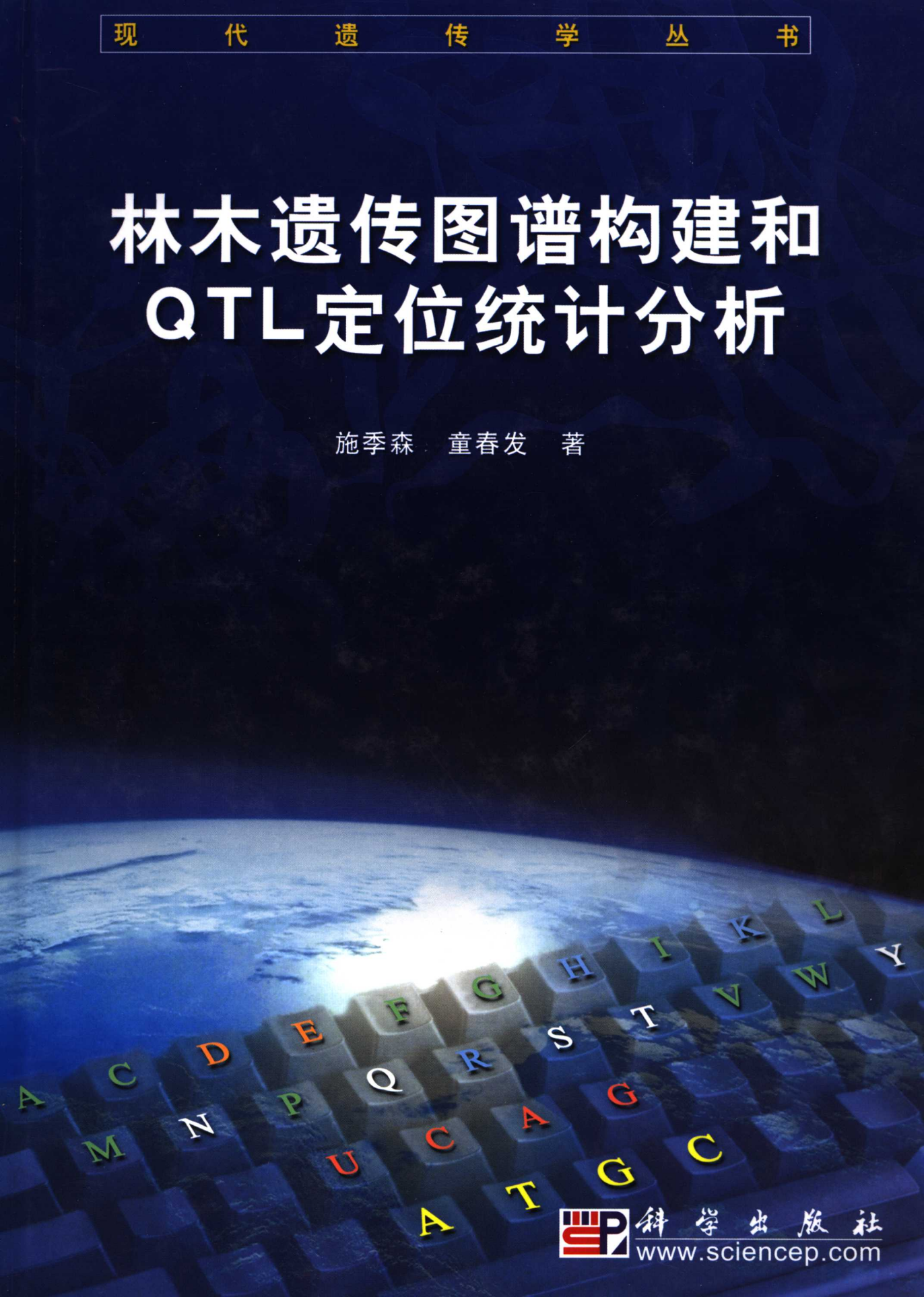
现 代 遗 传 学 丛 书

林木遗传图谱构建和QTL定位统计分析

施季森 童春发 著



科学出版社
www.sciencep.com

现代遗传学丛书

林木遗传图谱构建和QTL定位统计分析

施季森　童春发　著

科学出版社

北京

内 容 简 介

本书包括绪论、上篇和下篇，首次较系统、详细、全面地整理和介绍了近交群体（如回交群体、F_2 代群体等）的遗传图谱的构建和 QTL 定位的统计分析模型以及相关的数学方法。同时，针对目前林木高密度分子遗传图谱的构建和 QTL 精确定位的需要，着重研究了全同胞群体遗传图谱的构建和 QTL 定位的统计模型。本书紧扣遗传图谱构建和 QTL 定位的前沿研究展开，结构严密，理论性强，在内容安排上考虑到了读者使用的实用性和方便性。

