

生物学研究概说

群体遗传学

[英] L.M.库克 著



科学出版社

· 生物学研究概说 ·

群 体 遗 传 学

〔英〕 L. M. 库克 著

汉 槎 译

科 学 出 版 社

1 9 8 3

2D70/12

内 容 简 介

本书为“生物学研究概说”丛书之一，是一本基础理论读物。主要内容有群体与基因、基因频率的理论、多态性理论的应用、综合的基因组、变迁的代价及进化的模式等。可供生物学、遗传学、生物数学及生态学的科研及教学工作者参考。

L. M. Cook

Outline Studies in Biology
POPULATION GENETICS

Chapman and Hall 1976

· 生物学研究概说 ·

群 体 遗 传 学

[英] L. M. 库克 著

汉 槎 译

责任编辑 蒋伯宁

科 学 出 版 社 出 版

北京朝阳门内大街 137 号

石 家 庄 地 区 印 刷 厂 印 刷

新华书店北京发行所发行 各地新华书店经售

*

1981 年 7 月 第 一 版 开本：787×1092 1/32

1983 年 5 月 第二次印刷 印张：3 5/8

印数：10,301—14,700 字数：80,000

统一书号：13031·1612

本社书号：2212·13-10

定 价： 0.60 元

目 录

第一章	群体与基因·····	1
第二章	基因频率的理论·····	13
第三章	实际所见的情况——在一些自然群体中的基因 频率·····	35
第四章	在欧洲蜗牛的多态性问题中理论的应用·····	57
第五章	综合的基因组·····	72
第六章	变迁的代价·····	89
第七章	进化的模式·····	98
	为进一步阅读的建议·····	109

第一章 群体与基因

1.1 引言

群体遗传学所研究的是基因在群体中的行为。它涉及到对生物体适应稳定的或变动着的环境的情况的研究，因而也是对进化机制的研究。进化论虽可以从许多不同的水平用各种不同的方法加以研究，但归根结蒂，它无非就是一个基因频率变动的过程而已。所以，群体遗传学家的中心任务就是研究基因频率变化的动力学。

在群体遗传学中所使用的实验技术有许多类型，有的还相当精细，但为了知悉这门学科的范畴，并不一定需要对实验的细节详予了解。相反，所要了解的乃是其内在的概念与理论的发展。就在这个方面，群体遗传学跟遗传学的某些方面有所不同。例如在微生物遗传学中，由 Avery、MacLeod 和 McCarty 所作的细菌转化实验，以及 Lederberg 和 Tatum 所作的接合作用的实验都是主要的起点；要是没有这几位科学家的发现，这门学科就很难开展。在群体遗传学中，其对应的奠基石却要追溯到探讨遗传本质的遗传学创始之时。在三十年代中以此为基础所建立起来的各项理论，一直以这种或那种方式压倒一切地对群体遗传学发生影响。在此领域中的三个人物是 R. A. Fisher，他是 1929 年出版的《自然选择的遗传学理论》（*The Genetical Theory of Natural Selection*）一书的作者；S. Wright，他写了一篇论文《在孟德尔型群体中的进化》（*Evolution in Mendelian Populations*），发表在 1931 年出

版的刊物 *Genetics* 上;以及 J. B. S. Haldane, 他在 1932 年出版了一本书,名叫《进化的原因》(*The Causes of Evolution*)。所有这三位科学家都是在当时探究真理并且聚讼纷纭的气氛中进行研究工作的,他们所作的研究都以与他们同时代的学者与先驱的学者的杰出贡献为根据,可是在他们之间也为这门学科的前景刻划出他们各自的不同见解。在实验或观测数据还没有大量积累之前,理论早已有了发展,从那时候起就已经可能在一个已经设想好的概念框架之内设计一批实验,借以辨明各种理论间的真伪。

这本书也照样从理论的概述开始。然后讨论各项野外的和实验室的例子,以反映理论与观测之间的相互关联性。接着推导一些比较复杂的结论,最后对近代群体遗传学在进化过程的一般知识上所显示的苗头加以讨论。

