

第2章

群体遗传结构的定向改变

王建康

中国农业科学院作物科学研究所

wangjiankang@caas.cn

<http://www.isbreeding.net>

影响群体结构的因素

- 对于随机交配大群体来说，如果没有外来因素干扰，基因频率和基因型频率将代代相同。
- 但是，一种物种在繁衍过程中，总是要受到各种外界或内部因素的影响，致使群体的遗传组成在世代传递过程中发生变化，有时导致基因频率的变化，有时导致基因型频率的变化，有时还可能同时影响基因频率和基因型频率。
- 这些影响因素可以分成两大类型，一类称为系统过程（systematic process），如突变、迁移和选择。系统过程对遗传组成的影响，在数量和方向上是可以预期的。

影响群体结构的因素

- 另一类称为分散过程（dispersive process），分散过程与群体大小（容量）有关。例如，群体再分为亚群体（subpopulation）和亚群体隔离、有限群体中的随机交配、抽样效应和随机飘移等。对这类过程，群体结构变化的方向是不可预测的，但其变化幅度是可以预测的。
- 本章主要介绍系统过程对群体结构的影响，仍然假定大群体，即不考虑随机抽样的影响。随机交配大群体中，根据Hardy-Weinberg定律，可以从基因频率得到基因型频率。因此，这一章主要考察系统过程对基因频率的影响。

本章的主要内容

§ 2.1 突变和迁移对基因频率的影响

§ 2.2 选择对基因频率的影响

§ 2.3 突变和选择的联合效应

§ 2.4 基因的多态性和选择的连锁效应

§ 2.1 突变和迁移对基因频率的影响

- § 2.1.1 突变
- § 2.1.2 非逆突变
- § 2.1.3 可逆突变
- § 2.1.4 迁移

突变及其类型

- 突变（mutation）是新基因和遗传多样性的最终来源，也是进化和育种的最原始材料。
- 突变一般是指基因核苷酸序列的变化。例如，一个基因A在某碱基位置上发生了A到C的变化，DNA序列的改变有可能导致最终生化合成产物蛋白质结构的改变。因此，突变后的基因效应也会发生变化，从而产生出新的表型，结果就产生了一个新的等位基因*a*。
- 颠换（inversion）和移位（translocation）等染色体重组也能引起遗传变异。为了与染色体畸变相区别，我们将这种基因内部结构的变化称为基因突变。

突变的作用和发生频率

- 动植物育种中，突变也可能给人工选择带来更多机会。例如，绿色革命中发挥重要作用的小麦矮秆基因是一个自然突变，目前的矮秆和半矮秆小麦品种就是利用了这一突变。
- 自然条件下基因的突变频率总是比较低，单个基因在一个世代中的突变频率大约在 $10^{-5} \sim 10^{-6}$ 之间。尽管如此，由于一个群体往往包含很多个体，每个个体都携带大量有可能发生突变的基因，低突变频率也能产生很多新的突变等位基因。
- 另外，高温、低温、射线或化学诱变剂处理等外界因素，也会大大提高基因的突变频率。突变育种就是利用物理和化学的方法对育种材料（如种子）进行处理，以产生较多的突变体，从而为选择提供更多的遗传变异。

频发突变和非频发突变

- 根据发生频率的高低，可把突变分为频发突变（recurrent mutation）和非频发突变（non-recurrent mutation）两种类型。
- 例如，在某一座位只发生一次A到a的突变，即基因型为AA的群体中，有一个个体的基因型突变为Aa。此个体以后会遇到能否有机会成活，能否有机会产生配子等问题。即使有机会，突变基因在突变个体中也只有一半的机会进入下一代；在其后的世代中它又遇到同样的命运。这样，经过若干世代之后，突变基因一般就会不复存在。
- 因此，单个突变基因的丢失是永远的、经常性的。所以，遗传上一般考虑的是频发突变，即突变频率足以引起群体中基因频率的改变，在一个大群体中，突变基因不会因为随机抽样而丢失。

非逆突变和可逆突变

- 常态情况下的频发突变又可分为非逆突变（irreversible mutation）和可逆突变（reversible mutation）。
- 以一个座位上的两个等位基因 A 和 a 为例，在每个世代中，等位基因 A 以一定的频率突变为等位基因 a ，但等位基因 a 不会突变为等位基因 A ，称这种突变是不可逆的。
- 如果在等位基因 A 突变为等位基因 a 的同时，等位基因 a 也以一定的频率突变为等位基因 A ，称这种突变是可逆的。

非逆突变对基因频率的改变量

- 若 $A \rightarrow a$ 的突变频率为 u ，起始群体中等位基因 A 的频率为 p_0 。突变和随机交配发生 t 个世代后，等位基因 A 的频率为 p_t 。
- 相邻世代间基因 A 的频率与突变频率 u 的关系

$$p_t = p_{t-1}(1 - u)$$

- 经过 t 个世代后的频率与起始频率的关系

$$p_t = p_0(1 - u)^t$$

基因频率改变的世代数

$$p_t = p_0(1 - u)^t$$

- 利用上面的公式，在已知突变率的情况下，可以计算基因频率从 p_0 到 p_t 的时间。
- 例如 $u=10^{-4}$ ，基因频率减半的时间为

$$t_{0.5} = \ln(0.5) / \ln(1 - u) = 6931$$

突变频率的估计 $p_t = p_0(1-u)^t$

- 另一方面，在已知基因频率从 p_0 到 p_t 的时间 t 后，又可以用来估计突变频率。
- 例如，等位基因 A 相对于 a 为显性，在一个随机交配群体中的频率为1，经过30个世代后发现隐性纯合基因型 aa 的频率为0.05。如不考虑其它改变群体结构的因素，等位基因 a 的频率可以用隐性纯合基因型的频率的平方根来估计。突变率的估计如下：

$$p_0 = 1.0 \quad p_{30} = 1 - \sqrt{0.05} = 0.7764$$

$$u = 1 - \left(\frac{p_{30}}{p_0}\right)^{\frac{1}{30}} = 0.0084$$

可逆突变

- 非逆突变 $A \rightarrow a$ 会缓慢降低群体中等位基因 A 的频率 p ，提高等位基因 a 的频率。很多时候，两个等位基因 A 和 a 可以相互突变。
- 如果把 $A \rightarrow a$ 称为正向突变（forward mutation），则把 $a \rightarrow A$ 称为反向突变（reverse mutation）。
- 正向突变倾向于降低等位基因 A 的频率 p ，反向突变又倾向于提高等位基因 A 的频率 p ，最终的基因频率 p 会达到一个平衡点。
- 在这个平衡点上，反向突变产生的基因数正好弥补了正向突变损失的基因数。