本书适合从事动植物育种的科技工作者、相关专业的本科高年级学生和研究生阅读参考。

图书在版编目(CIP)数据

林木遗传图谱构建和 QTL 定位统计分析/施季森，童春发 著. —北京：科学出版社，2006

（现代遗传学丛书）

ISBN 7-03-016366-4

Ⅰ. 林… Ⅱ. ①施… ②童… Ⅲ. 树木学：遗传学 Ⅳ. ST18.46

中国版本图书馆 CIP 数据核字(2005)第 120122 号

责任编辑：莫结胜 卜 新 / 责任校对：陈丽珠

责任印制：钱玉芬 / 封面设计：王 浩

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

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2006 年 1 月第 一 版 开本：787×1092 1/16

2006 年 1 月第一次印刷 印张：10 3/4

印数：1—2 000 字数：255 000

定价：39.00 元

（如有印装质量问题，我社负责调换〈科印〉）

前　言

随着分子生物技术的迅速发展，人们能够获得多种多样的 DNA 分子标记，利用这些分子标记，已经构建了多种生物的分子遗传图谱。遗传图谱是系统进行基因组研究的基础，是在分子基础上进行动植物育种和人类遗传疾病诊断的依据。十多年来，林木遗传图谱的构建和数量性状基因座定位的研究进展非常迅速。但是，目前林木遗传图谱构建和数量性状基因座定位的统计理论方法都是套用为近交群体而建立的作图理论，这使得林木遗传图谱构建和数量性状基因座定位存在着诸如遗传图谱比较稀疏、数量性状基因座定位不太精确等许多问题。林木具有很多自身特有的复杂生物学特性，如生长周期漫长、异花授粉、具有很高的遗传杂合度等，因此必须研究和发展适合林木的遗传图谱构建和数量性状基因座定位的统计分析理论，本书正是基于这一点而产生的。本书取材于作者所在课题组的最新研究成果和国内外期刊上最近发表的文献。

全书分上、下两篇，共 8 章，首次较系统、详细、全面地整理和介绍了近交群体（如回交群体、F_2 代群体等）的遗传图谱的构建和数量性状基因座定位的统计分析模型以及相关的数学方法。同时，针对目前林木高密度分子遗传图谱的构建和数量性状基因座精确定位的需要，着重研究了全同胞群体遗传图谱的构建和数量性状基因座定位的统计模型。

上篇主要介绍和研究遗传连锁图谱构建的统计分析方法，包括不同作图群体的两点连锁分析、连锁群的划分、标记位点的排序和多位点的连锁分析等内容，提出了基因位点排序的启发式搜索法和全同胞群体的多位点连锁分析的隐马尔可夫链方法，为利用林木的 F_1 代构建高密度遗传图谱奠定了理论基础。最后，修订了杉木的遗传连锁图谱，结果比用拟测交法增加了 54 个标记。

下篇主要介绍了近交群体数量性状基因座定位的单标记分析、区间作图、复合区间作图等方法，同时还研究了异交群体的数量性状基因座定位方法。针对目前林木上所用的 F_1 代作图群体，考虑 1:1 分离位点的连锁相信息，提出了林木 F_1 代群体数量性状基因座区间作图和复合区间作图法，并利用新构建的杉木 AFLP 分子标记遗传连锁图谱对杉木的 16 个数量性状进行了数量性状基因座定位。

本书虽然以林木的 F_1 代群体为背景论述遗传图谱的构建和数量性状基因座定位的统计方法，但是也包含了用近交系所产生的作图群体进行图谱的构建和数量性状基因座定位的统计方法。同时，本书的统计分析方法也适用于其他异交植物的全同胞群体的遗传图谱的构建和数量性状基因座定位。

我们编写了全同胞遗传连锁图谱构建的应用软件 FsLinkageMap 1.0，附录 A 中给出了此软件的使用说明。

本书成稿过程中，得到王明庥院士、黄敏仁教授等许多师长的支持和帮助；盖钧镒院士、莫惠栋教授、章元明教授审阅了初稿全文，提出了许多重要建议；科学出版社莫结胜、卜新等编辑对本书的出版审校付出了大量辛勤的劳动；本书使用国家“973”林木育种分子基础研究第四课题、江苏省“六大人才高峰”资助课题、江苏省自然基金“杉木、杨树功能基

因组”课题、江苏省基础研究计划项目(BK2003098)的部分资金。在此,一并表示诚挚的谢意。

该领域尚处于起步和发展阶段,新发现、新思想和新问题并存,加之著者学术水平有限,敬请读者批评指正。

著　者

二〇〇五年十一月于南京

目　录

下篇 QTL 定位的统计方法

绪　论

遗传连锁分析是指将染色体上的基因位点进行排序并估计它们之间的遗传距离，而遗传距离则是依据统计规律确定的(Ott, 1999)。20 世纪初，Bateson 和 Punnet 在豌豆试验中发现了基因的连锁现象。后来，Morgan(1911)在果蝇试验中也发现了这一现象，Sturtevant 于 1913 年发表了第一篇有关基因图谱的论文。