首先有必要为几项术语下好定义。但生物学家在提出精确或非它性的定义时,往往冒有陷入语义纠缠的困境中去的危险。例如,在群体遗传学中对遗传漂移一词,以及在相邻学科生态学中对竞争和密度相倚性两词,都曾产生过这样的问题。因此在本章的其余部分中,只对群体和基因给出为一般性研究工作用的描述,作为进一步分析用的一个根据。

1.2 群 体

群体乃是生物体在空间里的一个组群(在本书中将只考虑那些二倍体的生物)。每个个体在出生时、或更好地说在配子融合成为合子时,都有一个随龄而定的存活和生殖的潜力,有一个在给定方向中移动给定距离的或然性、有一个增进或压抑其它个体的利益的或然性,等等。物种可能在它的整个生存区域中作均匀的分佈,但也可能东一堆、西一簇。其间所

涉及的各种过程如此之复杂，不可能以上面所说的作为依据而加以描述，因而我们只能简单地考虑发生这些过程的比率，诸如死亡率、出生率、扩散率等。这些比率都是从一个个体的组群的动性所推导出来的平均值。

生存、繁殖等都是生物个体的特征，而生物个体又是生态学和进化论的基础单元。可是在群体遗传学中，为了方便起见，却采用一种更趋极端的简化办法，那就是每次只考虑这些个体的一个性状，并对群体的范畴作出人为制定的界限。假如所考虑的性状是受基因遗传控制的话，那么就可以对下列术语下定义：

1. 群体

群体就是由一群个体所组成的一个品种间杂交的单元。如果性别有公母两种，并且每个个体与每个异性个体交配的机会均等的话，这个群体就称为随机交配群体。

2. 迁移

迁移就是个体从一个群体到另一个群体的流动。迁移跟基因流动这两个术语通常可以替换使用。

3. 在群体中的基因频率

群体中的基因频率就是在一个二倍体位点上一种等位基因占该位点上全部等位基因的比率。

这样就有了一个实际的系统可借以发展理论。要是是一个物种确实包含彼此隔离的随机交配单元，而且所研究的是某个主要基因的动性的话，则上述的图景就相当地现实。可是这种模式往往是理想的假定，即使当它在各方面都并不与真实情况接近也只能这样假定，例如在对人或小鼠或牧草的

研究中所作的那样。一般而论,从这样的简化所得出的结论可能没有什么差别,但是读者和研究工作者都应该经常留意,在把这一模型用之于他所研究的物种的生态学时究竟偏离到什么程度。

1.3 基 因

研究工作者可以从三个角度来理解基因。可以研究它在谱系或育种实验中的动性如何;研究它的化学组成如何;也可以研究它对产生一个表现型的作用如何。为了方便,也不妨去研究那些决定各种分离性状的主基因,它们与那些控制连续变异性状的多基因有区别。研究工作者又正在形成一些分域,而当所得资料并不能妥贴地安放到这些分域中时,就得随时引入好多合宜的术语与想法。对它们的含义予以理解是必不可少的。

1.3.1 主基因

这些都是在单一位点上以不同的等位基因态而存在的各种因素。一个等位基因在对另一个等位基因的效应上可能是显性的,也可能是隐性的或非显性的。这些等位基因显示出孟德尔型的分离作用。不同的位点或者互不相干、或者互相连锁地传代;在一位点上的等位基因对另一位点上的等位基因可能有上位或下位的作用(如果一项两点回交实验或杂交 F_2 代产生出三种表现型),并且上位关系与显性关系不相干。显性和上位是关系到一个基因的表的两个术语,并非指任何内在的性质而言。例如,小家鼠的白化基因对表型来说是隐性的,但对野灰色位点上的各等位基因却是上位的。普通果蝇(*Drosophila melanogaster*)的弯翅(Cy)基因在杂

