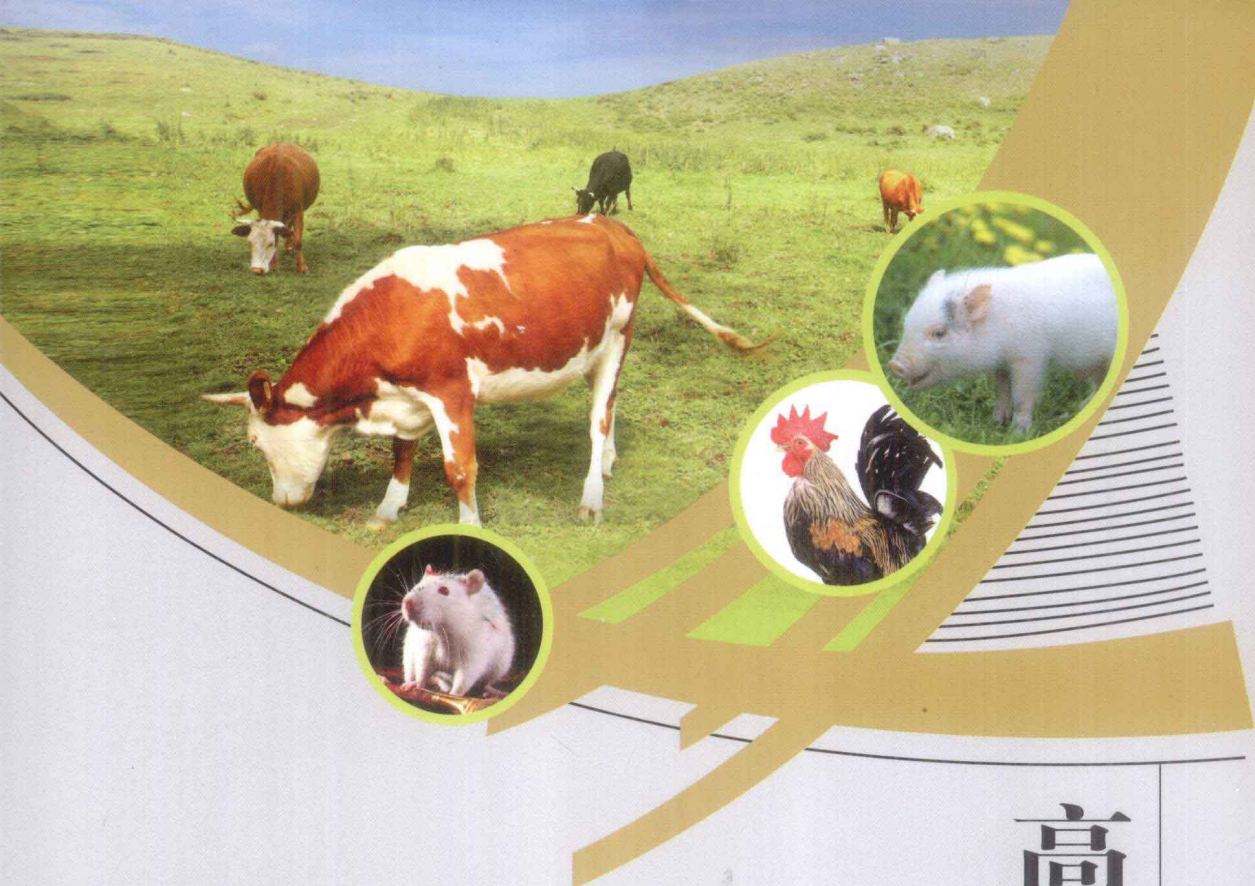


研究生创新教育系列丛书

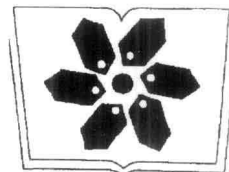


高级动物基因工程

李宁 主编



科学出版社



中国科学院科学出版基金资助出版

研究生创新教育系列丛书

高级动物基因工程

李 宁 主编



科学出版社

北 京

内 容 简 介

本书由 12 个专题构成,包括基因定位的原理和方法、小 RNA 原理与研究进展、表观遗传原理与研究、基因芯片原理与应用、高通量测序及荧光原位杂交技术、RNA 干扰原理与技术、基因打靶原理与技术、干细胞理论与研究方法、动物克隆原理与进展、转基因动物原理与进展、动物分子育种原理与进展及动物分子进化原理与研究方法。

本书的特点是“原理策略瞄准国际前沿,技术方法介绍深入浅出”。全书内容涵盖了目前动物基因工程研究的大部分研究重点及热点内容。例如,小 RNA 研究、表观遗传学、转基因动物、高通量测序研究及技术等,这些技术都是目前动物基因工程领域的最前沿技术。书中每一章系统地介绍一个专题,包括相关的技术原理、研究策略、研究方法 & 研究进展。其中在对研究进展的介绍中,作者综述了相关领域权威刊物的最新研究成果,介绍了各项专题技术的最前沿研究策略。而与此同时,每部分的专题所列的实验方法均在作者的研究实践中进行了验证。

