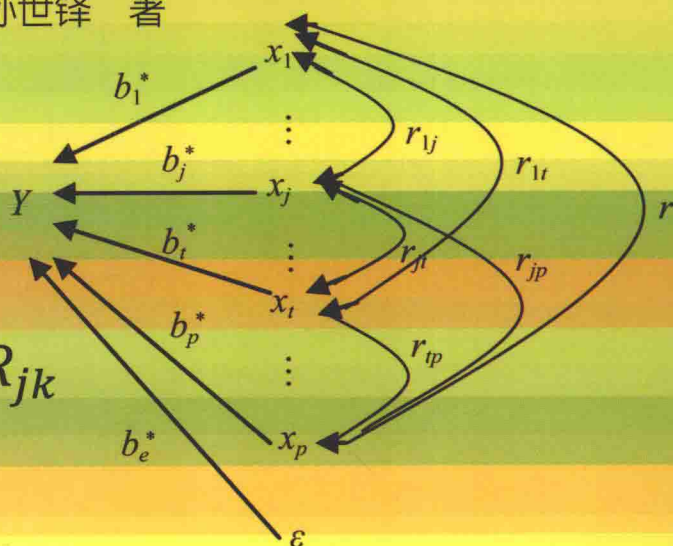


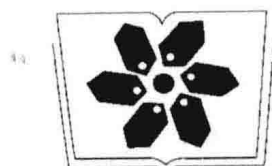
GENETIC ANALYSIS
OF QUANTITATIVE TRAITS

数量性状 遗传分析

袁志发 常智杰 郭满才 孙世铎 著

$$R_{(j)} = R_j^2 + \sum_{k \neq j} R_{jk}$$





中国科学院科学出版基金资助出版

数量性状遗传分析

袁志发 常智杰 郭满才 孙世铎 著

科学出版社

北京

内 容 简 介

本书的主要内容由两部分组成。一部分是基于经典遗传学的数量遗传学。首先,在多基因遗传体系假说下,讲述了数量性状多基因间的有关加性、显性和上位效应内含的遗传统计学模型;定义了基因在群体中的平均效应和个体的育种值,探讨了基因型值和育种值间的关系;讲述了品系间杂交有关近交群体的伯明翰学派的遗传分析特点,阐明了远交群体的北美学派的遗传分析方法;论述了遗传力、重复力和相关的定义、估计方法和假设检验;讲述了单性状选择及综合选择指数的有关概念、分析和应用;论述了作者提出的组合性状和组合性状对的遗传分析和应用;讲解了交配效应和配合力分析。在主基因-多基因遗传体系之下,简介了它的分离分析的理论和方法,并探讨它的综合选择指数的分析方法。第二部分是在数量性状的分子遗传学研究上,简述了QTL定位、分子标记辅助选择和基因组功能分析。在有关多性状选择基因组功能分析中,利用通径分析及其决策分析给出了各因子的直接作用、间接作用和总作用,并对各因子在选择和功能上的主次作用进行了判断。

本书适合遗传育种工作者、应用数学工作者,以及有关专业的研究生教师作为参考书或教材。

图书在版编目(CIP)数据

数量性状遗传分析/袁志发等著. —北京:科学出版社, 2015.8

ISBN 978-7-03-043582-6

I. ①数… II. ①袁… III. ①数量性状—遗传分析 IV. ①Q348

中国版本图书馆 CIP 数据核字(2015)第 206143 号

责任编辑:王 静 王丽平 李 悦 / 责任校对:邹慧卿

责任印制:徐晓晨 / 封面设计:北京铭轩堂广告设计公司

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2015 年 8 月第 一 版 开本:787×1092

2015 年 8 月第一次印刷 印张:28 1/4

字数:670 000

定价:148.00 元

(如有印装质量问题,我社负责调换)

前 言

数量遗传学是研究数量性状遗传规律并服务于动植物育种的一门学科。对数量性状进行遗传操纵的理论和方法既是数量遗传学的中心内容,也是数量遗传学应用于育种实践的桥梁和手段。因而,任一遗传学分支服务于育种的研究均可纳入数量遗传学的理论体系,可见数量性状遗传分析内涵的丰富和发展潜力。

接触和学习数量遗传学是源于我的一个朴素的想法:农学院的数学从主流上应走与农业科学相结合的道路。因而,除自学遗传学等基础生物学之外,吴仲贤教授的《统计遗传学》和马育华教授的《植物育种的量遗传学基础》几十年来始终伴随着我,其浓厚的育种机理和透彻的数理分析使我如临教诲。“文化大革命”后重招研究生起,我为研究生讲授“生物数学”、“试验设计与分析”、“数量遗传学”和“群体遗传学”课程,并协助赵洪章院士和邱怀教授指导研究生,使我受益匪浅。1992年,我被评为国家动物遗传育种与繁殖博士生导师,开始了以数量遗传学和群体遗传学为主要选题的研究生培养工作。1981~2014年,我和我的研究生发表了40余篇数量性状遗传分析的论文,提出了通径分析的决策分析、组合性状和组合性状对遗传分析、小麦生态型及其演变的统计分析方法和KEGG通路的通径分析等,结合数量遗传学的发展,写成了本书。全书共7章,前3章叙述数量遗传学的研究基础,近交群体和随机交配群体的遗传分析,重复力和遗传力等相关概念、估计和检验;第4章讲述单性状选择、多性状选择和组合性状与组合性状对分析;第5章介绍主基因-多基因混合遗传体系下的单分离世代和多世代联合的分离分析及综合选择指数;第6章叙述交配效应与配合力分析;第7章讲述数量性状遗传与分子遗传学,主要内容包括QTL定位、DNA标记辅助选择的标记值选择和指数选择、KEGG通路的通径分析和决策分析。

