

第一、二章内容要点

- 基因型频率，基因频率
- 自交和回交的遗传效应
- 随机交配 (Random mating)
- Hardy-Weinberg平衡及其检验
- 隐性有害基因携带者的比例
- 影响群体结构的原因
 - 系统过程 (systematic process)
 - 分散过程 (dispersive process)

连锁和连锁不平衡

➤ 连锁 (Linkage)

- A (A1, A2) — B (B1, B2) , 重组率 r

- 相引连锁 (Linkage in coupling)

 - A1B1, A2B2, $(1-r)/2$

- 互斥连锁 (Linkage in repulsion)

 - A1B2, A2B1, $r/2$

➤ 连锁不平衡 (Linkage disequilibrium)

配子别不平衡

- 平衡群体中，配子基因型的频率等于基因频率的乘积，因此群体偏离Hardy-Weinberg平衡的程度常用下面的公式去计算

$$D_{ij} = p(A_i B_j) - p(A_i) p(B_j)$$

连锁不平衡的度量(D)

Gene, frequency	A ₁ , p _A	A ₂ , q _A	B ₁ , p _B	B ₂ , q _B
Gamete (haplotype)	A ₁ B ₁	A ₁ B ₂	A ₂ B ₁	A ₂ B ₂
Frequency, equilibrium	p _A p _B	p _A q _B	q _A p _B	q _A q _B
Frequency, actual	u	s	t	v
Difference from equilibrium	+D	-D	-D	+D

- 基因频率: $p_A = u + s$; $q_A = t + v$; $p_B = u + t$; $q_B = s + v$.
- 对于A₁B₁来说, $p(A_1B_1) - p(A_1)p(B_1) = u - (u+s)(u+t) = u(1-u-s-t) - st = uv - st$

$$D = uv - st$$

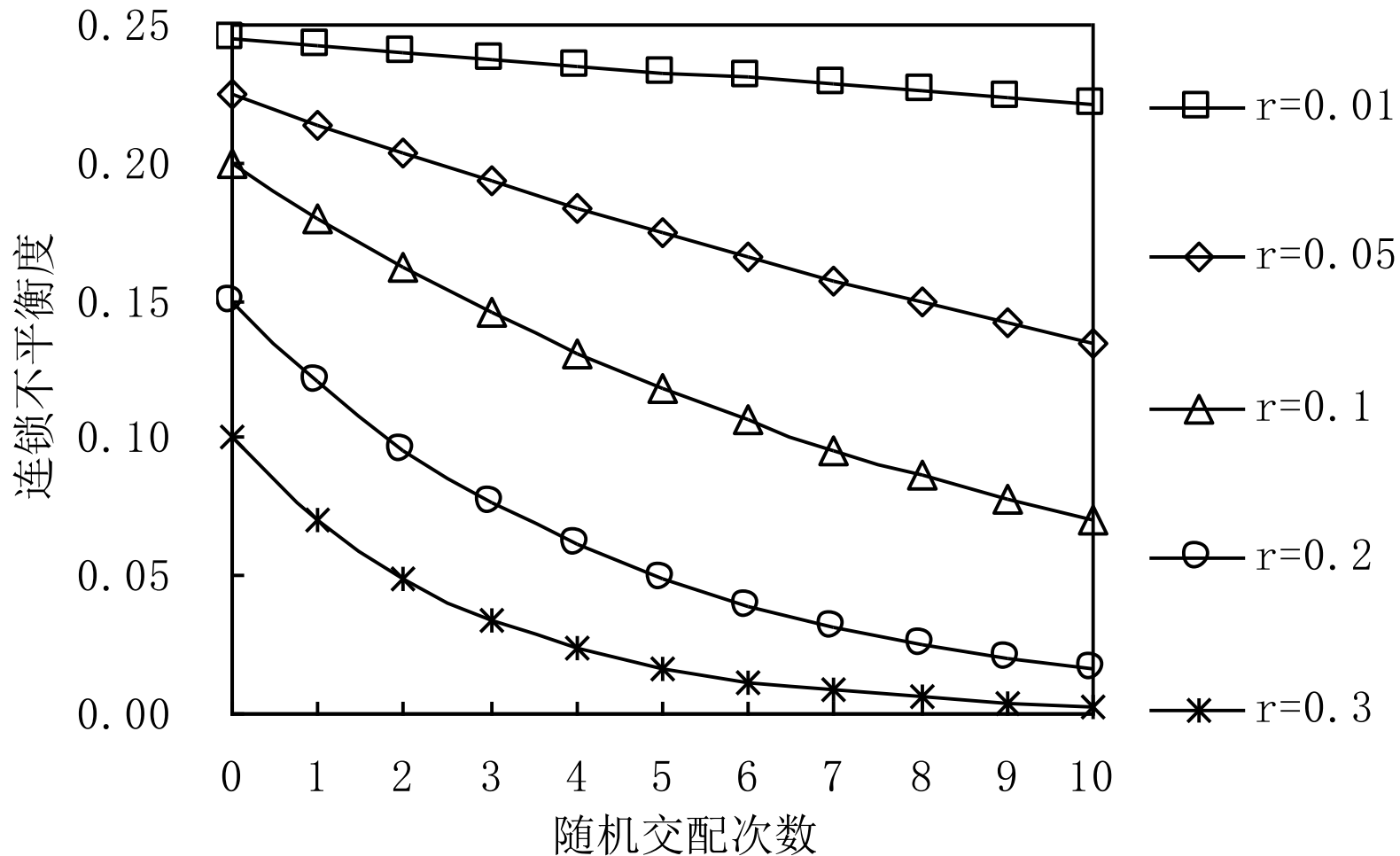
单倍体频率的另一种表示

Gene, frequency	A_1, p_A	A_2, q_A	B_1, p_B	B_2, q_B
Gamete (haplotype)	A_1B_1	A_1B_2	A_2B_1	A_2B_2
Frequency, equilibrium	$p_A p_B$	$p_A q_B$	$q_A p_B$	$q_A q_B$
Frequency, actual	u	s	t	v
Difference from equilibrium	$+D$	$-D$	$-D$	$+D$
Frequency, actual	$p_A p_B + D$	$p_A q_B - D$	$q_A p_B - D$	$q_A q_B + D$

注意1：连锁并非一定存在不平衡

➤ 随机交配可打破连锁不平衡

$$D_t = D_0 (1 - r)^t$$



果蝇，位点A和B，在同一条染色体，相差17.8kb

Haplotype		观测值	观测频率	期望频率	期望值
A	B				
+	+	4	0.085	0.06	2.8
+	-	4	0.085	0.11	5.2
-	+	13	0.277	0.30	14.1
-	-	26	0.533	0.53	24.9

➤ Chi-square=0.93, df=1, $P > 0.5$

➤ D=0.023

注意2：不平衡并非一定连锁

- 不同频率的随机交配群体的混合也会导致不平衡
 - 群体1和群体2 为随机交配群体，位点A和B独立遗传
 - 混和群体一半来自群体1，另一半来自群体2

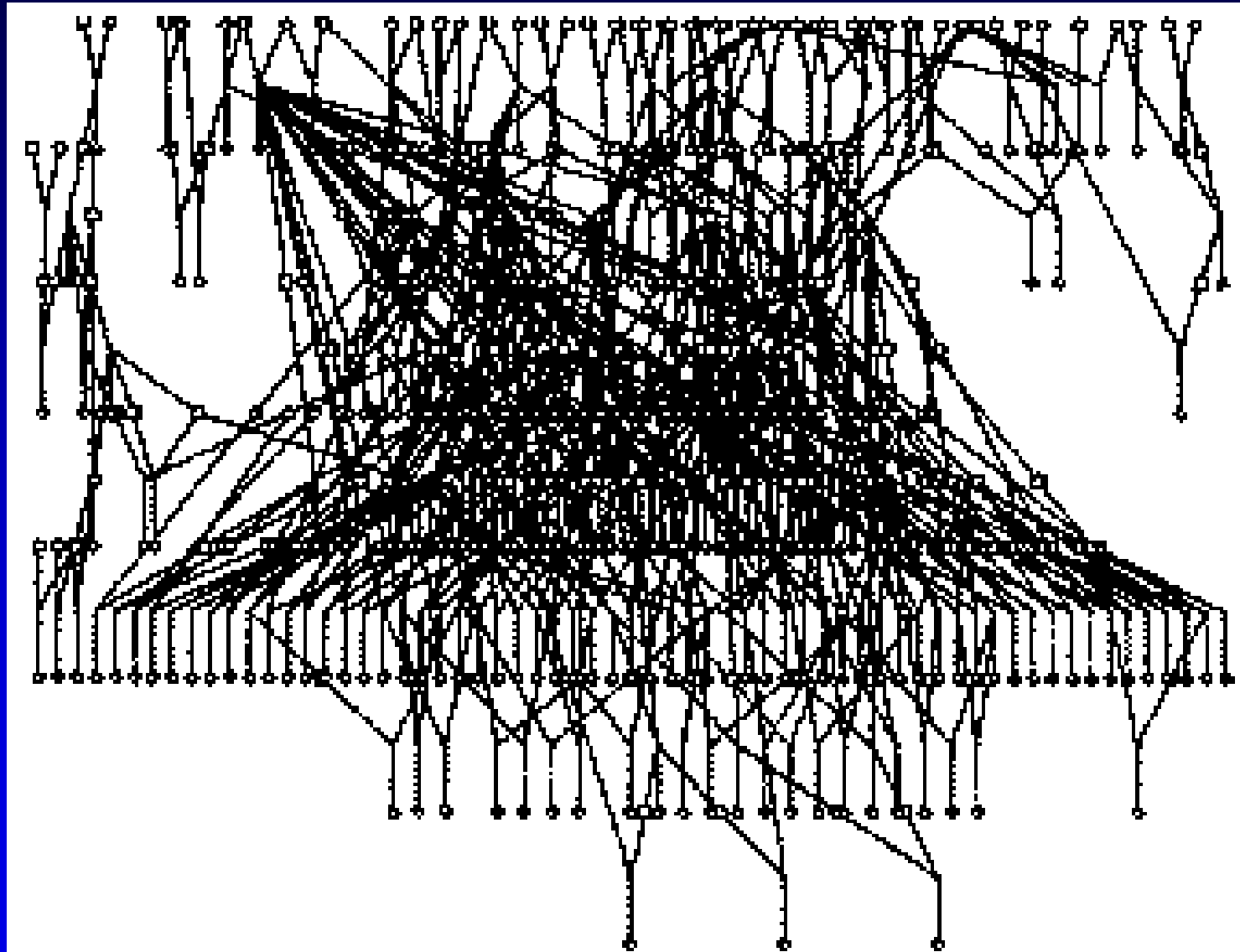
群 体	基因频率				基因型频率
	A1	A2	B1	B2	A1A1B1B1
群体1	0.7	0.3	0.7	0.3	0.2401
群体2	0.3	0.7	0.3	0.7	0.0081
混合群体	0.5	0.5	0.5	0.5	观测频率：0.1241 期望频率：0.0625