本书是国内第一本专门并全面介绍动物基因工程及其研究最新进展的学术著作,可作为高等院校动物生物技术、动物科学的教学参考书,也可作为相关科研工作者、研究生的专业参考书。

图书在版编目 (CIP) 数据

高级动物基因工程/李宁主编. —北京:科学出版社,2012

(研究生创新教育系列丛书)

ISBN 978-7-03-032854-0

I. ①高… II. ①李… III. ①动物-基因工程 IV. ①Q953

中国版本图书馆 CIP 数据核字 (2011) 第 240084 号

责任编辑:罗 静 李秀伟 李晶晶/责任校对:包志虹

责任印制:钱玉芬/封面设计:陈 敬

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>

装 订 厂 印 刷

科学出版社发行 各地新华书店经销

*

2012 年 1 月第 一 版 开本:787×1092 1/16

2012 年 1 月第一次印刷 印张:33 3/4 插页:4

字数:780 000

定价:118.00 元

(如有印装质量问题,我社负责调换)

《高级动物基因工程》编写组名单

主 编 李 宁 (中国农业大学)

副主编 胡晓湘 刘冉冉

参编者 (按姓氏笔画排序)

于胜利	王 弘	王少华	王青松	王晓波	冯春刚
吕小青	刘冉冉	刘晓辉	阮卫民	孙东晓	杜 茁
杜政霖	杨 惠	李 宁	李世杰	李向清	李庆贺
宋嘉哲	张 坤	张 然	张 瑾	林 莉	罗俊杰
郝 瑞	胡松年	胡晓湘	赵永辉	袁 晶	贾 博
顾晓荣	高 宇	高 宇			

前 言

基因工程是指利用现代生物技术手段，对生物体的遗传物质进行编辑和改造，产生、修饰或消除特定基因的功能，从而使生物体性状发生定向变异的技术过程。1982年美国华盛顿大学 Palmiter 教授等将大白鼠生长激素基因转移到小白鼠受精卵中，成功地育成一种个体比正常小白鼠大1倍的“超级小鼠”，在此之后动物基因工程开始了突飞猛进的发展。1996年，克隆羊“多莉”在英国诞生，世界为之震动，成为基因工程技术史上划时代的突破；2009年，重组人抗凝血酶Ⅲ（商品名：ATryn）获准进入美国市场，标志着转基因动物制药逐步走向商业化。与此同时，我国的动物基因工程也一日千里地飞速发展。2003年，中国农业大学成功获得了中国第一头体细胞转基因克隆牛；2005年，成功地获得我国第一头体细胞克隆猪。在动物转基因克隆领域，我们已经具备了独立自主、完备的知识和技术体系。2008年，我国启动了转基因生物新品种培育科技重大专项。作为《国家中长期科学和技术发展规划纲要（2006～2020）》确定的农业领域唯一的重大专项，转基因动物新品种培育已经得到了我国政府的高度重视。毫无疑问，在各方的共同努力下，中国动物基因工程的发展将极大地推动未来医药卫生、家畜改良等领域的更快发展。基于这样的技术发展背景和趋势，我们编写了本书。本书是至今为止最为系统和深入地介绍动物基因工程相关主题的专著，目的是更好地和同行相互交流，促进中国动物基因工程技术的发展和应用。

全书共12个专题，全面系统地阐述了动物基因工程研究的大部分研究重点和热点内容，包括基因定位的原理和方法（冯春刚、胡晓湘、高宇编写）、小RNA原理与研究进展（王晓波、杨惠、郝瑞编写）、表观遗传原理与研究（李庆贺、杜茁、高宇、林莉编写）、基因芯片原理与应用（胡晓湘、张瑾、顾晓荣编写）、高通量测序及荧光原位杂交技术（胡松年、杜政霖、贾博、赵永辉编写）、RNA干扰原理与技术（袁晶、王青松、王少华、吕小青编写）、基因打靶原理与技术（于胜利、李向清编写）、干细胞理论与研究方法（宋嘉哲、阮卫民编写）、动物克隆原理与技术（张坤、罗俊杰、李世杰编写）、转基因动物原理与进展（张然、王弘编写）、动物分子育种原理与进展（冯春刚、胡晓湘、孙东晓编写）及动物分子进化原理与研究方法（刘晓辉、刘冉冉编写）。书中每一章系统地介绍一个专题，包括相关的技术原理、研究策略、研究方法及研究进展。其中在对研究进展的介绍中，作者综述了相关领域权威刊物的最新的研究结果，介绍了各项专题技术的最前沿研究策略。而与此同时每部分专题所列实验方法均在作者的研究实践中进行了验证。总体来讲，本书内容丰富，论述深入浅出，表现形式图文并茂。我们深信这本书的出版，将更好地为在生物学、动物科学领域中开展动物基因工程相关研究的人员提供有用的参考书。

