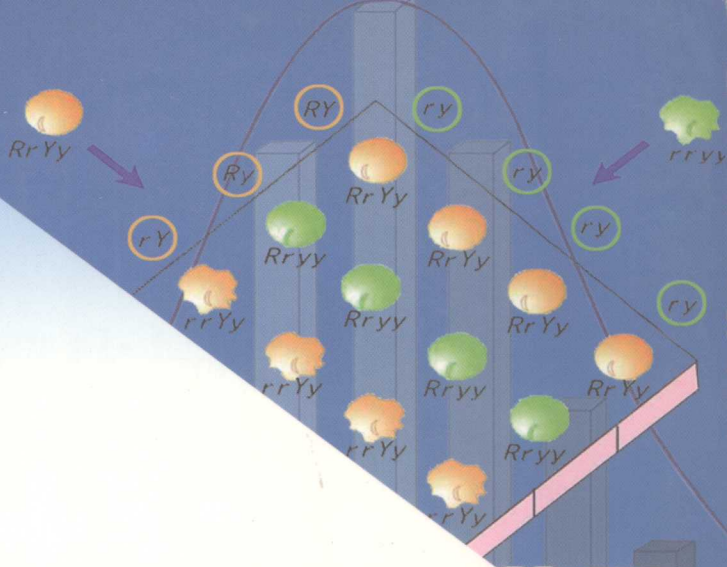




动物群体与数量遗传学

DONGWUQUNTIYUSHULIANGYICHUANXUE

李和平 编著

东北林业大学出版社

动物群体与数量遗传学

李和平 编著

东北林业大学出版社

图书在版编目 (CIP) 数据

动物群体与数量遗传学/李和平编著. —哈尔滨: 东北林业大学出版社,
2004.3

ISBN 7-81076-544-2

I. 动... II. 李... III. ①动物—群体遗传学②动物—数量遗传学 IV. Q953

中国版本图书馆 CIP 数据核字 (2004) 第 012882 号

责任编辑: 姜俊清

封面设计: 彭 宇



NEFUP

动物群体与数量遗传学

Dongwu Qunti Yu Shuliang Yichuanxue

李和平 编著

东北林业大学出版社出版发行

(哈尔滨市和兴路 26 号)

东北林业大学印刷厂印刷

开本 850 × 1168 1/32 印张 10.375 字数 260 千字

2004 年 3 月第 1 版 2004 年 3 月第 1 次印刷

印数 1—1 000 册

ISBN 7-81076-544-2

Q·102 定价: 17.80 元

前 言

遗传学是认识与阐明遗传与变异这一自然现象规律的一门自然科学。生命世界里的生物正是由于遗传才使生物体的特征得以延续，正是由于变异才形成了形形色色的生物，也正是由于生物具有遗传与变异的双重属性才构成了生物进化的基础。当今是生命科学蓬勃发展的时代，作为前沿学科之一的遗传学可以说发展最为迅速，其研究领域随着新技术、新方法的不断涌现更是越来越广泛，研究内容也不断深化。遗传学发展到今天，已经形成了诸多分支：经典遗传学、微生物遗传学、发育遗传学、免疫遗传学、生化遗传学、群体遗传学、数量遗传学、分子遗传学、分子数量遗传学等。

生物种群是以群体状态存在的，种群特征特性的传递和变异总是表现为群体水平的遗传和变异；物种的起源、演变也是以群体形式出现的，而群体遗传学正是一门根据遗传学原理，采用数学、统计学或其他方法研究生物群体的遗传结构、变化规律以及种群演化规律，即在群体水平研究生物遗传变异规律的科学。

生物的性状一般分成两大类：一类是性状的变异可以区分成若干种相对性状并可分别用形容词来描述的质量性状，另一类是性状的变异表现为连续而且只能用数量来表示的数量性状。而对于生物的大多数经济性状来讲都是数量性状，因此研究数量性状遗传规律的科学——数量遗传学在遗传学研究中同样居于极其重要的地位。

群体与数量遗传学是以数学和统计学方法为手段来研究生物的遗传变异的。因此，在学习本课程之前，不仅要具备遗传学特别是孟德尔遗传学的基础知识，而且同时要具备良好的数理统计