本书编写由袁志发和常智杰负责,并在杜俊莉博士、解小莉博士、刘建军副教授、曲高平硕士和马誉伟硕士的协助下完成。

在几十年的学习和研究中,周静芋教授是我最尊重的合作者。常智杰教授、郭满才教授、孙世铎教授、翟永功教授、贾青教授、雷雪芹教授、徐廷生教授、陈玉林教授、张恩平教授、秦豪荣教授、吉俊玲教授、王春平教授、郑惠玲副教授、宋世德副教授、解小莉副教授、刘璐副教授、刘建军副教授、杜俊莉博士、邵建成、王丽波、董晓萌、陈小蕾、曲高平、罗凤娟、马誉伟等,我们既是师生,又是良友。本书是我们共同学习、研究和写作的成果。

我于1962年毕业于兰州大学数学力学系,对于遗传学领域是后学,本书如有不足之处,恳请读者批评指正,以便日后修改。

感谢中国科学院出版基金对本书的支持。

袁志发

2014年6月于西北农林科技大学

目 录

前言

第 1 章 数量性状遗传研究基础	1
1.1 遗传学基础及其发展	2
1.1.1 质量性状遗传与经典遗传学的发展	2
1.1.2 数量性状遗传与数量遗传学的发展	6
1.1.3 DNA 标记与现代数量遗传学	14
1.1.4 数量遗传学在我国的发展	14
1.2 群体遗传学基础及其发展	15
1.2.1 孟德尔群体遗传结构的数学模型	15
1.2.2 随机交配下理想大孟德尔群体的平衡	17
1.2.3 群体在非随机交配下的遗传效应	19
1.2.4 改变群体基因频率的因素	29
1.3 统计学基础	31
1.3.1 正态分布及其参数估计	32
1.3.2 有限混合正态分布及其参数估计	35
1.3.3 线性统计模型基础	40
第 2 章 近交与随机交配群体的遗传分析	54
2.1 基因型与环境	54
2.1.1 表现型值与基因型值	54
2.1.2 基因型与环境互作的方差分析	55
2.2 基因型值分解、基因的平均效应和育种值	56
2.2.1 基因型值的分解	56
2.2.2 平衡群体中基因的效应	58
2.2.3 平衡群体中个体的育种值和育种效应	60
2.2.4 育种值和基因型值的关系	60
2.3 近交群体数量性状的世代均数分析	62
2.3.1 世代平均数的遗传组成(期望组分)	62
2.3.2 遗传参数估计与遗传模型检验	65
2.4 近交群体数量性状的世代方差分析	75
2.4.1 不分离世代的方差	75
2.4.2 分离世代集团混种(混养)的方差剖分	76
2.4.3 F_1 代连续自交的分支交配系统方差的逐级分解	77
2.4.4 S_3 、 S_4 世代方差的分解	81

2.4.5	显性度	83
2.4.6	自交选择的最适世代及分离极限估计	85
2.5	连锁对世代均数和方差的影响	86
2.5.1	加性-显性遗传模型下的两对与多对基因连锁的 F_2 代遗传分析	86
2.5.2	连锁对 F_1 代连续自交的分支交配系统方差的影响	90
2.5.3	连锁不平衡群体	92
2.6	随机交配群体中亲属间的相关与回归	93
2.6.1	随机交配平衡群体中亲属遗传协方差及其相似性	93
2.6.2	近亲交配群体的近交系数计算	101
2.6.3	亲属协方差的一般形式	105
2.6.4	常见亲属协方差组分中 u 和 r 的计算	106
第 3 章	重复力、遗传力、相关及其估计	109
3.1	重复力、遗传力和相关的概念	109
3.1.1	遗传力(率)和重复力(率)	109
3.1.2	表型相关、遗传相关和环境相关	111
3.1.3	亲属间的表型相似性	113
3.2	重复力、广义遗传力估计及重复力的应用	113
3.2.1	单因素遗传设计试验的组内相关估计方法	114
3.2.2	重复力的准确度及其应用	120
3.3	狭义遗传力的估计	123
3.3.1	用双亲本杂交类型遗传设计估计遗传力	123
3.3.2	用 NCI 遗传设计试验估计狭义遗传力	127
3.3.3	用 NCII 遗传设计试验估计狭义遗传力	141
3.3.4	用 NCIII 遗传设计估计狭义遗传力	145
3.4	多个数量性状的遗传力和相关估计	148
3.4.1	单因素遗传设计完全随机区组试验估计方法	148
3.4.2	NCI 遗传设计完全随机试验的多性状遗传力和相关估计	155
3.5	阈性状及其重复力、遗传力的估计	164
3.5.1	阈性状的重复力估计	165
3.5.2	阈性状的遗传力估计	169
3.6	遗传力、重复力显著性测验原理	175
3.6.1	均方的方差估计	175
3.6.2	均方比的方差估计	175
3.6.3	常用遗传设计试验中组内相关系数的假设检验	176
第 4 章	选择原理、方法和模型	183
4.1	选择的基本原理	183
4.1.1	选择及其预期选择反应	184
4.1.2	选择极限与世代选择效率指数	187
4.1.3	世代间隔、单位时间遗传进展与现实遗传力	190

4.2 单性状选择方法	192
4.2.1 个体选择	192
4.2.2 家系选择与同胞选择	193
4.2.3 家系内选择	194
4.2.4 合并选择	195
4.2.5 几种选择方法相对效率比较	196
4.3 间接选择和综合选择指数	199
4.3.1 间接选择	199
4.3.2 多性状的选择-综合选择指数	201
4.3.3 选择指数的通径分析和决策分析	205
4.3.4 相关遗传进展分解原理与决定系数遗传力	207
4.4 通用选择指数	211
4.4.1 通用选择指数	212
4.4.2 通用选择指数的通径分析和决策分析	218
4.5 组合性状与组合性状对分析	219
4.5.1 约束组合性状分析	219
4.5.2 约束组合性状对分析及其通径分析、决策分析	234
第 5 章 主基因-多基因遗传体系分析	243
5.1 单个分离世代数量性状的分离分析	244
5.1.1 单个分离世代表型分布的主要特征	244
5.1.2 F_2 代数量性状分离分析模型	245
5.1.3 回交世代的主基因-多基因混合遗传分离分析模型	251
5.1.4 单个分离世代数量性状分离分析的一步法和两步法	257
5.2 数量性状的多世代联合分离分析简介	263
5.3 主基因-多基因数量性状的综合选择指数	272
5.3.1 综合选择指数原理	273
5.3.2 应用举例	273
第 6 章 交配效应与配合力分析	278
6.1 交配效应分析	278
6.1.1 近交衰退	278
6.1.2 近交导致的群体遗传方差再分配	283
6.1.3 杂种优势	288
6.1.4 杂交引起的群体遗传方差再分配	293
6.1.5 近交、杂交与环境方差	298
6.1.6 杂种优势的机理	300
6.2 配合力分析	302
6.2.1 配合力概念及其方差	303
6.2.2 配合力估计的试验设计方法	305
6.2.3 双列杂交完全随机区组试验的配合力分析	306