本书得到了中国科学院科学出版基金的资助，在此表示最诚挚的感谢。全书的编写是在李宁院士的细心指导下，编写组全体成员历时四年的辛勤劳动成果。然而，虽然几经修改定稿，但鉴于我们编写经验及水平的有限，缺点和不妥之处在所难免。衷心希望广大读者不吝指正。

《高级动物基因工程》编写组

2011年10月

目 录

前言

第一章 基因定位的原理和方法	1
第一节 引言.....	1
第二节 数量性状及其遗传基础.....	3
第三节 QTL 定位的基本原理	6
第四节 候选基因分析.....	7
第五节 QTL 连锁定位的基本步骤	9
第六节 QTL 定位研究进展及成就.....	27
第七节 遗传基因组学	37
第八节 全基因组关联分析	39
第九节 基因定位的前景展望及其应用	43
参考文献	44
第二章 小 RNA 原理与研究进展	49
第一节 miRNA 的发现过程	50
第二节 miRNA 成熟过程	54
第三节 miRNA 特征	57
第四节 miRNA 功能	61
第五节 miRNA 研究策略	66
第六节 其他小分子 RNA——piRNA	82
第七节 miRNA 研究展望及应用	85
第八节 miRNA 资源	85
参考文献	86
第三章 表观遗传原理与研究	91
第一节 表观遗传信息——一类重要的“隐藏”遗传信息	91
第二节 高等动物表观遗传系统	92
第三节 表观信息对生命活动的调节.....	120
第四节 表观遗传学研究方法.....	140
第五节 表观遗传学前沿领域.....	145
参考文献.....	157
第四章 基因芯片原理与应用	162
第一节 基因芯片的概念、分类与应用.....	163
第二节 基因表达谱芯片.....	166
第三节 基因组 DNA 检测芯片	179

参考文献	189
第五章 高通量测序及荧光原位杂交技术	192
第一节 新一代高通量测序综述	192
第二节 新一代高通量测序技术 Solexa 简介	193
第三节 高通量测序仪 SOLiD 简介	200
第四节 454 的技术简介	206
第五节 单分子测序技术	209
第六节 纳米孔测序	214
第七节 数据分析的软件和生物信息工具	217
第八节 荧光原位杂交技术及其应用	218
参考文献	226
第六章 RNA 干扰原理与技术	230
第一节 RNA 干扰的发现	230
第二节 RNA 干扰的分子机制	232
第三节 RNA 干扰实现的技术路线	239
第四节 RNA 干扰的应用	245
第五节 实现基因表达抑制的方法	254
第六节 RNA 干扰实验的注意事项	266
参考文献	280
第七章 基因打靶原理与技术	287
第一节 基因打靶技术简介	287
第二节 基因打靶技术原理	292
第三节 基因打靶技术	299
第四节 新型的基因打靶技术	308
参考文献	317
第八章 干细胞理论与研究方法	323
第一节 干细胞概述	323
第二节 胚胎干细胞	328
第三节 成体干细胞	344
第四节 诱导性多能干细胞	363
参考文献	370
第九章 动物克隆原理与进展	376
第一节 动物克隆研究历程	376
第二节 体细胞核移植各环节及其影响因素	387
第三节 体细胞克隆技术的应用	399
第四节 体细胞核移植机理	404
参考文献	411

第十章 转基因动物原理与进展	417
第一节 转基因动物研究概述.....	417
第二节 转基因动物制备方法.....	420
第三节 提高转基因的策略.....	437
第四节 转基因动物的应用.....	446
第五节 动物转基因亟待解决的问题及展望.....	453
参考文献.....	455
第十一章 动物分子育种原理与进展	464
第一节 引言.....	464
第二节 常规育种方法.....	466
第三节 分子育种的制定和实施.....	473
第四节 基因组选择.....	477
第五节 分子育种在农业动物中的应用现状和前景.....	481
参考文献.....	486
第十二章 动物分子进化原理与研究方法	487
第一节 分子进化涉及的基本概念和基本理论.....	488
第二节 分子进化的基本研究方法.....	495
第三节 利用 DNA 进行系统演化分析	503
第四节 古 DNA 及其在动物分子进化研究中的应用	519
参考文献.....	524

