

# 第4章

## 有效群体大小和系谱分析

王建康

中国农业科学院作物科学研究所

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

<http://www.isbreeding.net>

# 一般群体的亲缘关系研究方法

- 理想群体需要满足一系列条件，对于大多数动植物育种群体来说，这些条件不一定都能够得到满足。但是，通过有效群体大小这一概念，可以把一个非理想群体看作理想群体进行研究。
- 另外，对于系谱关系明确或者按照特殊规则进行交配的群体，可以直接计算个体间共祖先系数和它们后代的近交系数。

# 本章的主要内容

- § 4.1 非理想群体的有效大小
- § 4.2 系谱群体中的共祖先系数和近交系数
- § 4.3 规则近交交配系统的近交系数
- § 4.4 群体遗传学在植物遗传资源保护中的应用

# § 4.1 非理想群体的有效大小

- § 4.1.1 有效群体大小的定义
- § 4.1.2 常见非理想情况下的有效群体大小
- § 4.1.3 建立者效应和瓶颈效应

# 非理想群体的有效大小

- 大多数动植物育种群体都不是理想群体，为了能够像理想群体那样，用基因频率的方差或是近交系数去研究这些群体的分散过程，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。有了这一概念，一个有效大小为 $N_e$ 的非理想群体，就可视为包含 $N_e$ 个个体的理想群体。
- 因此，在知道了有效大小后，非理想群体中分散过程造成的基因和基因型频率波动、以及近交系数的变化就与大小为 $N_e$ 的理想群体完全相同。

# 近交有效群体大小

- 例如，从一个育种群体的产生过程，有时容易计算每个世代的近交系数和近交系数的增长速率 $\Delta F$ ，这时的有效群体大小可以计算为：

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

- 上式利用近交系数估计出的有效群体大小，又称为近交有效群体大小（inbreeding effective population size）。

# 方差有效群体大小

- 下面的公式给出利用基因频率方差估计有效群体大小的方法，这样的 $N_e$ 又称为方差有效群体大小（variance effective population size）。

$$F_t = \frac{\sigma_{p_t}^2}{p_0 q_0} \quad \Delta F = \frac{1}{2N_e} = 1 - (1 - F_t)^{\frac{1}{t}}$$

# 群体有效大小为 $N_e$ 的随机漂移

- 有时，对一些已知育种结构的群体，有效群体大小 $N_e$ 还能够由实际群体大小计算出来。
- 在计算出一个群体的有效大小后，近交系数和基因频率方差与大小为 $N_e$ 的理想群体相同。即把第3章中的 $N$ 用 $N_e$ 代替即可。

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



# 一个随机交配大群体经过8个世代遗传漂移衍生的10个亚群体的基因频率

$$F_t = \frac{\sigma_{p_t}^2}{p_0 q_0}$$

亚群体	等位基因频率	杂合度
1	0.463	0.4973
2	0.565	0.4916
3	0.882	0.2082
4	0.721	0.4023
5	0.735	0.3896
6	0.839	0.2702
7	0.843	0.2647
8	0.910	0.1638
9	0.746	0.3790
10	0.358	0.4597

- 平均基因频率为0.7062，可以视为基础群体的基因频率。亚群体间基因频率的方差为0.0312，根据右上方公式得到亚群体的平均近交系数为  $F=0.1503$ 。因此得到  $\Delta F$  和  $N_e$  分别为：

$$\Delta F = 1 - (1 - F)^{\frac{1}{t}} = 0.0201$$

$$N_e = 24.82$$

# 一个随机交配大群体经过8个世代遗传漂移衍生的10个亚群体的基因频率

亚群体	等位基因频率	杂合度
1	0.463	0.4973
2	0.565	0.4916
3	0.882	0.2082
4	0.721	0.4023
5	0.735	0.3896
6	0.839	0.2702
7	0.843	0.2647
8	0.910	0.1638
9	0.746	0.3790
10	0.358	0.4597

- 近交系数还可以从相对于基础群体杂合度的下降方面进行估计。
- 基础群体杂合度 =  $2 \times 0.7062 \times (1 - 0.7062) = 0.4150$ 。亚群体平均杂合度 = 0.3526。
- 近交系数  $F = (0.4150 - 0.3526) / 0.4150 = 0.1503$

# 常见遗传/育种群体的有效大小

- 避免近亲交配的群体， $N$ 为实际大小

– 排除自交

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

– 排除自交和同胞交配

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

# 常见遗传/育种群体的有效大小

- 雌雄个体数不等的群体

$$\frac{1}{N_e} \approx \frac{1}{4N_m} + \frac{1}{4N_f} \quad N_e \approx \frac{4N_m N_f}{N_m + N_f} \quad \Delta F \approx \frac{1}{8N_m} + \frac{1}{8N_f}$$

- 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。
- 极端的例子是有无穷多个雌性个体但只有一个雄性个体（如动物中，一个雄性个体与大量雌性个体交配产生的后代群体），或者有无穷多雄性个体但只有一个雌性个体（如异花授粉植物中，一个植株上产生的大量种子），这时有效群体大小  $N_e$  大约只等于4。

# 常见遗传/育种群体的有效大小

- 世代间个体数不等的群体

- 假定世代 $t$ 的群体大小为 $N_t$  ( $t=0, 1, 2, \dots$ )，近交系数在上下代之间的关系是：

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

- 世代 $t$ 与世代0的近交系数关系是：

$$1 - F_t = \left(1 - \frac{1}{2N_{t-1}}\right) \cdots \left(1 - \frac{1}{2N_1}\right) \left(1 - \frac{1}{2N_0}\right) (1 - F_0)$$

# 常见遗传/育种群体的有效大小

- 世代间个体数不等的群体

- 有效群体大小 $N_e$ 为：

$$\left(1 - \frac{1}{2N_e}\right)^t = \left(1 - \frac{1}{2N_{t-1}}\right) \cdots \left(1 - \frac{1}{2N_1}\right) \left(1 - \frac{1}{2N_0}\right)$$

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

- 调和平均数受最小一个数值的影响最大。因此，具有最小个体数的世代对有效群体大小起决定作用。这种现象就是后面将要提及的瓶颈效应和建立者效应。

# 理想群体中亲代对子体的贡献

- 对于理想群体，我们假定每个个体为下一代平均贡献一个后代，或两个配子，以保持群体大小代代相同。但是，由于随机交配的缘故，有些亲本可能贡献多于1个的后代，有的亲本可能没有后代。没有后代的个体，它们携带的基因也就不能传递给后续的世代。

# 后代贡献的随机性

- 每个亲本个体贡献的后代个体数是一个随机变量，服从二项分布  $B(n=N, p=1/N)$ 。因此，作为随机变量，每个亲本贡献后代数的均值等于1，方差等于  $npq=1-1/N$ 。当亚群体容量较大的时候，这个随机变量近似服从  $\lambda=1$  的泊松分布，其均值等于1、方差也等于1。
- 如把亲本对后代的平均贡献看做两个配子的话，每个亲本贡献的配子数  $k$  是随机变量，近似服从  $\lambda=2$  的泊松分布，其均值等于2、方差也等于2。



# 家系大小非随机分布的群体

- 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。
- 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。

# 家系大小非随机分布的有效群体大小

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

- 理想群体中，家系大小这一随机变量 $k$ 可视为 $\lambda=2$ 的泊松（Poisson）分布，这时 $E(k)=V(k)=2$ ， $N_e=N$ 。
- 从这一公式还可以看出，如果家系大小的方差为0时， $N_e=2N$ ，即有效群体大小有时还可以是真实群体大小的两倍。

# 雌雄亲本贡献存在差异的情形

- 对于雌雄异体的交配，如大多数的动物物种，雌雄亲本贡献的配子数还可能有不同的方差。这时，可以利用下面的公式近似计算有效群体大小，其中， $V_{km}$ 和 $V_{kf}$ 分别是雌雄亲本贡献配子数的方差。

$$N_e \approx \frac{8N}{V_{km} + V_{kf} + 4}$$

# 建立者效应

- 从前面的公式可以看出，个体数最少的世代对有效群体大小的影响最大。当一个群体是由少数几个个体繁衍而来时，即繁衍过程中的某个阶段只有少数几个个体，这时的有效群体就会很小。
- 例如，某海岛上的人群，是由另一海岛或大陆上的少数几个人首先聚居，后来的群体是由这几个个体交配繁殖而来，那么，这几个祖先个体中的基因频率就决定了海岛群体的基因频率。由于随机漂移，海岛群体的遗传特性可能与原始群体存在较大差异。这种差异是由少数几个建立者造成的，因此也称为建立者效应（founder effect）。

# 瓶颈效应

- 另一情况下，少数个体在自然或人工的影响下（如自然灾害）存活下来，之后这些个体繁衍成一个新群体。当然新群体的基因频率就可能和原来的群体大不相同，而且基因的多样性也不丰富，交配产生新重组类型的可能性也不大，这种效应称为瓶颈效应（bottle-neck effect）。
- 建立者效应和瓶颈效应，归根结底是由小样本的抽样误差引起的。它们不仅降低后代群体的有效大小，提高了后代群体的近交系数，并且在很大程度上决定了后代群体在很多方面的遗传特性，包括等位基因个数的减少、基因频率的奇异分离、以及连锁不平衡度的增加等等。

# 建立者效应和瓶颈效应的细微差异

- 例如，最初从一陆地群体向一无人海岛迁入30人，经过几百年的繁衍，海岛上的人口增加到现今的10,000人。但是，这些人的祖先都是当初迁入的那30人，他们是陆地群体的一个有限样本，有着与陆地群体不同的基因频率。目前的海岛群体与陆地群体之间可能存在很大的遗传差异，这就是建立者效应。
- 又如，一个海岛上的人口一段时间内一直稳定在10,000人左右。后来由于一场突发灾难，如火山或地震等，人口迅速下降到30人。经过一段时间的恢复后，人口又发展到灾难前的水平。现在的人数与灾难前差不多，但是目前人群的遗传多样性要远低于灾难前，这就是瓶颈效应。

# § 4.2 系谱群体中的 共祖先系数和近交系数

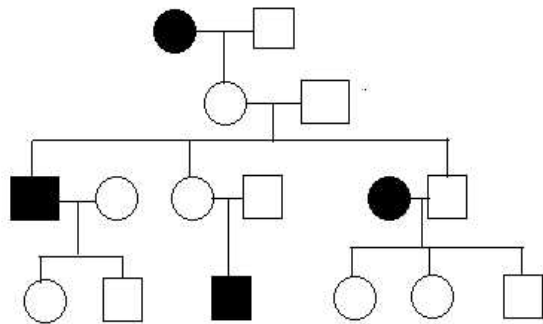
- § 4.2.1 系谱群体
- § 4.2.2 共祖先系数和近交系数的关系
- § 4.2.3 常见系谱的共祖先系数和近交系数
- § 4.2.4 系谱群体共祖先系数的列表计算

# 利用系谱计算亲缘关系和近交

- 前面介绍的是利用群体大小或有效群体大小，计算近交系数，研究群体中基因和基因型频率的波动。近交系数等于任意个体是后裔同样纯合基因型的概率，这样定义的近交系数其实是一个群体中所有个体近交系数的平均。
- 在实际的动植物育种群体中，个体之间存在复杂的系谱关系，这时难以估计群体的有效大小，试图通过有效群体大小研究近交系数就十分困难。但是，如果知道个体之间的系谱信息，个体间的共祖先系数及每个个体的近交系数也是可以计算的。