6.2.4 关于多元双列杂交的配合力分析.....	351
第 7 章 数量性状与分子遗传学	358
7.1 数量性状与遗传学.....	358
7.1.1 数量性状与经典遗传学.....	358
7.1.2 数量性状与分子遗传学.....	360
7.1.3 遗传标记与基因定位	362
7.2 QTL 的初级定位及其作图原理.....	365
7.2.1 作图群体的建立	366
7.2.2 图谱构建的理论基础	367
7.3 单标记分析法.....	375
7.3.1 单标计分析样本资料表.....	375
7.3.2 用重组率 r 表示的标记和 QTL 基因型频率表	375
7.3.3 基于标记分组的标记和 QTL 的共分离信息	376
7.3.4 均值差-方差分析法及参数估计	379
7.3.5 最大似然估计的 EM 算法.....	384
7.4 区间作图法和复合区间作图法.....	388
7.4.1 区间作图法	388
7.4.2 复合区间作图法	392
7.5 数量性状的标记辅助选择.....	396
7.5.1 标记值选择	396
7.5.2 指数选择	397
7.5.3 数量性状的分子标记辅助选择.....	399
7.6 KEGG 通路的通径分析及其决策分析	402
7.6.1 DIA 方法的 KEGG 通路	402
7.6.2 KEGG 通路 DIA 影响的通径分析及其决策分析	407
7.6.3 细胞过程路梯度分析、通径分析及其决策分析	408
7.6.4 梯度分析、通径分析及其决策分析与 DIA 方法的比较	418
参考文献	420

第 1 章 数量性状遗传研究基础

遗传学中把生物个体所表现出来的有关形态、生理等方面的特征称为性状(trait). 性状是由基因控制的, 有表现型(phenotype)和基因型(genotype)之分. 表现型是基因型和环境共同作用的结果, 是可以观察到的; 基因型是由双亲配子所携带的基因结合而成的, 现实中是观察不到的.

生物性状主要有两种不同的表现形式, 即质量性状(qualitative trait)和数量性状(quantitative trait). 例如, 花的颜色、牛角的有无、鸡冠的形状皆为质量性状. 数量性状的表现型值(phenotype value)可用一定的物理单位来测量, 如小麦的株高、千粒重、家畜的体尺、泌乳量等.

在动植物品种改良中, 涉及的目标性状有产量及产量因素性状、品质性状、逆境胁迫下的适应性状、抗病抗虫的抗性性状和加工性状等. 改良目标性状中大多数为数量性状, 少数为质量性状. 人们渴望了解和掌握各种性状的遗传变异机制, 尤其是数量性状的遗传变异机制, 以便有效地提高品种改良的效率.

生物性状由孟德尔基因来控制, 这是由试验资料和有关生物学成果抽象出来的一种假说或理论, 只有把基因定位在特定的染色体上才能有掌控它的可能. 对于质量性状的基因定位, 摩尔根已经实现; 人们认为数量性状由多基因控制, 1988 年才第一次实现了数量性状位点(quantitative trait loci, QTL)在染色体上的定位, 使人为控制数量性状基因成为可能.

研究数量性状遗传变异规律的学科称为数量遗传学(quantitative genetics), 也称为生统遗传学(biometrical genetics)或统计遗传学(statistical genetics), 前者突出了研究对象, 后者突出了研究方法, 其实质是一样的. 20 世纪 40~80 年代, 数量遗传学的遗传体系假设为微效多基因, 遗传模型为正态线性模型, 由此展开了数量性状的遗传分析和在动植物育种中的应用. 这个时期的数量遗传学称为经典数量遗传学或传统数量遗传学(classical quantitative genetics). 20 世纪 70 年代后期~2003 年, 数量性状的遗传体系发展为主基因-多基因混合遗传体系(mixed major gene and polygene inheritance), 其遗传模型扩展为有限混合正态分布. 20 世纪 90 年代后, 随着 QTL 定位和 DNA 标记辅助选择等研究的深入, 形成了现代数量遗传学(modern quantitative genetics), 又称为分子数量遗传学(molecular quantitative genetics). 数量遗传学是遗传学各分支服务于动植物育种的一门理论学科, 而且任一遗传分支要服务于育种都要通过它, 因而没有“传统”与“现代”之分, 仅有传承和深化的发展关系, 把二者统称为数量遗传学更符合学科发展规律. 数量遗传学的研究基础是经典遗传学、分子遗传学、群体遗传学和统计学, 本章予以简述.

1.1 遗传学基础及其发展

1.1.1 质量性状遗传与经典遗传学的发展

创立质量性状遗传基因理论的是孟德尔(Mendel, 1822~1884), 完善它的是摩尔根(Morgan, 1866~1945)学派. 孟德尔和摩尔根的遗传学称为经典遗传学.

