



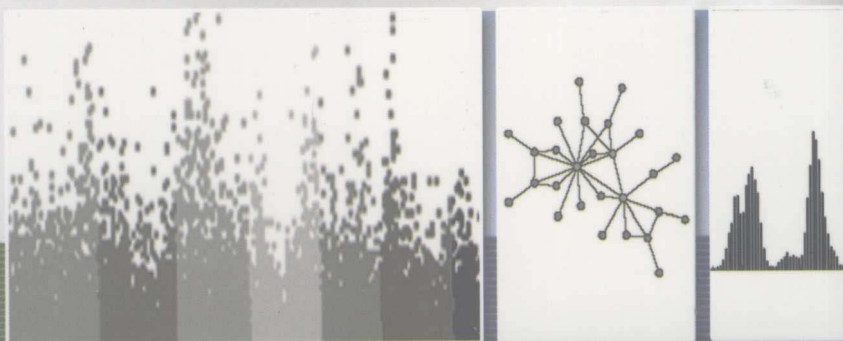
现代农业高新技术成果丛书

动物重要经济性状基因的分离与应用

Identification of Genes for Economically Important Traits
and Their Application in Animal Breeding

张 勤 主 编

张 沅 副主编
刘剑锋



中国农业大学出版社

CHINA AGRICULTURAL UNIVERSITY PRESS



现代农业高新技术成果丛书

国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

动物重要经济性状基因的 分离与应用

Identification of Genes for Economically Important Traits
and Their Application in Animal Breeding

张 勤 主 编
张 沅 副主编
刘剑锋



中国农业大学出版社
• 北京 •

内 容 简 介

本书总结了动物重要经济性状重要基因挖掘和利用方面国内外在理论、方法及应用上的研究进展,系统介绍了本研究团队取得的研究发现和重要成果。内容既涉及经典的复杂性状连锁分析、精细定位、候选基因分析、基因表达分析、标记辅助选择和生物信息学的理论与方法,又涵盖了近年来的研究热点:包括全基因组关联分析、基因组选择、基因聚合的理论与方法。同时本书阐述了上述理论方法在我国畜禽(包括奶牛、猪、鸡)分子育种和基因检测中的应用及主要结果。

本书可供高等院校、科研机构等从事动物遗传育种的科研人员和研究生参考。

图书在版编目(CIP)数据

动物重要经济性状基因的分离与应用/张勤主编. —北京:中国农业大学出版社,2012.2
ISBN 978-7-5655-0467-9

I. ①动… II. ①张… III. ①动物-经济性状-基因-研究 IV. ①Q953

中国版本图书馆 CIP 数据核字(2011)第 277943 号

书 名 动物重要经济性状基因的分离与应用

作 者 张 勤 主 编 张 沅 刘剑锋 副主编

策划编辑 宋俊果 高 欣

责任编辑 田树君

封面设计 郑 川

责任校对 陈 莹 王晓凤

出版发行 中国农业大学出版社

社 址 北京市海淀区圆明园西路 2 号

邮政编码 100193

电 话 发行部 010-62731190,2620

读者服务部 010-62732336

编辑部 010-62732617,2618

出 版 部 010-62733440

网 址 <http://www.cau.edu.cn/caup>

e-mail cbsszs @ cau.edu.cn

经 销 新华书店

印 刷 涿州市星河印刷有限公司

版 次 2012 年 2 月第 1 版 2012 年 2 月第 1 次印刷

规 格 787×1092 16 开本 24 印张 590 千字

定 价 98.00 元

图书如有质量问题本社发行部负责调换

现代农业高新技术成果丛书

编审指导委员会

主 任 石元春

副主任 傅泽田 刘 艳

委 员 (按姓氏拼音排序)

高旺盛 李 宁 刘庆昌 束怀瑞

佟建明 汪懋华 吴常信 武维华

编写人员

主 编 张 勤

副 主 编 张 沅 刘剑锋

参编人员 丁向东 孙东晓 俞 英

王志鹏 张 哲 巩元芳

卢 昕 刘 杨 姜 力

出版说明

瞄准世界农业科技前沿,围绕我国农业发展需求,努力突破关键核心技术,提升我国农业科研实力,加快现代农业发展,是胡锦涛总书记在 2009 年五四青年节视察中国农业大学时向广大农业科技工作者提出的要求。党和国家一贯高度重视农业领域科技创新和基础理论研究,特别是 863 计划和 973 计划实施以来,农业科技投入大幅增长。国家科技支撑计划、863 计划和 973 计划等主体科技计划向农业领域倾斜,极大地促进了农业科技创新发展和现代农业科技进步。