彩版

第一章 基因定位的原理和方法

第一节 引言

自然界中众多生物每个物种都具有复杂的多样性，人们一直在试图揭示不同物种及不同个体表型变异的本质，但是进展较为缓慢，直到遗传学诞生之后，这方面研究才获得了较快的发展。在遗传学中，表型是指一个特定基因型的生物个体在一定的环境条件下所表现出来的性状特征，是由基因和环境共同控制的。基因型是一个生物个体全部基因组合的总称，反映了生物个体的遗传组成。但是整个生物体的基因型是无法表示的，遗传学中使用的基因型往往是针对某一具体性状而言的。生物体的表型性状可以分为数量性状和质量性状，其遗传基础也是不同的，其中质量性状的遗传基础比较简单，数量性状比较复杂。基因定位在遗传学上是指确定基因在染色体上的位置和排列顺序的过程，我们此处介绍的基因定位是指从遗传学上确定控制生物表型的基因在染色体上的位置并分离基因突变的过程。质量性状因为遗传基础简单，通常是由单个基因控制的，而数量性状由数个甚至成百上千的基因控制，定位比较复杂，因而我们在一定程度上可以将质量性状的基因定位理解为简化的数量性状基因定位。数量性状基因定位又称为 QTL (quantitative trait locus) 定位 (QTL mapping)，下面主要以 QTL 定位的原理和方法来介绍基因定位。

一、什么是 QTL？为什么要研究 QTL？

影响数量性状表达的 DNA 或染色体片段被称为数量性状基因座。随着数量遗传学和分子生物学等的发展，QTL 定位研究取得了较大的进步，人们通过定位得到了大量影响重要经济性状的 QTL，并分离得到了部分效应较大的基因。

QTL 的定位及分离不仅具有重要的科学价值，而且在农业和医学研究中也具有重要作用：①有助于发展和完善数量遗传学理论，并促进进化理论的发展；②与分子生物学结合，共同揭示不同表型变异的形成机制；③通过标记辅助选择 (MAS) 等方法与传统育种理论相结合能够有效地提高育种效率，尤其是低遗传力性状、限性性状及抗病性状等；④更有效地与转基因、克隆等生物技术结合，共同促进农业生产；⑤预先检测和分离致病基因位点，有利于对致病基因携带者提前进行预防和治疗。

二、为什么利用农业动物进行 QTL 研究？

农业动物育种在过去将近 10 000 年中一直在进行类似的表型突变筛选，在驯化过程中将有利于行为、形态、生理及抗病性等性状的自然突变个体保留下来，同时也保留了在目前小鼠突变筛选中所忽略的大量的表型变化。在驯化过程中，长时间对大量突变的累计筛选导致了极其显著的表型变化，形成了宝贵的遗传资源。一个人工选择作用的例子就是在过去的几千年中驯化形成了约 400 个不同品种的狗，其中大部分是在过去的

几百年中驯化而成的，表型变异程度超过了野生犬科动物 1000 万年中自然产生的变异。农业动物的生产性能也在过去的 50 年中发生了明显的增长。例如，美国每头奶牛平均产奶量在 40 年中提高了 1 倍（图 1.1.1），其中遗传因素起了主要作用，主要就是借助数量遗传学的手段培育了专门的品种，并借助于数量遗传学和分子生物学进行选种的遗传改良。

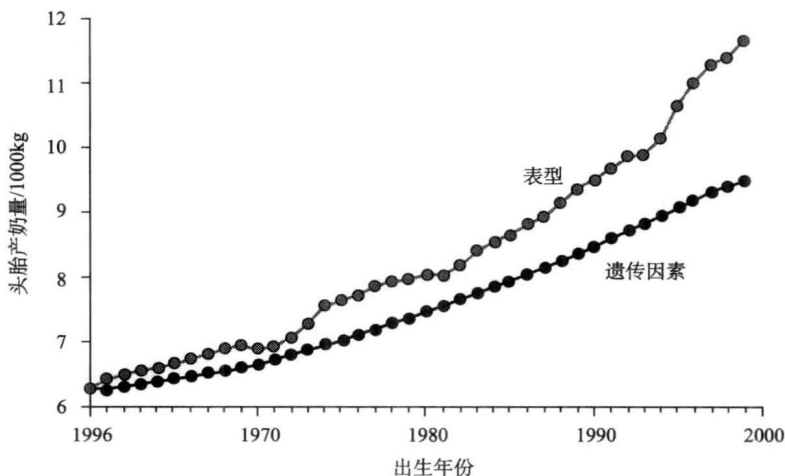


图 1.1.1 美国荷斯坦奶牛头胎产奶量增长趋势 (Jack et al., 2002)

20 世纪 80 年代分子遗传学进入到农业动物研究中，用于识别影响表型变化的 DNA 片段。分子遗传学最早主要是用于研究重要经济性状，以便更有效地实施标记辅助选择 (marker assisted selection, MAS)。农业应用依然是目前进行农业动物基因组研究的主要目标，但是基因组研究已不仅仅局限于此。比较基因组学能够促进基因组结构和功能的研究，通过研究影响表型变化的分子机制也能促进表型-基因型研究，并促进农业动物的驯化研究。

