

密级：

论文编号：

中国农业科学院 博士学位论文

玉米核心自交系群体结构及耐旱相关候选基因
*rab17*的等位基因多样性分析

**Population Structure of Core Inbred Lines and Allelic Diversity of
rab17, A Candidate Gene for Drought Tolerance in Maize**

博 士 研 究 生 ： 于永涛

指 导 教 师 ： 王天宇 研究员

黎 裕 研究员

申请学位类别 ： 农学博士

专 业 ： 作物遗传育种

研 究 方 向 ： 作物种质资源

培 养 单 位 ： 研究生院

作物科学研究所

提交日期 2006 年 6 月

Secrecy:

No.

Chinese Academy of Agricultural Sciences
Ph.D Dissertation

**Population Structure of Core Inbred Lines and Allelic Diversity of
rab17, A Candidate Gene for Drought Tolerance in Maize**

Ph.D Candidate: Yu Yongtao

Advisor: Prof. Wang Tianyu

Prof. Li Yu

Major: Crop Genetics and Breeding

Specialty: Crop Germplasm Resources

Chinese Academy of Agricultural Sciences

June 2006

独 创 性 声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国农业科学院或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

研究生签名：于永清

时间： 2006 年 6 月 20 日

关于论文使用授权的声明

本人完全了解中国农业科学院有关保留、使用学位论文的规定，即：中国农业科学院有权保留送交论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。同意中国农业科学院可以用不同方式在不同媒体上发表、传播学位论文的全部或部分内容。

(保密的学位论文在解密后应遵守此协议)

论文作者签名：于永清

时间： 2006 年 6 月 20 日

导师签名：王天序

时间： 2006 年 6 月 20 日

论文评阅人、答辩委员名单

论文题目	玉米核心自交系群体结构及耐旱相关候选基因 <i>rab17</i> 等位基因多样性分析				
论文作者	于永涛	指导教师	王天宇	培养单位	作物科学研究所

	姓名	职称职务	导师类别	单位	专业
评阅人	李建生	教授	博导	中国农业大学	作物遗传育种
	王建华	教授	博导	中国农业大学	作物遗传育种
	张增艳	研究员	博导	中国农科院作物所	生物技术
答辩主席	辛志勇	研究员	博导	中国农科院作物所	生物技术
答辩委员	李建生	教授	博导	中国农业大学	作物遗传育种
	王国英	教授	博导	中国农业大学	分子生物学
	孙传清	教授	博导	中国农业大学	种质资源
	肖兴国	教授	博导	中国农业大学	分子生物学
	张 辉	研究员	博导	中国农科院作物所	种质资源
	刘录祥	研究员	博导	中国农科院作物所	作物遗传育种
答辩时间与地址	2006 年 6 月 19 日，作物科学研究所南库会议室				
记录人员	王荣焕				

摘要

干旱是限制玉米生产的重要因素之一。发掘耐旱基因资源是提高玉米品种耐旱水平、培育耐旱新品种的根本途径。在玉米种质资源中蕴含着很多影响玉米耐旱性的功能等位基因变异。鉴定出这些功能等位基因不仅有助于深入了解耐旱分子机制，而且还可以开发与这些功能变异紧密关联的功能标记，从而通过分子标记辅助选择的手段来针对耐旱性进行种质创新和遗传改良。关联分析方法由于可以直接鉴定出与表型变异密切相关的功能等位变异，成为解决这一问题的有效手段。

本研究以我国玉米核心种质中的自交系及一些骨干自交系为基础，通过遗传多样性及群体结构分析，明确这些自交系的群体结构，并选出能够最大限度代表其遗传多样性的一组核心研究材料进行耐旱相关候选基因 *rab17* 的序列多样性和连锁不平衡（LD）结构分析；同时，对这套核心研究材料进行耐旱相关表型鉴定。最后，在充分考虑群体结构影响的基础上，进行 *rab17* 基因多态性与耐旱表型之间的关联分析，以求发掘与耐旱相关表型相关的 *rab17* 基因中的功能等位基因。获得的主要结果如下：

1. 利用 49 个 SSR 标记对主要来自玉米核心种质的 288 个自交系进行遗传多样性分析，共检测到总计 262 个等位位点，平均每个 SSR 位点等位位点数为 5.3469，多态性信息含量（PIC）平均为 0.5057，基因多样性平均值为 0.5663。群体结构分析发现，这些自交系按遗传相似性主要可分为四个种质来源，基本对应于塘四平头、Lancaster、Reid 和 P 群四个杂种优势群。总体上，四大种质来源所占的遗传相似性比例依次为 0.290（塘四平头）、0.267（Lancaster）、0.260（Reid）和 0.183（P 群）。以等位位点数目为选择标准，从这组自交系中选出了 94 个自交系作为核心研究材料。这组核心研究材料共拥有 228 个等位位点，占 288 份自交系等位位点总数（262）的 87%。

2. 通过对核心研究材料进行 *rab17* 基因的序列多态性分析，在总计 71 个序列中，共检测到 37 处核苷酸变异，包括 19 个 SNP（平均每 87bp 一个）和 18 个 Indel（平均每 92bp 一个）。在 37 个变异位点中，在非编码区中的有 30 个，占 81%。其中 Indel 在非编码区的发生频率是编码区的将近八倍。总体上来看，在 *rab17* 基因的启动子区及第一个外显子区域中核苷酸多样性较高，之后迅速下降。基因区域内连锁不平衡快速衰退达 800bp 以内。根据连锁不平衡结构，这些序列分为 14 个单体型。根据 *rab17* 基因序列的系统进化树分析表明，主要来自美国杂交种 P78599 的 9 个自交系与其它自交系间的遗传距离最远。

3. 通过 Structured Association（SA）测验，共检测到六个序列多态性与表型性状间的显著关联。其中，编码区的一个 Indel 分别与正常灌溉条件下的脱落酸（ABA）含量及干旱胁迫条件下的相对含水量存在显著关联，含有插入片段的等位基因对于干旱胁迫条件下的相对含水量具有增效作用，这个插入片段也导致了编码区的移码突变；5'-UTR 中的一个 Indel 与干旱胁迫条件下的开花—吐丝间隔期（ASI）间的关联达到显著，其中威风 322、Lo1125 和辽 2204 含有的等位基因使 ASI 平均比另一个等位基因延长了 6 天；另外，一个由 11 个多态性组成的单体型与三个产量因子

之间也存在显著关联,9个大部分来自 P78599 的自交系所含有的等位基因对这三个指标均起到了增效作用,暗示着该等位基因对耐旱性有正向贡献,可以应用于种质改良及育种实践中。

关键词: 玉米 (*Zea mays* L.), 耐旱性, 群体结构, *rab17*, 等位基因多样性

Abstract

Drought stress is one of the most important factors limiting maize production. Mining germplasm resources and discovering associated genes are fundamental approaches to enhancing drought tolerance in maize. There are much functional genetic variations for drought tolerance in maize germplasm. Identifying these functionally favorable alleles contributes to not only understanding molecular mechanism for drought-tolerance but also developing functional markers, and thus improving drought tolerance of maize new germplasm and hybrids by marker-assisted selection. Association analysis is a powerful approach for mining these alleles because its ability of identifying functional variations and association with phenotype.

Analysis of genetics diversity and population structure were conducted by using maize inbred lines from the established core collection and some elite lines, in order to clarify population structure of these lines which is necessary for further association analysis and choose a mini core set with maximal diversity for evaluating sequence diversity and linkage disequilibrium (LD) of *rab17*, a candidate gene for drought tolerance. In addition, phenotypes related to drought tolerance were characterized for the set of materials. Finally, association test between *rab17* polymorphisms with phenotype variation were performed in order to identify functional alleles that are responsible to phenotype variations with drought tolerance. The major results are as follows.

1. Totally 288 maize inbred lines from the established core collection and some elite lines were assayed for genetic diversity at 49 Simple Sequence Repeat (SSR) loci. The results showed that total 262 alleles were identified, the average of alleles per locus was 5.3469, the average of PIC was 0.5057 and gene diversity for the SSRs was 0.5663. Analysis of population structure assessed the membership of the 288 lines in four clusters that corresponded to major germplasm sources (Tangshipingtou, Lancaster, Reid and P), respectively. Overall proportion of membership of the lines in each of the four germplasm sources was: 0.290 (Tangshipingtou), 0.267 (Lancaster), 0.260 (Reid) and 0.183 (P group). Taking allele number as major criterion, 94 lines were chosen as a mini core set from the 288 inbred lines. The mini core set captured 228 alleles, accounting for 87% of the total alleles in the 288 inbred lines.

2. Sequence polymorphisms of *rab17* were evaluated using lines of the mini core set. Totally 37 polymorphic sites were identified in 71 accessions, including 19 SNP (one per 87 bp) and 18 Indel (one per 92 bp). Among these polymorphic sites, 30 sites were in noncoding regions, accounting for 81%. The frequency of Indel occurring in noncoding regions were eight times higher than that occurring in coding regions. Higher nucleotide diversity were observed in the promoter region and the first exon region, and reduced rapidly in the downstream regions. Linkage disequilibrium decayed rapidly within 800 bp in the *rab17* region. There were 14 haplotypes in these accessions based on linkage disequilibrium in this region. Based on the sequence data of *rab17*, analysis of phylogram tree showed that nine lines mostly deriving from P78599, an American maize hybrid, had the farthest genetic

distance from other lines.

3. Six significant associations between polymorphisms and phenotypic traits were identified by structured association test. An Indel in the coding region showed significant association with abscisic acid (ABA) content in the leaves under well water and relative water content (RWC) in the leaves under water stress. At this polymorphic site, the allele contain inserts which led to frameshift mutation in the coding region had an effect on increasing relative water content under water stress. A Indel in 5'-UTR showed significant association with ASI under water stress, and the allele present in three lines (Weifeng322, Lo1125 and Liao2204) led to a ASI increase of six days, compared with the other alleles. In addition, a haplotype, which consisted of eleven polymorphic sites, showed significant association with three yield related traits, and the allele present in nine lines (mostly deriving from P78599) had a positive effect on increasing the values of the three traits, implying that the allele had contribution to drought tolerance in maize and had the potential to be used in germplasm enhancement and breeding.

Key words: Maize (*Zea mays* L.), drought tolerance, population structure, *rab17*, allelic diversity

目 录

第一章 引 言	1
1.1 玉米耐旱性概述	1
1.1.1 耐旱性相关性状与常规鉴定	2
1.1.2 耐旱的生理生化基础	3
1.1.3 耐旱性分子遗传机制	4
1.1.4 ABA在玉米耐旱性中的作用	6
1.2 玉米自交系中的遗传多样性	6
1.2.1 遗传多样性的概念	6
1.2.2 我国玉米自交系遗传多样性及杂种优势群	7
1.2.3 玉米中的群体结构研究	8
1.3 玉米基因组中的序列多样性	8
1.3.1 SNP与连锁不平衡 (LD)	8
1.3.2 玉米基因组序列中的序列多样性	11
1.3.3 玉米及其它植物中的连锁不平衡结构	12
1.3.4 选择对玉米自交系LD及遗传多样性的影响	13
1.4 关联分析及在玉米研究中的应用	15
1.4.1 关联分析的概念	15
1.4.2 关联分析与连锁不平衡 (LD) 结构	16
1.4.3 关联分析与群体结构	17
1.4.4 关联分析中的统计方法	18
1.4.5 玉米中的关联作图	18
1.5 立题意义与技术路线	20
1.5.1 立题意义	20
1.5.2 技术路线	21
第二章 玉米核心自交系的遗传多样性及群体结构分析	22
2.1 前言	22
2.2 材料与方法	22
2.2.1 实验材料	22
2.2.2 基因型鉴定	34
2.2.3 遗传多样性及群体结构分析	36
2.3 结果与分析	38
2.3.1 遗传多样性分析	38
2.3.2 288 份自交系的聚类分析	40
2.3.3 288 份自交系的群体结构分析	41
2.3.4 Core Sets构建	46
2.4 讨论	48

2.4.1 遗传多样性的评价.....	48
2.4.2 遗传结构和玉米自交系的系谱来源.....	49
2.4.3 群体结构在评估自交系和杂交育种亲本选配中的作用.....	52
2.4.4 群体结构、聚类结果与系谱之间的比较.....	52
第三章 <i>rab17</i> 基因核苷酸多态性及连锁不平衡结构分析	54
3.1 前言	54
3.2 材料与方法.....	54
3.2.1 实验材料.....	54
3.2.3 <i>rab17</i> 基因的扩增与测序	55
3.2.4 序列多态性统计分析.....	56
3.3 结果与分析.....	57
3.3.1 <i>rab17</i> 基因的序列多态性	57
3.3.2 <i>rab17</i> 基因的序列多样性分析	60
3.3.3 <i>rab17</i> 基因的连锁不平衡结构分析	62
3.3.4 <i>rab17</i> 基因系统发育树构建	67
3.4 讨论	67
3.4.1 <i>rab17</i> 基因的序列多态性频率	67
3.4.2 功能序列中的多态性.....	67
3.4.3 <i>rab17</i> 基因序列系统发育树与自交系系谱关系	70
3.4.4 <i>rab17</i> 基因的LD结构及选择的影响.....	71
第四章 <i>rab17</i> 基因核苷酸多态性与耐旱相关性状间的关联分析	73
4.1 前言	73
4.2 材料与方法.....	73
4.2.1 实验材料.....	73
4.2.2 耐旱性相关性状调查及测定.....	73
4.2.3 <i>rab17</i> 基因序列多态性与表型性状间的关联分析	76
4.3 结果与分析.....	77
4.3.1 表型数据统计分析.....	77
4.3.2 <i>rab17</i> 基因序列多态性与表型性状间的关联分析	79
4.4 讨论	80
4.4.1 多态性变异的功能性.....	80
4.4.2 可能影响关联检测结果的一些因素	81
第五章 结 论	83
5.1 玉米核心自交系的遗传多样性及群体结构分析.....	83
5.2 <i>rab17</i> 基因的序列多样性及连锁不平衡结构分析	83
5.3 <i>rab17</i> 基因序列多态性和耐旱相关表型之间的关联分析	83
参考文献	84
致 谢	95

第一章 引言

随着全球性生态环境的不断恶化,干旱已经成为一种世界性的重大农业灾害。其对农作物造成的损失在所有的非生物胁迫中居首位,仅次于生物胁迫病虫害造成的损失。玉米对于干旱胁迫比较敏感。根据我国玉米带地区历年气象资料和玉米产量的相关分析,干旱是导致我国玉米产量波动的主要原因,同时,受全球温室效应的影响,水资源短缺矛盾将愈来愈突出。因此提高玉米自身的耐旱性才是减轻旱害,保持玉米稳产、高产的根本保证。应该把耐旱性作为重要的战略目标,努力挖掘耐旱资源,把玉米耐旱性水平提到一个新的高度。

提高玉米自身耐旱性,挖掘耐旱资源,归根到底要靠玉米自身的基因资源。因而,发掘基因资源就成为了玉米耐旱遗传研究中的一项重要工作。近年来,在耐旱 QTLs 定位(Frova et al., 1999; Quarrie et al., 1999; Ribaut et al., 1996, 1997)、干旱诱导基因的分离与鉴定(Yu and Setter, 2003; Zinselmeier et al., 2002; Riccardi et al., 1998)等方面做了大量的工作,鉴定、分离出了一些耐旱相关的候选基因。然而,现在人们越来越认识到,在植物体内蕴藏着大量的等位基因变异,这些功能变异直接导致了表型的变异。因而如果要全面地了解耐旱机理,仅仅分离出这些耐旱候选基因是不够的,必须鉴定出这些候选基因中蕴含的对耐旱性有正向贡献的功能等位变异。另外,发掘这些功能变异对育种工作也具有重要意义。可以通过分子标记辅助选择等手段来改良种质的耐旱性。

关联分析方法是一项能够把目标性状表型同基因的多样性结合起来进行分析,可以直接鉴定出与表型变异密切相关的功能等位变异的方法。本研究的主要目标就是在考虑自交系群体结构的基础上,采用关联分析方法,鉴定出在候选基因中蕴含的对耐旱性有正向贡献的功能等位变异。下面先就国内外在相关领域的研究进展作一综述。

1.1 玉米耐旱性概述

干旱对玉米所造成的危害涉及生长发育的各个阶段,影响生理生化代谢的多种途径。而玉米的耐旱性则是一个复杂的综合性状,是指在干旱胁迫环境下,植株细胞在结构、生理生化等方面发生一系列适应性反应后,最终在形态和产量上的集中表现。目前,国内的玉米耐旱研究主要是在形态指标、生理生化、分子标记等方面开展工作(霍仕平等, 1995a; 关义新等, 1996; 郑常祥等, 2002; 李新海等, 2002),取得了很大进展。众多的研究已经表明,玉米耐旱育种是提高玉米品种的耐旱性、减少干旱带来的损失的有效途径。多年来,植物生理学家和育种工作者们在这个领域做出了不懈的努力,通过常规育种手段进行耐旱育种取得了一些进展(Heisey & Edmeades, 1999)。然而,耐旱性的遗传机制相当复杂,所以进行耐旱性的常规改良比较困难。近年来兴起的以分子标记辅助选择和转基因育种为主要手段的分子育种可以在很大程度上克服常规耐旱育种中的一些问题,从而提高育种效率,加速耐旱玉米品种的培育和推广。但是,由于我们对耐旱性的分子遗传机制依然缺乏足够的了解,使得我们在分子育种中没有很好的优异基因资源可供利用。为此,迫切需要在基因或基因组水平上对耐旱的分子基础进行深入研究。近半个世

纪以来,国内外的玉米育种家及植物生理学家在这方面进行了广泛而深入的研究,并取得了显著的成绩,不仅为耐旱玉米种质鉴定提供了有效的参考指标,而且对玉米耐旱改良提供了理论指导。

1.1.1 耐旱性相关性状与常规鉴定

玉米耐旱相关性状主要包括形态性状、产量性状、生理生化性状等几大方面。

玉米的形态学指标主要包括根、茎秆、叶片、株型、开花至抽丝的时间(Anthesis-Silking Interval, ASI)、花粉败育率等。形态学性状的观察与测定,是最直接也是最简便的耐旱鉴定途径。在具体鉴定过程中,需要对其中的多个指标进行重复测定和综合评价,以提高鉴定结果的可信度。

生理生化指标主要有水分状况指标(包括水势、渗透势、相对含水量、叶片膨压等)、气孔导度和气孔扩散阻力以及蒸腾速率、水分利用效率、渗透调节能力、光合作用速率、呼吸作用、脯氨酸积累能力、脱落酸(ABA)含量、酶活力以及其它代谢物质等。其中较为重要的耐旱性指标是 ABA 含量和相对含水量。

干旱胁迫下,籽粒产量显著降低是玉米遭受伤害的最终体现。许多学者以玉米在干旱胁迫下的减产程度和产量构成性状的稳定性作为玉米耐旱性强弱的指标。籽粒产量是一个复杂的性状,取决于很多因素,如生长势、充足的水分及营养供给等。干旱对玉米产量有不同程度负面影响,减产程度取决于干旱胁迫的持续时间以及植株的发育阶段,也与种质的耐旱强弱有关。一般可使玉米减产 20%~30% (李新海等, 2002)。尽管如此,目前几乎所有的生产性育种项目仍将产量作为一个选择参数(Bruce et al. 2002),因为籽粒产量是玉米耐旱与否的最终体现。其鉴定指标主要包括株穗数、穗重、穗粒数、穗粒重、百粒重等等,通过比较干旱处理与正常灌溉下相应指标的比值来衡量耐旱性的高低。也有人依据耐旱指数(耐旱指数=(对照产量-干旱处理产量)/对照产量)来进行耐旱性判定的标准。

从玉米对水分敏感时期来看,苗期和开花期是对干旱最为敏感的两个时期,目前研究的也比较多。干旱胁迫强烈影响玉米幼苗的水分状况。苗期干旱最明显的变化是叶片萎蔫,植株生长受到抑制。而开花前期、开花期干旱则会使叶片过早衰老,导致同化作用降低。使得玉米穗和籽粒减少,同时也产生很多小的籽粒,最终造成减产。对异花授粉的玉米来说,当干旱在开花前和开花期发生,花粉自然散粉、传粉能力下降,会使抽丝延迟,从而导致 ASI 增加(Hall et al, 1982; Bolanos & Edmeades, 1993)。干旱胁迫还会造成开花-吐丝间隔期长,花期不遇,或者雌穗抽不出来,严重时减产达 50%以上(Westgate & Boyer, 1986)。这种现象具有很高的遗传力,在田间很容易测定(Ribaut et al, 2002)。CIMMYT 对热带玉米在干旱条件下,ASI 的变化连续进行了多年研究发现,干旱胁迫下 ASI 可由几天延长至 60 天左右(Bolanos et al, 1996; Chapman et al, 1999)。

在国内,近年来针对玉米耐旱性鉴定指标也曾进行过一些研究(罗淑平, 1990; 张宝石等, 1996; 孙彩霞等, 2004; 宋凤斌等, 2004),但始终还没建立起一套完整而又行之有效的玉米耐旱性鉴定指标体系。黎裕等(2004a)认为,对于玉米耐旱品种筛选和种质资源耐旱性鉴定,应采用不同的评价指标,对于耐旱品种筛选,耐旱指数是较好的指标,而耐旱系数及耐旱性指数则更适用于种质资源耐旱性鉴定评价。

1.1.2 耐旱的生理生化基础

（一）渗透调节

干旱胁迫下,植物的各种生理活动都会受到严重影响。大部分农作物都具有渗透调节的能力来减轻干旱的危害。细胞的这种机制,是调整体内的渗透势,缩小植物与环境之间的水势差,使得植物不仅能避免向高负水势的环境中逸失水分,而且还能从干旱环境中吸收水分,使之平衡细胞内水分亏缺,保持一定的膨压,维持正常的代谢活动。植物的渗透调节主要通过脯氨酸、甜菜碱等亲和性溶质的积累而实现 (Bray, 1993)。此外,离子和水分通道的变化调节着离子和水分进出细胞,也是渗透调节的重要方面。

（二）脱水保护

在严重干旱条件下,超出渗透调节范围时,有些植物能够存活,恢复供水后迅速恢复生长,其机理在于,这些植物在失水时产生的一些大分子或小分子有机物具有脱水保护功能,能够在水分亏缺时保护膜系统及其他生物大分子免受破坏,胚胎发生后期富集蛋白 (Lea 蛋白) 是这类脱水保护物质的代表。

Lea 蛋白 (Late embryogenesis abundant protein) 是指胚胎发生后种子中大量积累的一系列蛋白质,广泛存在于高等植物中。在植物个体发育的其他阶段,也能因 ABA 或脱水诱导而在其它组织中高水平表达。一般认为,Lea 蛋白在植物细胞中具有保护生物大分子,维持特定细胞结构,缓解干旱、盐害、寒害等环境胁迫的作用。

Lea 蛋白大多是高度亲水的。高度亲水性有利于 Lea 蛋白在植物受到干旱而失水时,能够部分替代水分子,蛋白质的多羟基能保持细胞液处于溶解状态,从而避免细胞结构的塌陷,稳定细胞结构,尤其是膜结构 (Baker et al, 1988)。

植物在受到干旱胁迫时,Lea 基因的高水平表达和 Lea 蛋白的大量积累也表明 Lea 蛋白基因具有干旱保护功能 (Hong et al, 1992)。Xu 等 (1996) 将 HVA1 cDNA 全序列导入水稻,获得了耐旱的转基因水稻,从而直接证实了 Lea 基因的干旱保护功能。

在 Lea 蛋白研究中目前研究最多的是脱水蛋白。脱水蛋白在胚胎发育晚期大量累积,对 ABA、低温以及一些起强制脱水作用的环境因子如干旱,细胞外冷冻或者盐害等进行响应,在植物干旱应答上起着积极作用 (Close et al, 1990,1996)。脱水蛋白的定义特征是一个亲水性 (amphipathic) 的 α 螺旋结构域,即 K-片段,该片段在高等和低等植物中具有较强的保守性。被子植物 K-片段中共有的氨基酸序列由十五个氨基酸组成: EKKGIMDKIKEKLP。另外,在脱水蛋白结构中还有一个丝氨酸束,称为 S-片段,以及一个称为 Y-片段的 N-端保守序列。

玉米中也发现了一系列 Lea 蛋白。RAB-17 即为此类蛋白中的一员,研究发现,该蛋白在胚胎发育晚期和叶片中被 ABA 和干旱胁迫诱导表达 (Pla et al, 1989; Vilardell et al, 1990)。

（三）抗氧化防御系统

20 世纪 80 年代以来,人们对于干旱胁迫下植物体内抗氧化防御系统进行了大量的研究,发现植物受到干旱胁迫后体内活性氧的积累程度与多种因子有关,在绿色植物组织中,光合电子传递系统是活性氧 (超氧离子 $O_2^{\cdot-}$) 的主要来源。超氧化物与 H_2O_2 反应生成一种潜在的强氧化剂—羟基,它攻击大部分的生物大分子,造成 DNA 损伤,从而影响蛋白质的合成,甚至导致细胞死亡。然而当超氧化物歧化酶 (SOD)、过氧化氢酶 (CAT) 以及不同的过氧化物酶 (POD) 存在时,

会中断过氧化物与 H_2O_2 反应,以达到对细胞的保护作用,而活性氧的直接清除剂是什么目前尚不完全清楚。王茅雁等(1995)研究发现,干旱胁迫下,SOD活力在不同耐旱性的玉米品种或杂交组合中均呈现下降---回升---下降的变化趋势;CAT活力在耐旱性强的品种中先升(或基本不变)后降,而在耐旱性弱的品种中一直降低;POD活力在所有品种均呈明显降低。同时,叶片质膜透性增大,叶绿体和线粒体超微结构受损,耐旱性强的品种受损程度较轻。SOD和CAT活力与质膜、叶绿体和线粒体膜系统结构的损伤程度呈负相关,与品种耐旱性呈正相关,可作为玉米耐旱鉴定的生化指标;而POD活力与上述膜系统的损伤程度及品种耐旱性相关不大。

(四) 气孔调节

气孔调节是干旱胁迫下植物适应环境、抵御干旱的机制之一。玉米通过气孔调节减少蒸腾,从而减少或避免旱害损伤。王畅等(1990)的研究结果表明,玉米对于旱适应的气孔调节机制,主要表现在气孔扩散阻力的改变上。玉米种质在干旱胁迫下,气孔阻力明显增大,以此减少体内水分的散失,不同品种之间其气孔阻力增加幅度不同,耐旱性强的玉米通过加大气孔阻抗,减少蒸腾,有效地减少水分的大量损失,保持植株一定的水分条件,使其新陈代谢得以正常进行,而耐旱性差的玉米品种增加幅度则较小。

1.1.3 耐旱性分子遗传机制

植物在干旱胁迫下,将诱导一系列基因表达,参与相关代谢途径的调控。这些基因主要对于旱胁迫有两个方面的作用:一方面,通过渗透调节物质的积累来提高细胞的渗透调节能力,保证水分子的正常运输,从而有利于维持细胞膜的完整性;另一方面,干旱胁迫使得一些基因过量表达,大大促进了一些参与耐旱防御的代谢产物的合成,以此提高其耐旱能力。

目前,在一些植物包括重要的农作物如豌豆、西红柿、水稻、小麦族(Cattivell et al, 2002)和模式植物拟南芥以及极端干燥环境下的复苏植物、盐生植物中都发现了一些与耐旱胁迫相关的基因(Bray, 1997; Bohnert et al, 1996; Ingram et al, 1996)。按功能划分,这些基因包括糖代谢调节酶基因,合成各种渗透调节因子(糖、脯氨酸和甜菜碱等)所需酶的基因,水通道蛋白基因和离子通道蛋白等转运因子基因,细胞壁结构调节酶基因,蛋白合成、加工和降解所需蛋白(酶)基因,去毒性蛋白基因(可溶性环氧化物水解酶基因、过氧化氢酶基因和抗坏血酸过氧化物酶基因等),信号转导和基因表达中所需的蛋白激酶基因、磷脂酶C基因和转录因子等调节蛋白基因等(Shinozaki et al, 1996;1997)。其中,渗透调节物质合成所需基因、控制离子及水分流动的蛋白基因以及合成清除氧自由基所需物质的基因具有特别重要的意义,因为它们是植物对于旱胁迫耐受的生化代谢的主要分子基础。对这些基因的表达特性及其调控机理的研究,是阐明植物对于旱胁迫响应的分子机理的重要内容,也是有效地通过转基因方式来改良植物性状的前提。

在玉米研究中,经过不断的探索和实践,近些年来在干旱胁迫下的基因表达、耐旱相关基因的克隆及功能鉴定等方面做了大量的工作,鉴定、分离出了一些耐旱相关的候选基因,为阐明耐旱的遗传机理提供了分子基础(Yu and Setter, 2003; Zinselmeier et al, 2002; Riccardi et al, 1998)。例如,Zinselmeier等(2002)以先锋玉米杂交种3732(耐旱材料)为实验材料,利用GeneChip技术对穗和粒发育阶段干旱诱导的1502个基因进行了研究。这些基因代表了27个调控和代谢途

径, 其中有 179 个基因呈现差异表达。

然而, 玉米的耐旱性机制相当复杂, 更多的是由数量性状位点 (Quantitative Trait Locus, QTL) 控制的 (Frova et al, 1999; Quarrie et al, 1999)。近几年来, 科学家们先后利用不同作图群体, 对控制耐旱性及其相关性状的 QTL 区域进行了定位和作图 (Ribaut et al, 1996, 1997; Sanguinetti et al, 1999; Sari-Gorla et al, 1999; Tuberosa et al, 2002; Li et al, 2003; 张吉民等, 2004)。这些 QTLs 涉及的性状主要包括形态学性状、产量性状及其产量构成因子 (籽粒产量、单株穗数、穗粒数及百粒重等)、开花-吐丝间隔期 (ASI)、ABA 含量、根拉力等, 以及在干旱胁迫下可能产生遗传变异的一些生理生化性状, 比如和干旱胁迫应答相关的一些代谢途径中的关键酶的活性等。

Lebreton 等 (1995) 利用 84 个 RFLP 位点对 POlj17 (耐旱型) \times F2 (干旱敏感型) 群体进行了耐旱相关性状 (如气孔导度、ABA、叶片水势、根数、膨压以及荧光特征等) 的 QTL 分析。结果表明, 在染色体 1、2、7 和 10 上有 4 个控制气孔导度的 QTLs; 染色体 3 上有控制 ABA 含量、叶片水势和叶片膨压以及根系拉力的位点。在染色体 2、3、5、6、和 8 上有 5 个与叶绿素含量有关的 QTLs。Agrama 等 (1996) 对 SD34 \times SD35 组合, 用 70 个 RFLP 位点进行了耐旱性的 QTL 分析, 在玉米染色体 1、3、5、6 和 8 上发现了影响产量的基因, 通过这 5 个推测的 QTLs 可以使与产量有关的 50% 的表型变异得到解释。Ribaut 等 (1996) 对与耐旱性密切相关的性状—ASI (开花至抽丝天数) 进行了 QTL 定位, 在染色体 1、2、5、6、8 和 10 上发现了 6 个位点。Tuberosa 等 (2002) 对玉米干旱胁迫下调控其形态生理、产量等性状的 QTLs 进行了研究, 揭示了这些相关性状复杂的遗传基础及其遗传背景和环境对 QTLs 效应的影响。该实验还研究了水培条件下, 根长、根重、根直径以及籽粒产量的 QTLs, 证实了通过该途径在玉米生长的早期鉴定控制根部性状和籽粒产量的 QTLs 位点是可行的。Sanguinetti 等 (1999) 在干旱条件下两年鉴定的玉米材料中, 随机选取 80 份 F₄ 家系进行研究后发现, 在不同的生育期和年份, 控制各个性状的 QTLs 的数目存在差异。Sari-Gorla 等 (1999) 和 Frova 等 (1999) 用 B73 \times H99 组合的 RIL 分别对其在干旱条件下进行了分子标记研究, 在几个染色体上分别发现了控制雄蕊及雌蕊开花时间、花药散粉时间的基因位点、控制株高以及产量的基因。

近几年来, CIMMYT 对干旱胁迫下, 玉米在开花前、开花期和开花后的耐旱情况也进行了广泛的研究。一些与产量组成和主要次级形态性状 (如 ASI) 表达的 QTLs 得到了鉴定。Ribaut 等 (1997) 利用 QTL 对产量组成和主要次级性状在干旱条件下的表现进行了研究, 发现这些性状是复杂的多基因性状。为了更加深入地探讨干旱胁迫条件下干旱诱导基因的表达机制, CIMMYT 还利用 QTL 对能够诱导耐旱表型的一些重要生理途径的差异表达进行了研究。

研究发现, 在玉米染色体 3 上的一个区间, 同时聚集有与玉米在干旱条件下的籽粒产量、开花-吐丝间隔时间及单株穗数有关的 QTLs (Agrama et al, 1996); 在玉米染色体 10 上有 4 个与转化酶活性和己糖浓度相关的 QTLs 也位于同一区域, 形成 QTLs 簇; 类似的情形还出现在染色体 1 和 5 上。对大麦的研究也曾发现类似的与渗透调节性状相关的基因簇 (Teulat et al, 1998)。一般认为, 这样的基因簇位于基因组的“热点”区域 (Hot Spot), 可能参与相关性状的调节。这使我们能够更好地认识耐旱相关性状的分子基础。

然而, 如何将所鉴定出的 QTL 与基因的功能、植物表型有机地结合起来进行研究, 仍是亟待解决的问题。功能基因组学的发展, 有助于解决上述问题, 将功能基因的研究与 QTL 检测相结合, 把着眼点放在那些既已检测到耐旱相关性状的 QTLs, 又已发现存在主效基因或者 EST 的

基因组区域上。这类区域应是蕴藏有重要耐旱基因资源的关键区域。分析该区域的基因组序列，找到其中的等位基因变异，与耐旱表型变异结合起来进行关联分析，就可能找到解析耐旱机制的突破口。另外，还可以将 QTL 分析与蛋白组学联系起来，在候选蛋白与 QTL、性状变异之间建立一种联系，再进而找到候选基因（黎裕等，2004b）。这些新的分子生物学技术的发展，为解析耐旱生理生化机制提供了一些新的思路。

1.1.4 ABA 在玉米耐旱性中的作用

脱落酸（ABA）是一种植物激素，在植物对逆境胁迫应答方面起着重要的作用。干旱胁迫可以诱导 ABA 的合成，使 ABA 含量的增加，继而调控多种相关基因的表达，这些基因的产物又可以保护细胞免受干旱胁迫带来的危害。在玉米上，干旱胁迫会使玉米幼苗叶片和花期叶片中的 ABA 成倍增加，ABA 含量的增加可以增加干旱胁迫下的根茎比，维持根细胞伸长，调节细胞某些渗透调节物质（如脯氨酸）的浓度，降低气孔导度，促进根对水分的吸收（Tuberosa et al, 2002; 霍仕平等, 1995b）。丁雷等（1993）也指出，干旱胁迫使叶片 ABA 含量增加，导致气孔关闭，这是植物对干旱的反应；来自叶肉细胞、根部和保卫细胞的 ABA 被保卫细胞质膜的受体所识别，促进气孔关闭和抑制气孔开放；ABA 还通过增加蛋白质合成和影响气孔的力学特性对气孔特性产生长期影响。利用外源 ABA 研究玉米耐旱性也取得了一些进展。张烈等（1998）对孕穗期的玉米植株喷施 ABA（2mg/kg）后，土壤进行轻度干旱胁迫。结果发现气孔阻力增加、蒸腾速率降低、细胞膜的破损程度降低、而叶绿素总量与施用 ABA 与否，差异不大。同时证实，在轻度胁迫下，ABA 处理植株与对照植株叶片的含水量差异显著，游离脯氨酸含量增加近 2 倍。

在分子水平上对 ABA 与耐旱性关系的研究也取得了一些进展。一些学者对控制 ABA 含量的 QTL 进行了检测，发现玉米中的 ABA 含量由多个基因控制，并且检测到大量相关的 QTLs 位点。Lebreton 等（1995）利用 84 个 RFLP 位点进行 QTL 分析，在染色体 3 上检测到了控制 ABA 含量的 QTL，Tuberosa 等（1998）对 80 个 F_{3,4} 玉米材料在干旱胁迫下叶片 ABA 含量进行了遗传分析，并进行 RFLP 作图得到了 16 个 QTLs。Ribaut 等在研究中检测到 16 个相关 QTLs，发现其中的 4 个能在不同时期显著影响叶片 ABA 含量。另外，研究中还发现了控制 ABA 含量的 QTL 与控制其它耐旱性状的 QTLs 的区域重叠，并且发现，ABA 含量的增加大多数情况下是和气孔导度的减小、产量的下降密切相关的（Sanguinetti et al, 1999）。正象前面提到的，将 QTL 检测与基因诱导表达、等位基因变异的发掘等手段结合起来，可能会成为揭示 ABA 调控耐旱机制的突破口。

