



数量性状遗传分析

之前所涉及的遗传变异规律是针对基因的作用引起生物体表型性状发生质量的变化。然而自然界中随处可见呈连续变异的性状,诸如,田间同一品种数万株水稻的植株高度从低到高有一个很宽的连续变化范围,人们不能明确地归类为矮株或高株;同一品种的玉米果穗的长度也有长短的差别;还有人类的身高、体重乃至体内的总胆固醇含量以及各种酶的活性等等。特别是那些具有经济价值的,大量的农作物与畜、禽产品的品质与产量性状的变异都有其复杂的遗传基础和特殊的遗传规律。遗传的差异与环境因素的影响共同造成了上述复杂性状的表现。本章将讲述呈连续变异的复杂性状的遗传规律及其基因理论。讨论了阈性状及其特征,近亲繁殖与杂种优势。涉及遗传率的估算方法时,本书介绍的是人类多基因遗传病的遗传度(率)的常用方法。此外,本章还重点讲述了运用通径分析的原理计算近交系数和亲缘系数的方法。

9.1 数量性状及其特性

9.1.1 数量性状的概念

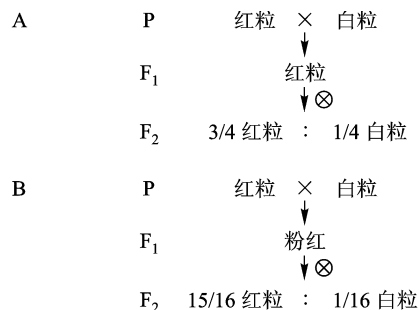
所有能够度量的性状都可称为数量性状(quantitative character或 quantitative trait QT)。这些性状呈连续变异,它不可以严格地分类,而是呈现出系列程度上的差异,带有这些差异的个体没有质的差别,只有量的不同。数量性状包括两大类:一是表现连续变异的性状,如牛的泌乳量、农作物的产量、棉花纤维、羊毛的长度等等;二是表型呈非连续变异,而遗传物质的数量呈潜在的连续变异的性状,即只有超越某一遗传阈值时才出现的性状,如动、植物甚至包括人类的抗病力、死亡率以及单胎动物的产仔数等性状,称为阈性状(threshold character或 threshold trait)。无论是牛的泌乳量还是单胎动物的产仔数,无论是玉米的果穗长度还是动、植物的抗病力,它们大多数对人类具有重要的经济价值,因而直接关系到人类自身的经济利益和生活质量。

数量性状表型的连续性是下列两个现象的结果。第一,一种基因型并不只表达为一种表型,而是影响一组表型的表现。其结果模糊了基因型所决定的不同表型之间的差异,因而不能将一个特定的表型归属于一个特定的基因型。第二,许多不同基因座的等位基因都能使某一种被观察的表型发生改变。现以一个简单例子来说明:假设5个同样重要的基因座影响一种植物每年开花的数目,每个基因座有两个等位基因(+或-),于是每个基因座有3种可能的基因型++、+-、-- ,等位基因“+”与“-”之间没有显隐性关系,每1个“+”表示增加一朵花,每个“-”不增加花。于是总共有 $3^5=243$ 种可能的基因型,但表型只有11种,因为具有“+”的数目为10,9,8,7,⋯,0,例如+++++/+++++,+++++/++++-中间经过+++++/-----一直到-----/-----。因为许多基因型具有相同数目的“+”和“-”。虽然其中只有一种基因型有10个“+”(+++++/+++++),对应的表型有10朵花。但是有51种不同的基因型有5个“+”和5个“-”,例如++++-/+-----,++-+-/+-----等等。所以,许多不同的基因型可能具有相同的表型。

9.1.2 数量性状的多基因学说

(1) 实验依据

毫无疑问,数量性状是由基因所决定,其基因同样具颗粒性并以线性方式排列在染色体上。数量性状的遗传机制是以多基因学说(polygene theory)为基础的基因理论。1909年,瑞典遗传学家 Nilsson-Ehle对小麦和燕麦中籽粒颜色的遗传进行了研究,他发现在若干个红粒与白粒的杂交组合中有如下A, B, C 3种情况。



5, 6...时的各项系数请读者自行推演。

对于 C 组实验的结果分析, Nilsson-Ehle 发现 F_2 中从白色到极深红有 7 种不同程度的红色籽粒, 中间颜色的麦粒最多, 而白色麦粒约占总数的 $1/64$ 。他认为至少有 3 对基因同时遗传。于是有 $3^3 = 27$ 种不同的基因型。其中 $R_1 R_1 R_2 r r r$ 、 $R_1 R_1 r r R_3 r$ 、 $r r R_2 R_2 R_3 r$ 、 $R_1 r R_2 R_2 r r$ 、 $R_1 r r r R_3 R_3$ 、 $r_1 r R_2 r R_3 R_3$ 、 $R_1 r R_2 r R_3 r$ 7 种基因型都应表现为“中等红色”, 因为它们同样只有 3 个 R 基因, 它们在 $(1:2:1)^3$ 的分布中占有 $2/64 + 2/64 + 2/64 + 2/64 + 2/64 + 2/64 + 8/64 = 20/64$ 的比率。同理可预期基因型 $R_1 R_1 R_2 R_3 R_3$ 具有 6 个 R 基因, 因而与亲本有相同的红色, 而基因型 $r r r r r r$ 的表型一定为白色, 因为不具有任何 R 基因, 这两种亲本的表型各占 $1/64$ 的比率。

随后美国学者 Edward M. East 进行了关于烟草 (*Nicotiana glauca*) 花冠长度的遗传学研究。他将花冠的平均长度为 40.5 mm 和 93.3 mm 的纯系亲本进行杂交, F_1 呈中等长度, 如所预期的一致, 但长度稍有变异, 这是由环境的变化所引起的。 F_2 得到 444 棵植株, 其长度的分布在两个亲本品种的平均长度之间, 但比 F_1 有较大范围的变异, 这也是他所预期的。但是 F_2 的花冠长度没有一株像短花冠亲本那样短, 也没有一株像长花冠亲本那样长 (图 9-1)。East 继续将 F_2 中的 3 类植株: 图中分布在左边的短花冠类型、分布在中央的中间类型以及分布在右边的长花冠类型分别进行繁殖, 各获得了 3 类 F_3 植株。结果显示, 来自短花冠的 F_2 植株的后代具有较其亲本更短的花冠平均值; 同样, 花冠长度较大的 F_2 的后代其花冠平均长度也超过其亲本类型。上述 F_3 的结果表明, F_2 的变异不单是环境的影响, 也有遗传的效应。

