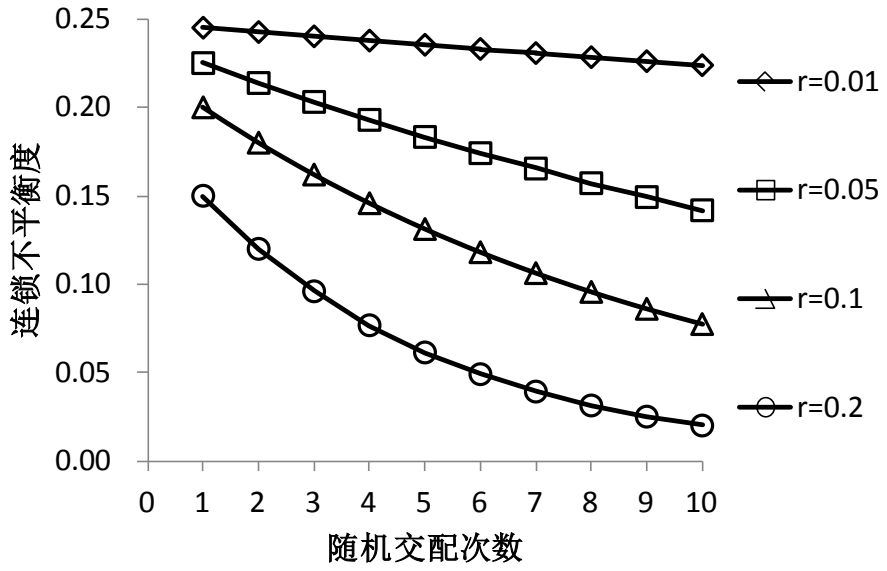

图2.6.1 连锁不平衡度随随机交配次数的变化

随机交配t代，四种配子AB，Ab，aB和ab的频率分别为$\frac{1}{4}+D_t$，$\frac{1}{4}-D_t$，$\frac{1}{4}-D_t$和$\frac{1}{4}+D_t$．如果定义一个累积重组率r'，四种配子的频率用重组率r'表示为$\frac{1}{2}(1-r')$，$\frac{1}{2}r'$，$\frac{1}{2}r'$和$\frac{1}{2}(1-r')$，容易看出$r'=\frac{1}{2}-2D_t$．利用Haldane作图函数把图距转换为重组率，或把重组率转换为图距，把累积重组率r'对应的图距称为累积图距．表2.6.2给出1cM，2cM，5cM和10cM四种图距下，随机交配群体中的累积图距．可以看出，随机交配一次，图距被近似放大50%．对于连续自交产生的RIL群体，图距被近似放大一倍 (表2.6.2)．例如，国际上得到广泛研究的IBM群体，是以著名玉米自交系B73作母本，Mo17作父本杂交，自F_2代开始随机交配4个世代后，再连续自交产生重组近交家系 (Lee et al., 2002)．利用这个群体构建连锁图谱的长度，将是一次交换重组率图谱的2.5倍左右．

表2.6.2 随机交配后的累积遗传图距

随机交配代数	累积图距 (cM)			
1 (等价于 F_2)	1.00	2.00	5.00	10.00
2	1.50	2.99	7.44	14.75
3	2.00	3.98	9.88	19.50
4	2.49	4.97	12.31	24.25
5	2.99	5.96	14.75	29.00
6	3.49	6.95	17.19	33.75
7	3.99	7.94	19.63	38.50
8	4.48	8.93	22.06	43.25
9	4.98	9.92	24.50	48.00
10	5.48	10.91	26.94	52.65
F1-RIL	1.98	3.92	9.55	18.33