为便于后续理解，在此先介绍几个基本概念：①等位基因 (allele)：在一对同源染色体的同一基因座上的两个不同形式的基因；②连锁不平衡 (linkage disequilibrium)：指不同基因座位上等位基因的非随机组合，它反映的是一种状态，即当位于某一座位的特定等位基因与同一条染色体另一座位的某一等位基因同时出现的概率大于群体中因随机分布而使两个等位基因同时出现的概率时，就称这两个座位处于 LD 状态。LD 是由突变产生的多态性形成的，因重组的发生而被打破，由此可知，突变和重组是影响 LD 的重要因素。除此之外，其他生物因素和历史因素也影响 LD 的程度和分布，如物种的异交率，即交配体系、染色体位置、群体大小、基因或染色体片段所受的选择强度、遗传漂变等。例如，两个相邻的基因 AB，它们各自的等位基因为 ab，假设 AB 相互独立遗传，则后代群体中观察得到的单倍体基因型 AB 中出现的 $P(AB)$ 的概率为 $P(A) \times P(B)$ ，实际观察到的群体中单倍体基因型 AB 同时出现的概率为 $P(AB)$ ，计算这种不平衡的方法为： $D = P(AB) - P(A) \times P(B)$ ；③单倍型 (haplotype)：在遗传学上是指在同一染色体上共同遗传的多个基因座上的等位基因的组合。例如，一对同源染色体

上的三个位点均有两个等位基因, Aa 、 Bb 、 Cc , 可能存在的 8 种单倍型分别为 ABC 、 ABc 、 AbC 、 Abc 、 aBC 、 aBc 、 abC 和 abc 。

遗传隔离群体在人类遗传学研究中具有非常重要的价值, 主要是由于长期维持在较小的群体规模或者遭遇瓶颈效应等原因造成。与一般群体相比, 隔离群体在基因位置克隆研究中具有以下优势: ①连锁不平衡 (linkage disequilibrium) 的范围较大, 即更大片段的染色体共同传递, 因此通过少量的分子标记即可检测到目标基因所在的区域; ②因始祖效应造成等位基因异质性 (allelic heterogeneity) 降低, 有利于分离相同的单倍型 (haplotype), 即后代中具有相同表型的个体通常是由相同的基因突变导致; ③基因座异质性 (locus heterogeneity) 降低, 遗传背景减小 (等位基因异质性是指同一基因的不同突变所引起的相同或相似的表型; 基因座异质性是指不同基因座位的不同基因突变所引起的相同或相似的表型); ④环境噪声降低, 即检测得到的遗传因素的真实性强, 而非外界环境造成 (Georges, 2007)。遗传隔离群体在单基因疾病致病基因克隆中已被证明具有极大的优势, 但在复杂性状分析中是否也具有优势目前仍存在争议。

农业动物群体具有遗传隔离群体的一些基本特征。动物育种中为获得纯系个体通常需要对亲缘个体进行交配, 以使品种特征纯合。大量的品种正是通过这种方法获得的, 尤其是在近年实施品种登记办法以后。优秀公畜的广泛使用, 特别是人工授精技术的推广, 使得农业动物有效群体大小急剧减小。例如, 全世界有几千万头荷斯坦奶牛, 但是其有效群体大小仅为几百头。

群体遗传结构比较研究表明, 农业动物与一般的人类群体相比, 处于完全连锁不平衡状态的染色体片段明显较长, 这有利于农业动物中 QTL 的初步定位, 但是定位的精度为 $5\sim 40\text{cM}$, 不利于进行精细定位, 而人类群体中精细定位的结果非常准确。

第二节 数量性状及其遗传基础

一、群体遗传结构

数量性状的研究对象是群体, 对于个体而言, 研究其数量性状无法开展。群体广义上是指在一定的时间和空间范围内, 具有特定的共同特征和特性的个体集合。这里所说的群体是指孟德尔群体, 即一群能够相互交配并进行有性繁殖的个体, 所有个体共享一个共同的基因库, 相互之间可以进行基因交流。

一个个体的存在时间是有限的, 除非发生突变, 否则其遗传基础终生不变。一个群体一般是长时间存在的, 群体可大可小, 分布可广可窄, 其遗传特征, 即遗传结构由所有个体的遗传基础决定, 并受到多种因素的影响 (如突变、选择、迁移、遗传漂变等) 而发生变化。群体遗传结构的主要指标是基因频率和基因型频率。

(一) 基因频率

位于同源染色体同一基因座位上控制某一性状的不同基因称为等位基因。基因频率就是在一个群体中某一基因的数量与相同基因座位的所有等位基因总数的相对比例, 取值为 $0\sim 1$, 一般用小数形式表示, 没有负值, 它是群体遗传特性的基本标志。同一基