中国农业大学出版社以 973 计划、863 计划和科技支撑计划中农业领域重大研究项目成果为主体,以服务我国农业产业提升的重大需求为目标,在“国家重大出版工程”项目基础上,筛选确定了农业生物技术、良种培育、丰产栽培、疫病防治、防灾减灾、农业资源利用和农业信息化等领域 50 个重大科技创新成果,作为“现代农业高新技术成果丛书”项目申报了 2009 年度国家出版基金项目,经国家出版基金管理委员会审批立项。

国家出版基金是我国继自然科学基金、哲学社会科学基金之后设立的第三大基金项目。国家出版基金由国家设立、国家主导,资助体现国家意志、传承中华文明、促进文化繁荣、提高文化软实力的国家级重大项目;受助项目应能够发挥示范引导作用,为国家、为当代、为子孙后代创造先进文化;受助项目应能够成为站在时代前沿、弘扬民族文化、体现国家水准、传之久远的国家级精品力作。

为确保“现代农业高新技术成果丛书”编写出版质量,在教育部、农业部和中国农业大学的指导和支持下,成立了以石元春院士为主任的编审指导委员会;出版社成立了以社长为组长的项目协调组并专门设立了项目运行管理办公室。

“现代农业高新技术成果丛书”始于“十一五”,跨入“十二五”,是中国农业大学出版社“十二五”开局的献礼之作,她的立项和出版标志着我社学术出版进入了一个新的高度,各项工作迈上了新的台阶。出版社将以此为新的起点,为我国现代农业的发展,为出版文化事业的繁荣做出新的更大贡献。

中国农业大学出版社

2010 年 12 月

前 言

动物的遗传改良从根本上说,是通过提高群体中的优良基因频率来改变群体的遗传结构,提高群体遗传水平。畜禽生产中的主要经济性状大多是复杂性状,受数量不详的多基因控制,传统的育种方法是基于微效多基因模型,将影响一个性状的所有基因当作一个整体,根据表型和系谱信息估计其总的遗传效应,在此基础上进行选择。这种方法在畜禽育种中取得了极大成功,但与此同时也存在一定的局限性。20 世纪 80 年代以来分子遗传学的兴起和快速发展,使得我们可以在基因组水平对重要经济性状的分子遗传机理进行研究,这些研究主要集中在对影响性状的重要功能基因进行挖掘,了解它们的作用机理,进而应用于动物育种。

中国农业大学动物分子数量遗传与育种研究团队在过去的 10 余年时间中,在国家 973、863、自然科学基金等课题的支持下,致力于动物重要经济性状重要基因的挖掘和利用的理论、方法和应用等方面的研究,取得了一些重要的研究成果,本书在总结国内外研究进展的基础上,对这些研究成果进行了系统的介绍,希望能为同行们提供参考和借鉴。

参加本书编写的人员为本团队的骨干研究人员和部分博士研究生,其中张勤撰写了第 1 章和第 7 章,刘剑锋撰写了第 2 章和第 3 章,丁向东撰写了第 4 章和第 5 章,孙东晓撰写了第 12 章,俞英撰写了第 13 章,王志鹏撰写了第 6 章,张哲撰写了第 8 章,巩元芳和卢昕撰写了第 9 章,刘杨撰写了第 10 章,姜力撰写了第 11 章。

动物的重要经济性状数量庞大而复杂,全面挖掘影响这些性状的重要基因是一项长期而又艰苦的工作,虽然我们做了大量工作,但也仅仅涉及其中的很小一部分,与国内外同行相比,我们在很多方面还存在差距。我们将继续努力,与国内外同行一道为实现我们共同的目标而奋力前行。