§2.6.2 基因型到配子的转移矩阵

对于任意一个遗传群体，其基因型频率用 $\mathbf{f}^{(0)}$表示，

$$\mathbf{f}^{(0)}=\begin{bmatrix} f_{\mathrm{AABB}}^{(0)} & f_{\mathrm{AABb}}^{(0)} & f_{\mathrm{AAbb}}^{(0)} & f_{\mathrm{AaBB}}^{(0)} & f_{\mathrm{AB/ab}}^{(0)} & f_{\mathrm{Ab/aB}}^{(0)} & f_{\mathrm{Aabb}}^{(0)} & f_{\mathrm{aaBB}}^{(0)} & f_{\mathrm{aaBb}}^{(0)} & f_{\mathrm{aabb}}^{(0)} \end{bmatrix} \quad (2.6.11)$$

两个座位上四种等位基因的频率分别为，

$$f_{\mathrm{A}}=f_{\mathrm{AABB}}^{(0)}+f_{\mathrm{AABb}}^{(0)}+f_{\mathrm{AAbb}}^{(0)}+0.5(f_{\mathrm{AaBB}}^{(0)}+f_{\mathrm{AaBb}}^{(0)}+f_{\mathrm{Aabb}}^{(0)})$$

$$f_{\mathrm{a}}=0.5(f_{\mathrm{AaBB}}^{(0)}+f_{\mathrm{AaBb}}^{(0)}+f_{\mathrm{Aabb}}^{(0)})+f_{\mathrm{aaBB}}^{(0)}+f_{\mathrm{aaBb}}^{(0)}+f_{\mathrm{aabb}}^{(0)}$$

$$f_{\mathrm{B}} = f^{(0)}_{\mathrm{AABB}} + f^{(0)}_{\mathrm{AaBB}} + f^{(0)}_{\mathrm{aaBB}} + 0.5(f^{(0)}_{\mathrm{AABb}} + f^{(0)}_{\mathrm{AaBb}} + f^{(0)}_{\mathrm{aaBb}})$$

$$f_{\mathrm{b}} = 0.5(f^{(0)}_{\mathrm{AABb}} + f^{(0)}_{\mathrm{AaBb}} + f^{(0)}_{\mathrm{aaBb}}) + f^{(0)}_{\mathrm{AAbb}} + f^{(0)}_{\mathrm{Aabb}} + f^{(0)}_{\mathrm{aabb}} \qquad (2.6.12)$$

假定随机交配群体充分大, 并且不存在影响群体结构的其他因素, 等位基因的频率在随机交配过程中将保持不变. 因此, 如果能够得到群体产生的四种配子型的频率, 就能根据公式 (2.6.7) 确定随机交配一代的连锁不平衡度 D_1, 根据公式 (2.6.9) 得到随机交配若干代的配子型连锁不平衡度, 根据公式 (2.6.8) 得到随机交配若干代各种配子型的理论频率, 配子间随机结合就能得到各种基因型的频率.

根据§2.1 计算基因型到基因型转移矩阵的计算方法, 首先确定基因型到配子型的转移矩阵 $\mathbf{T_{RM}}$ (公式 2.6.13). 按杂合座位的个数分以下三种情况讨论. 四种配子型 AB, Ab, aB 和 ab 称为配子型 1, 配子型 2, 配子型 3 和配子型 4.

(1) 无杂合座位, 即两个座位上的基因型都纯合. 纯合基因型只产生一种配子型. 基因型 AABB 只产生配子型 1, 基因型 AAbb 只产生配子型 2, 基因型 aaBB 只产生配子型 3, 基因型 aabb 只产生配子型 4. 因此, 转移矩阵 $\mathbf{T_{RM}}$ 第 1 行的第 1 个因素为 1, 其余因素为 0; 第 3 行的第 2 个因素为 1, 其余因素为 0; 第 8 行的第 3 个因素为 1, 其余因素为 0; 第 10 行的第 4 个因素为 1, 其余因素为 0 (公式 2.6.13).

(2) 一个座位纯合, 一个座位杂合. 产生两种配子型, 频率均为 $\frac{1}{2}$. 以基因型 AABb 为例, 产生频率均为 $\frac{1}{2}$ 的配子型 1 (AB) 和配子型 2 (Ab). 因此, 转移矩阵 $\mathbf{T_{RM}}$ 第 2 行的第 1 和 2 个元素为 $\frac{1}{2}$ 和 $\frac{1}{2}$, 其余为 0 (公式 2.6.13). 基因型 AaBB, Aabb 和 aaBb 与 AABb 类似.