自从 1980 年以来，分子遗传学和相关的技术已经变得越来越重要。1990 年 10 月美国启动了令人瞩目的人类基因组计划(HGP)，随后对生物基因图谱、基因组测序的研究成为了当代遗传学研究的热点之一。2000 年 6 月 26 日，美、英、日、法、德、中同时宣布人类基因组工作框架图构建完成，随后于 2001 年 2 月 12 日首次公布了人类基因组图谱及初步分析结果。2003 年初，HGP 宣布已完成了人类结构基因组的研究。除了人类基因组测序已完成之外，国际上还先后完成了大肠杆菌(Blattner et al., 1997)、酵母(Cherry et al., 1998)、果蝇(Adams et al., 2000)、秀丽线虫(Hodgkin et al., 1998)、小鼠(Copeland, 1993)和拟南芥(Wambutt, 2000)等生物的测序工作。2002 年 4 月，中国科学家宣告水稻全基因组测序工作已经完成，标志着我国水稻基因组研究已处于世界前列(Yu et al., 2002)。

重组 DNA 技术的出现使得几乎所有的生物学学科都被带入了分子生物学的大门，在遗传学领域最重大的进展之一就是利用 DNA 分子水平上的变异作为遗传标记进行遗传作图。1980 年，Botstein 首先提出了利用 RFLP 标记构建遗传图谱的设想。1987 年，Donis-Keller 等发表了第一张人类的 RFLP 连锁图，其饱和度远远超过了经典的图谱，之后许多生物的 RFLP 图谱相继问世。随着新的 DNA 标记技术的发展，很多生物都已构建了遗传图谱，图谱上标记的密度也越来越高。在植物中，已建立的遗传图谱有大麦(Ramsay et al., 2000)、小麦(Marion et al., 1998)、燕麦(Yu et al., 2000)、水稻(Harushimay et al., 1998)、玉米(Wilson et al., 1999)、大豆(Matthews et al., 2001)、番茄和马铃薯(Tanksley et al., 1992)等。在动物中，已建立的遗传图谱有小鼠(Blake et al., 2000)和家鼠(Steen et al., 1999)，还有猪、马、牛、羊、鸡等家畜和家禽，甚至还有家蚕(Tan et al., 2001)。最新的人类高密度遗传图谱是 Kong 等(2002)建立的。值得一提的是，过去被公认为难以开展遗传作图的木本植物，如松类、杨树、杉木、苹果、桃、茶树、桑树等，如今也涌现出了许多密度可观的分子标记连锁图。利用遗传图谱和分子标记可以寻找控制有关性状的基因，这在人类疾病研究中已获得了显著的成就。寻找控制数量性状基因座(quantitative trait locus, QTL)虽然比较困难，但是 20 多年来也进行了大量研究。

0.1　林木遗传图谱构建研究的现状

林木遗传连锁图谱构建研究最早从火炬松开始，使用的是同工酶标记，但只对部分基因组进行了作图。直到出现了以 DNA 为基础的分子标记，如 RFLP 和 RAPD，构建全部

基因组的遗传图谱成为可能。林木由于受自身生物学特性的限制，在利用分子标记构建遗传图谱方面滞后于农作物 10 多年，但林木遗传图谱研究进展却非常迅速，仅在几年时间内，就已在 30 多个树种中进行了图谱的构建工作(表 0.1)。其中，以利用 RAPD 标记技术构图的居多，占 90 %左右。主要针叶树种有湿地松、土耳其松、糖松、长叶松、海岸松、欧洲赤松、北美乔松、长叶松、欧洲云杉、白云杉、短叶红豆杉、花旗松等，阔叶树种主要有巨桉、尾叶桉等，而有一些树种如火炬松、辐射松、光亮桉、毛果杨×美洲黑杨、欧洲山杨等是利用多种标记联合构建图谱的。在现有林木遗传图谱中，具有标记数量最多的图谱是毛果杨×美洲黑杨，标记达 343 个；最少的是短叶红豆杉，仅有 41 个。图谱总图距最长的是欧洲云杉，达 3584cM；最短的仍是短叶红豆杉，达 305.8cM。随着林木遗传图谱构建研究的不断深入，现已开始利用图谱对控制重要质量性状和数量性状的基因进行定位和分离研究，进行分子标记辅助选择育种，这些将使林木遗传育种研究进入一个崭新阶段(尹佟明等，2000)。

