

前 言

植物群体遗传学是研究植物群体中等位基因分布、分布的维持和变化的科学,由此探讨植物进化的机制并为植物育种和植物资源保护与管理提供理论基础。群体遗传学起源于1908年George Hardy和Wilhelm Weinberg提出的遗传平衡法则,至今已有100多年的历史。一个世纪以来,在Wright、Fisher、Haldane和Kimura等著名群体遗传学家的努力下,群体遗传学作为遗传学的一个重要分支学科,已广泛应用于生物学各领域,显示出强大的生命力。

我国群体遗传学研究起步较晚,但发展很快。近年来,虽然国内科学家在这一领域某些方面的研究已处于国际领先地位,但是,中文版的群体遗传学仅有几本入门书籍,这与我国群体遗传学教学与研究的需求不适应。

本书是以近十年为中南林业科技大学硕士研究生讲授的《群体遗传学》讲义为蓝本,结合作者科学研究及在国外研修期间的学习心得,经过修改、补充而形成的。

全书共分11章。第一章和第二章是学习、理解和从事植物群体遗传学研究必备的理论基础;第三章至第八章,按照由理想群体逐步到现实植物群体的思路,由浅入深、循序渐进、较全面地阐述交配系统、选择、遗传漂移、突变及迁移对植物群体遗传结构的影响,重点强调如何利用遗传数学模型研究复杂的植物群体遗传中的问题,探讨植物进化机制。第九章,综合前面几章的内容,介绍现实植物群体遗传多态性的产生、维持机制及遗传多态性度量。第十章,介绍植物数量性状遗传,为理解复杂的植物适应性表型进化提供初步的数量遗传分析工具。第十一章,介绍群体遗传学知识在生物分子进化与物种形成中的应用,突出目前生物学中富有挑战性的课题——系统发育与性进化。

书中有关名词术语以《英汉遗传学词典》为准,《英汉遗传学词典》上没有收录的术语,以CNKI(http://dict.cnki.net/dict_result.aspx)翻译频率最高的译法为准。

随着现代分子生物学和信息科学研究进展,研究生物群体遗传信息流的群体遗传学目前正处在快速发展期间,新理论假说、新研究方法、新发现层出不穷。本书不刻意追求“深奥”,而注重实用性,重视群体遗传学分析中基本概念、基本原理与基本方法的准确把握。力求语言通俗、精炼,前后连贯,便于自学。用较简单的数学推导,系统地介绍植物群体遗传学领域中的理论与方法。对于需要用超越生物学本科生所掌握的数学知识来推导的较难的遗传数学模型,仅介绍结论,数学基础较好的读者,可参阅各章后的推荐读物,了解其推导过程。

感谢加拿大Alberta大学统计基因组与数量遗传学教授杨荣才博士无私地提供了群体遗传学、进化遗传学、生态遗传学、基因组学及数量遗传学等国外优秀教材和专著。南京林业大学施季森教授,华东师范大学陈小勇教授,北京林业大学续九如教授、李悦教授审阅了本书初稿,中南林业科技大学何钢、曾柏全、刘勉声、王爱云等仔细阅读了本书初稿第一章和第二章中各自所熟悉的内容并提出了很好的建议,博士生刘志祥,硕士生蒋淑

磊、占志勇、刘莹等参与了书中部分图片处理,在此深表谢意。

本书的出版得到了湖南省自然科学基金、中南林业科技大学教材建设基金与林木遗传育种学科建设基金资助。

由于作者水平有限,未必能把本学科领域的精髓及最新研究成果全面地反映出来,加之时间仓促,书中难免出现缺点和不足,敬请读者批评指正。若有任何宝贵意见或建议,请随时发邮件至 gangbiaoxu@csuft.edu.cn 或 gangbiaoxu@163.com。

徐刚标

2008年10月于长沙

目 录

前言		
绪论	1	
一、群体遗传学概念	1	
二、群体遗传学发展简史	2	
三、植物群体遗传学研究的内容	5	
四、植物群体遗传学研究模型	6	
第一章 分子生物学与生物统计学基础	9	
第一节 基因及其有关概念	9	
一、DNA 与染色体	9	
二、基因与基因表达	11	
三、遗传密码	14	
第二节 生物统计基础	15	
一、概率基本概念	15	
二、随机变量分布	16	
三、随机变量的数字特征	18	
四、抽样分布	19	
五、显著性检验及参数估计	21	
第二章 植物群体遗传变异	24	
第一节 植物群体变异来源	24	
一、环境条件	24	
二、基因重组	25	
三、突变	26	
四、表观遗传修饰	27	
第二节 遗传变异基本规律	28	
一、遗传学基本术语	28	
二、遗传变异基本规律	29	
第三节 植物群体遗传变异检测	32	
一、基于表型的遗传标记检测技术	33	
二、分子标记检测技术	36	
第三章 理想群体与 Hardy-Weinberg 法则	46	
第一节 Hardy-Weinberg 法则	46	
一、理想群体	46	
二、等位基因频率与基因型频率	46	
三、Hardy-Weinberg 法则	48	
四、理想二倍体群体平衡的性质	50	
五、Hardy-Weinberg 法则应用	51	
第二节 Hardy-Weinberg 法则推广	52	
一、复等位基因	52	
二、性连锁基因	53	
三、多基因座	56	
第三节 同源多倍体	60	
一、同源四倍体的等位基因分离	60	
二、理想同源四倍体植物群体平衡	62	
第四章 植物非随机交配系统	66	
第一节 植物交配系统	66	
一、植物繁殖与性别系统	66	
二、避免自交机制	67	
三、自交率	68	
第二节 近亲交配系统	69	
一、近交程度的度量	70	
二、自交	73	
三、近交	75	
四、混合交配	77	
第三节 遗传异型与同型交配系统	78	
一、配子体自交不亲和	78	
二、孢子体自交不亲和	79	