孟德尔是当时奥地利古老的布龙(Brünn)城的修道士, 这个地方现在是捷克斯洛伐克的布尔诺(Bron)城, 孟德尔以豌豆为材料进行了 8 年(1856~1864)杂交试验. 他把植株所表现的质量性状区分为各个单位(单位性状)作为研究对象, 如花色、种子性状、子叶颜色等, 并把同一单位性状的相对差异称为相对性状. 相对性状杂交中, F_1 代能表现出来的性状称为显性性状, 否则称为隐性性状. 然后用 F_1 代自交观察各单位性状的表现及比例, 由此提出用以解释试验结果的遗传理论. 孟德尔对 7 对相对性状的世代材料进行了详细记录(表 1.1.1), F_2 代显性性状与隐性性状的比例约为 3 : 1.

表 1.1.1 豌豆 7 对相对性状杂交试验结果

性状	杂交组合		F_1 代的表现	F_2 代的表现(株数)		
				显性性状	隐性性状	比例
花色	红花	白花	红花	705(红花)	224(白色)	3.15 : 1
种子性状	圆粒	皱粒	圆粒	5474(圆粒)	1850(皱粒)	2.96 : 1
子叶颜色	黄色	绿色	黄色	6022(黄色)	2001(绿色)	3.01 : 1
豆荚形状	饱满	不饱满	饱满	822(饱满)	299(不饱满)	2.95 : 1
未熟豆颜色	绿色	黄色	绿色	428(绿色)	152(黄色)	2.82 : 1
花着生位置	腋生	顶生	腋生	651(腋生)	207(顶生)	3.14 : 1
植株高度	高	矮	高	787(高)	277(矮)	2.84 : 1

孟德尔关于豌豆杂交试验的科学假设包括: ①花色的决定因子(今称基因)有两种形式(等位基因) A 和 a , 分别决定红花和白花; ②每个植株有两个这样的基因, 从亲本中各得一个, 这样便可有三种可能的基因型 AA 、 Aa 和 aa ; ③配子(卵细胞和花粉, 或动物的卵子和精子)只含一个基因, 是由亲本植株的两个基因中随机选取的; ④卵细胞和花粉不管带什么基因, 均随机结合而产生下一代植株. 这个理论的实质是形成配子时等位基因的分离, 然后异性配子随机结合而形成下一代的基因型. 这个理论归纳为孟德尔遗传第一定律, 即分离定律; 一对基因在杂合情况下并不互相影响和沾染, 在形成配子时按原样分离到不同的配子中. 这个定律说明, 在世代传递中, 连续的是基因而不是基因型. 孟德尔遗传理论对花色试验的解释如图 1.1.1 所示.

据图 1.1.1, 按孟德尔假设, F_2 花色的理论比例应为 3 : 1. 据表 1.1.1 资料, 可对其进行皮尔逊 χ^2 符合性检验, 其无效假设为 H_0 : 符合孟德尔比例 3 : 1.

$$\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(1) \quad (\text{自由度为 } 1, \text{需进行连续性校正})$$

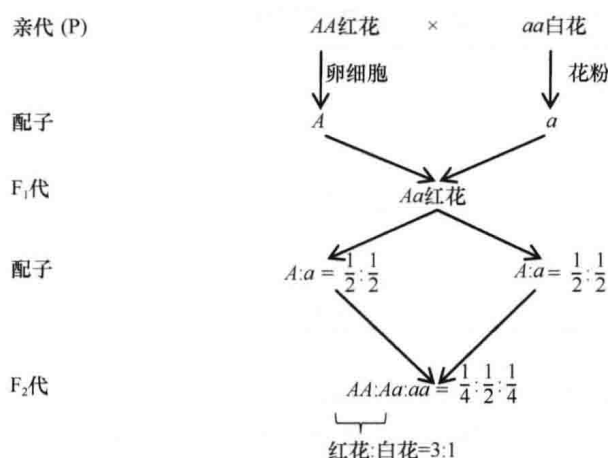


图 1.1.1 孟德尔关于花色试验的解释

其中, O_1 和 O_2 分别为 F_2 代显隐性观察株数($O_1=705$, $O_2=224$); E_1 、 E_2 分别为显隐性理论次数, $E_1 = 0.75n$, $E_2 = 0.25n$, n 为 F_2 代观察总株数(929); χ^2 的自由度 $f = \text{组数} - 1 - \text{独立参数个数}$. 本例组数为 2, 孟德尔比例不用估计, 故 $f = 1$, 需进行连续性校正

$$\chi^2 = \frac{(|705-929 \times 0.75| - 0.5)^2}{929 \times 0.75} + \frac{(|224-929 \times 0.25| - 0.5)^2}{929 \times 0.25} = 0.345 < \chi_{0.05}^2(1) = 3.841$$

表明花色 F_2 代的遗传符合孟德尔遗传定律.

通过交配结果 F_1 和 F_2 如何验证孟德尔基因的分离定律呢? 孟德尔用隐性回交($aa \times F_1$)的测交和用 F_2 自交产生 F_3 的方法来验证. 下面仅叙述测交法: 通过 $aa \times F_1$ 的结果 $\frac{1}{2}Aa + \frac{1}{2}aa$ 可预期红花:白花=1:1. 据回交资料用 χ^2 检验可验证分离定律的正确性.

孟德尔关于一对基因的分离定律, 实质上描述了杂交中由亲本基因型分离形成两性配子多项式(以配子频率为系数的配子式)相乘而形成下一代基因型多项式(以基因型频率为系数的基因型式)的数学遗传过程. 例如, $Aa \times Aa$, 各自的配子多项式均为 $\frac{1}{2}A + \frac{1}{2}a$, 故 F_2 代的基因型多项式为

$$\left(\frac{1}{2}A + \frac{1}{2}a\right) \delta \times \left(\frac{1}{2}A + \frac{1}{2}a\right) \varphi = \left(\frac{1}{2}A + \frac{1}{2}a\right)^2 = \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$$

即 Aa 的配子分离为 $A:a = 1:1$, 配子多项式为 $\frac{1}{2}A + \frac{1}{2}a$. F_2 代 $AA:Aa:aa = 1:2:1$, 当 A 对 a 为完全显性时, 表型($AA + Aa$): $aa = 3:1$.