花冠长度的遗传若由 4 对基因控制, 则预期 F_2 中落在每一亲本类型中的植株的表型频率为 $(1/2)^8 = 1/256$ 。由图 9-1 的分析可知, 在 444 棵 F_2 的植株中没有一个植株是落在亲本类型中, 推

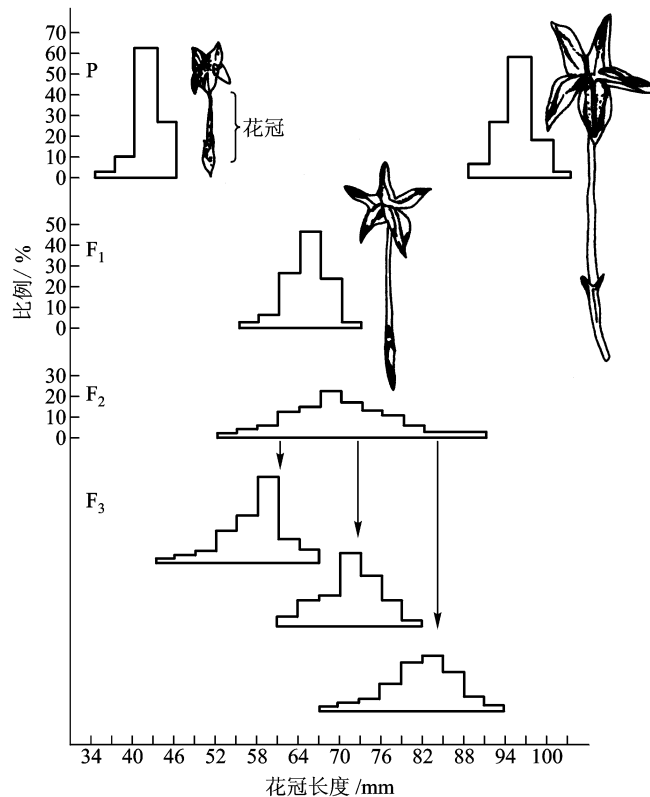


图 9-1 烟草花冠长度的遗传分析

P、 F_1 、 F_2 和 F_3 植株在各类中的比例分布; 花冠长度是连续变异的, 图中每相差 3 mm 分成一组; F_1 和 F_2 的分布在两个亲本的平均长度之间, 但 F_2 有较大变异; F_3 表明, 在 F_2 中观察到的部分方差是遗传的, 因为花冠长度不同的 F_2 植株在 F_3 中出现不同的分布情况, 其花冠的平均长度大致相当于产生它们的 F_2 植株的长度

论,这两种烟草亲本花冠长度的差异可能由 4对以上的基因所决定。

(2) 多基因学说的要点

N ilsson- Ehlk (1909)总结了上述实验分析的结果,提出了数量性状遗传的多基因学说。随后又经统计学家 Fisher及 East等人在玉米、烟草等植物的数量性状遗传的研究中进一步证明和完善,最终形成了解释和分析数量性状遗传的基因理论。其要点是:

- ① 数量性状是许多对微效基因(m inor gene)或多基因(p o lygene)的联合效应所造成的。
- ② 多基因中的每一对基因对性状表现的表现所产生的效应是微小的。多基因不能予以个别辨认,只能按性状的表现一并研究。
- ③ 微效基因的效应是相等而且相加的,故又可称多基因为加性基因(additive gene)。
- ④ 微效基因之间一般不存在显隐性关系。(通常用大写拉丁字母表示增效,小写字母表示减效。)
- ⑤ 微效基因对环境敏感,因而数量性状的表现容易受环境因素的影响而发生较大变化,难以识别个别基因的作用。

⑥ 多基因往往有多效性。多基因一方面对于某种数量性状起微效基因的作用,同时在其他性状上可以作为修饰基因(改变其他基因效果的基因)而起作用,使之成为其他基因表现的遗传背景。

⑦ 多基因与主效基因(m a j o r gene)一样都由染色体所携带,并同样具有分离、重组、连锁等性质。在个别情况下,多基因的效应不是累加而是累积的(如果实的体积),有时,某几对基因间也表现显性,以致表型分布呈现偏态。早在 20世纪 80年代,将分子遗传学所发展的方法和技术应用于数量性状遗传操作,形成了从分子遗传学的角度阐明数量性状遗传的本质的分子数量遗传学。采用现代分子生物学所建立的基因分离、克隆等方法对数量性状进行遗传分析,取得了令人瞩目的进展。当代的数量遗传学家不仅可以影响数量性状的多基因进行剖分并定位在染色体上,测定与之相关的 DNA序列,而且借助于已知的遗传标记追踪每一染色体片段的传递,对数量性状基因座(quantitative trait locus, QTL)作图,将其所在的染色体进行逐段的分析,而且也可以直接测量过去所不能鉴别的各染色体区段的效应,制作 QTL分子图谱和物理图谱,对 QTL进行克隆和序列分析(详见第 18章)。

根据多基因学说,我们可以这样来重新理解和定义数量性状:是由许多对微效基因(或称多基因)的联合效应造成的,表型具有正态分布特征的性状。具有这种性状的个体在正态分布中的位置决定于它们所拥有的微效基因的数目,这种基因多的个体处于分布的上端(或正极),这种基因数目少的个体则处于分布的下端(或负极)。

改良由多基因控制的性状不能单纯依靠孟德尔分离定律来获得有利基因的纯合体,因为其可能性随着控制性状的基因数目的增多而降低,甚至几乎在实际上变得不可能。如果在水稻中某数量性状由 10对基因所控制,则在杂交后代的分离中,期望获得 10对基因完全纯合的个体的机会将只有 $(1/4)^{10}=1/1\,048\,576$,即在 100万植株的群体中才可能出现一株这样理想的纯合体。

由于数量性状不同于质量性状,在研究方法上也具有下列特点:

- ① 在杂交后代中,个别或少数后裔所能提供的信息量很少。研究的单位必须扩大到群体和许多世系才可能获得对其遗传规律和动态变化的认识。
- ② 对个体的性状进行测量或称重,在阈性状方面则予以计数。
- ③ 利用生物统计学的方法,计算性状的表型参数:平均数、方差、协方差、相关系数、回归系数等等。在此基础上进而计算遗传参数:遗传率、遗传相关系数等。

9.1.3 阈性状及其特性

阈性状是指它们的遗传是由多基因决定的,而它们的表型是非连续性的一类性状。这类性状有两种分布,一种是造成这类性状的某种物质的浓度或发育速度潜在的连续分布(以 X表示),它一般是呈正态分布的,或经过统计学变换后成为正态的。另一种分布是表型的可以计数的间断分布(以 P