可逆突变过程中基因频率的变化



起始基因频率 p_0 q_0

世代 $t-1$ 基因频率 p_{t-1} q_{t-1}

世代 t 基因频率 p_t q_t

$$p_t = p_{t-1}(1-u) + (1-p_{t-1})v$$

$$\Delta p_t = p_t - p_{t-1} = vq_{t-1} - up_{t-1}$$

$$\Delta q_t = q_t - q_{t-1} = up_{t-1} - vq_t = -\Delta p_t$$

可逆突变的平衡频率

$$\Delta p_t = p_t - p_{t-1} = vq_{t-1} - up_{t-1}$$

- 从上式可知，当某一等位基因获得大于遗失时，则其频率增加；然而当频率增加后，其遗失也随着增加。所以，某一等位基因频率增加到一定程度后又会减少，直至最后达到平衡为止。等位基因的平衡频率分别为：

$$\tilde{p} = \frac{v}{u+v}, \tilde{q} = \frac{u}{u+v}$$

世代间基因频率的递推关系

$$p_t - \frac{v}{u+v} = \left(p_{t-1} - \frac{v}{u+v}\right)(1-u-v)$$

$$p_t - \frac{v}{u+v} = \left(p_0 - \frac{v}{u+v}\right)(1-u-v)^t$$

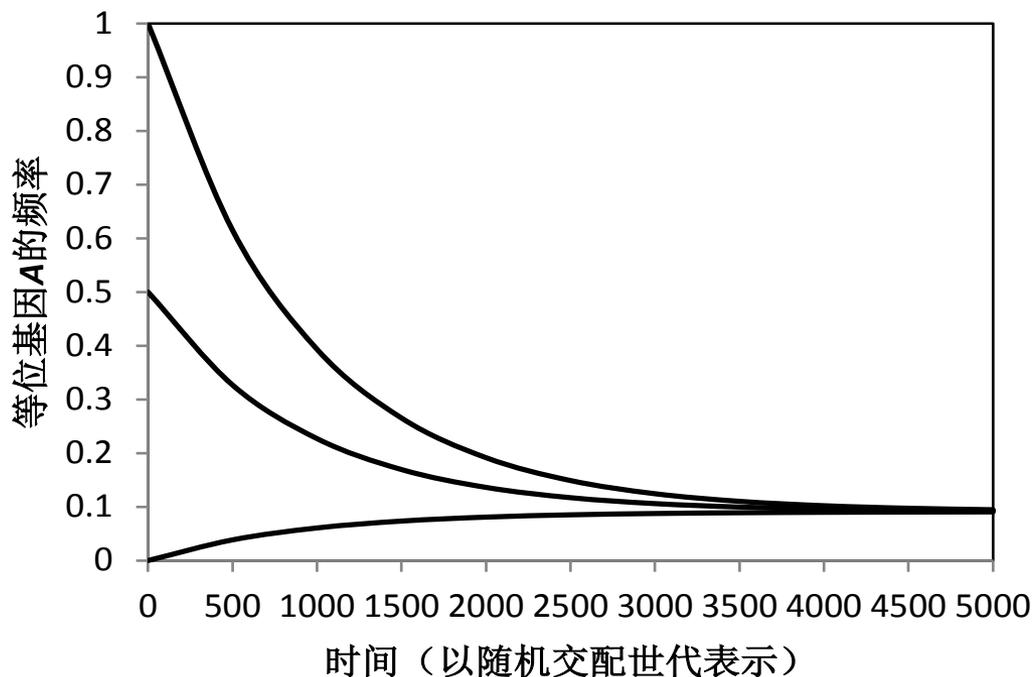
- 从中也可以得到前面的平衡频率。

基因频率到达给定值的时间

$$t = \frac{\ln(p_t - \tilde{p}) - \ln(p_0 - \tilde{p})}{\ln(1 - u - v)} \approx \frac{\ln(p_0 - \tilde{p}) - \ln(p_t - \tilde{p})}{u + v}$$

- 例如， $u=0.00003$ ， $v=0.00002$ ，平衡频率为0.40。根据上述公式2可以得到等位基因A的频率从0.90到0.80需要大约4463个世代。

随机交配群体中可逆突变座位上等位基因频率随时间变化曲线



- 三条曲线分别代表基因A频率为1.0、0.5和0的三个起始群体，基因A到a的突变频率为 10^{-3} ，基因a到A的突变频率为 10^{-4} ，平衡频率为0.0909。

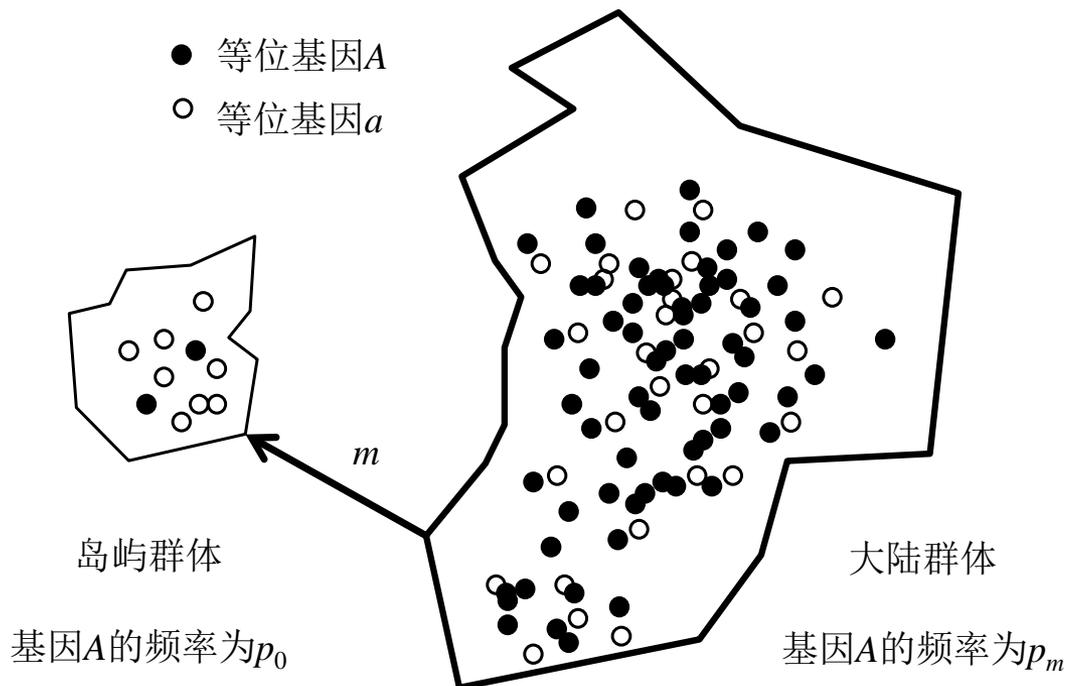
突变对群体结构的影响

- 短期内，突变对基因频率的影响不会很大（一些微生物除外）。但是，如果从进化的角度或从时间的角度来看，它又是非常重要的。
- 可逆突变往往是不对称的，即基因从野生型A到突变型a的突变频率一般要比从突变型a回复到野生型A的频率高得多，前者往往为后者的10倍左右，即 $u=10v$ 。
- 如果只有突变而没有其他改变基因频率的因素，最终的群体中突变型a的频率要远高于野生型A。果真如此，突变型会成为群体中的常见等位基因（common allele），而野生型则称为稀有等位基因（rare allele）。
- 但是，这种现象在很多自然群体中并没有出现，也就意味着可逆突变并不是改变基因频率的唯一因素，可能还有选择的影响。