限于我们的水平和能力,书中错误和缺陷在所难免,敬请读者批评指正,不吝赐教。

编 者

2011.12

目 录

第 1 章 绪论	1
参考文献	5
第 2 章 QTL 连锁分析理论和方法	7
2.1 QTL 连锁分析概述	7
2.2 Bayes 统计推断的基本原理	8
2.3 基于 RJ-MCMC 进行畜禽远交群体 QTL 连锁分析	10
2.3.1 遗传模型	10
2.3.2 标记基因型和性状表型的数据模拟	12
2.3.3 QTL 的 IBD 矩阵推断方法	12
2.3.4 基于 RJ - MCMC 的 QTL 连锁分析	14
2.3.5 连续性状 QTL 连锁分析	18
2.3.6 二级阈性状 QTL 连锁分析	22
2.3.7 主要结论	38
2.4 基于贝叶斯压缩(Bayes shrinkage)技术的 QTL 连锁分析	41
2.4.1 基于 Bayes 压缩技术的连锁分析模型	41
2.4.2 全条件后验分布推导和 MCMC 抽样	42
2.4.3 基于 MCMC 参数估计值的统计检验	44
2.4.4 贝叶斯压缩方法的模拟验证	44
参考文献	47
第 3 章 QTL 精细定位理论与方法	50
3.1 基于 IBD 方法的基因精细定位	51
3.1.1 IBD 定位方法基本原理	52
3.1.2 IBD 精细定位的分析方法	52
3.1.3 IBD 精细定位的模拟研究	54

3.2 结合连锁分析(LA)和连锁不平衡分析(LA/LD)的 QTL 精细定位策略 64

3.2.1 LD/LA 定位的理论基础和方法 65

3.2.2 基于模拟试验进行 LA 和 LA/LD 分析的比较研究 72

3.2.3 主要结论..... 82

参考文献 84

第 4 章 基于传递不平衡检验的关联分析方法 86

4.1 利用系谱传递不平衡检验进行阈值性状 QTL 定位 88

4.1.1 PTDT 方法..... 89

4.1.2 PTDT 的检验功效和 I 型错误..... 91

4.1.3 影响 PTDT 检验效力的因素 95

4.2 利用系谱传递不平衡检验进行数量性状 QTL 定位 100

4.2.1 数量性状转化为分类性状方法 100

4.2.2 选择性基因型测定 101

4.2.3 数据模拟 101

4.2.4 数量性状转化方法比较 101

4.2.5 PTDT 与 QTDT 和 PDT 效力比较 103

4.2.6 PTDT 选择性基因型测定基因定位效力 103

参考文献..... 104

第 5 章 单倍型推断..... 106

5.1 基本概念 106

5.1.1 单倍型(haplotype)、双倍型(diplotype) 106

5.1.2 单倍型推断 107

5.1.3 大片段染色体单倍型推断 107

5.2 利用单亲资料推断单倍型——SPFHAP 法 108

5.2.1 方法概述 108

5.2.2 单倍型推断效率 110

5.3 利用全同胞信息推断单倍型——FSHAP 法 111

5.3.1 方法概述 111

5.3.2 单倍型推断效率 113

5.4 复杂家系单倍型推断 115

5.4.1 零重组单倍型推断方法(ZRHI) 116

5.4.2 有重组单倍型推断方法——MRHI 122

5.4.3 讨论 127

参考文献..... 128

第 6 章 生物信息学在重要经济性状的基因分离中应用..... 132

6.1 牛全基因组转录因子数据库构建 134

6.1.1 牛全基因组预测转录因子识别原理 135

6.1.2 数据与方法 135

6.1.3	牛全基因组预测转录因子数据库的构架以及应用	135
6.1.4	数据库特征分析	138
6.2	后生动物转录因子 ETS 基因家族的分类与进化分析	141
6.2.1	数据来源和分析方法	142
6.2.2	分析结果	143
6.2.3	讨论	149
6.3	基因芯片的数据分析	151
6.3.1	微阵列数据的数学模型	152
6.3.2	非对数转换方法	153
6.3.3	检测差异表达的基因	155
6.3.4	cDNA 微阵列数据的模拟	156
6.3.5	模拟数据研究结果	158
6.3.6	模拟研究结果分析	162
6.3.7	实际数据分析应用	164
	参考文献	166
第 7 章	动物分子标记辅助育种技术	173
7.1	标记辅助选择	173
7.1.1	MA-BLUP	173
7.1.2	两阶段选择	177
7.2	标记辅助导入	181
7.3	标记辅助基因聚合	187
7.3.1	基因聚合的基本思路	188
7.3.2	动物群体的基因聚合方案	188
7.3.3	基因聚合所需群体规模的计算	189
7.3.4	不同基因聚合方案的比较	195
7.3.5	影响基因聚合方案效率的因素	196
7.3.6	基因聚合的果蝇模拟试验	199
	参考文献	200
第 8 章	基因组选择技术	202
8.1	基因组选择技术简介	202
8.1.1	基因组选择的基本原理	202
8.1.2	基因组育种值的计算方法	203
8.1.3	基因组选择的准确性及其影响因素	206
8.1.4	基因组选择的应用	208
8.2	利用性状特异关系矩阵的最佳线性无偏预测法 (TABLUP)	211
8.2.1	TABLUP 法的基本方法	211
8.2.2	TABLUP 法的性能	212