(3) 两个座位均杂合, 即基因型 AB/ab 和 Ab/aB. 基因型 AB/ab 产生的四种配子型的频率分别为 $\frac{1}{2}(1-r)$, $\frac{1}{2}r$, $\frac{1}{2}r$ 和 $\frac{1}{2}(1-r)$, 对应于转移矩阵第 5 行的四个元素 (公式 2.6.13). 基因型 Ab/aB 产生的四种配子型的频率分别为 $\frac{1}{2}r$, $\frac{1}{2}(1-r)$, $\frac{1}{2}(1-r)$ 和 $\frac{1}{2}r$, 对应于转移矩阵第 5 行的四个元素 (公式 2.6.13).

$$\mathbf{T}_{\mathrm{RM}}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2}(1-r) & \frac{1}{2}r & \frac{1}{2}r & \frac{1}{2}(1-r) \\ \frac{1}{2}r & \frac{1}{2}(1-r) & \frac{1}{2}(1-r) & \frac{1}{2}r \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.6.13}$$

确定了基因型到配子型的转移矩阵 $\mathbf{T}_{\mathbf{RM}}$ 之后, 就可以把群体中产生的 4 种配子基因型的频率表示为,

$$\left(f_{\mathrm{AB}}^{(1)} \quad f_{\mathrm{Ab}}^{(1)} \quad f_{\mathrm{aB}}^{(1)} \quad f_{\mathrm{ab}}^{(1)}\right)=\mathbf{f}^{(0)}\mathbf{T}_{\mathbf{RM}} \tag{2.6.14}$$

配子型连锁不平衡度 D_1 为,

$$D_1=f_{\mathrm{AB}}^{(1)}f_{\mathrm{ab}}^{(1)}-f_{\mathrm{Ab}}^{(1)}f_{\mathrm{aB}}^{(1)} \tag{2.6.15}$$

§2.6.3 随机交配若干代的配子型和基因型频率

用 t=0 表示基因型频率为公式 (2.6.11) 的群体, 随机交配 t (t>0) 代后的配子型连锁不平衡为,

$$D_t=D_1(1-r)^{t-1} \tag{2.6.16}$$

随机交配 t 代的 4 种配子基因型的频率为,

$$f_{\mathrm{AB}}^{(t)}=f_{\mathrm{A}}f_{\mathrm{B}}+D_t,\quad f_{\mathrm{Ab}}^{(t)}=f_{\mathrm{A}}f_{\mathrm{b}}-D_t,$$

$$f_{\mathrm{aB}}^{(t)}=f_{\mathrm{a}}f_{\mathrm{B}}-D_t,\quad f_{\mathrm{ab}}^{(t)}=f_{\mathrm{a}}f_{\mathrm{b}}+D_t \tag{2.6.17}$$

随机交配 t 代的 10 种基因型的频率分别为,

$$f_{\mathrm{AABB}}^{(t)}=[f_{\mathrm{AB}}^{(t)}]^2,\quad f_{\mathrm{AABb}}^{(t)}=2f_{\mathrm{AB}}^{(t)}\times f_{\mathrm{Ab}}^{(t)},\quad f_{\mathrm{AAbb}}^{(t)}=[f_{\mathrm{Ab}}^{(t)}]^2,$$

$$f_{\mathrm{AaBB}}^{(t)}=2f_{\mathrm{AB}}^{(t)}\times f_{\mathrm{aB}}^{(t)},\quad f_{\mathrm{AB/ab}}^{(t)}=2f_{\mathrm{AB}}^{(t)}\times f_{\mathrm{ab}}^{(t)},\quad f_{\mathrm{Ab/aB}}^{(t)}=2f_{\mathrm{Ab}}^{(t)}\times f_{\mathrm{aB}}^{(t)},\quad f_{\mathrm{Aabb}}^{(t)}=2f_{\mathrm{Ab}}^{(t)}\times f_{\mathrm{ab}}^{(t)},$$

$$f_{\mathrm{AABB}}^{(t)}=[f_{\mathrm{AB}}^{(t)}]^2,\quad f_{\mathrm{AABb}}^{(t)}=2f_{\mathrm{AB}}^{(t)}\times f_{\mathrm{Ab}}^{(t)},\quad f_{\mathrm{AAbb}}^{(t)}=[f_{\mathrm{Ab}}^{(t)}]^2 \tag{2.6.18}$$