迁移

- 迁移（migration）指的是个体在不同群体（也称为亚群体）间的移动。
- 迁移的最终结果是基因在亚群体间的流动，它可以是单向的，也可以是双向的，也可以是两个亚群体间的迁移，也可以是多个亚群体间的迁移。

从大陆向岛屿的单向迁移模型图



- 实心圆代表等位基因A，空心圆等位基因 a 。迁移开始前，等位基因A在岛屿和大陆上的频率分别为 p_0 和 p_m 。每个世代从大陆迁移的个体占岛屿群体的比例为 m 。

迁移对基因频率的影响

- 原来岛屿群体中基因A的频率为 p_0 。每个世代由大陆群体迁入比率为 m 的个体，大陆群体中基因A的频率为 p_m 。
- 迁移发生后，大陆群体对基因A频率的贡献为 $p_m \times m$ ；迁移前岛屿群体对基因A频率的贡献为 $p_0(1-m)$ 。
- 岛屿群体的基因频率及其改变量分别为：

$$p_1 = p_m \times m + p_0(1-m) = m(p_m - p_0) + p_0$$

$$\Delta p = p_1 - p_0 = m(p_m - p_0)$$

相邻世代间基因频率的关系

$$p_1 - p_m = (1 - m)(p_0 - p_m)$$

$$p_t = p_m + (1 - m)^t (p_0 - p_m)$$

- 从中可以明显看到单向迁移的长期结果，即岛屿群体中等位基因A的频率最终等于大陆群体中等位基因A的频率。

多个大陆群体向岛屿群体的迁移

$$p_1 = \sum_{i=1}^k m_i p_{m_i} + p_0 \left(1 - \sum_{i=1}^k m_i\right) = \sum_{i=1}^k m_i (p_{m_i} - p_0) + p_0$$

$$\Delta p = \sum_{i=1}^k m_i (p_{m_i} - p_0)$$

多个大陆群体向岛屿群体的迁移

$$p_1 = m\bar{p} + (1 - m)p_0$$

$$p_1 - \bar{p} = (1 - m)(p_0 - \bar{p})$$

- 其中， $m = \sum_{i=1}^k m_i$ 为总迁移率； $\bar{p} = \sum_{i=1}^k \frac{m_i}{m} p_{m_i}$ 为大陆群体频率的加权平均。
- 从递推公式可以看到，长期迁移使得岛屿群体的基因频率最终等于大陆群体基因频率按迁移比例的加权平均。

相互迁移及其结果

- 如果迁移是相互的，仍可以沿用前面的推导方法，只是分析起来要更复杂些。
- 但有一点是非常清楚的，随着相互迁移的发生，群体间基因频率的差异会变得越来越小。
- 经过若干代相互迁移之后，所有参与迁入和迁出的亚群体将具有相同的基因频率，即各群体基因频率按迁移比例得到的加权平均数。

突变和迁移之间的相似性

突变：
$$p_t - \frac{v}{u+v} = \left(p_0 - \frac{v}{u+v}\right)(1-u-v)^t$$

迁移：
$$p_t = p_m + (1-m)^t (p_0 - p_m)$$

- 对照突变和迁移的基因频率公式可以发现，如果把迁移方程中外来群体频率 p_m 视为突变方程中的平衡频率 $v/(u+v)$ ，迁移比例 m 视为突变频率之和 $u+v$ ，则两个公式是等价的。

突变和迁移的异同

- 因此，迁移和突变这两种看似截然不同的因素，它们对遗传结构影响的数学表达式是完全相同的，群体遗传研究中有时不区分二者的差异。
- 这一点直观上也是可以理解的，即群体中新增加的基因A是迁移而来，还是由基因a突变而成，它们对基因频率的影响没有本质区别。
- 但是，由于迁移的频率往往远高于突变的频率，二者对基因频率影响的程度存在明显的差异。迁移过程中，基因频率趋于平衡频率的速度要比突变快得多。

§ 2.2 选择对基因频率的影响

- § 2.2.1 适合度和选择系数
- § 2.2.2 不利隐性基因的部分选择
- § 2.2.3 不利隐性基因的完全选择
- § 2.2.4 有利于杂合子的选择
- § 2.2.5 选择的有效性

选择

- 选择是改变群体基因频率的一个最重要、也是最有效的手段
- 选择有自然选择和人工选择之分
- 选择的主要后果是群体内基因频率的改变
- 连续世代的基因频率间的关系
- 基因频率的变化
- 基因频率达到某一特定值所需的代数

自然选择

- 在一定自然条件下，群体内某些基因型个体比起另一些基因型个体具有更高的成活率和生殖率，因此这些基因型个体在群体中就会占居优势并逐渐代替其他基因型而形成新的群体或新的种；如果优势基因型和其他基因型都能生存下来，不同基因型就分布在它们最适宜的地域，成为亚群体或地理亚种。
- 这种在自然界的条件下，由于群体内不同基因型个体成活率和生殖率间的差异而造成的选择称为自然选择。

人工选择

- 在人工条件下，例如在作物杂交育种过程中，从杂种群体中选去具有优良性状的单株作育种材料，淘汰那些不符合育种目标的个体，这实质上是选择对人有利的基因型，淘汰对人不利基因型。所以，选择实际上是决定群体内不同基因型个体的繁殖比例。