1.2 玉米自交系中的遗传多样性

1.2.1 遗传多样性的概念

遗传多样性是生物多样性的的重要组成部分，也是物种和生态系统多样性的重要基础。植物遗传多样性包括植物不同种间，种内不同种群间以及同一种群内不同个体间的遗传多样性。遗传多样性研究是种质资源研究的一个重要方面。物种的遗传多样性越高或遗传变异越丰富，说明其对环境变化能力就越强，分布的范围就越广。

种内的遗传多样性是物种以上各水平多样性的最重要来源。多样性程度不仅是长期进化的产物，也是其生存和发展的前提，因此种内的遗传变异程度决定了其进化的潜力 (Solbrig, 1991)。所有的遗传多样性都发生在分子水平。这种分子遗传多样性程度与突变、重组、选择和遗传漂变等多种因素有关。其中突变是所有变异的来源，这些突变的累积形成了种内的大量遗传变异。经过自然选择，一些中性突变随机地整合到基因组中，形成了丰富的遗传多样性。通过对遗传多样性的研究，可以揭示物种的进化历史（起源的时间、方式），并为挖掘其遗传潜力奠定基础。

1.2.2 我国玉米自交系遗传多样性及杂种优势群

玉米 (*Zea mays* L.) 属于禾本科玉蜀黍族 (*Maydeae*) 玉蜀黍属。在玉米族中共有 7 个属，分别为玉蜀黍属 (*Zea* L.)、摩擦禾属 (*Tripsacum* L.)、薏苡属 (*Coix* L.)、流苏果属 (*Chionachne* R Br.)、硬皮果属 (*Schlerachne* R Br.)、三裂果属 (*Trilobachne* Henr.) 和多裔黍属 (*Polytoca* R Br.)。在玉蜀黍属中除栽培玉米外，其它均为大刍草的各个种和亚种，玉米即从大刍草进化而来。

在大约 6000~9000 年前，在经历了漫长的驯化历史后，在墨西哥南部开始形成了原始的玉米地方品种。进入 20 世纪后，育种家们又开始从对这些地方品种进行群体改良，从中选育自交系，用于组配具有强大杂种优势的杂交种。然而，由于人们只针对少数优良的农艺性状进行选择，不可避免地造成了玉米栽培种质中遗传多样性的降低。

大约 500 年前，玉米被引入中国，不同的种质被引种到了不同的地区，然而，这些种质的系谱关系和引种经过并没有被记录下来。之后经历了长期的培育选择，形成了我国大量的玉米地方品种。这些地方品种成为了我国玉米育种及生产的宝贵资源。长期以来我国玉米育种家利用这些地方资源育出了很多优良品种。20 世纪以来，我国的育种工作者开始选育自交系，组配杂交种。随着大量优良自交系的选育和大力推广，我国玉米种质中的多样性也在不断减少，吴景锋 (1983)、曾三省 (1990) 都认识到了这种趋势。他们认为由于遗传多样性下降所导致的亲本遗传基础狭窄现象，是当前作物育种难以取得突破性进展的重要原因。因而，近些年来，育种家又从国外引入了很多玉米的种质，如先锋种质、热带种质等。

随着杂交育种研究的深入，对我国的玉米自交系资源进行杂种优势群划分成为一项极其重要又亟待解决的工作。20 世纪 80 年代以来，通过分析系谱来源、地理分布等手段，育种家们对我国自交系的种质基础进行了初步的分析 (吴景锋, 1983; 曾三省, 1990)。王懿波等 (1998) 在对我国广泛应用的自交系的系谱进行系统分析以后，结合配合力以及杂种优势等因素，认为我国的自交系主要有四大类群：来自美国 Lancaster 和 Reid 群，以及国内起源的塘四平头与旅大红骨群，其中的 Lancaster 群包含 Mo17、自 330 两个亚群，此外还存在选自国外杂交种、综合品种、Suwan 和低纬度地区种质组成的一些亚群。

近几年来，随着 RFLP、RAPD、SSR、AFLP 等分子标记技术在种质资源研究中的大量应用，玉米杂种优势群的划分工作又取得了很大的进展 (表 1-1)。各种分子标记技术都在玉米杂种优势群划分上表现出了一定的适用性，其中，SSR 标记因为可靠性、重复性好而成为近几年来进行相关研究的首选。从划分结果看，袁力行等通过 RFLP 和 SSR 进行划分后，认为四大类群中的 Reid 群应包含 BSSS 和 PA 两个群 (袁力行等, 2001)。后来，选自美国先锋杂交种 P78599 的 P138、X178、齐 319、沈 137 等自交系在我国玉米生产中扮演了越来越重要的角色，在加入上述自交系

后进行杂种优势群划分时发现，这些自交系被划分到一个独立的杂种优势群，称为 P 群（赵久然等，1999；王元东等，2004）、PB 群（李新海等，2003）或者温热 I 群（滕文涛等，2004）。

表 1-1 对我国玉米自交系杂种优势群划分的一些研究

Figure 1-1 Studies about heterotic groups of inbred lines

方法	杂种优势群划分结果	参考文献
系谱、配合力、杂种优势综合分析	Reid、塘四平头、旅大红骨、Lancaster (Mo17、自 330)、其它	王懿波等，1998a,b
RAPD	Reid、塘四平头、旅大红骨、Lancaster、P 群	赵久然等，1999
AFLP	六个类群	吴敏生等，2000
RFLP	热带、Mo17、自 330、B73、地方 I、地方 II	黄益勤等，2001
RFLP、SSR	塘四平头、旅大红骨、Lancaster、BSSS、PA	袁力行等，2001
SSR	塘四平头、旅大红骨、Lancaster、BSSS、PA、PB	李新海等，2003
SSR	Reid、塘四平头、E28、Lancaster、温热 I、自 330、其它	滕文涛等，2004

1.2.3 玉米中的群体结构研究

利用群体结构的方法来研究玉米自交系群体内的遗传组成是近几年刚刚见诸报道的一种分析手段。Liu 等用 SSR 标记对 260 个由代表了美国温带玉米遗传多样性的自交系和重要的热带、亚热带自交系组成的一组玉米材料进行了遗传多样性及群体结构分析，发现热带亚热带自交系的遗传多样性要高于温带材料，而在温带自交系中，又以 SS (Stiff Stalk) 群的遗传多样性为最高，而且还发现玉米自交系含有的等位基因数还不到地方品种中的 80%，预示着地方品种还可以为玉米育种提供更多的遗传多样性 (Liu et al, 2003)。另外，也有用类似的方法对玉米野生近缘种大刍草进行的研究报道 (Fukunaga et al, 2005)。目前我国还没有相关研究被报道。

1.3 玉米基因组中的序列多样性

1.3.1 SNP 与连锁不平衡 (LD)

(一) 核苷酸多态性

基因组中的核苷酸多态性主要指 SNP 和插入/缺失 (Indel)。

单核苷酸多态性 (single nucleotide polymorphism, SNP)，主要是指在基因组水平上由单个核苷酸的变异所引起的 DNA 序列多态性。它是可遗传的变异中最常见的一种。占有已知多态性的 90% 以上。

SNP 所表现的多态性只涉及到单个碱基的变异，这种变异可由单个碱基的转换 (transition) 或颠换 (transversion) 所引起，也可由碱基的插入或缺失所致。但通常所说的 SNP 并不包括后两

种情况。理论上讲, SNP 既可能是二等位多态性, 也可能是 3 个或 4 个等位多态性, 但实际上, 后两者非常少见, 几乎可以忽略。因此, 通常所说的 SNP 都是二等位多态性的。SNP 主要有四种类型, 一种为转换: $C \leftrightarrow T$ ($G \leftrightarrow A$), 三种为颠换: $C \leftrightarrow A$ ($G \leftrightarrow T$)、 $C \leftrightarrow G$ ($G \leftrightarrow C$)、 $T \leftrightarrow A$ ($A \leftrightarrow T$)。由于核苷酸的 5-甲基胞嘧啶残基是基因组中最易发生突变的位点, 可自发地脱去氨基而形成胸腺嘧啶, 使得四种 SNPs 在基因组中出现的频率是不同的, 在生物体中约 2/3 是: $C \leftrightarrow T$ ($G \leftrightarrow A$), 而在拟南芥中: $C \leftrightarrow T$ ($G \leftrightarrow A$) 约占 52%。其它几种变异的发生几率相似。Wang 等 (1998) 的研究也证明了这一点。在基因组 DNA 中, 任何碱基均有可能发生变异, 因此 SNP 既有可能在基因序列内, 也有可能是在基因以外的非编码序列上。而 SNPs 在单个基因或整个基因组中的分布也是不均匀的, 多存在于非转录序列中, 而位于编码区内的 SNP (coding SNP, cSNP) 则比较少。

SNP 和 Indel 是大部分等位基因差异的基础。在一些物种基因组测序完成以后, 人们已经把注意力转移到了发现和分析种内序列差异上。由于测序技术的进步, 在 DNA 水平上对许多位点上的遗传变异进行直接分析近年来成为可能。DNA 芯片、等位基因特异 PCR 等高通量的基因型标定方法使 SNP 成为了很具有吸引力的遗传标记 (Lindblad-Toh et al, 2000; Bhattaramakki et al, 2001; Syvanen, 2001)。在人类研究中, 已经发现一百多万 SNP。它们能被作为最简单的分子标记, 鉴定与之连锁的功能基因。

在一些植物物种中, 如玉米和大豆 (Zhu et al, 2003), 发掘 SNP 相对简单一些, 因为这些物种中存在高水平的种内核苷酸多态性。对于这些物种, 可以直接读出 SNP 单体型。基于单体型分析比基于单个 SNP 的分析具有更多的信息量, 在与表型的关联分析中有更强的检测能力。如果 LD 足够的话, 就可以通过全基因组扫描鉴定与表型关联的基因组区域。随着技术上的进步, SNP 和 Indel 标记在分子标记辅助育种、EST 作图和遗传、物理图谱的整合上必将起到巨大的推动作用。同时, 在检测功能基因的等位基因变异与表型之间的关联方面, SNP 的使用更是潜力巨大。近来, 人群中的 LD 水平研究揭示, LD 水平比预想得还要高。该发现表明, SNP 单体型分析, 而不是单个 SNP 分析, 会提供一条进行等位基因与表型关联分析的更有效的途径。

(二) 连锁不平衡 (Linkage Disequilibrium) 与单体型 (Haplotype)

连锁不平衡 (LD) 也被称为配子期不平衡, 配子不平衡和等位基因关联。简而言之, LD 是“不同位点间的等位基因的非随机关联”。它是多态性 (例如 SNP) 之间的相关, 是进化中突变和重组的进程一起造成的 (Flint-Garcia, 2003)。如果一个大的位点独立分离的随机交配群体, 没有选择、突变或者迁移, 那么, 多态性位点将处于连锁平衡状态 (Falconer et al, 1996)。反之, 连锁、选择、混合将增加 LD 的程度。

所谓的单体型, 从遗传标记角度讲, 是指一组遗传标记在同一染色体上紧密连锁, 并且一起向后代遗传, 不容易通过重组发生分离。

通常 2-4 个 SNP 就可以确定一个单体型。而当存在连锁不平衡时, 完整地确定一个单体型所需要的 SNP 数更少。这样的用来限定一个单体型 SNP 被称为“单体型标签 (Haplotype tags)” (Johnson, 2001)。由于单位点遗传分析的局限性, 以及 SNP 检测技术的快速发展, 单体型的研究日益受到遗传学家的重视。特别在 LD 分析中, 单体型更被看作是处于连锁不平衡状态中的一个基本结构部分。

连锁和 LD 的概念经常被混淆。尽管 LD 和连锁是相关的, 但它们本质上是不同的。连锁指

的是一条染色体上的位点由于物理连接而产生的相关遗传，而 LD 则指一个群体内等位基因间的相关。混淆的产生是由于紧密连锁会导致高水平的 LD。例如，如果两个突变在几个碱基的距离内产生，那么进化过程中它们所承受的来自选择和遗传漂变的压力是相同的。因为两个相邻碱基间的重组是很罕见的，所以存在的这些 SNP 会高度相关，从而紧密连锁导致了高水平的 LD。与之相反，在分离的染色体上的 SNP 承受了不同的选择压，独立分离，因此这些 SNP 有很低的相关性或者低水平的 LD。

有多种统计方法可用于测量 LD。Delvin&Risch (1995) 以及后来的Jorde (2000) 都对每种方法的优缺点进行了总结。其中两种最常用的方法是 r^2 和 D' 。如果一个位点上的两个等位基因A和a，另一个位点上为B和b，等位基因频率分别是 π_A 、 π_B 、 π_a 、 π_b ，那么，则四种单体型频率为 π_{AB} 、 π_{aB} 、 π_{ab} 、 π_{Ab} 。所有LD统计的最基本成分是实测到的单体型频率与期望单体型频率之间的差异。

$$D_{ab} = (\pi_{AB} - \pi_A\pi_B)$$

而这些统计方法的不同之处在于对此差异的度量。其中 r^2 的计算公式是

$$r^2 = \frac{(D_{ab})^2}{\pi_A\pi_a\pi_B\pi_b}$$

LD 的统计显著性(P 值)通常可以用两种方法计算。在每个位点有两个等位基因时，用 Fisher 精确检验法 (Fisher RA, 1935) 比较；也可以用多因子变换分析 (Weir BS, 1996) 来对一个或两个位点上的多个等位基因进行比较。

另外一种 LD 统计值 D' 的计算公式是

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A\pi_b, \pi_a\pi_B)} \text{ for } D_{ab} < 0;$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A\pi_B, \pi_a\pi_b)} \text{ for } D_{ab} > 0.$$

由于 D' 是基于实测等位基因频率进行度量。所以即使位点间的等位基因频率不同的话，该值也将在 0 和 1 之间变化。

r^2 和 D' 这两种统计值反映了LD的不同方面，在不同的条件下采用不同的统计值。在样本规模较小或者等位基因频率较低时，这两种统计方法都不能很好地使用。但它们又有各自的优点。 r^2 既包括了重组史，又包括了突变史；而 D' 仅仅能测量重组史，因此是估算重组差异的最精确的统计值。然而，在样本规模较小时， D' 会因很低的等位基因频率而高度不稳定，这是因为，由于低频率的多态性，使得发现所有四种等位基因的可能性减少了。为了得到高分辨率的检测结果，在关联分析中应采用 r^2 统计，因为它指出了标记和目标QTL之间有怎样的联系。

因为等位基因频率和位点间的重组会对 LD 产生影响，所以群体遗传中观察到的大部分过程都会在 LD 模式中得到反映。突变提供了 LD 中多态性的原材料；重组是削弱染色体内部 LD 的主要因素；而染色体间的 LD 只能通过独立分配来打破。

当单体型内部因存在高度的 LD 而表现极低的重组率，同时与相邻单体型之间的区域又经常会发生重组时，那么在遗传过程中这几个保守的单体型（每个大约数十个 kb）就会高度结构化，

成为一个个分离而不连续的小岛 (Islands), 而它们之间的区域则是重组的热点区域 (Hot Spot) (Goldstein, 2001; Daly et al, 2001; Jeffreys et al, 2001)。

1.3.2 玉米基因组序列中的序列多样性

植物基因组中蕴含着丰富的序列多样性。这种序列多样性是基因多样性的遗传基础。研究这些序列多样性, 是发掘植物中蕴含的大量等位基因资源的重要途径。同时, 这些序列多样性的剖析, 还有助于阐明所包含的功能基因在长期的进化中是否承受了选择带来的压力, 对于揭示这些基因组序列的进化史具有重要意义。然而, 长期以来, 由于研究条件所限, 这方面的研究一直进展缓慢。直到近几年, 随着测序技术的发展, 可以在 DNA 水平上对许多位点上的遗传变异进行分析, 从而得到关于序列多样性的准确信息, 对序列多样性的研究才受到越来越大的关注。在玉米上, 已经有了一些研究报道。

这些研究揭示, 在相关基因的驯化过程中, 会导致分子遗传多样性的降低。这已经在 *c1* 位点 (一个花青苷合成调控位点) 中发现 (Hanson, 1996)。Wang (1999) 和 White (1999) 近来在两个玉米基因区域内也检测到了与驯化相关的改变。在控制分枝的位点 *tb1*, 启动子区域的序列多样性显著下降, 而编码区却没有发现这一现象。而 *terminal ear1* (*te1*) 等位基因的序列也显示出存在着一定的连锁不平衡结构, 与玉米祖先相比, 在栽培玉米中仅仅检测到少量的单体型。

近来的一项研究涉及到玉米一号染色体上的 21 个位点, 表明在当地品种中存在着高水平的多样性, 但在美国自交系中多样性有所降低。而且发现多样性和重组率存在相关性, 并且在 100~200bp 范围内检测到了连锁不平衡的衰退 (Tenaillon et al, 2001)。

通过对玉米种质中存在的序列差异的频率和分布的研究, 人们认识到, 进化过程本质上是一系列基因替换过程。在此过程中, 新等位基因以单个突变的形式产生, 继而增加其频率, 最终在群体中固定。而选择在其中起到了关键作用。近来在等位基因的产生、进化史方面也有人进行了研究。

Selinger 和 Chandler (1999) 在比较玉米 *b* 基因等位位点的系统树中三个彼此分开的进化枝的分歧后, 发现在进化枝之间有强烈的分离, 表明进化枝内最近可能产生了等位位点。在三个进化枝内都既发现了 *Zea mays*. 序列, 也发现了 *Zea mays. Parviglumis* 序列。对这个发现的一个可能的解释是, 这三个进化枝可能在玉米属分离之前就开始分离了。而 Selinger 和 Chandler 认为, 也可能是 *b* 基因上游区域的核苷酸置换率很高的缘故。

White 对玉米 *te1* 位点序列进行系统发育学分析后, 并没有把所有玉米序列放入一个单一的进化枝。玉米亚种的成员, 除了 *Z. huehuetenangensis* 以外, 都混在了进化枝内。在对其它玉米基因进行分析也得到了类似结果 (Hilton et al, 1998; Golubinoff et al, 1993; Hanson et al, 1996), 并用 *Zea taxa* 中的渐渗现象或者谱系划分来解释这个现象。玉米属内各个物种不能放入单一进化枝的问题也在 *c1* 和 *Adh2* 位点被发现 (Gaut et al, 2000)。与之不同的是, *glb1* 和 *Adh1* 却显示有不同的进化史, *Zea luxurians* 等位位点形成了一个不同的进化枝 (Hilton et al, 1998; Gaut et al, 1999)。

这些研究表明, 高分歧单体型广泛分布在玉米的种质资源中, 这说明玉米属内种间的基因流是显著的。由此可以推测, 不同位点拥有不同的进化史可能与等位位点的起源不同有关, 即等位位点不仅可能起源于单一的 *Zea* 物种内部, 也可能起源于在两个或者多个物种内, 并随之产生种

间渐渗 (Wilkes, 1977)。近来有研究发现, bz 位点的一些等位位点, 它们的基因组是不同的, 基因间的重复 DNA 区段的组成也是不同的, 这项研究有力地支持了这个假说 (Fu H et al, 2002)。

1.3.3 玉米及其它植物中的连锁不平衡结构

一般认为, 由于玉米自身的高水平的遗传多样性以及高重组率, 玉米的 LD 水平相对于其它物种而言, 表现出大幅度的 LD 衰退。然而, LD 水平受很多因素影响, 比如, 由于长期以来人们针对玉米的成熟期进行的选择, 那么控制成熟期性状的主效基因的基因组区域的 LD 水平就会有显著的提高。

在玉米中, LD 的衰退程度在位点间存在显著差异。例如, 在 *Y1* 基因 (编码八氢番茄红素合成酶), 观察到显著的大约 4kb 的 LD, 但同一群体中, 在 *PSY2* 基因 (一个假定的八氢番茄红素合成酶基因) 却只观察到 1kb (Palaisa et al, 2003)。随后的研究又发现, 在 *Y1* 基因周围 LD 延伸超过 800kb, 这个 LD 水平类似于在乙醇脱氢酶基因 *Adh1* 中观察到的 500kb (Jung et al, 2004)。在这么长的物理距离上存在这么高水平的 LD, 主要是近年来玉米育种实践中的人工选择造成的。

LD 的模式因选择的群体不同也有不同 (Flint-Garcia et al, 2003)。例如, 在玉米地方品种中 LD 延伸仅在 1000bp 以内 (Tenaillon et al, 2001), 在多样化玉米自交系中是大约 2000bp (Remington et al, 2001), 但在玉米核心自交系中, 由于长期定向选择造成的狭窄的遗传基础, LD 的衰退率较低。在骨干玉米群体中, 对于 *Adh1* 和 *y1* 位点, LD 延伸至超过 100kb (Rafalski A, 2002; Palaisa, 2003), 利用同样的骨干材料, Ching 和同事也在 300-500bp 内对 18 个基因的 LD 衰退进行研究, 实际上也没有发现 LD 衰退 (Ching et al, 2002)。

在用多样性高的遗传材料进行的研究中却相继发现了 LD 的快速衰退现象。Tenaillon 等用一组多样化的玉米种质调查了染色体 1 上的 21 个位点的序列多样性 (Tenaillon et al, 2001)。LD 通过 r^2 进行测量, 平均在 200bp 内减少了将近 0.25。位点间 LD 的分析表明, 如果忽略了所有位点在同一染色体的事实, 则在位点间几乎不存在 LD。Remington 等检测了基因内的以及遍及基因组的 SNP 间的 LD, 所用材料由 102 个自交系组成, 代表了全世界公众育种中普遍使用的育种材料中的遗传多样性 (Remington et al, 2001)。在对 6 个候选基因的调查中, 基因内的 LD 快速衰退 (在 1500bp 内 $r^2 < 0.1$)。这种快速衰退也在另一组 15 个基因的研究中观察到。然而, 不同位点的 LD 衰退速率存在很大差异, *su1* 位点在 12kb 内就几乎没发生 LD 衰退。为了评估基因组中的 LD, 他们还测量了 47 个 SSR 标记间的 LD (Remington et al, 2001)。用这些标记检测到的 LD 水平比用 SNP 检测到的高。

对于 LD 模式在这些材料中的差异有几种解释: 首先, 玉米中的大部分多样性来自一个大规模的野生近缘种。大部分观察到的重组单体型可能是在该野生种驯化前产生的。因此, 不同的 LD 衰退率反映了群体瓶颈的不同水平。例如, 在从多样性当地品种向多样性自交系再到骨干自交系的改良过程, LD 程度的不断增加。因而, 由于 Tenaillon 和 Remington 所用群体分别由育种种质和当地品种组成, 其种质基础相对较宽, 所以会检测到 LD 的快速衰退。

拟南芥中的 LD 模式与玉米中的模式形成鲜明对比。由于拟南芥是高度自交植物, LD 在拟南芥中的延伸区域较远 (Nordborg, 2000)。Hagenblad (Hagenblad et al, 2002) 对开花期位点 *FRIGIDA* 的一段 400kb 区域的 14 个短片段测序。他们发现 LD 在 250kb (相当于 1cM) 内发生

了衰退。在紧密连锁的座位之间发现了强的 LD。对来自 76 个序列的遍及基因组的 163 个 SNP 的分析也揭示出, LD 在 250kb 内才发生衰退 (Nordborg et al, 2002)。

为了调查拟南芥中 LD 的影响, 利用几个近来建立的群体对抗病位点 *RPM1* 周围的 LD 进行了分析, 在这些群体中, LD 延伸了数十个 cM (大约相当于 Mb 级)。这种 LD 延伸现象可能是由于在过去 200 年里发生的重组事件很少的缘故。

甘蔗也展现了广泛存在的长距离 LD, 大约 10cM (Jannoo et al, 1999)。如果考虑现代甘蔗栽培种的育种史中的瓶颈, 那么这种现象并不令人感到惊讶。大部分现代栽培种来自于 *S.offcinarum* 和 *S.spontaneum*, 然后与 *S.offcinarum* 多次回交。由于甘蔗为营养繁殖, 致使栽培种自从第一次种间杂交以来, 减数分裂少于 10 次。本研究中, 调查了 59 个栽培种中的 RFLP 标记间的 LD。这些存在显著 LD 的成对位点中大部分连锁在同一染色体上。然而, 另外 14% 涉及不同染色体上的位点。Jannoo 和同事相信, 由于甘蔗的多倍性, LD 的整体估测被夸大了。他们也指出, 这些存在显著 LD 的成对位点中大部分来自 *S.spontaneum* 亲本, 这反映了部分同源配对的现象。

此外, 在大麦 (Kraakman et al, 2004, 2006)、火炬松 (Brown et al, 2004)、高粱 (Hamblin et al, 2004)、白杨 (Ingvarsson, 2005) 中也已有类似报道。

1.3.4 选择对玉米自交系 LD 及遗传多样性的影响

突变是自然变异的来源, 当自然变异产生后, 导致了性状的变异, 而这些变异的个体对外界环境具有不同的适应性。由于自然选择或者人工选择的存在, 具有有利性状的个体得以生存并通过繁殖传递给后代, 具有不利性状的个体会逐渐被淘汰, 在选择的长期作用下, 分布在不同地区的同一物种就可能出现性状分歧, 甚至导致新物种的形成; 而同一地区的同一物种, 会越来越趋向于性状单一化, 使群体的遗传多样性下降。

在大约 6000~9000 年前, 印第安人从大刍草中经过驯化得到了原始的玉米地方品种。到了近代, 育种者们又开始对这些地方品种进行改良, 从中选育自交系以便用来进行杂交育种, 这两个过程 (驯化和作物改良) 都涉及到了对那些控制关键的形态及农艺性状的基因的特定等位基因的人工选择。这些人工定向选择过程直接导致了这些基因的遗传多样性的降低以及连锁不平衡程度的增加。初步研究表明, 玉米中大约 2%~4% 的基因, 即约 1200 个基因, 有可能在进化过程中受到了这种选择压力而导致了多样性的变化 (Wright et al, 2005)。

最近的研究表明, 驯化和改良过程涉及到的基因并不一致 (Yamasaki et al, 2005)。据此, 这些基因分为驯化基因、改良基因和中性基因 (图 1-1)。在大刍草向地方品种的驯化过程中因承受了选择压力而导致遗传多样性降低的基因称为驯化基因; 而另一部分基因在驯化进程中由于没有承受到选择压力, 其多样性并没有降低, 但在对地方品种改良的过程中却承受了选择压力而导致多样性降低, 这些基因称为改良基因; 在整个进化过程中一直未受到选择压力的基因称为中性基因。这些基因的分类不仅在遗传上有助于更好地理解选择的遗传效应, 而且在育种上也具有重要意义。因为, 对于改良基因, 尽管在改良后的核心自交系中已经失去了大部分变异, 但通过与地方品种杂交, 构建作图群体, 利用 QTL 检测等手段还可以重现遗传多样性, 而对于驯化基因, 则需要通过跟大刍草杂交来恢复其遗传多样性。

一般认为, 植物中 LD 程度如何, 由多种因素造成, 群体的瓶颈效应和近亲交配都会增加 LD

程度 (Hudson, 2001), 但主要依赖于群体的选择。长期针对功能基因的等位变异进行正向选择, 会极大地增加了目标基因区域功能变异的 LD 水平。因此, 如果候选基因是在长期人工选择中承受了选择压力的驯化基因和改良基因, 那么就会保持高强度的 LD 结构。因而, 通过对候选基因区域的 LD 进行估测, 就可以分析选择对该基因的影响 (Clark et al, 2004), 从而推断该基因在驯化和改良过程中是否承受了相应的选择压力。

在玉米的驯化和改良过程中, 由于人工定向选择, 那些与农艺性状密切相关的位点在多样性降低的同时, LD 程度相应增加, 形成了一些高分歧的单体型。研究表明, 玉米中的 *su1* 位点在 12kb 内几乎没有发生 LD 衰退, 因而成为了一个人工选择进程中的靶标 (Whitt SR et al, 2002)。除了 *su1* 位点以外, *c1* 和 *tb1* 两个位点也因遗传多样性相对较低而被看作是玉米人工选择过程中的靶标。

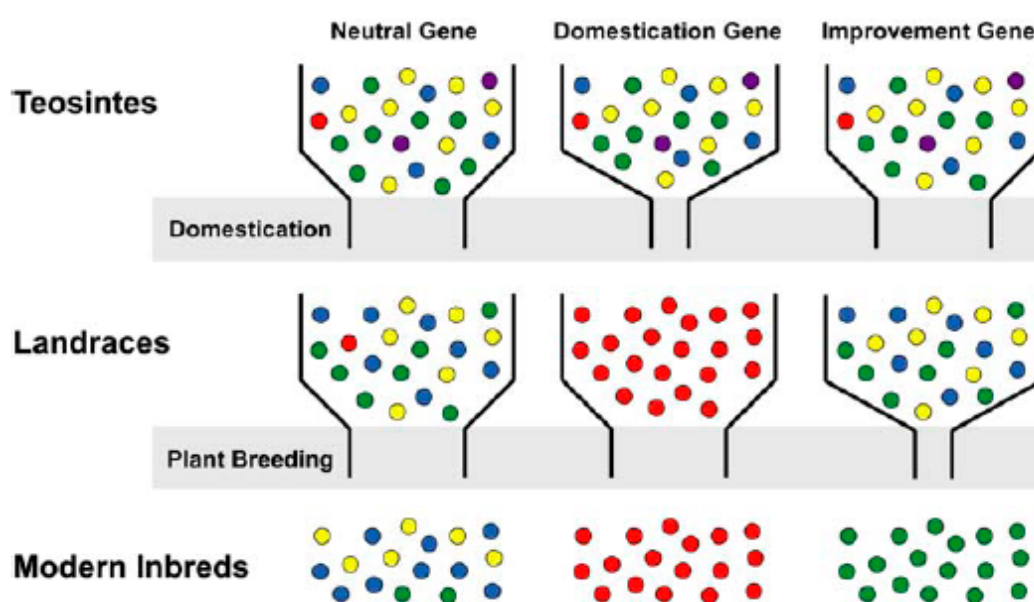


图 1-1 驯化和改良对玉米基因多样性的影响 (Yamasaki et al, 2005)

Figure 1-1 Effect of Domestication and Plant Breeding on Genetic Diversity of Maize Genes (Yamasaki et al, 2005)

对基因组中一个位点进行选择是否会影响到相邻位点的多样性, 这是从整体上了解基因组多样性的关键。在 Clark 对 *tb1* 位点的研究中发现, 针对 *tb1* 位点的选择并没有对其周围的基因组区域和相邻位点带来较大的影响; 另外在该区域, 观察到的基因间的 LD 也很低, 表明选择在该区域 LD 模式形成过程中起到的作用也很小。研究表明, 选择对基因变异和等位基因多样性的作用是因物种而异的, 同时受到重组率、群体大小、群体结构、育种系统等因素的影响。针对 *tb1* 位点的选择并没有对其周围的基因组区域和相邻位点带来较大的影响, 在某种程度上是由于在玉米驯化过程中群体的规模较大 (Buckler et al, 2001), 以及很短的世代数带来的群体瓶颈 (Eyre-Walker et al, 1998)。而且, 玉米是一个风媒异花授粉作物, 由此带来的高重组率也会减弱选择对基因组的影响。

人工选择的另一个后果是, 一些与人类需要密切相关的优良农艺性状基因, 却无法通过 QTL 定位和关联分析等方法鉴定出来, 这是由于长期针对目标性状的选择作用使得该基因失去了遗传

多样性。解决的办法只能是象上面提到的那样，用自交系去和地方品种杂交，甚至去和大刍草杂交，以在后代群体中恢复显现出其遗传多样性。

1.4 关联分析及在玉米研究中的应用

1.4.1 关联分析的概念

大部分重要的农艺性状都由复杂的多重数量性状位点控制的，通过对这些功能位点进行遗传作图和分子鉴定可以促进抗病性、耐旱性等性状的改良。

多年以来，一般用连锁分析测量位点间的遗传距离，进行数量性状作图以及QTL定位。在植物中，这些共分离分析类型中的大部分在那些高度结构化的，谱系清晰的群体中进行，如F₂群体、重组近交系群体（RIL）等。然而，这些群体有两个明显的局限性。首先，有限的重组事件导致了研究数量性状时分辨率很低。其次，对于任何给定的位点都仅仅只有两个等位基因能同时被研究。为了增加作图群体的分辨率，在几个植物物种中已经通过多次随机交配构建了大的重组近交系群体（RIL），例如玉米中的intermated B73×Mo17（IBM）群体。这些多代交配增加了重组事件的潜在数目。尽管如此，许多QTLs的分辨率仍然有几个cM，与数百个基因相对应。另外，在一个群体中对某个位点只能标定少量等位基因，这也使在各种植物中对遗传多样性大范围的检测难以进行。

关联作图（association mapping）是一项近几年才引入植物学研究的分析方法，曾广泛应用于人类遗传学研究中。该方法以自然种质群体为研究对象，能够把目标性状表型同基因的多样性结合起来进行分析，可以直接鉴定出与表型变异密切相关的具有特定功能的等位基因位点，是一种剖析复杂数量性状的很有前途的途径。连锁分析和关联作图是两种常用的研究综合性状的工具（Risch et al, 1996; Mackay, 2001）。由于关联作图探测的是更深一层的群体系谱而不是家系系谱，因而与连锁分析相比具有三大优势：拥有更高的作图分辨率；由于涉及材料广泛而使检测到的等位基因数目更多；另外，建立群体所需研究时间最短（Buckler et al, 2002; Flint-Garcia et al, 2003）。

然而，连锁分析和关联作图是互为补充的，比如在交叉验证、统计能力等方面（Wilson et al, 2004）。Hischhorn 等对这两种不同方法进行了全面的系统化比较（Hischhorn et al, 2005），植物中实施关联作图的进展也已被很好地总结过（Whitt et al, 2003），

该技术具有一些其它方法所不具备的优点：不需要构建专门的作图群体，而且所用的自然群体也无须知道其谱系背景，只要存在遗传变异即可，因而研究周期短；分辨率很高，比常用的F₂群体要高 5000 倍（Thornsberry et al, 2001）；可以同时对数十个等位基因进行检测；可以明确不同种质资源中所携带的等位基因及其对目标性状的贡献等（图 1-2, Yu et al, 2006a）。其不足之处在于由于群体遗传背景不清可能导致的假阳性较高，但这可以通过预先用SSR、SNP等标记分析其群体结构以去除假阳性。

关联分析的另一种用途是可以通过对全基因组进行扫描，从而鉴定出与目标表型有关联的基因组区域。基因组扫描过程中所需的最少位点数目取决于连锁不平衡（LD）的程度（Rafalski, 2002）。如果相关群体的连锁不平衡的数量足够的话，对 SNP 进行搜集以后，就使得以全基因组

扫描连锁不平衡为基础的方法能够被用来进行玉米中的性状剖析和基因作图。

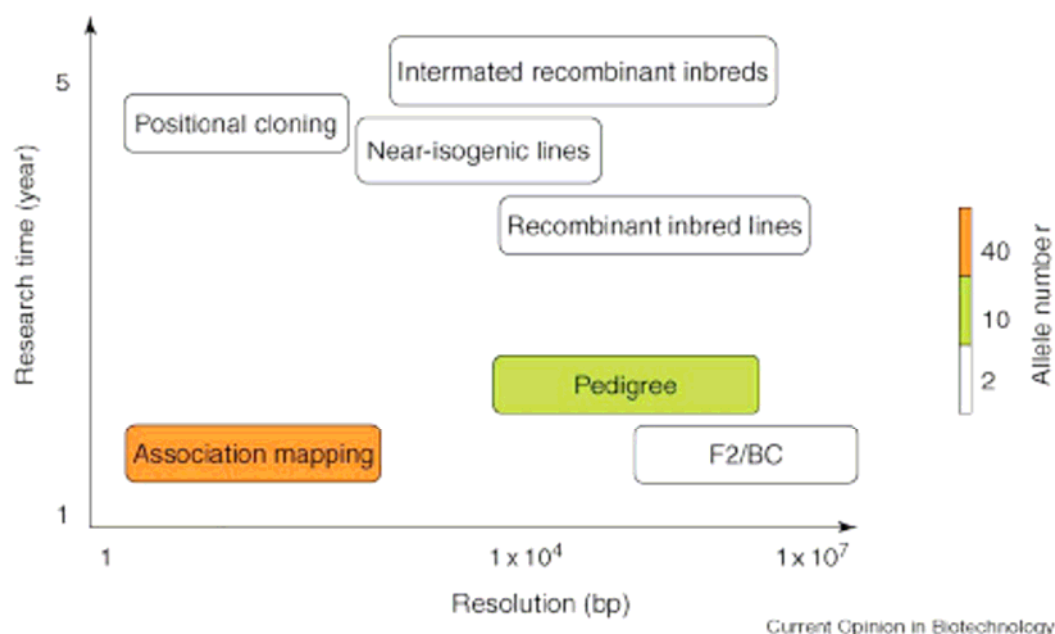


图 1-2 几种等位基因发掘途径的比较 (Yu et al, 2006a)