知识。考虑到大学相关课程对经典遗传学和数理统计的内容已经有过介绍，所以本教材对这些基础内容不做赘述。

针对林业院校本科生培养规格和课程基本要求，在多年教学实践和资料积累以及分析参阅国内外一些相关优秀教材的基础上，我们编写了这本教材。通过本教材，希望能对林业院校教材建设和教学质量提高有所促进；希望能使学生对群体与数量遗传学有一个基本的了解和掌握；希望能夯实学生的群体与数量遗传学基础知识，便于以后在更广、更深的层次学习和掌握遗传学知识，为遗传学在动植物育种及生产实践等领域的应用奠定基础。

本教材的编写，在内容上以群体与数量遗传学基本知识为主，力求深入浅出、简明扼要、推导尽量简单，在名词术语和符号使用上大多尊重习惯并且各章节尽量统一，避免代号混乱和重复。

本教材主要适合于林业院校的野生动物与自然保护区管理、生物科学、生命科学等专业本科学生使用，也可供其他相关专业本科生或研究生参考。

本教材的出版得到了东北林业大学优秀教材出版基金的资助，得到了有关领导、老师和同志们的关注与支持，同时东北林业大学出版社也积极推动本教材的编写与出版，在此一并表示感谢！

由于群体与数量遗传学领域知识的深奥和浩瀚，加之编者水平所限和时间匆忙，因此书中疏漏和错误在所难免，衷心希望大家提出宝贵意见，不吝赐教。

编 者

2003 年 8 月

目 录

第一章 群体的遗传结构	(1)
第一节 群体与基因库	(1)
第二节 基因频率	(2)
第三节 基因型频率	(4)
第四节 随机交配	(9)
第二章 Hardy - Weinberg 定律	(12)
第一节 Hardy - Weinberg 定律的含义及要点	(12)
第二节 Hardy - Weinberg 平衡群体的基本性质	(23)
第三节 基因频率与基因型频率的推算	(28)
第四节 群体遗传平衡的检验	(32)
第三章 Hardy - Weinberg 定律的扩展	(37)
第一节 复等位基因	(37)
第二节 伴性基因	(43)
第三节 多基因位点	(49)
第四章 影响 Hardy - Weinberg 平衡的因素	(62)
第一节 突变	(63)
第二节 选择	(72)
第三节 突变与选择的联合作用	(109)
第四节 迁移	(119)
第五节 选择与迁移的联合作用	(124)
第六节 遗传漂变	(128)
第七节 随机交配的偏移	(138)
第五章 理想与实际小群体	(143)
第一节 理想小群体	(144)
第二节 实际小群体	(161)

第六章 数量性状特征及统计学基础	(170)
第一节 数量性状及其特征.....	(170)
第二节 数量性状遗传分析的统计学基础.....	(184)
第七章 数量性状的表现型分量及其方差	(190)
第一节 数量性状表型值及其方差的分量.....	(190)
第二节 群体基因型值的平均数.....	(194)
第三节 一个基因的平均效应.....	(198)
第四节 基因型的育种值.....	(199)
第五节 显性离差.....	(199)
第六节 交互作用(上位)离差.....	(201)
第七节 方差的遗传分量.....	(203)
第八章 数量性状的遗传参数	(204)
第一节 遗传率.....	(204)
第二节 重复率.....	(219)
第三节 遗传相关.....	(226)
第九章 近亲繁殖与杂种优势	(231)
第一节 近交与杂交的遗传效应.....	(231)
第二节 杂种优势的遗传理论.....	(247)
第十章 数量性状基因位点分析	(250)
第一节 经典遗传学对数量性状基因数目的估计.....	(250)
第二节 QTL作图原理及步骤.....	(252)
第十一章 遗传与进化	(263)
第一节 自然群体中的遗传多态性.....	(263)
第二节 进化学说.....	(274)
第三节 群体遗传多态性的保持方式.....	(278)
第四节 物种形成.....	(283)
第五节 分子进化.....	(295)
第六节 种群间遗传分化的度量.....	(305)