三、显性同型交配受阻·····	80	六、世代重叠·····	127
第四节 表型同型与表型异型交配		七、群体地理散布·····	127
系统·····	82	第三节 有限群体中选择·····	128
一、表型同型交配·····	82	第七章 基因突变 ·····	132
二、表型异型交配·····	83	第一节 基因突变基础·····	132
第五章 自然选择 ·····	87	一、核苷酸替换·····	132
第一节 选择的遗传学基础·····	87	二、缺失、插入和倒位·····	133
一、选择类型·····	88	三、转座·····	136
二、适应度和选择系数·····	89	第二节 中性突变·····	137
三、选择效应一般遗传模型·····	90	一、不可逆突变·····	137
第二节 定向性选择·····	92	二、可逆突变·····	139
一、隐性纯合体净化选择·····	92	三、有限群体中可逆突变·····	141
二、显性个体净化选择·····	94	第三节 非中性突变·····	142
三、显性不完全选择·····	96	一、完全隐性有害突变·····	142
四、加性基因选择·····	97	二、不完全隐性有害突变·····	143
第三节 平衡选择·····	98	三、加性基因突变·····	144
一、超显性选择·····	98	四、显性有害突变·····	144
二、依频选择·····	100	五、有限群体中突变与选择联合作	
三、减数分裂驱动·····	102	用·····	145
四、依密选择·····	103	第四节 中性突变模型·····	146
五、环境异质性选择·····	104	一、无限等位基因模型·····	147
第四节 其他类型选择·····	106	二、Ewens 抽样方程·····	149
一、杂合体净化选择·····	106	三、无限位点模型·····	152
二、性别差异选择·····	107	四、逐步突变模型·····	154
三、配子选择·····	108	第五节 溯祖理论·····	155
第六章 遗传漂移 ·····	109	第八章 植物群体阶层结构与迁移	
第一节 遗传漂移的成因及特征		·····	160
·····	109	第一节 植物群体分化与阶层结构	
一、遗传漂移的成因·····	109	·····	160
二、遗传漂移的特征·····	112	一、Wahlund 效应·····	160
第二节 有效群体大小·····	121	二、群体遗传分化的度量·····	163
一、雌雄数目不等·····	122	三、遗传距离·····	168
二、群体大小波动·····	123	第二节 植物群体间迁移·····	171
三、家系大小变异·····	124	一、陆岛模型·····	171
四、排除近亲交配·····	126	二、岛屿模型·····	172
五、混合交配群体·····	126	三、距离隔离模型·····	174

第三节 迁移与选择、突变联合作用 176	第十章 植物数量性状遗传 207
一、迁移与选择 176	第一节 数量性状的遗传基础..... 207
二、岛屿模型有限群体的迁移 ... 179	一、数量性状的遗传机理 207
三、岛屿模型有限群体迁移与突变 181	二、数量性状表型值及其分解 ... 208
第四节 植物基因流..... 182	三、数量性状表型方差及其分解 210
一、基因流直接测量 183	四、亲属间遗传相似 212
二、基因流间接测量 187	第二节 数量性状进化潜力与遗传 力 215
第九章 植物群体遗传多态性 189	一、遗传力概念 215
第一节 遗传多态性产生及维持机 制..... 189	二、遗传力估算 216
一、超显性假说 189	第三节 选择应答 218
二、依频选择 190	第四节 数量性状的基因定位 222
三、时空异质性 191	一、QTL 定位原理和步骤 222
四、中性突变 191	二、QTL 定位基本方法 223
五、减数分裂分离比例偏移 192	第十一章 分子进化与物种形成 227
六、联合超显性 192	第一节 分子进化..... 227
第二节 植物遗传多态性度量..... 193	一、分子序列进化演变 227
一、多态基因座百分比与等位基因 数目 193	二、中性理论 232
二、基因多样性 195	第二节 分子系统发育 235
三、核苷酸多态性与核苷酸多样性 197	一、基本概念 235
四、Shannon 信息指数与多态信息 含量 198	二、分子系统发育分析 237
第三节 遗传负荷..... 200	第三节 性的起源与进化 244
一、突变负荷 202	一、性的起源 244
二、分离负荷 203	二、性的进化 245
三、置换负荷 205	第四节 物种的形成 250
	一、物种概念 250
	二、物种形成的方式 251
	主要参考文献 255