合体中产生弯曲的翅膀,但当成为纯合体时却是致死性的。因此它对翅形来说是显性的,但在致死性上却是隐性的。在对昆虫的抗杀虫药的主基因的试验工作中,往往可以通过逐渐增加杀虫剂浓度的实验把有敏感性的纯合体从有抗性的纯合体和杂合体分离出来。因此在实验室中,基因显示出无显性。可是在田间或野外,一处只能施用一单种浓度,于是随着所施杀虫药的浓度水准的不同而使基因或呈隐性或呈显性。

通常还有其它三个描述性的术语用之于主基因。一基因的载体如果清楚地与非载体有区别,则称这个基因具有高的外显率。在果蝇中的隐性基因残翅(vg)有高的外显率,总是使翅膀缩剩残端。要是表现型属可变,则表现度也变; vg 则有恒定的表现度。一个基因如果有相当范围的表现型出现,则称该基因是多效性的。以前有个时期曾提出这样的问题:一个基因能否产生多于一种的主产物?现在看来,所有的基因都很可能以多于一种的方式影响于表现型,因此,多效性的程度在一定范畴内有赖于观测者的处理加工。

为了把这几个术语用到一个特殊的例子上去,请看人类中一种叫结节性硬化的遗传性病症。此症引起智力缺乏、皮肤增生、和内脏长瘤。基因属显性遗传,因此所有杂合型的带病基因者都能予以断定。它是多效性的且有可变的表现度——某些带病基因者患这一组病症,另外的带病基因者却患另外不同的一组病症。在致死效应上则有不完全的外显率;有的个体早夭,有的则有较长的寿命。病症之有无虽然取决于一定位点上的分离因素,但是其表现型却属复杂而有变异,并且某些种变异可能还取决于带病基因者的基因型的其它方面的情况。

1.3.2 连续性变异

有许多表现型特征并不能作截然的分类。其中有的性状如高度、重量、叶子形状、或开花时间等都是连续性的；从某种意义上说，其类别必须由研究人员人为地予以制定，例如把长度归到最近的毫米数就是如此。还有诸如产卵数、鳞片数、伞形花序的小花数等都是离散性的，或计数变量，但它们也象连续变量的一类那样具有一个单众数频数分布的倾向。由于对它的研究有赖于度量和计数，所以有时候把它们称作生统变量。很早以前研究进化论的科学家就已经知道这种生统性状是遗传的，至少是部分地遗传，而且在适应性上它们往往还起着极其重要的作用。鉴于孟德尔型遗传学多数以稀少罕见的突变体，如在果蝇中所见到的那些种突变体为例，它们在野生群体中几乎无关紧要，然而象产卵数或开花时间那样的生统性状却有决定性的生态学重要意义。

在十九世纪末，F. Galton 为在双亲与子女的一些连续性状间的关系立出了数学关系式。他把这类关系称为祖先遗传定律。其确切的论点在不同的文章里有不同的说法，都已成为历史上的陈迹。这里不妨以其中一种形式说出这条定律如下：平均而论，在双亲之间为子女提供其总遗传的一半，四个祖亲提供四分之一，八个曾祖亲提供其八分之一，余类推。因此祖先遗传供量和级数 $0.5 + 0.5^2 + 0.5^3 + 0.5^4 + \dots = 1$ ，计及了全部的遗传。

Galton 在钻研“遗传”的统计分析的过程中还发明了回归系数。随着时间的推移，这种数理统计方法却成为高产的科学方法。如果说由 Galton 和继承他的生统学家们所研究的统计性状，在与自然界的适配性上要比由孟德尔学派所研究的那些性状有更加明显的成果，那真是一点也不假。可是在英国

的这两派学者之间关于哪类性状的研究更为重要这个问题上各持不同的看法,从而陷入许多无结果的争辩之中。

一位最强烈地支持生统学派而反对孟德尔学派的学者就是 W. F. R. Weldon。他在 1902 年的一篇论文中(发表在刊物 *Biometrika* 中)批评了孟德尔的豌豆实验结果。他指出了象圆或皱的性状在其本身的类别中就存在着变异,而且变异的数量或类型还可能随着豌豆品系的不同而异。结果,是当有一对个体进行杂交时,光是知道它们所带的圆的或皱的因素是什么还不足以预测其子代的情况,还须知道些关于它们的祖先的更多的情况才行。Weldon 还曾指出我们现在所称的主基因的表现度是可变的,而且它本身是以一种象 Galton 的生统性状那样的非分离方式而遗传的。那末,这类的遗传控制又是如何呢?