适合度与选择系数

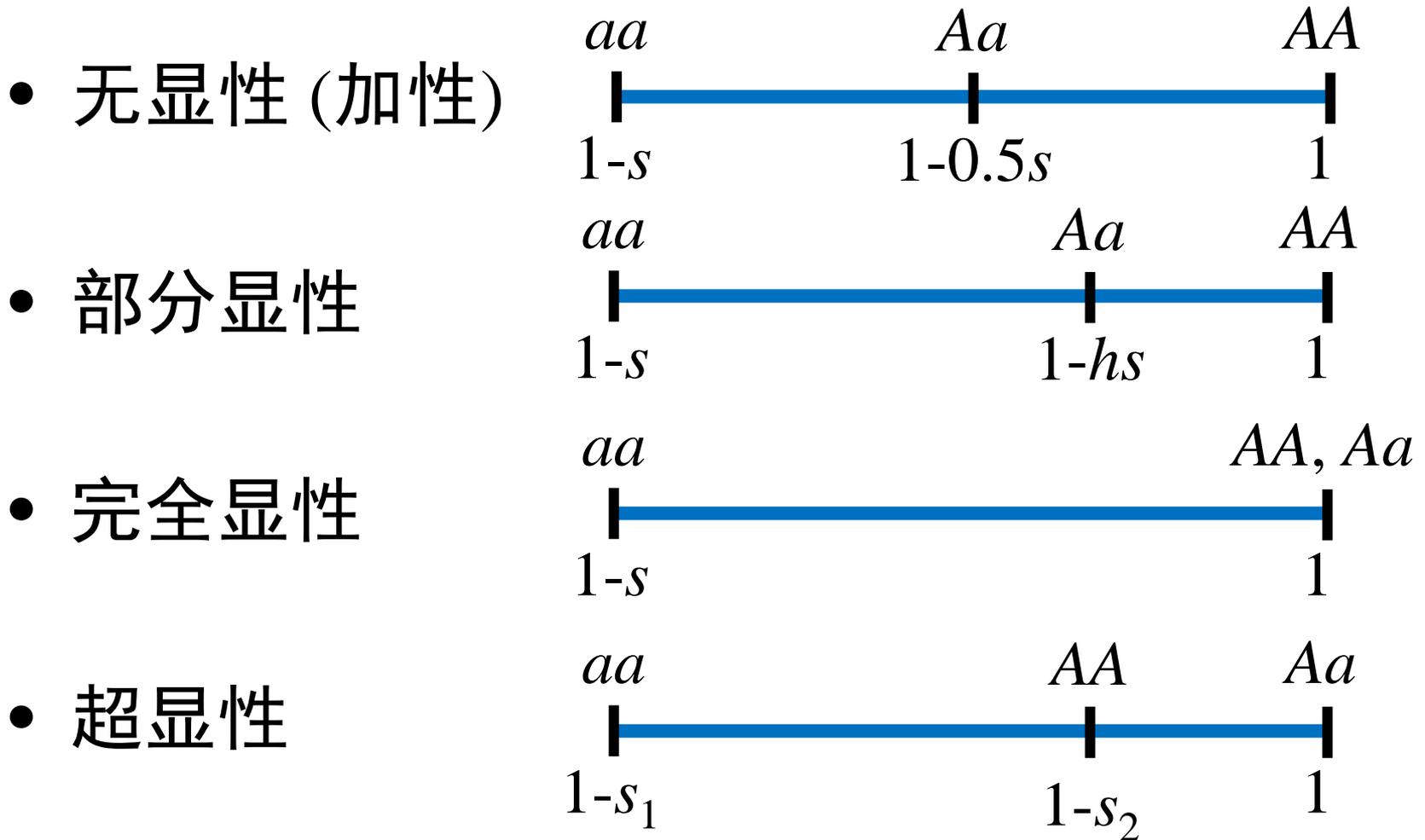
- 假定有两种基因型个体 AA 和 aa 各100个， AA 个体的繁殖成活率为1， aa 个体为0.9，那么我们可以说， aa 的适合度为 AA 的0.9。所以适合度（ W ）是指某基因型能繁殖成活后代的相对能力，其值在0和1之间。
- 在这里， AA 和 aa 的适合度分别为 $W_{AA}=1$ 和 $W_{aa}=0.9$ 。令 $s=W_{AA}-W_{aa}=1-0.9=0.1$ ，则一般称 s 为对 aa 基因型被淘汰而不能繁殖后代的个体在群体中所占的百分率，是选择强度的一个指标即选择系数。该例 $s=0.1$ 说明，基因型 aa 有的个体不能繁殖后代。

利用婴儿群体和成人群体的基因型 调查数据估计适合度和选择系数

基因型	AA	AS	SS
婴儿群体	189	89	9
成人群体	400	249	5
婴儿群体基因型频率	0.6586	0.3101	0.0314
成人群体基因型频率	0.6116	0.3807	0.0076
适合度 (成人和婴儿 基因型频率的比值)	0.9288	1.2278	0.2438
相对适合度	0.7565	1	0.1986
选择系数	0.2435	0	0.8014

适合度的显隐性

参数 s 、 h 、 s_1 、 s_2 在 $(0, 1)$ 之间取值，但不能等于0或1



适合度的类型与选择的结果

参数 s 、 h 、 s_1 、 s_2 在 $(0, 1)$ 之间取值，但不能等于0或1； p 和 q 分别为等位基因A和a的频率

适合度类型	AA	Aa	aa	选择一代后等位基因a频率	长期选择的最终结果
加性（或无显性），选择对基因型Aa和aa不利	1	$1-0.5s$	$1-s$	$\frac{q - \frac{1}{2}sq - \frac{1}{2}sq^2}{1 - sq}$	$p=1, q=0$
部分显性，选择对基因型Aa和aa不利	1	$1-hs$	$1-s$	$\frac{q - hspq - sq^2}{1 - 2hspq - sq^2}$	$p=1, q=0$
完全显性，选择对基因aa不利	1	1	$1-s$	$\frac{q - sq^2}{1 - sq^2}$	$p=1, q=0$
完全显性，选择对基因型AA和Aa不利	$1-s$	$1-s$	1	$\frac{q - sq + sq^2}{1 - s(1 - q^2)}$	$p=0, q=1$
超显性，选择对纯合基因型AA和aa不利	$1-s_1$	1	$1-s_2$	$\frac{q - s_2q^2}{1 - s_1p^2 - s_2q^2}$	$p = \frac{s_2}{s_1 + s_2}, q = \frac{s_1}{s_1 + s_2}$

加性适合度的最终结果

- 前表第二行中， aa 的选择系数为 s ， Aa 的选择系数恰好是 aa 的一半。基因型 AA 具有最高适合度， aa 的适合度最低，杂合型 Aa 的适合度正好介于两个纯合基因型之间。这时，就说这两个等位基因之间不存在显性，或称这两个基因为加性基因。
- 在这种情况下，选择对基因型 Aa 和 aa 来说都是不利的，选择降低了这两种基因型的频率，进而降低了等位基因 a 的频率。经过多个世代的选择和随机交配后，群体中基因 A 的频率为1， a 的频率为0，等位基因 A 在群体中被固定下来，等位基因 a 被丢失掉。