孟德尔对两种不同的相对性状植株杂交时所发生的情况也进行了研究. 他在确定了圆粒对皱粒是显性且黄色种子对绿色种子是显性后, 让圆形黄色种子与皱粒绿色的纯合品种杂交, 所得 F_1 代种子全是圆粒黄色. 然后让 F_1 代植株自交得到 F_2 代种子, 并让 F_1 代种子与皱粒绿色种子回交. 这两种试验结果列于表 1.1.2 中.

在表 1.1.2 中, F_2 代的比例为 3:1, 回交试验中为 1:1, 和预期的一样, 而且两个性状彼此独立, 并无互相变化和沾染的迹象. 由此可以断定, 决定这两个性状的基因在配子形成时是独立分离的. 把决定圆粒和皱粒种子的基因分别写成 A 和 a , 决定黄色和绿色种子的基因写成 B 和 b , 则圆粒黄色种子的基因型为 $AABB$, 皱粒绿色种子为 $aabb$.

表 1.1.2 孟德尔关于豌豆种子形状和颜色共同分离的数据

		黄色	绿色	共计
F ₂ 代	圆粒	315	108	423
	皱粒	101	32	133
	共计	416	140	556
回交	圆粒	55	51	106
	皱粒	49	52	101
	共计	104	103	207

$AABB \times aabb$ 产生F₁代, 基因型为 $AaBb$ 。 $AaBb$ 自交时将产生4种配子: AB 、 Ab 、 aB 、 ab 。这样, 表 1.1.2 的结果可作如下解释: 假定配子得到决定种子形状的两个基因中的一个和决定种子颜色的两个基因中的一个随机和彼此独立的, 则 AB 、 Ab 、 aB 、 $ab = 1:1:1:1$ 。因此, 表 1.1.2 中F₂代遗传过程可解释为图 1.1.2, 即圆粒: 皱粒和黄色: 绿色均为 3:1, 而两对性状的表型比例为 9:3:3:1。

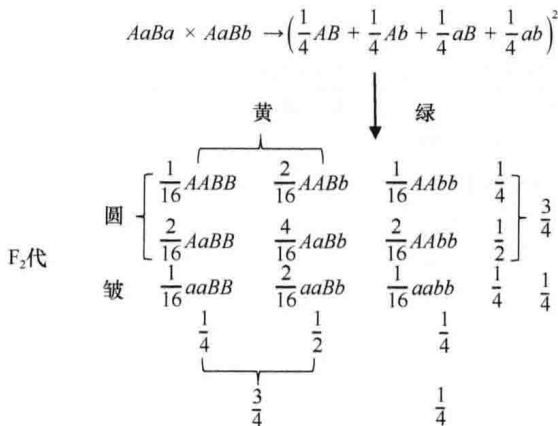


图 1.1.2 孟德尔关于两对相对性状的解释

关于回交的解释如图 1.1.3 所示, 圆粒: 皱粒和黄色: 绿色均为 1:1。

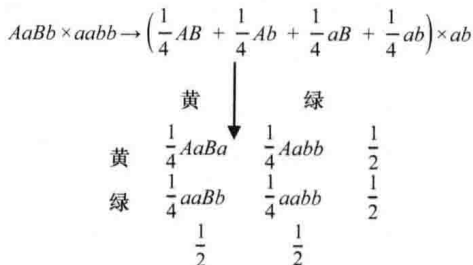


图 1.1.3 孟德尔关于表 1.1.2 中回交的解释

上述理论归纳为孟德尔第二遗传定律, 又称为多对基因的独立分配或自由结合定律: 当两对或多对基因处于杂合状态时, 它们在配子中的分离是彼此独立的。

孟德尔将其研究写成《植物的杂交试验》论文, 于 1865 年 2 月 8 日在 Brunn 自然

科学学会上宣读,并于1866年刊登在Brünn博物学会会刊上。他的论文默默无闻地被尘封了34年,直到1900年,荷兰的德弗里斯(De Vries, 1848~1935)、奥地利的切尔迈克(Tschermak, 1871~1962)和法国的柯伦斯(Correns, 1864~1933)在各自的研究中重新发现了孟德尔理论。从此,在孟德尔理论的基础上,经典遗传学得到了迅速发展。

在孟德尔开创的经典遗传学中,摩尔根是成就最大的人,是经典遗传学的旗帜,摩尔根学派是经典遗传学的主流学派。下面叙述摩尔根的主要贡献。

基因原是用以说明育种试验结果的假设结构。人们自然会以为基因是一种物质,这种物质应存在于细胞中的某个位置。孟德尔理论于1900年被重新发现前,人们已经发现每种二倍体生物每个细胞都有相同数目的染色体,并且成对存在。例如,不分种族、性别的人,每个细胞的染色体都是23对(46条);凡是水稻,不管粳、粳、糯,每个细胞都有12对(24条)染色体;凡是玉米,每个细胞都有10对(20条)染色体;凡是豌豆,每个细胞都有7对(14条)染色体;凡是猪,每个细胞都有19对(38条)染色体……性别是真核生物的共同特征,对于真核生物,细胞里的 n 对染色体中有 $n-1$ 对常染色体,每对内两条染色体完全一样,称为同源染色体,不属于同一对的染色体称为非同源染色体;剩下的一对为性染色体,两条性染色体虽然同源,但大小、形态和特性未必一样。如男性的两条性染色体,一条为X染色体,另一条为Y染色体;而女性的两条性染色体均为X染色体。多细胞生物的生长主要是通过细胞数目的增加和细胞体积的增大而实现的。在这个过程中细胞分裂称为有丝分裂或体细胞分裂。这种分裂是把细胞核内每条染色体准确复制而形成的两个子细胞。减数分裂又称为成熟分裂,是在性细胞成熟时发生的一种特殊的有丝分裂,其结果是形成配子。在减数分裂中,每个母细胞分裂成4个子细胞,但染色体只复制一次,因此每个子细胞(配子)中只含各同源染色体中的一条,这样的子细胞称为单倍体。如男性产生的精子有两种,一种是22条常染色体加X染色体,另一种是22条常染色体加Y染色体,两种精子数目相等;女性只产生一种卵子,就是22条常染色体加X染色体。如果卵子和精子有机会结合为一个合子(受精卵)时,它就具有22对常染色体和一对性染色体(共46条),然后经过多次有丝分裂发育成一个胎儿。如果合子中的精子是22条常染色体加X染色体,则胎儿为女性;若合子中的精子是22条常染色体加Y染色体,则胎儿为男性。