绪 论

一、群体遗传学概念

群体遗传学(population genetics)是遗传学的一个重要分支学科,是应用遗传学基本原理,采用数学、生物统计及其他方法,研究生物群体遗传结构及其在世代间变化规律的科学,并由此探讨生物进化的机制。

群体遗传学研究对象不是个体,而是孟德尔群体(Mendelian population)。孟德尔群体,简称为群体(population),不是一些生物个体的简单集合体,而是指特定时间和空间,能够自由交配、繁殖的同种生物个体群,通过个体间相互交配,孟德尔遗传因子以不同方式从一代群体传递给下一代群体。因此,群体中生物个体必须是同一个物种的个体,只有这样,才能通过相互交配而交流基因,组成共同的基因库(gene pool)。

最大的孟德尔群体是一个物种,不同物种分属于不同的孟德尔群体。例如,同一林地上红松林群体就是孟德尔群体,而一片禾本科植物植被就不是孟德尔群体,因为它包括了不同种植物,属于不同的基因库,并不是单一的交配、繁育单位。

交配是以有性繁殖为前提的,因此,群体遗传学一般以二倍体或多倍体并能进行有性繁殖的生物为研究对象,单倍体或只能进行无性繁殖的生物群体,通常不属于孟德尔群体。

自然生物大多数生活在复杂多样的环境中,无论在时间上还是在空间距离上都是异质的,存在着影响生物迁移、接触和分布的地理、地形以及其他物理环境因素隔离,或者,生活在同一地区的同种生物个体因生殖季节不同而产生季节隔离,因此,大多数情况下,一个物种是由许多地方群体(local population)组成的,不同地方群体的遗传结构组成可能存在着差异。

进化遗传学、生态遗传学、保育遗传学、数量遗传学也十分重视群体的遗传特征研究,它们与群体遗传学彼此之间交叉重复的内容相当多,但它们在研究的侧重点及运用的领域范围有所区别。

进化遗传学(evolutionary genetics),是研究物种进化的遗传变异规律,研究的范围更广,群体遗传学是进化遗传学的一个组成部分。

生态遗传学(ecological genetics),强调生物群体对生存环境的适应以及对环境改变作出反应的遗传机理,也就是适应性进化。它不仅研究自然条件下生物发生遗传变异的长期效应(进化),而且研究人工条件下发生遗传变异的短期效应(育种)。

保育遗传学(conservation genetics),是研究影响物种灭绝的遗传因素以及濒危物种的遗传管理,以降低物种灭绝的危险,特别关注小群体的遗传后果。

数量遗传学(quantitative genetics),又称统计遗传学(statistical genetics),侧重研究

生物经济性状遗传变异规律,这些经济性状多属于数量性状。它建立在微效多基因假说基础上,以群体遗传学为理论,讨论数量性状的世代平均数和基因型方差等参数估计方法、数量性状基因效应及其在染色体上的位置,而群体遗传学主要讨论群体中等位基因频率在世代间变化规律以及群体平均适应度对自然选择的应答。

植物群体遗传学与一些基础理论学科和应用学科有着密切的联系,有关的知识范围很宽,在理论研究和实际应用过程中,常涉及遗传学、植物学、生态学、分子生物学、细胞生物学、生物统计学等。因此,没有遗传学、分子生物学、生物统计学的基础,没有相关学科的密切结合,植物群体遗传学不可能得到迅速发展。同时,植物群体遗传学又是植物起源及进化、野生植物资源保护与管理、植物育种的理论基础。

二、群体遗传学发展简史

人类对群体遗传现象的认识有着悠久历史,农作物的种植就无意识地应用了群体遗传学知识。但是,群体遗传学的科学研究始于1859年达尔文的《物种起源》。达尔文提出的自然选择理论,揭示了生物的进化规律,从宏观角度用进化理论成功地解释了物种间的相互关系,强调微小变异的累积对物种进化的意义。

1889年, Galton 在发表的 *Natural Inheritance* 论文中,全面地总结了他首创的相关性概念、回归分析方法以及他首次发现的生物群体遗传变异趋势总是下一代比上一代更接近于群体平均值现象。

1900年, Correns 和 de Vries 等重新发现了孟德尔遗传定律,标志着遗传学的诞生,也为群体遗传学建立提供了理论基础。

1903年, de Vries 提出了突变学说,尽管他试图以突变的不连续性来概括生物进化学说的出发点不可取,但却指出突变对生物进化的重要意义。突变作为生物群体中遗传变异的最终来源,是改变群体遗传结构的进化因子之一。这对于群体遗传学理论的萌芽具有重要意义。同年, Castle 提出了理想群体遗传平衡法则,可惜这一发现当时没有发表。

1904年, Pearson 在 *Mathematical Contributions to the Theory of Evolution* 论文中,用数学方式研究生物遗传与进化问题。这样,以 de Vries 为首的孟德尔学派与以 Pearson 为代表的生物统计学派之间,围绕着微小的变异能否遗传产生了激烈的争论。

1908年, Hardy 在 *Mendelian Proportions in a Mixed Population* 和 Weinberg 发表的 *On the Demonstration of Heredity in Man* 论文中,以遗传学理论为基础,再次从生物群体遗传角度分别独立证明了 Hardy-Weinberg 法则,标志着群体遗传学的诞生。