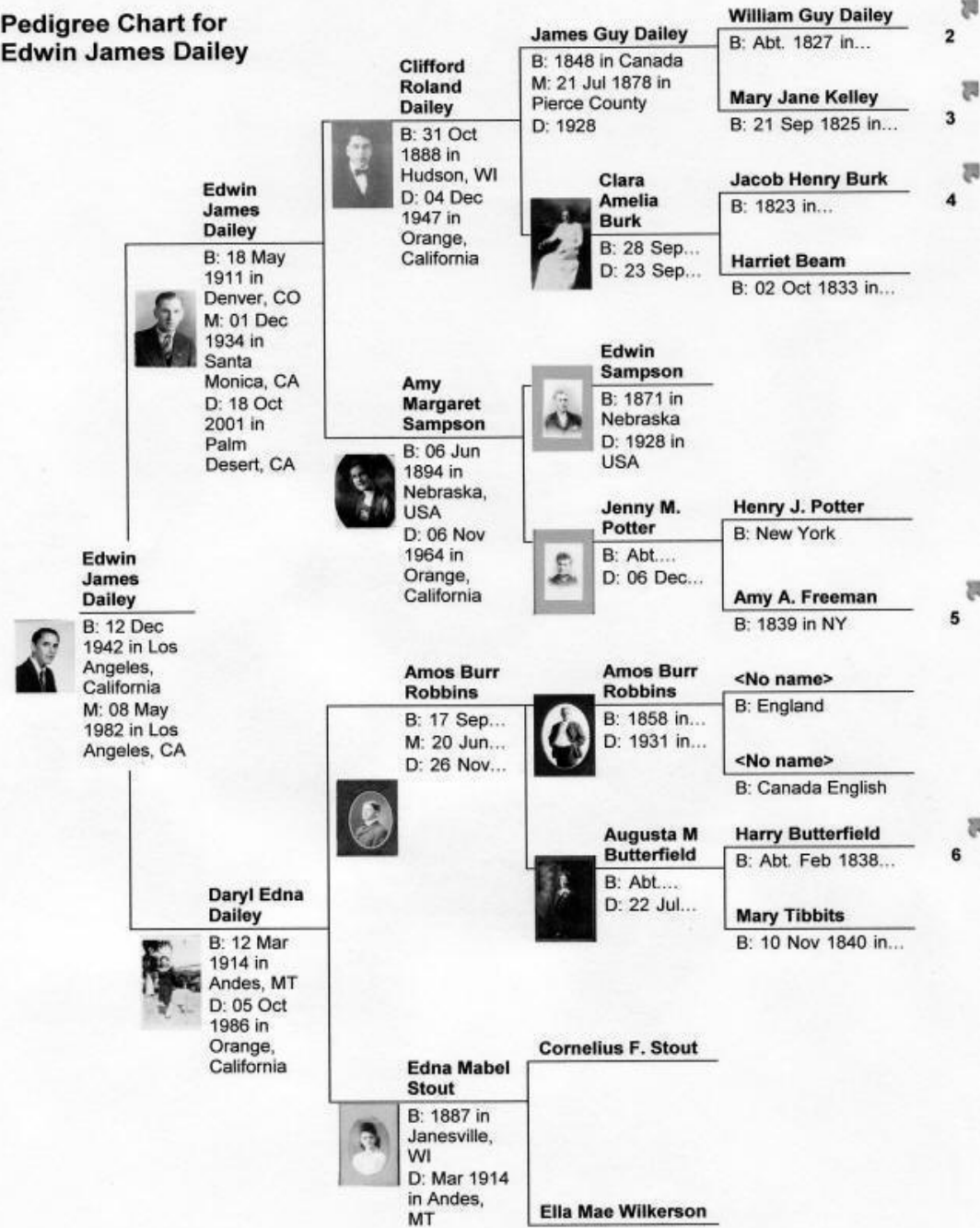


# 人类群体的家谱

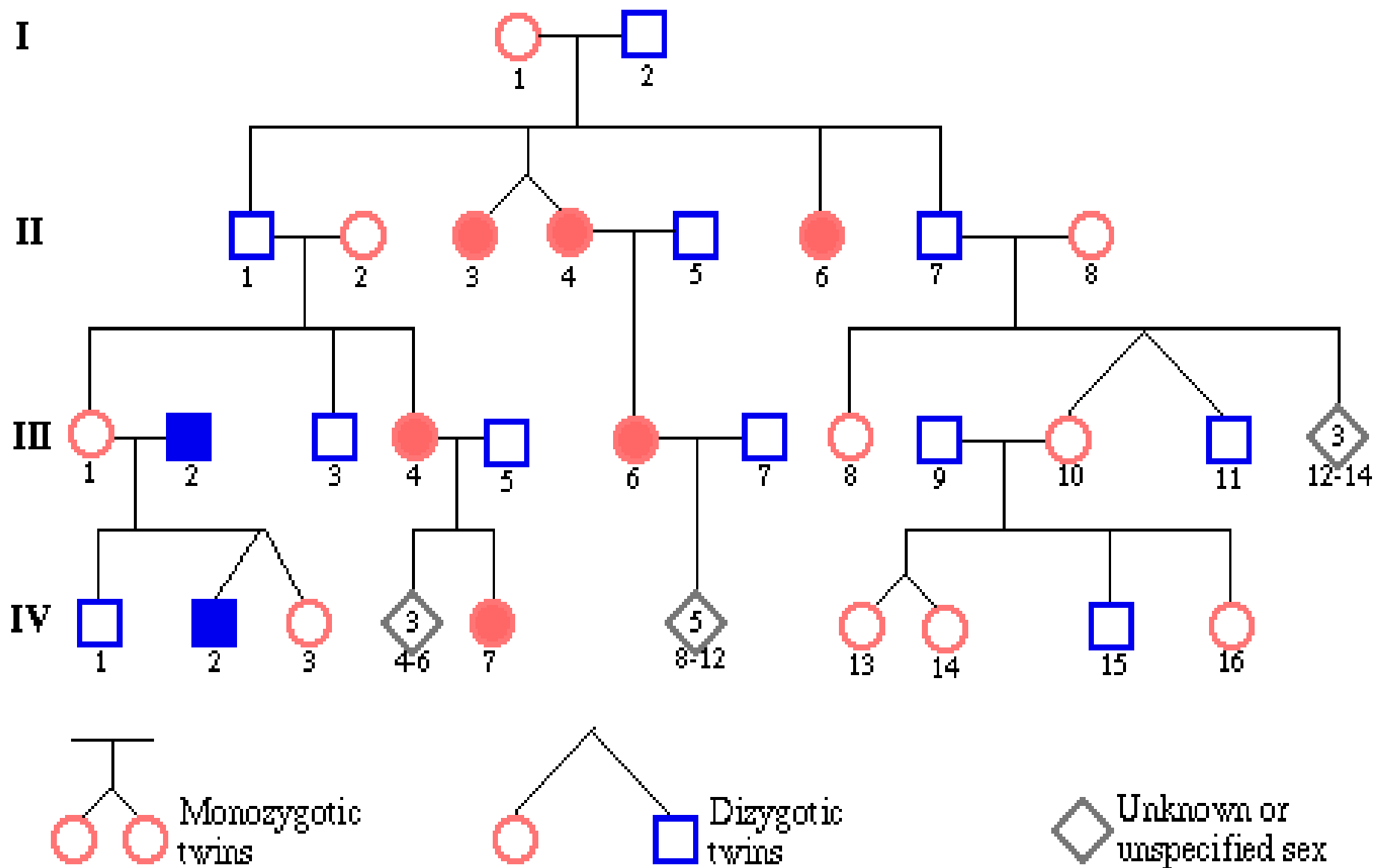


□ = male  
○ = female

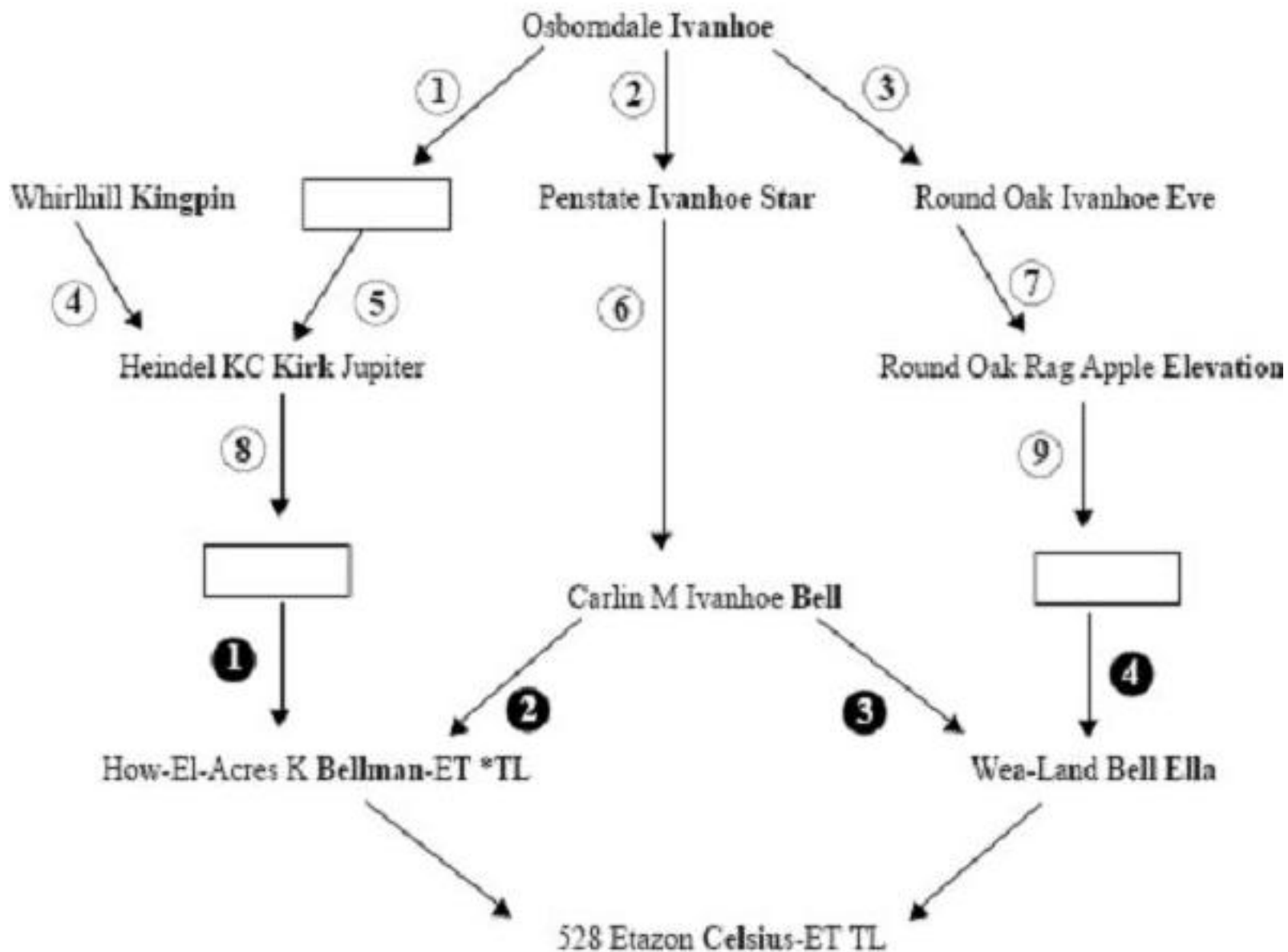
## Pedigree Chart for Edwin James Dailey