例如, 从 F_1 (t=0) 开始, 基因型的频率为,

$$\mathbf{f}^{(0)}=\begin{pmatrix}0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

等位基因的频率为,

$$f_{\mathrm{A}}=f_{\mathrm{a}}=f_{\mathrm{B}}=f_{\mathrm{b}}=\tfrac{1}{2}.$$

随机交配 t (t>0) 代的连锁不平衡度为,

$$D_t=\tfrac{1}{4}(1-2r)(1-r)^{t-1}$$

随机交配 t (t>0) 代的配子型频率为,

$$f_{\mathrm{AB}}^{(t)}=\tfrac{1}{4}+\tfrac{1}{4}(1-2r)(1-r)^{t-1},\quad f_{\mathrm{Ab}}^{(t)}=\tfrac{1}{4}-\tfrac{1}{4}(1-2r)(1-r)^{t-1},$$

$$f_{\mathrm{aB}}^{(t)}=\tfrac{1}{4}-\tfrac{1}{4}(1-2r)(1-r)^{t-1},\quad f_{\mathrm{ab}}^{(t)}=\tfrac{1}{4}+\tfrac{1}{4}(1-2r)(1-r)^{t-1}$$

随机交配 t (t>0) 代的基因型频率为,

$$f_{\mathrm{AABB}}^{(t)}=\tfrac{1}{16}[1+(1-2r)(1-r)^t]^2,\quad f_{\mathrm{AABb}}^{(t)}=\tfrac{1}{8}[1-(1-2r)^2(1-r)^{2t}],$$

$$f_{\mathrm{AAbb}}^{(t)}=\tfrac{1}{16}[1-(1-2r)(1-r)^t]^2,\quad f_{\mathrm{AaBB}}^{(t)}=\tfrac{1}{8}[1-(1-2r)^2(1-r)^{2t}],$$

$$f_{\mathrm{AB/ab}}^{(t)}=\tfrac{1}{8}[1+(1-2r)(1-r)^t]^2,\quad f_{\mathrm{Ab/aB}}^{(t)}=\tfrac{1}{8}[1-(1-2r)(1-r)^t]^2,$$

$$f_{\mathrm{Aabb}}^{(t)}=\tfrac{1}{8}[1-(1-2r)^2(1-r)^{2t}],\quad f_{\mathrm{aaBB}}^{(t)}=\tfrac{1}{16}[1-(1-2r)(1-r)^t]^2,$$

$$f_{\text{aaBb}}^{(t)}=\tfrac{1}{8}[1-(1-2r)^2(1-r)^{2t}],\quad f_{\text{aabb}}^{(t)}=\tfrac{1}{16}[1+(1-2r)(1-r)^t]^2 \tag{2.6.19}$$

随机交配 t (t>0) 代, 然后产生 DH 家系群体中, 4 种基因型的频率为,

$$f_{\text{AABB}}^{(t)-\text{DH}}=\tfrac{1}{4}[1+(1-2r)(1-r)^{t-1}],\quad f_{\text{AAbb}}^{(t)-\text{DH}}=\tfrac{1}{4}[1-(1-2r)(1-r)^{t-1}],$$

$$f_{\text{aaBB}}^{(t)-\text{DH}}=\tfrac{1}{4}[1-(1-2r)(1-r)^{t-1}],\quad f_{\text{aabb}}^{(t)-\text{DH}}=\tfrac{1}{4}[1+(1-2r)(1-r)^{t-1}] \tag{2.6.20}$$