1909年, Johansson 提出了纯系学说,明确地将生物变异区分为可遗传变异和不可遗传变异,并提出了基因型与表现型概念,丰富了群体遗传学的内容。同年, Nilsson-Ehle 根据小麦粒色遗传研究结果,提出了微效多因子假说。这一假说的提出,为通过统计方法对进化中微小变异的遗传效应分析提供了依据,从而奠定了群体数量性状遗传研究的基础。

1918年, Fisher 在 *The Correlation Between Relatives on the Supposition of Mendelian Inheritance* 论文中,将统计学与遗传学结合起来,研究亲属间性状相关,进一步从理

论上证明了多因子假说的合理性,并提出了数量性状遗传的加性模型,把基因型方差分解为相应的分量,用以研究自然选择的遗传学问题,解决了两大学派之间的争论,并为数量性状进化研究提供了理论基础和分析方法。

20世纪二三十年代,是群体遗传学形成时期。其中,Wright、Fisher和Haldane为群体遗传学的理论框架构建做出了卓越贡献。

1921年,Wright在*System of Mating II: The Effects of Inbreeding on the Genetic Composition of a Population*文章中概括了生物群体的交配系统,提出了近交系数、遗传漂移概念,并研究了不同交配系统的遗传效果。随后,他对遗传漂移进行了系统研究,并于1931年发表了*Evolution in Mendelian Population*,详细地论述了他的进化理论。

1926年,Haldane发表的*A Mathematical Theory of Natural and Artificial Selection*奠定了自然选择的数学基础,并首次研究了特定群体中等位基因频率变化的趋势和速率,突变、迁移与自然选择的联合作用。1932年,Haldane的*The Causes of Evolution*,系统地总结了他在自然选择和遗传学方面的数学工作,是现代进化综合论的重要组成部分。

1930年,Fisher的*The Genetical Theory of Natural Selection*,全面总结了同时代的研究成果,发展了性选择、拟态和显性演化的概念,指出生物体的适应度与突变率成比例的增长或减小,证明了大群体携带了更多的变异,这些变异使得大群体中的个体有更多的机会生存。这本书是对达尔文进化论和孟德尔遗传学的综合,是自然选择进化论经典著作。

这些理论研究,涉及复杂的数学理论推导,不是一般生物学家们所能理解的。而且只有理论没有实验,也很难被生物学家们所接受。因此,他们的研究工作,对当时的生物学界并没有产生太大的影响。直到后来,Dobzhansky与Wright合作,以果蝇为实验材料证实了Wright的理论。1937年,Dobzhansky发表了*Genetics and the Origin of Species*,在理论上和实验上统一了自然选择学说和孟德尔遗传学,对实验生物学家产生了巨大的影响,开创了应用群体遗传学基本理论解析有关群体遗传实验数据的方法。

接下去的20年,群体遗传学理论的普及与应用研究得到进一步深入。Mayr把群体遗传学理论应用于分类学研究,并提出了地理变异和隔绝条件下新种产生的模型;Simpson首次用群体遗传学理论解释化石记录;Stebbins成功地用群体遗传学原理解释了植物的进化问题。

由于人的寿命与进化时间相比极为短暂,没有办法探测长期进化过程中群体遗传结构变化或者基因的进化变异,当时,人们只能简单地用短期变化的延续来推测长期进化的过程,群体遗传学主要涉及群体遗传结构短期的变化。

20世纪60年代以后,生化技术日益成熟,利用生物同源大分子序列变异探讨群体遗传变异规律成为可能。通过检测生物群体中同源大分子序列变异,从数量上精确地推知群体的进化演变,并可检验以往关于长期进化或遗传系统稳定性推论的可靠程度。其中:

- (1) 1964年,Crow和Kimura提出了无限等位基因突变模型;
- (2) 1965年,Zucherkandl和Pauling提出了“分子钟”概念;

(3) 1968年, Kimura 提出了中性突变的随机漂移学说;

(4) 1971年, Kimura 和 Ohta 提出无限位点突变模型。

这些理论的提出, 将生物大分子的结构变化与群体遗传学理论紧密地结合起来, 初步形成了分子群体遗传学理论体系。对生物群体中同源大分子序列变异式样的研究, 也使人们开始重新审视达尔文的以“自然选择”为核心的生物进化学说。

1982年, Kingman 在 *On the Genealogy of Large Populations* 论文中构建了“溯祖”原理的基本框架, 从而使得用少量的样本来代表整个群体进行群体遗传结构的研究成为可能, 并可以进一步推断影响群体遗传结构形成的各种演化因素。溯祖原理的“回溯”分析使得对群体进化历史的推测更加合理和可信。

1983年, Kreitman 和 Aguade 发表的 *Excess Polymorphism at the ADH Locus in Drosophila melanogaster*, 开创了基于 DNA 序列变异检测手段的实验分子群体遗传学的研究先河。

1987年, Avise、Arnold、Ball 等在 *Intraspecific Phylogeography: the Mitochondrial DNA Bridge between Population Genetics and Systematics* 论文中, 提出了系统地理学 (phylogeography) 概念, 将物种个体看作是一个操作分类单元, 在种内水平上应用系统发育的概念和方法进行研究并且了解种内的线粒体谱系在地理上是如何形成的, 在群体遗传学和分类系统学这两大松散学科之间建立了一个概念性的、过程性的桥梁, 拓展了在传统上未曾尝试的、在历史作用和周期性发生过程之间进行区分的新研究领域。