第三章

有限大小的群体和近交

第一节 有限大小的理想群体

包含有近交的系谱图



有限大小群体

- 随机飘移 (random drift): 在小群体中对等位基因的频率产生较大的影响, 且其方向不可预测。
- 抽样方法
- 近交方法

两项分布成功概率的估计

- 两项分布：成功概率为 p ， n 次独立试验中成功次数 X 服从两项分布， $\{X=k\}$ 的概率为

$$P(X = k) = C_n^k p^k q^{n-k}, k = 0, 1, \dots, n$$

- X 的均值为 np ，方差为 npq ，其中 $q=1-p$
- 反过来，如在 n 次独立试验中，成功次数为 x ，可以证明成功概率 p 的极大似然估计为

$$\hat{p} = \frac{x}{n}$$

成功概率估计值的期望方差

- 反过来，如在一个n次独立试验中，成功次数为x，可以证明成功概率p的极大似然估计为

$$\hat{p} = \frac{x}{n}$$

- 如果 p_0 为真实成功概率，是未知的， $q_0=1-p_0$ ，那么，

$$V(\hat{p} | p_0) = \frac{V(x | p_0)}{n^2} = \frac{p_0 q_0}{n}$$

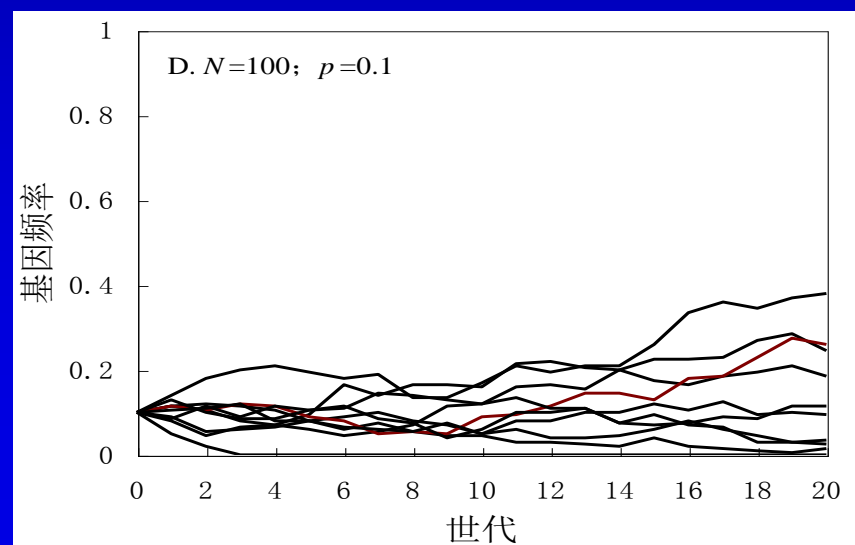
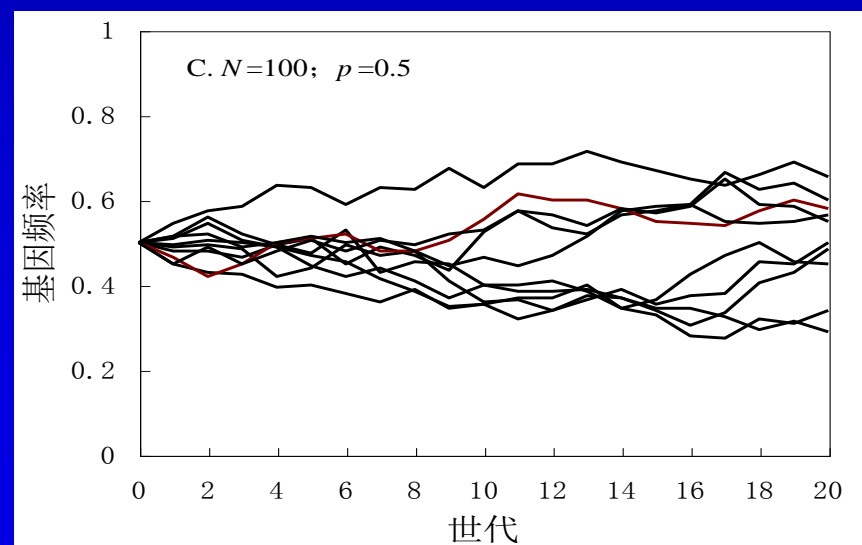
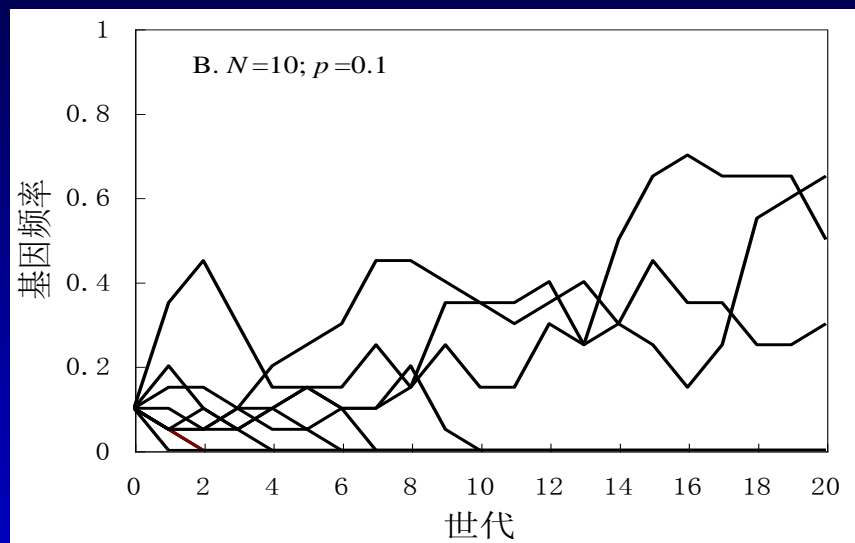
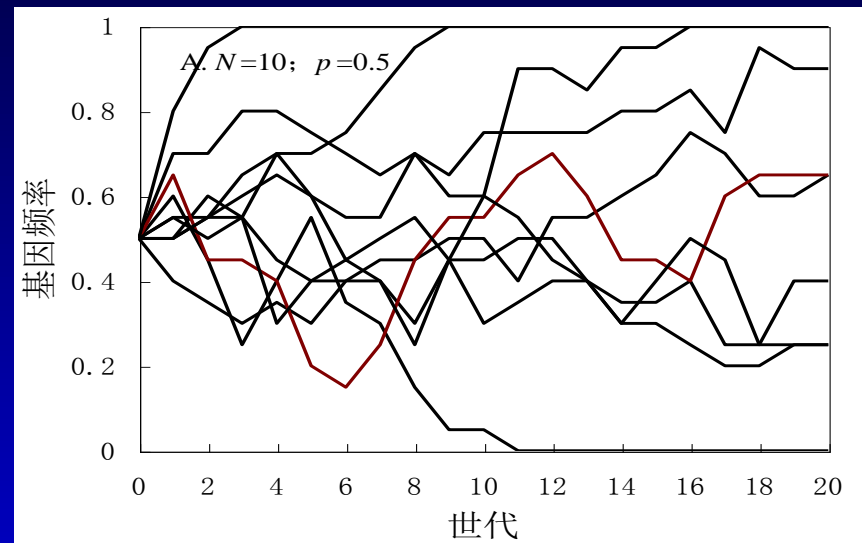
小群体中基因频率估计值的方差

- 两个等位基因A和a，a的频率q为群体中等位基因a的个数X与群体中A和a个数的比值。如N为群体大小，2N为群体中A和a的个数，因此，等位基因a的个数可以看作成功概率为 q_0 ，独立试验次数为2N的两项分布，

$$\sigma_q^2 = \sigma_{\Delta q}^2 = \frac{q_0(1-q_0)}{2N}$$

有限大小群体的随机飘变

$$\sigma_q^2 = \sigma_{\Delta q}^2 = \frac{q_0(1-q_0)}{2N}$$



分散过程的四大特点

- 随机飘变 (random drift)
- 亚群体间分化 (differentiation of sub-populations)
- 亚群体内遗传同质 (genetic uniformity within sub-populations)
- 纯合基因型频率的增加 (overall increase of homozygotes as a consequence of the dispersion of gene frequencies)

近交和近交系数

- 近交的定义：血缘关系较近或相同的个体间的交配，或者说近交是具有共同祖先的个体间的交配。
- 个体间的共祖先程度或者个体间血缘关系的密切程度与所在群体的大小有关。
 - 在两性生物的群体中，每个个体有两个亲本，四个祖先，八个曾祖先，追逆到第 t 个世代以前，应有 2^t 个亲本，当然在 t 较小、群体又相当地大时，还可以满足非近交的要求。
 - 在小群体中，因容量是有限的，所以群体内的个体间必定有某种程度的亲缘关系。一般地，在小群体中用不着追溯多久，就会发现某两个个体有共同的祖先。在小群体中，尽管交配是随机，但群体愈小，个体间的关系愈密切，所以有时小群体遗传性质可以近似看成为是自交交配系统。
- 近交系数：基因位点上2个等位基因后裔同样的概率。

理想群体中的近交

➤ 理想群体 (Idealized population) 定义

- 交配在小群体中发生，各小群体之间没有相互迁移；
- 没有世代重叠；
- 各个小群体中，每个世代有相同数量的个体或具有生殖力的个体数相等；
- 小群体内随机交配；
- 没有选择；
- 没有迁移和突变。