随机交配 t (t>0) 代, 然后产生重复自交产生的 RIL 家系群体中, 4 种基因型的频率为,

$$f_{\text{AABB}}^{(t)-\text{RIL}}=\tfrac{1}{4}[1+\frac{(1-2r)(1-r)^t}{1+2r}],\quad f_{\text{AAbb}}^{(t)-\text{RIL}}=\tfrac{1}{4}[1-\frac{(1-2r)(1-r)^t}{1+2r}],$$

$$f_{\text{aaBB}}^{(t)-\text{RIL}}=\tfrac{1}{4}[1-\frac{(1-2r)(1-r)^t}{1+2r}],\quad f_{\text{aabb}}^{(t)-\text{RIL}}=\tfrac{1}{4}[1+\frac{(1-2r)(1-r)^t}{1+2r}] \tag{2.6.21}$$

练习题

2.1 假定两个基因座位间的重组率为 r, 两个纯合亲本的基因型为 AABB 和 aabb, 计算两个亲本杂交 F2 产生的 DH 群体中, 4 种纯合基因型的理论频率.

2.2 下表给出表 2.2.4 在两个标记*Satt521 和*Satt549 座位上的原始数据. 群体为两个大豆品种间的杂交 F2, 对 60 个 F2 单株做标记型鉴定, 2 和 0 为亲本标记型, 1 为杂合 F1 标记型, -1 代表缺失标记. 即认为标记*Satt521 的 2 型为 AA, 0 型为 aa, 1 型为 Aa; 标记*Satt549 的 2 型为 BB, 0 型为 bb, 1 型为 Bb.

	1-20 个 F2 单株																			
*Satt521	2	0	2	1	1	1	1	1	0	1	2	0	0	1	1	1	0	1	0	2
*Satt549	2	0	2	1	1	-1	1	0	0	0	2	0	0	1	1	1	0	1	1	2
	21-40 个 F2 单株																			
*Satt521	1	2	1	1	1	0	1	2	0	1	2	0	0	0	1	2	0	2	0	0
*Satt549	1	2	1	1	1	0	1	2	0	1	2	0	0	0	1	2	0	1	0	-1
	41-60 个 F2 单株																			
*Satt521	1	1	2	0	1	0	-1	2	1	1	1	1	2	0	1	0	0	2	1	1
*Satt549	2	1	0	0	1	0	-1	2	1	0	1	1	1	0	-1	0	0	2	1	1

(1) 计算两个标记座位上四种等位基因的频率.

(2) 检验两个标记座位上是否存在显著的奇异分离, 即做 1:2:1 的分离比检验.

(3) 根据上述数据整理出表 2.2.4 中九种基因型的观测样本量.

(4) 利用 Newton 迭代算法计算重组率, 并检验重组率与 0.5 之间是否存在显著差异.

2.3 将表 2.2.4 的数据重新整理为下表, 第 4 列给出各种可分辨的基因型中, 重组单倍型的频率, 总样本量用 n 表示.

可分辨的基因型	观测值	理论频率	重组单倍型的频率
AAB_	n_1=572	$\frac{1}{4}(1-r^2)$	$p_1=\frac{r}{1+r}$
AAbb	n_2=3	$\frac{1}{4}r^2$	$p_2=0$
AaB_	n_3=1161	$\frac{1}{2}(1-r+r^2)$	$p_3=\frac{r(1+r)}{2(1-r+r^2)}$
Aabb	n_4=22	$\frac{1}{2}r(1-r)$	$p_4=\frac{1}{2}$
aaB_	n_5=14	$\frac{1}{4}r(2-r)$	$p_5=\frac{1}{2-r}$
aabb	n_6=569	$\frac{1}{4}(1-r)^2$	$p_6=0$