表示)。当只有一个阈值(threshold, th)时,它由连续分布的一个点(阈值)将个体分为两类:越过 th 者为一类表型,比如“死亡”;未越过 th 的为另一类表型,如“存活”,在这种情况下个体在表型分布中只有两个值,0或1。所以,可以认为阈性状是一种超越某一遗传阈值时才表现的性状。图 9-2 表示发生率为 20% 的阈性状的两种分布的特征。

阈性状是一类重要的数量性状。动、植物包括人类在内的抗病能力如“患病”或“正常”;“存活”或“死亡”;在通常单胎的哺乳动物中的双胎或三胎(在表型上可以简单地分为“单胎”或“多胎”);又如某些哺乳动物的前后肢的指(趾)数,多数个体有正常数目的指(趾),但少数个体可以出现多指(趾)等等,均属于阈性状。只含有一个阈值的阈性状又称为二者居一性状,或称全或无(all or none)性状。阈性状与非阈性状的数量遗传学分析的原理和方法基本相同,但在处理上有所差别。

人类多基因遗传病有唇裂、腭裂、脊柱裂、无脑儿、先天性心脏病、精神分裂症、原发性高血压、冠心病、糖尿病、哮喘等。一般认为是由遗传因素与环境效应共同决定个体是否容易患病,这在医学遗传学中称为易患(感)性(liability)。易患性的变异是呈连续变异的,它表示人体内由基因决定的某种抗体物质的浓度差异。易患性高的个体,抗病力低,当一个个体的易患性超过一定限度——阈值时,该个体即表现为“患病”,性状就表达。连续分布的易患性(X)就被阈值区分出不连续的“发病”与“正常”两类,未越过阈值(俗称门阂)者属于“正常”,个体对某种多基因病的易患性高达一定水平时,即越过阈值者则为“患病”。在一定的环境条件下,阈值标志着患病所必需的最低的相关基因的数目。

个体的易患性无法直接测量,只能根据每个个体婚后所生子女的发病情况作粗略估计。而群体的易患性平均值可以根据阈性状的正态分布原理,从该群体的发病率作估计。例如,当一种多基因病的群体发病率为 2.3% 时,其易患性的阈值与平均值的距离为 2σ 。研究发现,冠心病的群体易患性阈值与平均值的距离基本为 2σ 。当先天性畸形足的群体发病率为 0.13% 时,其群体的易患性与平均值的距离为 3σ 。由此可见,某种多基因病的易患性阈值与平均值相差愈小,则平均值高而阈值低,群体发病率愈高。反之,二者相距愈远,表明平均值低而阈值高,群体发病率愈低。例如某多基因病的群体发病率为 $1/1000$ 可以估计,其阈值较高,距平均值约为 3.1σ 。

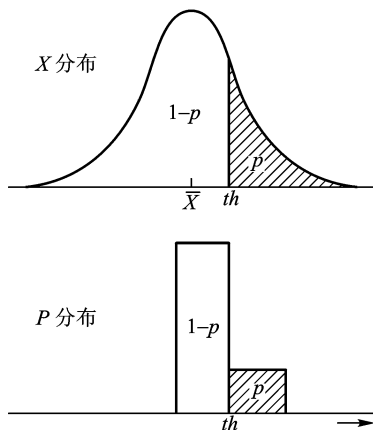


图 9-2 表示发生率为 20% 的一个阈性状的两种分布

9.2 数量性状遗传分析的基本方法

9.2.1 数量性状的遗传率

(1) 表型值及其方差的分量

① 表型值及其剖分 某数量性状的表型值就是实际所度量或观察到的数值。例如:测得某玉米的穗长为 10 cm,该数值就是该性状的表型值。这个表型值受许多外界因素如土壤、肥力、水分、光照、温度等的改变而发生变异,这种变异归因于环境因素。在相同的环境条件下,不同遗传组成的玉米的穗长不同,无疑,遗传组成是影响穗长表型值的一个重要因素。因而,任何一个数量性状的表现都是遗传和环境共同作用的结果。所以性状的表型值首先可以剖分为遗传和环境两个组成部分,用符号表示则有:

$$P = G + E$$

其中, P 表示性状的表型值, G 表示性状的基因型值, E 表示环境效应。

环境效应指的是不同于一般环境对表型值产生的影响, 若基因型值不变或变化很小时, 如果某一植株的生长发育环境优于一般环境, 则 E 为正值, 因而 P 值就大, 如果它比一般环境(正常的、基本的环境条件)要差, E 就是负值, 因而 P 值就小, 通常 E 值是正可负的数值, 同时考虑群体的许多植株时, 它们的正负 E 值相抵消时, 以致 $\sum E = 0$, 于是对于全部植株所构成的群体而言:

$$\sum P = \sum G + \sum E$$

同除以 N 时, 则有
$$\sum P/N = \sum G/N + \sum E/N \quad P = G$$

于是平均的表型值就等于平均的基因型值, 它表示的是全群的某数量性状的遗传水平。

如果对基因型值进一步剖分时, 它由下列 3 部分所组成:

基因的累加效应(additive effect, A)——许多微效基因效应的总和, 被认为性状在上下代遗传传递中可以固定的分量, 又称“育种值”, 即表示在动、植物育种工作中实际能够获得的效应。

显性离差(dominance deviation, D)——等位基因之间有显隐性的关系, 因而造成一些不同于累加效应的情况。如发生等位基因的相互作用, 属于非加性的成分, 称为显性离差, 在群体中它的数值可正可负, 因而 $\sum D = 0$ 。如果等位基因间不存在显性, 则基因型值与基因的累加效应相等。由于基因在不同世代中的分离和重组, 基因间的关系亦会发生变化, 显性离差被认为是能遗传而不能被固定的遗传效应。

交互离差或上位效应(interaction 或 epistatic deviation, I)——由于非等位基因之间的相互作用对于基因型值所产生的效应, 是一种非加性的基因作用, 称为交互或上位效应。若 $I = 0$, 则基因的作用只是加性的, 不存在非等位基因之间的交互。由于非等位基因间的关系比较复杂, 在作遗传分析时常常将其归于环境效应之内, 作为环境效应的一部分, 连同显性离差在内通称为剩余值(residual value, R), 即除育种值之外, 剩余的数值, 这样我们用符号表示时, 有:

$$G = A + D + I$$

因为 $R = D + I + E$, 所以

$$P = A + R$$

② 表型方差及其分量 在群体中, 某数量性状的遗传变异属于总的表型变异的一部分, 表型变异的其余部分是环境变异。在统计学上用方差作为变异的度量。在数量遗传学中从遗传的角度进行剖分, 总的表型方差首先剖分为遗传方差和环境方差, 于是我们有:

$$V_P = V_G + V_E$$

其中, V_P 表示表型方差, V_G 表示遗传(基因型)方差, V_E 表示环境方差。

如果, G 和 E 有相关时, 则:

$$V_P = V_G + V_E + 2 \text{cov}_{GE}$$

如果 G 与 E 不相关, 则:

$$V_P = V_G + V_E, \text{ 或 } \sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