对这项问题的答案的发现需追溯到 1900 到 1920 年的时期里。首先,应当指出的是在群体中一种生统性状的表型变异,具有单众数频数分布的形式。它往往是个正态分布,或者通过选用一种适当的变换,可使之化为正态分布。这种分布的特征因而不妨把它总结为分布的平均数与方差。到 1909 年,丹麦的遗传学家 Johannson 曾指出,如果在某性状上具有典型的表型分布的群体进行自交(那时候他正在研究菜豆群体的种子重量),则其方差趋于缩小。在平均数上的效应需视所选留的豆的重量是大的、是小的、还是众数值的而定,但是方差到达于减半时就显示出有不再下降的现象。现在看来,自交或近交的过程是一种使群体中个体的遗传均匀性增加的过程,因此,可以把在方差上的缩小作为在遗传变异性上的缩小看待。如以分数表达,我们不妨说如 V_0 是初始时的表型方差, V_i 是经自交了若干世代后的表型方差,则 V_i/V_0 就是在远系繁殖群体中由于环境效应所产生的方差的分数的分数,剩下的

$(V_0 - V_1)/V_0$ 。就是总方差中由于遗传效应所致的方差的分
数。这样就对连续性变异的遗传的了解已作出了一个开端。

下一步是要证明孟德尔型的分离，有时候也可能以一种
拟连续的方式而出现。另一位与 Johansson 同时的北欧遗传
学家 H. Nilsson-Ehle 研究小麦种子上所表现的红色与白色
的遗传。他证实了有三个位点对此性状有关。每个位点都为
红色与白色的等位基因而分离，并且杂合体的表现总是介乎
两种纯合体之间的粉红色。当把任两个位点使之成为白色纯
合体，只让剩下的一个位点作红白杂交实验时，则从其杂交 F_2
代会产生 1:2:1 的深红、粉红、和白色的分离比率。不仅如
此，在一位点上只要有一个红色等位基因的存在，就会增添由
其它位点的等位基因所控制的色素的强度，不论所涉及的等
位基因的组合如何。从这层意义上看，不妨说在不同位点上
的等位基因的效应具有可加性。因此，有可能当亲本植株在
三个位点上都属杂合体时其所产生的子代中有 1/64 呈白色，
有 1/64 呈深红，其余是深浅程度不同的且在数量上也各异的
中间颜色。事实上，按照孟德尔型的分离模式，其比率应该是
1:6:15:20:15:6:1，其中每个非白色的组在颜色上都比它的
上一组略深一点，但是由于非遗传性的变异把各组间的界限
弄模糊了，使得它们的分布有类似于一个连续性的正态分布
的倾向。因此这个连续变异的性状的遗传分量包括在三个位
点上可分离的等位基因，在它们之间既无显性，也无上位，而
且在效果上是有可加性的。

当然，上面说明连续性变异有可能由有相似效应的好几
个位点所控制的例子并不能证明通常所涉及的都是这类基
因。可是，一旦这项原理得到认可，就有可能设计一些预期性
的杂交实验来检验所作的假定。对各种受分离基因控制的生
统性状的最完整的分析是在 1918 年中由 R. A. Fisher 发表

的一篇杰出的论文中作出的。他在这篇文章里为方差这个术语下了定义(等于标准离差的平方),并且简述了以各种特殊系统的群体平均数和方差为据的各项遗传效应的后果,包括在有控制作用的基因部分中由显性、上位、和连锁等效应所导致的后果。

给一个简单的例子。如从基本群体取出两个品系,分别按某个数量性状的高值和低值而加以选择繁育。这样下去,它们的方差应趋于缩小,但它们的平均数应趋于分化。不容置疑,就内在的遗传控制而论,用其它可能的假定也会得出这一结论。如把高值的近交系与低值的近交系再度使之杂交,