8.3 利用低密度标记的基因组选择	214
8.3.1 高密度芯片下 GEBV 的准确性	215
8.3.2 标准参数下低密度芯片的 GEBV 准确性	216
8.3.3 遗传力对低密度标记 GEBV 准确性的影响	217
8.3.4 有效群体大小对低密度标记 GEBV 准确性的影响	217
8.3.5 筛选标记数和芯片密度的关系	218
8.3.6 低密度标记基因组选择的相关问题	219
参考文献	220
第9章 猪免疫性状基因定位	225
9.1 猪各种免疫性状的生物学意义	226
9.1.1 血液指标	226
9.1.2 细胞因子	227
9.1.3 T 淋巴细胞亚群	228
9.1.4 IgG 与溶菌酶	229
9.2 猪免疫(抗病)性状相关 QTL 和候选基因的研究现状	230
9.2.1 猪免疫(抗病)性状相关 QTL 的研究现状	230
9.2.2 猪免疫(抗病)性状相关候选基因的研究现状	232
9.3 我国猪群中免疫性状相关的 QTL 检测	234
9.3.1 试验猪群及免疫指标测定	234
9.3.2 微卫星标记的选择及 QTL 连锁分析	235
9.3.3 影响猪血常规指标 QTL 的定位结果	236
9.3.4 影响猪细胞因子 QTL 的定位结果	240
9.4 我国猪群中猪免疫性状的候选基因分析	242
9.4.1 猪 CD14 基因的克隆和组织表达	243
9.4.2 TLR4 的组织表达	244
9.4.3 LMP2 和 MECL-1 基因的克隆和组织表达	245
9.4.4 猪 LMP2、LMP7 基因与免疫性状的关联分析	249
参考文献	252
第10章 仔猪腹泻抗性基因分离与鉴定	263
10.1 致仔猪腹泻大肠杆菌简介	263
10.1.1 大肠杆菌疾病分类及主要特征	263
10.1.2 大肠杆菌的致病机理	264
10.1.3 ETEC 的血清分型	264
10.1.4 产肠毒素的毒力因子及其致病性	265
10.2 仔猪小肠上的大肠杆菌受体	266
10.2.1 受体的作用	266
10.2.2 受体的性质	266
10.2.3 肠道不同部位、日龄等对受体的影响	267

10.2.4	仔猪产肠毒素 F4ab/ac 受体的遗传特性	267
10.2.5	受体的检测以及影响因素	267
10.3	F4 受体候选基因的筛选	268
10.3.1	DDRT-PCR 技术筛选仔 F4 受体候选基因	268
10.3.2	抑制性消减杂交(SSH)技术筛选仔猪腹泻抗性候选基因	269
10.4	仔猪腹泻大肠杆菌 F4ab/ac 受体候选基因分析	271
10.4.1	转铁蛋白(Tf)	271
10.4.2	肿瘤相关钙离子信号转导因子 1(TACSTD1)	272
10.4.3	黏液素 4 基因(MUC4)	272
10.5	F4ab/ac 受体基因的精细定位	275
	参考文献	278
第 11 章	奶牛产奶性状全基因组关联分析	283
11.1	资源群体	284
11.2	SNP 基因型测定及质量控制	285
11.3	全基因组关联分析统计方法	286
11.3.1	关联分析的统计模型	286
11.3.2	对多重检验的校正	287
11.3.3	群体分层检验	289
11.4	与牛产奶性状显著关联的 SNPs	289
11.4.1	产奶量	291
11.4.2	乳脂量	292
11.4.3	乳蛋白量	293
11.4.4	乳脂率	294
11.4.5	乳蛋白率	296
11.5	主要结论	297
	参考文献	300
第 12 章	鸡生长和产蛋性状候选基因分析	305
12.1	鸡重要经济性状分子遗传机理研究现状	305
12.1.1	鸡的基因组图谱	305
12.1.2	鸡重要经济性状 QTL 定位研究进展	307
12.1.3	鸡重要经济性状的候选基因	308
12.1.4	北京油鸡生长和产蛋性状候选基因分析	310
12.2	TGFB2 和 IGF2R 基因的遗传效应分析	312
12.2.1	SNP 筛选和测定	313
12.2.2	SNP 与性状的关联分析	314
12.2.3	TGFB2 基因 SNP 的转录因子结合位点分析	319
12.3	IGFBP1、IGFBP3 和 STAT5B 基因遗传效应分析	320
12.3.1	SNP 筛选和测定	321