练习题.....	(309)
练习题参考答案.....	(319)
参考文献.....	(323)

第一章 群体的遗传结构

群体遗传学 (population genetics) 是研究群体的遗传结构及其变化规律的科学, 而体现群体遗传结构的主要标志就是基因频率和基因型频率, 基因频率和基因型频率的变化就表现了群体在遗传上的变化, 这些变化一般情况下要受到群体大小、交配方式、基因突变、选择 (自然选择或人工选择)、迁移及群体隔离状态等因素的影响。以下就与群体遗传学研究有关的基本概念加以阐明。

第一节 群体与基因库

经典遗传学中的孟德尔遗传定律是在“家庭水平”研究生命遗传现象的基础上而发现的遗传规律。“家庭水平”上的研究是仅限于特定亲本及其后代间遗传关系的研究, 但是我们知道在生物进化或动物育种过程中, 进化的单位或育种的对象不是个体或“某个家庭”, 而是群体, 因为一个个体或一个家庭的生命在时间的长短上是有限的, 并且个体的遗传结构, 除了发生突变之外, 在一生中都是固定的。相反, 一个群体实际上是不死的, 群体可能很大或者很小, 群体分布可能广泛或者狭窄, 一个群体能世代相传, 其存在不受群体成员中个体生命的限制。群体的遗传性, 由于遗传的机制使得群体在时间上有连续性, 群体的遗传结构, 由于可以受到多种因素的影响而一代一代地可能突然或逐渐地发生改变, 这些对于生物进化和育种都是有利的。

群体遗传学所针对的群体不是诸多个体的简单集合, 而是一

种特定的孟德尔群体 (Mendilian population), 即一个物种生活在某一地区内的、能相互交配繁殖的个体群。在群体遗传学中, 将一个群体中所有个体共有的全部基因或一个群体所具有的全部遗传信息定义成为一个基因库 (gene pool)。孟德尔群体和基因库这两个概念用杜布赞斯基 (Th. Dobzhansky, 1955) 的话说就是“一个孟德尔群体, 是一群能够相互繁殖的个体, 它们享有一个共同的基因库”。由此可知, 在有性繁殖的某一个生物物种就可以认为是一个最大的孟德尔群体。通常情况下, 分布在同一地区或空间的某一个物种, 如果个体间有基因的自由交流就可以看做组成了一个单一的孟德尔群体, 但是, 如果由于自然地理或人为因素的条件限制阻碍了其中个体间基因的自由交流, 那么有可能未必属于单一的孟德尔群体, 各自保持着各自不同的基因库, 这时就产生了同一地区共存着同一物种的几个孟德尔群体的状况。

第二节 基因频率

研究群体的遗传变化, 或者表达群体的遗传组成变化的机制, 首先要分析群体基因频率的变化, 基因频率是群体遗传组成的基本标志, 不同群体的同一基因的频率往往不相等; 其次是要分析由不同交配体制所导致的各种基因型在群体中数量上的变化。总之, 群体遗传学就是用基因频率和基因型频率的形式来分析、研究群体遗传和结构变化的机制的。

基因频率 (alleles frequency, gene frequency) 是指群体中某一基因的数量, 正是由于等位基因的存在才知道了某一基因的存在, 也才能计算其在群体中的频率, 因此, 基因频率准确的含义应该是等位基因频率, 指的是在一个二倍体的某特定基因位点上某一个等位基因占该座位 (位点) (注: 本书中使用的“基因位点”与“基因座”具有同等含义) 上等位基因总数的比率, 或者