因座位的全部等位基因的基因频率之和为 1。在只有一对等位基因（如 A 和 a ）时，其基因频率一般用 p 和 q 表示， $p+q=1$ 。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基因频率往往不同。

（二）基因型频率

基因型频率是指在生物的群体中，某一基因座位的特定基因型在全部基因型中所占的比例，取值为 0~1，一般用小数形式表示，没有负值。同一座位的全部基因型频率之和为 1。在只有一对等位基因（ A 和 a ）时，它们组成的基因型有三种： AA 、 Aa 和 aa ，可以分别用 D 、 H 和 R 表示其频率， $D+H+R=1$ 。

基因频率和基因型频率是密切关联的。只有一对等位基因（ A 和 a ）时，整个群体中， AA 基因型频率为 D ， A 基因比例为 $2D$ ； Aa 基因型频率为 H ， A 基因和 a 基因基因比例均为 H ； aa 基因型频率为 R ， a 基因比例为 $2R$ 。则有 A 基因频率 p 和 a 基因频率 q 分别为：

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

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

二、数量性状

生物的性状一般分为质量性状和数量性状两类。质量性状的变异可截然区分为若干种明显不同的类型，并能够分别用语言描述，如牛角的有无、人类的血型等；数量性状的变异呈连续状态，个体间的差异只能以数量来描述，如奶牛的产奶量、猪的日增重等，农业动物中的多数经济性状均为数量性状。

还有一类性状称为阈性状，它们的遗传是由多基因所决定的，但它们的表现却是非连续的。例如，单胎哺乳动物中，有单胎、双胎或三胎等；在抗病性状上，有死亡或存活、生病或健康等。但是在遗传学研究中较少单独对其进行分析，多数情况下将其作为数量性状进行研究。

以上三类性状可以根据以下四点进行区分：①性状的变异是描述性的或是可以度量的；②性状的变异是间断性分布或是连续性分布；③性状的表现是否容易受到环境的影响；④性状的遗传基础是单基因或是多基因。其差异见表 1.2.1。

表 1.2.1 质量性状、数量性状与阈性状的差异

项目	质量性状	数量性状	阈性状
考察方式	描述	度量	描述
变异分布	间断性	连续性	间断性
环境影响	不敏感	敏感	敏感
遗传基础	单个或少数几个基因	多基因	多基因
主要性状类型	外貌	生长、品质	疾病等

三、数量性状遗传基础

数量性状受到多个基因的控制，而且易受环境的影响，其表现不像质量性状一样易于观察、研究，在其早期的认识过程中存在较大的分歧，直到 1909 年瑞典遗传学家 Nilsson-Ehle (1909) 提出了多基因假说 (polygene hypothesis) 并得到遗传学界的普遍接受，才结束了孟德尔学派和统计学派关于连续变异是否符合孟德尔定律的长期论战，为数量遗传学奠定了重要的理论基础。之后，丹麦生物学家 Johannsen 于 1909 年提出了纯系学说，对连续变异的原因做了重要补充。目前，微效多基因假说仍然是动物育种中最有效的模型。

(一) 微效多基因假说

这一假说的主要论点是：数量性状是由大量效应微小、类似且可加的基因控制，这些基因在世代传递的过程中各自服从孟德尔规律，基因间一般没有显隐性区别。数量性状同时受到遗传和环境的影响，而且数量性状对于环境的影响相当敏感。

微效多基因系统仅仅是数量性状呈现连续变异的内在遗传基础，而影响数量性状表现的环境因素是数量性状呈现连续变异的外部原因。在遗传和环境的内外因素综合作用下，数量性状表现为连续变异，根据概率论的中心极限定理，这种分布应渐进于正态分布，这与实际情况相符，即数量性状表型值为钟形的正态分布。

目前，多个物种的基因组序列已经公布，人们在 DNA 等水平上对于各种生物也有了更深入的认识。人们发现各个物种的基因数目并不是无限的，而是非常有限的，如人的基因数目为 2.5 万~3.0 万个，线虫基因数目约为 1.5 万，数量有限的基因控制着如此众多的数量性状，可能的原因主要有：①基因与性状并非是一一对应的关系，而是一个基因影响多个性状，一个性状受多个基因的控制，并且基因作用包括等位基因间的加性效应和非等位基因间的非加性效应，而众多的基因相互之间又通过网络相互影响，进一步增加了基因和性状间关系的复杂性；②周围的环境效应通过与不同的基因之间的互作，影响着基因功能的行使，使得不同个体在表型上体现出更多的差异。