12.3.2	基因和基因型频率以及 Hardy - Weinberg 平衡检验	322
12.3.3	SNP 与生长和产蛋性状的关联分析	323
12.3.4	蛋白质结构预测.....	327
12.3.5	转录因子结合位点分析.....	327
12.4	GHRHR 和 JAK2 基因遗传效应分析	330
12.4.1	SNP 筛选和测定	330
12.4.2	SNP 与生长和产蛋性状关联分析	331
	参考文献.....	334
第 13 章	动物杂种优势分子遗传机理研究	341
13.1	试验鸡群组建.....	341
13.2	试验鸡群主要屠体性状及产蛋性状的杂种优势率.....	342
13.2.1	主要屠体性状的杂种优势.....	342
13.2.2	主要产蛋性状的杂种优势.....	343
13.2.3	主要蛋品质性状的杂种优势率.....	344
13.3	心肌和肝脏组织基因差异表达与鸡屠体性状杂种优势的关系.....	346
13.3.1	基因差异表达类型与鸡屠体性状杂种优势的关系.....	346
13.3.2	差异表达基因的表达量与屠体性状杂种优势的关系.....	351
13.3.3	主要研究结论.....	355
13.4	卵巢组织基因差异表达与鸡产蛋性状和蛋品质性状杂种优势的关系.....	356
13.4.1	基因差异表达模式与鸡产蛋性状和蛋品质性状杂种优势的关系.....	356
13.4.2	差异表达基因的表达量与产蛋性状和蛋品质性状杂种优势的关系.....	360
13.5	应用荧光定量差异显示技术探索鸡杂种优势的分子机理.....	362
13.5.1	利用持家基因的荧光定量分析标定 cDNA 池的浓度	362
13.5.2	促卵泡素受体基因(FSHR)的荧光定量检测结果	363
13.5.3	雌激素受体(ER)基因的荧光定量检测结果	365
13.5.4	主要研究结论.....	366
	参考文献.....	367

绪 论

动物育种主要有两个方面的工作,一是通过人工的定向选择,使现有的动物品种在主要生产性能上的遗传水平得到不断的提高,从而实现群体的遗传改良。二是通过杂交等手段将不同群体的优良基因整合,生产符合市场需求的产品或培育新的动物品种或品系。

选择的理论和方法历来是动物遗传育种工作者的研究重点。根据数量遗传学的理论,通过选择所获得的群体遗传进展的速率与选择强度、群体的遗传变异性和选择的准确性成正相关,与世代间隔成反相关,因而对选择理论和方法的研究也主要围绕着这 4 个方面进行,而其中的核心又是选择的准确性。选择的基础是个体遗传评估,因此遗传评估方法成为研究的重点。根据数量遗传学的微效多基因理论,数量性状的表现是由很多基因以及环境的共同作用所决定的,由于基因的数目很多,而每个基因的作用一般都较小,再加上环境因素的干扰,我们不能对每个基因单独地进行分析,而只能将所有的基因作为一个整体来考虑,实际上就是将基因的作用当做一个“黑匣子”。因此,遗传评估就是对所有基因的整体效应进行估计。遗传评估方法在过去的 50 余年中得到了很大的发展,尤其是自 20 世纪 70 年代后期以来,由 Henderson(1974)提出的最佳线性无偏估计(best linear unbiased prediction, BLUP)理论和方法在全世界范围内得到了广泛的应用,成为动物个体遗传评估的通用和标准方法。传统的遗传评估(包括 BLUP 方法)是根据个体的性状表型信息和系谱信息来估计个体育种值,BLUP(尤其是动物模型 BLUP)方法为我们提供了利用这种信息的最佳手段,使得动物个体育种值估计的准确性得到了很大提高,因而对于多数动物重要经济性状来说,基于 BLUP 的选择取得了显著的遗传改良效果。例如,美国荷斯坦奶牛平均 305 天的产奶量在过去的近 50 年中由约 6 000 kg 增加到约 12 000 kg,提高了近 1 倍,而其中由于群体遗传改良所带来的遗传进展约为 4 000 kg(图 1-1)。再如,加拿大长白猪、大白猪和杜洛克 3 个主要品种猪达 100 kg 体重日龄在过去的近 30 年中的累计遗传进展达到了近 30 天(图 1-2)。