# 人类群体系谱的表示方法

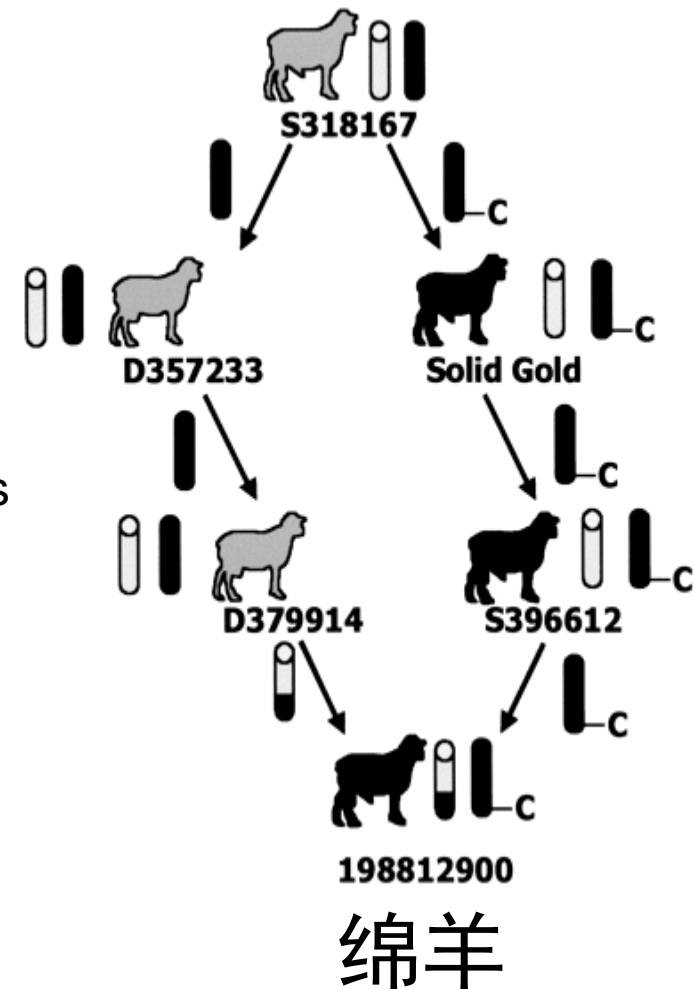


# 包含有近交的牛群系谱图



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

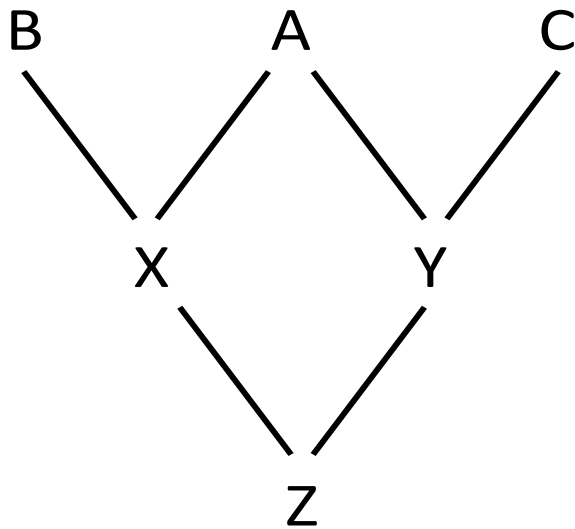
- Ram 198812900 exhibits an inbreeding path with the sire (S318167) of Solid Gold contributing on both maternal and paternal sides of the pedigree. We hypothesized that this inbreeding path has allowed the region to be identical-by-descent, except for the *CLPG* mutation, which we propose occurred in 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



# 半同胞个体X与Y交配的系谱图

- 假定个体B、A、C之间无祖先关联，个体Z的两个亲本X和Y具有共同祖先A。共同祖先A的基因型用 $a_1a_2$ 表示，它的近交系数用 $F_A$ 表示。

$$A: a_1a_2 \rightarrow 1/2 a_1 + 1/2 a_2$$



$$F_Z = \frac{1}{16} + \frac{1}{16} + \frac{1}{8} F_A = \left(\frac{1}{2}\right)^3 (1 + F_A)$$
$$= \left(\frac{1}{2}\right)^n (1 + F_A)$$

$n$ 表示路径XAY上的亲本个数

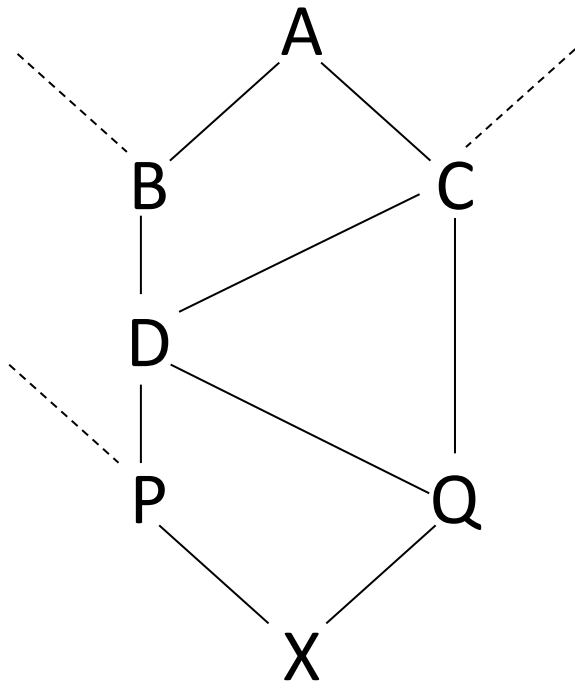
$$1/16 a_1a_1 + 1/8 a_1a_2 + 1/16 a_2a_2 + \dots$$

# 多个共同亲本的近交系数

- 对于具有多个共同亲本（均用A表示）的情形，只需要把每个共同亲本产生的近交系数累加在一起即可，其中， $n$ 表示后裔同样基因传递到Z的路径上的亲本个数。

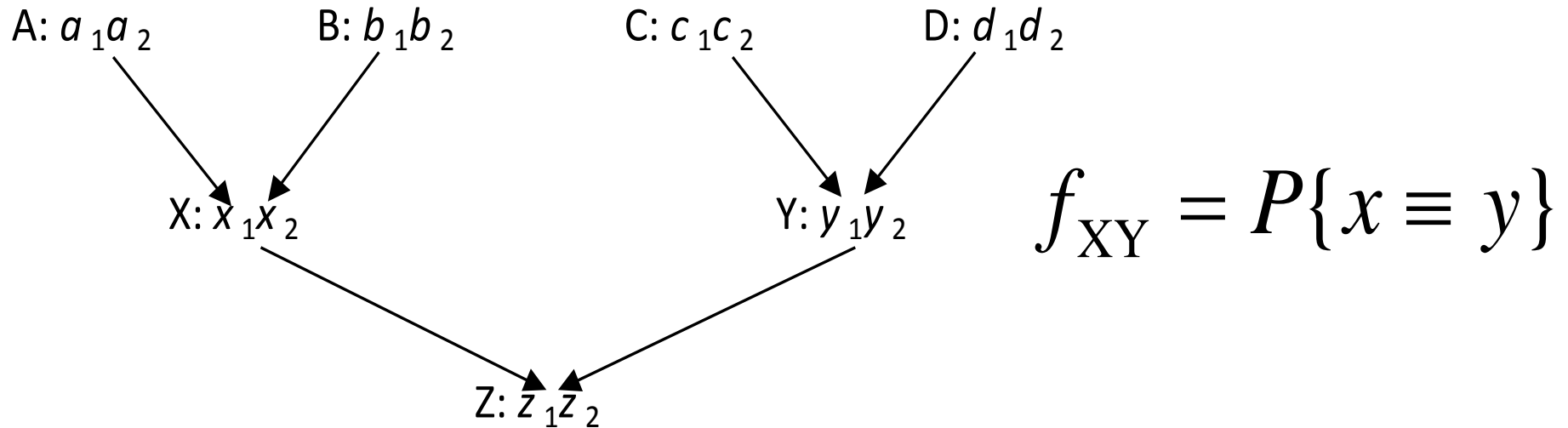
$$F_Z = \sum \left(\frac{1}{2}\right)^n \times (1 + F_A)$$

# 多个共同亲本的近交系数



- $F_A=0$ ,  $F_B=0$ ,  $F_C=0$
- $F_D=(1/2)^3=1/8$  (只有BAC一条路径)
- $F_P=0$ ,  $F_Q=(1/2)^4=1/16$  (只有DBAC一条路径)
- X的亲本是P和Q, 有D、C、A三个共同亲本。路径PDQ对近交系数的贡献是 $(1/2)^3(1+F_D)=9/64$ ; 路径PDCQ对近交系数的贡献是 $(1/2)^4=1/16$ ; 路径PDBACQ对近交系数的贡献是 $(1/2)^6=1/64$ 。因此,  $F_X=9/64+1/16+1/64=7/32$ 。

# 一个包含3个世代的系谱



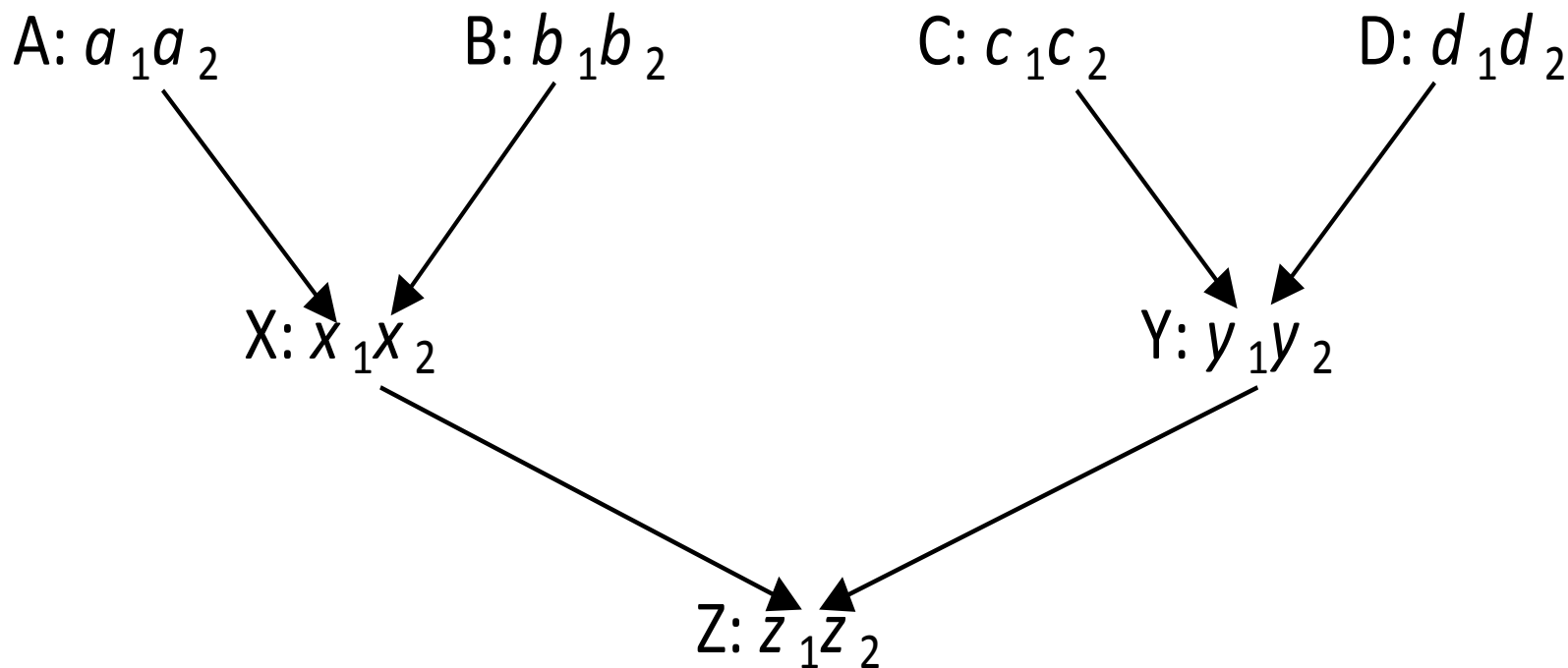
- 假定两个等位基因可以追踪它们的亲本来源。
- 等位基因 $x_1$ 来自亲本A、 $x_2$ 来自亲本B、 $y_1$ 来自亲本C、 $y_2$ 来自亲本D。
- 等位基因 $z_1$ 来自亲本X、 $z_2$ 来自亲本Y。