因为基因型方差的组分有育种值方差、显性方差、交互方差,

所以
$$V_P = V_A + V_D + V_I + V_E$$

其中, V_A 表示育种值方差或加性方差, 这是由于多基因的累加效应造成的遗传变异, 能遗传且固定的组分。 V_D 表示显性方差, 由于等位基因间的显隐性关系而造成的一部分非加性的遗传变异, 随着自交代数的增加而逐渐消失, 属于能遗传但不能固定的分量, 可视为杂种优势的组分。 V_I 表示交互方差, 由于非等位基因之间上下位关系所产生的非加性的遗传变异, 此分量也被认为能遗传而不能固定。

(2) 遗传率

数量性状的遗传分析必须解答所有观察到的数量性状表型值的变异是否完全由多基因的累加效应所决定。如果并非如此,则如何判断并量化影响一个数量性状表现的遗传效应和环境效应。进行上述遗传分析时必须借助于一个重要的遗传参数——遗传率(heritability)(在人类遗传学中又称遗传度)。

① 广义遗传率(broad heritability, H^2) 在表型方差即总方差中含有遗传方差的比率定义为广义遗传率,其定义公式为:

$$H^2 = \frac{\text{基因型方差}}{\text{表型方差}} \times 100\% = \frac{V_G}{V_P} \times 100\%$$

亦即表型值受基因型值决定的程度,故遗传率又可称为遗传决定系数(coeffcient of genetic),它是衡量基因型值在表型值中的相对重要性的一个遗传参数。

广义遗传率的概念的提出解决了数量遗传学中争论多年的问题,即在数量性状的表现上,究竟是遗传的作用大还是环境的作用大。对不同的数量性状而言 V_G 和 V_E 的相对大小显然会是不同的;例如,玉米的产量只有 20% 决定于遗传,80% 则决定于环境条件;在玉米株高性状的表型值中则有 70% 决定于多基因的遗传效应,只有 30% 决定于环境影响。所以问题的关键不在于哪一种因素的作用的大小,而在于研究的是哪一种性状。

② 狭义遗传率(narrow heritability, h^2) 表型方差中含有育种值方差的比率定义为狭义遗传率,是用来衡量育种值(基因累加效应)在表型值中的相对重要性的遗传参数。其定义公式为:

$$h^2 = \frac{\text{育种值方差}}{\text{表型方差}} \times 100\% = \frac{V_A}{V_P} \times 100\%$$

某数量性状的遗传率大,说明在该数量性状的表现中,由遗传所决定的比率较大,环境对它的影响较小。通常,与生物适应性无关的性状往往比与适应性有关的性状的遗传率要高一些。现列举几种动、植物包括人类的某些数量性状的遗传率,如表 9-1 所示。

表 9-1 某些数量性状的遗传率

| 种类 | 性状 | 遗传率 |
|------|-------|---------|
| 乳牛 | 泌乳量 | 0.30 |
| | 体重 | 0.37 |
| 猪 | 每窝仔数 | 0.15 |
| | 体长 | 0.55 |
| 来航鸡 | 产卵量 | 0.05 |
| | 成活率 | 0.10 |
| 小麦 | 粒重 | 0.10 |
| 黑腹果蝇 | 翅长 | 0.45 |
| | 产卵量 | 0.18 |
| 人 | 身高 | 0.8 |
| | 血压 | 0.6 |
| | 体质指数 | 0.5 |
| | 口语能力 | 0.7 |
| | 数学能力 | 0.3 |
| | 智力 | 0.5~0.8 |
| | 总胆固醇 | 0.6 |
| | 指纹总峰数 | 0.9 |

遗传率只在确定群体中和特定条件下测定才有意义,因为遗传率不是某个个体的而是个体所在群体的特性,是个体所处的环境的特性。例如,人类身高的遗传率是 0.8,这并不意味着某一个人的身高 80%由遗传因素决定,20%由环境决定,而是指在人类“身高”性状的总变异中,遗传变异占 80%,其他为环境条件的改变所造成的变异。

9.2.2 估计遗传率的方法

以人类群体中多基因遗传病的遗传率(或称遗传度)的计算方法为例说明之。为了掌握其计算方法,下列知识是必要的基础:

具有共同祖先的个体间亲缘程度用亲缘系数(coefficient of relationship, R或 r)来度量。亲缘关系越近,亲缘系数越大,如 $r=0$ 则可认为两个个体间在最近几个世代内没有共同祖先。双亲与其子女、同母同父的同胞间(全同胞)的亲缘系数为 $1/2$;人类遗传学上称这类亲缘关系为一级亲属。二级亲属是指某人及其叔、伯、姑、舅、姨、祖父母、外祖父母之间的亲属,其亲缘系数均为 $1/4$ 。三级亲属是指某人与其表兄妹、堂兄妹、曾祖父母之间,其 r 值均为 $1/8$ 。亲缘系数计算方法详见下节。

根据先证者亲属发病率和一般人群的发病率依下列公式可求 H^2 或 h^2 :

$$h^2 = b/r$$

其中 h^2 表示狭义遗传率, b 表示回归系数, r 表示亲缘系数。

已知一般人群的发病率时,回归系数(b)的计算公式为:

$$b = (x_g - x_r) / a_g$$

若缺乏一般人群发病率的资料时,可调查对照组亲属的发病率,求回归系数

$$b = p_c (x_c - x_r) / a_c$$

一般用 b 表示亲属易患性对先证者易患性的回归系数, r 表示亲缘系数, x_g 表示一般群体易患性平均数与阈值间的标准差, x_c 表示对照组亲属中的易患性平均数与阈值间的标准差, x_r 表示先证者亲属易患性平均数与阈值之间的标准差, a_g 表示一般群体易患性平均数与一般群体中患者易患性平均数之间的标准差, a_r 表示先证者亲属易患性平均数与先证者亲属中患者易患性平均数之间的标准差, q_g 表示一般群体发病率, q_c 表示对照组亲属发病率, $p_c = 1 - q_c$, q_r 表示先证者亲属发病率。