显隐性适合度的最终结果

- 前表第四行中，基因型 AA 和 Aa 有相同的适合度，选择只对纯合型 aa 不利，这时就说等位基因 A 相对于 a 表现为显性，有时也称 A 是显性基因， a 是隐性基因。经过多个世代的选择和随机交配后，显性基因 A 在群体中被固定下来，隐性基因 a 被淘汰掉。
- 前表第五行表示的也是完全显性的情况，只不过选择对纯合型 AA 不利，经过多个世代的选择和随机交配后，显性基因 A 在群体中被淘汰掉。

共显性适合度的最终结果

- 前表最后一行中，杂合型 Aa 有最高的适合度，选择对纯合型 AA 和 aa 不利，这时就说两个等位基因 A 和 a 之间存在超显性。在这种情况下，基因频率最终会达到一种平衡状态。

$$\tilde{p} = \frac{s_2}{s_1 + s_2}, \tilde{q} = \frac{s_1}{s_1 + s_2}$$

不利于隐性基因的部分选择

基因型	AA	Aa	aa	总和
选择前的基因型频率	p_0^2	$2p_0q_0$	q_0^2	1
适合度	1	1	$1-s$	
选择后的基因型频率	p_0^2	$2p_0q_0$	$(1-s)q_0^2$	$1-sq_0^2$

连续世代的基因频率间的关系

- 选择一代后基因a的频率为

$$q_1 = \frac{(1-s)q_0^2 + p_0q_0}{1-sq_0^2} = \frac{q_0(1-sq_0)}{1-sq_0^2}$$

- 任意两个连续世代

$$q_{n+1} = \frac{(1-s)q_n^2 + p_nq_n}{1-sq_n^2} = \frac{q_n(1-sq_n)}{1-sq_n^2}$$

基因频率的变化量

- 选择一代后基因a频率的变化为

$$\Delta q = q_1 - q_0 = \frac{q_0(1 - sq_0)}{1 - sq_0^2} - q_0 = -\frac{sp_0q_0^2}{1 - sq_0^2}$$

- 当 q_0 较小时， Δq 与 q_0^2 成比例， Δq 会很小，所以企图通过选择淘汰稀少的隐性基因几乎是不可能的。
- 在纯系内选择是无效的，因为这时 $q=0$ 或 $q=1$ ，所以 $\Delta q=0$ 。

a 基因频率减少至某一特定值所需的世代数

$$sn = \left[\frac{1}{q} + \ln \frac{1-q}{q} \right] \Big|_{q_0}^{q_n} = \frac{q_0 - q_n}{q_0 q_n} + \ln \frac{q_0 (1 - q_n)}{q_n (1 - q_0)}$$

实例

- 例如， $s=0.001$ ，求群体中隐性个体的比率由50%减少到1%所需的世代数 n 。

$$q_0 = \sqrt{0.50} = 0.7071 \quad q_n = \sqrt{0.01} = 0.10$$

$$sn = \left[\frac{1}{q} + \ln \frac{1-q}{q} \right] \Big|_{q_0}^{q_n} = \frac{q_0 - q_n}{q_0 q_n} + \ln \frac{q_0 (1 - q_n)}{q_n (1 - q_0)}$$

$$0.001 \times n = \frac{0.7071 - 0.10}{0.7071 \times 0.10} + \ln \frac{0.7071(1 - 0.10)}{0.10(1 - 0.7071)} = 11.665$$

- 因此， $n=11665$

不利于隐性基因的完全选择，即 $s=1$

基因型	AA	Aa	aa	总和
选择前的基因型频率	p_0^2	$2p_0q_0$	q_0^2	1
适合度	1	1	0	
选择后的基因型频率	p_0^2	$2p_0q_0$	0	$1-q_0^2$

连续世代的基因频率间的关系

- 选择一代后基因a的频率为

$$q_1 = \frac{(1-s)q_0^2 + p_0q_0}{1-sq_0^2} = \frac{q_0(1-sq_0)}{1-sq_0^2} = \frac{q_0}{1+q_0}$$

- 连续 n 个世代后

$$q_n = \frac{q_0}{1+nq_0} \quad n = \frac{q_0 - q_n}{q_0q_n}$$

实例

- 例如，已知一个油菜随机交配群体中，感病植株（ rr ）占9%，在苗期死亡。抗病植株（ RR 和 Rr ）占91%，可正常开花授粉。试问随机交配三代后，感病幼苗在群体中还占多大比例。
- 设 p_0 和 q_0 分别为原来群体中基因 R 和 r 的频率，
$$q_0 = \sqrt{0.09} = 0.3 \quad p_0 = 1 - q_0 = 0.7$$
$$q_3 = \frac{q_0}{1 + 3q_0} = \frac{0.3}{1 + 3 \times 0.3} = 0.1579$$
- 因此，三代随机交配和选择后，基因型 rr 的频率为 $0.1579^2 = 0.0249$ ，即抗病植株约占群体的2.5%。

有利于杂型合子的选择

基因型	AA	Aa	aa	总和
选择前的基因型频率	p_0^2	$2p_0q_0$	q_0^2	1
适合度	$1-s_1$	1	$1-s_2$	
选择后的基因型频率	$(1-s_1)p_0^2$	$2p_0q_0$	$(1-s_2)q_0^2$	$1-s_1p_0^2-s_2q_0^2$

连续世代的基因频率间的关系

- 选择一代后基因a的频率为

$$q_1 = \frac{(1-s_2)q_0^2 + p_0q_0}{1-s_1p_0^2 - s_2q_0^2} = \frac{q_0(1-s_2q_0)}{1-s_1p_0^2 - s_2q_0^2}$$

- 基因频率a的变化量为

$$\Delta q = q_1 - q_0 = \frac{p_0q_0(s_1p_0 - s_2q_0)}{1-s_1p_0^2 - s_2q_0^2}$$

平衡时的基因频率

$$\tilde{p} = \frac{s_2}{s_1 + s_2} \quad \tilde{q} = \frac{s_1}{s_1 + s_2}$$

- 平衡时的基因频率完全由两种同型合子的选择系数决定，而与原始群体的基因频率无关。因为 s_1 和 s_2 是常数，所以这个平衡是稳定的。

人类遗传研究中选择有利于杂型合子的经典例子

- 在非洲有些地区的镰刀形细胞贫血基因Hb^S的频率很高。例如根据调查，基因型Hb^SHb^S、Hb^AHb^S和Hb^AHb^A的频率分别为0.012、0.388和0.600，所以基因Hb^S的频率高达0.206。
- 而纯合体Hb^SHb^S的生存繁殖率都很低，据估计其适合度为杂合体的0.25，即Hb^SHb^S的选择系数是 $s=1-W=0.75$ 。