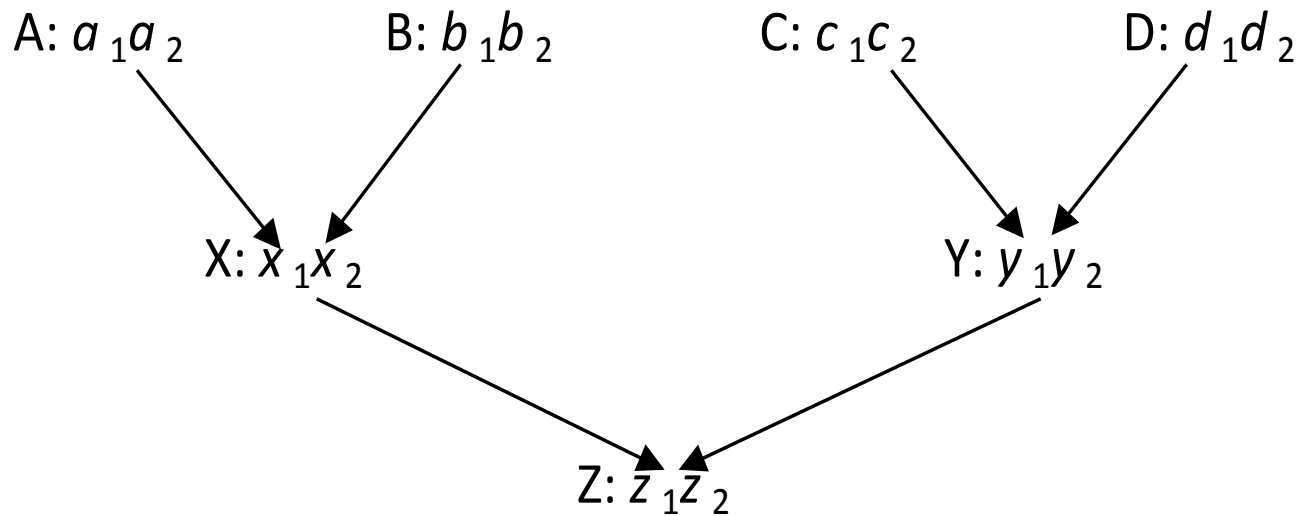
# 事件 $\{x \equiv y\}$ 的分解

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

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



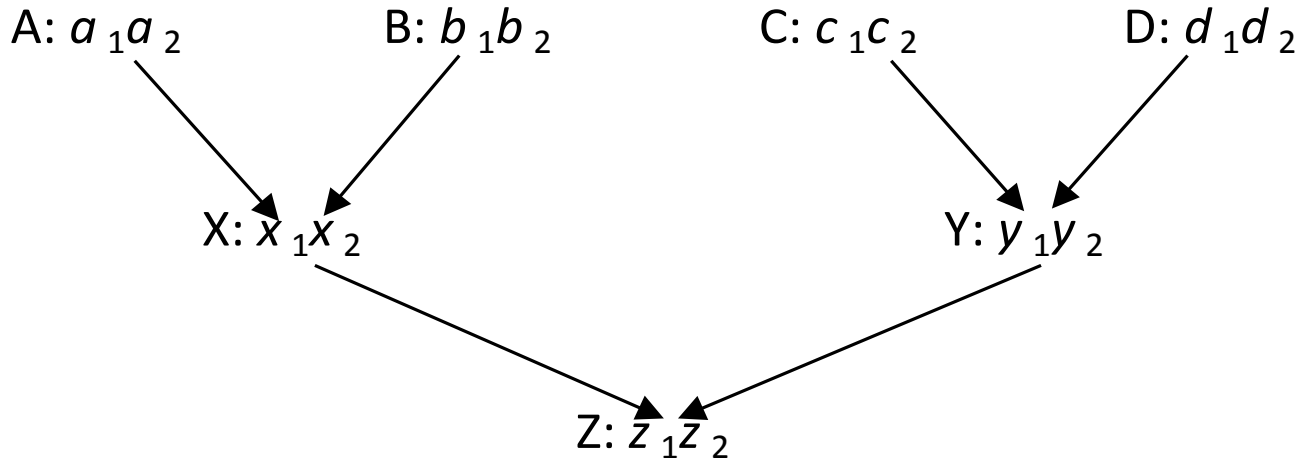
# 共祖先系数的一般表示



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

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

# 利用两个亲本X和Y的共祖先系数估计后代Z的近交系数



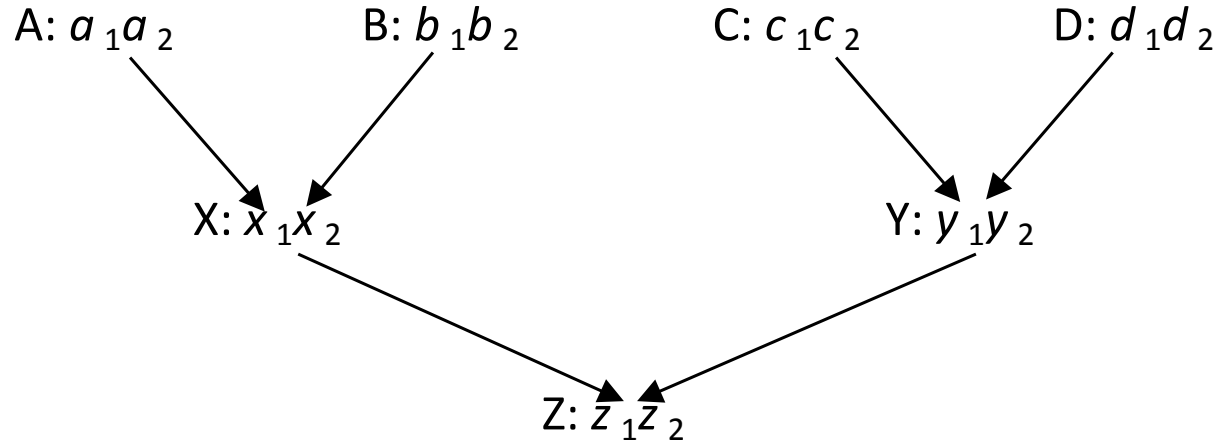
- 事件  $\{x \equiv y\}$  与事件  $\{z_1 \equiv z_2\}$  等价，因此

$$F_Z = \Pr\{z_1 \equiv z_2\} = f_{XY} = \Pr\{x \equiv y\}$$

- 注意：后代Z的近交系数是由亲本X和Y的共祖先系数决定的，而不是两个亲本的近交系数决定的。亲本X和Y的近交系数与后代Z的近交系数没有直接的关系。

# 利用4个祖先亲本A、B、C、D的共祖先系数估计后代Z的近交系数

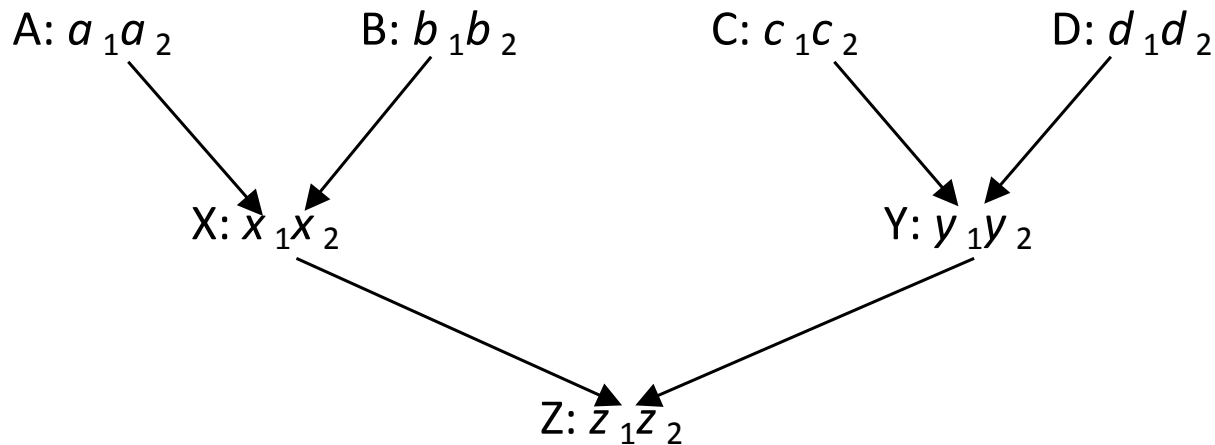
$$f_{XY} = \frac{1}{4} [\Pr\{x_1 \equiv y_1\} + \Pr\{x_1 \equiv y_2\} + \Pr\{x_2 \equiv y_1\} + \Pr\{x_2 \equiv y_2\}]$$



$$F_Z = f_{XY} = \frac{1}{4} (f_{AC} + f_{AD} + f_{BC} + f_{BD})$$

# 利用一个亲本的两个祖先估计 后代Z的近交系数

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



$$F_Z = f_{XY} = \frac{1}{2} (f_{XC} + f_{XD})$$

$$F_Z = f_{XY} = \frac{1}{2} (f_{AY} + f_{BY})$$

# 个体X和它自身的共祖先系数

A和B是X的亲本

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

$$= \frac{1}{4} [1 + F_X + F_X + 1] = \frac{1}{2} (1 + F_X) = \frac{1}{2} (1 + f_{AB}) \geq \frac{1}{2}$$

- 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 $f_{XX}=0.5$ 。
- 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 $f_{XX}=1$ 。

# 亲本X（或Y）和后代Z间的共祖先系数

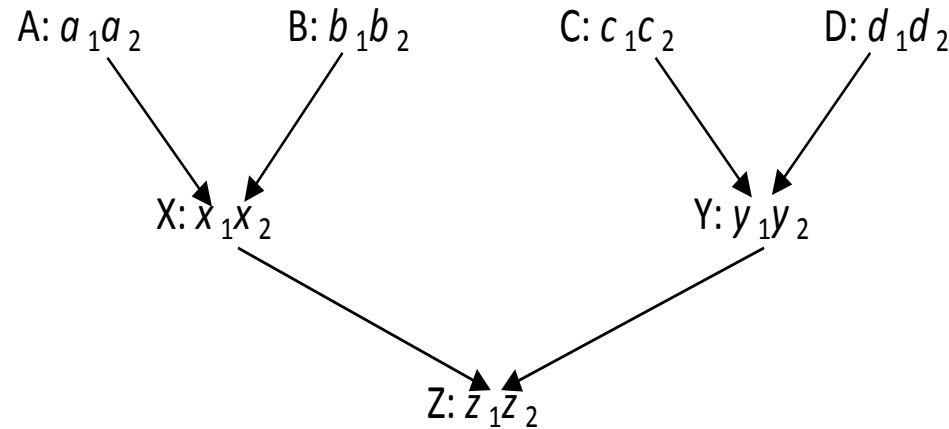
$$f_{XZ} = \frac{1}{2}[f_{XX} + f_{XY}] = \frac{1}{2}\left[\frac{1}{2}(1 + F_X) + f_{XY}\right] \geq \frac{1}{4}$$

$$f_{YZ} = \frac{1}{2}[f_{YX} + f_{YY}] = \frac{1}{2}\left[\frac{1}{2}(1 + F_Y) + f_{XY}\right] \geq \frac{1}{4}$$