Figure 1-2 Comparison of various methods for identifying alleles (Yu et al, 2006a)

尽管 SSR 是一种很优秀的分子标记 (Powell et al, 1996), 然而由于它们的多等位性和导致的高信息量, 一般不用于关联分析中。另外, SSR 等位基因的序列长度的非同源相似性也是在一些应用中存在的问题, 所谓非同源相似性即指同样大小的 SSR, 其进化起源却可能不同 (Viard et al, 1998; Estoup et al, 1995), 这也使得 SSR 标记无法用于关联分析。另一方面, SNP 具有二等位性, 其期望杂合度很低, 在骨干玉米种质中只有 0.263 (Ching et al, 2002), 远远低于 SSR 的 0.77。因此, 对全基因组进行扫描, 进行等位基因与性状间的关联分析, 就需要由 SNP 提供的高密度分子标记。玉米中多态性的高频率产生了大量 SNPs 和 Indels, 都可以作为遗传标记加以利用。这些标记可以通过基因相邻序列的直接测序来发现, 或者通过对从复合基因型中获得的可利用 EST 序列进行计算机分析后获得。在每个玉米基因的邻近区域都获得几个 SNP 标记在原则上是可行的, 这几个 SNP 会组成一个亚组, 来完整地确定一个单体型。

1.4.2 关联分析与连锁不平衡 (LD) 结构

关联作图提供了相当高的分辨率, 这依赖于基因组中的 LD 结构。许多遗传和非遗传因素, 包括重组、遗传漂移、选择、交配模式和混合程度 (例如一个群体内各个亚群拥有不同的等位基因频率), 都会影响 LD 结构 (Flint-Garcia et al, 2003; Gaut et al, 2003)。关联作图的关键是功能变异位点和与之存在物理连锁的标记间的 LD, 这是关联分析技术的核心, 单体型结构的确定是准确有效地进行关联分析的基础。它们决定着关联分析的分辨率高低。群体中一定物理距离内的 LD 的衰退决定了实施关联分析所需要的标记覆盖的密度, 以及实施关联分析所必需的试验设计。例如, 如果 LD 快速衰退, 那么就需要一个较高的标记密度, 以便捕获到与功能位点距离足够近的标记。所以对于在自然群体和育种群体中应用关联分析方法, 了解 LD 的结构与区域是必需的。

一般认为，由于自交方式的存在，自花授粉作物的 LD 延伸区域要长于异花授粉作物。

LD 的本质决定了关联分析的类型。LD 的衰退率决定了是进行基因组扫描，还是进行基于候选基因的关联分析。当基因组的 LD 水平比较高时，进行关联作图需要的标记数目少，但分辨率较低，因此适合于全基因组扫描；反之，如果 LD 快速衰退，那么就需要高密度的标记来进行关联作图，那么进行全基因组扫描的难度会相当大，同时，此时较高的分辨率也适合于候选基因区域的关联分析（图 1-3；Rafalski, 2002）。在基因组扫描时，用分布于全基因组的标记来同时评估所有基因。扫描基因组所需的最少标记数取决于 LD 程度（Kruglyak L, 1999）。例如，人类基因组需要 70000 个标记，拟南芥 2000 个，种类不同的玉米当地品种 750000 个，而由于长期的定向选择，玉米骨干自交系中具有高水平的 LD，因而评价玉米骨干自交系仅仅需要 50000 个标记，。尽管在可预见的未来，技术进步会使测定必需的标记数目成为可能，但对于除拟南芥以外的其它物种而言，这仍是一个庞大的数字。然而，比基因型鉴定更难的是，大量的资源需要进行表型鉴定和统计分析。当对遍及基因组的 50000 个 SNP 测序时，因为进行了数千个独立测验，所以存在一个很大的多次测验的问题。校正这些多次测验，就需要每次独立测验都采用很低的 P 值。在基因组扫描中检验解释适度变异的 QTL 时，只有用由数千个体组成的规模较大的样本时，才能获得统计的显著性。有两个办法可以避开这个问题：选择一个具有较高水平 LD 的群体，例如一个经历了瓶颈效应的群体，或者，仅仅对候选基因区域进行分析。

植物基因组学正在开始分子和生化方法与数量遗传学的交叉，LD 将可能在其中成为关键角色。关联作图可能会在大范围的物种中鉴定和评估数量变异基础的过程中起到关键性的作用。但是进行关联分析的关键在于对大范围物种内 LD 结构的全面了解。

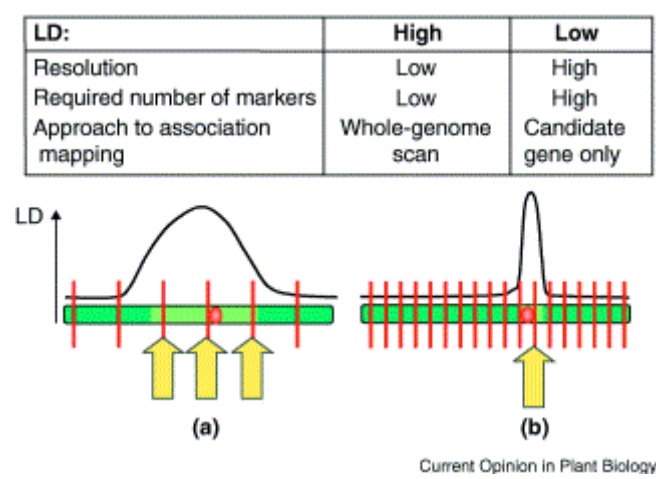


图 1-3 LD 程度和关联分析分辨率的关系（Rafalski，2002）

Figure 1-3 Relationship between the extent of LD and the resolution of association studies (Rafalski, 2002)

1.4.3 关联分析与群体结构

关联分析作为一种在人类研究中被广泛采用的方法，长期以来却很少应用于植物遗传研究中，很大程度上是由于植物中的群体结构可能带来的风险（Pritchard, 2001）。因为关联分析中所

用材料能够根据群体结构水平和群内家族谱系分成相应的亚群 (Yu et al, 2006b), 而亚群间的混合就能够造成 LD 结构, 如果在统计分析中不进行恰当的控制, 就会导致假阳性结果 (Pritchard et al, 2000a,b)。即检测到的多态性和性状间的关联不是由于序列的功能性造成的, 而是亲缘关系 (relatedness) 导致的。在玉米中由于自交系的系谱来源都有不同程度的共有比例, 因此这种假阳性存在的可能性更高 (Liu et al, 2003; Flint-Garcia et al, 2005)。其它物种中 (Nordborg et al, 2005; Garris et al, 2005) 复杂的进化及育种史也会造成这样的群体结构和复杂的系谱关系。为了降低这种风险, 在关联分析中必须考虑群体结构的影响。

然而, 如果功能等位基因的分布与群体结构高度相关, 那么对群体结构的统计控制就会导致假阴性, 特别是在小规模样本群体中。因此, 关联分析应该在大规模的独立群体中实施。两项近来的玉米研究可以用来解释上述情况。在一个对 *d8* 位点功能进行验证的研究中, 对 71 个欧洲核心自交系进行了 *d8* 序列多态性检测和开花期表型鉴定 (Andersen et al, 2005)。在没有控制群体结构影响的前提下检测到了显著关联, 然而当群体结构被控制后, 就没有再检测到关联的存在。相反, 在另外一个用 375 个玉米自交系组成的大规模关联作图群体进行的 *d8* 序列多态性与开花期之间的关联分析中, 这种关联却得到了验证 (Camus-Kulandaivelu et al, 2006)。

1.4.4 关联分析中的统计方法

对于不同的关联所用样本, 应当设计不同的统计方法来处理群体结构问题 (Yu et al, 2006b)。如果是基于家系的样本, 一般用 the Transmission Disequilibrium Test (TDT) 来研究其遗传基础, 例如在人类疾病的研究中就是如此, 而数量性状的检测却用数量 TDT (QTDT) 来进行。当研究基于群体结构的样本时, Genomic Control (GC) 和 Structured Association (SA) 是两种常用的方法, 这两种方法在人类研究中还是在植物研究中是通用的。当采用 GC 时, 先假定群体结构对所有位点的影响相当, 然后用一组随机标记来评估群体结构对测验统计产生的影响的程度 (Devlin et al, 1999)。相反地, SA 分析首先用一组随机标记来评估群体结构 (Q), 然后将这个评估结果合并到随后的统计分析中 (Pritchard et al, 1999, 2000a; Falush et al, 2003)。采用逻辑回归进行 SA 分析已经应用到先前的关联分析中 (Thornsberry et al, 2001; Wilson et al, 2004)。

目前, 已经又发展出一种一元化混合模型方法用于关联分析中, 该方法能够考虑多重水平的系谱关系 (Yu et al, 2006)。在这种方法中, 随机标记被用来评估 Q 和相应的系谱关系矩阵 K, 这两个结果都会被装配到一个混合模型框架中用于测验标记与性状间的关联。与其它方法相比, 在玉米数量性状和人类基因表达数据研究中应用这种新方法, 会提高对 I 类错误率和 II 类错误率的控制。由于该方法既考虑了群体结构又考虑了系谱关系, 因而对现有的关联分析方法是一种有力的补充。

另外, 在适当条件下, 关联分析中的似然性方法和方差分析方法能够得到类似的结果。所以在得到关联结果后可以用方差分析方法来验证关联分析结果。这种验证已经在关联研究中被采用 (Szalma et al, 2005)。

1.4.5 玉米中的关联作图

科学家们早已认识到从种质资源中蕴藏的巨大基因资源的重要价值 (Tanksley, 1997), 努力通过各种技术和途径来获得控制目标的未知基因, 并取得了很大的成绩。然而, 对于已知基因的未知等位基因的发掘, 却一直没有引起足够的重视。然而, 等位基因的差异的影响却是巨大的, 这已经在生产中得到证明。比如在第一次绿色革命时期, 对产量的大幅度提高起到关键作用的所谓“绿色革命基因”实际上就是控制株高基因的自然等位变异, 其矮秆与高秆的差异仅在编码区一个碱基的变异。

玉米基因组具有核苷酸多样性高、重组率高、连锁不平衡快速衰退等特点 (Remington et al, 2001; Tenaillon et al, 2001)。因而, 玉米成为了一种进行关联分析, 鉴定控制农艺性状相关基因的功能定位变异的模式植物 (Thornsberry et al, 2001; Rafalski and Morgante, 2004)。

2001 年, 美国科学家成功引入植物遗传学研究中, 首次发现玉米中控制株高的基因 *Dwarf8* 的多态性与开花时间的变异有关 (Thornsberry et al, 2001)。这以后, 这项技术开始受到广泛重视。在玉米上相继开展了一些研究, 如前面提到的对 *d8* 多态性与开花期之间的关联结果的两次验证研究。除此以外, 其它候选基因和涉及的性状也相继被报道。

Wilson 等对涉及籽粒淀粉生物合成的六个玉米候选基因: *ae1*、*bt2*、*sh1*、*sh2*、*su1* 和 *wx1* 进行了关联分析, 整体上, *bt2*、*sh1*、*sh2* 都和籽粒组分性状显示出显著的相关, *ae1*、*sh2* 和淀粉糊化特性表现出显著相关。*ae1*、*sh1* 都和直链淀粉水平存在相关 (Wilson et al, 2004)。另外, *su1* 也曾被发现与甜味相关 (Whitt et al, 2002)。

在对 *maysin* 和绿原酸含量的关联分析中, 共涉及到了四个位点: *pericarp color (p)*、*anthocyanin-less1 (a1)*、*colorless2 (c2)*、*white pollen1 (whp1)*, 其中 *p* 为主效位点, 且 *p* 对 *a1* 和 *c2* 具有上位性效应, 研究检测到, *a1* 启动子区域的两个序列多态性和 *maysin* 的合成存在显著关联, 而且发现在 *a1* 多态性关联显著的情况下, 可以检测到 *whp1* 启动子中的一个显著多态性位点, 该研究显示了把已知候选基因的信息合并到相关基因研究中的重要性, 并且发现上位性因素能够极大地增强关联测验的能力 (Szalma et al, 2005)。

另外, 还有一些类似的研究报道 (Fontaine and Barriere, 2003; Palaisa et al, 2003)。除玉米外, 在其它植物如马铃薯 (Gebhardt et al, 2004)、小麦 (Brescaglio et al, 2006) 上也有成功进行关联分析报告。

除了上述研究以外, 已经有一些新的候选基因成为了关联分析的目标。基于 SNP 分析, 另一项研究预测, 大约 1200 个玉米基因是在玉米驯化期间或者随后的现代育种改良过程中选择的靶标 (Wright et al, 2005)。在这些基因中, 已经发现几个与发育相关的候选基因在对玉米和野生近缘种大刍草之间的表型差异有贡献的 QTL 附近成簇存在。关联作图提供了一个有力的工具来继续这项必需的工作, 即验证这种候选基因与 QTL 之间的重叠。

还有两项玉米基因组方面的研究揭示了由于玉米自交系中序列的非同源性给关联作图造成的一些潜在的困难 (Fu et al, 2002; Brunner et al, 2005a)。在这两个研究中, 研究者发现不同的玉米自交系的逆转座子簇在定位和组成方面存在显著的不同。例如 *Helitron* 转座子造成的基因转移带来的单体型变异性 (Lai et al, 2004, 2005; Morgante et al, 2005; Brunner et al, 2005b)。这种序列非同源性降低了重组, 保留了 LD, 因此限制了关联分析的分辨率。

正如前面提到的, LD 结构是关联分析的基础, 它决定了关联作图的分辨率高低。而 LD 结构又与所选用的研究材料密切相关, 所用研究群体的遗传多样性越高, 这个群体中的 LD 衰退率

也就越高,进而,关联作图的分辨率也就越高。因而,选用一套能代表最大遗传多样性的玉米材料就成为了关联作图能否成功的关键因素之一。

迄今为止,在玉米中公开报道的关联作图群体的数目不多。第一个公开的玉米关联作图群体由 102 个多样化自交系组成 (Remington et al, 2001)。近来,另一项研究中所用群体包括了 302 个自交系,代表了世界上各个地区育种材料的多样性 (Flint-Garcia et al, 2005)。一个大规模的由 25 个多样化自交系和 B73 杂交后获得的包含 5000 个重组近交系 (RIL) 的玉米 QTL 作图群体当前正在构建中 (<http://www.panzea.org>)。这些自交系尽可能代表了玉米中的遗传多样性。该群体中的 LD 和连锁关系方面的信息也在同时被研究,因而该群体将作为一个玉米研究平台,成为连锁分析和关联作图间的桥梁。

1.5 立题意义与技术路线

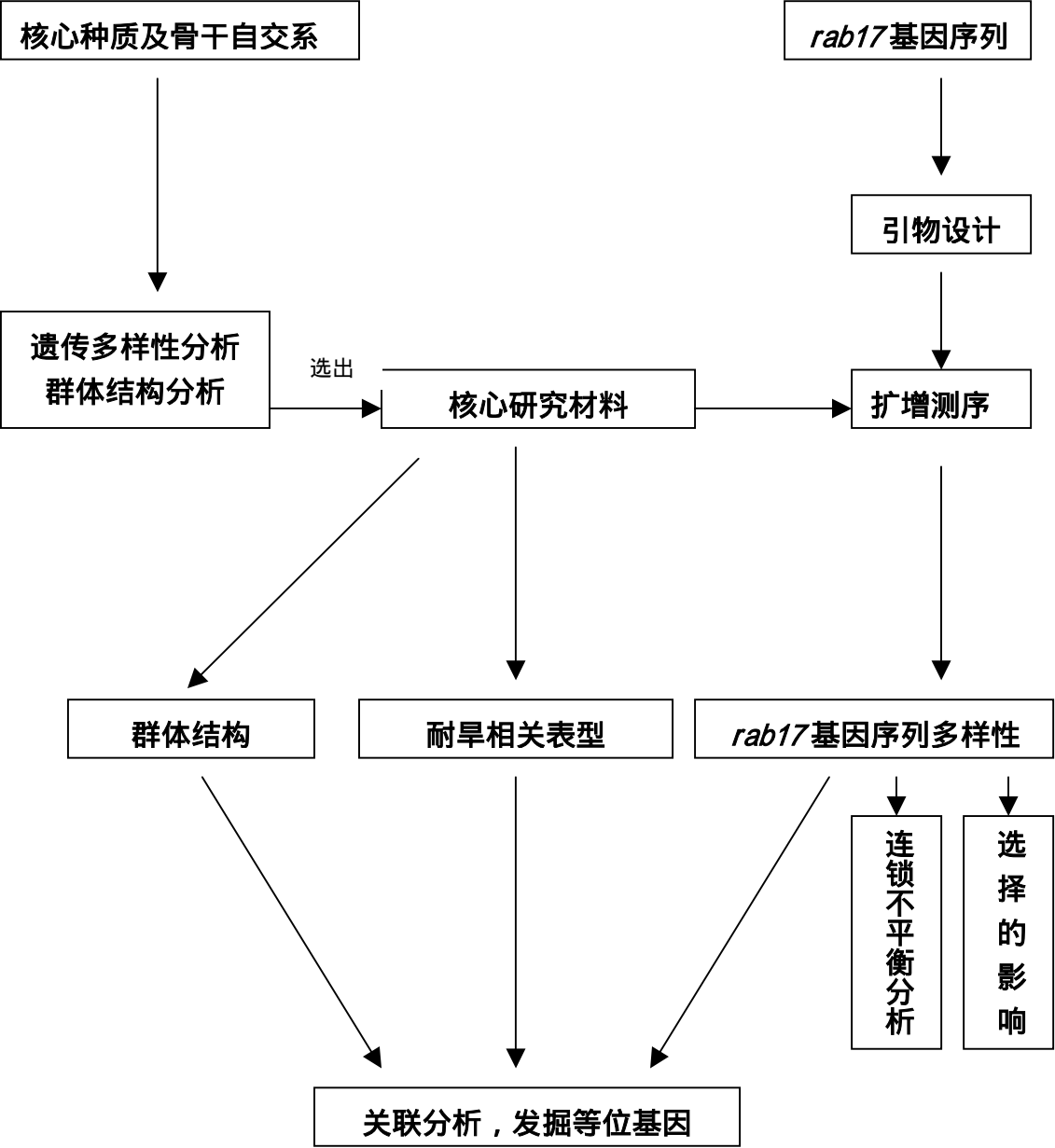
1.5.1 立题意义

干旱是限制玉米生产的重要因素之一。挖掘耐旱基因资源,提高玉米品种耐旱性水平已经成为了玉米育种者面临的一项重要课题。通过耐旱 QTL 定位、干旱诱导基因的表达等方面的研究,目前已经鉴定、分离出了一些耐旱相关候选基因和 QTLs,为我们弄清耐旱的遗传机理提供了分子基础。然而,在玉米种质资源中蕴藏着大量的耐旱相关等位基因变异,这些功能变异直接导致了耐旱相关表型的差异。因而鉴定出干旱相关基因还不能很好的解释耐旱机理,必须充分发掘出这些候选基因中蕴含的对耐旱性有正向贡献的功能等位变异。另外,了解这些功能变异对育种工作 also 具有重要意义。可以开发与这些功能变异紧密关联的 SNP 等分子标记,从而通过分子标记辅助选择的手段来针对耐旱性进行种质创新和遗传改良。关联分析方法由于可以直接鉴定出与表型变异密切相关的具有特定功能的等位基因位点,因而正是解决这一问题的有效手段。

为了提高关联分析的分辨率,需要选出一组遗传多样性高的种质材料作为研究群体。而且,由于群体结构可能导致假阳性关联结果,因而必须预先对自交系群体进行群体结构分析,以避免关联分析中的假阳性出现。

本研究以构建好的我国玉米核心种质中的自交系为基础,通过遗传多样性及群体结构分析,明确这些自交系的群体结构,并选出能够最大限度代表其遗传多样性,并包括了目前生产上常用骨干自交系的一组核心研究材料作为关联分析所用研究材料。并进一步确定耐旱相关基因 *rab17* 为候选基因,通过检测核心自交系中该基因区域内的序列多样性,分析 *rab17* 基因区域的连锁不平衡 (LD) 结构,并与耐旱相关表型鉴定数据相结合,在考虑群体结构影响的基础上,进行 *rab17* 基因多态性与耐旱表型之间的关联分析,以求发掘出与耐旱相关表型相关的 *rab17* 基因中的功能等位基因。得到的结果不仅有助于对耐旱性机制的进一步了解,而且也能对玉米耐旱性的改良提供亚基因水平上的基因资源。

1.5.2 技术路线



第二章 玉米核心自交系的遗传多样性及群体结构分析

2.1 前言

玉米是最重要的作物之一。它不仅是一个主要的食物来源，而且还是一个在植物基因组研究中的经典模式植物，这主要归功于玉米中高水平的遗传多样性 (Buckler et al, 2002, 2006)。

在我国，无论是从总产量还是种植面积来说，玉米都是仅次于水稻的第二大作物。2004 年，玉米总产量更高达 131, 860 千吨 (中国农业年鉴, 2005)。玉米大规模增产的一个重要因素是从 20 世纪 70 年代以来单交种的广泛应用 (黎裕, 1998)。而育种家们选育出的大量优良自交系则是推动杂交育种生产更上一层楼的种质基础。

在过去的二十年里，我国玉米种质的数量迅速增加。这些种质资源既包括地方品种和自交系，也包括许多从国外引入的一些材料。其中，大量的自交系，如黄早四、Mo17、自 330、478、丹 340、E28 等，都在我国玉米自交系选育和杂交育种中发挥了巨大作用 (黎裕, 1998)。这些自交系总体上可分为 Lancaster、Reid、塘四平头和旅大红骨四大杂种优势群 (王懿波等, 1998)。而近年来，一些来自先锋的杂交种，如杂交种 P78599、3382、3147 等，在国内育种工作中扮演了越来越重要的角色，这些自交系被认为是一个独立的杂种优势群，并被称为 P 群 (王元东等, 2004)。

为了更好的评估和利用这些自交系，构建核心种质是一种有效途径 (Brown, 1989)。依据用种质资源中最少的样本最大限度地来代表基础种质遗传多样性的原则，本实验室在以前的工作中已经构建了我国的玉米核心种质，包括 951 份地方品种和 242 份自交系 (黎裕等, 2004)。这些核心种质资源在进一步的基因多样性研究和基因发掘中将发挥重要作用。然而对这些核心种质还需要进行深入的遗传多样性分析。同时，玉米的杂交育种工作依赖于对玉米种质资源的遗传结构的深入了解。通过群体结构分析方法 (Pritchard et al, 2000a,b)，可以对各种自交系的谱系关系及遗传背景有更深入的理解。目前在玉米 (Liu et al, 2003) 及其野生近缘种大刍草 (Fukunaga et al, 2005)，以及水稻 (Garris et al, 2005) 中都已经有了类似的关于遗传多样性及群体结构分析的报道。

另外，在本研究中，为了随后的关联分析，也需要从大量核心自交系中进一步选出一组能最大限度地代表遗传多样性的核心研究材料 (Mini Core set) 作为关联分析的研究材料。因而，本部分研究的目的是：(1) 对构建好的我国玉米核心种质中的自交系的遗传多样性进行估测，以进一步阐明其系谱关系；(2) 分析其群体结构，了解这组自交系间的遗传相似程度；(3) 选出一组能够最大限度地代表遗传多样性的核心研究材料作为研究材料用于下一步的关联研究。

2.2 材料与方法

2.2.1 实验材料

本研究以本实验室在先前工作中构建的玉米核心种质 (黎裕等, 2004) 中所包括的所有 242 份自交系为基础，再加入一些近年来在杂交育种中应用的骨干自交系，共计 288 份自交系作为实验材料，材料清单及系谱来源见表 2-1。这些自交系代表了国内广泛的地理分布，另外还包括引自世界各地的一些种质。大部分自交系的种子来自国家种质库，一部分由育种工作者提供。

表 2-1 核心自交系的系谱及在四个种质来源中的遗传相似性比例

Table 2-1 pedigrees and membership in four germplasm sources for core inbred lines

ID	自交系名称	系谱来源	遗传相似性比例				先前研究中所	据群体结构划分
			Reid	P	塘四平头	Lancaster	属杂种优势群	结果
1	吉 63	(127-32×铁 84) ×(威 24×威 20)	0.558	0.105	0.060	0.276		
2	M14	BR10×R8	0.315	0.181	0.077	0.427		
3	矮金 525	武陟矮×金皇后	0.050	0.068	0.312	0.570		
4	C103	Noah Hershey	0.061	0.162	0.382	0.395	Lancaster	
5	多 229	濮阳石灰婆×小黄玉米	0.018	0.076	0.883	0.024		塘四平头
6	自 330	Oh43×可利 67	0.444	0.130	0.335	0.091	Lancaster	
7	塘四平头	塘四平头	0.104	0.027	0.825	0.045	塘四平头	塘四平头
8	获白	获嘉白马牙	0.040	0.250	0.654	0.056		
9	Mo17	C103×187-2	0.065	0.341	0.039	0.555	Lancaster	
10	旅 28	旅大红骨	0.048	0.388	0.326	0.239	旅大红骨	
11	原武 02	武 105×多 229	0.041	0.135	0.797	0.028	Reid	塘四平头
12	威风 322	W59E×风可	0.184	0.245	0.167	0.405		
13	华 160	华东二号	0.257	0.379	0.218	0.146		
14	太 183	混选 1 号	0.087	0.128	0.611	0.174		
15	黄早四	塘四平头	0.013	0.181	0.758	0.047	塘四平头	塘四平头
16	E28	(旅 9 宽×A619Ht) ×旅 9 宽	0.287	0.195	0.086	0.432	旅大红骨	
17	掖 8112	国外杂交种	0.926	0.013	0.044	0.016	Reid	Reid
18	5003	3147	0.818	0.024	0.132	0.026	Reid	Reid
19	掖 478	U8112×5003	0.920	0.039	0.027	0.014	Reid	Reid
20	丹 340	白轴旅 9×有稃玉米	0.592	0.138	0.096	0.174	旅大红骨	
21	444	A619×黄早四	0.032	0.012	0.942	0.014	塘四平头	塘四平头
22	X178	北美单交种 P	0.014	0.961	0.009	0.016	P	P
23	黄野四 3	(黄早四×野鸡红) ×黄早四	0.018	0.013	0.924	0.046	塘四平头	塘四平头

24	获唐黄 17	获唐白 42×白 917×Mo17Ht1	0.016	0.018	0.024	0.942	Lancaster	Lancaster
25	综 3	综合种	0.558	0.114	0.233	0.096		
26	郑 58	掖 478	0.632	0.205	0.027	0.137	Reid	
27	昌 7-2	潍 59×黄早四	0.031	0.054	0.455	0.460	塘四平头	
28	鲁原 92	原齐 123×1137	0.120	0.332	0.199	0.349		
29	齐 319	78599	0.030	0.571	0.351	0.048	P	
30	DH65232	DH6327×5003	0.749	0.017	0.103	0.131	Reid	
31	K12	黄早四×淮春	0.139	0.017	0.567	0.277	塘四平头	
32	沈 137	78599	0.020	0.945	0.014	0.020	P	P
33	繁荣 2	八道渠	0.035	0.028	0.826	0.111		塘四平头
34	中系 042	Across、7741、HEO2	0.168	0.025	0.230	0.577		
35	中 128	中 7490×尼 2118	0.112	0.078	0.326	0.484		
36	中 451	选自 R4513	0.398	0.254	0.041	0.307		
37	中黄 64	杂交种 64	0.069	0.229	0.526	0.175		
38	早 49	TW×peno×早 22	0.057	0.026	0.872	0.045		塘四平头
39	资玉 3	鲁资玉 3 号	0.139	0.304	0.074	0.484		
40	广优 5	广阔群体优株	0.210	0.035	0.650	0.105		
41	M3736	选自美国爆裂玉米	0.361	0.059	0.039	0.540		
42	京糯 2	不详	0.105	0.361	0.508	0.025		
43	遗 12	不详	0.704	0.057	0.110	0.129		
44	遗 49	不详	0.672	0.056	0.076	0.196		
45	遗 113	不详	0.224	0.217	0.385	0.174		
46	冀 69	2569×矮永 28	0.164	0.110	0.702	0.024		
47	张系 28	NODAK301	0.080	0.219	0.057	0.644		
48	张 21	多交种	0.020	0.041	0.503	0.436		
49	张金 6	金皇后	0.447	0.296	0.124	0.132		
50	多行金	金皇后	0.208	0.065	0.612	0.114		

51	冀甜 15	国内甜杂交种	0.255	0.074	0.202	0.468		
52	承 435	不详	0.728	0.019	0.201	0.051		
53	D729	D 团综合种	0.077	0.014	0.047	0.862		Lancaster
54	白 197	海东 013、黄早四	0.036	0.033	0.799	0.131	塘四平头	塘四平头
55	冀 432	2578、矮永 28 等 8 个系	0.170	0.114	0.558	0.159		
56	79028	昌×爱 90	0.031	0.269	0.030	0.671		
57	邢 26	黄 61、邢 11、金 03	0.033	0.132	0.646	0.188		
58	H2	美国杂交种	0.162	0.562	0.054	0.221		
59	邯 102	河南多穗	0.092	0.300	0.416	0.193		
60	太 184	混选 1 号	0.095	0.155	0.044	0.706		
61	太 1/Lg	Mo17×太 1/lg	0.060	0.020	0.027	0.893	Lancaster	Lancaster
62	辐射 A	不详	0.024	0.183	0.742	0.051		
63	紫多穗	不详	0.031	0.440	0.118	0.410		
64	枣 123	农家种	0.071	0.048	0.808	0.072		塘四平头
65	晋穗 54	自 330 变异	0.386	0.304	0.275	0.034	Lancaster	
66	GB	天津白玉米	0.050	0.532	0.384	0.033		
67	84-126-15-1	不详	0.139	0.334	0.272	0.254		
68	X.L9010-3/O2	5105×opaque2	0.386	0.158	0.067	0.389		
69	XZ19	金 0-14×忻 335	0.312	0.031	0.045	0.611		
70	VG85-5	5003×关 17	0.760	0.096	0.056	0.088		Reid
71	选 6	H95×Mo17	0.032	0.033	0.045	0.890	Lancaster	Lancaster
72	东 623/O2	不详	0.085	0.390	0.253	0.272		
73	13A/O2	不详	0.068	0.352	0.526	0.053		
74	85 黄 14	W59E×划筒	0.134	0.165	0.213	0.489		
75	85 白 64	85 黄 11×武 403	0.098	0.041	0.811	0.050		塘四平头
76	白鹤 43	白鹤	0.092	0.475	0.211	0.223		
77	337	(H95, 旅 28)BC	0.236	0.367	0.164	0.233	旅大红骨	

78	朝鲜白	朝阳白玉米	0.212	0.030	0.741	0.016		
79	抚 96	南充 5×矮 13-31	0.019	0.139	0.179	0.663		
80	抚 85	335 杂株	0.059	0.018	0.884	0.038		塘四平头
81	3514	齐 35、辽轮 814	0.122	0.061	0.224	0.592		
82	C107	美国	0.137	0.163	0.212	0.488		
83	丹 3116	107、E28	0.078	0.307	0.555	0.060	旅大红骨	
84	辽 2202	美国杂交种	0.574	0.201	0.078	0.147		
85	辽 7794	7922×8112/(32×5003)	0.959	0.018	0.012	0.012	Reid	Reid
86	吉 833	吉综 B717	0.070	0.511	0.330	0.090		
87	吉 846	(Mo17×吉 63)	0.269	0.026	0.061	0.643		
88	吉 880	哲 B77A	0.112	0.047	0.339	0.502		
89	九 03	不详	0.167	0.155	0.450	0.228		
90	吉 994	吉 63×Mo17	0.017	0.041	0.011	0.931	Lancaster	Lancaster
91	427	铁 133×Mo17	0.323	0.058	0.066	0.552	Lancaster	
92	D801	国外杂交种	0.333	0.484	0.102	0.081		
93	96201	美 P6 群体	0.079	0.560	0.270	0.091		
94	甘 41	不详	0.022	0.024	0.305	0.649		
95	辐 842	甸 11×意 210/辐 746	0.063	0.028	0.169	0.741		
96	辐 8502	吉单 118	0.510	0.042	0.154	0.294		
97	辐 8701	辐 746×F542	0.552	0.038	0.102	0.308		
98	龙抗 15	RC103×长 3	0.900	0.061	0.019	0.020		Reid
99	龙抗 279	不详	0.403	0.140	0.198	0.259		
100	2005	不详	0.203	0.102	0.578	0.116		
101	91af361	国外杂交种	0.079	0.225	0.641	0.056		
102	90Ha5O2	国外杂交种	0.043	0.458	0.170	0.329		
103	H205	国外杂交种	0.236	0.317	0.068	0.379		
104	5 黄苏 A	选自驯化杂交种	0.033	0.117	0.666	0.185		

105	87-20	不详	0.196	0.125	0.350	0.328		
106	85-25	选自驯化杂交种	0.228	0.154	0.558	0.061		
107	壳大 1	壳 1-1×大秋 36	0.082	0.761	0.084	0.072		P
108	金糯	金皇后×糯玉米	0.106	0.297	0.538	0.059		
109	鹿 65	北金 14×330	0.033	0.363	0.542	0.062		
110	皖系 23	Va35×B73	0.510	0.111	0.057	0.322	Reid	
111	岩 11	86-11	0.116	0.100	0.663	0.121		
112	岩赖(2)	奥帕克 2 号	0.045	0.294	0.298	0.362		
113	岩赖(3)	奥帕克 2 号	0.032	0.270	0.588	0.110		
114	岩 38	鲁单 38	0.020	0.029	0.824	0.128		塘四平头
115	岩 156	623-10	0.359	0.176	0.391	0.073		
116	岩 103	原武 02	0.021	0.201	0.611	0.166		
117	岩 172	浦城矮玉米	0.062	0.255	0.582	0.102		
118	黄小 162	黄县小粒黄	0.111	0.092	0.553	0.244		
119	凤可 1	可利 67×Oh43	0.039	0.095	0.752	0.113		塘四平头
120	107	XL80	0.127	0.029	0.025	0.820		Lancaster
121	黄金 63	黄小 162×金 131	0.168	0.332	0.450	0.050		
122	齐 404	黄早四×衡白 522	0.240	0.038	0.627	0.094	塘四平头	
123	齐 35	Mo17×太 183	0.017	0.038	0.278	0.667	Lancaster	
124	双 105	不详	0.276	0.020	0.500	0.204		
125	8112	国外杂交种	0.957	0.018	0.012	0.013	Reid	Reid
126	1029	XL80	0.309	0.119	0.035	0.537		
127	齐 318	78599	0.020	0.954	0.010	0.015	P	P
128	齐 410	8112×白棒粘	0.489	0.067	0.408	0.037	Reid	
129	200-24-13413	群体	0.385	0.067	0.357	0.190		
130	二南 24	南 107×南 55	0.090	0.028	0.018	0.864	Lancaster	Lancaster
131	无唐 448	塘四平头	0.062	0.105	0.717	0.116	塘四平头	

132	关 17-1	关 73×Mo17	0.017	0.023	0.014	0.946	Lancaster	Lancaster
133	32	3382	0.955	0.020	0.009	0.016		Reid
134	郑白 11	墨白	0.079	0.023	0.099	0.799		Lancaster
135	焦 05	旅 9 分离选	0.144	0.090	0.271	0.494	旅大红骨	
136	006	不详	0.344	0.025	0.022	0.609		
137	白 U8112	U8112 分离选	0.213	0.029	0.260	0.498	Reid	
138	77	双跃 147	0.026	0.026	0.884	0.064		塘四平头
139	长 7 大咸 1	不详	0.025	0.065	0.540	0.370		
140	总统 3	不详	0.562	0.031	0.111	0.296		
141	79131	金 261.白 313.辽金 23-1.威 22	0.294	0.048	0.551	0.107		
142	沙 1	金皇玉	0.079	0.111	0.258	0.553		
143	大 255	不详	0.029	0.052	0.491	0.428		
144	897	112×黄早四	0.098	0.033	0.843	0.026	塘四平头	塘四平头
145	112/O2	自 112×中系 091/O2	0.248	0.279	0.117	0.356		
146	中二/O2	中单 2 号	0.305	0.239	0.195	0.261		
147	75-24	恩单 2 号	0.032	0.237	0.537	0.194		
148	粤 89B-1	5003/抗旱大粒黄	0.263	0.421	0.259	0.057		
149	粤 267-1-1	5003/抗旱大粒黄	0.330	0.136	0.218	0.316		
150	粤 89E4-2	武 102×黄早四	0.024	0.188	0.713	0.075	塘四平头	
151	粤 89E5-3	武 102×黄早四	0.022	0.209	0.742	0.027	塘四平头	
152	粤 89A12-1	不详	0.128	0.488	0.305	0.080		
153	粤 20-3	掖单 13 变异	0.096	0.041	0.519	0.343		
154	粤 22-1	掖单 13 变异	0.129	0.064	0.318	0.490		
155	粤 39-4	四张三×47	0.035	0.140	0.737	0.088		
156	钦 8-22-1	不详	0.645	0.171	0.088	0.096		
157	阿蒂那	希腊	0.216	0.069	0.053	0.662		
158	双 M9B-1	CIMMYT	0.017	0.364	0.340	0.279		