说是在一个群体中某一基因对其等位基因的相对比率。这是群体遗传结构的一个最基本的测度或标志。

对于二倍体的某特定基因位点上的等位基因来讲，同一物种的不同群体的等位基因频率可能存在差异。例如某黑白花奶牛群体大多数个体有角，只有个别个体无角；某安格斯牛群体几乎全部个体无角；而一些杂种牛群体中有角个体与无角个体的数目比例更是千差万别。那么控制这些牛群在“角”这个性状上表现不同的原因是不是由于基因的不同而导致的呢？经过大量研究认为：不是的。研究结果表明：在大多数的牛群，控制“角”这个性状的基因都是相同的，都是由一对等位基因控制的；决定“无角”的基因是显性（dominant）的，用 P 表示；决定“有角”的基因是隐性（recessive）的，用 p 表示；只是在不同的牛群， P 基因与 p 基因的比率即基因频率各不相同而已。假如我们现在知道某牛群中控制“无角”的基因 P 的频率是 1%（或 0.01），控制“有角”的基因 p 的频率是 99%（或 0.99），那么我们就可以知道这个牛群中只有约 2% 的个体是无角，其余个体都是有角的。再假如另一个牛群中， P 基因占 100%（或 1.00）， p 基因占 0%（或 0），则这个牛群的所有个体都是无角的。

群体基因频率的表示一般用百分数（由于基因频率是一个相对比率）或小数的形式来表示，其变动范围在 0% ~ 100% 或 0 ~ 1 之间，但是，一般情况下基因频率都写成小数形式且没有负值。前文的 1%（或写做 0.01）和 100%（或写做 1）就是 P 的基因频率，99%（或写做 0.99）和 0%（或写做 0）就是 p 的基因频率。

群体中各等位基因的基因频率的总和等于 1，这是一条很重要的结论。例如前文述及的两个牛群中，前一群体的 $0.01 + 0.99 = 1$ 和后一群体的 $1 + 0 = 1$ 就说明了这一点。再例如人的 ABO 血型由三个等位基因 I^A 、 I^B 和 i 决定，据调查中国昆明人种群中 I^A

的基因频率约为 0.24, I^B 的基因频率约为 0.21, i 的基因频率约为 0.55, 三者的总和 $0.24 + 0.21 + 0.55 = 1$ 。

第三节 基因型频率

基因在二倍体生物的体细胞中都是成对存在的, 而且由一对或几对基因组成了控制某一性状的基因型, 因此在生物的遗传上, 一个性状的遗传特性不仅决定于基因, 而更重要的是决定于基因型。

基因型频率 (genotype frequency) 是指群体中某特定基因型个体的数目, 占个体总数目的比率, 或者说是一个群体中某一性状的各种基因型间的比率。例如前文提到的牛群中角的有无, 决定于一对等位基因 P 和 p , 由这一对等位基因组成的基因型有 PP 、 Pp 和 pp 三种, 前两种在性状上表现无角, 后一种表现为有角, 这三种基因型的频率在各群体中是有差别的。如果某一群体的 PP 占 0.01%, Pp 占 1.98%, pp 占 98.01%, 那也就是说, PP 基因型频率为 0.0001, Pp 基因型频率为 0.0198, pp 基因型频率为 0.9801, 三者之和等于 1, 即 100%。

我们知道, 基因型不等于表现型, 也就是说, 基因型频率并不是表现型的比率。例如上述三种基因型, 它们的表现型只有两种, 无角与有角, 两者的比率是 1.99% ($0.01\% + 1.98\%$) 和 98.01%。

基因型是由基因组成的, 所以两者的频率是密切关联的。一般地讲, 群体中某一基因频率可以从相关的基因型频率推算出来, 特别是对于共显性遗传的性状更容易从性状的表现来推算基因型频率, 进而计算出基因频率来, 如人类的 MN 血型遗传等。然而, 如果等位基因之间在表型效应有显隐性之分, 显性纯合子 (如 PP) 与杂合子 (如 Pp) 均显示显性性状而无法区分, 这

时想通过基因型来推算基因频率就不如共显性遗传基因的基因频率容易计算了，但是，如果群体的等位基因处于遗传平衡时，即使是表型效应有显隐性区分的性状，同样可以计算出群体的基因频率。

下面我们通过基因频率与基因型频率两者之间的密切关系，来推导两者的计算方法。现首先从一般的例子入手，假定某 100 个个体组成的二倍体群体的某一基因位点有一对等位基因 A、a，其三种基因型 AA、Aa、aa 的数目见表 1-1。