表 0.1　林木遗传图谱构建现状一览表

Table 0.1　List of current genetic maps in forestry trees

树名	作图群体	遗传标记	作图状况	文献
白云杉 (*Picea glauca*)	单倍体胚乳	RAPD	共产生 61 个标记，其中 47 个被分配到 12 个连锁群中，覆盖基因组 873.8 cM	Lomas et al.，1992
板栗 (*Castanea olisima*)	回交二代体	形态标记、同工酶、RFLP、RAPD	共得到 249 个多态性标记，有 197 个分配到 12 个连锁群上，其余标记未见连锁，覆盖基因组 660.9 cM	Kubisiak et al.，1996
巨尾桉 (*Eucalyptus grandis* ×*E. urophylla*)	拟测交群体	RAPD	巨桉有 240 个标记分配到 14 个连锁群中(1552 cM)；尾叶桉：有 251 个标记分配到 11 个连锁群中(1101 cM)	Grattapaglia et al.，1994
火炬松 (*Pinus taeda*)	4 株亲本、2 株 F_1 代及 95 株 F_2 子代	同工酶、RFLP	共产生 90 个 RFLP 标记位点，有 73 个 RFLP 标记和 2 个同工酶标记分配到 20 个连锁群中，覆盖基因组 632 cM	Devey et al.，1994
响叶杨 (*Populus adenopoda* M.)×银白杨 (*Populus alba* L.)	80 株 F_1 代	RAPD	响叶杨图谱上有 82 个标记，分配在 19 个连锁群中，总遗传跨度为 553 cM；银白杨图谱上有 197 个标记，遗传跨度为 2300 cM，覆盖基因组 87%	Yin et al.，2001；尹佟明等，1999
美洲黑杨(*Populus deltoids*)×青杨 (*Populus cathyana*)	F_2 群体	RAPD	产生 20 个连锁群，总图距为 1899.4 cM，覆盖基因组总长度 70.35%，平均间距为 17.27 cM	苏晓华等，1998
毛果杨 (*Populus trichoccarpa*)×美洲黑杨 (*Populus deltoids*)	F_2 群体	RFLP、STS、RAPD	产生 19 个连锁群，覆盖长度为 1527.3 cM，约占基因组总长度的一半，平均作图距离为 6.7 cM	Bradshaw et al.，1995

续表

树名	作图群体	遗传标记	作图状况	文献
橡胶 (*Hevae brasiliensis*)	F_2 子代	同工酶、RFLP、RAPD	共产生 111 个标记位点(4 个同工酶、92 个 RFLP 和 15 个 RAPD),分配到 23 个连锁群中,有 20 个标记未见连锁	Seguin et al., 1995
银杏 (*Ginkgo biloba*)	单倍体胚乳	RAPD	共得到 91 个标记,其中 62 个被分配到 20 个连锁群中,共覆盖基因组 829.1cM	谭晓凤等,1998
柑橘 (*Citrus sinensis*)	回交一代群体	同工酶、RFLP	群体 1:52 个标记被分配到 11 个连锁群中,作图距离 553cM。群体 2:32 个标记被分配到 8 个连锁群中,作图距离 314cM	Durham et al., 1992
苹果 (*Malus pumila*)	拟测交群体	同工酶、RFLP、RAPD	产生两个连锁图:一个有 233 个标记,24 个连锁群;一个产生 350 个标记,分配到 21 个连锁群中	Weeden et al., 1995
葡萄 (*Vitis vinifera*)	2 个种间杂交群体的 4 个亲本	同工酶、RFLP、RAPD	4 张图谱的标记数为 182～225,有 3 个亲本检测到 19 个连锁群,第 4 个亲本检测到 20 个连锁群,作图距离从 1059～1477 cM	Reisch et al., 1996
长叶松 (*Pinus palustris*)	88 颗胚乳和 86 株 F_1 代	RAPD	单倍体胚乳产生 137 个标记,有 75 个标记在 F_1 中能检测到,其中 49 个标记被分离	Kubisiak et al., 1995
湿地松 (*Pinus elliottii*)	一单株上 68 颗胚乳	RAPD	共产生 73 个 RAPD 标记,分配到 13 个连锁群和 9 个连锁对中,覆盖基因组 782 cM	Nelson et al., 1993
南欧海松 (*Pinus pinaster*)	F_1 单株	RAPD	共产生 463 个标记,分配到 12 个较大连锁群和 9 个较小连锁群中,覆盖基因组 1860 cM,平均图距为 8.3 cM	Plomion et al., 1995
欧洲赤松 (*Pinus sylvestric*)	单倍体胚乳	RAPD	共产生 261 个 RAPD 标记,分配到 14 个连锁群中,覆盖基因组 2638.6 cM	Yazdani et al., 1995
薄皮果松 (*Pinus edulis*)	一单株上的 40 颗胚乳	AFLP	338 个标记分配到 25 个连锁群中,覆盖基因组 2012cM,平均图距为 8.0 cM	Travis et al., 1998
欧洲云杉 (*Picea abies*)	一单株上的 72 颗胚乳	RAPD	共产生 96 个标记,解释 185 个 DNA 多态性,被分配到 17 个连锁群中,覆盖基因组 3584 cM,平均图距为 22 cM	Binelli et al., 1994
土耳其红松 (*Pinus brutia*)	4 株单株上各采集 30 颗胚乳,分别对单株作图	RAPD	1 号树产生 56 个标记,35 个分配到 9 个连锁群中,覆盖 661.8 cM;2 号树产生 44 个标记,27 个分配到 10 个连锁群中,覆盖 465.4 cM;3 号树产生 52 个标记,32 个分配到 13 个连锁群中,覆盖 555.2 cM;4 号树产生 27 个标记,13 个分配到 6 个连锁群中,覆盖 163.9 cM	Kaya et al., 1995