尽管传统的选择方法在动物育种实践中取得了巨大成就,但它并不完美,在有的情况下这种选择不能取得理想的效果。例如对于低遗传力的性状(如繁殖性状)和阈性状(如抗病性),

由于在表型信息中所包含的遗传信息很有限,除非有大量的各类亲属的信息,很难对个体做出准确的遗传评定。再如对于限性性状(如产奶、繁殖性状),对不能表达性状的个体一般只能根据其同胞和后裔的成绩对其进行评定,如果仅利用同胞信息,则由于同胞数有限,评定的准确性一般较低,如果利用后裔信息,而且后裔数很多(如在奶牛中的情形),评定的准确性可以达到很高,但世代间隔拖长,每年的遗传进展相对降低,且育种成本很高。还有对于胴体性状,因为必须屠宰后才能测定,一般也只能进行同胞或后裔测定,而且由于性状测定的难度和费用都很高,测定的规模受到限制,使得评定的准确性和世代间隔都受到影响。

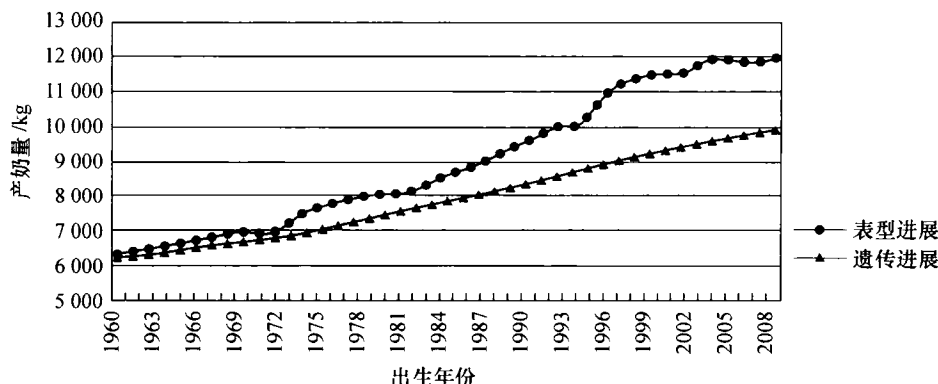


图 1-1 美国荷斯坦奶牛平均产奶量的表型和遗传进展

(数据来自 <http://aipl.arsusda.gov/>)

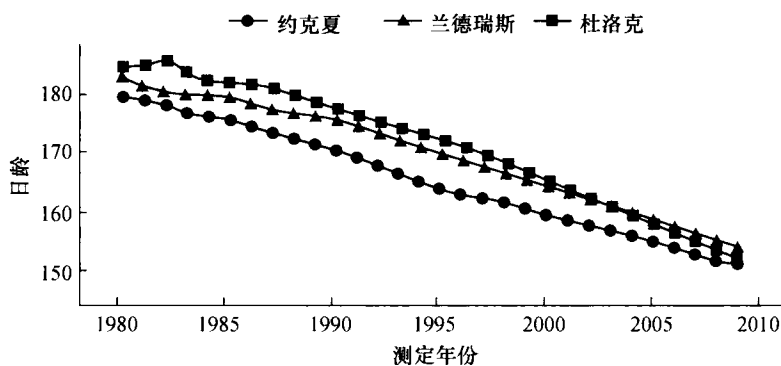


图 1-2 加拿大主要猪品种达 100 kg 体重日龄的遗传进展

(引自 <http://www.ccsi.ca/>)