表 1-1 等位基因 A、a 及其基因型的数目

基因	基因型			总数
	AA	Aa	aa	
个体数	10	60	30	100
A 基因数	20	60	0	80
a 基因数	0	60	60	120

对于二倍体生物来讲，每个个体的某一基因位点具有 2 个基因，那么该 100 个个体的群体共有 200 个基因。群体内的基因都存在于特定的基因型中，基因型 AA 个体具有 2 个 A 基因，Aa 个体具有 1 个 A 和 1 个 a 基因，aa 个体具有 2 个 a 基因。因此，该群体共具有 80 个 A 基因和 120 个 a 基因。于是，根据基因频率的定义，A 基因的基因频率 p 为

$$p = \frac{20 + 60}{200} = 0.4$$

同理可以计算 a 基因的基因频率 q 为

$$q = \frac{60 + 60}{200} = 0.6$$

通过上例，现假设在 N 个个体组成的群体中有一对等位基因 A、a 在常染色体遗传，那么其可能的基因型为 AA、Aa、aa，

共 3 种。如果群体中 AA、Aa、aa 基因型的个体数分别是 n_1 、 n_2 、 n_3 ，即 $n_1AA + n_2Aa + n_3aa$ ，且 $n_1 + n_2 + n_3 = N$ 。那么，可以用一般化的形式，将基因型所具有的个体数与基因频率之间的关系列于表 1-2。

表 1-2 基因型个体数与基因频率之间的关系

基 因	基因型			总 数
	AA	Aa	aa	
个体数	n_1	n_2	n_3	N
频率	$D = n_1/N$	$H = n_2/N$	$R = n_3/N$	1
A 基因数	$2n_1$	n_2	0	$2n_1 + n_2$
a 基因数	0	n_2	$2n_3$	$n_2 + 2n_3$
等位基因总数				$2N$

根据表 1-2，如果用 D 、 H 、 R 分别表示三种基因型 AA、Aa、aa 的频率，那么此三种基因型的频率为

$$AA:D = \frac{n_1}{N}; \quad Aa:H = \frac{n_2}{N}; \quad aa:R = \frac{n_3}{N}$$

$$\text{而且 } D + H + R = \frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} = \frac{n_1 + n_2 + n_3}{N} = \frac{N}{N} = 1$$

A 和 a 基因各自的基因频率为

$$p = \frac{2n_1 + n_2}{2N} = \frac{2n_1}{2N} + \frac{n_2}{2N} = D + \frac{1}{2}H$$

$$q = \frac{n_2 + 2n_3}{2N} = \frac{2n_3}{2N} + \frac{n_2}{2N} = R + \frac{1}{2}H$$

上式表达了群体基因频率与基因型频率之间的关系，即两者关系的公式：

$$p = D + \frac{1}{2}H$$

$$q = R + \frac{1}{2}H$$

而且，群体中同一基因位点的基因频率之和： $p + q = D + \frac{1}{2}H + R + \frac{1}{2}H = 1$ 。

由公式 $p = D + \frac{1}{2}H$ 、 $q = R + \frac{1}{2}H$ 可知，群体中某一基因（A 或 a）的基因频率（ p 或 q ）是该基因的纯合体（AA 或 aa）频率（ D 或 R ）与杂合体（Aa）频率（ H ）的 $1/2$ 之和。因而，在计算群体基因频率时，只要知道各基因型的频率就可以了，不一定需要各种基因型在群体中的个体数；而且由于 $p + q = 1$ ，所以只要计算出 p 或 q 之一，则另一个由 1 减去它的值即可得出；也就是说，只要计算出群体的基因型频率，就可以计算出群体的基因频率。但是，反过来，只知道基因频率，却不一定能知道基因型频率，这一点是显而易见的；例如，现群体中有一对等位基因 A、a，两者的基因频率相等，即 $p = q = 0.5$ ，那么，群体中 AA、Aa、aa 三种基因型的频率可以是 0.25、0.5、0.25 或 0.3、0.4、0.3，也可以是 0.0、1.0、0.0 等。为了能用基因频率计算基因型频率，往往还需要其他一些条件。