理想群体中的近交

世代0

基础群体（无限大、随机交配）

配子

$2N$

$2N$

$2N$

$2N$

世代1

N

N

N

N

配子

$2N$

$2N$

$2N$

$2N$

世代2

N

N

N

N

配子

$2N$

$2N$

$2N$

$2N$

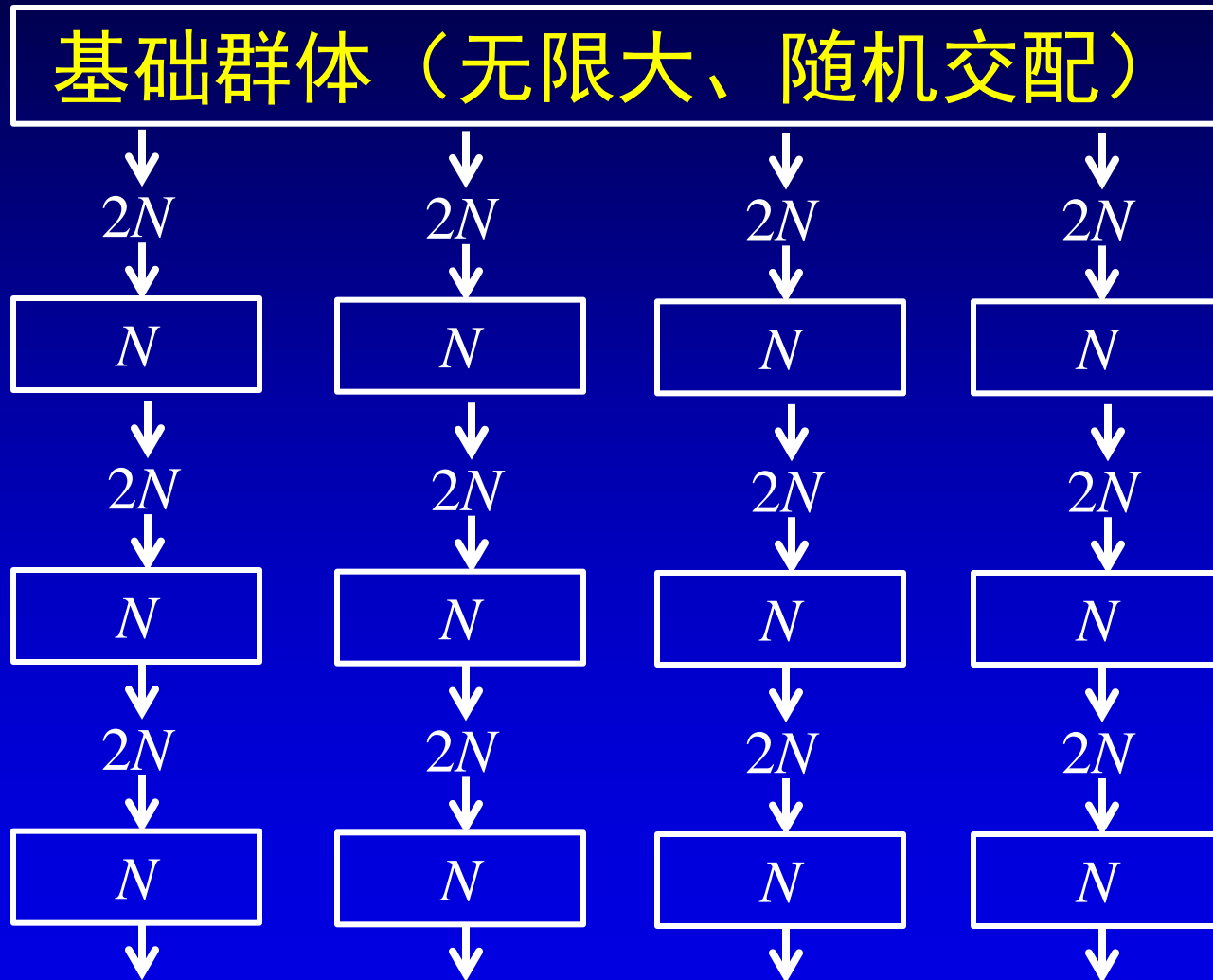
世代3

N

N

N

N



近交系数的估计

- 考虑某一位点的两个等位基因
- 小群体的容量为 N
- 那么一共有 $2N$ 个等位基因
- 那么在雌雄交配中，每个配子与其本身完全一样的配子结合的概率为 $1/2N$
- $1/2N$ 定义为理想群体的近交系数

$$\Delta F = \frac{1}{2N}$$

理想群体近交系数的变化

$$F_0 = 0$$

$$F_1 = \frac{1}{2N}$$

$$F_2 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_1$$

理想群体近交系数的变化

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{t-1}$$

$$F_t = \Delta F + (1 - \Delta F)F_{t-1}$$

➤ 用1减去等式的左右两边，得到

$$1 - F_t = (1 - \Delta F)(1 - F_{t-1})$$

多世代后的近交系数

$$1 - F_t = (1 - \Delta F)^t (1 - F_0)$$

$$F_t = 1 - (1 - \Delta F)^t (1 - F_0)$$

➤ 如果0世代群体的近交系数为0，t个世代后，

$$F_t = 1 - (1 - \Delta F)^t$$

有限大小理想群体中的近交系数

$$F_t = 1 - (1 - \Delta F)^t \xrightarrow{t \rightarrow \infty} 1$$

- 在小群体中，经过若干世代，其交配形式的实质是大部分配子完全相同，即自交，其结果为基因的固定或丢失。另一方面，这个过程可能导致亚群体内的差异在缩小，或者在一个小群体中纯合个体增加而杂合体减少。

三种大小随机交配群体中近交系数的变化

群体大小	随机交配世代				
	1	5	10	20	40
10	0.050	0.226	0.401	0.642	0.871
20	0.025	0.119	0.224	0.397	0.637
30	0.017	0.081	0.155	0.285	0.490

建立者效应 (Founder effect)

- 当群体是由少数几个个体繁衍而来时，群体容量的大小是非常重要的。例如，某海岛上的的人群，是由于另一海岛或大陆上的少数几个人首先聚居，而后来的群体则是由这几个个体交配繁殖而来，那么海岛群体的基因频率特征决定于几个祖先个体中的频率，海岛群体的遗传特性可能和原群体有较大差异，这种差异是由建立者效应所引起的。

瓶颈效应 (Bottle-neck effect)

- 另一情况下，可能的少数个体在自然或人工的影响下（如大的自然灾害）能够存活下来的，以后这些个体繁衍一个新群体，当然新群体基因频率可能和原来的群体大不相同，而且基因的多样性也不丰富，交配产生新类型的可能性不大，这种效应叫瓶颈效应。

第二节 有效群体大小及其在植物遗传资源保护中的应用

理想群体的近交系数和 基因频率变化的度量

➤ 理想群体中近交系数的变化

$$\Delta F = \frac{1}{2N}$$

➤ 理想群体中基因频率的方差

$$\sigma_q^2 = \sigma_{\Delta q}^2 = \frac{q_0(1-q_0)}{2N}$$

一般的群体，

- 育种群体大多并不是理想群体，为了能够像理想群体那样用基因频率的方差或近交系数去研究育种群体的分散过程，特提出有效群体大小（Effective population size）这一概念，有效群体大小常用 N_e 表示，这样任一有效群体大小为 N_e 的群体均可视为包含 N_e 个个体的理想群体，因此，群体的分散过程具有与理想群体同样的基因频率方差和近交系数。

有效群体大小的计算

➤ 近交有效群体大小
(Inbreeding effective
population size)

$$N_e = \frac{1}{2\Delta F}$$

➤ 方差有效群体大小
(Variance effective
population size)

$$N_e = \frac{q_0(1-q_0)}{2\sigma_{\Delta q}^2}$$

常见育种群体的有效群体大小

➤ 避免近亲交配的群体

■ 排除自交

$$N_e \approx N + \frac{1}{2}$$

$$\Delta F \approx \frac{1}{2N+1}$$

■ 排除自交和同胞交配

$$N_e \approx N + 2$$

$$\Delta F \approx \frac{1}{2N+4}$$

常见育种群体的有效群体大小

➤ 雌雄个体数不等的群体

$$\frac{1}{N_e} \approx \frac{1}{4N_m} + \frac{1}{4N_f}$$

$$N_e \approx \frac{4N_m N_f}{N_m + N_f}$$

$$\Delta F \approx \frac{1}{8N_m} + \frac{1}{8N_f}$$

- 可以看出，这时的近交系数主要由雌雄个体数 N_m 和 N_f 中的较小的一个决定。一个极端的例子是有无穷多雌性个体但只有一个雄性个体，这时有效群体大小 N_e 大约只等于4。

常见育种群体的有效群体大小

➤ 世代间个体数不等的群体

$$\frac{1}{N_e} \approx \frac{1}{t} \left(\frac{1}{N_1} + \frac{1}{N_2} + \dots + \frac{1}{N_t} \right)$$

- 可以看出，具有最小个体数的世代对有效群体大小 N_e 的影响最大，这种效应就是前面提及的瓶颈效应或建立者效应。

常见育种群体的有效群体大小

➤ 家系大小非随机分布的群体

- 家系大小非随机分布的群体是偏离理想群体的最常见的形式，家系大小定义为每个个体为下一个世代群体贡献的子代的数目。
- 对理想群体，我们假定每个个体有等同的机会为下一代贡献基因或后代，亲本个体对后代的贡献是一个随机分布。
- 对实际的育种群体来说，由于亲本个体在繁殖能力上的差异，以及它们的后代个体在生存能力上的差异，亲本个体很少具有相同的机会为下一代贡献基因或后代。这种差异会造成家系大小差异的增加，其结果是后代群体中的大部分个体来自于少数的一些亲本，从而造成有效群体大小的下降。
- 相反，特殊的育种方法也可以把家系大小的差异降低，从而增加有效群体大小。

常见育种群体的有效群体大小

➤ 家系大小非随机分布的群体

$$N_e \approx \frac{4N}{V_k + 2}$$

- 对于理想群体，平均家系大小： $\bar{k} = 2$
- 对于理想群体，家系大小这一随机变量 k 在理想群体中可视为泊松（Poisson）分布，泊松（Poisson）分布的方差等于其均值，即 $V_k = \bar{k}$
- 对于理想群体， $N_e = N$
- 还可以看出，在方差为0时， $N_e = 2N$