159	237	获白×秋 23	0.060	0.046	0.038	0.856		Lancaster
160	秋 23	南充秋子	0.352	0.045	0.552	0.052		
161	材 11-8	成单 5 号	0.285	0.047	0.197	0.471		
162	75-14 高	75-14 杂株	0.141	0.037	0.769	0.052		塘四平头
163	材 48-1111	中单 2 号×合川拖翻山	0.039	0.330	0.073	0.558		
164	48-2	选自综合种	0.322	0.122	0.523	0.033		
165	综西 241	普通玉米×大刍草	0.063	0.115	0.198	0.624		
166	030-1	高赖玉米杂交种	0.085	0.371	0.072	0.472		
167	Pop-1111	美国爆裂玉米	0.402	0.042	0.115	0.441		
168	宗 548-1521	综 5×材 48	0.456	0.091	0.183	0.269		
169	关岭花山	关岭花山	0.075	0.102	0.332	0.491		
170	707	77 异交株	0.049	0.150	0.762	0.038		塘四平头
171	Q2509	墨西哥群体	0.241	0.194	0.541	0.024		
172	美爆	美国爆粒种	0.332	0.056	0.451	0.160		
173	云 113	(泽大黄×金皇后×晋杂 1 号)	0.134	0.138	0.556	0.172		
174	武 202	武 403×Bup29	0.461	0.086	0.367	0.086		
175	武 302	野鸡红	0.046	0.253	0.618	0.083		
176	C 武 215B	RC264Ht1×78(9-2)	0.281	0.103	0.428	0.188		
177	武 312	78479	0.333	0.428	0.186	0.053	P	
178	武 308	武 109 杂株	0.274	0.067	0.117	0.542		
179	L005	陕综 5 号	0.017	0.948	0.014	0.021		P
180	S001	黄早四×302D	0.469	0.078	0.429	0.024	塘四平头	
181	L069	78599	0.011	0.960	0.016	0.014	P	P
182	55113-3-3-5	ZPDC551B	0.206	0.239	0.337	0.218		
183	峰 263	华亭白马牙	0.091	0.145	0.209	0.556		
184	峰 273	春杂 1 号	0.135	0.069	0.583	0.213		
185	武系 205	辽东白	0.047	0.223	0.609	0.122		

186	酒 138-5	330×马尔 102	0.346	0.048	0.218	0.388	Lancaster
187	宁 37	延 753×7601	0.265	0.193	0.424	0.118	
188	宁 45	矮旅 28/辽粤 606+Va35+宁 8+塘 四平头+矮自 330	0.517	0.156	0.160	0.167	
189	宁 55	中单 2 号×宁 34/ (长 220×B68) ×宁 34	0.464	0.114	0.385	0.037	
190	白色爆裂	不详	0.232	0.317	0.337	0.114	
191	中引 15	日本	0.478	0.051	0.029	0.442	
192	BC4B	南斯拉夫	0.902	0.031	0.033	0.033	Reid
193	72-125	意大利	0.017	0.010	0.963	0.011	塘四平头
194	F520	法国	0.048	0.022	0.070	0.860	Lancaster
195	F522	法国	0.536	0.104	0.082	0.278	
196	MP 704	尼日利亚	0.542	0.118	0.020	0.320	
197	CH705-8	加拿大	0.304	0.036	0.630	0.029	
198	A509	美国	0.124	0.060	0.597	0.219	
199	C103R	美国	0.073	0.045	0.109	0.773	Lancaster
200	L105	美国	0.032	0.114	0.482	0.372	
201	RCL 64	美国	0.260	0.302	0.293	0.145	
202	WD	美国	0.159	0.089	0.586	0.167	
203	407sel11:182Ht4	美国	0.270	0.431	0.259	0.040	
204	3025	美国	0.076	0.128	0.045	0.751	Lancaster
205	FR218	美国	0.732	0.017	0.119	0.132	
206	P39/su	美国	0.174	0.226	0.060	0.539	
207	Za5125/su	美国	0.243	0.106	0.091	0.560	
208	CML58	墨西哥	0.016	0.204	0.395	0.384	
209	CML67	墨西哥	0.149	0.349	0.430	0.071	
210	CML125	墨西哥	0.074	0.406	0.063	0.457	

211	大 MO	不详	0.012	0.034	0.041	0.913		Lancaster
212	京糯 3	不详	0.074	0.236	0.450	0.240		
213	303WX	不详	0.338	0.352	0.084	0.226		
214	遗 15	不详	0.782	0.108	0.067	0.044		Reid
215	中自 05	78599	0.136	0.664	0.078	0.122	P	
216	遗 72	不详	0.374	0.020	0.574	0.032		
217	遗 50	不详	0.491	0.204	0.161	0.143		
218	遗 88	不详	0.657	0.068	0.110	0.165		
219	遗 55	不详	0.487	0.281	0.149	0.083		
220	遗 67	不详	0.355	0.298	0.084	0.264		
221	遗 85	不详	0.055	0.420	0.440	0.085		
222	478	U8112×5003	0.857	0.112	0.014	0.018	Reid	Reid
223	84112	不详	0.207	0.308	0.259	0.227		
224	92 黄 7	W153×C103	0.066	0.251	0.285	0.398		
225	Sh21	不详	0.047	0.883	0.048	0.022		P
226	海 9-21	不详	0.254	0.489	0.178	0.079		
227	太系 112	不详	0.408	0.074	0.262	0.256		
228	太系 113	不详	0.038	0.278	0.070	0.615		
229	运 402	不详	0.037	0.116	0.823	0.024		塘四平头
230	运 422	晋单 34	0.339	0.027	0.421	0.213		
231	750	N9×553	0.434	0.107	0.431	0.028		
232	7152	粗 61×330	0.431	0.046	0.495	0.029		
233	5168	白粒杂交选育	0.344	0.229	0.166	0.261		
234	785	330×黄早四	0.662	0.294	0.026	0.018		
235	91 黄 4	旅 28×赤 120	0.496	0.309	0.029	0.166	旅大红骨	
236	91 黄 5	甸 11A×旅 9	0.692	0.128	0.027	0.152	旅大红骨	
237	92 黄 40	853×Mo17	0.145	0.236	0.091	0.528	Lancaster	

238	y673	英 64×吉 63	0.092	0.045	0.036	0.828		Lancaster
239	赤黄 29	旅 9×Ci7	0.762	0.055	0.093	0.091	旅大红骨	Reid
240	赤黄 32	旅 9×Ci7	0.034	0.041	0.334	0.592	旅大红骨	
241	兴 60	478×J852	0.817	0.110	0.015	0.057	Reid	Reid
242	兴 83	144×147	0.468	0.119	0.045	0.368		
243	吉资 5A	5003×黄早四	0.253	0.010	0.726	0.011	塘四平头	
244	辽 2203	美国杂交种	0.529	0.238	0.054	0.179	P	
245	辽 2204	美国杂交种	0.593	0.270	0.066	0.072	P	
246	辽 3398	综合种	0.215	0.230	0.046	0.509		
247	辽 3409	综合种	0.950	0.014	0.016	0.019		Reid
248	辽 3413	综合种	0.434	0.110	0.021	0.435		
249	辽 4358	7922×3352	0.339	0.149	0.076	0.436		
250	辽 4366	8112×3352	0.787	0.132	0.045	0.036	Reid	Reid
251	辽 4476	7922×3358	0.779	0.047	0.079	0.095		Reid
252	辽 5031	综合种	0.257	0.059	0.077	0.607		
253	辽 5045	N-46×N-22	0.714	0.031	0.027	0.227		
254	辽 5110	7922×5003	0.854	0.017	0.094	0.035	Reid	Reid
255	吉 1015	吉综 D	0.159	0.182	0.144	0.515		
256	吉 1025	不详	0.351	0.076	0.093	0.481		
257	吉 992	吉 63×Mo17	0.043	0.024	0.026	0.907	Lancaster	Lancaster
258	吉 996	吉 63×Mo17	0.091	0.017	0.020	0.872	Lancaster	Lancaster
259	新 846	吉 8462×Mo17	0.339	0.154	0.024	0.483	Lancaster	
260	吉七 8	7884—7HT×E28	0.362	0.143	0.034	0.461	旅大红骨	
261	甸骨 11A	桦甸红骨子	0.020	0.037	0.138	0.805		Lancaster
262	龙抗 1	辽 1311(甸 11×大风 22)	0.031	0.033	0.384	0.553		
263	84	W641A×黄早四	0.170	0.030	0.727	0.072	塘四平头	
264	Lo1067	Pioneer 3780×Lo87602	0.681	0.063	0.181	0.075		

265	武 134	78599	0.059	0.871	0.040	0.030	P	P
266	丹 3502	POSC205	0.164	0.045	0.410	0.382		
267	64C103	Bup64×C103	0.125	0.056	0.054	0.765	Lancaster	Lancaster
268	黄 C	(黄小 162×自 330/O2) ×Tuxpeno	0.618	0.191	0.094	0.098		
269	De811	[B68×[B73Ht×(C103×Mp3204 双 交种)Sel.]]	0.808	0.010	0.045	0.137		Reid
270	A632	(Mt42×B14) ×B14(3)	0.889	0.017	0.022	0.073	Reid	Reid
271	Lo1125	Pioneer ADA (PR3374)	0.686	0.163	0.021	0.130		
272	单 330	不详	0.318	0.377	0.076	0.230		
273	HR962	黄早四	0.011	0.012	0.965	0.012	塘四平头	塘四平头
274	IAB078	意大利	0.197	0.124	0.065	0.613		
275	K36	不详	0.528	0.401	0.020	0.051		
276	遵 90110	78599	0.037	0.928	0.012	0.023	P	P
277	CN165	不详	0.097	0.263	0.551	0.090		
278	P138	78599	0.017	0.950	0.011	0.021	P	P
279	沈 135	78599	0.019	0.953	0.014	0.015	P	P
280	H21	黄早四×H84	0.381	0.024	0.566	0.030	塘四平头	
281	7922	美国杂交种 3382	0.859	0.024	0.014	0.104		Reid
282	2196	不详	0.193	0.749	0.029	0.029		
283	92—3	不详	0.399	0.531	0.046	0.025		
284	Q114-1	不详	0.022	0.105	0.056	0.817		Lancaster
285	资 3R	不详	0.448	0.046	0.381	0.125		
286	488	8112×5003	0.958	0.013	0.014	0.015	Reid	Reid
287	承 18	顶上玉米×(公 70×60—22)	0.150	0.023	0.260	0.568		
288	四 444	A619×黄早四	0.024	0.024	0.891	0.062	塘四平头	塘四平头

2.2.2 基因型鉴定

本研究采用 TP-M13 荧光标定方法进行基因型鉴定 (Schuelke et al. 2000)。

具体实验流程如下：

1) 提取 DNA：

在玉米幼苗 5 叶期时，剪取幼嫩叶片 3-5g。采用 CTAB 法提取基因组 DNA (Saghai-Maroo et al, 1984)。具体步骤如下：

具体如下：

- a) 将叶片在液氮条件下迅速研磨成粉末。
- b) 将粉末放入到 15ml 的离心管中，然后加入缓冲液约 7ml，充分混匀。
- c) 将离心管置于 65℃ 水浴中 1.5-2h，水浴过程中温和地混匀几次。
- d) 取出离心管，每管加入等体积的氯仿-异戊醇 (V:V=24:1) 溶液，温和地混匀 30min 后离心，8000rpm，20min。
- e) 用塑料吸管小心吸出上清液移入另一离心管中，加入等体积氯仿-异戊醇 (V:V=24:1)，混匀后离心 8000rpm，20min。
- f) 将上清液移入另一离心管中，加入等体积预冷的异丙醇，轻轻混匀，静止一段时间，然后用玻璃钩勾出 DNA，并用 70% 的乙醇冲洗 2-3 次。
- g) 勾出 DNA，室温短时间干燥至无乙醇味，将其溶于含 1×TE 500μl 的 1.5ml 离心管，加入 6μl RNase 溶液在 37℃ 恒温箱中保温 1h。
- h) 每管加入等体积的氯仿-异戊醇 (V:V=24:1)，轻轻混匀，8000 rpm，离心 20min。
- i) 取上清液，加入等体积氯仿-异戊醇 (V:V=24:1)，轻轻混匀后，8000 rpm，离心 20min。
- j) 取上清液，加入 1/10 体积的 3M 醋酸钠，混匀后，加入 2 倍体积预冷的无水乙醇 (或 95% 乙醇)。
- k) 轻轻混匀静止一会儿 (或放置于 -20℃)，用移液器吸头勾出 DNA，用 70% 乙醇冲洗 2-3 次后，空气中干燥过夜。
- l) 依所提 DNA 量加入适量的 1×TE 溶解 DNA。
- m) 检测 DNA 的浓度和纯度：用 Pharmacia 公司的 LKB ultrospec III 紫外分光光度计测定，并用少量样品通过 0.8% Agarose 胶电泳测定 DNA 的质量。
- n) 取适量 DNA 原液稀释成 10ng/μl 的使用浓度，在 4℃ 下保存备用；其余 DNA 原液在 -20℃ 下保存备用。

附：提取缓冲液配方：2% CTAB

1 M Tris.Cl

5 M NaCl

0.5 M EDTA

10 g PVPP/1L

100×TE 配方：800ml H₂O 中加 121.1g Tris，37.2g EDTA·Na₂·2H₂O，用 HCl 调 pH 至 8.0，定容至 1000 ml，灭菌。

3 M 醋酸钠配方 (pH5.2)：600ml H₂O 中加入 408.24g NaAc·3H₂O，溶解后，用冰

醋酸调pH至 5.2, 然后定容至 1000 ml, 灭菌。

RNase配方: 用超纯水溶解RNase, 终浓度 10mg/ml, 煮沸 20min, 缓慢冷却, 分装, -20℃下保存备用。

0.8% Agarose 胶配方: 0.8 g Agarose加入至 100ml 1×TAE中, 加热至沸腾, 加入 3μlEB, 冷却后倒胶。

50×TAE配方: 500ml H₂O中加 242g Tris., 溶解后加入 100 ml 0.5M EDTA (pH8.0), 和 57.1 ml冰醋酸定容至 1000 ml, 灭菌。

0.5M EDTA (pH8.0) 配方: 800ml H₂O中加 186.1g EDTANa₂·2H₂O, 用NaOH调pH至 8.0, 定容至 1000 ml。

2) 引物选取与合成:

我们采用 49 对SSR引物进行玉米材料的基因型鉴定, 这些SSR广泛分布于玉米的基因组中, 先前曾用于玉米遗传分析(表 2-2)。在这组SSR中, 其中大约 50%为三碱基重复, 37%为四碱基重复。重复种类和引物序列等相关信息可以从MaizeGDB (<http://www.maizegdb.org>) 上获得。

在引物合成时, 每对 SSR 引物中的正向引物按正常序列合成, 而在反向引物的 5' 端加上 M13 序列 (5' -CACGACGTTGTAAAACGA-3'), 另外再合成一条 5' 端带有荧光标记的 M13 引物序列。本研究中采用 FAM (蓝色) 和 TET (绿色) 两种荧光标记。

3) PCR 扩增

然后进行 PCR 扩增。PCR 产物在 ABI Prism 3700 DNA Sequencer 上进行电泳。用 GeneScan 软件查看 PCR 片段数据并在 EXCEL 表中手工记录, 然后基于 SSR 重复类型确定 SSR 等位位点片段大小, 将数据转换为特定的等位位点数据。

4) PCR 扩增采用以下体系:

5ng/μl 模板 DNA	5μl
10×PCR buffer (含Mg ²⁺)	1μl
10 mM dNTP	0.6μl
2U/μl Taq 酶	0.4μl
SSR 正向引物 (1μM)	1μl
SSR 反向引物 (0.5μM)	1μl
荧光标记 M13 引物 (1μM)	1μl
Total	10μl

所用试剂均购自北京天根生化公司。

5) PCR 反应程序如下:

Step1	95℃	1min
Step2	94℃	30s
Step3	52℃	1min
Step4	72℃	2min
Step7	34 times to 2	
Step8	72℃	8min

6) PCR 产物的纯化:

取 5μl 的 PCR 产物+2 倍体积的无水乙醇混匀, 3000 rpm 离心 30 min, 倒置 600 rpm 离心 1 min, 100 μl 70%乙醇水洗, 3000 rpm 离心 1 min, 倒置 600 rpm 离心 1 min, 加 50μl 的去离子水溶解, 于 4℃冰箱保存。

7) 荧光检测:

提取纯化的 PCR 产物 2 μ l 于 96 孔板上, 然后在每孔中分别加上 8.5 μ l 的甲酰胺 (含 1.7%的内标), 离心, 95 $^{\circ}$ C 变性 5min, 在 ABI3700 上进行自动荧光检测。另外, 进行 5% 变性聚丙烯酰胺凝胶电泳, 验证自动荧光检测的结果, 以确保数据的准确性。

8) 数据分析:

荧光检测结果用 GeneScan 软件分析。读取 SSR 扩增片段的长度, 手工录入到 EXCEL 表中。然后参考 SSR 重复序列的寡核苷酸数目, 将数据转换成对应的等位位点数据。

2.2.3 遗传多样性及群体结构分析

1) 遗传多样性统计

应用 PowerMarker 软件 (Liu et al. 2005) 进行一般统计量分析, 并获得等位位点数目、等位位点频率、基因型数目、实测核苷酸杂合度、基因多样性 (期望核苷酸杂合度) 以及多态性信息含量 (PIC) 等参数值。

其中, 基因多样性指群体内两个随机选择的等位位点之间存在差异的可能性, 第 l 个位点的基因多样性的无偏估计值为

$$\hat{D}_l = (1 - \sum_{u=1}^k \tilde{p}_{lu}^2) / (1 - \frac{1+f}{n})$$

其中, f 为近交系数。

核苷酸杂合度 (Heterozygosity) 是指群体内杂合个体所占的比例, 其计算公式为

$$\hat{H}_l = 1 - \sum_{u=1}^k \tilde{P}_{luu}$$

多态性信息含量 (PIC) 值是另外一个常用的多样性衡量参数 (Botstein et al, 1980)。其计算公式为

$$\widehat{PIC}_l = 1 - \sum_{u=1}^k \tilde{p}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^k 2\tilde{p}_{lu}^2 \tilde{p}_{lv}^2$$

2) 聚类分析

利用 PowerMarker 软件 (Liu et al. 2005) 构建聚类图。其步骤是: 首先通过“compute frequency”功能计算每个等位位点在相应自交系中的频率, 如果自交系中该位点是纯合的, 则频率为 1, 如果是杂合的, 则两个等位位点频率各为 0.5。然后基于所得频率通过 Shared-Allele 方法计算遗传距离, 然后用 Neighbor-Joining 算法建树。

3) 群体结构分析

为了估测玉米自然群体的遗传结构, 应用 STRUCTURE 软件 (Pritchard 2000a,b) 把自交系材料分到几个相应的 cluster 中, 并确定自交系的祖先组成。具体分析过程是, 首先需要假定存在一个群体数目 (K), 并认为位点都是独立的, 每个 cluster 内都遵循 Hardy-Weinberg 平衡。 K 值设定为 2~10, 然后将参数 iterations 设为 1000,000, burn-in period 设为 300,000。每个 K 值都独立运行三次。然后选择一个适当的 K 值确定为群体数目,

在特定 cluster 内祖先比例大于 75% 的自交系被指派到相应的 cluster 中。在任何一个 cluster 中的祖先比例都小于 75% 的品系被认为拥有混合的祖先血缘。因为有大量的自交系的系谱是未知的，所以我们没有把自交系的系谱信息加入到分析过程中。

表 2-2 用于基因型鉴定的 SSR 引物

Table 2-2 SSR Primers for Genotyping

Primer	Repeat Sequence	Bin	Primer	Repeat Sequence	Bin
Nc130	AGC	5.00	Phi121	CCG	8.04
Nc133	GTGTC	2.05	Phi123	AAAG	6.07
Phi008	GGC	7.02	Phi213984	ACC	4.01
Phi011	AGC	1.09	Phi233376	CCG	8.09
Phi015	AAAC	8.09	Phi299852	AGC	6.07
Phi032	AAAG	9.04	Phi308707	AGC	1.10
Phi046	ACGC	3.08	Phi328175	AGG	7.04
Phi050	AAGC	10.03	Phi374118	ACC	3.02
Phi059	ACC	10.02	Phi423796	AGATG	6.01
Phi062	ACG	10.04	Phi448880	AAG	9.06
Phi063	TATC	10.02	Phi453121	ACC	3.01
Phi064	ATCC	1.11	Phi96342	ATCC	10.02
Phi069	GAC	7.05	Umc1109	(ACG)4	4.10
Phi072	AAAC	4.01	Umc1122	(CGT)7	1.06
Phi073	AGC	3.05	Umc1136	(GCA)5	3.09
Phi076	AGCGGG	4.11	Umc1143	AAAAT	10.02
Phi078	AAAG	6.05	Umc1152	(ATAG)6	10.01
Phi079	AGATG	4.05	Umc1153	(TCA)4	5.09
Phi084	GAA	10.04	Umc1169	(TTA)4	1.04
Phi087	ACC	5.06	Umc1196	CACACG	10.07
Phi100175	AAGC	8.03	Umc1277	(AATA)5	9.08
Phi101049	AGAT	2.10	Umc1279	(CCT)6	9.00
Phi108411	AGCT	9.05	Umc1304	(TCGA)4	8.02
Phi112	AG	7.01	Umc1545	(AAGA)4	7.00
Phi114	GCCT	7.02			

4) 核心研究材料选择:

核心研究材料是从自然种质群体中选出的一组能够最大限度地代表该自然群体遗传多样性的种质。这组种质在应用关联分析方法发掘优异等位基因的工作中是理想的研究材料。为了选出

核心研究材料用作下一步研究的材料，我们在保留了一些农艺性状优良的骨干自交系的基础上，应用 PowerMarker 软件（Liu et al. 2005）中的 Line Selection 功能，基于 simulated annealing 算法选择了 94 个能捕获最大量 SSR 等位位点数目的自交系作为关联研究的实验材料。

2.3 结果与分析

2.3.1 遗传多样性分析

多样性统计指标，包括等位位点数目、等位位点频率、基因型数目、实测核苷酸杂合度、基因多样性（期望核苷酸杂合度）以及多态性信息含量（PIC）等指标值，参见表 2-3。

共检测到总计 262 个等位位点，平均每个 SSR 位点等位位点数为 5.3469。其中，23% 的等位位点，其基因频率小于 1%，其中更有 22 个等位位点只在一个自交系中检测到。

实测核苷酸杂合度、基因多样性以及多态性信息含量（PIC）是重要的多样性衡量指标。图 2-1 显示了本研究中几个指标的关系。大部分位点表现出了高 PIC 值，平均 0.5057。基因多样性表现出了和 PIC 类似的变化趋势，平均 0.5663。而对于实测核苷酸杂合度，虽然变化趋势与前两个指标也有一些类似，但其总体水平明显低于基因多样性，大部分位点的值在 0.20 以下，仅仅 nc130 和 phi063 两个位点杂合度较高，分别达到了 0.6585 和 0.4982。

表 2-3 SSR 的多样性统计

Table 2-3 Diversity statistics for maize microsatellites

标记 Marker	主要等位位点的 频率 Major Allele Frequency	基因型数目 No. of Genotype	等位位点 数目 No. of Allele	基因多样性 Gene Diversity	实测核苷酸杂合度 Observed Heterozygosity	多态性信息 含量 PIC
Nc130	0.4547	10	6	0.5931	0.6585	0.5081
Phi032	0.4983	8	5	0.5895	0.0972	0.5052
Nc133	0.6805	3	2	0.4348	0.0830	0.3403
Phi015	0.3902	17	7	0.6772	0.2474	0.6196
Phi96342	0.7517	9	5	0.4044	0.1285	0.3696
Phi308707	0.4983	11	5	0.6411	0.1840	0.5800
Phi213984	0.7578	6	4	0.3711	0.0662	0.3085
Phi108411	0.9167	5	3	0.1561	0.0764	0.1500
Phi121	0.8944	7	5	0.1942	0.0493	0.1847
Phi453121	0.4844	9	5	0.6354	0.1181	0.5675
Phi084	0.4790	9	4	0.5430	0.1189	0.4379
Phi100175	0.5727	10	6	0.5512	0.1702	0.4705

Phi374118	0.4791	12	6	0.6316	0.1533	0.5614
Phi079	0.8369	8	4	0.2897	0.0638	0.2757
Phi078	0.5327	3	2	0.4979	0.0764	0.3739
Phi059	0.5874	12	7	0.5255	0.1434	0.4359
Phi076	0.3798	9	4	0.6668	0.1916	0.5955
Phi233376	0.4379	14	6	0.7045	0.1773	0.6582
Phi423796	0.6941	6	3	0.4698	0.1049	0.4221
Phi087	0.4854	18	11	0.6558	0.1168	0.6007
Phi114	0.3579	13	6	0.7377	0.0807	0.6921
Phi101049	0.3316	30	13	0.7660	0.2465	0.7304
Phi062	0.6198	8	6	0.4803	0.1181	0.3762
Phi069	0.4773	7	4	0.6391	0.0804	0.5697
Phi063	0.3434	23	8	0.7586	0.4982	0.7229
Phi072	0.5712	17	8	0.6176	0.1875	0.5791
Phi064	0.3059	29	9	0.8167	0.1923	0.7942
Phi008	0.6855	5	4	0.4421	0.0618	0.3594
Phi011	0.4211	10	6	0.7204	0.0466	0.6781
Phi046	0.5052	6	3	0.5773	0.0903	0.4888
Phi050	0.6175	6	3	0.5388	0.0632	0.4757
Phi112	0.7465	12	8	0.4174	0.0524	0.3889
Phi073	0.4216	8	4	0.6393	0.0523	0.5690
Phi328175	0.3449	13	6	0.7246	0.1568	0.6764
Phi448880	0.7083	10	6	0.4682	0.0590	0.4376
Phi299852	0.2430	22	10	0.8226	0.1084	0.7985
Umc1143	0.6449	5	3	0.5046	0.0725	0.4374
Umc1136	0.3482	21	8	0.7682	0.1786	0.7345
Umc1152	0.4007	9	5	0.6523	0.0993	0.5812
Umc1196	0.5509	10	4	0.6215	0.0807	0.5730
Umc1109	0.8031	11	7	0.3343	0.0836	0.3075
Umc1304	0.5818	3	2	0.4866	0.0800	0.3682
Umc1279	0.7358	8	5	0.4139	0.0638	0.3653
Umc1169	0.3978	14	7	0.7104	0.1864	0.6598
Umc1545	0.5573	6	3	0.5914	0.0799	0.5257

Umc1277	0.5091	3	2	0.4998	0.0985	0.3749
Umc1153	0.4132	12	5	0.6973	0.1736	0.6443
Umc1122	0.5320	10	5	0.6171	0.1637	0.5566
Phi123	0.6519	3	2	0.4538	0.0954	0.3508
Mean	0.5437	10.6122	5.3469	0.5663	0.1342	0.5057

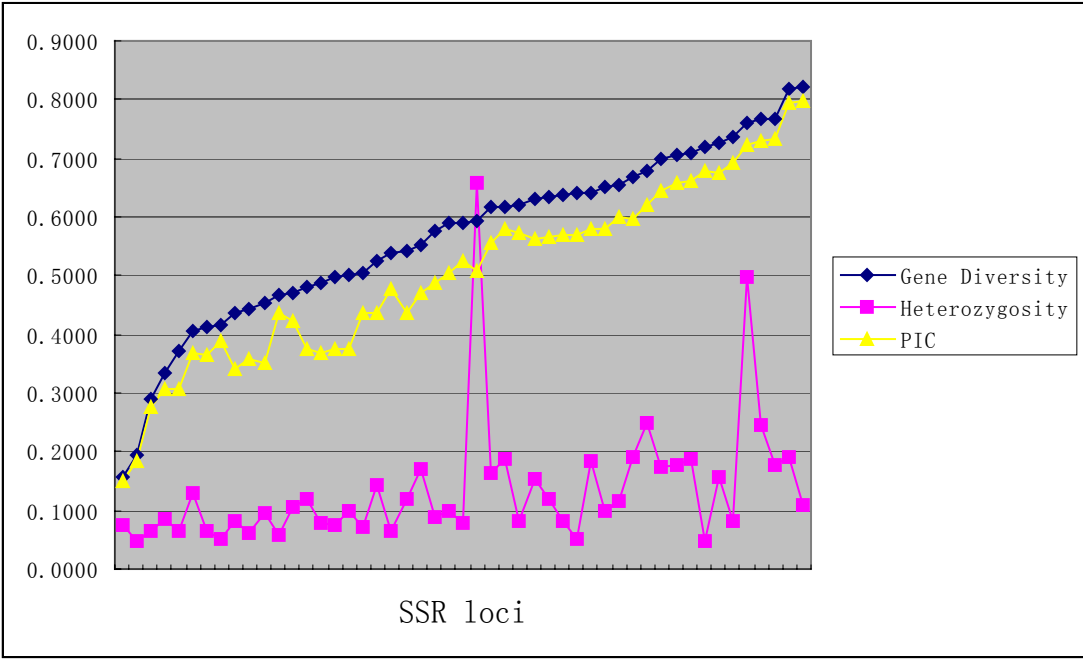


图 2-1 SSR 位点的核苷酸杂合度、基因多样性及 PIC 值

Fig 2-1 Heterozygosity, gene diversity (expected heterozygosity) and PIC for the SSR loci

2.3.2 288 份自交系的聚类分析

本研究通过 PowerMarker 软件构建的 N-J 树见图 2-2。所有 288 份自交系被聚到三个大的分枝下，类群 1、2、3 分别包含 41、171、76 个自交系。各类群又可以进一步根据遗传距离的不同分别被分成若干个分枝。

通过与自交系的系谱关系进行比较后发现，从总体上看，自交系的聚类结果呈现一个“大杂居，小聚居”的特点。即，在小的分枝中，亲缘关系最近的自交系都能聚到一起，其结果同系谱关系大体一致。然而，从大的类群看，三个大类群并没有与我国的几大杂种优势群很好的对应起来。

在第一个大的类群中，既包括被认为是 Reid 群的自交系（如 A632、辽 4366），又包括丹 340 这样公认的旅大红骨群自交系，其余大部分自交系或者系谱不清楚，或者选自地方品种、综合种。

第二个大的类群中由六个分枝组成，被认为属于 Reid 群的一些自交系，如辽 7794、掖 8112 等与一些系谱不详的品系及一些外引系位于其中一个分枝上，而另外一个分枝则包含了由 478、

488、兴 60、5003、VG85-5、DH65232 等另外一部分 Reid 系组成的小分枝以及大部分由 P 群自交系（遵 90110、齐 318、P138、沈 137、X178、L069、武 134 等）构成的另一个小分枝相邻。与这两个小分枝距离较近的还有自 330、Mo17 等骨干自交系，有趣的是，两个自交系并没有象被期望的那样，与其众多的衍生系聚在一起。与自 330 邻近的只有晋穗 54 是其衍生系，而显示与 Mo17 这个典型的 Lancaster 群自交系遗传距离最近的却是一个选自 78599 的自交系中自 05。在这个类群中另一个较大的分枝中，则包括了大部分属于塘四平头群的自交系：黄野四 3、塘四平头、四 444、444、897、S001、白 197、HR962、吉资 5A 等。这些黄早四衍生系自交系紧密地聚在一起，然而黄早四却没有包含在内，相邻的小分枝上则有旅 28、E28、宁 45、337 等旅大红骨系。

第三个大的类群由两个分枝组成。其中一个分枝以 Mo17 衍生系为主，如关 17-1、太 1/Lg、吉 996、吉 994、获唐黄 17、92 黄 40，此外还包括其它一些自交系。而另外一个分枝则显得混杂一些，既有黄早四与无唐 448、Yue89E4-2、Yue89E5-3 等塘四平头群自交系，也有糯性品系京糯 2 以及岩赖 2、岩赖 3 等高赖氨酸玉米。P39/su、Za5125/su、冀甜 15 等三个甜玉米自交系和白色爆裂也共同组成了其中一个小小的分枝，暗示着白色爆裂与甜玉米品系之间有极近的亲缘关系。

综合分析后可以发现，得出的分析结果尽管存在一定的规律性，但并不是象所期望的那样与系谱关系高度一致，其可能原因有以下几个方面：一是我们所用的 SSR 密度还不足，二是在基因型鉴定过程中可能存在的统计误差，还有一个原因就是，我国自交系的遗传基础和背景相当复杂，大量的系谱不详的自交系，与外引系、地方品种选系等交错在一起，使分析变得很困难。

2.3.3 288 份自交系的群体结构分析

为了估测我国自交系的历史来源和种质中的混合程度，我们采用 STRUCTURE 软件(Pritchard 2000a,b)进行了群体结构分析。经过 STRUCTURE 软件 30 次运行后 ($K=2\sim 10$ ，三次重复)，并结合种质的系谱知识，最终选定 K 值为 4。根据运行结果将所研究材料分成四个 cluster (表 2-1)，在特定 cluster 内遗传相似性比例大于 75% 的自交系被划分到相应的 cluster 中。

以在特定 cluster 内遗传相似性比例大于 75% 为标准，在 288 份自交系中，总计有 74 个自交系被划分到相应 cluster 中。结合已知系谱的自交系的系谱背景进行分析，发现划分结果与这些自交系的已知系谱大部分一致。四个 cluster 对应于四大种质来源。种质来源 I 拥有 21 个品系，大部分属于 Reid 种群 (表 2-4)；种质来源 II 拥有 11 个品系，其中 7 个来源于先锋杂交种 78599，在近年来的研究中这些品系被单独地归为 P 群 (表 2-5)；种质来源 III 拥有 22 个品系，其中 8 个已知属于塘四平头优势群 (表 2-6)，其它的品系系谱未知；种质来源 IV 拥有 20 个品系，其中仅前 10 个遗传相似性比例最高的品系中，就有 7 个是 Mo17 的后代，所以这个 Cluster 被看作是对应于 Lancaster 群 (表 2-7)。所有品系在这四大种质来源中的整体的遗传相似性比例见图 2-3，这暗示了在这组核心自交系中四大祖先的血缘所占的总体比例。