续表

树名	作图群体	遗传标记	作图状况	文献
日本柳杉（*Cryptomeria japonica*）	2 株亲本 F_1 代，73 株 F_2 子代	同工酶、RFLP、RAPD	共产生 164 个标记，145 个分配到 20 个连锁群中，覆盖基因组 887.3 cM	Mukai et al.，1995
日本柳杉（*Cryptomeria japonica*）	F_1	AFLP、CAPS	母本和父本两品种的连锁图上分别有 91 和 132 个标记，相应分配在 19 和 23 个连锁群上，跨度为 1266.1 cM 和 1992.3 cM，占基因组的 50% 和 80%	Nikaido et al.，2000
花旗松（*Pseudotsuga menziesti*）	48 株 F_2 子代	RFLP、RAPD	共产生 266 个 RFLP 标记，其中 208 个分配到 17 个连锁群中，覆盖基因组 1075 cM	Krutovskii et al.，1995
乌饭树（*Vaccinium darrowi*）	38 株测交子代	RAPD	共产生 89 个 RAPD 标记，其中 72 个分配到 12 个连锁群中，覆盖基因组 950 cM，平均图距为 16 cM	Rowland et al.，1994
红豆杉（*Taxus brevifolia*）	一单株的 39 个胚乳	RAPD	共产生 102 个 RAPD 标记，其中 41 个分配到 17 个连锁群中，覆盖基因组 305. 8cM	Gocmen et al.，1996
油棕（*Elaeis guineensis*）	F_1 群体	RAPD	用 50 个经过筛选的引物共产生 75 个标记，其中 61 个分配到 14 个连锁群中	Moretzsohn et al.，1997
桃树（*Prunus parsica*）	71 株 F_2 子代	形态标记、RFLP、RAPD	共产生 65 个标记（46 个 RFLP，12 个 RAPD，7 个形态标记），有 47 个连锁到 8 个连锁群中，覆盖基因组 332 cM	Abbott et al.，1995
蓝桉（*Eucalyptus globules*）×细叶桉（*Eucalyptus tereticornis*）	拟测交群体	AFLP	蓝桉：200 个标记分配到 16 个连锁群，总图距为 967 cM。细叶桉：268 个标记分配到 14 个连锁群，总图距为 919 cM	Marques et al.，1997；Marques et al.，1998
可可（*Theobroma cocao*）	100 株 F_1 代	同工酶、RFLP、RAPD	产生 193 个标记，其中 190 个分配到 10 个连锁群中，作图距离为 759 cM	Clamau et al.，1995
马占相思（*Acacia mangium*）	F_1	RFLP、微卫星	得到一张整合的连锁图谱，219 个 RFLP 和 33 个微卫星标记，分配在 13 个连锁群上，覆盖基因组的 966 cM。最大的连锁群为 103 cM，最小的为 23 cM	Butcher et al.，2000
蓝桉（*Eucalyptus globules*）	F_1	同工酶、微卫星、RFLP、EST	通过 98 个共线性的位点整合为一张图谱。包括 249 个标记，总长度为 1375 cM，平均长度约为 5.5 cM	Thamarus et al.，2001

续表

树名	作图群体	遗传标记	作图状况	文献
欧洲栗（*Castanea sativa* Mill.）	F_1	同工酶、RAPD、ISSR	共有 381 个标记被分配到连锁图上，其中母本和父本分别有 187 和 148 个，分别覆盖基因组长度 720 cM 和 721 cM	Casasoli et al.，2001
日本黑松（*Pinus thunbergii*）	71 个大配子体的群体	AFLP、RAPD	有 157 个 AFLP 和 50 个 RAPD 标记，覆盖基因组长度 2085.5 cM，约占总长度的 77.1%～78.4%，标记间平均长度为 10.1 cM	Hayashi et al.，2001
尾叶桉（*Eucalyptus urophylla*）×细叶桉（*Eucalyptus tereticornis*）	F_1 代的 82 个个体	RAPD	尾叶桉连锁图包括 16 个连锁群，包含 129 个框架标记和 65 个侧连标记，总图距为 1741.3 cM，对基因组的覆盖率为 91.0%；细叶桉连锁图也包括 16 个连锁群，包含 96 个框架标记和 43 个侧连标记，总图距为 992.1 cM，对基因组的覆盖率为 81.8%	甘四明等，1999
杉木（*Cunninghamia lanceolata*）	F_1 代的 88 个个体	RAPD	母本连锁图谱包括 11 个连锁群，总图距为 1132.2 cM；父本连锁图也包括 11 个连锁群，遗传跨度为 1057.8 cM	郎亚琴等，2000
马尾松（*Pinus massoniana* Lamb.）	66 个胚乳	RAPD	包括 13 个连锁群，48 个标记位点，覆盖基因组 692.5 cM	尹佟明等，1999

0.2 QTL 定位统计方法的研究进展

0.2.1 数量遗传学的发展历史

20 世纪初数量遗传学作为一门独立的学科而产生，研究实验群体或自然群体的数量性状的遗传变异。它主要研究两个基本问题：①总的表型变异中有多少比例属于遗传的影响；②有多少遗传因子参与这种遗传变异（Wu et al.，1999）。1909 年 Johannsen 提出的遗传力概念可以反映遗传因子对表型变异的贡献。利用亲属间的相关性，如父母间、全同胞间或半同胞间的相关性（Fisher，1918；Cockerham，1954），研究者能够估计出遗传力的大小。现在已有大量的有关数量性状遗传力估计的报道，所涉及的生物体从微生物到植物、动物甚至到人类自身（Lynch and Walsh，1998）。这些估计在决定适当的育种计划以及在说明生物性状的表型进化中都起着极其重要的作用。