纯合型 $Hb^A Hb^A$ 的选择系数

- 基因 Hb^S 的频率 $q=0.206$

$$s_1 = \frac{s_2 \hat{q}}{\hat{p}} = \frac{0.75 \times 0.206}{0.794} = 0.195$$

- 这就是说，杂型合子 $Hb^A Hb^S$ 的适合度为1时，同型合子 $Hb^A Hb^A$ 的适合度是 $1 - s_1 = 0.805$ 。
- 适合度表现出超显性。

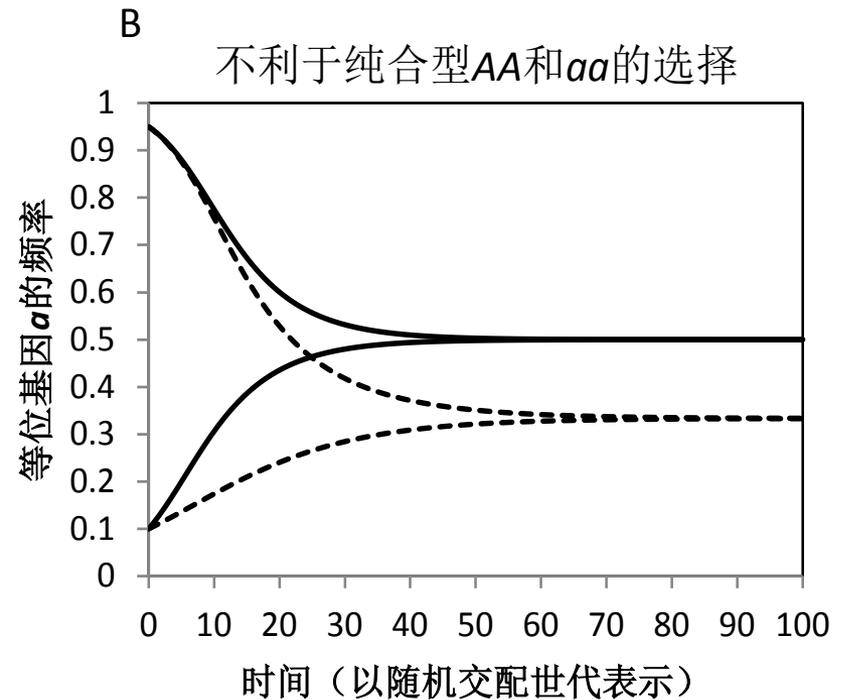
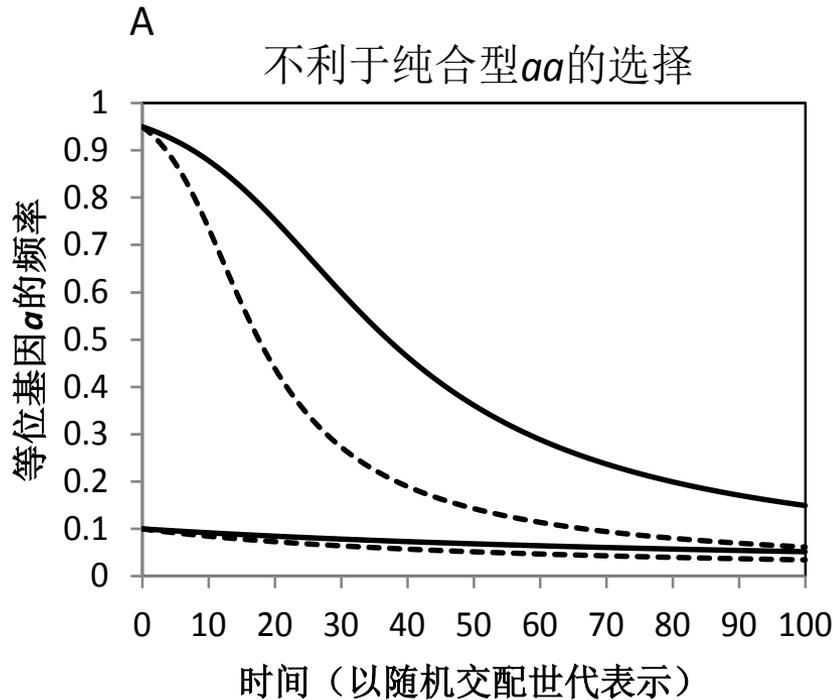
选择有利于杂型合子 $Hb^A Hb^S$ 的原因

- 为什么 $Hb^A Hb^S$ 的适合度高于 $Hb^A Hb^A$ 呢？据调查，前者的红血球内含有两种血红蛋白：一种是抵抗疟疾感染的血红蛋白，一种是阻止溶血的正常血红蛋白。在疟疾盛行的地区，自然 $Hb^A Hb^S$ 比 $Hb^A Hb^A$ 更适于生存。在疟疾被控制后， Hb^S 基因频率也就随着迅速下降。
- 这里的超显性，其实是由基因的多效性（pleiotropic effect或pleiotropy）引起的。纯合基因型 $Hb^S Hb^S$ 因贫血而减低了适合度，纯合基因型 $Hb^A Hb^A$ 因抵抗疟疾感染能力差减低了适合度，杂合基因型 $Hb^S Hb^S$ 不会患严重的贫血症且抵抗疟疾感染能力又强，因此具有最高的适合度。

选择的有效性

- 选择系数一般要远高于自然条件下基因的突变频率。因此，选择对基因频率的影响程度也远高于突变。下图给出隐性基因选择和超显性选择两种情况下，基因频率随时间的变化。与突变引起的基因频率变化相比较，选择的影响在少数几个世代后就能显现出来，而突变的影响要经过数百甚至数千个世代后才能显现。

选择对基因频率的影响



两种起始频率0.95和0.1下不利于基因型 aa 的选择，即等位基因 a 表现为隐性，实线和虚线分别代表选择系数0.2和0.1。

两种起始频率0.95和0.1下不利于纯合基因型 AA 和 aa 的选择，即超显性，实线表示 AA 和 aa 选择系数均为0.2，虚线表示 AA 和 aa 选择系数分别为0.2和0.1

§ 2.3 突变和选择的联合效应

- § 2.3.1 突变和选择的平衡基因频率
- § 2.3.2 基因突变频率的估计

突变和选择的联合效应

- 在之前考虑选择对基因频率的影响时，实际上假定了突变基因是中性的这一前提条件，即突变基因对携带它们的个体的生存和繁殖既无利又无害。但是，在自然界中，这两种影响有时会同时存在。同时考虑选择和突变对群体基因频率的影响，则更接近实际情况。
- 显然，如果突变和选择对基因频率影响的方向相同，那么基因频率的变化相比突变或选择单独存在时的变化要更快些。但是，如果方向相反，那么它们的效应就会相互抵消，最后形成一个稳定的平衡状态。