- 个体X和后代Z、个体Y和后代Z的共祖先系数都不会低于0.25。
- 如果X和Y均是非自交系，即 $F_X = F_Y = 0$ ，X与Y之间无祖先关联，即 $f_{XY} = 0$ ，这时X和后代Z、以及Y和后代Z的共祖先系数都是0.25。
- 如果X和Y都是自交系，即 $F_X = F_Y = 1$ ， $f_{XZ} = f_{YZ} = (1 + f_{XY})/2$ 。如果X和Y都是自交系，并且X与Y之间无祖先关联，即 $F_X = F_Y = 0$ ， $f_{XY} = 0$ ，这时X和后代Z，以及Y和后代Z的共祖先系数都是0.5。

# 利用一个个体的两个亲本时，要注意

- $f_{XY}$  是X与Y的两个亲本C和D间的共祖先系数的平均。
  - 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。
  - 条件：X不是Y的后代
- 例如：  $f_{XZ} \neq \frac{1}{2}[f_{AZ} + f_{BZ}]$



如果Z是X和Y的后代，那么X一定不是Z的后代。这时，

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



# 全同胞间（A和B是亲本）的共祖先系数

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

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

- 与亲本和后代的关系一样，全同胞个体X和Y的共祖先系数都不会低于0.25。
- 如果亲本A和B均是非自交系，无祖先关联，即  $F_A = F_B = 0$ ， $f_{AB} = 0$ ，则全同胞之间的共祖先系数为0.25。
- 如果亲本A和B均是自交系，无祖先关联，即  $F_A = F_B = 1$ ， $f_{AB} = 0$ ，则全同胞之间的共祖先系数为0.5。

# 半同胞间的共祖先系数

(A和B是X的亲本， A和D是Y的亲本)

$$f_{XY} = \frac{1}{4} [f_{AA} + f_{AD} + f_{BA} + f_{BD}]$$

$$= \frac{1}{4} \left[ \frac{1}{2} (1 + F_A) + f_{AD} + f_{BA} + f_{BD} \right] \geq \frac{1}{8}$$

- 半同胞的共祖先系数都不会低于0.125。
- 如果亲本A、B、D均是非自交系，无祖先关联，则半同胞之间的共祖先系数为0.125。
- 如果亲本A、B、D均是自交系，无祖先关联，则半同胞之间的共祖先系数为0.25。

# 系谱群体共祖先系数的列表计算

## 1. 确定一个基础群体

- 假定一个系谱群体包含 $n$ 个个体，把所有个体按照在系谱中的亲子关系排序，并用 $1, 2, \dots, n$ 对个体进行编号。亲代必需排在子代的前面，排在最前面的 $p$ 个无祖先关联的个体组成一个基础群体。基础群体中的个体，无法再追踪到它们的亲本信息，个体之间有无亲缘关系也不清楚。一般就认为它们之间不存在任何祖先关联。

# 系谱群体共祖先系数的列表计算

## 2. 计算基础群体的共祖先系数

- 共祖先系数的计算从基础群体开始。基础群体中的个体都是非近交的，其近交系数为0；两两个体之间没有任何亲缘关系，共祖先系数也为0。

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

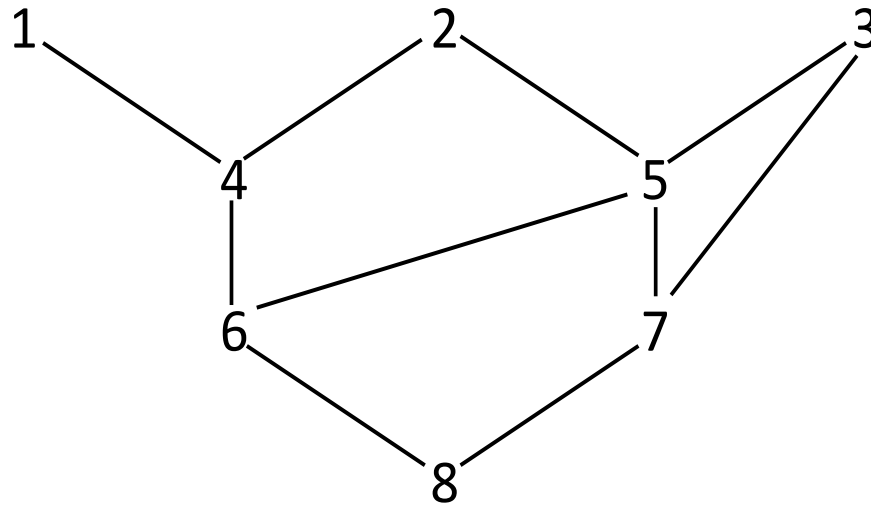
# 系谱群体共祖先系数的列表计算

## 3. 按顺序依次计算其余个体与它前面的个体以及与其自身的共祖先系数

- 基础群体之外的个体，按照在系谱中的先后顺序，依次计算个体 $i$  ( $i > p$ ) 和个体 $1, 2, \dots, i$ 间的共祖先系数。设个体 $i$ 的亲本编号为 $g$ 和 $h$ 。由于个体按先后次序排列，因此一定有 $g < i$ 和 $h < i$ 。

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

# 一个包含有8个个体的系谱



- 个体1、2、3的亲本信息未知，这三个个体构成一个基础群体。个体4的亲本是1和2，个体5的亲本是2和3，它们之间有一个共同亲本，即个体2。个体6的亲本是4和5，个体7的亲本是3和5，个体8的亲本是6和7。显然，个体6和7之间也存在祖先关联，它们的后代个体8也因此存在近交。

# 基础群体中个体之间的共祖先系数

$$f_{11} = \frac{1}{2}(1 + F_1) = \frac{1}{2} \quad f_{12} = 0 \quad f_{13} = 0$$

$$f_{22} = \frac{1}{2}(1 + F_2) = \frac{1}{2} \quad f_{23} = 0 \quad f_{33} = \frac{1}{2}(1 + F_3) = \frac{1}{2}$$

# 个体4的亲本是1和2 它与个体1~4之间的共祖先系数

$$f_{14} = \frac{1}{2} f_{11} + \frac{1}{2} f_{12} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0 = \frac{1}{4}$$

$$f_{24} = \frac{1}{2} f_{21} + \frac{1}{2} f_{22} = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$f_{34} = \frac{1}{2} f_{31} + \frac{1}{2} f_{32} = \frac{1}{2} \times 0 + \frac{1}{2} \times 0 = 0$$

$$f_{44} = \frac{1}{2} (1 + F_4) = \frac{1}{2} (1 + f_{12}) = \frac{1}{2}$$



# 个体5的亲本是2和3 它与个体1~5之间的共祖先系数

$$f_{15} = \frac{1}{2} f_{12} + \frac{1}{2} f_{13} = \frac{1}{2} \times 0 + \frac{1}{2} \times 0 = 0$$

$$f_{25} = \frac{1}{2} f_{22} + \frac{1}{2} f_{23} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0 = \frac{1}{4}$$

$$f_{35} = \frac{1}{2} f_{32} + \frac{1}{2} f_{33} = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$f_{45} = \frac{1}{2} f_{42} + \frac{1}{2} f_{43} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times 0 = \frac{1}{8}$$

$$f_{55} = \frac{1}{2} (1 + F_5) = \frac{1}{2} (1 + f_{23}) = \frac{1}{2}$$

# 个体6的亲本是4和5

## 它与个体1~6之间的共祖先系数

$$f_{16} = \frac{1}{2} f_{14} + \frac{1}{2} f_{15} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times 0 = \frac{1}{8}$$

$$f_{26} = \frac{1}{2} f_{24} + \frac{1}{2} f_{25} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4} = \frac{1}{4}$$

$$f_{36} = \frac{1}{2} f_{34} + \frac{1}{2} f_{35} = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

$$f_{46} = \frac{1}{2} f_{44} + \frac{1}{2} f_{45} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{8} = \frac{5}{16}$$

$$f_{56} = \frac{1}{2} f_{54} + \frac{1}{2} f_{55} = \frac{1}{2} \times \frac{1}{8} + \frac{1}{2} \times \frac{1}{2} = \frac{5}{16}$$

$$f_{66} = \frac{1}{2} (1 + F_6) = \frac{1}{2} (1 + f_{45}) = \frac{1}{2} (1 + \frac{1}{8}) = \frac{9}{16}$$

# 个体7的亲本是3和5 它与个体1~7之间的共祖先系数

$$f_{17} = 0 \quad f_{27} = \frac{1}{8} \quad f_{37} = \frac{3}{8} \quad f_{47} = \frac{1}{16}$$

$$f_{57} = \frac{3}{8} \quad f_{67} = \frac{7}{32} \quad f_{77} = \frac{5}{8}$$

# 个体8的亲本是6和7 它与个体1~8之间的共祖先系数

$$\begin{array}{cccc} f_{18} = \frac{1}{16} & f_{28} = \frac{3}{16} & f_{38} = \frac{1}{4} & f_{48} = \frac{3}{16} \\ f_{58} = \frac{11}{32} & f_{68} = \frac{25}{64} & f_{78} = \frac{27}{64} & f_{88} = \frac{39}{64} \end{array}$$

# 每个个体的近交系数

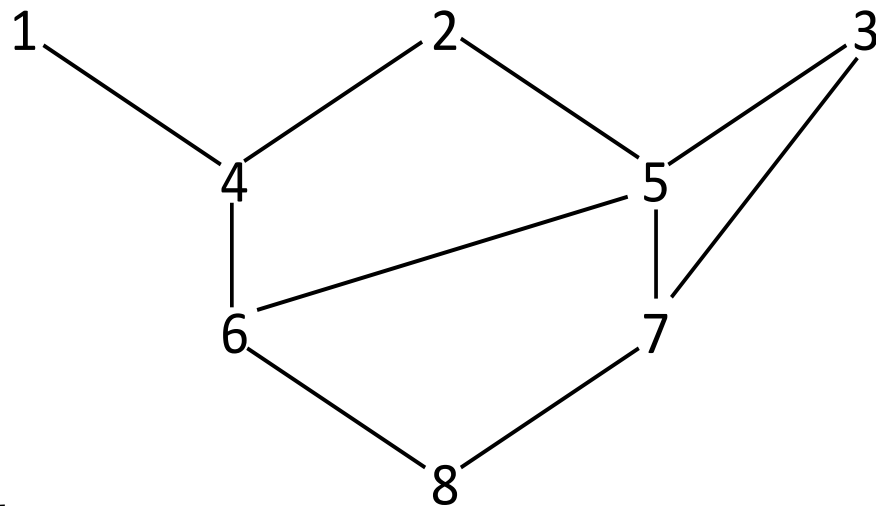
$$f_{XX} = \frac{1}{2}(1 + F_X) \quad F_X = 2f_{XX} - 1$$

$$F_1 = F_2 = F_3 = F_4 = F_5 = 0$$

$$F_6 = \frac{1}{8} \quad F_7 = \frac{1}{4} \quad F_8 = \frac{7}{32}$$

# 个体8近交系数的计算

- 个体2、3、5的近交系数均为0，而且都对个体8的近交系数有贡献。它们通过三条路径将同一个基因传递到个体8。



- 一是8642578（下划线表示亲本），除个体8外，这个路径上包含5个个体；二是865378，除个体8外，这个路径上包含4个个体；三是86578，除个体8外，这个路径上包含3个个体。因此，

$$F_8 = \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^3 = \frac{7}{32}$$