虽然已经为估计影响一个数量性状的基因数目的研究付出了巨大的努力，但是它们不能完美地解释数量变异的本质。Wright（1921）首次提出了估计基因数目的方法，这种方法是基于双亲和它们的 F_2 代或回交子代之间的差异。由于 Wright 的方法只涉及一些简单的假设，如所有位点均具有加性效应、显性效应、非连锁和等效性，因此它从一开始就受到批评，随后得到改进（Comstock and Enfield，1981；Lande，1981；Cockerham，1986；Zeng，1992；Wu，1996）。

Wright 的方法没有能够得到很好的应用，究其原因是多方面的。其中一个最重要的原因，就是没有能够将基因间复杂的关系考虑进去。几乎与 Wright 方法同时出现的，是 1923 年 Sax 运用孟德尔的遗传定律把数量的变异分解成单因子的方法。Sax 的方法是当今 QTL 作图方法的原始模型，但是这一方法在 Sax 时代没有能够得到很好的发展，这是因为当时没有足够多的所谓标记来分离数量性状(Thoday，1961)。随着当今分子生物学的发展，基于 PCR 的生物技术几乎能提供无限多数目的 DNA 多态性标记，这些标记可用于构建遗传图谱，进而可研究和确定生物上、医学上和农业上重要数量性状的位点。

在过去的十多年中，人们已提出了很多的使用分子标记进行 QTL 定位的统计方法。在植物中，QTL 作图的方法一般基于起始于两个近交系杂交后的 F_2 代和回交子代等群体。在动物中，尤其是家畜中，现已发展了许多 QTL 作图方法，由于动物作图群体一般为异交群体，其 QTL 作图的统计方法比较复杂。林木有其自身特有的复杂的生物学特性，如生长周期长、遗传负荷高、一些树种有自交不亲和或近交衰退的现象，因此到目前为止，还没有建立比较好的针对林木的 QTL 作图方法，现在林木的 QTL 作图方法一般都借鉴近交群体的 QTL 作图方法(Groover et al.，1994；Bradshaw and Stettler，1995；Grattapaglia et al.，1995，1996；Wu，1998)。

0.2.2 近交群体的 QTL 作图方法

利用两个极端的纯系作为亲本进行杂交产生 F_1 代，然后进行自交或回交产生 F_2 代或 BC 等群体，应用这些群体和遗传连锁图谱可进行 QTL 定位分析。现已发展了不少比较成熟的统计分析方法，常见的有单标记分析法、区间作图法和复合区间作图法等。

0.2.2.1 单标记分析法

单标记分析法检测一个标记的位点是否与某个 QTL 间存在着连锁关系。一般通过方差分析、回归分析或似然比检验比较不同标记基因型数量均值的差异。如果差异显著，那么说明 QTL 与该标记有连锁。由于单标记分析法不需要完整的连锁图谱，因而早期的定位研究多采用这种方法(Soller et al.，1976；Weller，1986；Edwards et al.，1987；Stuber et al.，1987；Tanksley et al.，1982；Luo and Kearsey，1989)。

单标记分析法存在许多缺点(Lander and Botstein，1989)：①不能确定标记与多少个 QTL 连锁；②无法确切估计 QTL 的位置；③由于遗传效应与重组率混合在一起，导致低估了 QTL 的遗传效应；④容易出现假阳性；⑤检测效率不高，需要较多的个体。

0.2.2.2 区间作图法

Lander 和 Botstein(1989)首次提出了基于两个相邻标记的区间作图法，使得能够在整个基因组上搜索 QTL。Lander 和 Botstein 建立了正态混合分布的似然函数，用以计算任一相邻标记之间的任一位置上所对应的似然函数比的对数(LOD)，当区间中最大的 LOD 超过某一给定的临界值时，就认为该区间中存在着 QTL，而且 QTL 的位置也被确定下来。

1992 年 Haley 和 Knott 以及 Martinez 和 Curnow 相继提出了区间作图的回归分析方法。研究表明回归作图方法同 Lander 和 Botstein 的极大似然法作图相比，两者的效果差别不大。从理论上来说，由于 QTL 的分离使得对标记基因型的正态性假设不成立，从而使回归方法不成功。不过大多数信息包含在标记基因型的均值差别上，只有很少一部分

信息来自标记基因型的分布(Haley and Knott,1992)。在计算速度上,回归方法比极大似然法要优越得多,而且回归方法能应用常用的统计软件包进行数值计算。

与单标记分析方法相比,区间作图法有许多优点(Lander and Botstein,1989):①能推断 QTL 在区间中的位置;②如果染色体只有一个 QTL,则 QTL 的位置和效应的估计渐近无偏;③能使 QTL 检测所需的个体数减少。

0.2.2.3 复合区间作图法

Rodolphe 和 Lefort(1993)提出了用多个标记同时检测多个 QTL 的线性模型,该方法能够检测同一连锁群中的一组 QTL,但是它不能提供有关这些 QTL 的数目、位置和效应的精确信息。