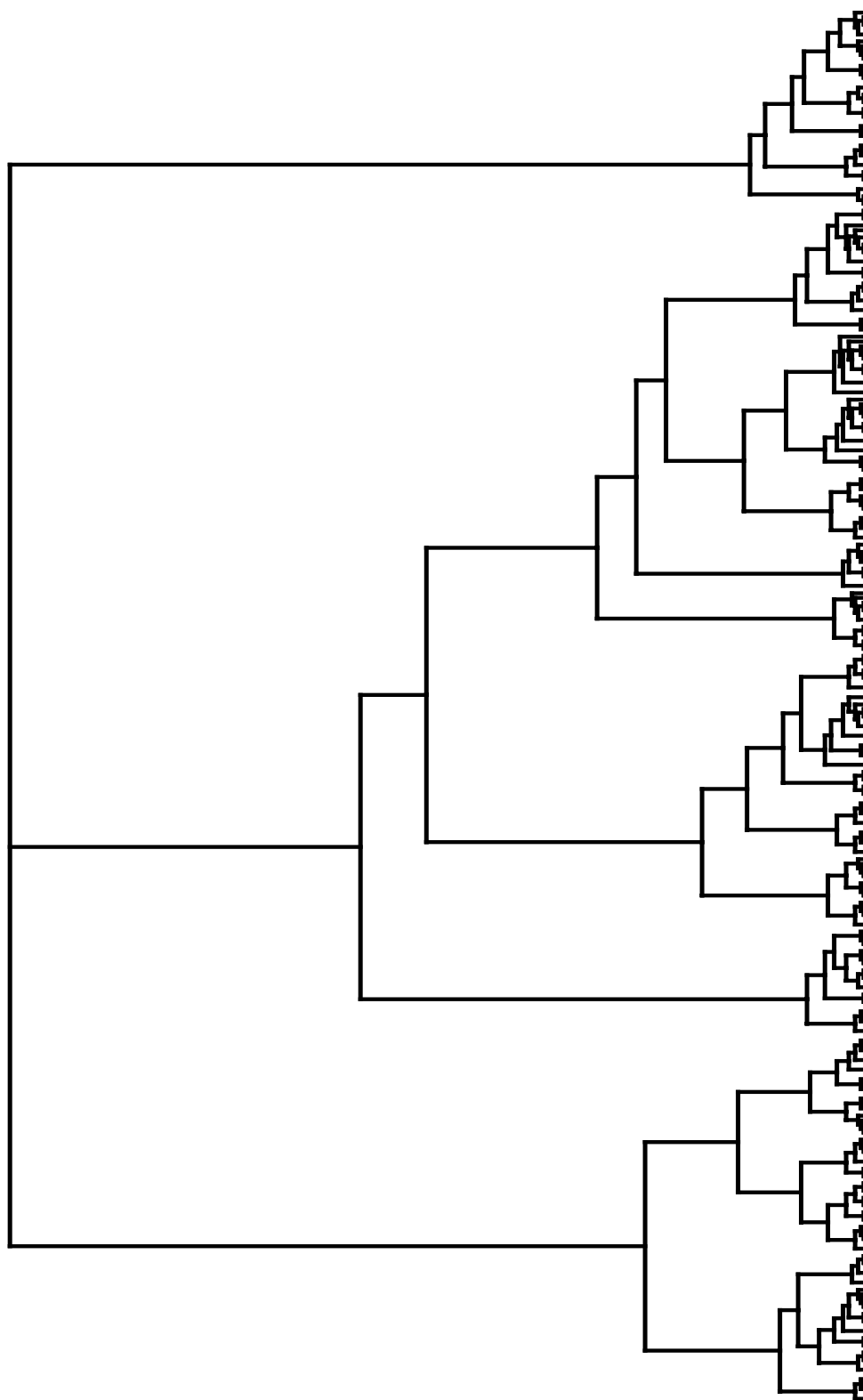


图 2-2 288 份自交系的聚类图

Fig 2-2 Cladogram tree of 288 inbred lines

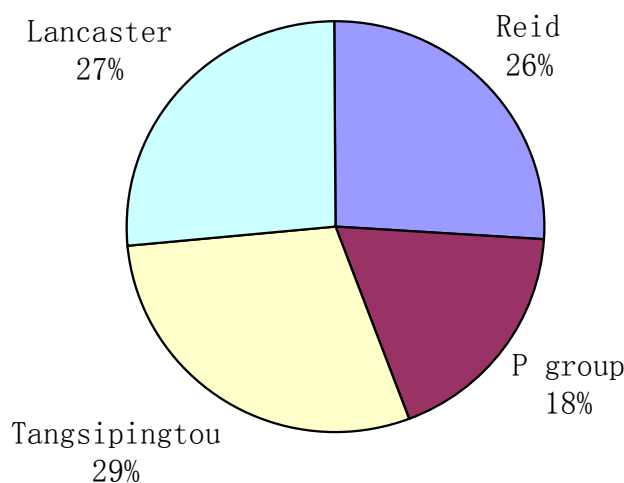


图 2-3 核心自交系中四个群所占的总体遗传相似性比例

Fig 2-3 Overall proportion of membership of the lines in each of the four groups

在我们的材料中,大量品系的系谱还不清楚,通过群体结构分析,可以推断出这些品系的可能遗传结构和祖先组成。在引自国外的品系中,A632 和 De811 被归入 Reid 群,这与二者的血缘关系是一致的。A632 和 De811 是美国两个骨干自交系,A632 属于 BSSS 血缘,De811 则比较复杂,不仅包括 BSSS 血缘,也包含来自 Lancaster 群的 C103 的血缘 (Gerdes et al,1993)。另外,BC4B (南斯拉夫) 也被归入了 Reid 群;此外,72-125 (意大利) 归入塘四平头群,F520 (法国) 归入 Lancaster 群。在来源于地方品种和综合品种的品系中,辽 3409 被归入 Reid 群,L005 归入 P 群,77、多 229、繁荣 2 和运 402 被归于塘四平头群,237、D729 和甸骨 11A 被归入 Lancaster 群。当然,被归入特定的杂种优势群并不意味着一定含有该群祖先的血缘,而只是表明了该自交系与相应种质来源间存在的高度遗传相似性。

而剩余的 214 个品系在任何一个群中的遗传相似性比例都小于 75%,因此被看作是混合品系。它们占到了 288 份品系的 75%。在各个群中遗传相似性比例介于 60%~75%之间的品系个数分别是:11 (Reid), 2 (P), 22 (塘四平头), 15 (Lancaster)。如果统计在各群中遗传相似性比例大于 60%的品系个数,则结果是 35 (Reid), 13 (P), 42 (塘四平头), 35 (Lancaster)。显然,在四个群中,P 群品系的个数是最少的,而塘四平头群则占的比例最大。这个结果同自交系在四个群中整体上所占的遗传相似性比例是一致的 (图 2-3)。其可能原因是,塘四平头群来源于国内的地方品种,大量其它的地方品种与该群都可能存在一定的遗传相似性。而其它三个都是由在不同时期从美国引入的种质繁衍而来,其中 P 群的引进时间要晚于 Reid 和 Lancaster 群,所以其所占比例也最小。

表 2-4 对应于 Reid 群的 Cluster 中自交系的遗传相似性比例

Table 2-4 Membership of the lines of the cluster corresponding to Reid group

自交系名称	遗传相似性比例 (Membership)				亲本来源
	Reid	P	塘四平头	Lancaster	
辽 7794	0.959	0.018	0.012	0.012	7922×8112/(32×5003)
488	0.958	0.013	0.014	0.015	8112×5003
8112	0.957	0.018	0.012	0.013	国外杂交种
32	0.955	0.02	0.009	0.016	3382
辽 3409	0.95	0.014	0.016	0.019	综合种
掖 8112	0.926	0.013	0.044	0.016	国外杂交种
掖 478	0.92	0.039	0.027	0.014	U8112×5003
BC4B	0.902	0.031	0.033	0.033	南斯拉夫
龙抗 15	0.9	0.061	0.019	0.02	RC103×长 3
A632	0.889	0.017	0.022	0.073	(Mt42×B14) ×B14(3)
7922	0.859	0.024	0.014	0.104	3382
478	0.857	0.112	0.014	0.018	U8112×5003
辽 5110	0.854	0.017	0.094	0.035	7922×5003
5003	0.818	0.024	0.132	0.026	3147
兴 60	0.817	0.11	0.015	0.057	478×J852
De811	0.808	0.01	0.045	0.137	[B68 × [B73Ht × (C103 × Mp3204 双交种)Sel.]]
辽 4366	0.787	0.132	0.045	0.036	8112×3352
遗 15	0.782	0.108	0.067	0.044	不详
辽 4476	0.779	0.047	0.079	0.095	7922×3358
赤黄 29	0.762	0.055	0.093	0.091	旅 9×Ci7
VG85-5	0.76	0.096	0.056	0.088	5003×关 17

除了这些品系外,其余自交系大约占 57%,它们被认为是高度混合的。其中一部分自交系的遗传相似性比例主要在两个群中,表明这些自交系与相应的这两个群的亲缘关系要较为亲密;还有一部分自交系,其遗传相似性构成主要来自三个 Cluster;最典型的,比如在 87-20、55113-3-3-5、84-126-15-1、84112、Yue267-1-1 品系中,在四个群中的遗传相似性比例是相当的,都在 20%~30%之间,这可能暗示着其选育、繁衍历史的复杂性。由于传统的杂种优势理论认为,两个亲本自交系相互间遗传距离越远,组配的杂交种可能拥有的杂种优势就越大,因而在选用这些自交系进行杂交种亲本选配时更应慎重一些。

另外需要指出的是,一些自交系如 CML58、CML67、CML125 等,都含有一定的热带血缘,但却没有被划分到任一群中,而是表现为混合来源,这些自交系还需要进一步研究。

表 2-5 对应于 P 群的 Cluster 中自交系的遗传相似性比例

Table 2-5 Membership of the lines of the cluster corresponding to P group

自交系名称	遗传相似性比例 (Membership)				亲本来源
	Reid	P	塘四平头	Lancaster	
X178	0.014	0.961	0.009	0.016	北美单交种 P
L069	0.011	0.96	0.016	0.014	78599
齐 318	0.02	0.954	0.01	0.015	78599
沈 135	0.019	0.953	0.014	0.015	78599
P138	0.017	0.95	0.011	0.021	78599
L005	0.017	0.948	0.014	0.021	陕综 5 号
沈 137	0.02	0.945	0.014	0.02	78599
遵 90110	0.037	0.928	0.012	0.023	78599
Sh21	0.047	0.883	0.048	0.022	不详
武 134	0.059	0.871	0.04	0.03	78599
壳大 1	0.082	0.761	0.084	0.072	壳 1-1×大秋 36

表 2-6 对应于塘四平头群的 Cluster 中自交系的遗传相似性比例

Table 2-6 Membership of the lines of the cluster corresponding to Tangsipingtou group

自交系名称	遗传相似性比例 (Membership)				亲本来源
	Reid	P	塘四平头	Lancaster	
HR962	0.011	0.012	0.965	0.012	黄早四
72-125	0.017	0.01	0.963	0.011	意大利
444	0.032	0.012	0.942	0.014	A619×黄早四
黄野四 3	0.018	0.013	0.924	0.046	(黄早四×野鸡红)×黄早四
四 444	0.024	0.024	0.891	0.062	A619×黄早四
抚 85	0.059	0.018	0.884	0.038	335
77	0.026	0.026	0.884	0.064	双跃 147
多 229	0.018	0.076	0.883	0.024	濮阳石灰簪×小黄玉米
早 49	0.057	0.026	0.872	0.045	TW×peno×早 22
897	0.098	0.033	0.843	0.026	112×黄早四
繁荣 2	0.035	0.028	0.826	0.111	八道渠
塘四平头	0.104	0.027	0.825	0.045	塘四平头
岩 38	0.02	0.029	0.824	0.128	鲁单 38
运 402	0.037	0.116	0.823	0.024	不详
85 白 64	0.098	0.041	0.811	0.05	85 黄 11×武 403
枣 123	0.071	0.048	0.808	0.072	农家种
白 197	0.036	0.033	0.799	0.131	海东 013、黄早四
原武 02	0.041	0.135	0.797	0.028	武 105×多 229

75-14 高	0.141	0.037	0.769	0.052	75-14 杂株
707	0.049	0.15	0.762	0.038	77 异交株
黄早四	0.013	0.181	0.758	0.047	塘四平头
凤可 1	0.039	0.095	0.752	0.113	克利 67×Oh43

表 2-7 对应于 Lancaster 群的 Cluster 中自交系的遗传相似性比例

Table 2-7 Membership of the lines of the cluster corresponding to Reid group

自交系名称	遗传相似性比例 (Membership)				亲本来源
	Reid	P	塘四平头	Lancaster	
关 17-1	0.017	0.023	0.014	0.946	关 73×Mo17
获唐黄 17	0.016	0.018	0.024	0.942	获唐白×Mo17Ht1
吉 994	0.017	0.041	0.011	0.931	吉 63×Mo17
大 MO	0.012	0.034	0.041	0.913	不详
吉 992	0.043	0.024	0.026	0.907	吉 63×Mo17
太 1/Lg	0.06	0.02	0.027	0.893	Mo17×太 1/lg
选 6	0.032	0.033	0.045	0.89	H95×Mo17
吉 996	0.091	0.017	0.02	0.872	吉 63×Mo17
二南 24	0.09	0.028	0.018	0.864	南 107×南 55
D729	0.077	0.014	0.047	0.862	D 团综合种
F520	0.048	0.022	0.07	0.86	法国
237	0.06	0.046	0.038	0.856	获白×秋 23
Y673	0.092	0.045	0.036	0.828	英 64×吉 63
107	0.127	0.029	0.025	0.82	XL80
Q114-1	0.022	0.105	0.056	0.817	不详
甸骨 11A	0.02	0.037	0.138	0.805	桦甸红骨子
郑白 11	0.079	0.023	0.099	0.799	墨白
C103R	0.073	0.045	0.109	0.773	美国
64C103	0.125	0.056	0.054	0.765	Bup64×C103
3025	0.076	0.128	0.045	0.751	美国

2.3.4 Core Sets 构建

本研究在保留了一部分农艺性状优良的骨干自交系的基础上,应用 PowerMarker 软件 (Liu et al. 2005) 中的 Core Set 功能,以等位点数目为选择标准,选择了 94 个自交系作为微核心研究材料。这组自交系及其系谱关系见表 2-8。初步分析表明,这组自交系共拥有 228 个等位点,占 288 份自交系等位点总数 (262) 的 87%。另外,这组自交系的基因多样性是 0.5591,略低于 288 份实验材料的基因多样性 (0.5663) (表 2-9)。因而,这组自交系最大限度地代表了原 288 份自交系中尽可能多的多样性,可以作为核心研究材料应用于随后的关联分析之中。

表 2-8 核心研究材料的系谱来源

Table 2-8 Pedigree sources for mini core sets

ID	名称	亲本来源	ID	名称	亲本来源
1	吉 63	(127-32×铁 84)×(威 24×威 20)	48	A632	(Mt42×B14)×B14(3)
2	矮金 525	武陟矮×金皇后	49	Lo1125	Pioneer ADA (PR3374)
3	C103	Noah Hershey	50	K36	不详
4	自 330	Oh43×可利 67	51	遵 90110	78599
5	塘四平头	塘四平头	52	P138	78599
6	获白	获嘉白马牙	53	H21	黄早四×H84
7	Mo17	C103×187-2	54	488	8112×5003
8	旅 28	旅大红骨	55	X178	北美单交种 P
9	威风 322	W59E×风可	56	昌 7-2	潍 59×黄早四
10	华 160	华东二号	57	DH65232	DH6327×沈 5003
11	黄早四	塘四平头	58	沈 137	78599
12	E28	(旅 9 宽×A619Ht) BC	59	H2	美国杂交种
13	掖 8112	国外杂交种	60	晋穗 54	自 330 变异
14	5003	3147	61	GB	天津白玉米
15	掖 478	U8112×5003	62	X.L9010-3/O ₂	5105×opaque2
16	丹 340	白轴旅 9×有稃玉米	63	龙抗 15	RC103×长 3
17	黄野四 3	(黄早四×野鸡红)×黄早四	64	皖系 23	Va35×B73
18	获唐黄 17	获 唐 白 42 × 海 1917 × Mo17Ht1	65	关 17-1	关 73×Mo17
19	综 3	综合种	66	粤 267-1-1	5003×抗旱大粒黄
20	郑 58	掖 478	67	粤 89E4-2	武 102×黄早四
21	鲁原 92	原齐 123×1137	68	粤 20-3	掖单 13 变异
22	齐 319	78599	69	武 202	武 403×Bup29
23	K12	黄早四×淮春	70	CML67	墨西哥
24	资玉 3	鲁资玉 3 号	71	辽 2204	美国杂交种
25	京糯 2	不详	72	黄 C	(黄小 162×自 330/O ₂) × Tuxpeno
26	冀甜 15	国内甜杂交种	73	CN165	不详
27	D729	D 团综合种	74	沈 135	78599
28	太 184	混选 1 号	75	原武 02	武 105×多 229
29	XZ19	金 0-14×忻 335	76	444	A619×黄早四
30	13A/O ₂	不详	77	早 49	TW×peno×早 22
31	抚 96	南充 5×矮 13-31	78	朝鲜白	朝阳白玉米

32	辽 7794	7922×8112/(32×5003)	79	H205	国外杂交种
33	吉 846	Mo17×吉 63	80	87-20	不详
34	吉 880	哲 B77A	81	岩 172	浦城矮玉米
35	九 03	不详	82	中二/O ₂	中单 2 号
36	辐 842	甸 11×意 210/辐 746	83	峰 273	春杂 1 号
37	鹿 65	北金 14×330	84	P39/su	美国
38	岩 103	原武 02	85	92 黄 7	W153×C103
39	齐 318	78599	86	785	330×黄早四
40	48-2	综合种	87	92 黄 40	853×Mo17
41	FR218	美国	88	赤黄 32	旅 9×Ci7
42	大 MO	不详	89	兴 83	144×147
43	遗 67	不详	90	龙抗 1	辽 1311(甸 11×大风 22)
44	91 黄 5	甸 11A/旅 9	91	Lo1067	Pioneer 3780×Lo87602
45	辽 5110	7922×5003	92	HR962	黄早四
46	邯 102	河南多穗	93	IAB078	意大利
47	De811	[B68 × [B73Ht × (C103 × Mp3204 双交种)Se1.]]	94	承 18	顶上玉米× (公 70×60 —22)

表 2-9 选出的核心研究材料与 288 份自交系间的遗传多样性比较

Table 2-9 Comparison of diversity between the mini core sets and 288 lines

	等位位点数目	基因多样性	实测核苷酸杂合度	PIC
288 份自交系	262	0.5663	0.1342	0.5057
94 份核心研究材料	228	0.5591	0.1129	0.4992

2.4 讨论

2.4.1 遗传多样性的评价

在遗传多样性分析中，等位位点数目和基因多样性是两个重要的参数，它们与 SSR 重复的类型及样本规模密切相关。本研究中，采用主要来自我国玉米核心种质中的 288 份自交系作为研究群体，检测到的每个位点平均的等位位点数目是 5.3469，基因多样性是 0.5663。这两个指标值都与用其它地区的玉米材料得到的相应参数值接近：6.6 和 0.76 (Taramino *et al.* 1996), 5.21 和 0.59 (Senior *et al.* 1998), 6.9 和 0.62 (Matsuoka *et al.* 2002)。

然而，在 Liu 等的观测结果中，等位位点数目和基因多样性分别为 21.7 和 0.82 (Liu *et al.* 2003)，与之相比，本研究中的这两个参数相对要低一些。这可能暗示着，还有一些多样性更高的玉米种质有待于被搜集和收入到国家种质库中。当然，从另一方面考虑，这种多样性差异也可能是由于

研究中所使用的 SSR 类型的差异造成的。Vigouroux (2002) 和 Liu (2003) 等都发现双核苷酸重复的 SSR 比三核苷酸重复及更多核苷酸重复的 SSR 多态性更高。本研究中所采用的这组 SSR 中, 大约 50% 是三核苷酸重复, 37% 是四核苷酸重复, 而在 Liu 等的研究中所用的 SSR 组中, 大部分是双核苷酸重复类型, 这也可能是两研究中检测到的多样性不同的原因。

2.4.2 遗传结构和玉米自交系的系谱来源

我国玉米自交系的种质基础十分复杂, 20 世纪 80 年代以来, 通过分析系谱来源、地理分布等手段, 育种家们进行了初步的类群划分(吴景锋, 1983; 曾三省, 1990)。王懿波等(1998) 在对我国广泛应用的自交系的系谱进行系统分析以后, 结合配合力以及杂种优势等因素, 认为我国的自交系主要有四大类群: 来自美国 Lancaster 和 Reid 群, 以及国内起源的塘四平头与旅大红骨群, 其中的 Lancaster 群包含 Mo17、自 330 两个亚群, 此外还存在选自国外杂交种、综合品种、Suwan 和低纬度地区种质组成的一些亚群。而袁力行等则认为 Reid 群包含 BSSS 和 PA 两个群(袁力行等, 2001)。后来, 选自美国先锋杂交种 P78599 的 P138、X178、齐 319、沈 137 等自交系在我国玉米生产中扮演了越来越重要的角色, 因而, 在利用分子标记进行杂种优势群划分时加入了上述自交系, 并被划分到一个独立的杂种优势群, 称为 P 群(赵久然等, 1999; 王元东等, 2004)、PB 群(李新海等, 2003) 或者温热 I 群(滕文涛等, 2004)。

虽然经过了大量研究, 但仍有许多玉米自交系的系谱是未知的。我国自交系的复杂系谱背景使得对这些自交系进行遗传结构分析变得相当复杂。应用 STRUCTURE 软件(Pritchard *et al.* 2000a) 进行群体结构分析, 能够有助于阐明自交系的历史来源, 并且度量这些种质的混合程度。

本研究中, 通过设定 $K=2\sim10$ 进行运算, 结合已知的我国自交系系谱信息, 最终发现, 当 $K=4$ 时, 大部分自交系的分析结果同系谱信息是一致的, 于是选择了 $K=4$, 并通过运行程序, 得到四个群(Cluster) 及所有 288 份自交系在四个 Cluster 中各自的遗传相似性比例。在任一群中遗传相似性比例 ≥ 0.75 的自交系被归入相应群中。根据这些自交系的系谱及先前获得的其它信息, 我们发现这四个群分别对应于我国的四个杂种优势群(塘四平头、Reid、P 群以及 Lancaster 群(表 2-4-2-7))。然而, 应该指出的是, Liu 等在进行相关研究时将在任一群中遗传相似性比例均小于 80% 的自交系视为混合来源(Liu *et al.* 2003), 而本研究将在任一群中遗传相似性比例均小于 75% 的自交系视为混合来源, 这是由于用于分析的种质的遗传复杂程度不同, 因而采用了不同的选择标准。

另外, 正如前面提到的, 自交系在四个种质群中的遗传相似性比例并不一定意味着该自交系的系谱来自该杂种优势群, 而只是暗示着遗传背景的相似性。例如, C103, 一个来自美国的自交系, 在塘四平头群中拥有一定的遗传相似性比例(38.2%), 这意味着 C103 的遗传背景同塘四平头种质群有 38.2% 的遗传相似性, 并不是表示 C103 的一部分血缘来自塘四平头优势群。类似的, 国内骨干自交系黄早四在 P 群中拥有 18.2% 的遗传相似性比例, 这也不意味着黄早四的系谱与 P 群有关系。

(1) 塘四平头群

塘四平头是一个来自河北唐山的硬粒地方品种, 适应性广。1974 年, 育种者从中选育出了著

名的骨干自交系黄早四。自此以后的几十年间，该自交系又先后衍生出了 100 多个自交系，共同组成了一个杂种优势群，塘四平头群由此形成（高翔等，2005）。黄早四具有较高的一般配合力，适应性广，生长势强。由于其特殊的紧凑叶型，导致了我国出现了大量的拥有紧凑型或者半紧凑型的杂交种。因此，黄早四被广大育种者视作一个划时代的自交系。本研究的结果显示，该自交系在塘四平头群中拥有 76% 的遗传相似性比例，而另有 18% 在 P 群中。有趣的是，由黄早四衍生而来的自交系在塘四平头群中都有更高的遗传相似性（444，94%；黄野四 3，92%；四 444，89% 等）。与之相对应的是，在前面构建的聚类树中，黄早四与其衍生系也没有出现在同一分枝上，而衍生系彼此之间却能很好的聚在一起。另外，另一个典型自交系塘四平头，是直接从塘四平头地方种中选育而来，在本群中也仅有 82% 的遗传相似性，这两个自交系与衍生系之间遗传相似性产生差异的原因需要进行深入探讨。

表 2-10 与塘四平头群具有较高遗传相似性的地方品种选系

Table 2-10 Lines developed from landraces with high membership in the Tangsipingtou group

自交系	地方品种起源	在塘四平头群中的遗传相似性比例
获白	获嘉白马牙	65.4%
多 229	濮阳石灰篓×小黄玉米	88.3%
繁荣 2	八道渠	82.6%
邯 102	河南多穗	41.6%
枣 123	地方品种，名字不详	80.8%
朝鲜白	朝阳白玉米	74.1%
岩 172	浦城矮玉米	58.2%
秋 23	南充秋子	55.2%
武 302	野鸡红	61.8%
武系 205	辽东白	60.9%

还应该注意到，从本研究获得的数据中可以发现，大量从地方品种中选育来的自交系都在塘四平头群中拥有较高的遗传相似性比例（表 2-10）。这可能是由于我国地方品种的遗传多样性本来就不高的缘故。因而有必要对我国的地方品种资源进行类似的群体结构分析，以求对这些地方品种中的遗传多样性结构有更全面的了解。

(2) 旅大红骨群

旅大红骨是一个由辽宁旅大地区的农家种与引入的大红骨在混种条件下天然杂交，然后又经不断选择而来的农家品种（鞠方成等，2003）。许多自交系是其中选育而来，如著名自交系旅 28、丹 340 等。因而，这些从旅大红骨衍生而来的自交系长期以来一直被育种者认为是一个独立的杂种优势群。然而，根据本研究中获得的结果，这些自交系却表现为混合的种质来源（表 2-11）。在近几年的一些聚类分析研究中，也曾把这些自交系划分到不同的群中（黎裕等，2002；滕文涛等，2004）。在滕文涛的研究中，用了 111 个 SSR 标记对 84 份自交系进行了聚类分析，发现丹 340 被划分到了自 330 群中，旅 28 被划分到了塘四平头群中。本研究中的群体结构分析结果显示，

这些自交系在塘四平头群中的遗传相似性比例要低于在其它三个群中的比例，这暗示着，旅大红骨的祖先之一大红骨与 Reid、P 群、Lancaster 群的亲缘关系可能更近一些。

表 2-11 几个旅大红骨自交系在各群中的遗传相似性比例

Table 2-11 Memberships proportion of several presumed Ludahonggu lines

自交系	遗传相似性			
	Reid	P	塘四平头	Lancaster
旅 28	0.048	0.388	0.326	0.239
丹 340	0.592	0.138	0.096	0.174
337	0.236	0.367	0.164	0.233
赤黄 29	0.762	0.055	0.093	0.091
赤黄 32	0.034	0.041	0.334	0.592
91 黄 4	0.496	0.309	0.029	0.166
E28	0.287	0.195	0.086	0.432

(3) Lancaster 群

Lancaster 群在以前的研究中被分成两个亚群：自 330 亚群和 Mo17 亚群（王懿波等，1997，1998）。然而，另一研究中，认为自 330 亚群是一个独立的杂种优势群（滕文涛等，2004）。根据群体结构分析结果，本研究中的 Cluster 4 中包含了 20 个自交系，分析其系谱来源后发现，其中拥有本群遗传相似性比例最高的前 8 个自交系中，有 7 个是从 Mo17 衍生而来（表 2-7）。因而这组自交系对应于 Lancaster 杂种优势群。然而，Mo17 所拥有的本群遗传相似性比例却只有 56%。这与聚类树中 Mo17 与其衍生系相距较远的结果是一致的，这可能是由于，经过多年选育，与原来的 Lancaster 群种质相比，这些属于 Lancaster 群的自交系的遗传基础已经有了很大的变化（郑大浩等，2002）。

然而，另一个自交系自 330 却分别在 Reid 和塘四平头群中拥有 44% 和 34% 的遗传相似性比例。自 330 的两个亲本是 Oh43 和 Keli67。Oh43 选自(W8 × Oh40B)的后代，因而与原来的 Lancaster 种质已经就有了一些遗传变异；而 Keli67 选自可利（Creole），该种质与 Lancaster 种质并没有直接的关系。因而，自 330 不应被看作 Lancaster 群中的一员（李艳天等，2002）。在其它研究中，曾将自 330 划入旅大红骨群（袁力行等，2000、2001）。另外，那些衍生系在任一群中都没有大于 75% 的遗传相似性。因此，基于本研究结果和系谱分析，自 330 和其衍生系似乎不应被划归到 Lancaster 杂种优势群。

(4) P 群

自从 20 世纪 70 年代开始，育种家们从美国引入了先锋杂交种 78599，并进而从中选育出了一批优良自交系如齐 319、P138、X178、沈 137 等。育种家们以这些自交系为亲本，先后育成了农大 108、鲁单 50 等一大批表现优异的杂交种。迄今为止，应用这些自交系已育成 100 多个杂交种。鉴于这些自交系在育种生产中的显著表现，王元东等（2004）把这些自交系单独列为一个杂种优势群并称之为 P 群。滕文涛则推测这些自交系很可能包含一些热带种质，因而称作温热群。

本研究支持了这一观点,即这些 P 群自交系属于一个单独的种质来源,很有可能是一个独立的杂种优势群。

2.4.3 群体结构在评估自交系和杂交育种亲本选配中的作用

通过利用分子生物学手段对自交系的遗传结构和祖先来源进行的分析,我们对遗传组成和杂种优势群有了新的理解。在我国,由于各种原因,许多玉米自交系的系谱并不清楚,有的甚至是错的。而且,随着玉米杂交种生产的商品化,许多育种者不愿意将所用亲本自交系的真实系谱公之于众,甚至自交系名称都不是真实的。本研究中所用的这种遗传结构分析方法可以用来推断这些自交系的祖先来源。

例如,吉 63 是一个在我国东北地区广泛应用的重要自交系。王懿波等(1998)认为它来自一个从综合种选出的亚群。该自交系是从双交种“吉双一号”选育而来,该双交种的母本杂交种的两个亲本 127-32 和铁 84 均来自地方品种,而父本杂交种的亲本威 24 和威 20 则是 20 世纪 50 年代从美国引入的自交系(吴景锋,1983),然而,其具体的遗传背景还不清楚。研究表明,吉 63 有更多的遗传相似性比例来自 Reid 与 Lancaster 优势群。另一个例子是原武 02,王懿波等(1998)将其划分到了 Reid 群,但在本研究中却有很高比例的遗传相似性比例来自塘四平头群。

祖先来源分析在杂交育种的亲本选配过程中所扮演的角色可以用下面的例子来说明。第一个例子是玉米杂交种“中单 2 号”。该杂交种是“Mo17 \times 自 330”的杂交 F_1 代。这个杂交种 20 世纪 70 年代后被广泛推广,由于其具有的广泛适应性,甚至被推广到了广大的边远地区。本研究中,Mo17 在 Lancaster 群中拥有 55.5%的遗传相似性比例,另有 34.1%来自 P 群。而自 330 则有 44.4%的遗传相似性比例来自 Reid 群,另有 33.5%来自塘四平头群。第二个例子是玉米杂交种“郑单 958”。它是郑 58 与昌 7-2 的杂交 F_1 代。郑单 958 在近年来被大面积推广,2005 年年播种面积已超过 5 千万亩。郑 58 中 63.2%的遗传相似性比例来自 Reid 群,20.5%来自 P 群,而昌 7-2 却有 45.5%的遗传相似性比例来自塘四平头群,另有 46.0%来自 Lancaster 群。这不仅证实了已被普遍接受的杂种优势理论,即两个相互间遗传距离较远的自交系组配的单交种拥有较高的杂种优势,而且也为在杂交育种工作中的亲本选配提供了一个新的思路。

2.4.4 群体结构、聚类结果与系谱之间的比较

在本研究中,采用 SSR 分子数据构建了一个聚类树,对比聚类结果与群体结构分析结果,并与系谱结合起来分析,发现三者之间既呈现很大程度上的一致性,又存在许多差异。对于一些自交系组,比如,近年来在育种中发挥了重要作用的几个选自 P78559 的自交系(齐 318、P138、沈 137、X178、L069、武 134),在聚类图中紧密地聚在一起,在群体结构中也自成一个独立的 Cluster。类似地,Mo17 衍生系、黄早四衍生系、Reid 系等也都表现出系谱分析、聚类分析与群体结构分析结果的高度一致。然而,还有相当一部分自交系的分析结果存在一些差异,比如 Mo17、黄早四这两个骨干自交系,它们在聚类时没有与其衍生系聚在一起,在群体结构中则表现为在本群中的遗传相似性略有降低,其原因还需要进一步研究。

由于这些基于分子数据得到的分群结果和先前基于杂种优势、配合力以及系谱关系分析划分

的杂种优势群之间存在着较大的差异，因此，本研究中的结果暗示着，如果用结构分析结果来将自交系指定到特定的杂种优势群也并不是十分合适的。这种不足可能与我国自交系的遗传基础的复杂程度有关，也可能是基因型数据统计过程中的误差造成的，然而，更可能是 SSR 标记密度的限制。因为所用自交系群体规模较大，如果加大标记密度，也许遗传结构分析结果会更理想。

与聚类分析相比，群体结构分析的优势在于可以暗示出自交系的内在遗传结构。由于多种原因，我国自交系中存在大量系谱不详的自交系，以及近几年来引入的一些外引系，系谱不清使得育种者在应用这些自交系进行亲本选配时存在一定的盲目性。进行遗传结构分析，有助于阐明其遗传组成。而且，如前所述，该结果也能为育种者在未来的杂交育种工作中的亲本选配提供一些有益的思路。

第三章 *rab17* 基因核苷酸多态性及连锁不平衡结构分析

3.1 前言

在玉米基因组中, 普遍存在高水平的核苷酸多样性 (Tenaillon et al, 2001)。近几年来, 玉米基因组中的序列多样性分析已经成为了玉米基因组研究中的一个热点。

玉米基因组包括大约 59000 个基因, 占基因组的 7.5% (Messing et al, 2004)。其中, 2~4% 的基因 (大约 1200 个) 是在玉米驯化期间或者随后的种质改良过程中人工选择的靶标 (Wright et al, 2005)。通过检测基因的序列多样性, 度量其连锁不平衡结构, 就可以分析选择在该基因进化中产生的影响。

另外, 由于玉米是异交作物, 重组率高 (Fu, 2002), 这些导致了自交系中的 LD 的快速衰退 (Remington et al, 2001; Tenaillon et al, 2001), 从而能够提高关联分析时的分辨率。因而, 玉米是一个实施关联分析, 鉴定控制农艺性状相关基因的功能等位变异的模式作物 (Thornsberry et al, 2001; Rafalski and Morgante, 2004)。

rab17 (responsive to ABA) 是玉米体内一个受 ABA 和干旱胁迫诱导表达的基因, 在 1990 年由 Vilardell 等首先分离得到 (Vilardell et al, 1990)。该基因是涉及植物对 ABA 应答的一组基因的成员之一 (Gomez et al, 1988; Mundy et al, 1988), 其 mRNA 不仅在 ABA 诱导下在胚胎发育期间累积, 干旱胁迫条件下也在叶片中累积 (Pla et al, 1989)。该基因由于在序列结构和功能表达方面都类似于水稻 *rab21* 基因 (Mundy et al, 1988), 其表达蛋白的分子量大小为 17kD, 因此而得名。

对 *rab17* 的表达产物的定位和磷酸化研究后发现, *rab17* 编码一个甘氨酸富集蛋白, 包含有一个 8 个丝氨酸残基组成的丝氨酸簇, 其中 7 个是连续的。该蛋白分布在细胞核和胞质中, 通过与核定位信号肽结合在核蛋白转运中起作用。它的表达活性可以通过丝氨酸残基上的磷酸化作用进行调节 (Goday et al, 1994)。

研究表明, *rab17* 是一个与耐旱性相关的基因 (Kizis et al, 2002)。位于染色体 6 长臂上的 6.05 区段, 在先前报道的耐旱相关 QTL 研究中, 在该位点附近相继鉴定出一些控制干旱胁迫下的重要生理性状的 QTLs, 这些性状包括开花—吐丝间隔期 (ASI) 和产量性状等 (Ribaut et al, 1996, 1997)。那么这是否暗示着该基因内部存在与这些表型变异相关联的功能等位变异呢? 为了鉴定出是否存在这种功能变异, 有必要对 *rab17* 的序列多样性进行分析。

因此, 本研究以前一部分工作中选出的一组微核心自交系为材料 (第二章), 对 *rab17* 基因的序列多样性进行了分析, 对连锁不平衡程度进行了检测, 以期对该基因中的序列多样性分布有一个全面了解, 并为随后的关联分析提供分子基础。

3.2 材料与方法

3.2.1 实验材料

采用第一部分工作中（参见第二章）从 288 份核心自交系中选出的 94 份核心研究材料作为这部分研究的实验材料。

94 份自交系的 DNA 提取、纯化等见第二章。

3.2.3 *rab17* 基因的扩增与测序

1) 引物设计:

以 NCBI 中的 *rab17* 基因序列 (X15994) 为参照, 设计引物。最终选用的引物上游序列在转录起始位点上游-346 处, 下游序列在+1322 处。

Forward Primer: 5'-TGT TAT GCA GTT CGC TCT GGA C-3'

Reverse Primer: 5'-GCC AAT TGC ATG CGT CTA ATC-3'

2) PCR 扩增:

为了最大限度的减少人为误差, 避免由于 PCR 导致的序列错误, 本部分实验采用具有高保真特性的 Pfu DNA 聚合酶进行 PCR, 并且每个样品扩增三个重复。主要试剂来自天根生化公司。

PCR 体系如下:

DNA 模板 (5ng/μl)	5μl
Forward Primer (10μM)	1μl
Reverse Primer (10μM)	1μl
2×Pfu PCR Master Mix	12.5μl
ddH ₂ O	5.5μl
Total	25μl

PCR 扩增程序为:

Step1	94℃	4min
Step2	94℃	1min
Step4	57℃	1min
Step6	72℃	2min
Step7	34 times to 2	
Step8	72℃	10min

3) PCR 产物的回收、纯化:

对 PCR 产物, 先取 2μl 进行 1% agarose gel 电泳检测, 如果存在亮度较高的目标带, 将其余部分 PCR 产物进行 1% agarose gel 电泳, 电泳完毕后进行 PCR 片段的挖胶回收和纯化。回收纯化试剂盒选用 QIAGEN 公司的 QIAquick Gel Extraction Kit, 具体操作步骤如下:

- 用干净的刀片将单一的 DNA 片段从琼脂糖凝胶上切下, 尽量切除多余凝胶, 并称取其重量。
- 将称重后的胶块放入一个无色的 1.5ml 离心管, 加入 3 倍体积的 Buffer QG (100ml~100μl)。
- 50℃水浴 10min, 直到胶块完全溶解。水浴过程中每隔 2~3min 不断旋转混匀离心管。
- 待胶块完全溶解后, 检查混合液颜色是否为黄色 (与 Buffer QG 类似), 如果为桔黄色或