(二) 数量性状基因座

尽管微效多基因假说在育种中取得了巨大的成就，但是它毕竟只是一个假说，人们并没有据此而真正地阐明数量性状的遗传基础。如果想从遗传上真正地描述数量性状就必须研究清楚影响数量性状的基因特征，包括基因数目、基因类型、基因频率、效应大小、基因分布等。但是限于生物学发展的时代背景，各种理论和技术都不能满足需求，因此数量性状遗传机制的阐明进展比较缓慢。

随着分子生物学的发展，分子标记的出现改变了对数量性状遗传机制的研究策略。在研究基因组结构时发现，控制生物性状的基因在基因组中并不是随机分布的，往往是有相关功能的成簇地分布，基于这一认识，Geldermann (1975) 在 1975 年提出了控制数量性状的数量性状位点的概念。数量性状基因座 (QTL) 是指基因组中影响目标性状的染色体片段，可以是单个的基因，也可以是共同影响同一性状的多个紧密连锁的基

因。从广义上来说,数量性状基因座位是所有影响数量性状的基因座(不论效应大小),但通常只将那些可被检测出的、有较大效应的基因或染色体片段称为 QTL,而那些不能检测出的基因仍被当作微效多基因来看待。一般认为如果一个基因的效应(两种纯合基因型的基因型值之差)达到 $0.5 \sim 1.0$ 表型标准差时,即可看作是主效基因。同一性状一般由多个 QTL 共同决定,而同一个 QTL 也可能影响到多个性状。

第三节 QTL 定位的基本原理

QTL 定位主要是一种将 QTL 定位到基因组位置的方法和策略,就是通过连锁分析或关联分析将遗传变异和表型变异联系起来,即寻找表型变异和基因型差异之间的关联。QTL 定位主要包括:①确定 QTL 的位置,即将 QTL 定位到染色体的特定位置上;②估计 QTL 的效应值,即估计不同的 QTL 及同一 QTL 的不同基因型的效应值。目前,在一般的连锁分析和关联分析中,我们通常更关心 QTL 的位置。随着学科的发展,将来能够阐明 QTL 的作用机制,即从分子水平上彻底阐明 QTL 影响数量性状的作用机制。

分子生物学技术的发展以及较高密度遗传连锁图谱的构建成功,允许研究人员通过一定的试验杂交群中分子标记的分析来进行 QTL 定位。利用遗传标记进行 QTL 定位的遗传基础是个体带有连锁信息,遗传标记必须与控制数量性状的 QTL 位于同一染色体上,呈现较紧密连锁,处于连锁不平衡状态。QTL 的检测及定位通常是建立在下述假定的基础之上:①遗传标记和数量性状基因位点遵循孟德尔遗传规律;②数量性状基因位点在整个基因组上分布稀少,如发现显著的效应,这就是由单个数量性状基因位点所影响的;③数量性状基因只有两个等位基因(Q 及 q)。QTL 和附近的标记是紧密连锁在一起的,并在传递过程中发生共分离,我们就能够检测不同的标记基因型并计算其对应的表型值,如果不同基因型的表型值之间有差别,则能够判断出这个位置存在一个 QTL。检测出此 QTL 的强度取决于此 QTL 和标记之间的重组率,即连锁程度,以及该 QTL 效应的大小。

连锁分析最早在 1920 年被应用到果蝇单因子遗传研究中(Altenburg and Muller, 1920),同时 Fisher (1918)提出了统计理论将孟德尔遗传因子和数量性状结合起来。到 20 世纪 80 年代末,在模式生物中通过杂交实验来进行数量性状的 QTL 定位逐渐变得可行。

目前,对 QTL 进行检测和分析的方法主要有两种,一种是关联分析,另一种是标记——QTL 连锁定位(QTL linkage mapping),也称为基因组扫描,即通常所说的 QTL 定位(QTL mapping)也即 QTL 连锁定位。其中关联分析主要是利用标记和目的基因之间的连锁不平衡来进行的,随着技术的发展和标记数量的不同,主要包括早期的候选基因分析和近几年发展起来的全基因组关联分析(genome wide association analysis, GWAS),其中候选基因分析只针对具体的某个基因与表型进行关联分析,而全基因组关联分析是利用全基因组高密度分子标记从全基因组范围内寻找影响某个表型的基因或突变。连锁分析主要是利用家系结构的资源群体,利用染色体之间的重组而定位

目的基因。两种方法各有优势，同时也存在着一些不足，下面将对两种方法分别进行介绍。其中候选基因分析方法的发展已非常成熟，在此仅作简单介绍；全基因组关联分析方法近几年发展很快，仅对其原理作简单介绍。本章节重点介绍标记——QTL 连锁分析方法。

第四节 候选基因分析

候选基因分析主要是从可能影响目标性状的基因中挑选并验证某基因的功能，进而判断该基因是不是影响目标性状的 QTL。所选择的基因可能是结构基因、调节基因或者是在生化代谢过程中影响性状表达的基因。候选基因的方法在早期的定位研究中因其操作性强曾发挥了重大作用，但是现在已被证明不是最优的方法，结果不可靠，尤其是对于数量性状来说更是如此，在此仅进行简要介绍。