Zeng(1993)证明了表型对标记的偏回归系数只取决于两个相邻标记所包括的区间所存在的 QTL,而且与其他区间的 QTL 无关。应用这一特性,Zeng(1994)提出了将多元线性回归与区间作图结合起来的复合区间作图方法。复合区间作图的主要优点是:①仍采用 QTL 似然图来显示 QTL 的可能位置及显著程度,从而保留了区间作图法的优点;②一次检验一个区间,把对多个 QTL 的多维搜索降低为一维搜索;③假如不存在上位性和 QTL 与环境互作,QTL 的位置和效应的估计是渐近无偏的;④充分利用了整个基因组的标记信息;⑤以所选择的多个标记为条件,在较大程度上控制了背景效应,提高了作图的精度和效率(Zeng,1994)。正是由于这些优点,Zeng 的复合区间作图法目前已被广泛应用。

0.2.3 异交群体的 QTL 作图方法

家畜类(尤其是牛和猪)的中密度遗传标记图谱以及大群体的获得使得科学家们能够在基因组上的任何地方搜索 QTL。表型性状一般受到数目较多的基因和环境的影响。依据实验群体的研究,如果蝇(Shrimpton and Robertson,1988)、家畜中的主基因分析(Hanset,1982)以及目前发表的数目有限的有关家畜 QTL 作图研究,可以假定只有少数几个基因的效应较大,有一小部分的基因具有中等效应,而大部分基因的效应却很小。较大效应和中等效应的基因是统计基因作图方法的目标,现在有好几种方案可识别这些基因(Andersson et al.,1994; Georges et al.,1995)。

现已发展了许多统计基因作图方法,可分为五类。第一类是使用单个标记或多种连锁标记的线性回归方法;第二类是使用单个标记或多个连锁标记对假定的两个等位基因的 QTL 进行极大似然分析;第三类是亲子对表型差的平方对某个位点的共祖系数的期望值的回归分析;第四类是基于混合线性模型的残差(或限制)极大似然分析法,此法假定 QTL 等位基因的效应服从正态分布,并应用于已观察到的标记数据条件下的协方差矩阵;第五种是利用单个标记或多个连锁标记拟合两个等位基因或无穷个等位基因的 QTL 的精确贝叶斯连锁分析方法。

这些方法在计算量上有所差别,第一种和第三种计算量最小,第四种稍大,第二种和第五种计算量最大。有些方法是拟合固定的 QTL 效应(第一种),有的是拟合 QTL 的方差(第三—五种),而另一些则是拟合两个等位基因的纯合子的差和基因频率(第二、五种)。计算量少的方法(第一种和第三种)允许数据置换(Churchill and Doerge,1994),可用来计算整个基因组上统计量的阈值,而且容易扩展到多个 QTL 和多个性状的情形。

然而这些方法只能用特定的群体(如半同胞或全同胞),并且第一种方法不能提供除了QTL位置以外其他参数的估计。参数的标准误差和置信区间可通过蒙特卡罗和投影寻踪的抽样技术而获得。

0.2.3.1 线性回归方法

很多研究使用简单的线性回归方法来进行QTL作图,这里只讨论使用多个标记的回归方法。Zeng(1993)给出了用多个连锁标记进行回归分析的理论基础。Knott等(1994)、Spelman等(1996)和Uimari等(1996)提出了用于半同胞群体QTL作图的最小二乘法(LS)。其假设条件是:①父本互不相关,母本互不相关,随机交配,每个母本只有一个子代;②标记图谱(位置和等位基因频率)是已知的。这种LS方法已经编写了Fortran程序(Knott和Spelman之间的私人通信),在一个连锁群里使用所有的标记的情况下,一般可计算出一个亲本中的标记基因型的可能连锁相。

0.2.3.2 极大似然法

在异交群体中,极大似然法已用于半同胞群体的QTL作图(Weller,1986;MacKinnon and Weller,1995)。在给定QTL基因型的条件下,有关假定和模型同LS分析一样。MacKinnon和Weller(1995)给出了半同胞群体关于单个标记的似然函数,而Georges等(1995)则给出了对于多个连锁标记的似然函数。

最小二乘法(LS)和极大似然法(ML)有几点不同。极大似然法一般假定QTL只有两个等位基因,最小二乘法不直接给出QTL参数的估计。在极大似然法里,表型的分布是一个正态的混合分布,不同的QTL基因型其均值不同,而在最小二乘法里假设表型是单一的正态分布,其均值为QTL基因型均值的加权平均,而权则是在标记信息已知条件下的QTL基因型(或等位基因)的概率。对于单个标记,用最小二乘法,QTL的位置和效应不能独自估计出来,而用极大似然法,它们却能被估计出来。用极大似然法,表型的非正态性的假设可能会错误地得出QTL存在的结论。如果QTL就在标记位点的话,那么最小二乘法和极大似然法则是等价的。