## § 4.3 规则近交交配系统

- § 4.3.1 自交系统的近交系数
- § 4.3.2 回交系统的近交系数
- § 4.3.3 全同胞系统和亲子系统
- § 4.3.4 半同胞系统
- § 4.3.5 混合自交和异交系统
- § 4.3.6 自交系系谱的共祖先系数

# 规则近交交配系统

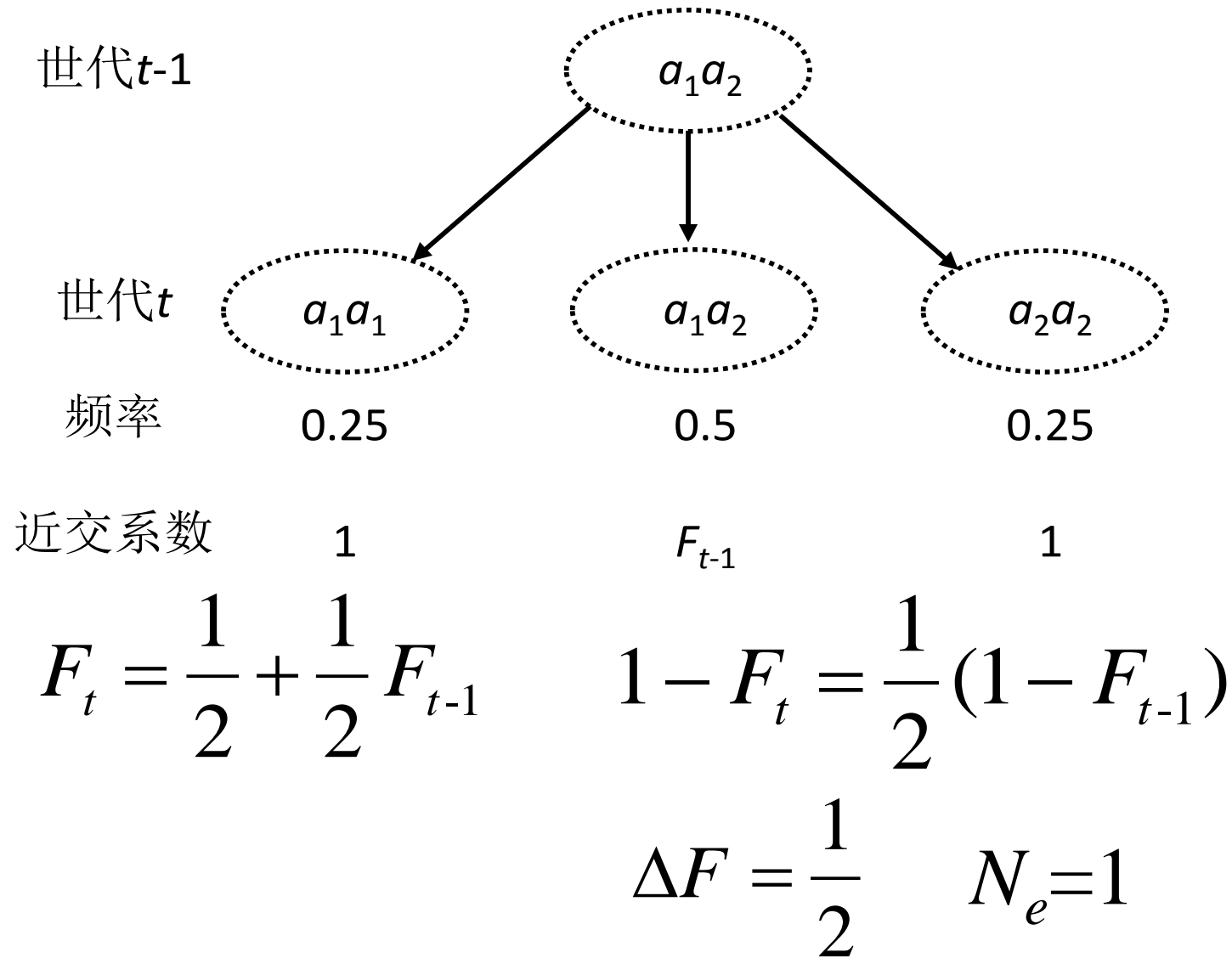
- 随机交配大群体中，具有亲缘关系的个体相互交配产生后代的概率非常低。因此可以认为，随机交配大群体中不存在近交。自交和回交系统中，交配发生在具有亲缘关系的个体之间。因此，自交和回交群体中都存在近交。
- 前面介绍的系谱群体中，同一个世代的不同个体之间可能有不同的共祖先系数，后代个体也可能具有互不相同的近交系数。规则近交交配系统（简称规则近交系统）（regular system of inbreeding）是指在所有世代中重复同一种交配方式，每个世代中，不同个体具有相同的近交系数。



# 规则近交系统的作用

- 规则近交系统是动植物遗传和育种研究中，通过具有亲缘关系个体之间连续不断的交配，快速产生近交家系的常用方法。
- 重复自交和重复回交都属于规则近交系统。此外还有亲子交配、全同胞交配和半同胞交配，是动物遗传育种中经常采用的规则近交系统。

# 自交系统的基因型构成和近交系数



# 自交系统的近交系数变化

- 如果基础群体是两个纯系亲本X与Y杂交产生的F<sub>2</sub>群体，X与Y之间的共祖先系数 $f_{XY}$ 等于基础群体的近交系数 $F_0$ ，自交后代的近交系数由公式4.30给出。可见，亲本间的祖先关联会进一步提高自交过程中的近交系数。

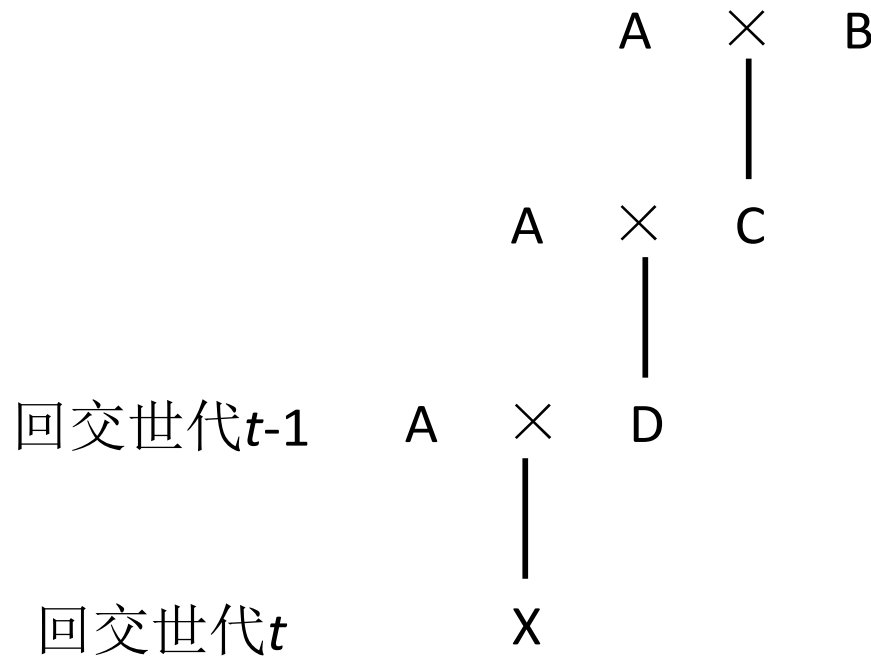
$$F_t = 1 - \left(\frac{1}{2}\right)^t (1 - F_0) = 1 - \left(\frac{1}{2}\right)^t (1 - f_{XY})$$

- 如果双亲X与Y之间无祖先关联，从杂种F<sub>2</sub>群体开始的重复自交过程中，近交系数的变化与任意随机交配群体中的重复自交是一样的，区别在于两个群体的基因频率。纯系双亲杂交产生的基础群体中，两个等位基因的频率各占0.5。任意随机交配群体中，两个等位基因的频率不一定等于0.5。

# 自交系统中近交系数和杂合型频率的变化

群体	近交系数F	杂合型频率
$S_0$	0	H
$S_1$	0.5	0.5H
$S_2$	0.75	0.25H
$S_3$	0.875	0.125H
$S_4$	0.9375	0.0625H
$S_n$	$1-0.5^n$	$0.5^n H$
$S_\infty$	1	0

# 回交系统的基因型构成和近交系数



$$F_X = f_{AD} = \frac{1}{2}(f_{AA} + f_{AC}) = \frac{1}{2}\left[\frac{1}{2}(1 + F_A) + F_D\right]$$

$$F_t = \frac{1}{4}(1 + F_A + 2F_{t-1})$$

# 回交系统中近交系数的变化

$$F_t = \frac{1}{4}(1 + F_A + 2F_{t-1})$$

$$\tilde{F} = \frac{1}{2}(1 + F_A)$$

$$F_t - \tilde{F} = \frac{1}{2}(F_{t-1} - \tilde{F})$$

$$F_t = \tilde{F} + \left(\frac{1}{2}\right)^t (F_0 - \tilde{F}) \quad F_0 = f_{AB}$$

# 回交系统中目的基因染色体长度

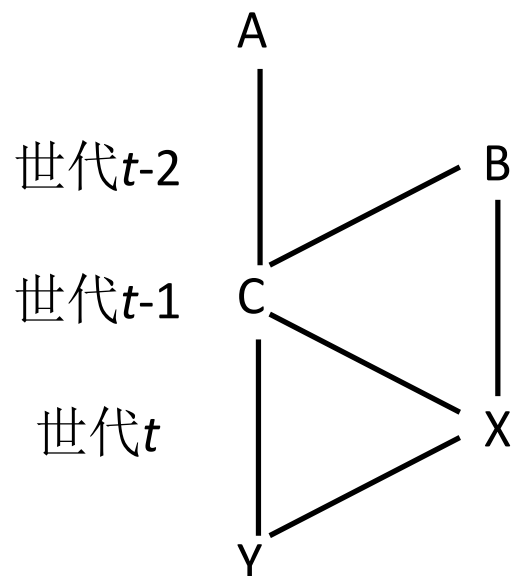
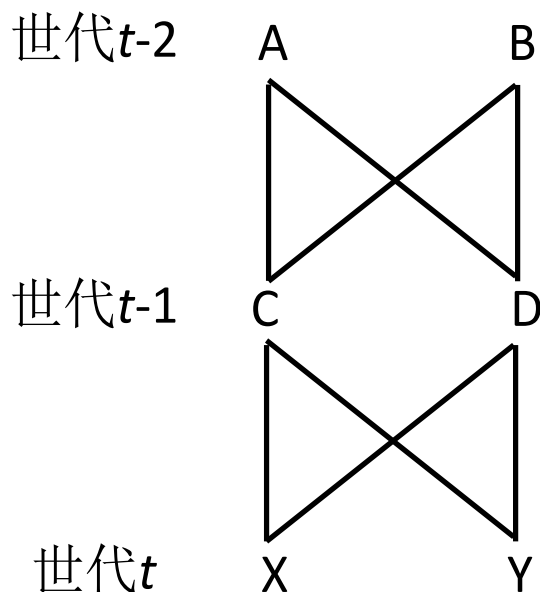
- 对于显性基因控制的性状来说，回交群体中的显性性状个体携带有目的基因，利用这些个体与轮回亲本继续回交即可。可以证明，回交 $t$ 代后，携带显性目的基因的单侧外源片段平均长度为 $100/t$  cM，或双侧片段平均长度为 $200/t$  cM。
- 对于隐性基因控制的性状，回交群体中只有隐性一种表型，需要自交一代后才能选择出携带目的基因的隐性纯合个体，然后与轮回亲本继续回交。Bartlett和Haldane（1935）证明了回交 $t$ 代后，携带隐性目的基因的单侧外源片段平均长度为 $200/t$  cM，或双侧片段平均长度为 $400/t$  cM。