常见育种群体的有效群体大小

➤ 常异花授粉物种的有效群体大小

- 有些植物以常异花授粉方式繁殖后代，用 s 表示自花授粉的比例， $1-s$ 为异花授粉的比例，平衡时群体的近交系数为 F 。常异花授粉物种的有效群体大小为

$$F = \frac{s}{2-s}$$

- 家系大小不等时的有效群体大小，

$$N_e = \frac{2N}{\frac{V_k}{\bar{k}}(1+F) + (1+F)}$$

植物遗传资源保护

➤ 遗传资源的搜集

➤ 遗传资源的再生繁殖

遗传资源群体的搜集

➤ 对自交率为 s 的群体，从 n 个植株上搜集 N 粒种子

$$N_e = \frac{2N(2-s)}{(1+s)^2 \frac{N}{n} + (3+s)(1-s)}$$

➤ 对随机交配群体，即 $s=0$ ，从 n 个植株上搜集 N 粒种子

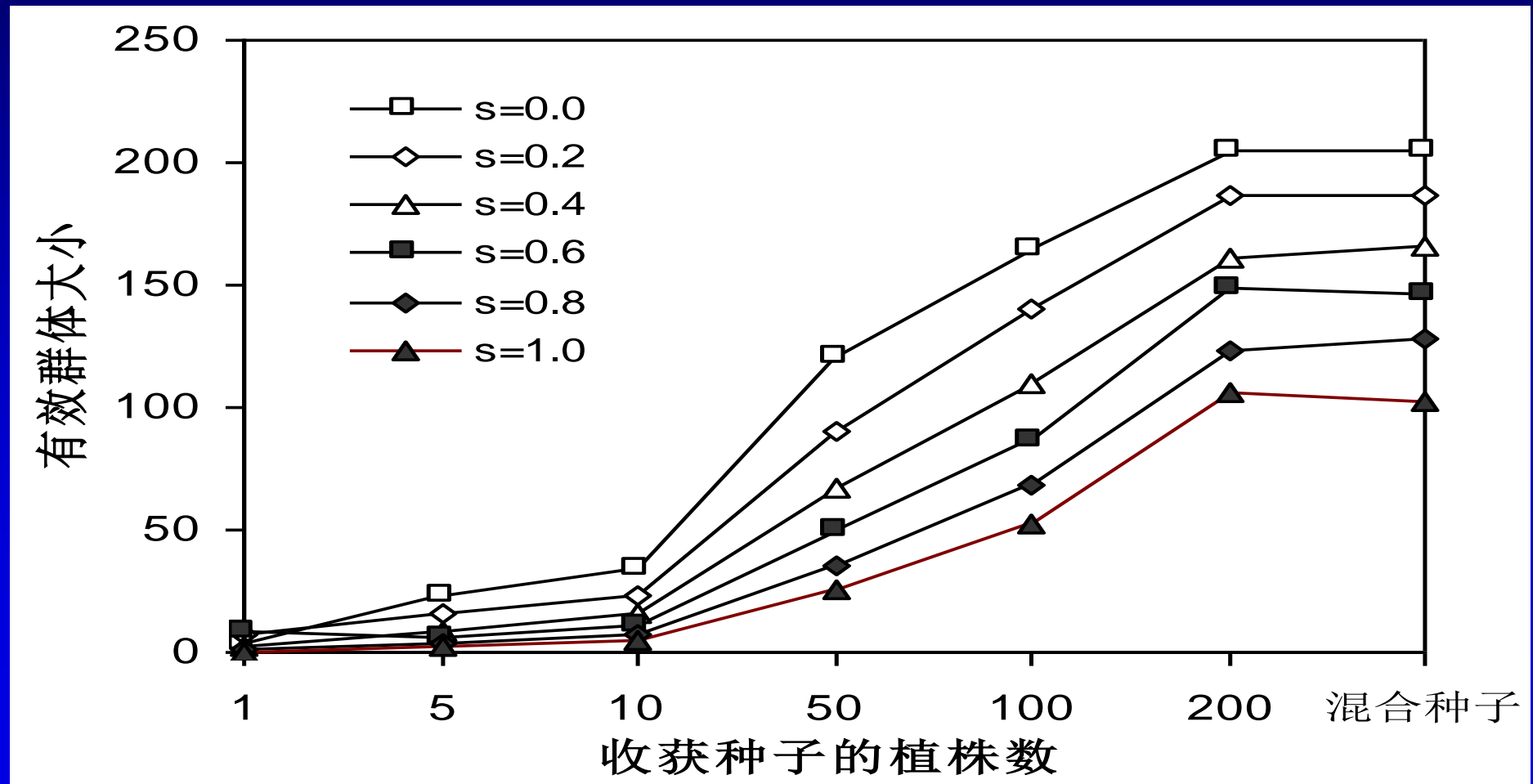
$$N_e = \frac{4N}{\frac{N}{n} + 3}$$

➤ 从自交率为 s 的混合群体中搜集 N 粒种子

$$N_e = \frac{N}{1 - \frac{1}{2}s} = \frac{N}{1 + F}$$

遗传资源的搜集

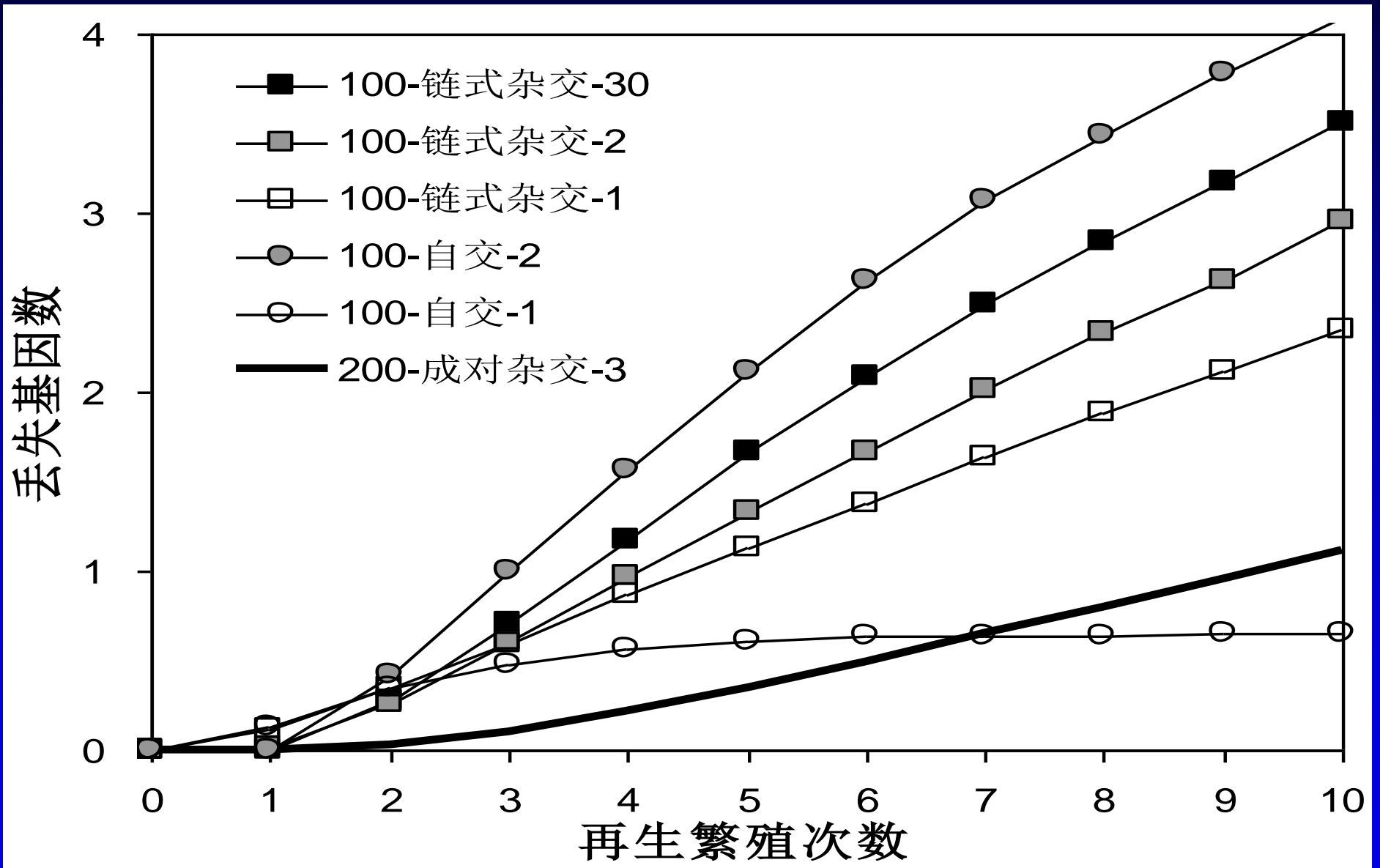
从不同数量的植株等量采集200粒种子的有效群体大小 (s 为自交率)



植物遗传资源的再生繁殖的杂交方式

- 自交
- 随机交配
- 混合花粉随机交配
- 成对杂交： $1 \times 2, 3 \times 4, 5 \times 6, \dots, (n-1) \times n$
- 链式杂交： $1 \times 2, 2 \times 3, 3 \times 4, \dots, n \times 1$

再生繁殖过程中基因的丢失



以一定概率保证至少有一棵成株时， 需要种植的种子粒数

种子发芽率	至少有一粒发芽的概率		
	0.999	0.99	0.95
0.9	3	2	2
0.8	5	3	2
0.7	6	4	3
0.6	8	6	4
0.5	10	7	5

第三节 亲属关联和近交系数

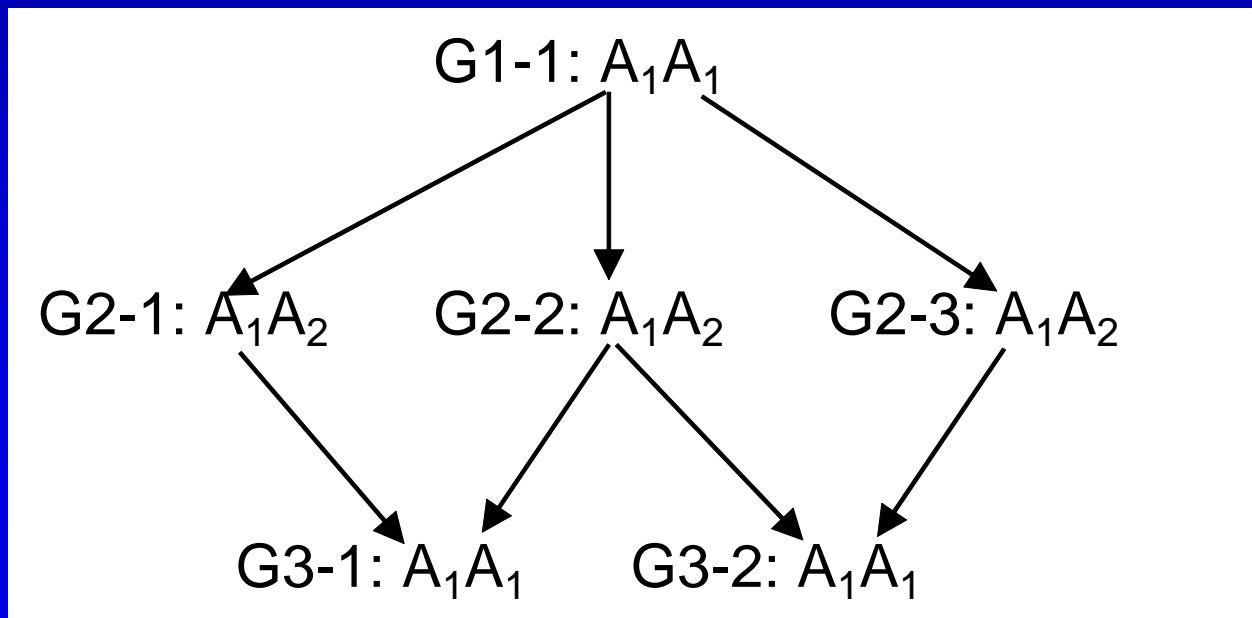
状态相同和后裔同样基因

➤ 状态相同 (alike in state)