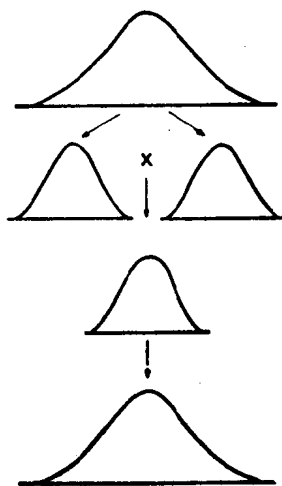


图 1.1 对连续性变异的遗传控制的说明图。从一个基本群体导出两个品系,分别按性状的高值与低值而选育。在此过程中随着那些有控制作用的位点越来越纯化而使方差越来越缩小。在这两品系间的杂交 F_1 代也是遗传均匀的,因为多数个体在这些位点上都属同样的杂合型。但从 F_1 所衍生的 F_2 代则由于发生了遗传重组之故,使方差重新扩大。

则所产 F_1 代的平均数应当介乎两亲本之间,但其方差应保持

不变,因为在此群体中的个体在遗传基因型上都是一致的亲本基因的杂合体。如果从这个 F_1 代再传出一个 F_2 代,其结果将是平均数保持不变,可是由于遗传重组之故,把变异性放出来了,因而使方差扩大(图 1.1)。象这种类型对平均数和方差的改变都可予以记录的实验,其设计的方式应以证实所作的关于遗传控制的本质如何的假定为目的。

1.4 方差的分量

到此为止我们所谈论的都是那些有可加性效应的不相连锁的基因。但在现实的情况中可能还有显性、上位交互作用、连锁的效应、以及其它可能有的效应。从这几种来源所导致的在总的表现型变异中所起作用的部分称为方差的分量。这个数量遗传学的过程成为一个把总的方差分解为它的各项分量的过程。

上位的效应很难与显性的效应分开,这是因为其间的差别取决于表现型的频数分布的形状之故,但这两种来源不妨合并在一起作为方差的非加性分量。它们对遗传模式的影响如何可以回顾 Galton 所创的研究遗传的回归方法去看出来。如果以随机的方式进行一系列的交配,然后把子代的平均值对它们的双亲的平均值在一张坐标纸上作图,则其回归系数就是那根描述两者之间关系的直线的斜率(图 1.2)。不难看出,要是环境对问题中的量纲没有效应,而且要是遗传分量全部由加性基因所组成,使得每个个体都可以精确地度量的话,那么这条直线的斜度应该是 45° ,斜率应是 1,回归系数也是 1。以另一极端而论,要是遗传分量是零,则此回归线的斜率也应为零。因此回归系数就成为方差的加性分量占总的表型方差的分数的多少的一个度量。它被称为该性状的遗传力。由

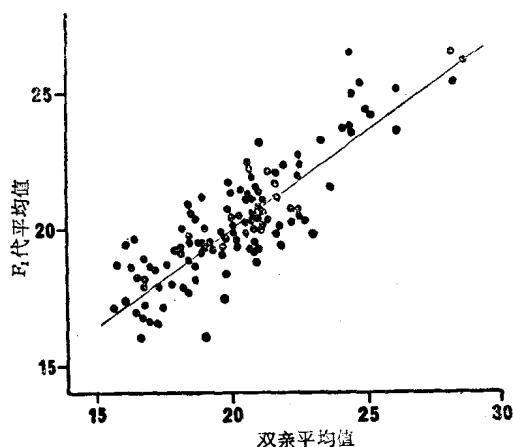


图 1.2 为一种 *Arianta arbustorum* 的螺的壳宽性状所测得的亲代平均与子代 F_1 平均之间的关系(据 Cook, 1965, *Evolution*, 19: 86—94)。其回归系数或遗传力为 0.7, 是个高值。在此例中的亲体是挑选配对的, 这使回归系数略有所增加。如要准确地估计遗传力, 亲体应当取自基因频率已达于平衡(亲代的基因频率等于子代的)的群体并以随机的方式而交配。