至此，我们可以清楚地知道群体的遗传组成由基因型频率和基因频率来表示，但需要说明的是，在实际分析群体时，往往首先考虑的用什么方法知道群体表现型的频率，在查明了其性状遗传方式的基础上，再根据表现型频率来计算基因型频率或基因频率。如果控制某性状基因的遗传属于不完全显性或共显性的遗传方式，那么其表现型与基因型是对应的，即其表现型频率就等于基因型频率，这样的话就为计算群体基因频率和基因型频率带来了便利。

例如，在人类群体中，由常染色体上的一对等位基因 T、t 遗传的性状，表现为对苯硫脲的尝味能力不同，而且基因 T 对 t

为不完全显性，即其表型与基因型是完全对应的。在某次抽样调查中国汉族人群中 1 000 人对苯硫脲的尝味能力时，发现其中 TT、Tt、tt 3 种基因型（表型）的分布见表 1-3。

表 1-3 中国汉族人群中 PTC 尝味能力的分布

表 型	基因型	人数	基因型频率	基 因	
				T	t
尝味者	TT	(n_1) 490	(D) 0.49	980	
味觉杂合体	Tt	(n_2) 420	(H) 0.42	420	420
味 盲	tt	(n_3) 90	(R) 0.09		180
总 计		1 000	1	1 400	600

根据表 1-3 中的每种基因型的人数除以总人数得到相应的基因型频率。如 TT 基因型频率是 $490/1\,000 = 0.49$ ，余类推。

基因 T 在群体中的频率为： $p = 1\,400/2\,000 = 0.7$ ，等位基因 t 的频率 $q = 600/2\,000 = 0.3$ 。

等位基因频率也可根据前面的公式 $p = D + \frac{1}{2}H$ 和 $q = R + \frac{1}{2}H$ ，由基因型频率求得：

$$p_{(T)} = D + \frac{1}{2}H = 0.49 + \frac{1}{2} \times 0.42 = 0.7$$

$$q_{(t)} = R + \frac{1}{2}H = 0.09 + \frac{1}{2} \times 0.42 = 0.3$$

而且有： $p + q = 0.7 + 0.3 = 1$

再例如：MN 血型是共显性遗传的性状，它决定于 4q28 ~ q31 上的等位基因 M 和 N，M 血型决定于纯合的基因型 MM，N 血型决定于纯合的基因型 NN，MN 血型决定于杂合的基因型 MN，因此，人类群体遗传学一直用它来作为群体的遗传结构研

究的对象。假设有人曾对 1 000 人的 MN 血型进行了调查, 结果发现 M 血型的有 360 人, 占 36%; N 血型的有 120 人, 占 12%; MN 血型的有 520 人, 占 52%。如果用 p 表示 M 的基因频率, 用 q 表示 N 的基因频率, 那么根据基因型频率和基因频率的定义可知:

三种基因型 MM、MN、NN 的频率分别是 0.36、0.52、0.12。
等位基因 M、N 的频率分别是:

$$p = MM + \frac{1}{2} MN = 0.36 + \frac{1}{2} \times 0.52 = 0.62$$

$$q = NN + \frac{1}{2} MN = 0.12 + \frac{1}{2} \times 0.52 = 0.38$$

等位基因 M、N 的基因频率之和 $p + q = 0.62 + 0.38 = 1$

另外, 在这里值得一提的是当基因伴性遗传 (详见后面有关章节) 时, 我们可以把雌雄分别看做是两个群体, 现假设配子异型的是雄性 (如哺乳动物), 在这种属于雄异型生物的雌性群体中, 基因频率与基因型频率的关系与前文叙述的常染色体基因一样, 只是在雄性群体中, 基因型只有两种 ($X^A Y$ 、 $X^a Y$), 这时雄性群体的基因频率和基因型频率是相等的。如果配子异型是雌性 (如禽类), 那么情况就正好相反。

群体遗传学, 最重要的是群体中的等位基因频率, 因此以等位基因频率作为表示群体的遗传组成的基本度量。

第四节 随机交配

在群体遗传学研究中, 随机交配和“泛交”是同义语, 随机交配 (random mating) 是一个重要的原则和基石, 在后面叙述的群体遗传学内容时, 除非说明; 否则都是针对的随机交配的情况。随机交配的经典定义是: 在一个有性繁殖的生物群体中, 任