1902年,美国学者Sutton(1877~1916)等发现孟德尔遗传定律与性细胞的减数分裂和受精形成合子的行为是一致的,并提出了细胞的染色体可能是基因的载体学说(摩尔根,《基因论》,2007)。

这个学说引起了广泛重视,并引发了一系列猜想,基因就是染色体?一条染色体就是一个基因?这是不可能的,因为生物的性状很多,基因很多,而染色体是有限的。进一步的思考是,一条染色体上载有许多基因,一个特定的基因位于特定的染色体上。这个问题的解决和遗传学家摩尔根的研究联系在一起。

1910年,摩尔根在哥伦比亚大学的试验发现一种奇特的果蝇,它眼睛的颜色不是野生型的红色而是白色,经过缜密思考和试验,他发现决定眼色的基因与决定性别的基因是连锁遗传的,于是第一次把一个特定的果蝇白眼基因和一条特定的染色体(雄性性染色体)联系起来,并于1912年提出了基因的连锁与交换定律(经典遗传学第三定律),完善了经典遗传学的理论体系,导致细胞遗传学诞生。同源染色体上的两个基因的重组率

r 的相对恒定, 说明同源染色体上各基因的位置相对稳定. 摩尔根于 1911 年指出, 重组率的大小反映了基因位点在染色体上距离的远近. 根据摩尔根的设想, 由重组率确定同一染色体上不同基因的相对位置和排列顺序的过程称为基因定位(gene mapping). 由此绘出的线性示意图(chromosome map)称为基因连锁图(linkage map)或遗传图(genetic map). 两个基因在染色体上的距离称为图距(map distance). 若重组率 $c = m\%$, 则称它们之间的图距为 $m \text{ cM}$ (厘摩), 即将重组率“1%”定义为图距单位(map unit). 将图距单位称为“厘摩”(centimorgan, cM), 这是为了纪念摩尔根. 1913 年, Sturtevant 通过连锁分析, 成功地在果蝇的 X 染色体上定位了 5 个基因, 确定了遗传学的染色体理论和遗传作图的基本原理. 果蝇有 4 对染色体, 摩尔根小组花费了巨大的精力, 测定了果蝇许多基因之间的连锁强度, 形成了 4 个连锁群, 并在他于 1926 年出版的巨著《基因论》第一章的“基因的直线排列”一节中绘出了 4 个连锁群的遗传图.

《基因论》是经典遗传学最权威的著作. 在《基因论》中, 摩尔根证实了孟德尔的遗传学理论, 即生物的性状是由基因决定的; 基因是长期稳定的、颗粒性的, 可以区分为一个单位; 证明了基因存在于染色体上; 发现了基因的连锁与交换定律, 位于同一对染色体上的许多基因存在不同程度的连锁, 这些基因呈线性排列, 其相对位置由相邻基因间的重组率($0 \leq r < 0.5$)决定, 形成了一个连锁群(linkage group). 连锁群的数目等于单倍体染色体数目(n); 重组发生的原因是同源染色体间发生了局部交换的结果. 《基因论》给出了关于生物性状遗传的一幅清晰而直观的经典基因理论的物理模型.

摩尔根的理论使孟德尔的两个遗传定律的机理明朗起来, 这就是他所研究的豌豆 7 对相对性状分别位于豌豆仅有的 7 对染色体上, 不连锁, 故配子会独立分离和重组($r = 0.5$), 其实质是独立遗传而不是连锁遗传. 《基因论》虽然完善了经典遗传学的理论体系, 但也提出了遗传学中未曾解决的一些核心问题. 例如, 未曾回答基因是什么物质, 基因的结构是什么样的等. 但摩尔根在《基因论》的最后一章指出: “我们仍然很难放弃这个可爱的假设, 就是基因之所以稳定, 是因为它代表着一个有机的化学实体”, 并认为“至少它不失为一个良好的试用假说”.

综上所述, 质量性状遗传研究的数学模型运用与它在杂交后代的表现特点有关.

质量性状在杂交后代中的表现包括: ①对环境不敏感, 个体间有明显的归属类型. F_1 代仅有一种类型的个体. 分离世代有多种类型个体, 呈离散变异, 可用计数方法估计各种类型的比例. 如表 1.1.1 所示 F_2 代的 7 对相对性状均为两种类型个体, 其比例均约 3 : 1; ②在遗传上可假定它受少数主基因控制, 遵从孟德尔的分离和自由组合遗传定律及摩尔根的连锁交换遗传定律; ③从遗传的数学模型上讲, 可用离散型随机变量描述基因分布和基因型分布; ④在遗传学研究中, 遗传标记主要应用于连锁分析、基因定位、遗传作图及基因转移等. 遗传标记可以明确标示遗传多态性的特征. 在经典遗传学中, 遗传多态性是指等位基因的变异. 在摩尔根的研究中, 果蝇白眼遗传标记的发现导致了连锁交换定律的产生, 并实现了质量性状的基因定位、遗传作图等.

1.1.2 数量性状遗传与数量遗传学的发展

孟德尔关于质量性状遗传学的论文《植物杂交试验》虽然于 1866 年发表, 但并未

被人们所认知. 1859 年达尔文(Darwin, 1809~1882)的生物进化论巨著《物种起源》出版后, 引起了科学界的巨大反响. 高尔登(Galton, 1822~1911)为了解决生物进化中的数量分析问题, 研究了人群中双亲与其子女身高的关系等, 并分别于 1886 年、1889 年用相关和回归表达这种关系, 把统计方法引入数量性状研究. 由于没有科学的遗传机理, 无法阐明相关和回归中的基因内涵, 导致生物统计学的产生和发展.