当已知一般群体发病率、对照组亲属发病率和先证者亲属发病率时, x_g 、 x_r 、 a_g 和 a_r 均可查 Falconer 表(表 9-2)得到。其中的 x_g 、 x_r 、a 及其与阈值的关系见图 9-3。

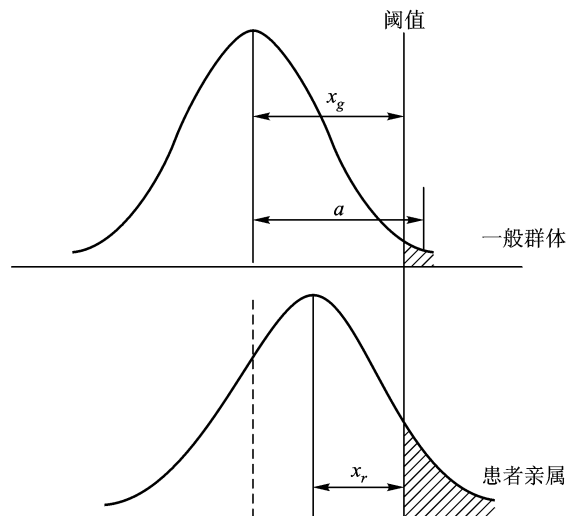


图 9-3 一般群体与患者亲属易患性平均数的比较(引自陈竺, 2004)

现举例说明上述公式的应用。

例 1 某次调查先天性房间隔缺损的疾病在一般人群中的发病率为 0.1%, 在该病 100 个先证者的家系中, 先证者的一级亲属 669 人(其中双亲 200 人, 全同胞 279 人, 子女 190 人)中有 22 人发病。于是可求 $q_r = 22/669 = 3.3\%$ (先证者一级亲属的发病率)。查 Falconer 表(部分数据显示在表 9-2 中, 完整

表 9-2 正态分布的部分 x 和 a 值表(节选自李璞, 2004)

| $q/\%$ | x | a | $q/\%$ | x | a | $q/\%$ | x | a |
|--------|-------|-------|--------|-------|-------|--------|-------|-------|
| 0.01 | 3.719 | 3.960 | 1.02 | 2.319 | 2.658 | 18.2 | 0.908 | 1.452 |
| 0.02 | 3.540 | 3.790 | 1.03 | 2.315 | 2.655 | 18.3 | 0.904 | 1.449 |
| 0.03 | 3.432 | 3.687 | ∴ | ∴ | ∴ | 18.4 | 0.900 | 1.446 |
| 0.04 | 3.353 | 3.613 | 12.5 | 1.150 | 1.647 | 18.5 | 0.896 | 1.443 |
| 0.05 | 3.291 | 3.554 | 12.6 | 1.146 | 1.643 | 18.6 | 0.893 | 1.440 |
| 0.06 | 3.239 | 3.507 | 12.7 | 1.141 | 1.639 | 18.7 | 0.889 | 1.437 |
| 0.07 | 3.195 | 3.464 | 12.8 | 1.136 | 1.635 | 18.8 | 0.885 | 1.434 |
| 0.08 | 3.156 | 3.429 | 12.9 | 1.131 | 1.631 | 18.9 | 0.882 | 1.431 |
| 0.09 | 3.121 | 3.397 | 13.0 | 1.126 | 1.627 | 19.0 | 0.878 | 1.428 |
| 0.10 | 3.090 | 3.367 | 13.1 | 1.122 | 1.623 | 19.1 | 0.874 | 1.425 |
| 0.11 | 3.062 | 3.341 | 13.2 | 1.117 | 1.620 | 19.2 | 0.871 | 1.422 |
| ∴ | ∴ | ∴ | 13.3 | 1.112 | 1.616 | ∴ | ∴ | ∴ |
| 0.47 | 2.957 | 2.911 | 13.4 | 1.108 | 1.612 | 32.0 | 0.468 | 1.118 |
| 0.48 | 2.590 | 2.905 | 13.5 | 1.103 | 1.608 | 33.0 | 0.440 | 1.097 |
| 0.49 | 2.583 | 2.898 | 13.6 | 1.098 | 1.605 | 34.0 | 0.412 | 1.078 |
| 0.50 | 2.576 | 2.892 | 13.7 | 1.094 | 1.601 | 35.0 | 0.385 | 1.058 |
| 0.51 | 2.569 | 2.886 | 13.8 | 1.089 | 1.597 | 36.0 | 0.358 | 1.039 |
| 0.52 | 2.562 | 2.880 | 13.9 | 1.085 | 1.593 | 37.0 | 0.332 | 1.020 |
| 0.53 | 2.556 | 2.873 | 14.0 | 1.080 | 1.590 | 38.0 | 0.305 | 1.002 |
| 0.54 | 2.549 | 2.868 | 14.1 | 1.079 | 1.586 | 39.0 | 0.279 | 0.984 |
| 0.55 | 2.543 | 2.862 | 14.2 | 1.071 | 1.583 | 40.0 | 0.253 | 0.966 |
| 0.56 | 2.536 | 2.856 | 14.3 | 1.067 | 1.579 | 41.0 | 0.228 | 0.948 |
| 0.57 | 2.530 | 2.850 | 14.4 | 1.063 | 1.570 | 42.0 | 0.202 | 0.931 |
| ∴ | ∴ | ∴ | ∴ | ∴ | ∴ | 43.0 | 0.176 | 0.913 |
| 0.93 | 2.353 | 2.690 | 17.3 | 0.942 | 1.479 | 44.0 | 0.151 | 0.896 |
| 0.94 | 2.349 | 2.686 | 17.4 | 0.938 | 1.476 | 45.0 | 0.126 | 0.880 |
| 0.95 | 2.346 | 2.683 | 17.5 | 0.935 | 1.473 | 46.0 | 0.100 | 0.863 |
| 0.96 | 2.342 | 2.679 | 17.6 | 0.931 | 1.470 | 47.0 | 0.075 | 0.846 |
| 0.97 | 2.338 | 2.676 | 17.7 | 0.927 | 1.467 | 48.0 | 0.050 | 0.830 |
| 0.98 | 2.334 | 2.672 | 17.8 | 0.923 | 1.464 | 49.0 | 0.025 | 0.814 |
| 0.99 | 2.330 | 2.669 | 17.9 | 0.919 | 1.461 | 50.0 | 0.000 | 0.798 |
| 1.00 | 2.326 | 2.665 | 18.0 | 0.915 | 1.458 | | | |
| 1.01 | 2.323 | 2.662 | 18.1 | 0.912 | 1.455 | | | |