突变和隐性选择的联合效应

- 假定选择对隐性个体不利，选择系数为 s ；突变发生的方向是由A至a，每一世代的突变率为 u 。选择的作用是减少a基因频率，即减少 q 值；突变的作用是增加 q 值。在随即交配群体中每一世代 q 值的净变化是：

$$\Delta q = u(1 - q) - sq^2(1 - q)$$

突变和隐性选择的平衡

$$\Delta q = u(1 - q) - sq^2(1 - q) = 0 \quad \hat{q} = \sqrt{\frac{u}{s}}$$

- 这是一个稳定的平衡，能很好解释如下事实：对隐性有害个体通常总有一定的比例存留在自然群体中而不能全部消失。因为这样的群体中，是通过突变才把隐性基因保存下来， u 值通常是很小的。

隐性致死基因的突变率

- 当隐性个体aa为致死时，即s=1，则aa基因型的频率等于突变率：

$$\hat{q} = \sqrt{u}$$

突变和显性选择的联合效应

- 如存在部分显性，即 AA 、 Aa 、 aa 的选择系数分别为 1 、 $1-hs$ 、 $1-s$ ， h 在 0 和 1 之间但不等于 0 或 1 ，这里不加证明地给出平衡频率的近似表达式。

$$\tilde{q} \approx \frac{u}{hs}$$

- 可以看出，对杂合基因型有选择和无选择的巨大差异。仅对 aa 的完全选择 $s=1$ 时，平衡频率等于突变频率的平方根。如对 Aa 和 aa 同时选择，对于任何接近但不等于 1 的 h ，平衡频率近似等于突变频率，远低于突变与隐性选择的平衡频率。

突变和选择的平衡与基因多态性

- 对隐性有害基因来说，通常总有一定的比例存留在自然群体中而不能全部消失。这样的群体中，是通过频发突变才把隐性基因保存下来，突变频率 u 通常都很小。例如，一个随机交配群体中由 A 到 a 的突变率 $u=1.8 \times 10^{-7}$ ，对隐性基因的纯合个体 aa 的选择系数 $s=0.02$ ，平衡频率为0.03。因此，该群体中纯合子 aa 的频率是0.0009，杂合子 Aa 的频率是0.0582，约为隐性纯合子的65倍。
- 所以，大部分隐性有害基因存在于杂型合子中，选择没有什么效果。甚至当平衡频率小到隐性个体的频率可以忽略不计时，隐性基因仍可能存留在群体中（即隐藏在杂型合子内）。所以，在发生隐性突变的条件下，自然群体中实际上不可能把隐性基因完全选择掉。

突变率的估计 $\hat{q} = \sqrt{\frac{u}{s}}$

- 只要测定出 s ，就根据基因频率可估算出 u 。人类各种基因的自发突变率大多是根据这个公式估得的。
- 例如，人类中全色盲是属常染色体隐性遗传。据调查，约8万人中有一个全色盲，全色盲的平均子女数约为正常人的一半（ $s=0.5$ ）。

$$q^2 = \frac{1}{80000} \quad u = sq^2 = 6.25 \times 10^{-5}$$

§ 2.4 基因的多态性和选择的连锁效应

- § 2.4.1 自然群体中基因的多态性
- § 2.4.2 选择的连锁效应
- § 2.4.3 遗传平衡及其分类

由于选择的作用，有害突变基因在自然群体中的频率一般都很低

- 野生型到有害突变一般属于正向突变，正向突变的频率远高于反向突变。如果没有选择的存在，群体中的大部分基因最终都是有害的突变基因。
- 在自然群体中，有害突变基因相对于野生型有较低的适合度。在突变和选择平衡公式中，有害基因的选择系数一般远远高于突变频率。在突变和选择的共同作用下，平衡群体中有害突变的基因频率并不会很高。如 $u=10^{-5}$ 、 $s=0.5$ 时，突变基因的平衡频率仅为0.0014。考虑显性效应时，平衡频率将更低。突变和选择的共同作用，很好地解释了自然群体中有害突变比较稀有这一常见现象。

多态性座位

- 当一个座位上，多于一个的等位基因频率高到突变和选择平衡理论不能解释的时候，如突变基因的频率超过0.01，称这个座位具有多态性（polymorphism）。
- 几乎所有的遗传群体，都能在DNA序列到表型性状的不同水平上观测到多态性。有时，基因引起的可观测形态变异比较少。但是，利用分子标记技术仍可以检测到DNA水平上的广泛变异。例如，动植物的指纹图谱鉴定，就是利用DNA水平的多态性来区分不同基因型的个体。

稀有基因和常见基因

- 对于大多数自然群体，只要样本足够大，很多座位上都会发现多于一个的等位性变异。样本越大，发现的等位基因也就越多。但是，有的频率会比较高，有的频率可能非常低。
- 为研究方便，人们常把这些等位基因按照频率的高低分为稀有和常见两类。稀有基因（rare allele）和常见基因（common allele）的划分带有主观性。
- 人们一般倾向于把频率低于0.5%的基因称为稀有基因，高于这个频率时则称为常见基因。根据Hardy-Weinberg定律，稀有基因主要存在于杂合子中。例如，一个基因的频率为0.5%，该基因的杂合子在群体中的存在比例近似等于1%。因此，可以认为稀有基因在100个个体中至多有一个携带者。

多态性座位和单态性座位

- 人们常用最高的等位基因频率来判断一个座位是否多态，常用的划分标准是0.95。也就是说，等位基因的最高频率不超过0.95的座位，则认为多态性座位。
- 当两个等位基因频率相等时，杂合子的频率最高；基因频率差异越大，杂合子的频率越低。这个结果对于复等位基因也是成立的。对两个等位基因的座位来说，0.95的划分标准相当于杂合子频率不低于10%，即 $2 \times 0.95 \times 0.05$ 约等于10%。例如，最高频率超过0.95，说明群体中只有一个等位基因的频率很高，大部分个体都是这个等位基因的纯合体，称这样的座位是单态的（monomorphism）。

产生多态性的原因

- 自然群体中，产生多态性的原因有很多，除了前面介绍的有利于杂合子的选择外，还有依赖于基因型频率的选择、一因多效性、上位性互作、环境效应、中性突变、奇异分离、配子选择、性选择等等。

选择的连锁效应

- 对于遗传上存在连锁的座位来说，选择在改变一个座位上基因频率的同时，也可能改变连锁座位上的基因频率。

$D=0$ 时，选择不产生连锁效应

- 假定一个群体中四种配子 AB 、 Ab 、 aB 、 ab 的频率分别为0.25、0.25、0.25、0.25，四种等位基因的频率均为0.5，两个座位处于平衡状态，即连锁不平衡度 $D=0$ 。
- 如果选择使得所有的 A 都保留下来，而 a 只有60%保留下来，那么选择后四种配子还有0.25、0.25、0.15、0.15，标准化后的频率为0.3125、0.3125、0.1875、0.1875。
- 重新计算基因频率发现， A 和 a 的频率分别变为0.625和0.375。但是， B 和 b 的频率并未发生任何改变，仍然为0.5。因此，对一个座位的选择，不会影响另一个平衡座位上的基因频率。