- 如果这两个基因具有相同的物理结构（即有相同的DNA序列，相同的功能等），则称为状态相同

➤ 后裔同样 (IBD: identical by descent)

- 如果两个等位基因是共同亲本中的同一个基因的拷贝，则称它们是后裔同样的



祖先关联和近交

➤ 祖先关联(related by ancestry)

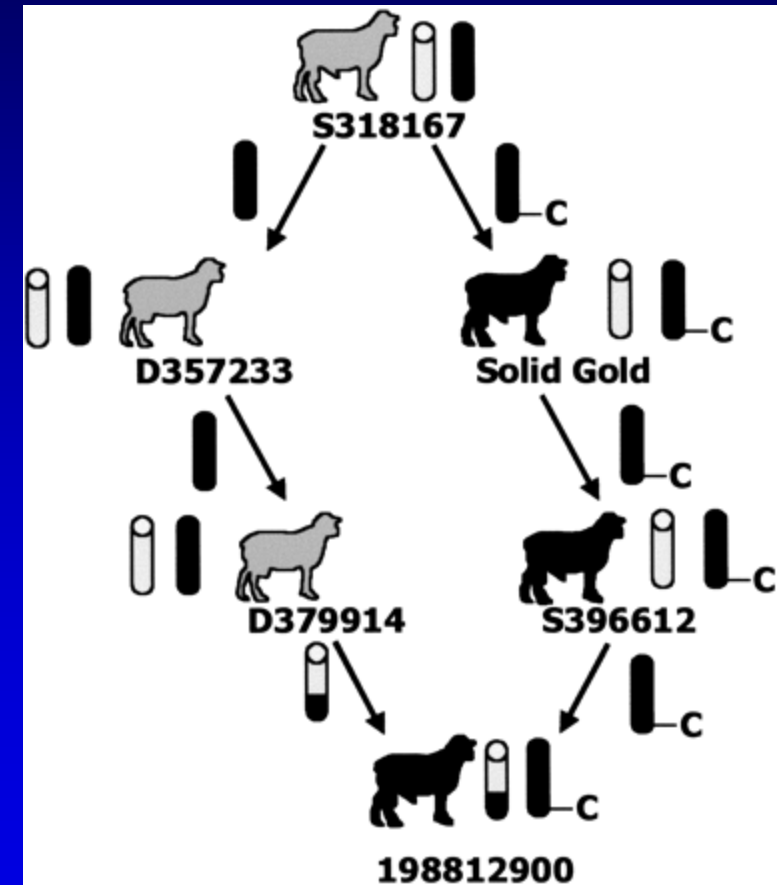
- 如果两个个体具有共同的祖先，或者说具有至少一个共同的祖先，则称这两个个体是祖先关联的。

➤ 近交 (inbreeding)

- 有祖先关联的两个个体间的交配称为近交，近交导致近交系数的增加，近交系数的增加意味着两个等位基因有相同来源的概率的增加。

Identification of the Single Base Change Causing the Callipyge Muscle Hypertrophy Phenotype, the Only Known Example of Polar Overdominance in Mammals

➤ Ram 198812900 exhibits an inbreeding path with the sire (S318167) of Solid Gold contributing on both maternal and paternal sides of the pedigree. We hypothesized that this inbreeding path has allowed the region to be identical-by-descent, except for the *CLPG* mutation, which we propose occurred in the gamete that produced Solid Gold. The C identifies the mutated allele, and individuals with the muscle hypertrophy phenotype are identified as **solid black sheep**; those with normal phenotype are identified as **gray sheep**. Wild-type chromosomes are represented as gray chromatids. A recombination event is depicted to have occurred on the maternal side of the pedigree; however, the event could have occurred on either or both sides to generate marker informativeness in the centromeric region of chromosome 18.



➤ Genome Res. 2002. 12(10):1496-1506

共祖先系数

- 共祖先系数 (coefficient of coancestry): 在一个基因座位上, X的一个随机等位基因与Y的一个随机等位基因后裔同样的概率, 又称亲本系数 (coefficient of parentage)
- 它针对的是两个个体
- 个体X和Y之间的共祖先系数常用 f_{XY} 表示。

近交系数

- 近交系数：基因位点上2个等位基因后裔同样的概率
- 针对的是一个个体，个体的近交系数常用 F_x 表示。
 - 近交系数 F_x 等于个体 X 中一个基因座位上的两个等位基因是后裔同样的概率。
 - $F_x=0$ ，意味着无近交（如无限大小随机交配群体）
 - $F_x=1$ ，意味着完全近交（如自交系）
 - 近交系数等于由近交引起的杂合型频率相对于Hardy-Weinberg平衡的杂合型频率的变化。

$$F = \frac{P_{12} - P_{12(F)}}{P_{12}} = 1 - \frac{P_{12(F)}}{P_{12}}$$

近交系数为F，基因型的频率

➤ 杂合基因型A1A2

$$P_{12(F)} = 2pq(1-F)$$

➤ 纯合基因型A1A1

$$P_{11(F)} = p^2 + pqF$$

➤ 纯合基因型A2A2

$$P_{22(F)} = q^2 + pqF$$

自交后代的近交系数

- 考虑到杂交亲本X和Y间的亲本系数为 f_{XY} ，自交后代的近交系数为

$$F = 1 - \left(\frac{1}{2}\right)^{n-2} (1 - f_{XY})$$

- 相邻两个世代间，近交系数 F_n 和 F_{n-1} 的关系为

$$F_n = \frac{1}{2} (1 + F_{n-1})$$

- 亲本间的祖先关联会提高近交系数。

不同自交世代的近交系数

世代类型		杂合型频率	近交系数 F
单株	家系		
F_2 或 S_0	F_3 或 S_1	P_{12}	0
F_3 或 S_1	F_4 或 S_2	$0.5 \times P_{12}$	0.5
F_4 或 S_2	F_5 或 S_3	$0.25 \times P_{12}$	0.75
F_5 或 S_3	F_6 或 S_4	$0.125 \times P_{12}$	0.875
F_n 或 S_{n-2}	F_{n+1} 或 S_{n-1}	$(\frac{1}{2})^{n-2} P_{12}$	$1 - (\frac{1}{2})^{n-2}$
F_∞ 或 S_∞	F_∞ 或 S_∞	0	1

混合自交和随机交配

➤ 混合自交和随机交配 (Mixed selfing and random mating)

➤ 根据两个相邻自交世代间的近交系数 F_n 和 F_{n-1} 的关系, 有:

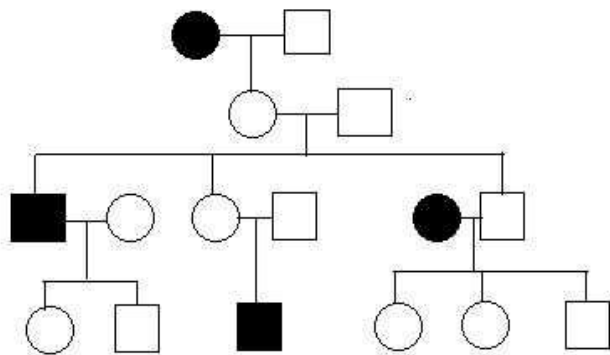
$$F_n = \frac{1}{2}(1 + F_{n-1})s$$

➤ 在平衡状态时, $F_n = F_{n-1}$, 因此

$$F = \frac{s}{2-s}$$

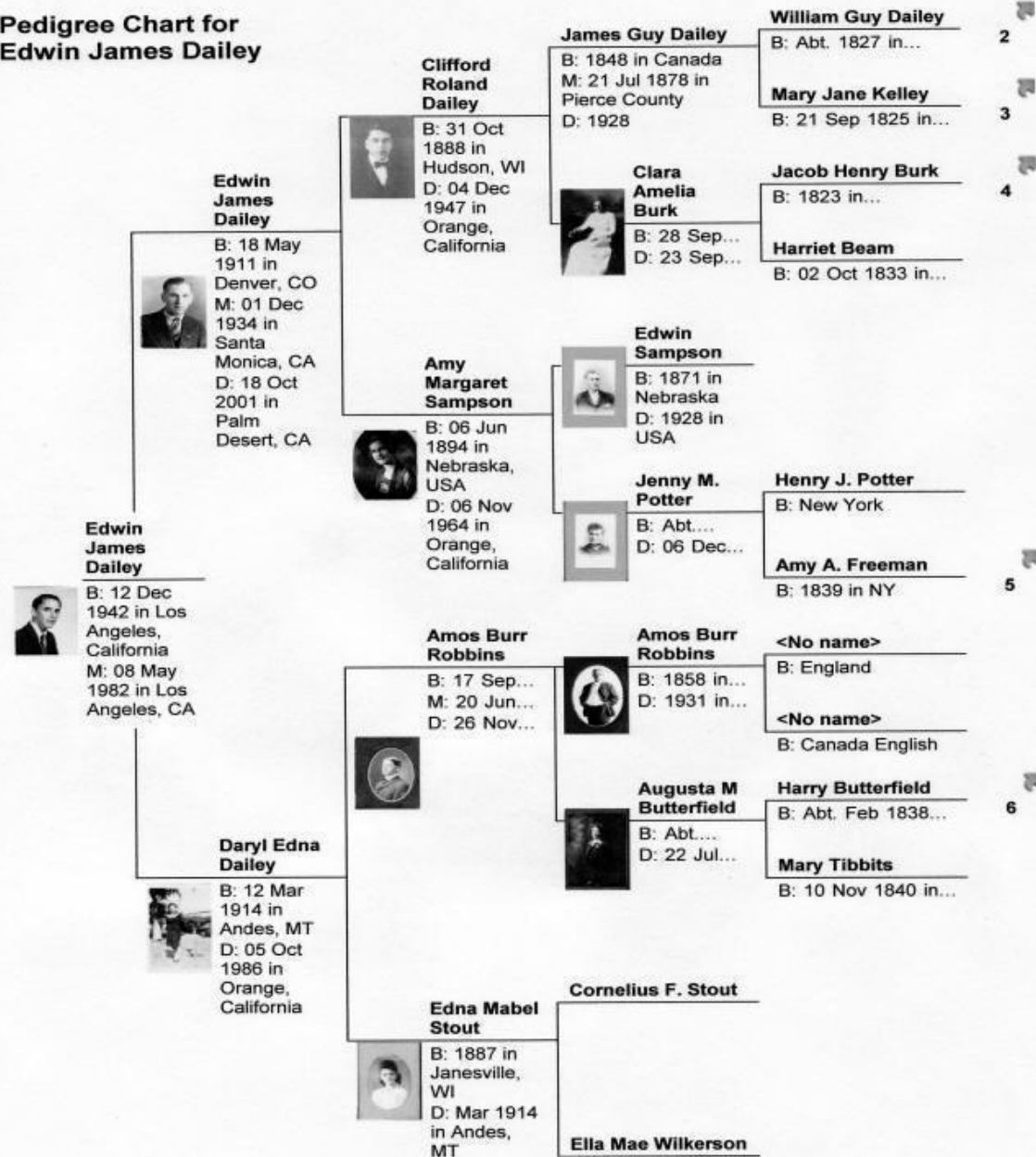
(式中 s 表示自花授粉比例, $1-s$ 为异花授粉比例)