一、候选基因分析的基本步骤

(一) 原 理

候选基因分析主要是利用标记和基因之间的连锁不平衡，通常直接选取候选基因内的分子标记进行检测，然后分析标记与表型之间的关联，若达到显著性水平，则表明该基因影响了所分析的表型。

(二) 选择候选基因

候选基因选择的主要依据包括以下几点。

(1) 已知的生理生化知识进行判断，例如，生长激素、类胰岛素等在动物生长发育过程中起到重要作用，因而相应的生长激素基因、类胰岛素生长因子等基因可以作为生长性状的候选基因进行研究；

(2) 其他物种中发现的类似的功能基因，例如，将目前在小鼠等模式生物中发现的影响疾病的基因作为候选基因来研究相同基因在人类疾病中的作用；

(3) QTL 定位结果，这和后面提到的标记——QTL 连锁分析结果相结合，从 QTL 定位区间中挑选基因作为候选基因进行研究；

(4) 结合生物信息学手段，目前多个物种的基因组序列已经公布，并有多多个包括序列、结构预测、调控元件分析、代谢网络分析等信息的数据库，以及物种表达谱芯片等数据库，研究者可以从中查询和预测所研究的基因信息。

以上四个方面并不相互冲突，挑选候选基因时可以综合各方面的信息，尤其是目前各种数据库中包含了非常多的信息，充分利用生物信息学的手段提前进行预测分析，可以达到事半功倍的效果。

(三) 获得候选基因序列

如果所研究物种的序列信息已经公布，则通常可以直接从中得到较多的信息。如果

所研究物种的序列未知,那么通常可以比较其他相近物种的同源基因序列等而获得部分基因序列等。可以根据获得的序列设计引物进行扩增、测序进行验证和补充,进而得到比较完整、可靠的整个基因序列或能够满足研究目的的部分序列。

(四) 基因内多态性的寻找和大规模检测

在一个基因内部及其 UTR 区域等通常都有较多的多态位点,它们通常和候选基因的功能位点处于紧密的连锁不平衡状态,因而这些多态位点的检测和分析对于这个基因功能的研究是非常有帮助的。目前应用最广泛的分子标记(多态位点)主要是单核苷酸多态性(single nucleotide polymorphism, SNP),即在 DNA 序列的单个位点发生的核苷酸的插入、缺失、替换等突变,可以通过限制性片段长度多态性(restriction fragment length polymorphism, RFLP)、单链构象多态性(single strand confirmation polymorphism, SSCP)、测序、芯片等方法进行大规模的 SNP 检测。

(五) 选择用于分析的群体

原则上任何具有表型记录和足够的 DNA 样品的群体均可用作候选基因分析,包括商业群体、地方品种自然群体、实验杂交群体等。但是一些两个近交系或品种杂交得到的子一代群体等群体中,大范围的多个相邻的基因处于连锁不平衡状态,即这些等位基因在同一个个体中出现的概率较大,因此不利于判断个体的表型变异是由所研究的候选基因引起的还是由该候选基因相邻的紧密连锁的基因引起的。

(六) 候选基因与表型性状的关联分析

根据候选基因进行分析,可以分析单个多态位点与性状的关系,也可以分析多个位点或者整个候选基因与性状的关系,后者会增加测定基因型的成本,但是能够提高统计检验的效率和特异性。是根据候选基因的单个多态位点还是根据候选基因的单倍型(haplotype)进行分析,在检测 QTL 时必须首先进行选择。进行分析时可以选用一个线性模型,将候选基因的等位基因或基因型作为固定效应,通过最小二乘分析检测候选基因与表型性状的关系,并估计候选基因的效应大小及比较不同基因型的差别。

进行关联分析时,通常结合单位点和单倍型同时进行法分析。如果由 n 个 SNP 多态位点构建单倍型,则理论上存在 2^n 种单倍型(单个 SNP 位点的突变一般包括两种核苷酸)。例如,三个多态位点 $AaBbCc$ 理论上存在 ABC 、 ABc 、 AbC 、 Abc 、 aBC 、 aBc 、 abC 和 abc 8 种单倍型,但是实际分析中一般只会存在其中的几种。因为我们所选取的分子标记通常并不是真正造成该基因功能差异的突变,单个标记与功能性突变之间通常会存在历史重组等因素而并不能通过标记完全代表功能性突变,因此通过对该基因的多个位点构建单倍型进行分析能够更准确地代表所研究的候选基因及其等位基因,提高相关性分析的效力。

(七) 转录及翻译等水平进行分析

大多数基因是通过转录成 RNA,然后翻译成蛋白质,进而行使功能的,因而除了