$D > 0$ 时，选择将产生连锁效应

- 如对于另外两个座位来说，配子 AB 、 Ab 、 aB 、 ab 的频率分别为0.4、0.1、0.1、0.4，四种等位基因的频率均为0.5，但两个座位处于不平衡状态（这里的 $D=0.15$ ）。
- 与前面类似，选择使得所有的 A 都保留下来，而 a 只有60%保留下来。选择后四种配子还有0.4、0.1、0.06、0.24，标准化后的频率为0.5、0.125、0.075、0.3。
- 与前面一样， A 和 a 的频率分别变为0.625和0.375。不一样的是， B 和 b 的频率也发生了改变，分别变为0.575和0.425。

选择中的搭车效应

- 因此，尽管基因 B 和 b 是中性的，选择在提高等位基因 A 频率的同时，也提高了连锁座位上等位基因 B 的频率。这种现象称为搭车效应（Hitch-hiking effect）。
- 在随机交配群体中，座位间的不平衡程度随着随机交配世代的增加而逐渐消失，或者说如果有足够的时间，遗传上连锁座位间的不平衡将消失，搭车效应也随之停止。
- 在小群体中，随机漂移也将影响座位间的不平衡。在连锁非常紧密的情况下，可能发生搭车效应，此时只要被选择的基因固定下来，搭车效应随即停止。一般来说，连锁越松散，搭车效应停止越早。反之，连锁越紧密，搭车效应停止越迟。选择强度越大，对搭车基因的影响也越大。

选择清除和连锁累赘现象

- 当一个有利突变发生后，这个突变基因的适合度越高，选择将其固定下来的速度就越快。当这个突变基因被快速固定之后，与此座位紧密连锁的染色体区域由于搭车效应也被固定下来，大片紧密连锁的染色体区域也因此失去多态性。由于搭车效应而引起多态性下降的现象，遗传上称为选择清除（selective sweep）。被清除区域的大小依赖于重组率，低重组率的染色体区域、失去多态性的区域就会很长。
- 此外，如果座位A的有利基因和另一个座位B的不利基因紧密连锁，在选择座位A有利基因的同时，也选择了座位B的不利基因，这种现象称为连锁累赘（linkage drag）。

矮秆基因利用与“绿色革命”

- 例如，在绿色革命中发挥重要作用的小麦矮秆基因，一开始被发现于农艺性状表现很差的遗传材料中。最初的几轮回交转育，在导入矮秆基因的同时，也导入了矮秆基因附近的大片染色体片段。因此，最初获得的矮秆育种材料农艺性状也很差。这就是连锁累赘现象。
- 但是，随着回交、自交和选择的继续，矮秆基因与紧密连锁的不利基因间有更多重组的机会。在不断重组的过程中，矮秆基因所在的原始染色体片段不断变小，最终打破了矮秆基因与不利农艺性状基因间的连锁，选育出既矮秆又农艺性状优良的小麦新品种。

遗传平衡及其分类

- 遗传平衡主要指的是群体结构的平衡，也就是说群体基因频率和基因型频率保持不变的状态。平衡一般分为稳定和不稳定两种类型。
- 稳定平衡可以是全局稳定（globally stable），也可能是局部稳定（locally stable）。全局稳定的平衡与基因的起始频率无关，不管基因的起始频率如何变化，平衡群体的基因频率都保持不变。局部稳定的平衡与基因的起始频率有关，不同的起始频率对应于不同的平衡频率。

稳定平衡 $p_t = p_0(1 - u)^t$

- 对于稳定平衡（stable equilibrium）来说，当基因频率接近于平衡频率时，群体随世代逐渐趋近于平衡状态。在平衡群体中，当基因频率发生微小波动后，仍然能够回到平衡状态。
 - 以非逆突变为例，当群体的基因频率 $p > 0$ 时，后代群体的基因频率将逐渐下降，最终趋近于0。当群体的基因频率 $p = 0$ 时，所有后代群体的基因频率也都是0。
 - 因此，频率 $p = 0$ 是一个平衡状态。同时，平衡频率与起始频率没有任何关系，因此 $p = 0$ 还是一个全局平衡。

不稳定平衡 $\Delta q = -\frac{sp_0q_0^2}{1-sq_0^2}$

- 对于不稳定平衡（unstable equilibrium）来说，平衡群体的基因频率发生微小波动时，后代群体就不能回到平衡状态。
 - 对于不利隐性基因的选择来说，如果群体中全部都是不利等位基因，即基因型 aa 的频率为1，这时选择将不起任何作用，后代群体中也将全部是纯合基因型 aa 。因此， $q=1$ 是一个平衡状态。但是，如果 q 偏离了1，由于选择的作用，后代群体中的频率 q 将不断下降，最终趋近于0。
 - 因此， $q=1$ 是一个不稳定平衡；而频率 $q=0$ 只是一个局部稳定的平衡，而不是一个全局稳定的平衡。

平衡点依赖于起始基因频率的情况

- 假定起始群体中等位基因A和a的频率分别为 p 和 q ，Aa的适合度为0，AA和aa的适合度为1。选择发生后，等位基因A的频率和频率改变量分别为：

$$p' = \frac{p^2}{p^2 + q^2} \quad \Delta p = p' - p = \frac{pq(2p - 1)}{p^2 + q^2}$$

- 这样，如果起始频率 p 大于0.5，等位基因A在后代群体中将不断增加，最终趋近于1；如果起始频率 p 小于0.5，等位基因A在后代群体中将不断减少，最终趋近于0；如果起始频率 p 等于0.5，等位基因A在后代群体中一直为0.5。
- 因此，对应于起始频率小于、等于和大于0.5，分别存在三个平衡点 $p=0$ 、 $p=0.5$ 和 $p=1$ ，其中 $p=0.5$ 是不稳定平衡，另外两个是局部稳定平衡。

半稳定平衡

- Hardy-Weinberg定律描述的随机交配大群体，是一种更为特殊的遗传平衡。在这个平衡中，后代群体的基因频率与起始群体始终保持相同。每个起始频率代表一个平衡点，有无穷多个平衡点。这样的平衡有时也成为半稳定平衡（semi-stable equilibrium）。
- 对于自交交配系统来说，重复自交无穷多代后，纯合基因型的频率等于起始群体的等位基因频率，并保持不变。这种平衡并未改变基因频率，因此也可以看作是半稳定平衡。