利用系谱数据估计后裔同样概率

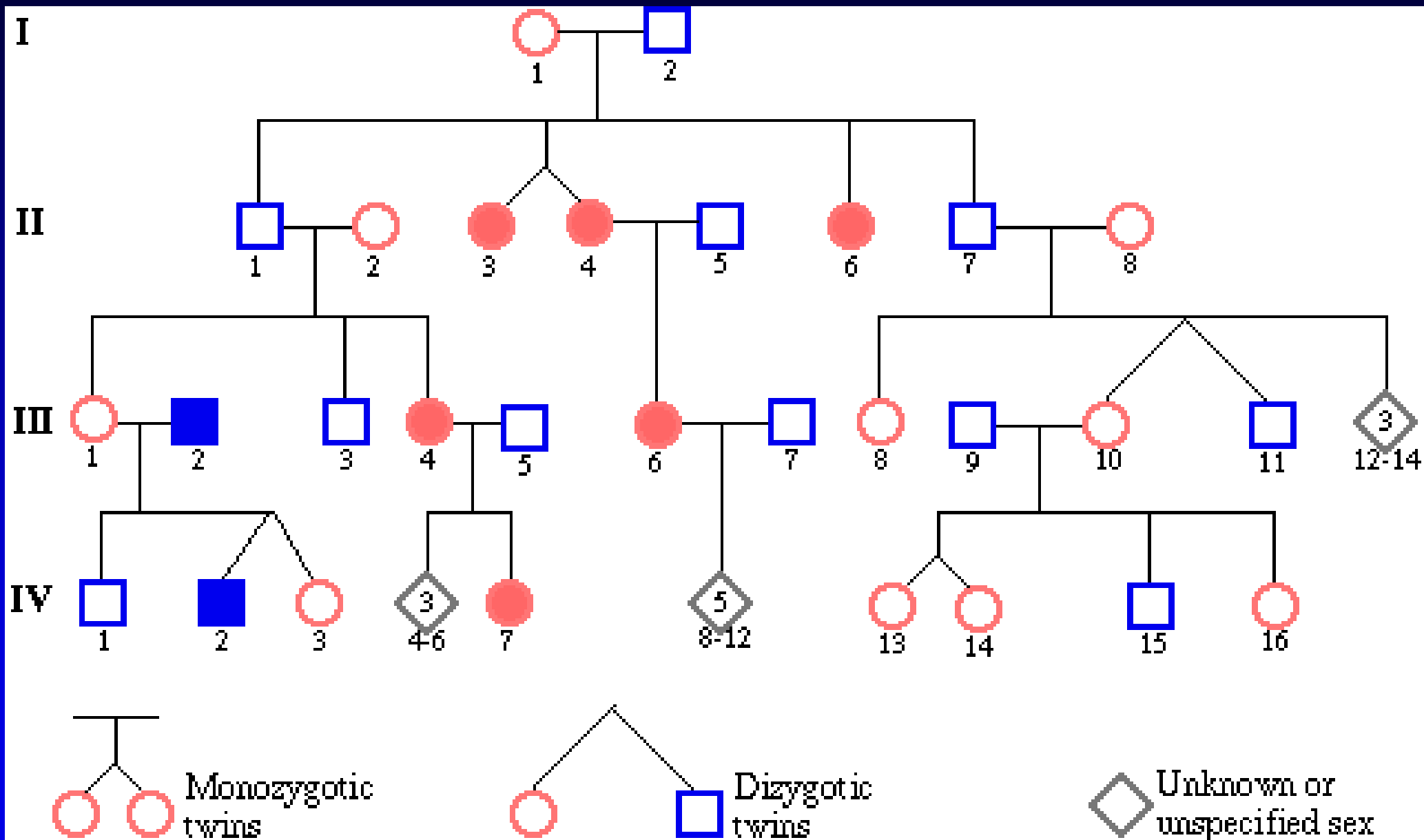


□ = male
○ = female

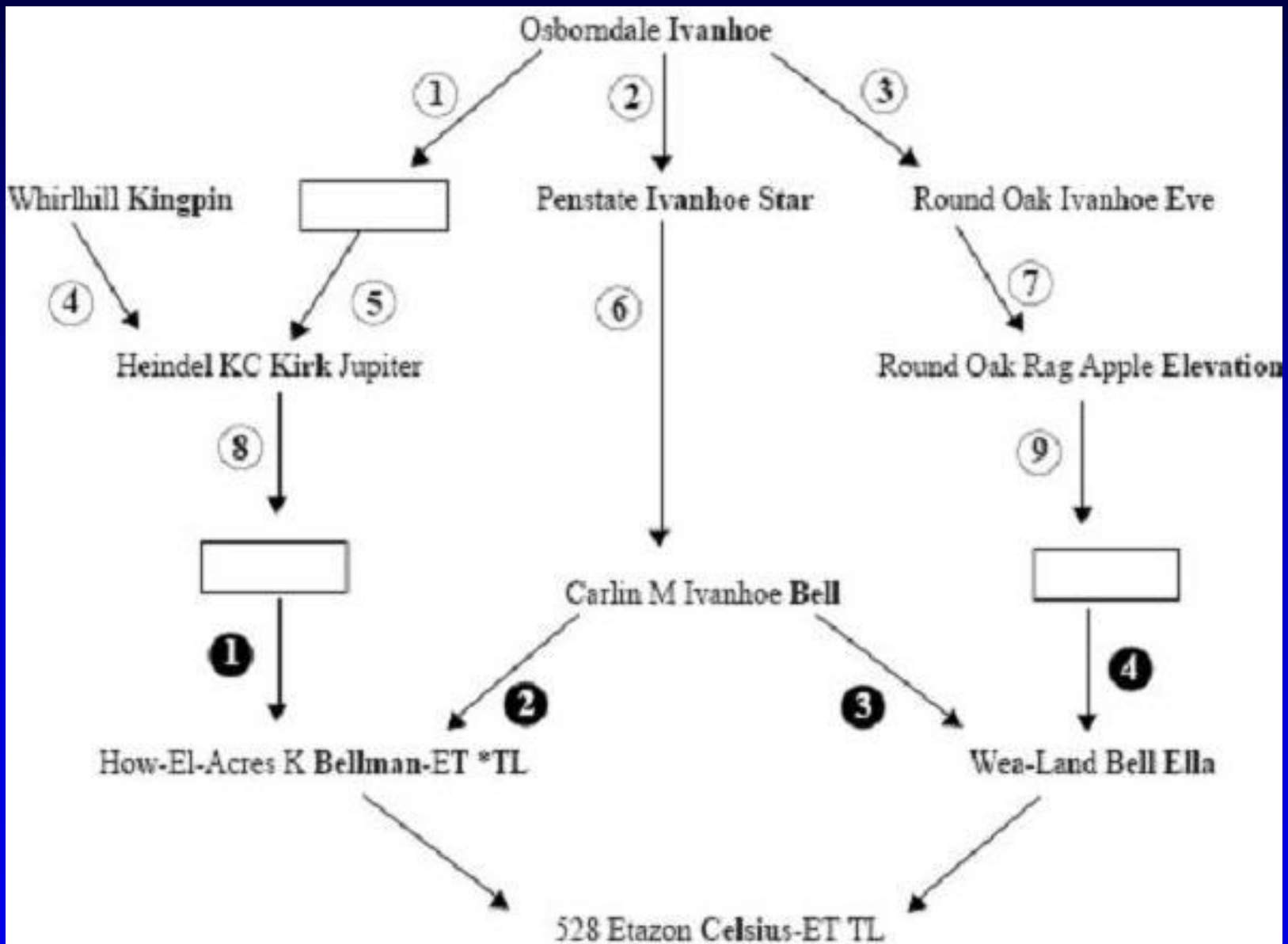
Pedigree Chart for Edwin James Dailey



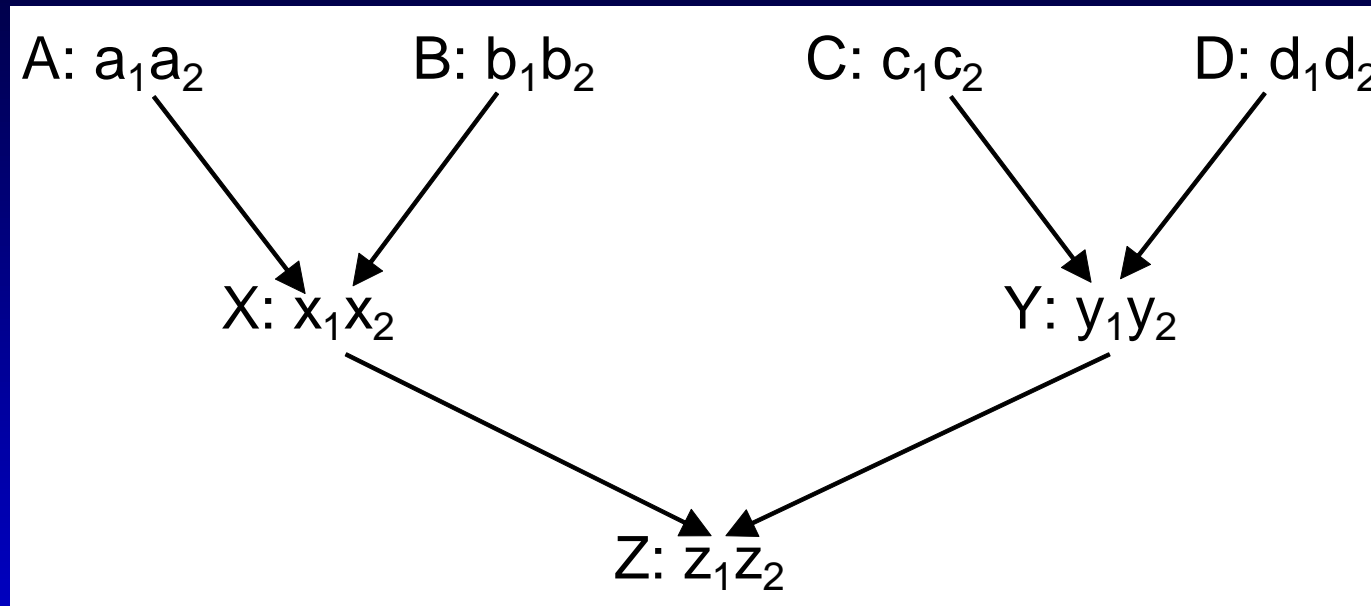
人类系谱图



包含有近交的系谱图(牛)