数据的查找请读者参阅相关的文献)。按照一般群体发病率查得 x_g 和 a_g ; 亲属发病率查得 x_r 和 a_r , 代入

$$b = \frac{x_g - x_r}{a_g} = \frac{3.090 - 1.838}{3.367} = 0.372$$

所以 $h^2 = b / r = 0.372 / 0.5 = 74.4\%$

求得的遗传率说明多基因的遗传效应对该遗传病的影响约为 74%, 环境因素的影响占约 26%。

如果缺乏一般人群的发病率的数据时, 可以选择与病例组相匹配的某对照人群, 调查该对照人群亲属的发病率, 利用先证者亲属和对照组亲属的发病率计算遗传率(度)。

在人类遗传学中, 也可利用一(单)卵双生子(monozygotic twin)即指由一个受精卵发育形成的双生子, 和二(双)卵双生子(dizygotic twin)即指由两个受精卵发育形成的双生子的资料, 按下列公式求遗传率:

$$h^2 = \frac{C_{Mz} - C_{Dz}}{1 - C_{Dz}}$$

其中 C_{Mz} 表示一卵双生子的同病率 C_{Dz} 表示二卵双生子的同病率。

例 2 在对躁狂抑郁性精神病 15 对单卵双生子中的调查, 发现共同患该病者有 10 对, 在 40 对双卵双生子中, 共同患该病者有 2 对。分别算得其同病率为 67% 和 5%, 代入下式求得该病的遗传率:

$$h^2 = \frac{67\% - 5\%}{1 - 5\%} = 0.65 = 65\%$$

下表给出了人类中某些常见多基因遗传病的发病率和遗传率(表 9-3)。

表 9-3 多基因遗传病的发病率及遗传率(摘自陈竺, 2001)

| 疾病 | 群体发病率/% | 先证者一级亲属发病率/% | 遗传率/% |
|----------|---------|--------------|-------|
| 唇裂±腭裂 | 0.17 | 4 | 76 |
| 腭裂 | 0.04 | 2 | 76 |
| 脊柱裂 | 0.3 | 4 | 60 |
| 无脑儿 | 0.5 | 4 | 60 |
| 各型先天性心脏病 | 0.5 | 2.8 | 35 |
| 精神分裂症 | 0.5~1.0 | 10~15 | 80 |
| 原发性癫痫 | 0.36 | 3~9 | 55 |
| 原发性高血压 | 4~10 | 15~30 | 62 |
| 冠心病 | 2.5 | 7 | 65 |
| 青少年型糖尿病 | 0.2 | 2~5 | 75 |
| 哮喘 | 1~2 | 12 | 80 |
| 消化性溃疡 | 4 | 8 | 37 |
| 原发性肝癌 | 0.05 | 5.45 | 52 |

9.3 近亲繁殖与杂种优势

9.3.1 近交及其遗传学效应

在遗传学中, 异型交配(nonassortative mating)是指基因型不同的纯合子之间的交配。相同基因型

之间的交配称为同型交配(assortative mating)。而近交(inbreeding)也称近亲繁殖或近亲婚配,是完全的或不完整的同型交配。近亲繁殖按亲缘关系的近、远程度一般可分为:全同胞(full sib)(同父、母的兄妹)、半同胞(half sib)[同父(或母)异母(或父)的兄妹]和表兄妹(first cousins)之间的交(婚)配。植物的自花授粉、动物的自体受精(self-fertilization)或称自交(selfing),由于其雌、雄配子来自于同一植株(个体)或同一朵花,因而是近亲繁殖中最极端的方式。

(1) 近交使基因纯合, 杂交使基因杂合

以一对等位基因为例,同型交配仅有3种交配类型: $AA \times AA$, $aa \times aa$ 和 $Aa \times Aa$ 。第一、二两种交配类型产生的子代全部都是与亲本相同的纯合体,第三种交配类型属于杂合体间的同型交配。在完全由杂合体之间自交时,其后代将分离为 $1/4AA + 1/2Aa + 1/4aa$ 。经过 n 代连续的同型交配的过程和最终结果显示于表 9-4。

表 9-4 $Aa \times Aa$ 连续 n 代同型交配的结果

| 世代 | 基因型频率 | | | 基因频率 | |
|----------|---------------------|-------------|---------------------|------|-----|
| | AA | Aa | aa | A | a |
| 0 | 0 | 1 | 0 | 0.5 | 0.5 |
| 1 | 1/4 | 2/4 | 1/4 | 0.5 | 0.5 |
| 2 | 3/8 | 2/8 | 3/8 | 0.5 | 0.5 |
| 3 | 7/16 | 2/16 | 7/16 | 0.5 | 0.5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| n | $(2^n - 1)/2^{n+1}$ | $2/2^{n+1}$ | $(2^n - 1)/2^{n+1}$ | 0.5 | 0.5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| ∞ | 1/2 | 0 | 1/2 | 0.5 | 0.5 |

表 9-4显示:① 群体中杂合体(Aa)的基因型频率随同型交配世代的增加而按 $H_n = 1/2H_{n-1}$ 的规律每代减少前一代的 $1/2$ 而迅速递减,可以证明其极限值为 0。② 同型交配群体中纯合体基因型频率每代都有增加,它们以 $\left[\frac{2^n - 1}{2^{n+1}} \right]$ 的规律变迁。可以证明,当 $n \rightarrow \infty$ 时,AA和aa的纯合体频率为 1。故连续进行同型交配的群体最终由两种基因型 AA、aa的个体所组成。如果可以区分它们的话,则成为 AA及aa的两个纯系群体,其频率各占 $1/2$ 。杂合体通过自交导致后代基因的分离,使后代群体中的遗传组成迅速趋于纯合化。另一方面,杂合体通过自交,导致等位基因的纯合,使隐性性状得以表现,从而可以借此淘汰有害的隐性个体,改良群体的遗传组成。③ 完全同型交配的群体中的基因频率并没有改变。当 $H_0 = 1$ 时,每代同型交配的群体中 A、a基因频率永远为 $p = q = 1/2$ 。

在家养动物近交过程中,由于近交个体有限,加上严格的选择。因此,实际上基因频率会发生显著变化,这并非近交自身的遗传效应,而是遗传漂变和选择作用的结果。

杂交的遗传效应则使基因杂合,增加杂合体的频率。杂交实质上是纯合子间的异型交配,子代必然都是杂合体。

(2) 近交系数与亲缘系数

近交系数(coefficient of inbreeding F 或 f)从统计遗传学角度理解,个体(X)的近交系数(F_x 或 f_x)就是结合的配子(即产生个体X的双亲的配子)间的遗传相关系数。如果个体X的双亲是随机交配(婚配)的,则表明双亲配子间的遗传相关系数为 0。如果个体X的双亲是非随机婚配的亲属,则双亲的配子间的相关系数是大于 0而小于 1的某正小数,表明其双亲有共同祖先,因而是有亲缘关系的近亲(或称亲属)。近交系数是度量个体近交程度的重要遗传参数。