- 者紫罗兰色，加 10 μ l 3M 醋酸钠 (pH5.0)，混匀，使颜色转为黄色。(因为黄色时 pH 小于等于 7.5，此时 DNA 吸收效果最好，颜色深表明 pH 偏高)。
- e) 把一个 QIAquick spin column 放在提供的 2ml 收集管中。
 - f) 把混合液转入 column 中，12000rpm 离心 1 min。
 - g) 弃掉废液，把 column 放回收集管，12000rpm 离心 1 min。
 - h) 加 0.75ml Buffer PE 于 QIAquick spin column，12000rpm 离心 1 min，以漂洗 DNA。
 - i) 弃掉废液，13000rpm 离心 1min。
 - j) 把 QIAquick spin column 放在干净的 1.5ml 离心管中。
 - k) 为了洗提DNA，加入 30 μ l 65℃预热的Buffer EB或者ddH₂O到QIAquick膜的中心，静置 1min，12000rpm离心 1min。
 - l) 取 2 μ l 回收产物进行 1% agarose gel 电泳检测，如果成功回收，将其余部分置于-20℃备用。
- 4) 本实验所测序列中的一部分是采用直接测序方法得到，即将经过回收纯化的 PCR 片段直接送公司进行测序；另一部分连接到质粒载体上，再通过培养，选择阳性克隆进行测序，简要步骤如下：
- a) 连接：连接选用 TaKaRa 公司的 pMD18-T 载体试剂盒。按载体与插入片段摩尔数比 1 : 3 的比例配制总量 10 μ l 的连接液，16℃反应 2h；
 - b) 全量 (10 μ l) 加入至 100 μ l Top10 感受态细胞中，冰浴 30min；
 - c) 42℃水浴 45s；
 - d) 快速将离心管转移至冰浴中，冷却细胞 2-3min；
 - e) 每管加 890 μ l 未加 Amp 的 LB 培养基，37℃振荡培养 60min 使细菌复苏；
 - f) 取 200~300 μ l，均匀涂布在含有 X-gal、IPTG、Amp 的 LB 固体培养基上，37℃倒置培养 14~16h；
 - g) 挑取白色克隆于 1.5ml 离心管中，加入 1ml 含 Amp 的 LB 培养基，37℃振荡培养 12-16h；
 - h) 通过菌液 PCR 检测的方法，每个样品选择出三个阳性克隆，送测序公司测序。
- 5) 测序：
- 对 PCR 片段或者阳性克隆进行双向测序，其中对仍未测通的序列，以得到的序列做模板，设计并合成中间引物（设计引物时注意与已测序列的重叠与衔接），继续进行测序 PCR，直至完成整条序列的测定。
- 6) 序列的比对与拼接：
- 将同一材料测出的不同区域的 Contig，用 DNASTAR 软件包内的 SeqMan 程序进行序列拼接，拼接过程中查看测序图，以手工编辑辅助，保证序列的准确可靠。对测序质量不好的序列，重新进行测序，最终完成所有品系的序列拼接工作。然后进行所有品系的序列比对，对于只在一个自交系中检测到的多态性，进行重复扩增和测序，确保序列的准确无误。

3.2.4 序列多态性统计分析

(一) 序列比对

所有自交系的序列比对工作采用 ClustalX 软件进行 (Thompson et al,1997)。用 Do Complete Alignment 功能对序列进行比对, 对认为需要局部比对的部分, 通过 Realign Selected Sequences 或者 Realign Selected Residue Range 进行局部调整; 将完成比对的序列保存为 PHYLIP 格式文件。

(二) 系统发育树的构建

系统发育树的构建用 PHYLIP 软件包 (Felsenstein, 1993) 完成。replicate 数目设成 1000, 评估方法采用 Bootstrap, 算法采用最大简约性法, 所得到的 TRE 文件用 TREEVIEW 打开, 即为所得进化树。

(三) 序列多态性分析及 LD 分析

用 Arlequin 软件 (Excoffier et al, 2005) 进行 Tajima'D 测验 (1989)。该值基于一个不考虑重组的 infinite-site 模型, 通过比较突变参数 θ 的两个估测值而得来 ($\theta = 2M\mu$, 在二倍体群体中 $M=2N$, N 为有效的群体尺寸, μ 为突变率)。D 测验值的计算公式如下:

$$D = \frac{\hat{\theta}_{\pi} - \hat{\theta}_S}{\sqrt{\text{Var}(\hat{\theta}_{\pi} - \hat{\theta}_S)}}$$

其中 S 为分离位点的数目。

单体型分析也用 Arlequin 软件进行, 并构建 Minimum Spanning Tree (MST) (Kruskal,1956; Prim, 1957)。

用 TASSEL 软件进行序列多样性分析及 LD 分析。

多样性分析采用 Sliding Window 分析方法, Window 设为 300bp, Step 设为 50bp;

LD分析时计算相应测量值 D' 和 r^2 , 选择Rapid Permutations, 将Permutations值设为 1000。

3.3 结果与分析

3.3.1 *rab17* 基因的序列多态性

在扩增过程中, 94 份自交系中有六个自交系的扩增片段略大于目标片段, 另有四个自交系的扩增片段小于目标片段, 这十个自交系的 PCR 产物经测序发现, 与 *rab17* 基因无同源性。可能是这些自交系在引物结合区域发生了较大的序列变异, 如大的插入/缺失 (Indel), 使得引物无法正常结合而产生了非特异性扩增。另外, 还有 14 个自交系因扩增较弱, 难以回收得到高质量的 PCR 产物。因而本研究最终得到了 70 个自交系的 *rab17* 基因序列, 与参照序列 X15994 一起, 共计 71 个序列用来进行序列多态性统计及分析。

rab17 基因的转录区域包括两个外显子, 之间被一个 137bp 的内含子分隔, 另外在两端分别有 5'-UTR 和 3'-UTR。另外在上游的非转录区 (5'-NTR) 有一个高 GC 含量的启动子, 再往上游是一段富含 AT 的序列, 下游还有一段 400 多 bp 的非转录区 (3'-NTR)。在序列比对、拼接后, 去掉两端测序质量不好的序列, 本研究最终得到的序列长度 1654bp 左右, 约为参照序列 (X15994) 长度的 81%, 包含了 *rab17* 基因完整的转录区、调控序列区域以及 5' 上游和 3' 下游的大部分序列 (图 3-1)。



图 3-1 rab17 基因的结构

Figure 3-1The structure of rab17 gene

（注：本研究中所得到的 1654bp 的序列对应于参照序列 X15994 的 274~1898bp，因 Indel 的存在，对应长度略有不同。此文中除特别注明外，所涉及到的关于 rab17 基因序列中的位置的描述均为本研究中 1654bp 片段中的相对位置。）

表 3-1 rab17 基因序列多态性分布

Table 3-1 Polymorphisms of rab17 sequence in the lines

	序列长度（bp） No.of bp	变异位点数目 No.of polymorphic sites	备注 Comments
非编码区 Noncoding region	1144		
SNP		13	1 个 SNP / 88bp
Indel		17	1 个 Indel / 67bp
Total		30	1 个 polymorphic site / 38bp
编码区 Coding region	510		
SNP		6	1 个 SNP / 85bp
Indel		1	1 个 Indel / 510bp
Total		7	1 个 polymorphic site / 73bp
整体序列区域 All sequenced region	1654		
SNP		19	1 个 SNP / 87bp
Transitions/ Transversions ratio			0.9
Indel		18	1 个 Indel / 92bp
Total		37	1 个 polymorphic site / 45bp

经过比对和统计，在总计 71 个序列中，共检测到了 37 个多态性位点。平均每 45 个 bp 检测到一个多态性位点。其中，SNP 19 个，平均每 87bp 检测到一个。另外还检测到 18 个 Indel，平均每 92bp 检测到一个（表 3-1）。在 37 个变异位点中，在非编码区中的有 30 个，占 81%，平均每 38bp 就检测到一个，而编码区只检测到 7 个，平均每 73bp 才检测到一个。其中，SNP 在非编码区与编码区的发生频率大体相当，而 Indel 在非编码区的发生频率则是编码区的将近八倍。

如果单独统计 5'-UTR、3'-UTR 及 Intron 中的多态性，会发现与整体上的结果相比，这三个区域有着更高的多态性发生频率。在 5'-UTR 区域，仅仅不到 100bp 的序列长度内，就检测到了 3 个 Indel 和 3 个 SNP，平均每 16bp 就有一个多态性变异；在长 262bp 的 3'-UTR 区域，并没有检测到 SNP，但检测到的 Indel 多达 5 个，平均每 52bp 就有一个 Indel 产生。类似的情况也出现在 Intron 中，仅仅 141bp 的序列中就有 3 个 SNP 和 3 个 Indel 产生。这三个区域总计 500bp，却有 11 个 Indel 和 6 个 SNP，平均达到了 29bp 一个多态性变异。

另外，对检测到的 19 个 Indel 的长度分布进行统计后发现，各种长度类型的 Indel 发生频率不同。单核苷酸 Indel 发生频率最高，为 5 个，2 核苷酸和 6 核苷酸 Indel 分别检测到两个，其它长度的各检测到一个，其中，大于 10bp 的 Indel 检测到 3 个，长度分别为 16、19、23bp（图 3-2）。

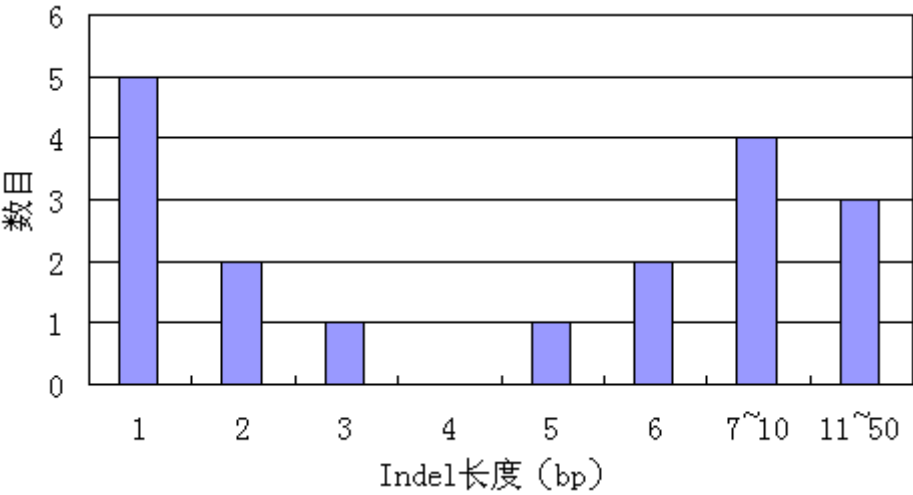


图 3-2 各种长度类型的 Indel 分布

Figure 3-2 Distribution of Indels with different length

需要特别指出的是，在总计 37 个核苷酸变异中，有两个 SNP 和 3 个 Indel 仅仅在参照序列中存在（表 3-2）。其中，755bp 处 X15994 的 16 个碱基片段的插入与相邻序列相同，实质上是多了一个拷贝。

表 3-2 核心自交系与 X15994 之间的序列多态性

	424	425	755	1074	1258
X15994	C	G	AATCCGTGGGTTTCGT	—	—
70 accessions	G	C	—	G	T

3.3.2 *rab17* 基因的序列多样性分析

研究以核苷酸多样性 π 值和核苷酸多态性 θ 值为指标，进行了多样性统计， θ 值是一个衡量群体突变率的参数，与核苷酸变异占序列位点数的比率有关，而 π 值则是衡量同一位点不同序列两两之间的差异的一个参数。分析后发现，*rab17* 基因的序列多样性在基因区域中呈现不均匀分布。

本检测到的 SNP 多样性分布见图 3-3 和图 3-4。由图中可以看出， π 和 θ 值的变化趋势较为一致。在前 800bp 内，存在着较高的 SNP 多样性。这种多样性在 800bp 后迅速下降，直至 0.0005 以下。

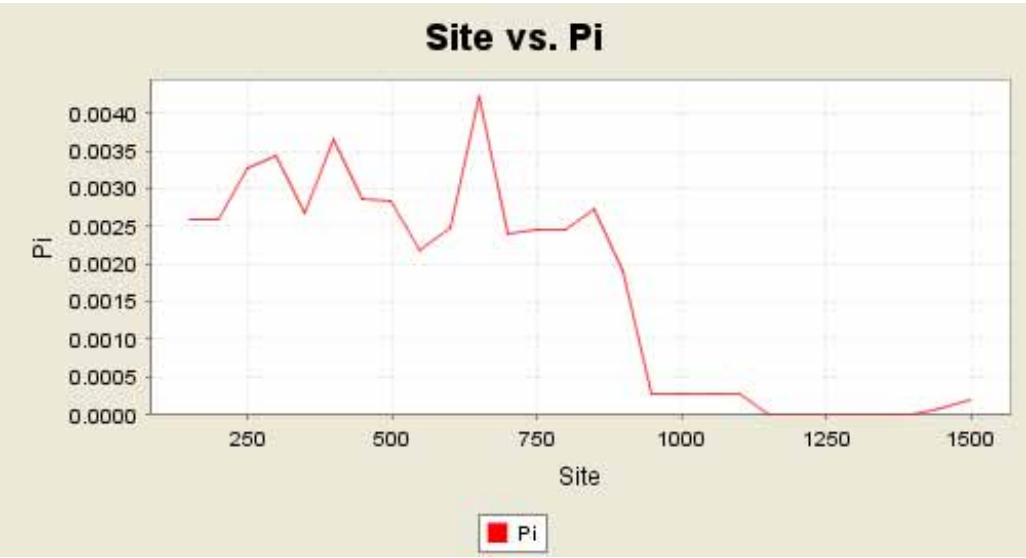


图 3-3 *rab17* 基因区域内的序列多样性 (SNP) (π)

Figure 3-3 The sequence diversity of *rab17* (SNP) (π)



图 3-4 *rab17* 基因区域内的多样性变化 (SNP) (θ)

Figure 3-4 The sequence diversity of *rab17* (SNP) (θ)

Indel 出示的变化（图 3-5、图 3-6）不同于 SNP。 π 值也象 SNP 一样，在 800bp 区域内一直处于 0.0015 以上的高水平，在 800bp 以后开始下降，但其下降幅度要小于 SNP 的相应变化幅度，一直在 0.001~0.0015 之间波动，至 1200bp 以后才又开始持续下降。 θ 值的变化则明显与 SNP 变化规律不同，而是一直在 0.003 上下波动，并在 1200bp 左右达最大值。

另外，本研究也对选择是否对 *rab17* 基因产生了影响进行了分析。通过对 *rab17* 序列进行 Tajima'D 测验，得到的 Tajima'D 值为 -1.16547， $P(D_{\text{simul}} < D_{\text{obs}})$ 为 0.10700，不显著。暗示着 *rab17* 基因在玉米驯化和改良进程中可能没有承受选择带来的压力。但这个结果只是根据这组自交系内的序列多样性分析得到的，还需要在以后的工作中通过与玉米地方品种、大刍草等材料的序列进行比较分析来验证。

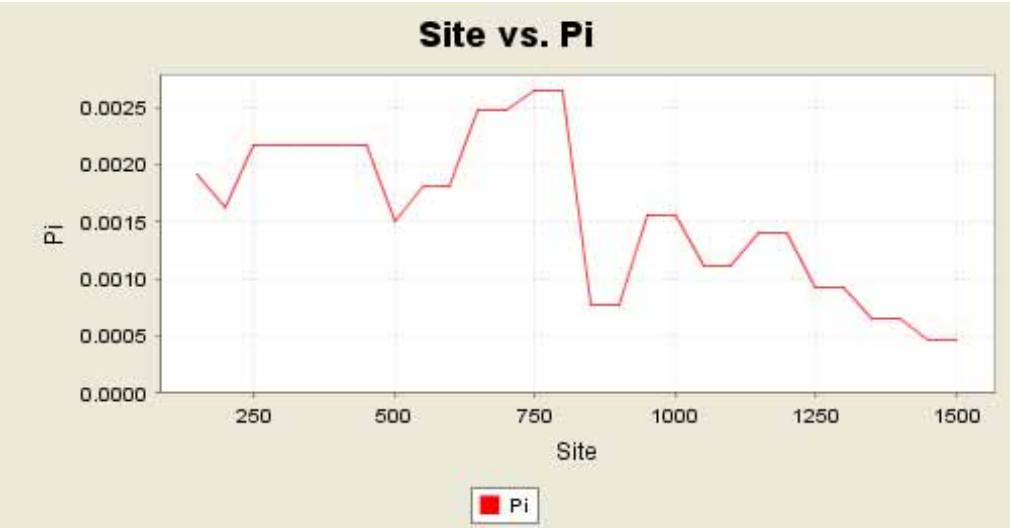


图 3-5 *rab17* 基因中的多样性分布 (Indel) (π)

Figure 3-5 The sequence diversity of *rab17* (Indel) (π)



图 3-6 *rab17* 基因中的多样性分布 (Indel) (θ)

Figure 3-6 The sequence diversity of *rab17* (Indel) (θ)

3.3.3 *rab17* 基因的连锁不平衡结构分析

经过对*rab17* 基因的序列多态性进行连锁不平衡结构分析,我们发现了在基因区域的前 800bp 内存在较强的LD结构, 包括 5'-NTR、5'-UTR、Exon1 及Intron, 而在Exon2 及 3'下游区域, r^2 值迅速降至 0.1 以下 (图 3-7)。

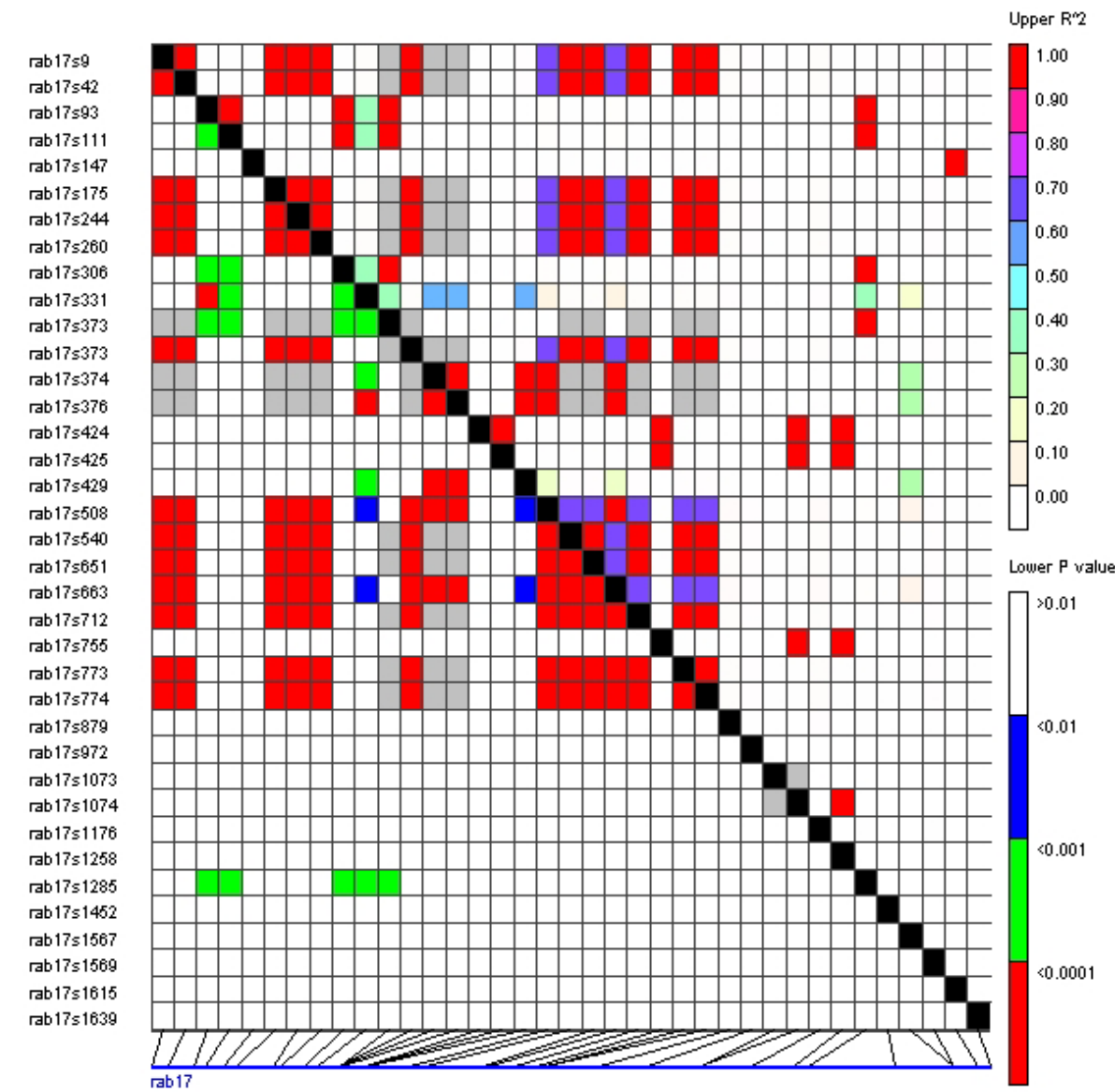


图 3-7 *rab17* 基因多态性位点间的连锁不平衡

Figure 3-7 Linkage disequilibrium between pairs of *rab17* sequence polymorphic sites

另外，本研究还对*rab17* 基因区域LD的衰退程度进行了检测，用 r^2 为衡量指标，在该基因区域检测到了明显的LD的衰退，由图 3-8 可以看出，在*rab17* 基因区域内，LD已衰退至 800bp左右。

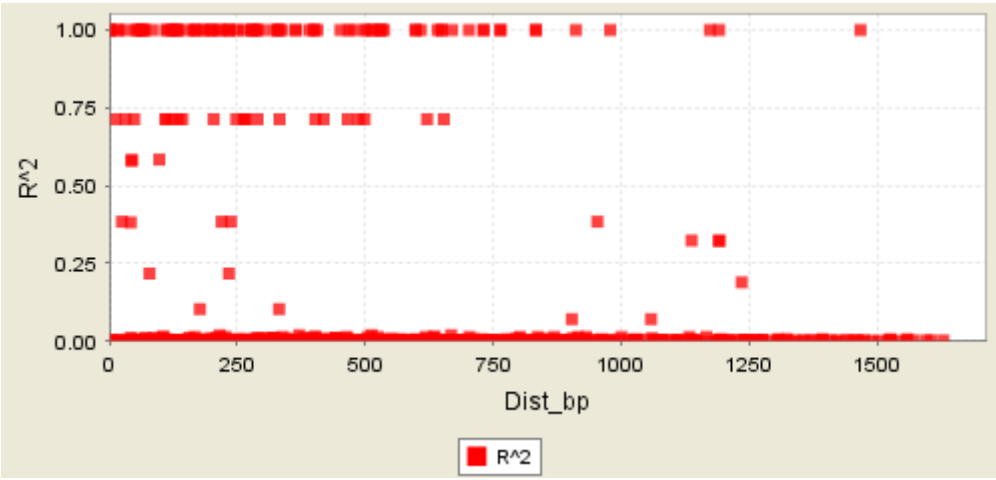


图 3-8 *rab17*基因区域内的LD衰退 (r^2)

Figure 3-8 Decay of linkage disequilibrium in *rab17* sequences

根据各核苷酸变异位点间的连锁不平衡结构，所有序列可以分为 14 个单体型 (Haplotypes)。各单体型频率不同 (表 3-3)。从单体型包含的自交系数来看，6 个单体型为只包括一个自交系的稀有单体型。有两个单体型各包含两个自交系，包含 3 个自交系和 6 个自交系的单体型各为两个，还有一个单体型包含了 9 个自交系，余下的自交系则同属于单体型 5 (表 3-3)。对这些单体型通过 Minimum Spanning Tree (MST) 方法 (Kruskal, 1956; Prim, 1957) 构建的最小生成树见图 3-9。

表 3-3 *rab17* 基因序列中的单体型
Table 3-3 Haplotypes in the *rab17* sequences of the accessions

	自交系	Haplotype	序列多态性																		
	数目	频率	9	42	93	111	147	175	244	260	306	331	373	373	374	376	424	425	429	508	540
Haplotype 1	1	1.41%	0	0	6	C	C	C	T	C	0	0	A	0	T	G	G	C	T	G	C
Haplotype 2	2	2.82%	0	0	6	C	C	C	T	C	0	0	A	0	T	G	G	C	T	G	C
Haplotype 3	1	1.41%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	C	G	G	A	C
Haplotype 4	1	1.41%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 5	34	47.89%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 6	3	4.23%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 7	6	8.45%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 8	1	1.41%	0	0	6	C	T	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 9	1	1.41%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 10	2	2.82%	0	0	0	G	C	C	T	C	10	4	G	0	G	C	G	C	G	A	C
Haplotype 11	6	8.45%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 12	1	1.41%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 13	3	4.23%	0	0	6	C	C	C	T	C	0	2	A	0	G	C	G	C	G	A	C
Haplotype 14	9	12.68%	1	1	6	C	C	T	C	G	0	2	-	9	-	-	G	C	G	G	A

表 3-3 *rab17* 基因序列中的单体型 (续)

Table 3-3 Haplotypes in the *rab17* sequences of the accessions (continued)

	自交系 数目	Haplotype 频率	序列多态性																	
			651	663	712	755	773	774	879	972	1073	1074	1176	1258	1285	1452	1567	1569	1615	1639
Haplotype 1	1	1.41%	1	0	T	16	T	T	7	G	0	0	6	0	2	0	A	0	C	0
Haplotype 2	2	2.82%	1	0	T	16	T	T	7	G	0	0	6	0	2	0	C	0	C	0
Haplotype 3	1	1.41%	1	3	T	0	T	T	7	G	0	1	6	1	2	0	C	0	C	0
Haplotype 4	1	1.41%	1	3	T	16	T	T	7	G	0	0	6	0	2	23	C	0	C	0
Haplotype 5	34	47.89%	1	3	T	16	T	T	7	G	0	0	6	0	2	0	C	0	C	0
Haplotype 6	3	4.23%	1	3	T	16	T	T	7	G	0	0	6	0	2	0	C	19	C	0
Haplotype 7	6	8.45%	1	3	T	16	T	T	7	G	0	0	0	0	2	0	C	0	C	0
Haplotype 8	1	1.41%	1	3	T	16	T	T	7	G	0	0	6	0	2	0	C	0	T	0
Haplotype 9	1	1.41%	1	3	T	16	T	T	7	G	5	?	6	0	2	0	C	0	C	0
Haplotype 10	2	2.82%	1	3	T	16	T	T	7	G	0	0	6	0	0	0	C	0	C	0
Haplotype 11	6	8.45%	1	3	T	16	T	T	0	G	0	0	6	0	2	0	C	0	C	0
Haplotype 12	1	1.41%	1	3	T	16	T	T	7	G	0	0	6	0	2	0	C	0	C	9
Haplotype 13	3	4.23%	1	3	T	16	T	T	7	A	0	0	6	0	2	0	C	0	C	0
Haplotype 14	9	12.68%	0	0	C	16	C	A	7	G	0	0	6	0	2	0	C	0	C	0

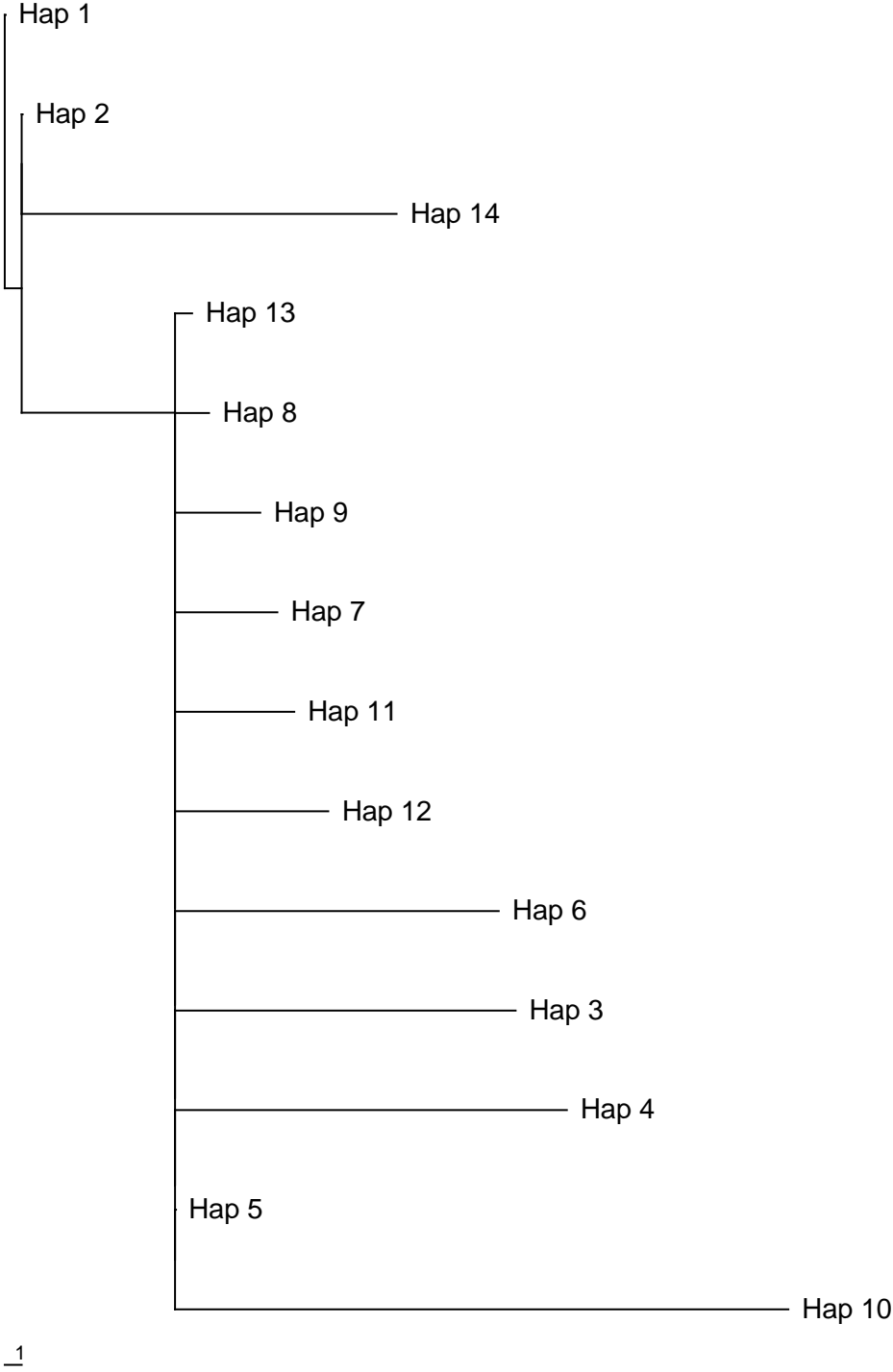


图 3-9 单体型的最小生成树

Figure 3-9 The Minimum Spanning Tree (MST) for Haplotypes

3.3.4 *rab17* 基因系统发育树构建

利用 PHYLIP 软件包中的 SEQBOOT、DNAPARS、CONSENSE 等程序,对序列比对结果进行系统发育树构建,最终得到的进化树见图 3-10。在图中,所有 71 个序列(含 70 个自交系与参照序列 X15994)被分成 14 个分枝,对应于所确定的 14 个单体型,各分枝包含的自交系数目各不相同。分枝间的距离也有很大差异,表明了各组自交系间的遗传距离的远近。从图中可以看出,齐 319、齐 318、遵 90110、P138、X178、沈 137、H2、沈 135 和 87-20 等 9 个自交系与其它材料间的遗传距离最远。其次是辽 2204、Lo1125、威风 322 三个自交系。

3.4 讨论

3.4.1 *rab17* 基因的序列多态性频率

本研究在 *rab17* 基因区域的多态性检测中,每 87 bp 检测到一个 SNP,与 *Adh1* 位点周围区域的 SNP 发生频率相当(每 83~104 bp 一个, Jung et al, 2004),但要低于 Ching 在 18 个位点的研究中检测到的(每 61 bp 一个, Ching et al, 2002),更是远远低于染色体 1 上的 21 个位点的序列多样性(每 28 bp 一个, Tenaillon et al, 2001)。

在 Indel 的检测中,每 92 bp 检测到了一个 Indel,则要高于 *Adh1* (每 186 bp 一个, Jung et al, 2004) 和 Ching 检测的位点(每 126 bp 一个, Ching et al, 2002)。

Indel 作为一种序列多态性,在其它作物中产生的频率较低,因此在一些作物的序列多样性分析中只统计 SNP,而忽略掉了 Indel。然而,与其它作物不同,玉米中存在高频率的 Indel (Bhattaramakki et al, 2002)。在本实验中所检测到的核苷酸变异位点中,SNP 和 Indel 比例相当,各占约 50%。然而,两者在基因区域的分布规律不尽相同,SNP 在编码区与非编码区的发生频率大体相当,但 Indel 在非编码区的发生频率则要远远高于编码区,18 个 Indel 中,只有一个发生在编码区,发生频率是非编码区的将近八分之一。类似的结果在 Ching 的研究中也曾被检测到,该研究中,在编码区序列中 2.35 kb 内仅仅有一个 3 bp 的 Indel,而非编码区平均每 85 bp 就有 Indel 产生(Ching et al, 2002)。这很可能与编码区内的 Indel 会强烈影响表达蛋白的功能有关。

3.4.2 功能序列中的多态性

(一) 编码区中的多态性与相应的编码蛋白结构变化

在检测到的 *rab17* 基因区域 37 个多态性变异位点中,有 7 个位于 *rab17* 基因的编码区 (Vilardell et al, 1990),包括 6 个 SNP 和 1 个 Indel (表 3-4)。除 429 bp 处和 540 bp 处的 SNP 属于同义突变外,其它几个变异位点均导致了编码产物的变化。

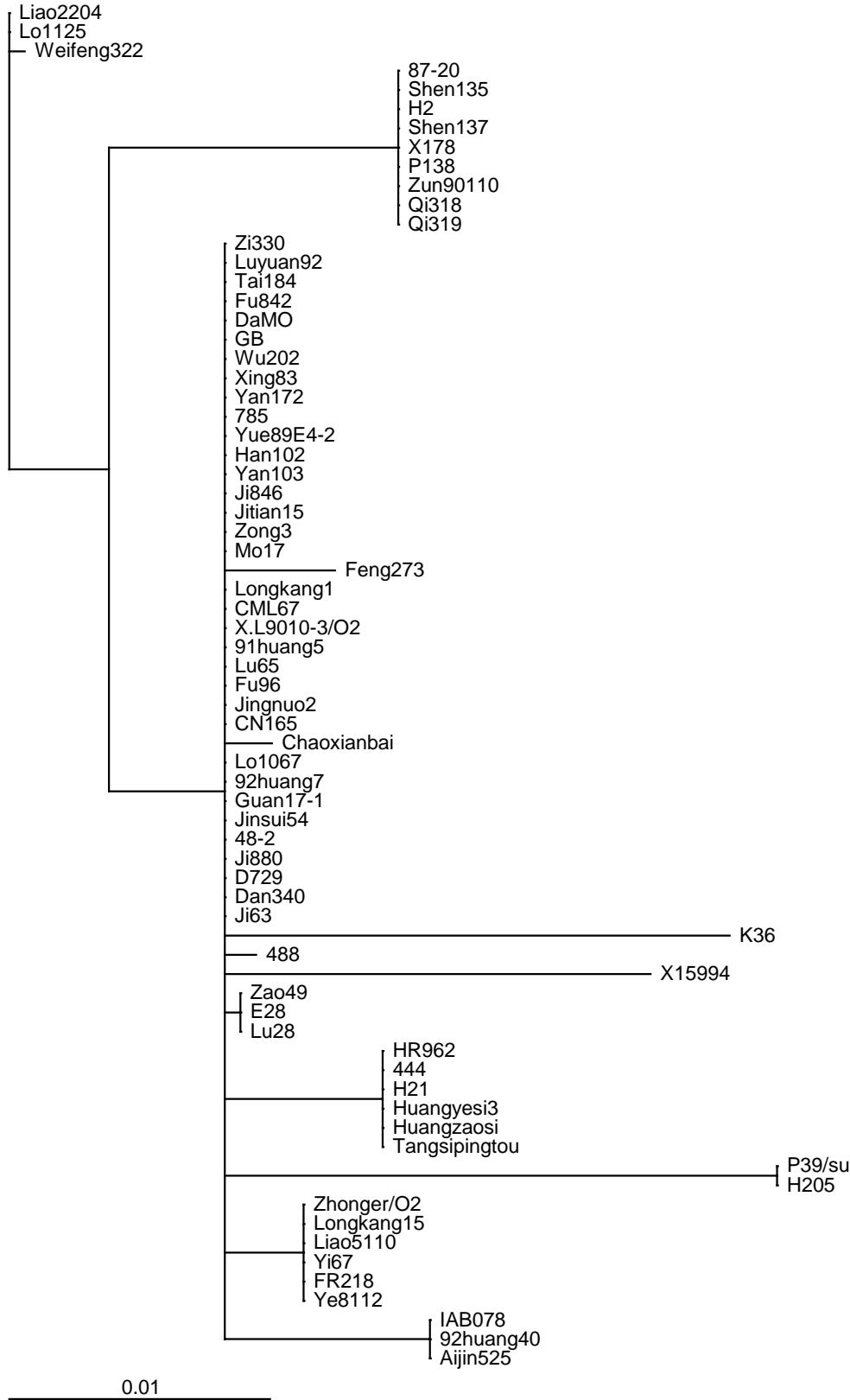


图 3-10 *rab17* 基因序列的系统发育树

Figure 3-10 Phylogram tree of *rab17* sequences

表 3-4 *rab17* 基因编码区内的核苷酸变异Table 3-4 Nucleotide polymorphisms in the coding region of *rab17*