0.2.3.3 差方回归法

这种方法是基于分析亲子对表型差的平方,最初是由Haseman和Elston(1972)为分析全同胞对而提出的。该方法对于性状在不同的分布条件下是稳健的(Amos and Elston,1989),并可用于不同类型的子代对(Amos and Elston,1989;Gotz and Hamann,1995)。Gotz和Ollivier(1994)发现对于家猪群体,这种方法至少同最小二乘法一样有效。该方法的假设前提是随机交配和连锁平衡,并且只用到特定类型的亲子对。在该方法中,QTL可有任意的等位基因数目(Fulker and Cardon,1994)。

Haseman和Elston(1972)的方法的缺点是不能区分QTL的方差和QTL与标记之间的距离。

0.2.3.4 基于混合模型的残差极大似然法

Grignola等(1996a,b;1997)发展了一种用于半同胞设计的QTL作图方法,称为残差极大似然法(REML),此法使用多个连锁标记并允许应用家系间的关系。Van Arendonk等(1994)指出当在半同胞群体中使用单个标记时,QTL的方差和位置不能独立地得到估计。

REML方法能使我们使用全谱系信息,它不局限于特定的遗传设计,适合于分析一般的或比较复杂的谱系。因此,它最有可能成为人类谱系QTL作图的有用方法

(Blangero and Almasy,1996),同时也可能是异交植物群体 QTL 作图的有用方法。

0.2.3.5 贝叶斯分析方法

在贝叶斯分析里,统计推断不是使似然函数极大化,而是基于在给定观测值的条件下所有未知变量的联合后验分布。在动物 QTL 作图中,贝叶斯连锁分析首先由 Hoeschele 和 Van Raden(1993a,b)提出,然后由 Thaller 和 Hoeschele(1996a,b)对单个标记通过马尔科夫链蒙特卡罗(MCMC)方法实现。Uimari 等(1996)考虑了用多个连锁标记进行 QTL 作图,Uimari 和 Hoeschele(1997)考虑了多个 QTL 连锁的问题。贝叶斯分析方法还用于植物(Satagopan et al. ,1996)和人类(Thomas and Cortessis,1992)等群体的作图。

早期贝叶斯分析 QTL 作图一般假定在特定的染色体上只出现一个 QTL,并认为有以下几种可能:①标记位点的顺序以及它们之间的遗传距离是已知的;②标记位点的顺序是已知的,但遗传距离是未知的;③标记位点的顺序和遗传距离均未知。目前连锁分析一般采用假设①,即使使用 MCMC(Satagopan et al. ,1996)也是这样。Uimari 等(1996)的分析是基于假设②的,同时假设③的情况也可以包括在内。

0.2.4 林木 QTL 作图

林木具有许多自身特有复杂的生物学特性。首先,林木是多年生植物,生长周期漫长,短则十年,长则数百年。其次,林木是异花授粉植物,具有很高的遗传杂合度。最后,长期的异交使林木具有大量的遗传负荷,自交或近交会导致林木生殖力的降低,形成所谓的近交衰退。

目前林木 QTL 定位的遗传群体主要有以下三种(易能君等,1998)。

(1) 拟近交谱系。种间高度杂合的亲本(P_1、P_2)杂交产生第一子代群体(F_1),两个 F_1 全同胞交配或 F_1 代回交于亲本 P_1 或 P_2 产生第二子代群体(拟 F_2 或拟 BC 群体)。这类设计适合于遗传负荷较低的树种,如毛果杨×美洲黑杨(Bradshaw and Stettler,1995)。

(2) 种间 F_1 设计。种间两高度杂合的亲本(P_1、P_2)杂交产生全同胞 F_1 家系,继而进行无性繁殖,QTL 定位使用这些 F_1 无性系的表型数据。这类设计适合于全同胞交配困难、无性繁殖容易而且杂种优势明显的树种,如巨桉×尾叶桉(Grattapaglia et al.,1995)。

(3) 三代全同胞谱系。在这种设计中,作图群体的两亲本(P_1、P_2)及 P_1 的亲本 GP_{11}、GP_{12},P_2 的亲本 GP_{21}、GP_{22}均存在,其中四个祖亲无亲缘关系,因而 P_1 与 P_2 也无亲缘关系。每对亲本的表型应有极端的分化,即其一有较高的表型值,而另一个则有较低的表型值。表型的分化可增加其子代标记位点杂合的概率,而祖亲的遗传型数据可提供连锁相的信息,因此可提高标记间及标记与 QTL 间连锁检测的效率。

除了以上三种常见的遗传群体,还可利用其他遗传材料进行 QTL 定位,如半同胞家系(Grattapaglia et al. ,1996)和针叶树的大配子体(Wu et al. ,1997)等。

林木 QTL 作图的统计分析方法主要是利用为近交群体而发展的 QTL 作图理论。Grattapaglia 和 Sederoff(1995)提出的"拟测交"方法,使人们可直接使用 F_1 代来定位 QTL。考虑到林木的高度杂合性,Knott 等(1997)提出一种定位具有复等位基因的 QTL 方法。但是这些基于单个或几个家系的定位策略受到不少批评(Muranty,1996),因为它们没有利用目前林木育种研究所常用的交配设计。到目前为止,没有一种方法考虑到林木的复杂生物学特性、它们的群体进化史以及多年育种所积累的育种群体(邬荣领等,2000)。