具有一个或一个以上共同祖先的个体称为亲属,最基本而又最为重要的亲属关系是亲子关系,即亲代(父、母)与子女的关系。两个个体亲缘程度的度量,定义为亲缘系数(coefficient of relationship R)。亲缘系数愈大,亲缘关系愈近。亲缘系数为 0 则可认为个体(X、Y)间在近期几个世代内没有共同祖先。

(3) 运用通径分析方法计算近交系数和亲缘系数

一种计算近交系数和亲缘系数的简单而又十分便捷的方法是由 S Wright 提出并加以发展的通径分析(path analysis),现简要说明其原理与方法:

① 通径与通径链 在一个相关变量的网络系统中,连接“原因”与“结果”的每一个单箭头线条称为一条通径(path)。比较每一条通径中的“原因”对于某一个“结果”所起作用的相对大小时,需要用通径系数(path coefficient)来表示,即度量各“原因变量”对“结果变量”的直接影响的系数称为通径系数。由一条或一条以上的通径所组成的完整的通道称为通径链。它是从“原因”到“结果”的各条通径的总称。例如,在图 9-4 中,个体 A、B 为“原因变量”(以○表示),个体 X、Y 为“结果变量”(以●表示)。由“原因变量”A、B 指向“结果变量”X、Y 的单箭头分别为 4 条通径,而连接个体 A、B 与个体 X、Y 的有两条通径链即:由 $X \leftarrow A \rightarrow Y$ 及 $X \leftarrow B \rightarrow Y$ 。显然该通径图(图 9-5)即由具有同父、同母的全同胞系谱图(图 9-4)转换而成。

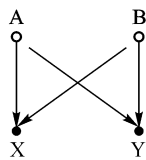


图 9-4 全同胞系谱

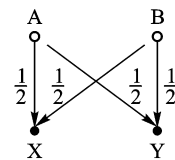


图 9-5 全同胞通径图

② 通径分析的理论证明,在随机交配群体中个体世代的每一条通径的通径系数 = $1/2$ 。而各类亲属关系的特定个体 X、Y 间的亲缘系数的计算公式由通径系数的理论推导证明如(D)式所示:

$$R_{XY} = \sum \left(\frac{1}{2}\right)^L \dots\dots\dots (D)$$

其中, R_{XY} 表示 X、Y 个体的亲缘系数, L 表示沿着某两个特定亲属间(X、Y 间)的连接通径链条中的箭头数, \sum 表示所有这类链条之和。

例如:当我们计算亲兄妹的亲缘系数时,将全同胞系谱图转换成通径图(图 9-5)后,可按上述公式进行快速而准确的计算。

在图 9-5 通径图中由 $X \leftarrow A \rightarrow Y$ 及由 $X \leftarrow B \rightarrow Y$ 两条通径链连接 X、Y 两个特定个体(以●表示),一条经过共同祖先 A,另一条经过共同祖先 B,而 A、B 两个个体不存在共同祖先,所以属随机婚配。每条通径链的系数由各通径的系数相乘,然后对每条通径链的系数求和,则可有

$$R_{XY} = (1/2)^2 + (1/2)^2 = 2(1/2)^2 = 1/2$$

同理可求表(堂)兄妹间的亲缘系数(图 9-6 及图 9-7)。

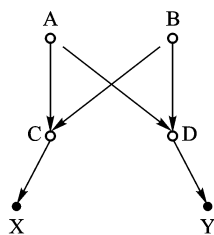


图 9-6 亲表(堂)兄妹婚配

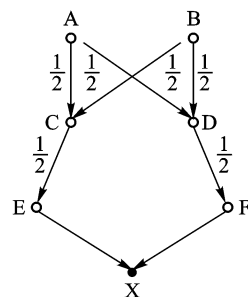


图 9-7 亲表(堂)兄妹婚配的后代 X 的通径图

或根据(1)式先求:

$$R_{II} = (1/2)^5 + (1/2)^5 = 1/16$$

然后依(2)式可求:

$$F_x = R_{II} \times 1/2 = 1/16 \times 1/2 = 1/32$$

两者结果相同。

在此,请读者注意公式(3)中的“N”与公式(1)中的“L”含义不同。

各类亲属的亲缘系数及婚配后代的近交系数列于表 9-5。

表 9-5 几类亲属间亲缘系数及婚配后代的近交系数(仿自杜若甫,2005)

| 亲属关系 | 亲缘系数 | 近交(婚)系数 |
|--|-----------|------------|
| 亲子(图 9-10) | 1/205 | 1/4025 |
| 全同胞(图 9-4及图 9-5) | 1/205 | 1/4025 |
| 舅甥女、叔侄女、姨甥、姑侄(图 9-11) | 1/4025 | 1/80125 |
| 同父(母)异母(父)同胞(即半同胞)(图 9-12) | 1/4025 | 1/80125 |
| 亲表(堂)亲(图 9-6图 9-7) | 1/80125 | 1/1600625 |
| 隔山亲表(堂)亲(图 9-13) | 1/1600625 | 1/32003125 |
| 表舅甥女、表(堂)叔侄女、表(堂)姑侄、表姨甥 [又称隔代表(堂)亲](图 9-8图 9-9) | 1/1600625 | 1/32003125 |

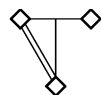


图 9-10 亲子婚(交)配

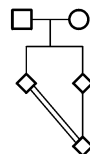


图 9-11 舅甥女等 4种婚配类型

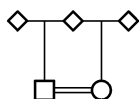


图 9-12 半同胞婚配

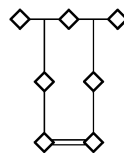


图 9-13 隔山亲表(堂)兄妹婚配

(4) 近交降低群体基因型值的平均值, 杂交则提高群体均值

由前已知,一个数量性状的基因型值是由基因的累加效应值和非累加效应值两部分组成。非累加效应值中的显性效应和大部分上位效应都存在于杂合体中,因此也可大致称非加性效应值为杂合效应值。

(5) 近交使群体分化, 杂交使群体一致

一对等位基因(A, a)遗传的群体中,同型交配的最终结果是使群体分化成两个纯系,AA系与aa系。在两对基因(A, a及B, b)的情况下,分化成4个不同的纯系,即AABB系、AAbb系、aaBB系和aabb系。依此类推。纯系内遗传差异小,方差也小;纯系间差异大,方差也加大,全群的方差也变大。杂交则相反,它能使个体的基因型杂合化,因而使群体趋向一致。因为通过杂交,杂合子频率增加,相当于增加了相互间没有差异的杂合子在群体中的比率,从而加大了群体的一致性。最为典型的情况是两个纯系AA与aa间的杂交,F₁群体中100%为Aa的杂合体,群体达到了完全的一致性。