20世纪90年代, 随着 DNA 分子标记技术和 DNA 测序技术的广泛应用, 以及基于溯祖理论的各种计算机软件分析程序的开发和应用, 分子群体遗传学应用得到了迅速的发展。

2000年, Nei 和 Kimura 合著了 *Molecular Evolution and Phylogenetics*, 将群体遗传学理论在分子进化中的应用推向新的高潮。

我国群体遗传学的引进始于20世纪60年代, 发展于90年代至今。60年代初到90年代初, 国内先后翻译出版了 *Population Genetics* (Li, 1955), *Molecular Population Genetics and Evolution* (Nei, 1975), *The Neutral Theory of Molecular Evolution* (Kimura, 1983) 以及日文版《群体遗传》(大羽兹, 1978) 等国外教材。90年代, 王喜忠和杨玉华 (1992) 编写了《群体遗传学原理》, 郭平仲 (1993) 编著了《群体遗传学导论》, 张芳和李玉奎 (1999) 合编了《群体遗传学概论》等。这些出版物对我国群体遗传学教学与科学研究, 作出了积极的贡献。近10年来, 我国植物群体遗传学已由引进介绍发展到与国际接轨的崭新阶段, 部分研究已处于国际领先地位。研究内容涵盖了植物多倍体群体遗传平衡定律的拓展、群体遗传结构(同源 DNA 分化式样)、各种进化力量(如突变、重组、连锁不平衡、选择等)对植物群体遗传结构的影响、群体内基因进化方式(中性或者适应性进化)、群体间遗传分化及变异模式、植物基因流, 以及通过对栽培群体与野生群体或野生近缘种的 DNA 多态性比较研究等。植物群体遗传理论与方法在植物育种方面也取得了重要的进展。例如, 植物驯化和林木种子园建立的“奠基者效应”, 人工选择对植物群体中多态性的影响作用, 利用群体间遗传变异的林木种源选择等。

三、植物群体遗传学研究的内容

植物群体遗传学的研究对象是植物。尽管动物和植物都是生物,存在着许多生物学共同特性,但它们之间还存在着诸多差异。例如,植物不能如动物般自由移动逃避天敌,也不能在一年中不利的季节迁移到别处,无法逃避不良影响,通常表现出对生长环境(如土壤化学成分、气候因子等)的遗传适应性;动物的减数分裂发生于形成配子时期,只有二倍体的动物体,没有单倍体的动物体,因此没有世代交替。而高等植物的减数分裂都发生于孢子时期,既有二倍体的植物体,也有单倍体的植物体,两种植物体互作出现形成了世代交替。种子植物的配子体寄生于孢子体上,这就使得植物,特别是高等植物的性别概念不同于动物,性别决定也就更复杂;多倍体植物广泛存在,但多倍体动物比较少见;植物既有单性繁殖,又有有性繁殖方式,植物性系统及交配方式远比动物多样化;植物虽然固着不动,却可以通过花粉和种子传播产生基因流,但大多数情况下,被散布的种子主要集中在母株的周围,虽然存在远距离的花粉传播,然而,当花粉到达另一个群体时有可能失去生活力,也可能由于为时过早或过晚,以致不能有效地与当地的花粉竞争,植物基因流局限于邻近的植物之间。

由于植物本身固有的生物学特征,植物群体遗传学主要侧重于以下五个方面内容。

(一) 理想群体遗传特征

理想群体的遗传特征是植物群体遗传学研究的基础。包括单基因座上一对等位基因的理想群体遗传平衡法则,单基因座上的复等位基因、性连锁基因、多基因座的群体连锁不平衡以及理想多倍体群体遗传平衡法则的拓展。

(二) 影响植物群体遗传结构的进化因子

影响植物群体遗传结构的进化因子包括植物交配系统、选择、遗传漂移、突变和迁移,它们是植物群体遗传学研究的主要内容。

植物交配系统决定了植物群体遗传变异的特点,植物交配系统的不同会产生不同的后代群体遗传结构。植物交配系统分为随机交配、遗传同型交配、遗传异型交配、表型同型交配和表型异型交配系统,也可划分为近交、异交和混合交配系统,利用近交系数和自交率概念,可推导各种交配系统群体的遗传后果。

选择和遗传漂移作为进化因子,其后果都是导致植物群体中等位基因的固定或丢失。但选择是外界环境因子对群体遗传结构的影响,是确定性进化动力的系统过程,影响结果是可预测的,其遗传模型是决定模型;而遗传漂移是生物小群体固有的属性,是或然性进化动力的离散过程,不能预测小群体后代中等位基因频率的变化方向和大小,其遗传模型是随机模型。

突变和迁移具有相似的群体遗传效应,作为确定性系统因素,都是引入新的等位基因进入群体中,对植物群体影响后果都可预测,遗传模型很相似,属决定模型。但是,它们不是相同的进化因子。突变为进化提供其所必需的最原始的新的突变基因,而迁移是群体间等位基因的流动。突变率通常很低,而迁移率可能相当大,而且迁移是从一个群体中传