一个假定的系谱中的个体及其基因型



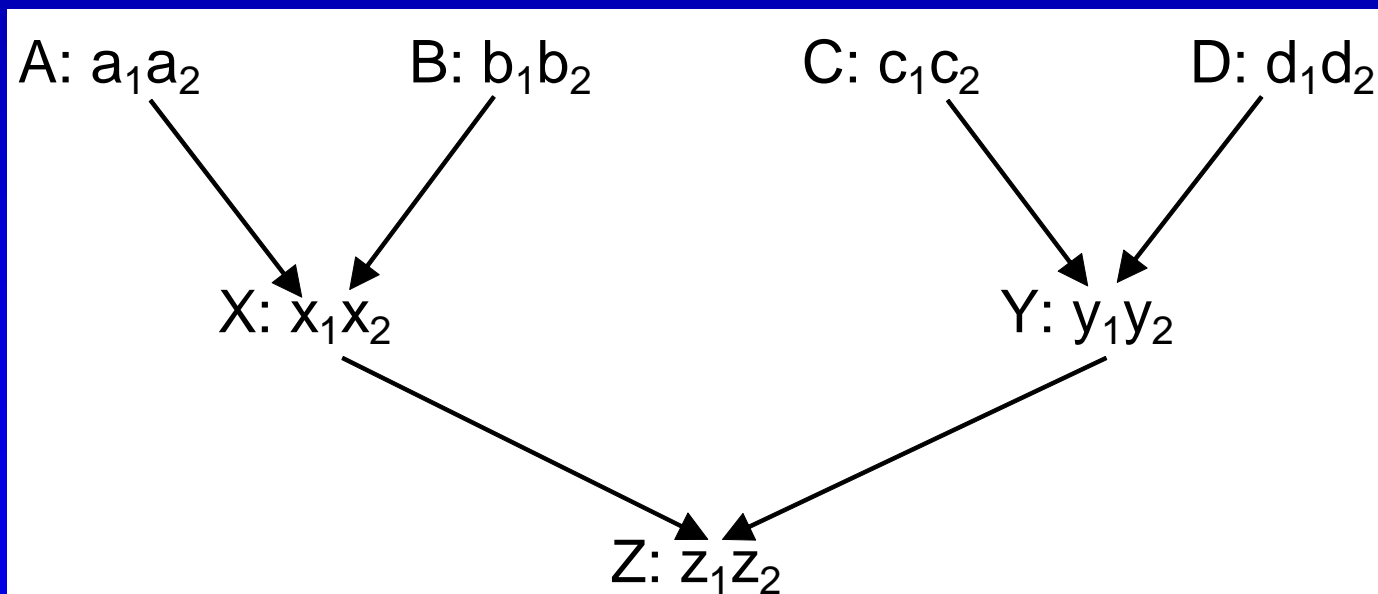
- 由上图可见，个体X和Y间的共祖先系数 f_{XY} 为等位基因x和y后裔同样的概率为：

$$f_{XY} = P\{x \equiv y\}$$

$\{x \equiv y\}$ 的分解

$$\{x \equiv y\} = \{x = x_1, y = y_1, x_1 \equiv y_1\} \cup \{x = x_1, y = y_2, x_1 \equiv y_2\}$$

$$\cup \{x = x_2, y = y_1, x_2 \equiv y_1\} \cup \{x = x_2, y = y_2, x_2 \equiv y_2\}$$



二倍体物种的共祖先系数

$$P\{x = x_1\} = P\{x = x_2\} = P\{y = y_1\} = P\{y = y_2\} = \frac{1}{2}$$

$$f_{XY} = \frac{1}{4}[P\{x_1 \equiv y_1\} + P\{x_1 \equiv y_2\} + P\{x_2 \equiv y_1\} + P\{x_2 \equiv y_2\}]$$

- 上式是共祖先系数的一般数学表述，是利用系谱计算共祖先系数的基础。

个体X与其自身的共祖先系数

$$f_{XX} = \frac{1}{4}[P\{x_1 \equiv x_1\} + P\{x_1 \equiv x_2\} + P\{x_2 \equiv x_1\} + P\{x_2 \equiv x_2\}]$$

- 如果个体X是非自交系，即 $F_X=0$ ，则个体X和它自己的共祖先系数 $f_{XX}=1/2$ 。
- 如果个体X是自交系，即 $F_X=1$ ，则个体X和它自己的共祖先系数 $f_{XX}=1$ 。

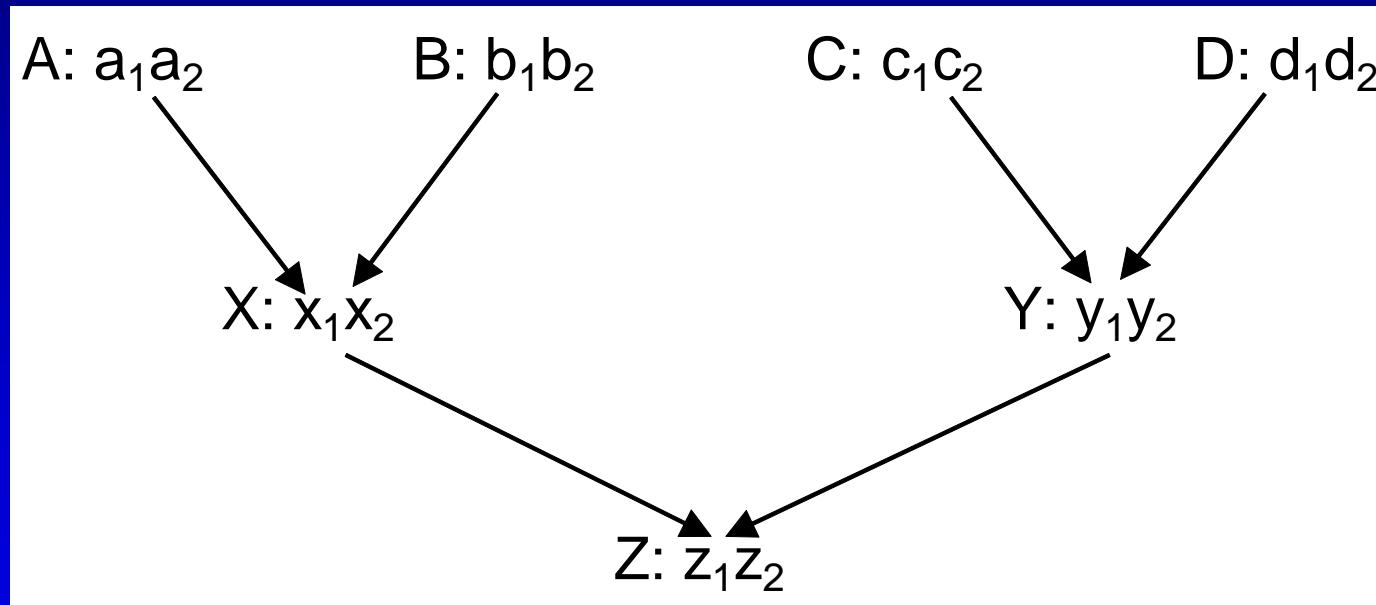
双亲后代的近交系数

$$F_Z = \frac{1}{4}[P\{x_1 \equiv y_1\} + P\{x_1 \equiv y_2\} + P\{x_2 \equiv y_1\} + P\{x_2 \equiv y_2\}] = f_{XY}$$

- 等式表明，一个个体的近交系数等于它的两个亲本间的共祖先系数。即有祖先关联的亲本会引起后代的近交。

利用四个亲本的共祖先系数估计

- ▶ 个体间的共祖先系数为亲本间共祖先系数的平均。



$$f_{XY} = \frac{1}{4} [f_{AC} + f_{AD} + f_{BC} + f_{BD}]$$

利用个体和另一个体的两个亲本的共祖先系数

➤ f_{XY} 是X与Y的两个亲本C和D间的共祖先系数的平均。

$$f_{XY} = \frac{1}{2} [f_{AY} + f_{BY}]$$

▪ 条件：Y不是X的后代

➤ f_{XY} 是Y与X的两个亲本A和B间的共祖先系数的平均。

$$f_{XY} = \frac{1}{2} [f_{CX} + f_{DX}]$$

▪ 条件：X不是Y的后代

如果Z是X和Y的后代，那么X一定不是Z的后代

$$f_{XZ} \neq \frac{1}{2} [f_{AZ} + f_{BZ}]$$

$$f_{XZ} = \frac{1}{2} [f_{XX} + f_{XY}] = \frac{1}{2} \left[\frac{1}{2} (1 + F_X) + f_{XY} \right]$$

共祖先系数的特点

- 亲本X和它的后代Y间的共祖先系数不小于1/4

$$f_{XZ} = \frac{1}{2}[f_{XX} + f_{XY}] = \frac{1}{2}\left[\frac{1}{2}(1 + F_X) + f_{XY}\right] \geq \frac{1}{4}$$