在动物新品种培育方面,在过去的几十年中,并没有取得显著的成绩。由于长期以来对产量的片面追求,世界各国的肉、蛋、奶、毛的生产越来越依赖少数几个品种,而这些品种基本上都是在 100 年前育成的。经过长期的选育,这些品种的主要生产性能有了很大提高,在很大程度上满足了畜产品市场的需求,但也使得畜产品生产的遗传基础越来越窄。在过去的半个多世纪中,世界各国也陆续地培育了一些新的畜禽品种,这些品种虽然在某些方面具有一定优良特性,但大多总体经济价值不高,未能成为广泛使用的主流品种。目前消费者对畜产品质量(如营养成分与口味)和安全性(如病原微生物含量与药物残留量)要求越来越高,同时生产者也希望通过提高动物的繁殖力、抗病性和长寿性等来进一步提高生产效率,而这些需求仅仅依

靠目前的少数几个品种是很难满足的。一方面是因为这些性状用目前的选择方法很难实现遗传改良,二是在这些品种中缺少一些相关基因。虽然目前在畜牧生产中广泛使用的品种只有少数几个,但世界上的畜禽地方品种资源是相当丰富的,据联合国粮农组织(FAO)统计(FAO,2007),至2006年,世界上有记录的哺乳类地方家畜品种共计4 068个(其中牛品种 897 个,山羊品种 512 个,绵羊品种 995 个,猪品种 541 个),地方家禽品种共1 644个(其中鸡品种 1 077 个,鸭品种 186 个,鹅品种 158 个)。这些品种中蕴藏着丰富的基因资源,而其中必然有我们目前需要的基因。例如,我国的很多地方畜禽品种在肉(蛋)品质、繁殖力、抗病(抗逆)性等方面具有显著优势。如何挖掘并有效利用这些资源是摆在动物育种工作者面前的一大难题。

不可否认的是,数量遗传学的多基因理论并没有完全揭示数量性状的遗传本质。自20世纪70年代以来,人们在经过长期选择的群体中陆续发现了一些对数量性状有较大作用且仍然处于分离状态的基因,例如影响鸡的体型大小的矮小(dwarf)基因(Merat and Ricard, 1974),影响猪的瘦肉率和肉质的氟烷(Halothane)基因(Smith and Bampton, 1977; Webb *et al.*, 1982,它同时也是造成猪应激敏感的主要基因),影响肉牛的臀部肌肉丰满程度的双肌(double muscling)基因(Rollins *et al.*, 1972),以及影响羊的产羔数的Booroola基因(Piper and Bindon, 1982)。人们将这些基因称为主效基因(major gene)。在这里主效是相对于微效而言的,也就是说它们并不能决定数量性状的所有变异。这些基因的发现大多是偶然的,但它们的发现使得人们对数量性状的遗传机制有了新的认识,可以想象在其他数量性状中也完全可能存在类似的主效基因,如何主动地去发掘这些基因就成为数量遗传学和家畜育种学的新的研究课题。在80年代中期以前,人们主要借助一些统计学方法来检测主效基因,例如,多众数分布(multimodal distribution)检验法,非正态分布(non-normal distribution)检验法,杂合方差(heterogeneity of variance)检验法,亲子相似性(offspring-parent resemblance)检验法,复合分离分析(complex segregation analysis)法。遗憾的是这些方法的检测效率都较低,而且一般只能检测出是否有主效基因存在,而不能告诉我们主效基因在染色体上的位置。

自20世纪80年代以来,随着分子遗传学和分子生物技术的飞速发展,大量的以分子遗传标记为形式的DNA水平的分子遗传信息被发现,其中包括限制性酶切片段长度多态性(restriction fragment length polymorphism, RFLP)、扩增片段长度多态性(amplified fragment length polymorphism, AFLP)、随机扩增多态性DNA(randomly amplified polymorphic DNA, RAPD)、微卫星(minisatelite)以及单核苷酸多态(single nucleotide polymorphisms, SNP)。利用这些标记,我们可以真正从DNA水平上对影响数量性状的单个基因或染色体片段进行分析,人们将含有这些单个的基因或染色体片段的区域称为数量性状基因座[quantitative trait locus (loci), QTL]。从广义上说,数量性状基因座是所有影响数量性状的基因座(不论效应大小),但通常人们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,而将那些不能检测出的基因仍当做微效多基因来对待。当一个QTL就是一个单个基因时,它就是主效基因。