播出去,只对迁入群体发生遗传影响,是单向的。

因为大多数植物交配系统十分复杂,交配主要发生在相邻的植株之间,其周围的个体常常是近缘植株,即使是在连续分布的植物群体中也会产生遗传斑块(genetic patchiness)而导致小群体的遗传漂移。因此,交配系统和遗传漂移导致植物群体遗传分化,是植物群体遗传学中最为关注的基本问题,也是当前植物群体遗传学研究的热点。

(三) 植物适应性性状进化

与植物进化密切相关的性状是植物的生殖健康,包括植株的生存力、繁殖率、交配能力、寿命等,这些适应性性状都是受微效多基因控制并易受环境影响的数量性状。群体的直接进化潜力由生殖健康性状遗传力决定,通常采用群体对选择压力的适应性应答速度和程度来度量,但植物群体生殖健康这些数量性状的微效多基因遗传结构世代间变化规律异常复杂,目前面临的挑战很多。

(四) 分子进化与植物物种形成机制

植物群体遗传学研究的根本目标是研究生物进化(微进化)规律。随着现代分子生物技术的飞速发展,利用这些分子数据遗传信息、构建分子进化树、探讨植物分子进化和物种形成,已成为植物遗传学研究的主要内容之一。

(五) 植物群体遗传学研究的一些实用技术

这些技术包括检测植物群体遗传变异的遗传标记技术、群体遗传学中各种模型建立方法、群体遗传结构参数统计分析方法及其计算机模拟程序等。

四、植物群体遗传学研究模型

群体遗传学研究,主要是基于对生命现象的观察、研究,通过归纳、推理,建立一种模型,以对观察到的一种生物群体遗传现象进行解释。

模型(model),就是对要研究的现实情况进行有意识地简化,只要假定的条件大体满足,在现实群体中可能会出现。因此,模型十分适合于对大体符合条件的数据进行分析,在植物群体遗传学研究中具有十分重要的意义。

植物群体遗传学研究模型主要有实验模型、数学模型以及计算机模拟。

(一) 实验模型

实验模型(experimental model),是指在容易控制的实验条件下研究事物发展的内在规律。实验模型是实验技能的一种,是一种系统、科学认识事物的思路、角度和方法,有利于了解事物发展的各要素之间的内在联系。

自然条件下,开展植物群体遗传学研究往往是相当困难的,因为要考虑群体的大小、地理分布、交配方式、突变、迁移及选择等一系列因子,要连续进行若干年才能搜集到大量可信的遗传数据。即使搜集了资料,由于这些因子太多、经常改变,而且它们之间又以一种复杂的方式发生互作,往往难以进行解释。因此,为了把握问题的实质,设计一种仅考

虑少数几个可辨别的重要因子,而对其他因子忽略不计的简单模型,显得尤为重要。

植物实验模型就是将符合研究目的的植物群体长期保持在人工控制的环境下,改变各种可控制的环境因子,观察、研究、分析各种环境因子改变所引起的群体中特定基因型频率或等位基因频率的变化,从而模拟在自然界植物群体中发现的等位基因频率波动的原因。这种人工控制环境下生长的群体称为实验群体。例如,改变苜蓿(*Medicago sativa*)实验群体的土壤条件,苜蓿群体中特定基因型频率在一定时期内表现出明显波动,而这种波动与以前连续多年在自然群体中发现的波动相同,那么,就可以认为苜蓿群体中这种基因型的相对选择优势或劣势的波动是由这种土壤因子造成的,由此,可推断影响苜蓿群体中等位基因频率的环境因子。

(二) 数学模型

数学模型(mathematical model),是数学在近代科学研究中广泛应用的一种方法。为了研究选择、突变、迁移和遗传漂移对群体进化的影响,必须构建简单的数学模型,从而定量地揭示群体进化规律,预测性检测群体遗传变异,为那些很难或不可能确定的群体遗传学参数提供计算式,预测那些仍没有发现的现象及现象之间不可预测的相关性。通过建立植物群体遗传数学模型,将植物群体遗传特征的可测量数值或参数之间关系用数学表达式来进行描述,有助于理解各种生物和非生物因子是怎样影响植物群体进化的。

植物群体遗传学中的数学模型研究,人们最感兴趣的不是特定模型表达式的数学推导细节,而是模型中各个变量的生物学意义,否则,就会导致只看见“数学”的树木,而看不见真正生物学意义“森林”的危险。

数学模型通常要比现实需要解释的植物群体简单得多。数学模型中,有意识地省略现实植物群体的许多特征,只包含其中主要因子,否则会导致数学模型过于复杂而不实用。但是,数学模型太简单,就有可能与现实植物群体相差太大而无用。理想的数学模型应包括现实植物群体所具有的全部基本特征因子而排除所有非基本特征因子。例如,将所研究的植物以最简单生命史来处理,群体中所有植株在其下一代出生之后全部死亡,即世代间不重复。这样处理,任何世代植物群体中所有植株同时出生,同时成熟,同时交配,而且在其子代产生之后立即死亡。关键性的简化就是群体中所有个体处于相同年龄,没有任何植株能跨越世代存在而繁殖(详见第三章)。尽管这种数学模型看起来过于简单,但对研究植物群体遗传学中很多问题是足够的,甚至能满足生命周期长、生活史复杂的林木群体遗传学研究。