(6) 近交与人工选择相结合是提高杂种优势的重要手段之一

通过人工选择保留理想的纯合体,近交与人工选择结合能较快地加大群体间基因频率的差异,因

而也成为提高杂种优势的有力手段。玉米自交系间杂种优势比品种间杂种优势高,因为自交系是通过连续自交使双亲基因型的纯合程度都很高, F₁ 群体的基因型才能具有整齐一致的异质性,表现明显的优势。玉米双杂交育种是杂种优势在农业生产中应用的典范。

9.3.2 杂种优势及其遗传理论

杂种优势 (**heterosis hybrid vigor**) 是指杂交子代在生长、成活、繁殖等能力或生产性能等方面优于双亲平均值的现象。目前,有下列两种理论解释杂种优势形成的机制。

(1) 显性说 (**dominance hypothesis**)

由于在基因库中存在不少隐性有害基因或不利基因,它们不同程度地影响生物体的生活力、繁殖力、抗病性等。在数量性状中表现为表型值低于显性的等位基因。这些不利作用可以在杂合体中由于显性等位基因的存在而被不同程度消除。所以,显性说认为杂种优势是由于双亲的显性基因集中在杂种中所引起的互补作用,导致杂合体优于纯合体。

(2) 超显性说 (**overdominance hypothesis**)

此又称“等位基因异质结合假说”。该假说认为杂种优势并非由显性有利等位基因互补作用引起,而是由于杂合体 (**b₁b₂**) 在适应性、生理生化反应能力方面均优于纯合体 **b₁b₁** 或 **b₂b₂**。这一理论认为等位基因间没有显隐性关系。杂种优势来源于双亲基因型的异质结合所引起的基因间的互作。例如,某数量性状由两对基因决定, **a a**、**a a**、**b b**、**b b** 4种纯合体的基因型值分别为 200个单位,而 **a a** 和 **b b** 的基因型值各为 300个单位,显然杂合体的基因型值超过了任何纯合体。其遗传机制有可能由于基因杂合时具有更强的适应环境能力,其功能产物多于纯合体;可能具有可变的合成相关产物的途径;杂合体也可能产生某特殊的杂种物质而有利于杂种的生长发育。

众说纷纭的杂种优势的遗传理论,虽各有其相应的事实依据,可以解释一些遗传现象,但由于各种假说自身的不完善性,有待进一步深入研究。

知识窗 9-1

连锁不平衡及其应用

若同时考虑在同一染色体上两个或多个基因座的等位基因在随机交配群体中的行为与非连锁的基因一样,经过足够多的世代仍然可以达到 **Hardy-Weinberg** 平衡(参阅第 20章),即等位基因间的组合是按照它们各自的频率进行自由组合的,则这些基因被认为是处于连锁平衡 (**linkage equilibrium**) 的。而连锁不平衡 (**linkage disequilibrium**) 则是指一个群体的配子中的连锁基因之间是非随机组合的。亦即,一个基因座上的某些等位基因与同一染色体上另一基因座的某些等位基因一同出现的频率大于单独由机会所产生的频率。例如,在同一染色体上两对等位基因 **A₁A₂**、**B₁B₂**,按基因的自由组合原理可形成 4种不同的配子类型,有时称配子型,扩展的概念又称为单体形。**A₁B₁**、**A₁B₂**、**A₂B₁**、**A₂B₂** 其相应的频率应为相关基因频率的乘积。如果实际的情况恰好如此,则说明该群体中的这两对基因处于平衡状态。换言之,平衡状态下的配子型频率只由相应的基因频率决定。列下表说明连锁不平衡现象。

表 1 两个基因座的连锁不平衡表现

| | | | | |
|------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 基因 | A ₁ | A ₂ | B ₁ | B ₂ |
| 基因频率 | p _A | q _A | p _B | q _B |
| 配子类型 | A ₁ B ₁ | A ₁ B ₂ | A ₂ B ₁ | A ₂ B ₂ |
| 平衡频率 | p _A p _B | p _A q _B | q _A p _B | q _A q _B |
| 实际频率 | r | s | t | u |

由上表可看出： $p_A + q_a = 1$, $p_B + q_b = 1$, 且有 $p_A p_B + p_A q_b + q_a p_B + q_a q_b = 1$, 互斥型配子 ($A_1 B_2$ 、 $A_2 B_1$) 频率的乘积等于互引型配子 ($A_1 B_1$ 、 $A_2 B_2$) 的乘积, 即

$$p_A p_B \times q_a q_b = p_A q_b \times q_a p_B$$

如果这两对等位基因是非随机结合的, 则互斥型与互引型配子的频率的乘积不相等, 则表明产生了连锁不平衡。

设 d 为连锁不平衡程度

$$d = (A_1 B_1 \text{ 的频率}) \times (A_2 B_2 \text{ 的频率}) - (A_1 B_2 \text{ 的频率}) \times (A_2 B_1 \text{ 的频率}) = nr - st$$

若 $nr - st = 0$, 则有 $nr = st$ 则 $d = 0$, 连锁平衡。

若 $nr \neq st$ 或 $nr - st \neq 0$, 即 $d \neq 0$ 时, 则群体处于连锁不平衡状态。

连锁不平衡原理应用于人类复杂疾病(多基因病)的致病基因与遗传标记之间的关联分析和基因定位。在对动、植物数量性状基因座(QTL)定位过程中也有重要作用。

思考题

1. 如何理解阈性状? 常见的阈性状有哪些? 全或无性状的阈值有几个? 为什么?
2. 什么是 QTL? 如何鉴定和定位数量性状基因座? (参阅本书第 18 章作答)
3. 如何理解亲缘系数 (R_{xy}) 及近交系数 (F_x)? 两者的区别与联系是什么?
4. 果蝇腹部刚毛数是一种数量性状。一项对雌性果蝇腹部刚毛数目的研究得到以下方差的估计值:

$$V_P = 6.08 \quad V_G = 3.17 \quad V_E = 2.91$$

试求该性状的广义遗传率 (H^2)。

5. 为什么说某数量性状的狭义遗传率 (h^2) 小于该数量性状的广义遗传率 (H^2)?

6. 试运用途径分析的方法计算下列系谱图中的特殊个体(以“•”表示)的亲缘系数和近交系数(注意将系谱图转换成途径图)。