- 全同胞间的共祖先系数不小于1/4

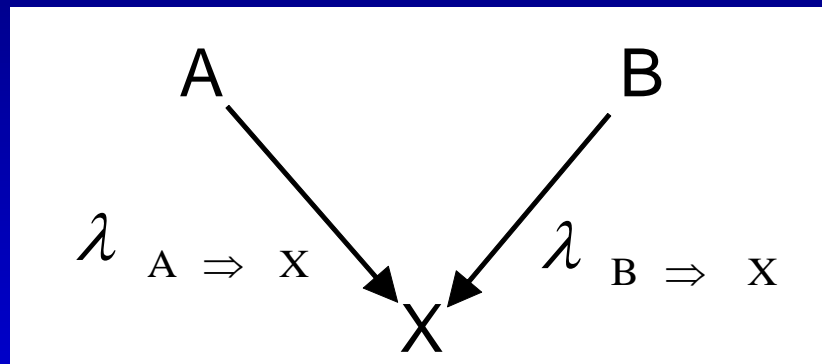
$$f_{XY} = \frac{1}{4}[f_{AA} + f_{AB} + f_{BA} + f_{BB}]$$

$$= \frac{1}{4}\left[\frac{1}{2}(1 + F_A) + 2f_{AB} + \frac{1}{2}(1 + F_B)\right] \geq \frac{1}{4}$$

- 半同胞间的共祖先系数不小于1/8

亲本贡献（仅限于植物）

- 亲本贡献 (Parental contribution) 的定义
- 亲本对自交系后代的贡献



- 亲本贡献可以用分子标记数据来估计

$$\lambda_{A \Rightarrow X} = \frac{S_{AX} - S_{BX}S_{AB}}{1 - (S_{AB})^2}$$

$$\lambda_{B \Rightarrow X} = \frac{S_{BX} - S_{AX}S_{AB}}{1 - (S_{AB})^2}$$

利用分子标记估计亲本贡献

➤几个自交系间共同谱带的比例

自交系	P ₁	P ₂	LBC1-1	LBC1-2	LF2
P ₁	1	0.498	0.775	0.813	0.702
P ₂	0.498	1	0.600	0.603	0.723
LBC1-1	0.775	0.600	1	0.833	0.712
LBC1-2	0.813	0.603	0.833	1	0.654
LF2	0.702	0.723	0.712	0.654	1

亲本P1和P2对几个自交系后代的亲本贡献

自交系	系谱信息		124 个标记的信息		
	$\lambda_{P1 \rightarrow X}$	$\lambda_{P2 \rightarrow X}$	$\lambda_{P1 \rightarrow X}$	$\lambda_{P2 \rightarrow X}$	总和
LBC1-1	0.75	0.25	0.63	0.28	0.91
LBC1-2	0.75	0.25	0.68	0.26	0.94
LF2	0.25	0.50	0.45	0.50	0.95

根据自交系的系谱估计亲本系数

- 假定有 n 个自交系，把自交系按系谱中的先后排序，近交系数和亲本系数的计算从基础群体开始。

$$f_{ij} = f_{ji} = \begin{cases} 0, & 1 \leq i < j \leq p \\ \frac{1}{2}(1 + F_i) = 1, & 1 \leq i = j \leq p \end{cases}$$

根据自交系的系谱估计亲本系数

- 依次计算自交系 i ($i > p$) 和自交系 1、2、...、 i 间的亲本系数，自交系 i 的亲本为 g 和 h 。

$$f_{ij} = f_{ji} = \begin{cases} \frac{1}{2} f_{gj} + \frac{1}{2} f_{hj}, & p < j < i \\ \frac{1}{2} (1 + F_i) = 1, & p < j = i \end{cases}$$

根据动物系谱估计亲本系数

- 假定有 n 个个体 $i > p$ ，把个体按系谱中的先后排序，近交系数和亲本系数的计算从基础群体开始。

$$f_{ij} = f_{ji} = \begin{cases} 0, & 1 \leq i < j \leq p \\ \frac{1}{2}(1 + F_i) = \frac{1}{2}(1 + 0) = \frac{1}{2}, & 1 \leq i = j \leq p \end{cases}$$

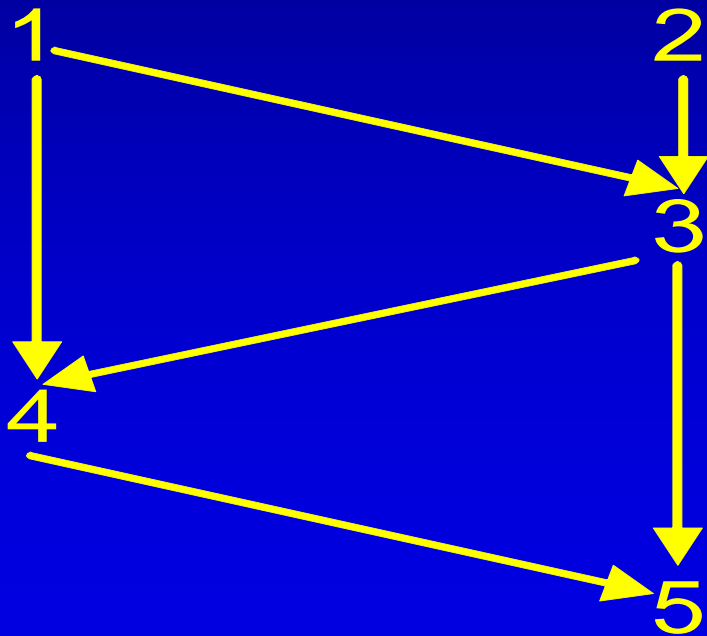
根据动物系谱估计亲本系数

- 依次计算个体 i ($i > p$) 和个体1、2、...、 i 间的亲本系数
- 设个体 i 的亲本为 g 和 h ，由于个体按先后次序排列，因此有， $g > i$ ， $h > i$

$$f_{ij} = f_{ji} = \begin{cases} \frac{1}{2} f_{gj} + \frac{1}{2} f_{hj}, & p < j < i \\ \frac{1}{2} (1 + F_i) = \frac{1}{2} (1 + f_{gh}), & p < j = i \end{cases}$$

一个有5个非自交个体的系谱中共祖先系数矩阵

- 一个包含有5个个体的系谱如下图：
- 个体1和2的来源未知



$$F_1 = F_2 = f_{12} = f_{21} = 0$$

$$f_{11} = \frac{1}{2}(1 + F_1) = \frac{1}{2}$$

$$f_{22} = \frac{1}{2}(1 + F_2) = \frac{1}{2}$$

个体3的亲本为个体1和2

$$f_{13} = f_{31} = \frac{1}{2} f_{11} + \frac{1}{2} f_{21} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0 = \frac{1}{4}$$

$$f_{23} = f_{32} = \frac{1}{2} f_{12} + \frac{1}{2} f_{22} = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$f_{33} = \frac{1}{2} (1 + F_3) = \frac{1}{2} (1 + f_{12}) = \frac{1}{2}$$

个体4的亲本为个体1和3

$$f_{14} = f_{41} = \frac{1}{2} f_{11} + \frac{1}{2} f_{31} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{4} = \frac{3}{8}$$

$$f_{24} = f_{42} = \frac{1}{2} f_{12} + \frac{1}{2} f_{32} = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

$$f_{34} = f_{43} = \frac{1}{2} f_{13} + \frac{1}{2} f_{33} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{8}$$

$$f_{44} = \frac{1}{2} (1 + F_4) = \frac{1}{2} (1 + f_{13}) = \frac{1}{2} (1 + \frac{1}{4}) = \frac{5}{8}$$

个体5的亲本为个体3和4

$$f_{15} = f_{51} = \frac{1}{2} f_{31} + \frac{1}{2} f_{41} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{8} = \frac{3}{16}$$

$$f_{25} = f_{52} = \frac{1}{2} f_{32} + \frac{1}{2} f_{42} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{8} = \frac{3}{16}$$

$$f_{35} = f_{53} = \frac{1}{2} f_{33} + \frac{1}{2} f_{43} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{8} = \frac{7}{16}$$

$$f_{45} = f_{54} = \frac{1}{2} f_{34} + \frac{1}{2} f_{44} = \frac{1}{2} \times \frac{3}{8} + \frac{1}{2} \times \frac{5}{8} = \frac{1}{2}$$

$$f_{55} = \frac{1}{2} (1 + F_5) = \frac{1}{2} (1 + f_{34}) = \frac{1}{2} (1 + \frac{3}{8}) = \frac{11}{16}$$

亲本系数矩阵

	1	2	3	4	5
1	0.5	0	0.25	0.375	0.1625
2		0.5	0.25	0.125	0.1625
3			0.5	0.125	0.0625
4				0.625	0.5
5					0.6875

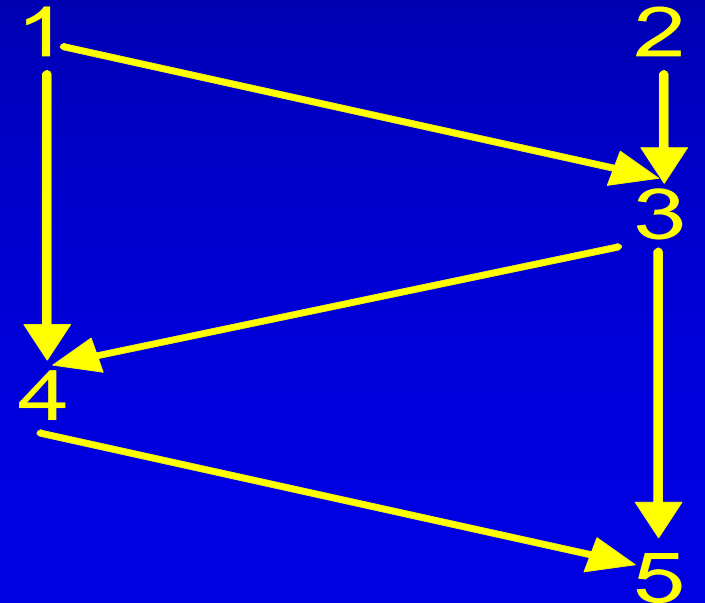
个体的近交系数等于 2个亲本的亲本系数

$$f_{13} = f_{31} = \frac{1}{4} \quad f_{44} = \frac{1}{2}(1 + F_4) = \frac{5}{8}$$

$$F_4 = 2f_{44} - 1 = \frac{1}{4} = f_{13}$$

$$f_{34} = f_{43} = \frac{3}{8} \quad f_{55} = \frac{11}{16}$$

$$F_5 = 2f_{55} - 1 = \frac{3}{8} = f_{34}$$



第三章内容要点

- 随机飘移 (random drift)
- 有限大小群体中的近交
- 近交和近交系数
- 理想群体和有效群体大小
- 有效群体大小与遗传资源的保护
- 后裔同样基因
- 近交系数和亲本系数
- 亲本系数矩阵的计算