植物群体遗传学研究中,有决定模型(deterministic model)和随机模型(stochastic model)两种最基本的数学模型。决定模型较简单。它假定植物群体中等位基因频率从上一代到下一代的改变以唯一的方式发生,可以根据初始植物群体遗传结构知识准确地预测以后各世代植物群体遗传结构(详见第四章、第五章、第七章、第八章)。严格来说,这种模型只有当植物群体在大小上是无限的、环境或者不随时间而变化或者按决定性规则而变的前提下,才能应用。这些前提条件,在自然界中显然是不能满足的。因此,纯决定模型对于描述植物群体中等位基因频率随时间变化可能是不够的。随机的或不可预料的等位基因频率波动必须被考虑进去。

随机模型是假定植物群体中等位基因频率的变化按或然性方式发生。即根据当代条件,不能预料下一代植物群体中等位基因频率,而只能决定下一代群体中可能出现的等位基因频率分布的概率(详见第六章)。显然,随机模型比决定模型更可取,因为它们建立在更为现实的假定之上。不过,决定模型在数学处理上要容易得多,而且在某些条件下它们也能得出足够精确的结果。

目前,还没有用来确定植物群体遗传参数的通用方法。用数学模型确定的主要群体遗传参数有:突变率(突变-选择平衡模型)、选择压(群体中等位基因频率因选择力作用随世代变化模型)、迁移率(群体中等位基因频率因迁移变化模型)、基因流(不同群体中等位基因频率模型)、近交系数及自交率(混合交配模型)等。

(三) 计算机模拟

计算机模拟(computer simulating),是根据数学模型,编写计算机程序,利用计算机进行群体遗传参数计算,并模拟群体各世代遗传过程的一种方法。

人的一生中所能观察到的植物世代数是很有有限的,不可能对千百年生态条件作用下的植物群体遗传结构变化进行实际观察,而用数学模型研究,随机因子、选择、迁移和突变对自然群体的综合影响相当复杂,似乎得不到一种综合的数学模型,只能对影响群体遗传结构的进化因子做简化处理,或研究几个限制的影响因子。因此,根据遗传学原理,编写计算机程序,用以模拟不同的选择压、突变率、迁移率、繁殖率、近交和随机漂移等因子作用下遗传变异的长期效应,从而能够对物种(群体)进化(或退化)和植物遗传资源的保护、利用作出预测。

计算机模拟的基本原理是利用计算机所提供的随机函数,结合生物群体遗传规律,产生模拟数据,模拟生命过程的随机现象。随机函数的数值范围为 $[0,1]$,每次产生的随机数均匀分布,但各不相同。计算机模拟非常简便,大大提高了研究的效率。应用计算机技术模拟复杂的植物群体随机过程或计算变量的值,是植物群体遗传学研究的重要辅助手段。

将计算机应用于群体遗传学研究,主要归功于 20 世纪 80 年代微型计算机的普及和应用。植物群体遗传学研究中常常利用计算机模拟群体中等位基因频率变化的所有确定性和或然性,将自然植物群体的复杂系统分解为若干个较为简单的子系统,再借助它们高效运算的优势,揭示植物群体遗传变异的实质。例如,植物群体很小时,群体中等位基因频率会发生随机波动,产生遗传漂移。这种波动会导致植物有限群体中等位基因固定或丢失。遗传漂移来源于两个随机过程的发生,亲本基因型和子代基因型(见第七章),因此,利用计算机随机函数来模拟遗传漂移现象极为有用。近年来,群体遗传结构模拟研究不断深入,主要涉及针对特殊群体设计模拟模型,提高群体遗传模拟性能和算法的研究,并将研究结果应用于分子进化研究。

第一章 分子生物学与生物统计学基础

尽管基因的分子生物学和生物统计学不是本书的主题,但为了便于理解和掌握本书的内容,有必要先介绍这方面的基础知识。

第一节 基因及其有关概念

一、DNA 与染色体

生物的遗传信息主要由脱氧核糖核酸(deoxyribonucleic acid, DNA)分子携带。DNA 分子通常是由两条平行但反向相互缠绕的互补链组成,形成一种右手螺旋状结构(图 1-1)。每条链都是由四种核苷酸(nucleotide)组成的线性多聚核苷酸。每个核苷酸由一个糖基、一个磷酸基团和一个环状含氮碱基三部分组成。DNA 的糖基是脱氧核糖,碱基有腺嘌呤(adenine, A)、鸟嘌呤(guanine, G)、胞嘧啶(cytosine, C)、胸腺嘧啶(thymine, T)四种类型。碱基与脱氧核糖结合构成脱氧核苷(deoxynucleoside),脱氧核苷与磷酸基团结合构成脱氧核苷酸。两条核苷酸链靠彼此之间的氢键联系在一起。其中,腺嘌呤与胸腺嘧啶通过两个氢键配对,鸟嘌呤与胞嘧啶通过三个氢键配对(图 1-2)。