1900 年, 孟德尔遗传定律(遗传因子假说)被重新发现, 其遗传因子和统计学相结合的研究方法给人们以启示, 促进了数量性状遗传的研究. 1909 年, 约翰森(Johannsen)据其菜豆遗传实验提出了纯系假说, 并提出了基因型和表现型的概念, 认为表现型是基因型与环境相互作用的结果, 把数量性状的表现型变异区分为可遗传的变异和不可遗传的变异, 阐明了表现型、基因型和环境效应三者间的关系, 为数量性状遗传所表现出的连续性变异研究和建立数学模型提供了依据.

1. 微效多基因假说

1909 年, 尼尔森-埃尔(Nilsson-Ehle)根据小麦粒色的遗传研究提出了数量性状遗传的多基因假说(multiple-factor hypothesis), 认为数量性状的遗传效应是许多彼此独立的遗传因子共同作用并累加的结果, 而每个遗传因子的效应甚微且相等, 但其遗传机理仍为孟德尔式遗传, 后人称其为微效多基因假说或多基因假说(polygene hypothesis), 用数量性状的微效基因(minor gene)区分质量性状中的主基因(major gene). 这一假说于 1913 年被 Emerson 和 East 的玉米穗长研究所证实, 又被 East 于 1916 年的烟草花冠长度遗传试验所证实. Nilsson-Ehle 等的遗传试验均为两纯系杂交试验, 其世代资料的表现均相似. 下面以 East 的玉米穗长资料为例, 说明微效多基因假说的内容及其分析方法. East 的资料如表 1.1.3 和图 1.1.4 所示. 其中 P_2 为短穗亲本, P_1 为长穗亲本.

由表 1.1.3 和图 1.1.4 可以看出, 数量性状在亲代和其杂交后代群体中的表现和质量性状是不同的, 因而其遗传研究方法应不同于质量性状遗传研究的经典遗传学方法. 这些不同之处包括: ①对环境敏感, 在不分离群体(双亲和 F_1)和分离群体(F_2)中, 个体间无明显的归属类型, 均呈连续变异, 因而无法用离散型随机变量(如二项式分布)来描述其遗传规律, 只能用连续型随机变量来研究它的遗传; ②从统计分析的角度看, 各群体的均值和方差有以下表现, $\bar{F}_1$ 和 $\bar{F}_2$ 均接近双亲均值的平均, 即接近 $\frac{1}{2}(\bar{P}_1 + \bar{P}_2)$; 不分离群体的方差 $S_{P_1}^2$ 、 $S_{P_2}^2$ 和 $S_{F_1}^2$ 基本上同质; 分离群体的方差 $S_{F_2}^2$ 大于不分离群体的方差.

表 1.1.3 玉米穗长的遗传 (单位: cm)

世代	长度																				n	$\bar{x}$	s	s^2
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21							
P ₂	4	21	24	8																57	6.632	0.816	0.656	
P ₁									3	11	12	15	26	15	10	7	2	101	16.802	1.887	3.561			
F ₁					1	12	12	14	17	9	4							69	12.116	1.519	2.307			
F ₂			1	10	19	26	47	73	68	68	39	25	15	9	1			401	12.888	2.252	5.072			

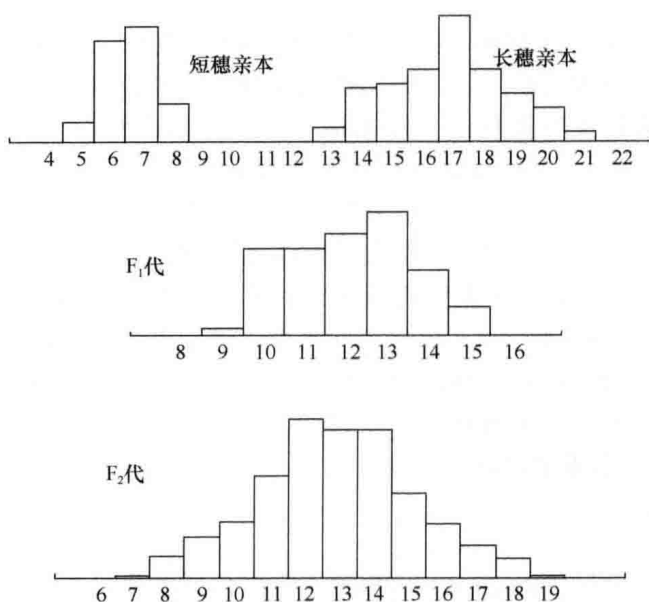


图 1.1.4 玉米穗长遗传的柱形图

Nilsson-Ehle 在小麦子粒颜色的试验中发现了与表 1.1.3 数据相同的特点: 红色子粒与白色子粒杂交中, F_1 代子粒颜色为红与白的中间类型, 不能区别显隐性; F_2 代子粒颜色表现出由红到白的不同程度的类型. 据此他提出数量性状由许多彼此独立的基因作用的结果, 每个基因对性状表现的效果较微, 但其遗传方式仍然服从孟德尔遗传规律, 而且假定: ①各基因的效应相等; ②各等位基因的表现为不完全显性 or 无显性, 并表现为增效和减效作用; ③各基因的作用是累加性的. 这就是数量性状的微效多基因假说或多基因假说.