	424 (C/G)	425 (G/C)	429 (G/T)	508 (A/G)	540 (C/A)	879 (Indel)	972 (G/A)
同义			Thr		Gly		
Synonymous							
非同义	Arg↔Ala	Arg↔Ala		Thr↔Ala		7bp	Gly↔Asp
Nonsynonymous							

以 X15994 为参照序列, 我们所检测的 70 个自交系均在 424 bp 和 425 bp 处存在核苷酸变异, 这两处 SNP 共同导致了该处精氨酸转变成了丙氨酸; 另外, 508 bp 处的 SNP (A→G) 导致了苏氨酸转变成了丙氨酸, 这个变异存在于 9 个自交系中 (齐 319、齐 318、遵 90110、P138、X178、沈 137、H2、87-20 和沈 135); 972 bp 处的 SNP (G→A) 使甘氨酸转变成了天冬氨酸, 该变异存在于三个自交系早 49、E28 和旅 28 中。

最值得关注的是 879 bp 处的一个 7 bp 的 Indel, 它不仅使 6 个属于塘四平头类群的自交系 (塘四平头、黄早四、黄野四 3、H21、444、HR962) 在该处编码的氨基酸发生了变化, 而且直接导致了随后的编码区发生了移码突变, 如果按照编码规律推断, 编码区将延长直至下一个终止密码子, 与其它自交系相比, 这 6 个自交系的编码蛋白将增加氨基酸达 25 个, 而实际中这 6 个自交系的 *rab17* 基因的编码蛋白结构是否真的发生了这样的变化, 还需要通过其它研究加以验证。

除两个同义突变外, 其它几个使蛋白质组成改变的变异有可能对编码蛋白的结构造成一定的影响 (特别是导致移码突变的 Indel 的存在), 甚至可能会影响基因的正常表达。

(二) 基因调控区域的多态性

除了编码区的序列变异有可能具有生物功能以外, 分析基因调控序列中的核苷酸变异同样具有重要的意义, 因为这些序列上的突变也可能引起表型的变异。例如, *tb1* 启动子中的序列变异和栽培玉米进化过程中的植株形态、发育以及基因表达等方面发生的变化相关 (Wang et al, 1999; Clark et al, 2004); *anthocyanin-less1* (*a1*) 位点启动子区域的两个序列多态性和 maysin 的合成存在显著关联 (Szalma et al, 2005); 其它作物中也有类似发现, 如小麦中 AP1 启动子的一个顺式元件中的序列变异被认为是春化作用的变异形成的原因 (Yan et al, 2003)。

在 *rab17* 的启动子区域, 包含着 9 个转录因子结合位点, 这些位点作为顺式元件在种子和营养生长阶段调节对 ABA 和脱水应答并调控基因的表达 (表 3-5, Busk et al, 1997)。其中, 5 个是 ABREs (abscisic scia responsive elements)。这些元件的序列一般来说具有较高的保守性。但是在本研究中也检测到了一些核苷酸变异。例如, 在其中一个与叶片中的 *rab17* 转录密切相关的 GRA 元件 (CACTGGCCGCCC) 序列中, 齐 319、齐 318、遵 90110、P138、X178、沈 137、H2、87-20 和沈 135 等 9 个自交系的序列中检测到一个核苷酸变异 [CACTGGCCG (C→T) CC]。而在上游相邻的 ABRE3a/3b 元件中, 自交系朝鲜白也检测到一个 SNP [TAC (C→T) GTGTACGTG]。

表 3-5 *rab17* 基因调控区域中的顺式元件

Table 3-5 Cis-elements in regulatory sequences of *rab17*

所测序列中的相对位置	元件	序列
Sites in the accession sequences	Elements	Sequences
76	DRE1	ACCGAGA
81	ABRE1	GACGTG
105	DRE2	ACCGAC
131	ABRE2	CACGTC
146	ABRE3a	TACGTG
152	ABRE3b	TACGTG
167	GRA	CACTGGCCGCCC
187	SPH	CATGCATG
197	ABRE4	CACGTA

另外，在启动子区域，还存在着一些虽然没有经过生化分析验证，但在公共数据库中被认为是顺式调控元件的区域（Buchanan et al, 2004）。比较发现，本研究中的一些自交系在这些区域也发生了变异。例如，上述齐 319 等 9 个自交系不仅在 9 bp 处的 DRE-LIKE 元件（ACCGAC）中缺失了一个 A，在 42 bp 处的 DRE-CORE（ACCG）元件中缺失了一个 C，而且也因为在 373 bp 处一个 9 bp 序列（AGCCGTGCA）的缺失而丢掉了 C-ABRE（CGTG）元件。在该 C-ABRE 元件处发生变异的还有威风 322、Lo1125 和辽 2204，它们的序列中元件的第一个 C 突变成了 G [(C→G) GTG]。类似的变异也发生在 331 bp 处，此处存在一个 DBF-CORE（CACAAG）元件，本研究中 2 个自交系（H205 和 P39/su）在该位置缺失了一个 CA，与之相反，威风 322、Lo1125 和辽 2204 却有一个 CA 的插入。这些变异是否与表型变异有关，还需要进一步研究。

3.4.3 *rab17* 基因序列系统发育树与自交系系谱关系

根据构建的系统发育树，71 个序列被分成 14 组，对应于 14 个单体型，将进化树与自交系的系谱来源进行比较后发现，有的单体型与系谱之间存在一定的对应关系。

9 个自交系（齐 319、齐 318、遵 90110、P138、X178、沈 137、H2、沈 135 和 87-20）在树中处于单独的一个分枝，且与大部分自交系有较远的遗传距离，仅与辽 2204、Lo1125、威风 322 三个自交系相对近一些。分析 *rab17* 基因序列的单体型结构后发现，这 9 个自交系以一个较强的单体型（单体型 14）存在，这个单体型由 11 个存在较强 LD 的核苷酸变异位点构成，包括 7 个 SNP 和 4 个 Indel，跨越序列长度达 700 多 bp。再从系谱来源来看，齐 319、齐 318、遵 90110、P138、沈 137 和沈 135 都选自美国先锋杂交种 P78599，而 H2 和 X178 也是从美国杂交种中选育而来，可以推断，这两个自交系可能是从 78599 或者与之亲缘关系极相近的其它杂交种选育而来，87-20 的系谱不清楚。在我们用 SSR 数据对这组核心自交系进行群体结构分析时，这 9 个自交系中除 87-20 外，其它 8 个拥有相似的遗传结构，并且组成了一个单独的 Cluster（第二章）。综合

以上这些结果可以推断出, 这些来自 P78599 或相近杂交种的自交系是一个独立存在的杂种优势群, 而且与其它类群之间有较远的亲缘关系。

辽 2204 与 Lo1125 同属一个单体型, Lo1125 来自 Pioneer ADA (PR3374), 辽 2204 也是选自一个美国杂交种, 这两个自交系与选自 P78599 的自交系遗传距离近, 是容易理解的,

单体型 11 由 6 个塘四平头群的自交系组成(塘四平头、黄早四、黄野四 3、H21、444 和 HR962), 而旅 28 与 E28 同属于单体型 13 中, 也很可能跟它们都含有旅大红骨的部分血缘有关。

这些对应关系暗示着, 基因的序列多样性在一定程度上可以反映出自交系的遗传血缘。

3.4.4 *rab17* 基因的 LD 结构及选择的影响

近来关于玉米基因组 LD 结构的报道很多, 一般认为, 由于玉米自身属异交作物, 重组率高, 因而在玉米基因组中会存在 LD 的快速衰退。然而, 驯化和改良过程中的人工选择还会对基因区域的 LD 水平产生影响。长期针对功能基因的等位变异进行正向选择, 会极大地增加了目标基因区域功能变异的 LD 水平。因此候选基因是在长期人工选择中承受了选择压力的驯化基因和改良基因, 那么这些基因位点就会依然保持高强度的 LD 结构。因而, 通过对候选基因区域 LD 进行估测, 就可以分析选择对该基因的影响 (Clark et al., 2004), 从而推断该基因在驯化和改良过程中是否承受了相应的选择压力。

在本研究中, 利用一组代表了我国最大遗传多样性和广泛地理分布的核心自交系, 对 *rab17* 基因将近 1700bp 的区域进行了分析。该区域包括了全部的转录区和主要的调控序列, 以及 5' 上游和 3' 下游的大部分序列。检测到了 LD 的快速衰退, 已衰退至 800bp 以内。这个结果类似于玉米中其它基因位点的研究结果 (Remington et al, 2001; Tenaillon et al, 2001)。Remington 等检测了基因内的以及遍及基因组的 SNP 间的 LD, 所用材料由 102 个自交系组成, 代表了育种中普遍使用的遗传材料中的遗传多样性。在对 6 个候选基因的调查中, 其中的 4 个基因 (*d3*、*id1*、*tb1*、*sh1*) 内的 LD 快速衰退 (在 1500bp 内 $r^2 < 0.1$)。这种快速衰退也在另一组 15 个基因的研究中观察到。然而, 有的位点 LD 衰退速率存在很大差异, *su1* 位点在 12kb 内就几乎没发生 LD 衰退 (Remington et al, 2001), 据推测可能与这个控制籽粒糖分含量的位点承受了长期选择压力有关。

尽管在本研究中, 检测到了 *rab17* 基因区域内的 LD 快速衰退, 然而, 并不能由此认为该基因在进化过程中没有承受到选择的压力。因为选择对基因区域的作用也可能是局部性的, 比如 *tb1* 位点, 一个曾承受选择压力的控制分枝的基因, Wang 等发现, 选择只对其调控区域产生了影响 (Wang et al, 1999)。

为此, 本研究对所得序列数据进行了 Tajima'D 测验, 测验结果不显著, 表明该基因在选择上是中性的, 选择没有对该基因的多样性与 LD 造成影响。

另外, LD 检测结果不仅与候选基因是否承受了选择压力有关, 也与选用的研究材料有关。在玉米核心自交系中, 由于其狭窄的遗传基础, LD 的衰退率较低 (Rafalski A, 2002)。一项对 *y1* 位点的研究表明, 在骨干玉米群体中 *y1* 位点的 LD 延伸至超过 100kb (Palaisa et al, 2003), 而在 *Adh1* 位点附近, LD 延伸则超过 500kb (Jung et al, 2004)。利用同样的骨干种质, Ching 和同事也在 300-500bp 内对 18 个基因的 LD 衰退进行研究, 实际上也没有发现 LD 衰退 (Ching A et al, 2002)。

然而, 如果选用的材料由代表着广泛的地理分布和较高的遗传多样性的自交系种质组成, 那

么依然会检测到 LD 的快速衰退。除 Remington 等的研究以外, Tenaillon 等也用一组多样性玉米种质调查了染色体 1 上的 21 个位点的序列多样性, 均存在 LD 的快速衰退, 甚至低至几百个 bp 以内 (Tenaillon et al, 2001)。我们选用的材料是来自玉米核心种质的一组能最大限度代表我国玉米遗传多样性的自交系。因而, 这组自交系应该具有较高的遗传多样性。进而, 利用这组种质检测到的候选基因的 LD 水平也应相应地较低一些。本研究中检测到 LD 衰退至 800bp 以内, 也与这样的推断相符合。

基因组中的 LD 结构对在玉米关联分析中采用哪一种关联方法有直接的影响。如果玉米基因组中普遍存在 LD 的快速衰退, 那么对玉米全基因组进行 LD 扫描来检测与表型性状关联的基因组区域的方法就会因需要的标记数目过高而难以实施。与之相反, 基于候选基因 LD 结构的关联分析却会因由于位于功能多态性两侧的 LD 衰退而使分辨率大大提高。本研究中 *rab17* 基因区域检测到的 LD 快速衰退, 也暗示着利用该候选基因的序列多态性进行关联研究, 将具有较高的分辨率。

第四章 *rab17* 基因核苷酸多态性与耐旱相关性状的关联分析

4.1 前言

关联分析方法是一项近几年才引入植物学研究的分析方法，曾广泛应用于人类遗传学研究中。该方法以自然种质群体为研究对象，能够把目标性状表型同基因的多样性结合起来进行分析，可以直接鉴定出与表型变异密切相关的具有特定功能的等位基因位点。而且不需要花费时间构建作图群体。因而相对于 QTL 而言，具有分辨率高、鉴定周期短、鉴定等位基因数目多等优点。

正如前面提到的，由于玉米基因组中具有核苷酸多样性高、重组率高、连锁不平衡快速衰退等特点（Remington et al, 2001; Tenaillon et al, 2001），而使玉米成为了一种进行关联分析，鉴定控制农艺性状相关基因的功能等位变异的模式植物（Thornsberry et al, 2001; Rafalski and Morgante, 2004）。

玉米的耐旱性是由多重数量性状位点（QTL）控制的数量性状，遗传机制相当复杂。近几年来，在耐旱性相关 QTL 定位、干旱胁迫诱导基因克隆等方面都取得了一定的进展。然而，许多耐旱相关表型变异可能是由于相关功能基因中的等位基因变异导致的，因而，充分发掘种质资源中蕴含的控制耐旱表型变异的相关基因中的功能变异，对于挖掘种质资源中的耐旱遗传资源，将具有极大的推动作用。而关联分析正是解决这一问题的有效手段。

rab17 基因是玉米体内一个受 ABA 和干旱胁迫诱导表达的基因（Kizis et al, 2002）。而且其基因位点和先前报道的许多控制干旱胁迫下的重要生理性状的 QTLs 处于同一区域，这些性状包括开花—吐丝间隔期（ASI）和产量性状等（Ribaut et al, 1996, 1997）。因而，该基因的一些等位基因变异很可能与上述耐旱相关性状有关。为此，我们利用前一部分工作中选出的一组核心自交系（第二章）对该基因的序列多样性进行了分析（见第四章），并在考虑这组核心自交系群体结构（见第三章）的基础上，对基因的序列多态性与耐旱相关表型进行了关联分析，以期鉴定出与耐旱性相关的 *rab17* 基因功能变异，为后继研究做一些基础性工作。

4.2 材料与方法

4.2.1 实验材料

采用在前一部分工作中选出的 94 份玉米自交系核心研究材料作为实验材料。

实验材料于 2005 年种植于新疆农科院安宁渠试验场国家灰漠土长期定位观测站内，于 5 月 7-8 两日播种。相同环境下设干旱胁迫区和正常灌水区，每个处理区种植三个重复，四周设水分隔离区。随机区组设计。行距 0.6 米，3 米行长，每行 10 株，株距 0.3 米。

4.2.2 耐旱性相关性状调查及测定

干旱处理区在大约开花前 10-15 天（中晚熟区 6 月 26 日，早熟区 6 月 22 日）停水，开始进行干旱胁迫。7 月 16 日有 10mm 降水，但影响不大，至 8 月 5 日，降水 34.5 mm，解除胁迫。

（一）形态性状及产量性状统计

胁迫开始后每天记录胁迫区及正常灌水区的各自交系的抽雄期、散粉期、吐丝期，计算开花一吐丝间隔期（ASI）；成熟期田间测量株高及穗位高；收获后考种，记录株穗数、单株穗重、单株粒重、百粒重等指标。

（二）相对含水量测定

在干旱胁迫 30 天后测量，同时在田间随机取点，于 20cm 土深，测量土壤含水量。同期对正常灌水区的自交系相对含水量也进行测定。

- a) 取样部位：上部第三片叶。
- b) 取样时间：上午 11 点至下午 1 点间。
- c) 取样方法：每行随机取 5 株，共取 5 片叶，约 1 克左右。
- d) 测量方法：叶片称量鲜重后加水饱和 4 个小时，然后称量饱和重，再用烘箱烘干，称量干重。依照下式计算相对含水量（RWC）：

$$RWC=100 \times (FW-DW) / (TW-DW)$$

其中，FW 代表鲜重；DW 代表干重；TW 代表饱和重。

（三）ABA 含量测定

ABA 含量采用酶联免疫吸附测定法（Enzyme-Linked ImmunoSorbent Assay, ELISA）测定（吴颂如等, 1988）。激素测定试剂盒购自中国农业大学。按照 ELISA 说明书，用酶联免疫检测仪进行测定。具体步骤如下：

1. ABA 的提取：

- a) 每行自交系材料随机取三株，每株剪取少量叶片混合在一起，共约 1g 左右。取回后用液氮速冻后保存在-78℃的超低温冰箱中备用；
- b) 提取时首先从取回的叶片中称取 0.5g 左右，加入 2ml 样品提取液，在冰浴下研磨成匀浆，转入 10ml 试管，再用 2ml 提取液分次将研钵冲洗干净，一并转入试管中，摇匀后放置在 4℃冰箱中；
- c) 样品在 4℃下提取 4h，1000 rpm 离心 15min，取上清液。沉淀中加 1ml 提取液，搅匀，4℃下再提取 1h，离心，合并上清液，弃去残渣；
- d) 将上清液过 C-18 固相萃取柱。具体步骤是：80%甲醇（1ml）平衡柱→上样→收集样品→移开样品后用 100%甲醇（5ml）洗柱→100%乙醚（5ml）洗柱→100%甲醇（5ml）洗柱→循环；
- e) 将过柱后的样品转入 10ml 塑料离心管中，置于 40℃水浴中，用氮气吹干，以除去提取液中的甲醇，随后用样品稀释液定容至 1ml。

附：

样品提取液配方：80%甲醇，内含 1 mmol/L BHT（二叔丁基对甲苯酚，为抗氧化剂），配制时先用甲醇溶解BHT，再配成 80%的浓度。

2. 样品 ABA 测定：

- a) 包被: 在 10ml 包被缓冲液中加入一定量的包被抗原(最适稀释倍数见试剂盒标签), 混匀, 在酶标板每小孔中加 100 μ l。然后将酶标板放入内铺湿纱布的带盖瓷盘内, 置于 37 $^{\circ}$ C 下 3h;
- b) 洗板: 将包被好的板取出, 放在室温下平衡。然后甩掉包被液, 将下口瓶内的洗涤液通过乳胶管均匀加入板上, 使每孔充满洗涤液, 放置约 0.5min, 再甩掉洗涤液。重复 3 次, 将板内残留洗涤液在吸水纸上甩干;
- c) 竞争: 即加标准物、待测样和抗体
- i. 加标样及待测样: 取样品稀释液 0.98ml, 内加 20 μ l 激素的标样试剂 (100 μ g / ml), 即为 2000ng/ml 标准液, 然后再依次稀释 1000ng/ml, 500ng/ml, 250ng/ml, 125ng/ml \cdots , 0ng/ml。将系列标准样加入 96 孔酶标板的前三行, 每个浓度加 3 孔, 每孔 50 μ l, 其余孔加待测样, 每个样品重复三孔, 每孔 50 μ l;
- ii. 加抗体: 在 5ml 样品稀释液中加入一定量的抗体(最适稀释倍数见试剂盒标签), 混匀后每孔加 50 μ l, 然后将酶标板加入湿盒内开始竞争;
竞争条件 37 $^{\circ}$ C 左右 0.5h;
- d) 洗板: 方法同包被之后的洗板。同时注意: ①加洗涤液时一定要从标准曲线的低浓度一边向高浓度一边加, 并且酶标板要向高浓度的一边倾斜。②第一次加入洗涤液后要立即甩掉。然后再接着加第二次。目的是防止各孔的交叉反应;
- e) 加二抗: 将一定量的酶标羊抗兔抗体, 加入 10ml 样品稀释液中(最适稀释倍数见试剂盒标签), 混匀后, 在酶标板每孔加 100 μ l, 然后将其放入湿盒内, 置 37 $^{\circ}$ C 下, 温育 0.5h;
- f) 洗板: 方法同竞争之后的洗板(步骤 d), 洗五次;
- g) 加底物显色: 称取 10-20mg 邻苯二胺(OPD)溶于 10ml 底物缓冲液中(小心勿用手接触 OPD), 完全溶解后加 2~4 μ l 30% H_2O_2 。混匀, 在每孔中加 100 μ l (在暗处操作), 然后放入湿盒内, 当显色适当后(即 0ng/ml 孔与 2000ng/ml 孔的 OD 差值为 1.0 左右时), 每孔加入 50 μ l 2mol/L 硫酸终止反应;
- h) 比色: 用 2000ng/ml 浓度孔(即标准曲线最高浓度孔)调 0, 在酶联免疫分光光度计上依次测定标准物各浓度和各样品 490nm 处的 OD 值。

附:

包被缓冲液: 称取 1.5g Na_2CO_3 , 2.93g $NaHCO_3$, 0.2g NaN_3 (可不加), 用量筒加 1000 ml 蒸馏水, pH 为 9.6;

磷酸盐缓冲液 (PBS): 称取 8.0g $NaCl$, 0.2g KH_2PO_4 , 2.96g $Na_2HPO_4 \cdot 12H_2O$, 用量筒加 1 000 ml 蒸馏水, pH 为 7.5;

样品稀释液: 100 ml PBS 中加 0.1 ml Tween-20, 0.1g 明胶 (稍加热溶解);

底物缓冲液: 称取 5.10g $C_6H_8O_7 \cdot H_2O$ (柠檬酸), 18.43g $Na_2HPO_4 \cdot 12H_2O$, 溶解定容至 1 000ml, 再加 1 ml Tween-20, pH 为 5.0;

洗涤液: 1000ml PBS 加 1ml Tween-20;

终止液: 2mol/L H_2SO_4 ;

3. 结果计算:

用 logit 曲线计算 ELISA 结果。曲线的横坐标用激素标样各浓度 (ng/ml) 的自然对数表示, 纵坐标用各浓度显色值的 logit 值表示。Logit 值的计算方法如下:

$$\text{Logit}(B/B_0) = \ln \frac{B/B_0}{1-B/B_0} = \ln \frac{B}{B_0-B}$$

其中 B_0 是 0ng/ml 孔的显色值, B 是其它浓度的显色值。

作出的 logit 曲线在检测范围内应是直线。待测样品可根据其显色值的 logit 值从图上查出其所含激素浓度 (ng/ml) 的自然对数, 再经过反对数即可知其激素的浓度 (ng/ml)。

求得样品中激素的浓度后, 用下式计算样品中激素的含量 (ng/g·fw):

$$A = \frac{N \cdot V_2 \cdot V_3 \cdot B}{V_1 \cdot W}$$

其中, A 表示激素的含量 (ng/g·fw);

V_2 表示提取样品后, 上清液的总体积;

V_1 表示进行浓缩干燥的上清液体积; (本实验中因为提取的上清液全部进行了浓缩干燥, 所以 $V_1 = V_2$)

V_3 表示用样品稀释液定容的体积;

W 表示样品的鲜重;

N 表示样品中激素的浓度 (ng/ml);

B 表示样品的稀释倍数 (样品稀释液定容以后的稀释倍数)。

4.2.3 *rab17* 基因序列多态性与表型性状间的关联分析

为了检测 *rab17* 基因序列多态性同耐旱相关性状之间是否存在关联, 在考虑群体结构影响的基础上, 进行序列多态性与耐旱相关性状间的关联分析。

关联分析用 TASSEL 软件中的 Structured Association (SA) 测验进行。

由于在序列多样性分析中发现, *rab17* 基因序列的多态性中, 大约 50% 为 Indel。因而, 在关联分析中, 将 SNP 和 Indel 一起作为多态性数据。选择等位基因频率大于 5% 的 SNP 和 Indel 用于关联分析。在控制群体结构影响的条件下, 用逻辑回归比率测验来估测关联程度 (Thornsberry et al, 2001)。在无效假设 H_0 中, 认为候选基因多态性与表型变异是两个相互独立的变量, 在假设 H_1 中, 则认为候选基因多态性和表型变异间存在关联。每个假设的可能性按下列方法进行比较:

$$\Lambda = \frac{\text{Pr}_1(C; T, \hat{Q})}{\text{Pr}_0(C; \hat{Q})}$$

其中, C 表示自交系的候选基因的基因型, T 是自交系的性状值, Q 为群体结构。而测验统计量 Λ 则表示两种模型 (Pr_1 和 Pr_0) 的自然对数似然性的差异。

为了得到准确的关联结果, 在运行时, Permutations 的数目设定为 1000。当多态性的 P 值小于 5% 时, 就认为该多态性和相应的性状值之间存在显著关联。

4.3 结果与分析

4.3.1 表型数据统计分析

本试验共获得了株高、穗位高、ASI 等形态指标, 株穗数、单株穗重、单株粒重、百粒重等产量指标以及相对含水量、ABA 含量等生理指标, 其中相对含水量和 ABA 含量取样时的土壤绝对含水量在胁迫区为 6.21%, 正常灌水区为 11.35%。经过方差分析, 各重复间的差异不显著, 因此, 在各指标估算时, 以三个重复的平均值为该指标的衡量值。

分析发现, 与对照相比, 总体上自交系群体干旱胁迫下的耐旱指标变化比较明显。在胁迫正常灌溉条件下, 各自交系的 ASI 在 0.3~14.3 天之间变化, 胁迫条件下在 0~19.5 天之间变化; 株高正常灌溉条件下在 134.3~263.3cm 之间变化, 胁迫条件下在 93.5~214.3cm 之间变化; 穗位高正常灌溉条件下在 30.3~147.3cm 之间变化, 胁迫条件下在 20.5~103cm 之间变化; 株穗数正常灌溉条件下在 0.44~3.34 之间变化, 胁迫条件下在 0.22~1.25 之间变化; 单株穗重正常灌溉条件下在 15.44~233.28g 之间变化, 胁迫条件下在 6.24~96.57g 之间变化; 单株粒重正常灌溉条件下在 10.86~186.39g 之间变化, 胁迫条件下在 5.28~67.95g 之间变化; 百粒重正常灌溉条件下在 8.38~39.67g 之间变化, 胁迫条件下在 10.05~33.83g 之间变化; 相对含水量正常灌溉条件下在 89%~106% 之间变化, 胁迫条件下在 76%~99% 之间变化; ABA 含量正常灌溉条件下在 81.12~2050.80 ng/g·fw 之间变化, 胁迫条件下在 161.131~2402.997 ng/g·fw 之间变化。

值得注意的是, 胁迫条件下, 一些自交系没有能够正常吐丝并结实, 造成了 ASI 与产量指标中一部分数据的缺失, 给结果带来了一些影响, 另外, 也存在试验过程中的误差, 如正常灌溉条件下有的自交系相对含水量测定值超过了 100%, 以后的类似试验中应改善试验条件, 尽量减小误差。还有一些数据误差是难以避免的, 如胁迫条件下造成雄穗及上部叶片的枯死给株高的测量带来的一些误差。

从群体内所有自交系的平均值来看, 各性状在胁迫条件下也均发生了一定的差异, 但变化幅度不同 (表 4-1)。其中开花-吐丝间隔期 (ASI) 变化幅度最大, 间隔天数相对于正常灌溉条件下延长达 76.84%, 单株穗重和单株粒重两个指标的变化幅度略低一些, 分别为 62.22% 和 64.66%。变化幅度最小的指标是相对含水量, 降低约 10% 左右。

然而, 以上统计只是胁迫后耐旱指标总体上的变化情况。而作为自交系本身固有的形态特征和生理特征, 群体内各自交系间的性状值本来就存在很大的差异, 在遭受干旱胁迫后, 各自交系间相应指标的变化也各不相同。有的自交系的 ASI 延长较为明显, 有的自交系则更多的表现为植株变矮或者叶片萎蔫, 这些变化幅度差异与耐旱性的强弱密切相关。在 ASI 统计中, 有的自交系胁迫后 ASI 变短, 在 ABA 含量的测定中, 有的自交系胁迫后含量低于对照, 这些是自交系自身的生理特性还是试验误差造成的, 还有待于进一步研究。

为了更多地了解这些表型鉴定指标与 *rab17* 基因表达间的关系, 本研究将正常灌溉条件下、干旱胁迫条件下的性状值均列为关联分析中的表型性状值, 并增加胁迫条件下的变化幅度为表型性状值。然后将这些表型性状与 *rab17* 基因多态性一起, 在考虑群体结构的基础上进行关联分析。



图 4-1 昌 7-2 在胁迫处理与正常灌溉条件下 ASI 的差异

Figure 4-1 Difference of ASI in Chang7-2 line between droughts stress and well water situation



图 4-2 自交系 GB 在胁迫处理与正常灌溉条件下植株高度的差异

Figure 4-2 Difference of Plant Height in GB line between droughts stress and well water situation

表 4-1 遭受干旱胁迫后的表型变化

Table 4-1 Variations of phenotype after drought stress

性状	正常灌溉（群体平均值）	胁迫处理（群体平均值）	胁迫后的变化幅度
ASI(天)	3.80	6.72	76.84%
株高 cm	198.75	152.47	23.29%
穗位高 cm	82.25	62.86	23.57%
株穗数	1.14	0.73	35.96%
单株穗重 g	104.62	39.53	62.22%
单株粒重 g	82.57	29.18	64.66%
百粒重 g	27.95	23.30	16.64%
ABA (ng/g-fw)	629.39	713.32	13.34%
相对含水量	0.98	0.89	9.18%

4.3.2 *rab17* 基因序列多态性与表型性状间的关联分析

因为在 *rab17* 基因序列多样性分析过程中只得到了 70 个自交系的序列，因而在关联分析时，从群体结构数据（第二章）和总计 29 个耐旱表型性状数据中抽出相应的 70 个材料的数据，和 70 个 *rab17* 基因序列数据（第三章）一起，利用 TASSEL 软件进行 Structured Association (SA) 测验，测验中选择等位基因频率大于 5% 的多态性（共 16 个），采用逻辑回归方法，将序列数据视为双等位性数据，复等位形式的等位位点将被合并。1000 次 Permutations 后，当多态性的 P 值小于 5% 时，就认为该多态性和相应的性状值之间存在显著关联。最终获得的关联分析结果见表 4-2。

一共检测到六个与表型存在显著关联的序列多态性位点。其中，879 bp 处的 7 bp 的 Indel 分别被检测到与正常灌溉条件下的 ABA 含量及干旱胁迫条件下的相对含水量存在显著关联，前者显著水平达到极显著；331 bp 处的 2 bp 的 Indel 与干旱胁迫条件下的开花—吐丝间隔期 (ASI) 的关联也达到显著性水平；另外，由 11 个多态性组成的单体型（单体型 14，见第三章）与干旱胁迫下单株粒重、干旱胁迫下单株穗重之间存在极显著关联，与正常灌溉下单株穗重之间存在显著关联，这 11 个多态性之间存在高水平的 LD ($r^2 > 0.9$; $P < 0.0001$)。

分析这几个多态性位点的遗传效应后发现，不同的等位位点均引起了相应的表型上的变异。

879 bp 处的 Indel 处，含有该插入片段[CCACGGC (CCACGGC) GAC]的等位基因所对应的正常灌溉下 ABA 含量平均值为 348.62 ng/g-fw，干旱胁迫下相对含水量 III 的平均值为 93.35%，而缺失该 7 bp 片段的等位位点[CCACGGC (-----) GAC]对应的正常灌溉下 ABA 含量平均值为 660.32 ng/g-fw，干旱胁迫下相对含水量 III 的平均值为 88.81%，含有该插入片段[CCACGGC (CCACGGC) GAC]的等位位点对应的正常灌溉下 ABA 含量相对较低，同时干旱胁迫下仍能够保持较高的相对含水量。

单体型 14（含 9 个自交系，大部分来自 P78599）对应的胁迫下单株穗重、对照单株穗重、胁迫下单株粒重分别是 49.79g、129.22g 和 34.03g，与之相比，另一个等位基因对应的胁迫下单株穗重、对照单株穗重、胁迫下单株粒重则分别是 38.80g、101.05g 和 28.82g，相应值均相对较

低，表明，那些主要来自 P78599 的自交系中含有的相应等位基因对这三个产量指标均起到了增值效应。

另外检测到的一个显著关联的多态性是 331 bp 处的 Indel，该处存在三个等位基因，三个等位点之间依次相差一个 CA，分别是[AGCACA (----) AGCA]、[AGCACA (CA--) AGCA]和[AGCACA (CACA) AGCA]，对应的干旱胁迫条件下的 ASI 依次为 5.17、6.13 和 11.89。三个等位点中以[AGCACA (CA--) AGCA]频率最高，而含有[AGCACA (----) AGCA]和[AGCACA (CACA) AGCA]的自交系分别仅有 2 个和 3 个。由于 TASSEL 软件在进行关联测验时还只能进行双等位基因形式测验，所以多个等位基因中的低频率等位点会被合并，而这三个等位基因中，低频率位点[AGCACA (----) AGCA]会被合并掉，又因其性状值与频率最高的[AGCACA (CA--) AGCA]性状值差异不大，因而，对于干旱胁迫条件下的 ASI 起到显著增值效应的应该是[AGCACA (CACA) AGCA]，该变异位点对应的性状值比另一个等位基因高 94%。

需要指出的是，以各鉴定指标胁迫后的变化幅度为表型数据，均没有检测到显著的关联。这暗示着耐旱指标的胁迫后变化幅度，可能不适于应用到关联分析中。

表 4-2 *rab17*多态性或者单体型与性状间的关联
Table 4-2 Association between *rab17* sequence polymorphisms and traits

多态性 or 单体型	性状	DiffLL	PermuteP
879 (+/-)	正常灌溉下 ABA 含量	12.2076	0.0020
879 (+/-)	干旱胁迫下相对含水量	5.3526	0.0321
单体型 14	干旱胁迫下单株粒重	9.7639	0.0041
单体型 14	干旱胁迫下单株穗重	9.7642	0.0071
单体型 14	正常灌溉下单株穗重	7.3063	0.0131
331 (+/-)	干旱胁迫下 ASI	6.1579	0.0302

注：DiffLL表示两种模型（ H_1 和 H_0 ）的自然对数似然性的差异；PermuteP表示基于Permutations的P值。

4.4 讨论

4.4.1 多态性变异的功能性

本研究中一共检测到三个与表型显著关联的多态性变异。这些多态性变异可能导致了一定的生物学功能上的变化，而这些变化有可能是导致表型变异的原因。

其中，879 bp 处的变异是一个 7 bp 的 Indel，含有插入片段[CCACGGC (CCACGGC) GAC]的等位点对应的自交系在干旱胁迫下仍能够保持较高的相对含水量（93.35%），而缺失该片段的等位基因对应的相对含水量为 88.81%。一般情况下，胁迫下保持较高的相对含水量意味着较强的耐旱性，因而，这暗示着该含有插入片段的等位基因对耐旱性具有正向贡献。与系谱结合分析，发现含有该等位基因的 6 个自交系均来自公认的塘四平头群（塘四平头、黄早四、黄野四 3、H21、444、HR962）。黄早四是我国从地方品种中选育出的优良骨干自交系，耐旱性强，这与本研究中

得到的结果是一致的。而且, 结合上一章中分析出的序列信息, 这个 Indel 位于编码区中, 7 bp 片段的插入(相当于增加了一个 CCACGGC 拷贝)不仅使得该处编码的氨基酸发生了变化, 而且直接导致了随后的编码区发生了移码突变。那么, 这个移码突变是否使 *rab17* 基因的表达发生了变化, 以及这种变化是否能够从内在机制上解释干旱胁迫条件下仍能保持高水平的相对含水量, 还有待于进一步研究。

同时我们也注意到, 该多态性还与正常灌溉条件下 ABA 含量显著相关, 并且含有插入片段的等位基因使相应的 ABA 含量较低。然而, 当胁迫条件下 ABA 含量增加后, 却没有再检测到这个多态性与 ABA 含量的显著相关。其内在机理还不清楚, 有可能跟 ABA 的代谢及调控途径有关。