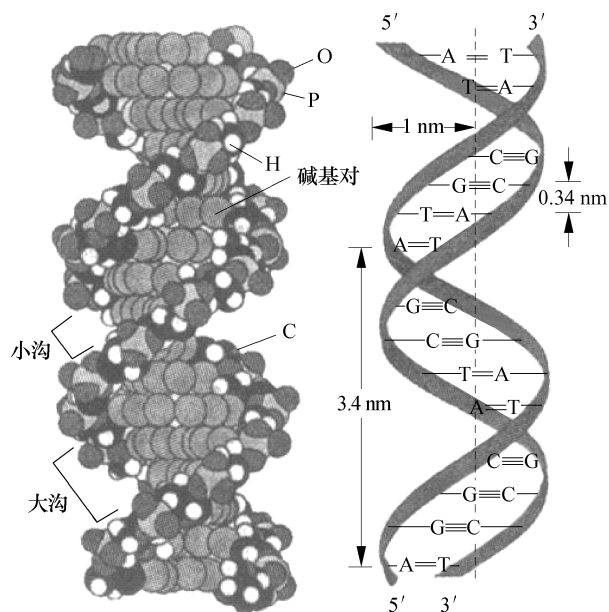


图 1-1 DNA 分子的双螺旋结构模型(Russell,1998)

DNA 分子有成千上万个碱基对(base pair, bp),但 DNA 链之间碱基配对只有 A-T、

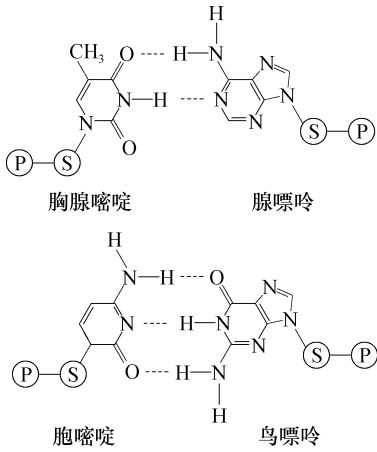


图 1-2 DNA 分子四种脱氧核苷酸 (Russell, 1998)

P. 磷酸根; S. 糖

C-G、T-A、G-C 四种碱基互补配对 (complementary base pairing) 形式。DNA 分子通过不对称的 5'-磷酸二酯键 (phosphodiester bond) 共价连接在一起, 一端是一个游离的 5'-三磷酸根 (-P), 称 5' 端; 另一端是一个游离的 3'-羟基 (-OH), 称 3' 端。因此, DNA 分子具有极性。

生物细胞核中 DNA 分子以染色体 (chromosome) 形式存在。染色体, 是由与蛋白质和少量 RNA 相结合的双链 DNA 组成的, 易被碱性染料着色的一种丝状或杆状物。每条染色体一般包括着丝粒 (centromere)、端粒 (telomere)、DNA 复制起始点 (origin of replication) 以及基因 (gene) 和非编码序列 (non-coding sequence)。着丝粒, 是细胞分裂过程中染色体分离的一种装置, 也是姐妹染色单体 (sister chromatid) 在分开前相互联结的位置, 在染色体的形态上表现为一个缢痕 (constriction)。端粒, 是染色体的两端特殊组成, 使染色体保持其个体性 (individuality), 不会引起染色体之间端与端融合或端部松散而解体。二倍体生物体细胞中, 存在着—对形态和结构相同的染色体, 这对染色体互称为同源染色体 (homologous chromosome)。

DNA 复制首先是两条链间的氢键断裂并使双链解旋、分开, 然后以每条链为复制模板 (replicate template), 按碱基互补配对原则, 由一条链形成互补的两条链 (图 1-3)。新

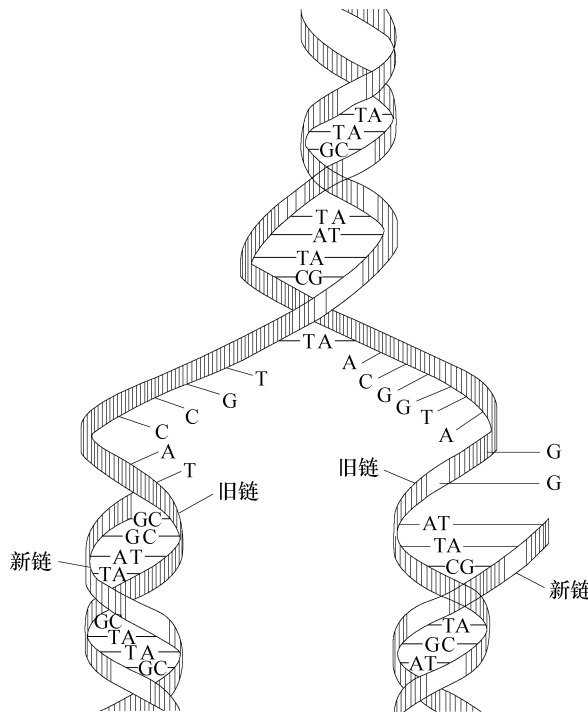


图 1-3 染色体半保留复制 (Russell, 1998)