下面介绍 Nilsson-Ehle 据其假说对小麦子粒颜色试验进行的遗传分析. 假定由 m 对基因 (R_i, r_i) 决定, 红粒亲本 P_1 的基因型为 $R_1R_1R_2R_2 \cdots R_mR_m$, 白粒亲本 P_2 的基因型为 $r_1r_1r_2r_2 \cdots r_mr_m$. F_1 代子粒表现为红与白的中间色, 基因型为 $R_1r_1R_2r_2 \cdots R_mr_m$. F_2 代中有 4^m 个基因型. 假定所有的 (R_i, r_i) 均等效为 (R, r) , 即基因型中多一个 R 就会使子粒更红一些, 这样 F_2 代的基因型就会按 R 的个数 $k=0, 1, 2, \cdots, 2m$ 分为由白色开始的红色程度不同 $(2m+1)$ 种类型, 当 $k=2m$ 时最红. F_1 的基因型为 $RrRr \cdots Rr$, 配子式仍为 $(\frac{1}{2}R + \frac{1}{2}r)$, F_2 代基因型为 $(\frac{1}{2}R + \frac{1}{2}r)^{2m}$. F_2 代中, 基因型中有 k 个 R 的概率为

$$P(k) = C_{2m}^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{2m-k} = \frac{1}{4^m} C_{2m}^k \quad k = 0, 1, 2, \cdots, 2m \quad (1.1.1)$$

即 k 服从二项分布 $(\frac{1}{2} + \frac{1}{2})^{2m}$, 其均值和方差分别为

$$\begin{cases} \mu_k = \sum_{k=0}^{2m} \frac{1}{4^m} C_{2m}^k = m \\ \sigma_k^2 = \sum_{k=0}^{2m} \frac{(k-m)^2}{4^m} C_{2m}^k = \frac{m}{2} \end{cases} \quad (1.1.2)$$

在上述假定下, P_1 、 P_2 和 F_1 的基因型值(R 和 r 的累加值)分别为各世代的均值

$$\bar{P}_1 = 2mR \quad \bar{P}_2 = 2mr \quad \bar{F}_1 = m(R+r) \quad (1.1.3)$$

$$d = \frac{1}{2m}(\bar{P}_1 - \bar{P}_2) = R - r \quad (1.1.4)$$

则有 k 个 R 的基因型值为 $2mr + kd$, F_2 代基因型值的平均值和遗传方差分别为

$$\begin{cases} \mu_{F_2} = \sum_{k=0}^{2m} \frac{(2mr + kd)}{4^m} C_{2m}^k = 2mr + \mu_k d = m(R+r) \\ \sigma_{F_2g}^2 = \sum_{k=0}^{2m} \frac{(2mr + kd - 2mr - \mu_k d)^2}{4^m} C_{2m}^k = \frac{m}{2} d^2 = \frac{m}{2} (R-r)^2 \end{cases} \quad (1.1.5)$$

式(1.1.5)表明, 在微效多基因假说之下 F_2 的基因型值服从以中亲值 $\frac{1}{2}(\bar{P}_1 + \bar{P}_2) = m(R+r)$ 为中心的对称的二项分布. 这个结果较好地解释了表1.1.3中 F_2 代的表现: $\bar{F}_1$ 和 $\bar{F}_2$ 约等于 $\frac{1}{2}(\bar{P}_1 + \bar{P}_2)$; 由于 P_1 、 P_2 和 F_1 均为只含一个基因型的群体, 故它们的方差均由环境影响引起, 为环境方差 σ_e^2 , 而 F_2 代中除了环境方差 σ_e^2 之外还有遗传方差 $\sigma_{F_2g}^2 = \frac{m}{2}(R-r)^2$, 故 F_2 代的表型资料方差大于各不分离世代的方差.

多基因假说产生于两个纯系的杂交试验分析. 在杂交后代群体中, 个体的数量性状表现型测定值 P 称为表现型值. P 中由基因型决定的值 G (不能直接观测)称为基因型值. 根据约翰森纯系学说, 表现型是基因型和环境共同作用的结果. 即 P 中除 G 外还有环境的直接作用 E 和基因型和环境互作效应 $G \times E$. 将 P 、 G 和 E 作为描述杂交各世代群体中数量性状的总体连续型变量, 就得到约翰森意义下的数量性状遗传的数学模型

$$P = G + E + G \times E \quad (1.1.6)$$

$G \times E$ 对育种很重要, 应另行分析. 为了分析简便, 假定 $G \times E = 0$, 并假定 P 和 G 的平均值(期望值)均等于 μ , E 的期望值为零, 且 G 和 E 独立, 还可令 $G = \mu + g$, g 的期望值为零, 则式(1.1.6)可写成

$$P = G + E \text{ 或 } P = \mu + g + e \quad (1.1.7)$$

其中, P 的方差 σ_P^2 称为表型方差, G 或基因型效应 g 的方差 σ_g^2 称为遗传方差, 环境效应 E 或 e 的方差 σ_e^2 称为环境方差, 则有

$$\sigma_P^2 = \sigma_g^2 + \sigma_e^2 \quad (1.1.8)$$

正态分布 $N(\mu, \sigma^2)$ 产生于测量误差的研究. 测量对象的真值和实测值分别为 μ 和 x , ε 是测量的随机误差, 即 $x = \mu + \varepsilon$. ε 时大时小, 时正时负. 偏离 μ 大时出现的机会小, 偏离 μ 小时出现的机会大, 正误差与负误差出现的机会是相等的, 这就形成了中间高两边低左右对称的正态曲线. 多基因假说下的 F_2 代小麦子粒颜色基因型是以 $m(R+r)$ 为中心的对称的二项分布, 因而可认为在环境作用下按中心极限定理 F_2 代的表现型及基因型值服从正态分布. 这样在式(1.1.7)、式(1.1.8)中, 一般假定

$$P \sim N(\mu, \sigma_P^2) \quad G \sim N(\mu, \sigma_g^2) \quad g \sim N(0, \sigma_g^2) \quad E \text{ 或 } e \sim N(0, \sigma_e^2) \quad (1.1.9)$$

这就形成了以正态分布为中心的数量性状多基因模型式(1.1.7)~式(1.1.9)的遗传统计分析方法. 事实上, 模型表明: 它是数量性状多基因整体的群体遗传学, μ 、 σ_g^2 和 σ_e^2 的相对重要性为其群体遗传特征.