在过去的近30年中,对QTL及主效基因的检测主要有2种策略,一是利用遗传标记通过连锁分析或连锁不平衡分析来进行QTL检测和定位,二是候选基因分析,即通过对某个候选基因与性状表型的关联分析,来判断该基因是否与某个性状有关。在QTL检测和定位方面,世界各国的动物遗传育种工作者做了大量工作,发掘了大量的QTL,在国际学术刊物上报道

的 QTL 被汇集到了一个 QTL 数据库中(Hu,2009)。截止到 2011 年 3 月,该数据库中已收集的 QTL 情况见表 1-1。

表 1-1 家畜(禽)中已报道的 QTL
(数据引自 <http://www.animalgenome.org/QTLdb/>)

物种	QTL 总数	论文数	涉及性状数	QTL 数量最多的 5 个性状
猪	6 344	281	593	滴水损失(945),平均背膘厚(158),眼肌面积(126),胴体长(122),日增重(82)
牛	4 682	274	376	乳蛋白率(201),产奶量(175),剩余采食量(159),胴体重(127),体细胞评分(122)
鸡	2 451	125	248	腹脂重(155),体重(130),马立克病相关性状(115),42 日龄重(122),胸肌重(90)
羊	348	47	137	Haemonchus Contortus FEC2(18),骨密度(14),Haemonchus Contortus FEC1(13),Ultrasound Fat Depth(13),睾丸重(8)

要说明的是,①这些 QTL 中,有些是重复的,这是因为由不同研究小组对同一性状进行分析,在相同或相近的染色体区域发现的 QTL,都被分别收集到数据库中,而不考虑它们是否是相同的 QTL;②并不是每个报道的 QTL 都一定是真实的 QTL,由于 QTL 存在与否是通过统计学检验而定的,因而存在犯错误的可能,也就是说有些 QTL 可能是假的;③在这些 QTL 中,绝大部分 QTL 是可能含有主效基因的染色体区域,而且这些区域的置信区间大多较大(~ 20 cM),也就是说,对这些 QTL 来说,我们仅知道在某些染色体区域内可能存在主效基因,并知道有哪些分子标记是与之连锁的,但基因本身却是未知的。

无论是 QTL 定位还是候选基因分析,最终目的是要找到影响性状的主效基因,但寻找主效基因是一项十分艰巨的工作。直接的候选基因分析虽然简单易行,但效率很低,因为可以作为候选基因的基因数量太大,很难从中筛选合适的候选基因。QTL 定位的结果通常只能得到一个 QTL 区域,基因本身却是未知的。比较合理的做法是,先用连锁分析进行 QTL 定位(置信区间为 $10\sim 20$ cM),而后用连锁不平衡分析进行 QTL 精细定位(置信区间为 $1\sim 2$ cM),再从精细定位的区间中筛选候选基因进行关联分析。而最后确定一个基因是否为主效基因,则一方面要在多个群体中进行反复的关联分析验证,另一方面还要对其进行生物学功能的验证。这一过程需要相当长的时间,并要花费大量的人力和财力。虽然在文献中报道的候选基因很多,但经过验证确认为主效基因的非常少。除了前面提到的氟烷基因、双肌基因、Booroola 基因和矮小基因外,近年来确认的主效基因主要有影响奶牛乳脂率的 DGAT1 基因(Grisart *et al.*, 2002)、影响猪肌肉生长的 IGF2 基因(van Laere *et al.*, 2003)、影响羊肌肉发育的 MSTN 基因(Clop *et al.*, 2006)。近年来,随着全基因组高密度 SNP 芯片的问世,诞生了一种新的分析策略,即全基因组关联分析(genome-wide association study, GWAS)。目前对于主要的家畜(禽)物种,如牛、猪、鸡、羊、马、犬,都已有了 SNP 芯片,在这些芯片上含有 $50\,000\sim 60\,000$ 个 SNP,SNP 间的平均距离为 $40\sim 50$ kb。利用这些芯片可以进行高通量、大规模的 SNP 标记分型。GWAS 就是通过对每个 SNP 与性状的关联分析,来寻找与性状高度关联的 SNP,这些 SNP 可能就是某个主效基因中的功能突变,也可能是与功能突变处于高度