# 回交系统中杂合基因组比例

- 例如，要从一个供体亲本转育一个显性基因，该基因所在染色体的长度为200cM。经过5个世代的回交后，携带目的基因外源片段的平均长度为 $200/5=40\text{cM}$ 。如果整个基因组的长度是1000cM，则外源染色体所占比例为 $40/1000=4\%$ 。
- 在重复回交过程中，外源片段始终处于杂合状态，外源片段的纯合型在自交后才能获得。其他与目的基因无关的基因组，在回交过程中会不断地纯合。根据回交次数计算近交系数 $F$ ，那么 $1-F$ 就可以视为目的基因之外杂合基因组所占的比例。



# 全同胞系统和亲子系统



$$F_X = f_{CD} = \frac{1}{4}(f_{AA} + 2f_{AB} + f_{BB})$$

$$F_X = f_{CB} = \frac{1}{2}(f_{AB} + f_{BB})$$

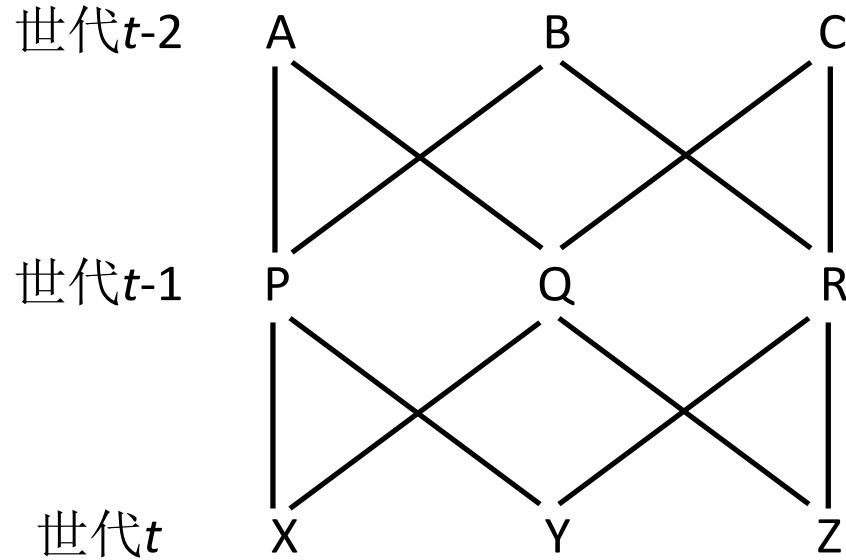
$$F_t = \frac{1}{4}(1 + 2F_{t-1} + F_{t-2})$$

$$F_t = \frac{1}{4}(1 + 2F_{t-1} + F_{t-2})$$

# 全同胞系统和亲子系统的有效群体大小

- 如果世代0的两个亲本A和B是非近交系，并且不存在祖先关联。这时， $F_0=0$ ， $F_1=0$ ， $F_2=0.25$ ， $F_3=0.375$ ， $F_4=0.50$ ， $F_5=0.59$ 。根据公式3.35计算近交系数的变化速率，发现世代5之后的速率就一直稳定在0.191左右，得到 $N_e=2.62$ 。也就是说，全同胞系统的近交过程与大小为2.62的理想群体是等价的，由近交系数得来的 $N_e=2.62$ 属于近交有效群体大小。
- 如果世代0的两个亲本A和B是近交系，但不存在祖先关联。这时， $F_0=1$ ， $F_1=0$ ， $F_2=0.5$ ， $F_3=0.5$ ， $F_4=0.625$ ， $F_5=0.6875$ 。这些近交系数高于非近交亲本的情形，但是世代9之后近交系数的速率仍稳定在0.191左右，有效群体大小也是 $N_e=2.62$ 。

# 半同胞近交系统



$$F_X = f_{PQ} = \frac{1}{4}(f_{AA} + f_{AC} + f_{BA} + f_{BC})$$

$$F_t = \frac{1}{8}(1 + 6F_{t-1} + F_{t-2})$$

# 半同胞近交系统的有效群体大小

- 如果世代0的三个亲本A、B、C是非近交系，并且不存在祖先关联。这时， $F_0=0$ ， $F_1=0$ ， $F_2=0.125$ ， $F_3=0.219$ ， $F_4=0.305$ ， $F_5=0.381$ 。根据公式3.35计算近交系数的变化速率，发现世代5之后的速率稳定在0.11左右，得到近交有效群体大小为 $N_e=4.56$ 。
- 如果世代0的三个亲本A、B、C是近交系，但是不存在祖先关联。这时， $F_0=1$ ， $F_1=0$ ， $F_2=0.25$ ， $F_3=0.313$ ， $F_4=0.391$ ， $F_5=0.457$ 。近交系数高于非近交系亲本，但是世代5之后近交系数的速率仍稳定在0.11左右，近交有效群体大小也是 $N_e=4.56$ 。

# 不同规则近交系统中近交系数的变化

世代	自交I	自交II	回交I	回交II	全同胞I	全同胞II	半同胞I	半同胞II
0	0	0.5	0	0	0	1	0	1
1	0.5	0.75	0.25	0.5	0	0	0	0
2	0.75	0.875	0.375	0.75	0.25	0.75	0.125	0.25
3	0.875	0.9375	0.4375	0.875	0.375	0.875	0.2188	0.3125
4	0.9375	0.9688	0.4688	0.9375	0.5	0.875	0.3047	0.3906
5	0.9688	0.9844	0.4844	0.9688	0.5938	0.9063	0.3809	0.4570
6	0.9844	0.9922	0.4922	0.9844	0.6719	0.9219	0.4487	0.5166
7	0.9922	0.9961	0.4961	0.9922	0.7344	0.9375	0.5092	0.5696
8	0.9961	0.9980	0.4980	0.9961	0.7852	0.9492	0.5630	0.6168
9	0.9980	0.9990	0.4990	0.9980	0.8262	0.9590	0.6109	0.6588
10	0.9990	0.9995	0.4995	0.9990	0.8594	0.9668	0.6535	0.6962

# 基础群体的构建

- 自交系统I的基础群体中，近交系数为0。这样的群体包括任意随机交配大群体、两个非近交系亲本的杂种 $F_1$ 群体、多个非近交系亲本的互交群体等。
- 自交系统II的基础群体中，近交系数为0.5。两个近交系亲本的杂种 $F_2$ 群体（或双向杂交群体）就属于这种类型，它相当于近交系数为0的基础群体中，单个个体的自交一代群体。
- 类似地，如果基础群体是4个近交系相互杂交而成，这样的群体又称为四向杂交群体。这样的基础群体中，近交系数为0.25，相当于近交系数为0的基础群体中两个个体的全同胞后代。

# 基础群体的构建

- 回交系统I中，轮回亲本的近交系数为0，如随机交配大群体的一个个体。回交系统II中，轮回亲本的近交系数为1，如一个纯系。
- 全同胞系统I中，两个初始亲本的近交系数为0，无祖先关联，如随机交配大群体的两个个体。全同胞系统II中，两个初始亲本的近交系数为1，无祖先关联，如两个纯系。
- 半同胞系统I中，三个初始亲本的近交系数为0，无祖先关联，如随机交配大群体的三个个体。半同胞系统II中，三个初始亲本的近交系数为1，无祖先关联，如三个纯系作为起始亲本。

# 不同近交相同的近交系数对比

- 自交系统中，个体之间有着最紧密的亲缘关系，近交系数增长速度也最快，是植物遗传和育种研究中最常用的纯系产生方法。近交系数为1的群体称为纯系，之后的自交繁殖能够维持纯系的基因型不变。
- 值得一提的是，随着技术的进步，有些植物物种中还能通过加倍单倍体技术产生纯系。加倍单倍体携带的两个基因是由配子体的同一个基因复制而来，因此近交系数自然就等于1。显然，利用这样的技术只需要一个世代，近交系数就能达到1。
- 回交系统II的近交系数，同时也可看作是轮回亲本基因组的恢复程度。考虑到选择目标基因时的连锁效应，恢复程度要略低于表中的数字。



# 不同近交相同的近交系数对比

- 全同胞系统I和II，实际上可以看作是等位基因个数的差异。当两个亲本来自一个随机交配大群体时，它们携带了4个非后裔同样基因，可以看作是4个等位基因。当亲本来自两个纯系群体时，它们的后代只携带了二个非后裔同样基因，可以看作只有二个等位基因。
- 等位基因个数的增加，降低了后代个体中两个基因的后裔同样概率，因此也降低了后代的近交系数。同理，半同胞系统I和II也可以看作是等位基因个数的差异，半同胞系统I的三个亲本共携带6个等位基因，而半同胞系统II的三个亲本只携带三个等位基因。

# 混合自交和异交系统

- 在自然条件下，自交繁殖的物种，有时也存在一定比例的异交。例如，利马豆、高粱、棉花等作物以自花授粉为主，但异交率有时可以达到5至50%。同样，异交繁殖的物种，有时也存在一定比例的自交。
- 用 $C$ 表示异交比例，不考虑群体大小的影响，异交个体的近交系数为0。用 $S=1-C$ 表示自交比例。