以 AAB_为例说明重组单倍型频率的计算. 从表 2.1.5 可以看出, AAB_由 AABB 和 AABb 两种类型组成, 理论频率分别为 $\frac{1}{4}(1-r^2)$ 和 $\frac{1}{2}r(1-r)$, 频率之和为 $\frac{1}{4}(1-r^2)$. 因此, AAB_的期望观测值为 $\frac{1}{4}(1-r^2)n$, 包含 $\frac{1}{2}(1-r^2)n$ 个单倍型. AABB 的两个单倍型都是亲本型, 因此, 不包含任何重组单倍型. AABb 中的一个单倍型是亲本型, 一个是重组型, 因此, 重组单倍型有 $\frac{1}{2}r(1-r)n$ 个, AAB_中重组型的频率为 $\frac{\frac{1}{2}r(1-r)n}{\frac{1}{2}(1-r^2)n}=\frac{r}{1+r}$. 根据第 4 列的频率, 可以得到重组单倍型的个数为 $\sum_{i=1,6}2n_ip_i$, 总的单倍型为 $2n$. 因此, 得到新的重组率的估计值 $r'=\frac{1}{n}\sum_{i=1,6}n_ip_i$. 对新的重组率估计值 r' 重复上述过程, 就是估计重组率的 EM 算法.

(1) 推导基因型 AaB_中重组单倍型的频率.

(2) 推导基因型 aaB_中重组单倍型的频率.

(3) 给定重组率的一个初始值 0.25, 计算 EM 算法迭代 10 次后的重组率.

(4) 给定重组率的一个初始值 0.10, 计算 EM 算法迭代 10 次后的重组率.

2.4 对表 2.3.2 的数据, 如果分子标记也表现为显性, 4 种基因型的观测值和期望频率列于下表.

基因型	A_B_	A_bb	aaB_	aabb
样本量	1733	25	14	569

期望频率	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(2-r)=\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(2-r)=\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$

(1) 令 $\theta=(1-r)^2$，给出 θ 的似然函数.

(2) 计算 θ 的极大似然估计 $\hat{\theta}$.

(3) 统计理论表明，如果 $\hat{\theta}$ 是参数 θ 的极大似然估计，$g(\theta)$ 是参数 θ 的一个单调函数，那么 $g(\hat{\theta})$ 也是参数 $g(\theta)$ 的极大似然估计. 利用 θ 的极大似然估计 $\hat{\theta}$ 和关系式 $\theta=(1-r)^2$，计算重组率 r 的极大似然估计.

(4) 利用 Fisher 信息量，计算重组率 r 极大似然估计的方差和标准差.

(5) 利用 EM 算法估计重组率 r.

2.5 对练习 2.4 的数据，如果 aa 相对于 A_的选择系数为 0.5，利用选择后的数据计算重组率的极大似然估计.

2.6 以 QTL IciMapping 软件中提供的大麦 DH 作图群体，

(1) 构建大麦的 7 条连锁图谱

(2) 输出大麦的 7 条连锁图谱

(3) 试把某一条染色体从最长的一个标记区间处拆分成两条

2.7 确定一个群体自交无穷多代后的基因型频率，关键是要估计双杂基因型自交无穷多代后的基因型频率. 现以双杂型 AB/ab 为例，自交后代中，根据频率是否相等可把基因型可合并分为 5 类: (1) AABB 和 aabb，为亲本纯合型; (2) AAbb 和 aaBB; 为交换纯合型; (3) AABb, aaBb, AaBB 和 Aabb，为单杂合型; (4) AB/ab，为相引双杂合型; (5) Ab/aB，为互斥双杂合型. 自交世代矩阵可合并为,

$$\mathbf{T}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 & 2r(1-r) & \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 \\ \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 & 2r(1-r) & \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 \end{bmatrix}$$

(1) 对转移矩阵进行分块，用下面的分块矩阵表示，给出分块矩阵 **I, R, O** 和 **Q** 的具体形式

$$\mathbf{T}=\begin{bmatrix}\mathbf{I}_{2\times2} & \mathbf{O}_{2\times3}\\ \mathbf{R}_{3\times2} & \mathbf{Q}_{3\times3}\end{bmatrix}$$