另外, 本研究还检测到由 11 个处于 LD 状态的多态性组成的单体型 14 与干旱胁迫下单株粒重、干旱胁迫下单株穗重之间存在极显著关联, 与正常灌溉下单株穗重之间存在显著关联, 这三个指标都是产量因子。9 个自交系(大部分来自 P78599)所含有的等位基因对这三个指标均起到了增效作用, 暗示着该等位基因也可能是对耐旱性具有正向贡献。单体型 14 包含的 11 个多态性中, 有的可能是真正导致表型变异的功能多态性, 其它的则是由于 LD 结构而表现出显著相关。分析序列后发现, 这 11 个多态性中的一个 SNP (C→T) 位于启动子区域的 GRA 元件内, 该元件与叶片中的 *rab17* 转录密切相关 (Busk et al, 1997), 另外还有三个多态性位于虽然没有经过生化分析验证, 但在公共数据库中被认为是顺式调控元件的区域 (Buchanan et al, 2004) (详见第三章)。有的研究中已经发现, 基因调控序列中的核苷酸变异也可能引起表型的变异 (Yan et al, 2003; Clark et al, 2004; Szalma et al, 2005)。因而, 本研究中这几个位于调控元件区域内的多态性更可能是导致表型变异的功能多态性。

331 bp 处也存在一个 DBF-CORE 元件, 此处的 Indel 被检测到与干旱胁迫条件下的 ASI 的关联, 该多态性位点存在三个等位基因, 其中威风 322、Lo1125 和辽 2204 含有的等位基因表现出了强烈的增效作用, 使干旱胁迫下的 ASI 比另一个等位基因延长了 6 天, 因而该等位基因被看作是对耐旱性有负向效应, 相应地, 另一个等位基因则可能是对耐旱性有正向贡献的等位基因。

4.4.2 可能影响关联检测结果的一些因素

由于本研究中尽量保证了实验数据的准确可靠性, *rab17* 基因的 LD 快速衰退也保证了关联分析的高分辨率, 而且也通过考虑群体结构的方法尽可能地避免了自交系的群体结构可能带来的假阳性关联结果, 因而, 关联结果应具有较高的可信度。尽管如此, 仍然有一些因素可能影响关联结果:

首先是表型数据, 尽管我们在表型鉴定过程中努力使可能存在的误差减到了最小限度, 但由于胁迫处理在大田条件下进行, 一些自然因素依然是无法控制的, 这些因素或多或少都会对关联结果造成一定的影响。另外, 由于时间、条件所限, 我们的实验只拥有一年一点的统计数据, 而多年多点的表型鉴定应该更为准确和可靠。

其次, 统计方法也可能带来一定的影响。比如, 本研究所用的 TASSEL 软件目前在进行 Structured Association (SA) 测验时, 仍只能考虑双等位基因形式测验, 所以多个等位基因中的低频率等位点会被合并。本研究中检测到 331 bp 处的三个等位基因与干旱胁迫下的 ASI 存在显著关联, 并经分析, 找出了其中解释 ASI 增加的等位基因, 然而, 另外一个低频率的等位基因仍

可能会对关联结果产生一定的影响。更灵敏、全面的关联检测能力依赖于统计软件的进一步完善。

再次,群体结构也会影响关联分析结果。尽管本研究通过考虑群体结构的方法尽可能地避免了自交系的群体结构可能带来的假阳性,然而,随之可能另外一个问题,即假阴性的出现,也就是说,一些与群体结构高度相关的多态性也可能是能够解释表型变异的功能多态性或者与之存在关联,但在控制群体结构影响时,这种多态性被认为是群体结构造成的影响而忽略掉了。由于我们所用的群体大小的限制,关联分析时也可能存在假阴性。也就是说,可能存在与耐旱表型相关关联的多态性,因为该多态性与群体结构高度相关,在排除群体结构影响时已经被滤掉,没有在关联结果中体现出来。以前的类似研究中已经发现了这种可能,例如,在近来一项验证 *d8* 位点功能的研究中,对 71 个欧洲核心自交系进行了 *d8* 序列多态性检测和开花期表型鉴定(Andersen et al, 2005)。在没有控制群体结构影响的前提下检测到了显著关联,然而当群体结构被控制后,就没有再检测到关联的存在。相反,在另外一个用 375 个玉米自交系组成的大规模关联作图群体进行的 *d8* 序列多态性与开花期之间的关联分析中,这种关联却得到了验证(Camus-Kulandaivelu et al, 2006)。因而,减少这种假阴性的最有效的办法就是增大群体规模,群体规模越大,这种假阴性存在的可能性就越小。另外还有一种方法,就是用其它的群体来对该候选基因进行重新估测。

最后,无法检测到低频率的等位基因的作用。本研究中,为了保证测验时关联结果的可信度,等位基因频率小于 5% 的多态性变异被过滤掉,不参加关联测验,这样也使一些可能与耐旱相关表型存在显著关联的变异位点无法通过关联分析检测出来。通过增大群体规模使功能多态性变异频率增加,会增加相应的关联检测能力。

总之,尽管存在以上有待改进和完善的因素,但本研究的结果仍然具有较高的可靠性和实用性,检测到的几个与耐旱表型相关的 *rab17* 等位基因为开发相应的功能分子标记提供了分子基础,也为深入研究 *rab17* 基因在玉米耐旱机制中的作用提供了重要的思路。

第五章 结 论

5.1 玉米核心自交系的遗传多样性及群体结构分析

利用 49 个 SSR 标记对来自玉米核心种质及骨干自交系的 288 个自交系进行遗传多样性分析, 共检测到总计 262 个等位位点, 平均每个 SSR 位点等位位点数为 5.3469。PIC 值平均为 0.5057, 基因多样性平均值为 0.5663。群体结构分析发现, 这些自交系按遗传相似性主要可分为四个种质来源, 基本对应于塘四平头、Lancaster、Reid 和 P 群四个杂种优势群。总体上, 四大种质来源所占的遗传相似性比例依次为 0.290 (塘四平头)、0.267 (Lancaster)、0.260 (Reid) 和 0.183 (P 群)。以等位位点数目为选择标准, 从这组自交系中选出了 94 个自交系作为核心研究材料。这组核心材料共拥有 228 个等位位点, 占 288 份自交系等位位点总数 (262) 的 87%。

5.2 *rab17* 基因的序列多样性及连锁不平衡结构分析

对核心研究材料进行 *rab17* 基因的序列多态性分析。在总计 71 个序列中, 共检测到了 37 处核苷酸变异, 包括 19 个 SNP (平均每 87bp 一个) 和 18 个 Indel (平均每 92bp 一个)。在 37 个变异位点中, 在非编码区中的有 30 个, 占 81%。其中 Indel 在非编码区的发生频率是编码区的将近八倍。总体上来看, 在 *rab17* 基因的启动子区及第一个外显子区域中核苷酸多样性较高, 之后迅速下降。连锁不平衡分析显示在该基因区域内连锁不平衡快速衰退达 800bp 以内。根据连锁不平衡结构, 这些序列分为 14 个单体型。根据 *rab17* 基因序列的系统进化树分析表明, 主要来自 P78599 的 9 个自交系与其它自交系间的遗传距离最远。

5.3 *rab17* 基因序列多态性和耐旱相关表型之间的关联分析

通过 Structured Association (SA) 测验, 共检测到六个序列多态性与表型间的显著关联。其中, 编码区的一个 Indel 分别与正常灌溉条件下的 ABA 含量及干旱胁迫条件下的相对含水量存在显著关联, 含有插入片段的等位基因对于干旱胁迫条件下的相对含水量具有增效作用, 这个插入片段也导致了编码区的移码突变; 5'-UTR 中的一个 Indel 与干旱胁迫条件下的开花—吐丝间隔期 (ASI) 间的关联达到显著, 其中威风 322、Lo1125 和辽 2204 含有的等位基因使 ASI 平均比另一个等位基因延长了 6 天; 另外, 由 11 个处于 LD 状态的多态性组成的单体型与三个产量因子之间也存在显著关联, 9 个大部分来自 P78599 的自交系所含有的等位基因对这三个指标均起到了增效作用, 表明该等位基因也对耐旱性具有正向贡献。

参考文献

1. 丁雷,王学臣.干旱胁迫下 ABA 对气孔运动的作用机制.干旱地区农业研究, 1993. 11(2):50-56
2. 高翔, 曹绍书, 罗仕文, 王进, 祝云芳, 沈建华, 陈泽辉. 浅析塘四平头杂优群选系及其改良系在中国玉米育种和生产中的作用. 玉米科学, 2005, 13 (增刊):20-22,25
3. 关义新, 戴俊英, 徐世昌. 土壤干旱下玉米叶片游离脯氨酸的累积及其与抗旱性的关系. 玉米科学, 1996, 4:43-45
4. 黄益勤, 李建生. 利用 RFLP 标记划分 45 份玉米自交系杂种优势群的研究. 中国农业科学, 2001, 34:244-250
5. 霍仕平, 晏庆九, 宋光英, 许明陆. 玉米抗旱鉴定的形态和生理生化指标研究进展. 干旱地区农业研究, 1995a, 13: 67-73.
6. 鞠方成, 王春莲. 旅大红骨杂种优势群及其在玉米育种上的应用. 丹东纺专学报. 2003, 10:7-8
7. 黎裕, 王天宇, 石云素, 宋燕春. 玉米抗旱性的 QTL 分析研究进展和发展趋势. 干旱地区农业研究. 2004b, 22:32-39
8. 黎裕, 王天宇, 刘成, 石云素, 宋燕春. 玉米抗旱品种的筛选指标研究. 植物遗传资源学报, 2004a, 5:210-215
9. 李新海, 高根来, 梁晓玲等. 我国主要玉米自交系开花期耐旱性及改良. 作物学报, 2002, 28: 595-600.
10. 李新海, 袁力行, 李晓辉, 张世煌, 李明顺, 李文华. 利用 SSR 标记划分 70 份我国玉米自交系的杂种优势群. 中国农业科学, 2003, 36:622-627
11. 李艳天, 丁朝瑞, 张运伟, 李继昌, 黄玉禄, 赵文龙 玉米“自 330”的血缘与系谱分析. 玉米科学 2002, 10:39-40
12. 罗淑平. 玉米抗旱性及鉴定指标的相关分析. 干旱地区农业研究, 1990, 3: 72-78.
13. 宋凤斌, 徐世昌. 玉米抗旱性鉴定指标的研究. 中国生态农业学报. 2004, 12:127-129
14. 孙彩霞, 武志杰, 张振平, 陈利军. 玉米抗旱性评价指标的系统分析. 农业系统科学与综合研究. 2004, 20:43-47
15. 滕文涛, 曹靖生, 陈彦惠, 刘向辉, 景希强, 张发军, 李建生. 十年来中国玉米杂种优势群及其模式变化的分析. 中国农业科学. 2004, 37(12): 1804-1811
16. 王畅等. 夏玉米的干旱适应性及其生理机制的研究. 华北农学报, 1990. 5:54-60
17. 王茅雁, 邵世勤, 张建华, 耿庆汉. 水分胁迫对玉米保护酶系活力及膜系统结构的影响. 华北农学报. 1995. 10:43-49
18. 王懿波, 王振华, 陆利行, 王永普, 张新, 田曾元. 中国玉米种质基础、杂种优势群划分与杂优模式研究. 玉米科学. 1998, 6:9-13
19. 王懿波, 王振华, 王永普, 张新, 陆利行. 中国玉米主要种质杂种优势利用模式的研究. 中国农业科学. 1997, 30(4): 16-24.
20. 王懿波, 王振华, 王永普, 张新, 陆利行. 中国玉米主要种质杂种优势群的划分及其改良利

- 用。华北农学报, 1998, 13:74-80
21. 王元东, 段民孝, 邢锦丰, 王继东, 张春原, 郭景伦, 赵久然, 陈绍江。P 群种质在玉米杂种优势利用和种质创新中的作用及展望。玉米科学, 2004, 12:10-12,15
 22. 吴景锋。我国主要玉米杂交种种质基础评述。中国农业科学。1983, 16:1-8
 23. 吴敏生, 王守才, 戴景瑞。AFLP 分子标记在玉米优良自交系优势群划分中的应用。作物学报, 2000, 26:9-13
 24. 吴颂如, 陈婉芬, 周燮 酶联免疫法(ELISA)测定内源植物激素。植物生理学通讯。1988。5:53-57.
 25. 袁力行, 傅骏骅, Warburton M, 李新海, 张世煌等。利用 RFLP、SSR、AFLP 和 RAPD 标记分析玉米自交系遗传多样性的比较研究。遗传学报, 2000, 27:725-733
 26. 袁力行, 傅骏骅, 张世煌, 刘新芝, 彭泽斌, 李新海。利用 RFLP 和 SSR 划分玉米自交系杂种优势群的研究。作物学报, 2001, 27:149-156
 27. 曾三省。中国玉米杂交种的种质基础。中国农业科学, 1990, 23:1-9
 28. 张宝石, 徐世昌, 宋凤斌, 张威, 戴俊英。玉米抗旱基因型鉴定方法和指标的研究。玉米科学, 1996, 4:19-22
 29. 张吉民, 刘成, 石云素, 宋燕春, 白宝璋, 黎裕, 王天宇。2004。干旱胁迫和正常灌溉条件下玉米开花相关性状的 QTL 分析。植物遗传资源学报, 5(2):161-165
 30. 张烈, 沈秀瑛, 孙彩霞, 刘金刚。ABA 与玉米抗旱性关系的研究。玉米科学。1998. (增刊): 42-44.
 31. 赵久然, 郭景伦, 郭强, 尉德铭, 孔艳芳。应用 RAPD 分子标记技术对我国骨干玉米自交系进行类群划分。华北农学报, 1999, 14:32-37
 32. 郑常祥, 杨文鹏, 舒世德, 傅良中。干旱条件下玉米产量的主要相关性状研究。玉米科学。2002. 10(增刊):59-61
 33. 郑大浩, 李艳茹, 金锋学, 蒋基建。中国玉米 Lancaster 杂种优势群自交系的系谱、种质基础分析。中国农业科学。2002, 35:750 – 757
 34. 中国农业年鉴。中国农业出版社。2005 年, 总 26 卷
 35. Agrama HA S, Moussa M E. Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). Euphytica, 1996. 91: 89- 97.
 36. Andersen J.R., Schrag T., Melchinger A.E., Zein I. and Lubberstedt T., Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.), Theor Appl Genet. 2005. 111:206–217
 37. Baker J, Steele C, Dure L III. Sequence and characterization of 6 LEA proteins and their genes from cotton. Plant Mol Biol 1988. 11: 277-291.
 38. Bhattaramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, Tingey SV, Rafalski A: Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 2002. 48:539-547
 39. Bhattaramakki D, Rafalski A: Discovery and Application of Single Nucleotide Polymorphism Markers in Plants. In: Plant Genotyping: The DNA Fingerprinting of Plants (Edited by: Henry RJ). Wallingford Oxon, UK: CABI Publishing 2001.

40. Bohnert H J, Jensen R G. Strategies for engineering water-stress tolerance in plants. *Trends Biotechnol*, 1996, 14: 89-97.
41. Bolanos J, Edmeades GO. Eight cycles of selection for drought tolerance in lowland tropical maize . Responses in reproductive behavior. *Field Crops Res*, 1993, 31: 253-268.
42. Bolanos J, Edmeades GO. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. *Field Crops Res*.1996, 48:65-80
43. Botstein D, White RL, Skolnick M and Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*. 1980. 32:314-331
44. Bray EA. Molecular responses to water deficit. *Plant Physiol*, 1993. 103:1035-1040
45. Bray EA. Plant responses to water deficit. *Trends Plant Sci*, 1997, 2: 48-54.
46. Breseghello F and Sorrells ME. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics* 2006. 172: 1165–1177
47. Brown AHD. Core collections: a practical approach to genetic resource management. *Genome*. 1989, 31:818-824
48. Brown GR, Gill GP, Kuntz RJ, Langley CH, and Neale DB. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. USA*. 2004. 101:15255-15260
49. Bruce W, Edmeades G O, Barker T C. Molecular and physiological approaches to maize improvement for drought tolerance. *J Exp Botany*, 2002, 53: 13-25.
50. Brunner S., Fengler K., Morgante M., Tingey S. and Rafalski A. Evolution of DNA sequence nonhomologies among maize inbreds, *Plant Cell*. 2005a. 17:343–360
51. Brunner S., Pea G. and Rafalski A. Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize, *Plant J*. 2005b. 43:799–810
52. Buchanan C.D., Klein P.E., Mullet J.E. Phylogenetic Analysis of 5'-Noncoding Regions From the ABA-Responsive *rab16/17* Gene Family of Sorghum, Maize and Rice Provides Insight Into the Composition, Organization and Function of cis-Regulatory Modules. *Genetics*. 2004. 168:1639-1654
53. Buckler ES IV, Thornsberry JM, Kresovich S. Molecular diversity, structure and domestication of grasses. *Genet. Res. Camb*. 2001. 77:213-218
54. Buckler ES, Thornsberry JM. Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol*. 2002, 5:107-111
55. Buckler ES, Gaut BS and McMullen MD. Molecular and functional diversity of maize. *Current Opinion in Plant Biology* 2006, 9:172–176
56. Busk P.K., Jensen A.B. and Pagès M. Regulatory elements in vivo in the promoter of the abscisic acid responsive gene *rab17* from maize. *Plant J*. 1997. 11: 1285–1295
57. Camus-Kulandaivelu L, Veyrieras J-B, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A. Maize adaptation to temperate climate: relationship with population structure and polymorphism in the *Dwarf8* gene. *Genetics* 2006; published ahead of

- print [doi: 10.1534/genetics.105.048603](https://doi.org/10.1534/genetics.105.048603)
58. Cattivell L, Baldi P, Crosatti C, et al. Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. *Plant Mol Biol*, 2002, 48: 649-665.
 59. Chapman SC, Edmeades GO. Selection improves tolerance to mid/late season drought in tropical maize populations. II. Direct and correlated responses among secondary traits. *Crop Science*, 1999, 39:1315-1324
 60. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, and Rafalski AJ. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *Biomed. Central Genetics* 2002. 3:19-32
 61. Clark RM, Linton E, Messing J, Doebley JF. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA*. 2004. 101:700-707.
 62. Close TJ, Chandler PM. Cereal dehydrins: serology, gene mapping and potential functional roles. *Aust J Plant Physiol*. 1990. 17:333-344
 63. Close TJ. Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physio Plant*. 1996. 97:795-803
 64. Daly NJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet*, 2001. 29:229-232
 65. Delvin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995. 29:311-322
 66. Devlin B. and Roeder K., Genomic control for association studies, *Biometrics*. 1999. 55:997-1004
 67. Estoup A, Tailliez C, Cornuet JM, Solignac M. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol Biol Evol*, 1995. 12:1074-1084
 68. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS: Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 1998, 95:4441-4446.
 69. Excoffier L, Laval G, and Schneider S. Arlequin ver 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005. 1:47-50
 70. Falconer DS, Mackay TF. *Introduction to Quantitative Genetics*. Essex, UK: Longman Group Ltd. 1996. 464pp
 71. Falush D., Stephens M. and Pritchard J.K., Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics*. 2003. 164:1567-1587
 72. Felsenstein J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
 73. Fisher RA. The logic of inductive inference. *J. R. Stat. Soc. Ser. A*. 1935. 98:39-5
 74. Flint-Garcia SA, JM Thornsberry, ES Buckler IV, Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol*. 2003. 54:357-374.
 75. Fontaine A, Barriere Y, Caffeic acid O-methyltransferase allelic polymorphism characterization and analysis in different maize inbred lines. *Mol. Breed*. 2003. 11:69-75.

76. Frova C et al., Genetic analysis of drought tolerance in maize by molecular markers. I. Yield components. *Theor. Appl. Genet.* 1999. 99:280-288
77. Fu H, Dooner HK. Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* 2002. 99:9573-9578
78. Fu H, Zheng Z, Dooner HK. Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Natl. Acad. Sci. USA* 2002. 99:1082-1087
79. Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez-G J, Liu K, Buckler ES, Doebley J. Genetic diversity and population structure of teosinte. *Genetics*. 2005, 169:2241-2254
80. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. Genetic structure and diversity in *oryza sativa* L. *Genetics*. 2005, 169:1631-1638
81. Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC: Maize as a model for the evolution of plant nuclear genomes. *Proc Natl Acad Sci USA* 2000, 97:7008-7015.
82. Gaut BS, Peek AS, Morton BR, Clegg MT: Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). *Mol Biol Evol* 1999, 16:1086-1097.
83. Gaut BS and Long AD. The lowdown on linkage disequilibrium, *Plant Cell*. 2003. 15:1502-1506
84. Gebhardt C, A Bellvora, B Walkemeier, P Oberhagemann, K Schuler, Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breed.* 2004. 13:93-102.
85. Gerdes, J.T., C.F. Behr, J.G. Coors, and W.F. Tracy. 1993. Compilation of North American maize breeding germplasm. CSSA, Madison, WI.
86. Goday A, Jensen A.B., Culiáñez-Macia F.A., M. Mar Alba, Mercè Figueras, Joan Serratosab, Margarita Torrent, and Montserrat Pagès. The Maize Abscissic Acid-Responsive Protein *Rab17* Is Located in the Nucleus and Interacts with Nuclear Localization Signals. *The Plant Cell*, 1994. 6:351-360
87. Goldstein DB. Islands of linkage disequilibrium. *Nat Genet*, 2001. 29:109-111
88. Golubinoff P, Paabo S, Wilson AC: Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archeological specimens. *Proc Natl Acad Sci USA* 1993, 90:1997-2001.
89. Gomez J, Sanchez-Martinez D, Steifel V, Rigau J, Puigdomenech P, Pages M. A gene induced by the plant hormone abscissic acid in response to water stress encodes a glycine-rich protein. *Nature*. 1988. 334:262-264
90. Hagenblad J, Nordborg M. Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. *Genetics*. 2002. 161:289-298
91. Hall AJ, Vilella f, Trapani N, Chimenti C. The effects of water stress and genotype on the dynamics of pollen-shedding and silking in maize. *Field Crops Research*, 1982. 5:349-363
92. Hamblin M.T., Mitchell S.E., White G.M., Gallego J., Kukatla R., Wing R.A., Paterson A.H. and Kresovich S. Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of sorghum bicolor, *Genetics*. 2004. 167:471-483

93. Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics*, 1996. 143:1395-1407
94. Heisey PW & GO Edmeades, 1999. Maize production in drought-stressed environments: technical. Options and research resource allocation Part1 of CIMMYT 1997/98 World Maize Facts and Trends.
95. Hilton H, Gaut BS: Speciation and domestication in the maize and its wild relatives: evidence from the globulin-1 gene. *Genetics* 1998. 150:863-872.
96. Hirschhorn J.N. and Daly M.J., Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005. 6:95-108
97. Hong B, Barg R, Ho DT. Developmental and organ specific expression of an ABA and stress induced protein in barley. *Plant Mol Biol*, 1992. 18:663-674.
98. Hudson RR. Linkage disequilibrium and recombination. In *Handbook of Statistical Genetics*. Edited by Balding DJ, Bishop M, Cannings C. Chichester: John Wiley and Sons, Ltd; 2001:309-324
99. Ingram J, Bartels D. The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol Plant Mol Biol*, 1996. 47:377-403.
100. Ingvarsson P.K., Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae), *Genetics*. 2005. 169:945-953
101. Jannoo N, Grivet L, Dookun A, D'hont A, Glaszmann JC. 1999. Linkage disequilibrium among modern sugarcane cultivars. *Theor. Appl. Genet*. 99:1053-1060
102. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*, 2001. 29:217-222
103. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F. et al. Haplotype tagging for the identification of common disease genes. *Nat Genet*. 2001. 29:233-237
104. Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res*. 2000. 10:1435-1444
105. Jung M., Ching A., Bhatramakki D., Dolan M., Tingey S., Morgante M. and Antoni Rafalski. Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. *Theor and Appl Genet*. 2004. 109:681-689
106. Kizis D., Pages M., Maize DRE-binding proteins DBF1 and DBF2 are involved in *rab17* regulation through the drought-responsive element in an ABA-dependent pathway. *Plant J*. 2002. 30: 679-689
107. Kraakman ATW, Martínez F., Mussiraliev B., Eeuwijk F.A. van and Niks R.E. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Molecular Breeding*. 2006. 17:41-58
108. Kraakman ATW, Niks RE, Petra M. M. M. Van den Berg, Piet Stam and Fred A. Van Eeuwijk.

- Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. *Genetics* 2004. 168: 435–446
109. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet.* 1999. 22:139-144
110. Kruskal J.B. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* 1956. 7:48-50
111. Lai J., Li Y., Messing J. and Dooner H.K. Gene movement by Helitron transposons contributes to the haplotype variability of maize, *Proc Natl Acad Sci USA* 2005. 102:9068–9073
112. Lai J., Ma J., Swigonova Z., Ramakrishna W., Linton E., Llaca V., Tanyolac B., Park Y.J., Jeong O.Y. and Bennetzen J.L. et al. Gene loss and movement in the maize genome, *Genome Res* 2004. 14:1924–1931
113. Lebreton C, Lazic-Jancic V, Steed A, et al. Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. *J Exp Bot*, 1995. 46:853- 865.
114. Li XH, XD Liu, MS Li, SH Zhang, 2003. Identification of quantitative trait loci for anthesis-silking interval and yield components under drought stress in maize. *Acta. Bot. Sin.* 45:852-857.
115. Li Y, Du J, Wang T, Shi Y, Song Y, Jia J. Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. *Maydica.* 2002, 47:93-101
116. Li Y, Shi YS, Cao YS, Wang TY. Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. *Genetic Resources and Crop Evolution.* 2004, 51: 845-852
117. Li Y. Development and germplasm base of maize hybrids in China. *Maydica* 1998. 43:259-269.
118. Lindblad-Toh K, Winchester E, Daly M, Wang D, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES: Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 2000, 24:381-386.
119. Liu K, Muse SV. PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics.* 2005, 21:2128-2129
120. Liu KJ, Goodman M, Muse S, Smith JS, Buckler E, Doebley J. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics.* 2003, 165:2117-2128
121. Mackay T.F., The genetic architecture of quantitative traits, *Annu Rev Genet.* 2001. 35:303–339
122. Matsuoka Y., Mitchell S.E., Kresovich S., Goodman M., and Doebley J. Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. *Theor Appl Genet* 2002. 104:436-450.
123. Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF et al. Sequence composition and genome organization of maize. *Proc. Natl. Acad. Sci. USA* 2004. 101:14349-14354
124. Morgante M., Brunner S., Pea G., Fengler K., Zuccolo A. and Rafalski A. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet.*

2005. 37:997-1002
125. Mundy J, Chua HH. Absciscic acid and water-stress induce the expression of a novel rice gene. EMBO J. 1988. 7:2279-2286
126. Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J et al. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 2002. 30:190-193
127. Nordborg M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial selffertilization. Genetics. 2000. 154:923-929
128. Nordborg M., Hu TT, Ishino Y., Jhaveri J., Toomajian C., Zheng H., Bakker E., Calabrese P., Gladstone J. and Goyal R. et al. The pattern of polymorphism in *Arabidopsis thaliana*, PLoS Biol. 2005. 3:e196
129. Palaisa KA, Morgante M, Williams M, Rafalski A, Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell. 2003. 15:1795-1806.
130. Pla M, Goday A, Vilardell J, Gomez J, Pages M. Differential regulation of the ABA induced 23-25 kD proteins in embryos and vegetative tissues of the *viviparous* mutants of maize. Plant Mol Biol. 1989, 13:385-394
131. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) marker for germplasm analysis. Mol Breeding, 1996. 2:225-238
132. Prim R.C. Shortest connection networks and some generalizations. Bell Syst. Tech. J. 1957. 36:1389-1401
133. Pritchard J.K. and Rosenberg N.A. Use of unlinked genetic markers to detect population stratification in association studies, Am J Hum Genet. 1999. 65:220-228
134. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000a, 155:945-959
135. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000b, 67:170-181
136. Pritchard JK. Deconstructing maize population structure. Nature. 2001. 28:203-204
137. Quarrie SA et al., Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J. Exp. Bot. 1999. 50:1299-1306.
138. Rafalski A, Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. 2002. 5:94-100.
139. Rafalski A, Morgante M. Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. 2004, 20:103-111
140. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES, Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA. 2001. 98:11479-11484.
141. Ribaut JM, Banziger M, Hoisington D. Genetic dissection and plant improvement under abiotic

- stress conditions: drought tolerance in maize as an example. JIRCAS Working Report, 2002, 85-92
142. Ribaut J-M, Hoisington DA, Deutsch JA, Jiang C, Gonzalezde-Leon D. Identification of quantitative trait loci under drought conditions in tropical maize. I. Flowering parameters and the anthesis-silking interval. *Theor and Appl Genet.* 1996. 92: 905-914
143. Ribaut J-M, Jiang C, Gonzalez-de-Leon D, Edmeades GO, Hoisington DA. Identification of quantitative trait loci under drought conditions in tropical maize. II. Yield components and marker-assisted selection strategies. *Theor and Appl Genet.* 1997.94: 887-896
144. Riccardi F, Gazeau P, Vienne D de, Zivy M, Protein changes in reponse to progressive water deficit in maize. Quantitative variation and polypeptide identification. *Plant Physiol.* 1998. 117:1253-1263.
145. Risch N. and Merikangas K. The future of genetic studies of complex human diseases, *Science.* 1996. 273:1516-1517
146. Romero-Severson J, Smith JSC, Ziegler J, Hauser JL, Hookstra G. Pedigree analysis and haplotype sharing within diverse groups of *Zea mays* L. inbreds. *Theor Appl Genet.* 2001, 103:567-574
147. Saghai-Maroo MA, Soliman KL, Jorgensen RA et al. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc Natl Acad Sci USA*, 1984,81:8014-8018
148. Sanguineti MC, Tuberosa R, Landi P, Salvi S, Maccaferri M, Casarini E, Conti S. QTL analysis of drought-related traits and grain yield in relation to genetic variation for leaf abscisic acid concentration in field-grown maize. *J Exp Bot*, 1999, 50:1289-1297
149. Sari-Gorla M et al., Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. *Theor. Appl. Genet.* 1999. 99:289-295.
150. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology.* 2000, 18:233-234
151. Selinger DA, Chandler VL: Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize. *Proc Natl Acad Sci USA* 1999, 96:15007-15012.
152. Senior M.L., Murphy J.P., Goodman M.M., and Stuber C.W. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci* 1998. 38:1088-1098.
153. Shinozaki K, Yamaguchi-Shinozaki K. Gene Expression and Signal Transduction in Water-Stress Response. *Plant Physiol.* 1997. 115:327-334.
154. Shinozaki K, Yamaguchi-Shinozaki K. Molecular responses to drought and cold stress. *Curr Opin Biotechnol.* 1996. 7:161-167
155. Soltis PS. and Soltis DE. Genetic variation in endemic and widespread plant species: examples from Saxifragaceae and Polystichum. *A liso.* 1991, 13:215-223.
156. Syvanen AC: Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001, 2:930-942.

157. Szalma S.J., Buckler E.S., Snook M.E. and McMullen M.D., Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks, *Theor Appl Genet.* 2005. 110:1324-1333
158. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989. 123:585-595
159. Tanksley SD, McCouch SR, Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 1997. 277:1063-1066.
160. Taramino G. and Tingey S. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 1996. 39:277-287.
161. Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS, Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). *Proc. Natl. Acad. Sci. USA* 2001. 98:9161-9166.
162. Teulat B, This D, Khairallah M. et al. Several QTLs involved in osmotic-adjustment trait variation in barley (*Hordeum vulgare* L). *Theor Appl. Genet.* 1998. 96:688-698
163. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. and Higgins D.G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research.* 1997.25:4876-4882
164. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. *Dwarf8* polymorphisms associated with variation in flowering time. *Nat. Genet.* 2001. 28:286-289.
165. Tuberosa R, Salvi S, Sanguineti M C, Landi P, Maccaferri M, Conti S. Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. *Ann Bot*, 2002, 89: 941- 963.
166. Tuberosa R, Sanguineti M C, Landi P, et al. RFLP mapping of quantitative trait loci controlling abscisic acid concentration in leaves of drought-stressed maize (*Zea mays* L.) . *Theor Appl Genet*, 1998. 97: 744- 755.
167. Viard F, Franck P, Dubois MP, Estoup A, Jarne P. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J Mol Evol*, 1998. 47:42-51
168. Vilardell J, Goday A, Freire MA, Torrent M, Martinez MC, Torne JM, Pages M. Gene sequence, developmental expression, and protein phosphorylation of RAB-17 in maize. *Plant Mol Biol.* 1990. 14:423-432.
169. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 1998. 280:1077-1082
170. Wang R.L., Stec A., Hey J., Lukens L. & Doebley J. The limits of selection during maize domestication. *Nature*.1999. 398:236-239
171. Weir BS. *Genetic Data Analysis II*. Sunderland, MA: Sinauer. 1996. 376 pp.
172. Westgate M E, Boyer J S. Reproduction at low silk and pollen water potentials in maize. *Crop Sci*, 1986, 26: 951-956
173. White SE, Doebley JF. The molecular evolution of terminal ear 1, a regulatory gene in the genus

- Zea. Genetics. 1999. 153:1455-1462
174. Whitt SR, Buckler ES IV. Using natural allelic diversity to evaluate gene function. *In* Plant Functional Genomics: Methods and Protocols (E Grotewold ed.). Humana Press, Inc., Totowa, NJ. 2003. pp.123-139.
175. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. USA*. 2002. 99:12959-12962
176. Wilkes HG. Hybridization of maize and teosinte in Mexico and Guatemala and the improvement of maize. *Economic Bot*. 1977. 31:254-293
177. Wilson LM, Whitt SR, Ibañez AM, Rocheford TR, Goodman MM, Buckler IV ES. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. *Plant Cell*. 2004, 16:2719-2733
178. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. The Effects of Artificial Selection on the Maize Genome. *Science*. 2005. 308:1310-1314
179. Xu D, Duan X, Wang B, et al. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol*, 1996. 110:249-257.
180. Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS. and McMullen MD. A Large-Scale Screen for Artificial Selection in Maize Identifies Candidate Agronomic Loci for Domestication and Crop Improvement. *The Plant Cell*. 2005. 17:2859-2872
181. Yan L., Loukoianov A., Tranquilli G, Helguera M., Fahima T et al. Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. USA* 2003. 100:6263-6268
182. Yu J and Buckler ES. Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*. 2006a. 17: 155-160
183. Yu J., Pressoir G., Briggs W.H., Bi IV, Yamasaki M., Doebley J.F., McMullen MD, Gaut BS, Nielsen D.M. and Holland J.B. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006b. 38:203-208
184. Yu LX, TL Setter, Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. *Plant Physiol*. 2003. 131:568-582.
185. Zhu YL, Song QJ, Hyten DL et al. Single-Nucleotide Polymorphisms in Soybean. *Genetics* 2003. 163: 1123-1134
186. Zinselmeier C, Y Sun, T Helentjaris, M Beatty, S Yang, H Smith, J Habben, The use of gene expression profiling to dissect the stress sensitivity of reproductive development in maize. *Field Crops Res*. 2002. 75:111-121

致 谢

时光匆匆，转眼间三年的博士学习生活结束了，回首三年来，每一个普普通通的日子仿佛都历历在目。在这里首先要感谢的是我的导师王天宇研究员和黎裕研究员。我的学位论文是在两位老师的悉心指导下完成的，从论文的选题，试验的设计、实施，到文稿的修改，处处都凝聚着两位老师的心血。他们开阔的学术视野、严谨的科研态度和忘我的工作精神使我终身受益！两位老师三年来的培养和启迪，学生会铭记在心！

还要特别感谢石云素副研究员和宋燕春老师，两位老师几年来无论是在学业上还是在生活中，都给了我无微不至的关怀和帮助。真诚地祝福她们！

论文得以顺利完成，得益于玉米资源研究室这个团结友爱的集体。我的部分实验是在王荣焕博士协助下完成的，此外，路运才博士、黄素华博士、李会勇博士、王阳博士、李永祥博士、贾小平博士、刘志斋博士、吴建伟硕士、公强硕士、赵丽娟硕士、张中保硕士、王秀堂硕士、毛毅辉硕士、赵曦硕士等以及程春明、杜月红两位访问学者都曾经在实验技术或者具体工作中给予了热心的帮助。另外，刘金辉、张美琴和张淼等同学在实验过程中也曾经协助做了大量的工作，在此一并表示衷心地感谢！

我的部分实验是在新疆农科院实验基地进行的。在这里要特别感谢新疆农科院的刘成老师在工作上 and 生活中给予的大力支持和热心帮助。

此外，还要感谢同宿舍的康俊根博士和李瑞博士三年来在学业上和生活中的关心和帮助。李孟军博士、陈吉宝博士等在实验技术上也曾经给予了指导和建议，这里一并表示感谢！

还要感谢研究生院和所科研处的各位老师三年来在学业上的培养！

本项研究得到了国家攻关项目“玉米、高粱、粟类基因资源保护与创新利用研究”（2004BA525B04）、国家自然科学基金“玉米抗旱相关基因的关联分析与功能标记的开发”（30571133）、国际合作项目“玉米种质资源抗旱新基因和等位基因发掘及其利用”（0502）等项目的资助，也在此表示感谢！

在此还要郑重感谢我的父亲、母亲、姐姐和爱人吴红霞女士，我能够顺利完成学业，与他们在生活上的支持和精神上的鼓励是分不开的！

最后，把感谢和祝福送给三年中所有曾经关心和帮助过我的朋友们！

于 永 涛

2006.6