# 近交系数的上下代关系

- 世代 $t$ 中，
  - 异交的近交系数为0， 比例为 $C$
  - 自交的近交系数等于 $(1+F_{t-1})/2$ ， 比例为 $(1-C)$

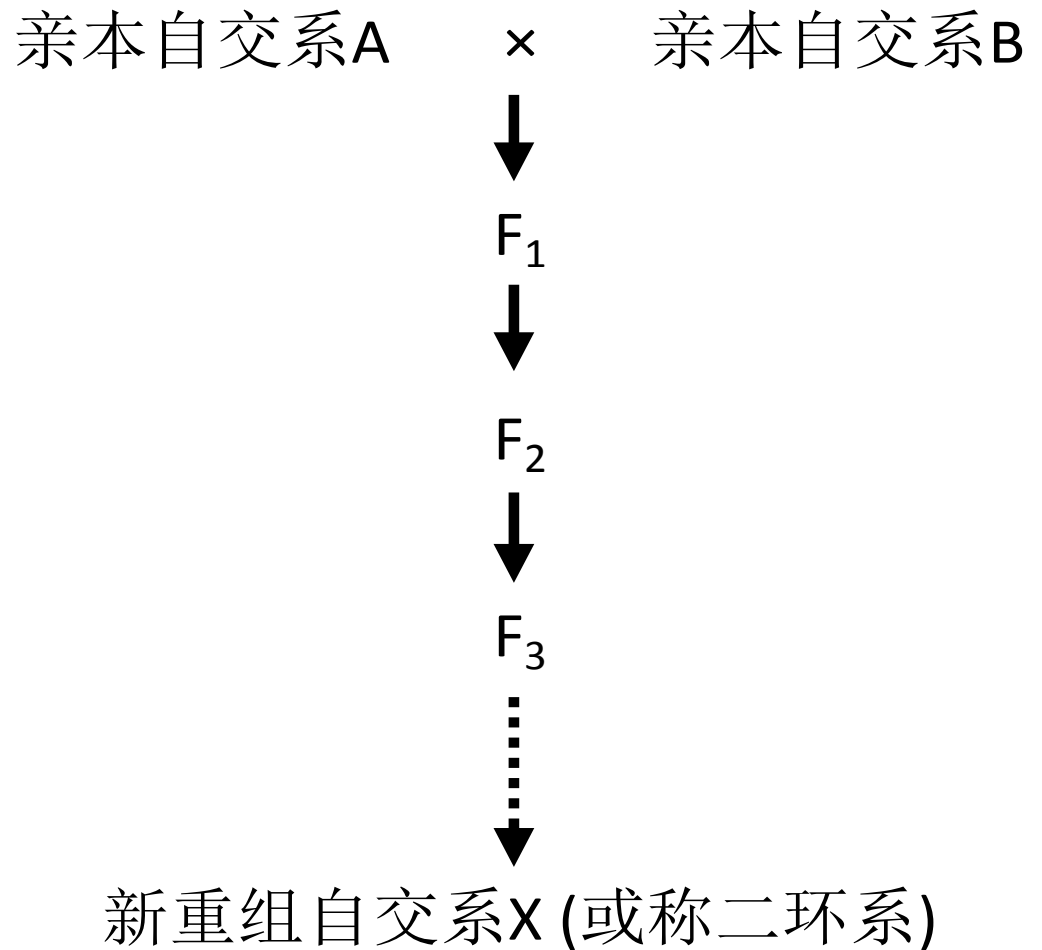
$$F_t = 0 \times C + \frac{1}{2}(1 + F_{t-1})(1 - C) = \frac{1}{2}(1 + F_{t-1})(1 - C)$$

- 平衡近交系数为（ $S$ 为自交率）

$$\tilde{F} = \frac{1 - C}{1 + C} = \frac{S}{2 - S}$$

# 双亲衍生重组自交系的系谱 (仅限于植物)

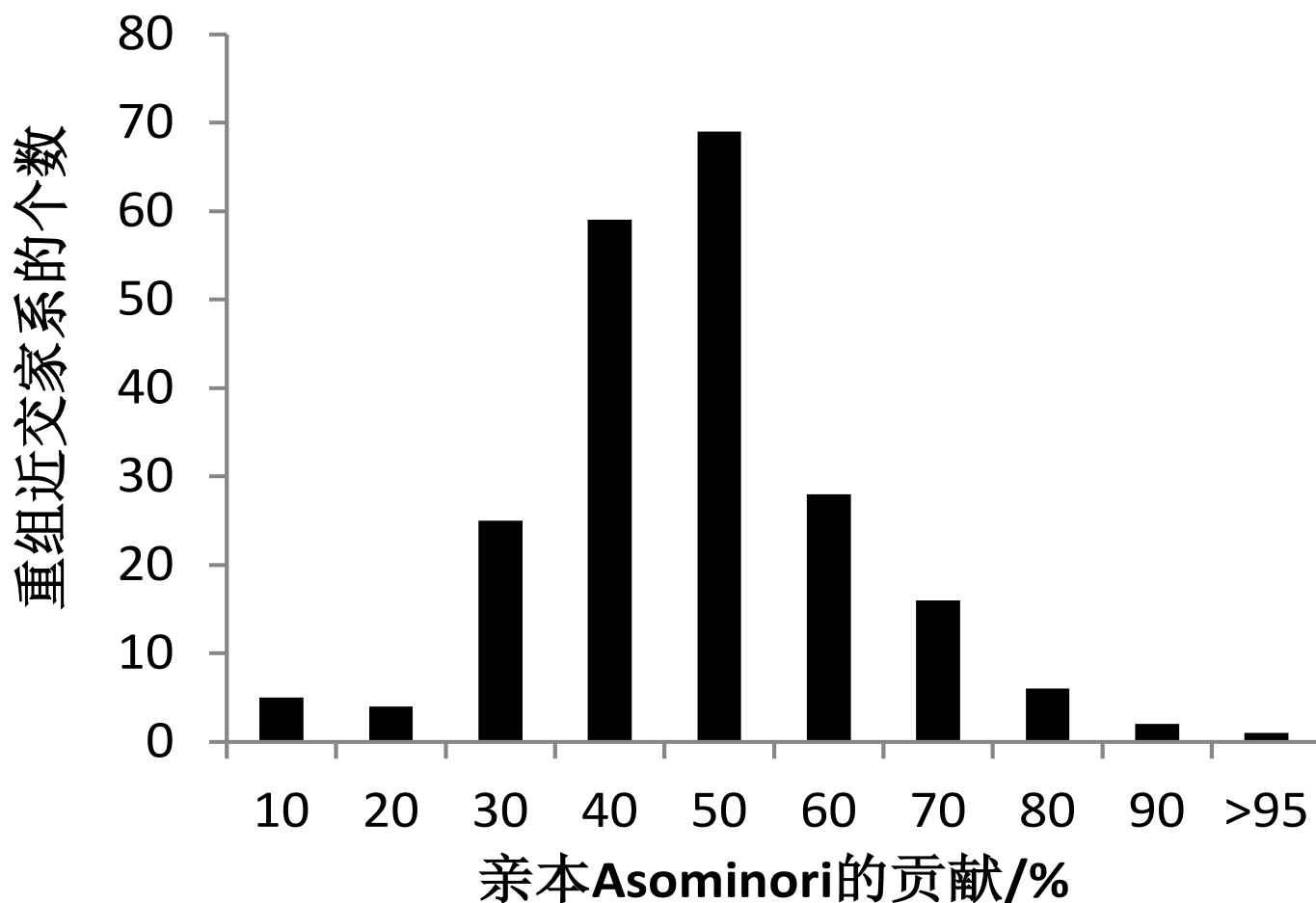
- 自交系或纯系是通过两个或多个亲本杂交，经过连续自交和选择得到的基因型纯合一致的家系。不论是亲本自交系还是后代重组自交系，一个座位上的两个等位基因都是后裔同样的，因此近交系数均为1。



# 重组自交系的亲本贡献

- 在单交组合产生的重组自交系中，由于自交过程中的随机漂变和选择等因素，每个亲本对后代自交系的贡献可能偏离50%。在回交一代群体产生的重组自交系中，轮回亲本对自交系的贡献可能偏离75%。
- 两个亲本对自交系的具体贡献，可以根据分子标记的基因型数据进行估计。如无基因型数据，可以用无漂变、无选择的理论贡献来代替。

# 粳稻品种‘Asominori’和籼稻品种‘IR24’为亲本、通过单粒传衍生的215个RIL家系中，亲本‘Asominori’遗传贡献(%)的次数分布



# 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 $h$ 。

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



# § 4.4 群体遗传学在植物遗传资源保护中的应用

- § 4.4.1 植物遗传资源的搜集
- § 4.4.2 植物遗传资源的再生繁殖

# 植物遗传资源保护

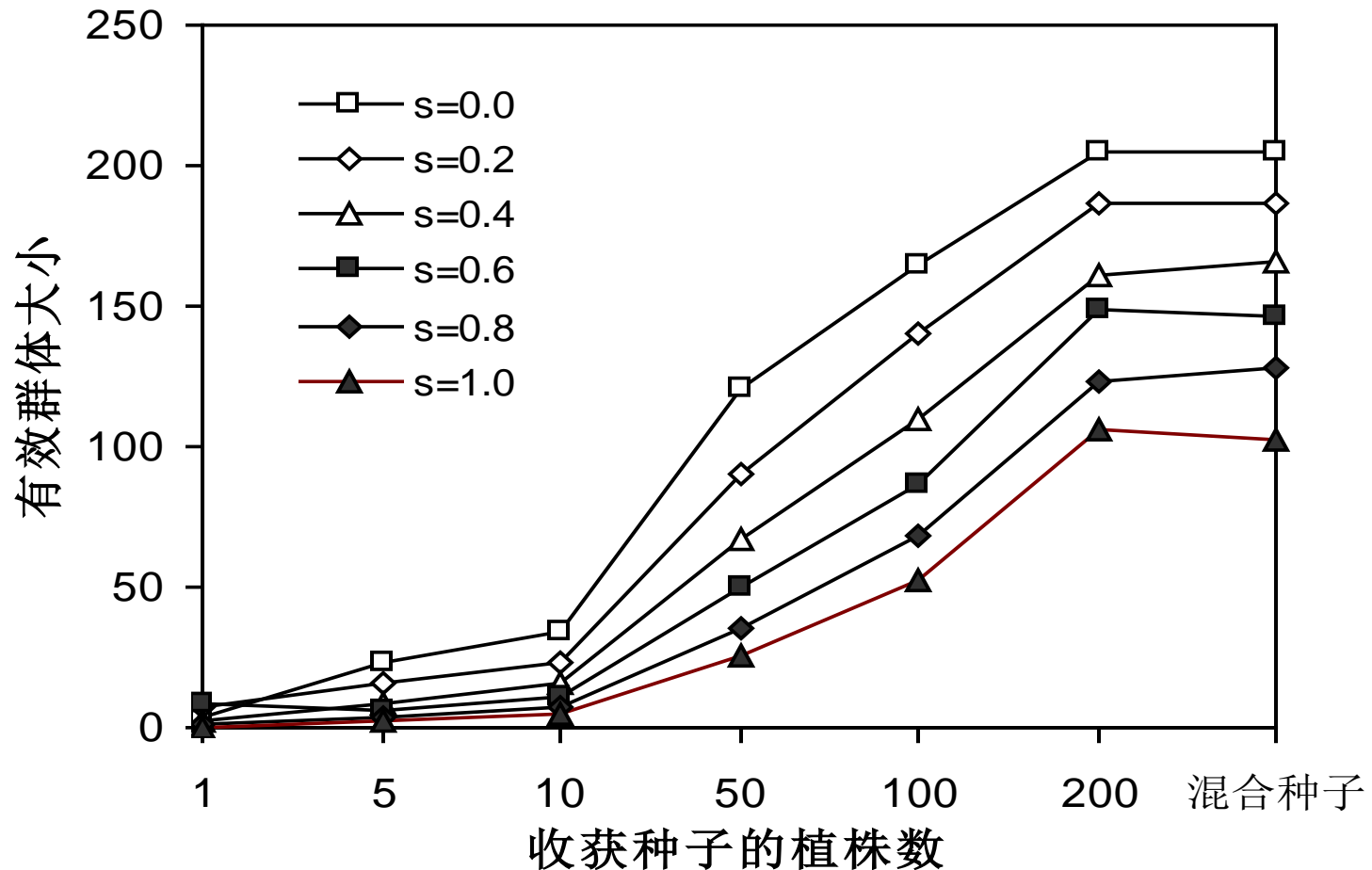
- 遗传资源保护主要涉及遗传资源搜集和再生繁殖两大过程。当然，遗传资源在种质库中的保存也很重要，但只要措施得当，对群体的遗传结构不会造成很大的影响。
- 这里，我们仅对搜集和再生过程中如何提高群体的有效大小、避免丢失基因、尽可能地保持原始群体的遗传完整性作一些探讨。

# 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