(2) 自交过程中, 五种基因型频率的变化可看作一个马尔可夫链. 类型 (1) 和 (2) 称为吸收态 (Absorbing states), 一旦进入, 将不会再转移到其他状态. 类型 (3), (4), (5) 称为瞬时态 (Transient states). 利用随机过程的有关理论可以证明, 最终由瞬时态类型 j+2 (j=1, 2, 3) 进入吸收态类型 i (i=1, 2) 的概率由矩阵 $\mathbf{R}(\mathbf{I}-\mathbf{Q})^{-1}$ 中的元素 (i, j) 表示. 证明,

$$(\mathbf{I}-\mathrm{Q})^{-1}=\begin{bmatrix}\mathbf{I}_{2\times2} & \mathbf{O}_{2\times3}\\ \mathbf{R}_{3\times2} & \mathbf{Q}_{3\times3}\end{bmatrix}$$

(3) 基因型 AB/ab 属于类型 (4). 自交无穷多代后, 由瞬时态类型 j+2 (j=1, 2, 3) 进入吸收态类型 i (i=1, 2) 的概率由矩阵 $\mathbf{R}(\mathbf{I}-\mathbf{Q})^{-1}$ 中的元素 (i, j) 表示. 由此证明基因型 AB/ab 进入类型 (1)的概率为 $R=\dfrac{2r}{1+2r}$, 进入类型 (2) 的概率为 $1-R=\dfrac{1}{1+2r}$.

2.8 在 P1BC1F1 群体中, 两个座位上 10 种基因型的频率为,

$$\mathbf{f}^{(0)}=\left[f^{(0)}_{AABB}\quad f^{(0)}_{AABb}\quad f^{(0)}_{AAbb}\quad f^{(0)}_{AaBB}\quad f^{(0)}_{AB/ab}\quad f^{(0)}_{Ab/aB}\quad f^{(0)}_{Aabb}\quad f^{(0)}_{aaBB}\quad f^{(0)}_{aaBB}\quad f^{(0)}_{aaBB}\right]$$

$$=\left[\tfrac{1}{2}(1-r)\quad \tfrac{1}{2}r\quad 0\quad \tfrac{1}{2}r\quad \tfrac{1}{2}(1-r)\quad 0\quad 0\quad 0\quad 0\quad 0\right]$$

(1) 证明四种等位基因的频率分别为 $f_{\mathrm{A}}=\frac{3}{4}$, $f_a=\frac{1}{4}$, $f_B=\frac{3}{4}$ 和 $f_b=\frac{1}{4}$.

(2) 证明随机交配一代四种配子型的频率分别为,

$$f^{(1)}_{\mathrm{AB}}=\tfrac{1}{2}+\tfrac{1}{4}(1-r)^2,\quad f^{(1)}_{\mathrm{Ab}}=\tfrac{1}{4}-\tfrac{1}{4}(1-r)^2,$$

$$f^{(1)}_{\mathrm{aB}}=\tfrac{1}{4}-\tfrac{1}{4}(1-r)^2,\quad f^{(1)}_{\mathrm{aB}}=\tfrac{1}{4}(1-r)^2$$

(3) 证明随机交配一代的配子型连锁不平衡度 D_1 为,

$$D_1=\tfrac{1}{16}(3-8r+4r^2)$$

(4) 证明随机交配 t+1（t>=0）代的配子型连锁不平衡度 D_{t+1} 为,

$$D_{t+1}=\tfrac{1}{16}(3-8r+4r^2)(1-r)^t$$

(5) 证明随机交配 t+1（t>=0）代的四种配子型频率为,

$$f^{(t+1)}_{\mathrm{AB}}=\tfrac{9}{16}+D_{t+1},\quad f^{(t+1)}_{\mathrm{Ab}}=\tfrac{3}{16}-D_{t+1},\quad f^{(1)}_{\mathrm{aB}}=\tfrac{3}{16}-D_{t+1},\quad f^{(1)}_{\mathrm{aB}}=\tfrac{1}{16}+D_{t+1}$$