于遗传力度量一个育种方案产生响应的能力。因此它在应用数量遗传学中颇为重要。如果子代的平均值不是对双亲的平均值而是对单个亲体的数值作图, 则遗传力为所得回归系数的两倍。

如果群体已经经过许多个世代的近交, 那么在原则上就应已有足够的资料去估计出加性分量 V_A 、非加性分量 V_D 和环境分量 V_E 这三个分量对总的方差 V_P 的相对供量了。从近亲交配的考虑已经看到我们能借以定出分数 $(V_A + V_D)/V_P$ 以及剩下的分数 V_E/V_P 。从回归分析又可以定出另一个分数 V_A/V_P 。从这些资料就可以计算出 V_A 、 V_D 和 V_E 的相对供量。数量遗传学的方法虽然远比这些复杂而细致, 但总是根据以上的原理来安排交配实验, 使其方式足以允许总的表型方差得以划分为各项分量。

在各项分量的相对大小与所研究的性状的表观适应值之间的关系是值得注意的。要是在性状上的变异对拥有这种性状的个体在适应性上并无多大的区别，例如人类指纹图形的性状就是如此，则其遗传力就很可能高。要是一种性状对生命至关重要，例如鸟类的脚爪大小，则其遗传力很可能低，大部分的方差于是归因于非加性的遗传分量。这类性状因此不随环境的变化而变，从而不易因选择而发生变更。要是一种性状可以在生态型上朝有利的方向变更的话，例如某些种分布很广的植物的叶子开裂的程度，则环境分量很可能高。

这类的可预期性使得群体遗传学家们对数量性状的方差的分量分析甚感兴趣。可是实验的结果有时令人失望，这是因为我们虽然认为性状是由在不同位点处相似基因的一个遗传系统所控制的，这些位点有的杂合、有的纯合，但是所得的只不过是些间接的证据。不仅如此，所得的数值随进行该项研究的环境而变。在一种环境中所得出的遗传力的值不一定适用于另一种环境，甚至还可能引入歧途。

一般而论，方差分析虽有其不够完善之处，但它是我们在研究那些控制着连续性变异性状的基因中力所能及的最接近它们的方法。但也有少数的研究工作允许对有关的基因定位，其中突出的有 Thoday 和他的助手们所做的研究工作。他们对在果蝇群体中刚毛数的变异和选择跟孟德尔型的连锁研究结合起来进行实验。把对选择的响应跟在标志位置的频数上的变化相关起来，从而可以鉴定出基因的数目和位置。

第二章 基因频率的理论

微观进化论包括在群体中的变化或者使它在一个变动着的环境中适应于一个新的地点或时间，或者在一个恒定的环境中使它改善其生存能力。其所发生的饰变都属遗传性的。S. Wright 在 1949 年所写的一篇文章中指出：“在一个合理地大的均匀群体中，其基本的进化过程不妨借在基因频率上的变化来研究。”他为在第一章中所简述的简化条件下因过程而变化着的基因频率提供最佳的一般性分类，这在下文中予以使用。

2.1 Hardy-Weinberg 平衡

最有用的起点莫过于零情况，其中有两个等位基因存在于一个群体的一个常染色体位点上，而且并无外力作用以变更它们的频率。如果群体的含量无限，交配属随机，且无选择、突变、或迁移，则不论初始的基因型和基因频率是什么，在经过传一代之后，群体就能获得一个平衡的基因和基因型频率。在本世纪之初曾由许多位科学家对这项结论进行过讨论。为了统一应用，这种平衡最终称为 Hardy-Weinberg 平衡，以一位数学家 G. H. Hardy 和一位遗传学家 W. Weinberg 的姓为名，他们曾在 1908 年在这个问题上不约而同地发表了论文。

假定所研究的位点有两个等位基因 A 、 A' ，其中 A 的基因频率是 p ， A' 的频率是 q ，所以 $p + q = 1$ 。总共有三