(6) 证明随机交配 t+1（t>=0）代的 10 种基因型频率为,

$$f_{\mathrm{AABB}}^{(t+1)}=(\tfrac{9}{16}+D_{t+1})^2,\quad f_{\mathrm{AABb}}^{(t+1)}=2(\tfrac{9}{16}+D_{t+1})(\tfrac{3}{16}-D_{t+1}),$$

$$f_{\mathrm{AAbb}}^{(t+1)}=(\tfrac{3}{16}-D_{t+1})^2,\quad f_{\mathrm{AaBB}}^{(t+1)}=2(\tfrac{9}{16}+D_{t+1})(\tfrac{3}{16}-D_{t+1}),$$

$$f_{\mathrm{AB/ab}}^{(t+1)}=2(\tfrac{9}{16}+D_{t+1})(\tfrac{1}{16}+D_{t+1}),\quad f_{\mathrm{Ab/aB}}^{(t+1)}=2(\tfrac{3}{16}-D_{t+1})^2,$$

$$f_{\mathrm{Aabb}}^{(t+1)}=2(\tfrac{3}{16}-D_{t+1})(\tfrac{1}{16}+D_{t+1}),\quad f_{\mathrm{aaBB}}^{(t+1)}=(\tfrac{3}{16}-D_{t+1})^2,$$

$$f_{\mathrm{aaBb}}^{(t+1)}=2(\tfrac{3}{16}-D_{t+1})(\tfrac{1}{16}+D_{t+1}),\quad f_{\mathrm{aabb}}^{(t+1)}=(\tfrac{1}{16}+D_{t+1})^2$$

(7) 证明随机交配 t+1（t>=0）代的加倍单倍体家系群体中, 四种纯合基因型的频率为,

$$f_{\mathrm{AABB}}=f_{\mathrm{A}}f_{\mathrm{B}}+(1-r)D_{t+1}=\tfrac{9}{16}+\tfrac{1}{16}(3-8r+4r^2)(1-r)^{t+1},$$

$$f_{\mathrm{AAbb}}=f_{\mathrm{A}}f_{\mathrm{b}}-(1-r)D_{t+1}=\tfrac{3}{16}-\tfrac{1}{16}(3-8r+4r^2)(1-r)^{t+1},$$

$$f_{\mathrm{aaBB}}=f_{\mathrm{a}}f_{\mathrm{B}}-(1-r)D_{t+1}=\tfrac{3}{16}-\tfrac{1}{16}(3-8r+4r^2)(1-r)^{t+1},$$

$$f_{\mathrm{aabb}}=f_{\mathrm{a}}f_{\mathrm{b}}+(1-r)D_{t+1}=\tfrac{1}{16}+\tfrac{1}{16}(3-8r+4r^2)(1-r)^{t+1}.$$

(8) 证明随机交配 t+1（t>=0）代的连续自交 RIL 群体中, 四种纯合基因型的频率为,

$$f_{\mathrm{AABB}}=f_{\mathrm{A}}f_{\mathrm{B}}+(1-R)D_{t+1}=\tfrac{9}{16}+\tfrac{1}{16}\frac{3-8r+4r^2}{1+2r}(1-r)^{t+1},$$

$$f_{\mathrm{AAbb}}=f_{\mathrm{A}}f_{\mathrm{b}}-(1-R)D_{t+1}=\tfrac{3}{16}-\tfrac{1}{16}\frac{3-8r+4r^2}{1+2r}(1-r)^{t+1},$$

$$f_{\mathrm{aaBB}}=f_{\mathrm{a}}f_{\mathrm{B}}-(1-R)D_{t+1}=\tfrac{3}{16}-\tfrac{1}{16}\frac{3-8r+4r^2}{1+2r}(1-r)^{t+1},$$

$$f_{\mathrm{aabb}}=f_{\mathrm{a}}f_{\mathrm{b}}+(1-R)D_{t+1}=\tfrac{1}{16}+\tfrac{1}{16}\frac{3-8r+4r^2}{1+2r}(1-r)^{t+1}.$$