

大豆遗传图谱研究进展及对应的几个问题^{*}

许占友 常汝镇 邱丽娟 李向华

(中国农业科学院作物品种资源研究所 北京 100081)

中图分类号 S565.1 文献标识码 A 文章编号 1000-9841(2001)02-0133-05

大豆是由古四倍演变而来的二倍体作物(Lackey, 1980),其基因组大小为 129×10^9 181×10^9 bp 之间(Arumanagathan, 1991)。由于大豆的基因组较大,染色体又很小,而且细胞在有丝分裂时染色体压缩,使细胞不易观察,因此,大豆的细胞遗传学研究难度较大。另外,由于大豆是严格的自花授粉作物,其遗传变异程度低,导致大豆遗传研究,尤其是遗传作图研究,与其它作物相比较为落后。

DNA 分子标记出现以后,以其数量丰富,遗传稳定及操作简单等特性,大大促进了大豆的遗传作图,使连锁图谱的标记密度越来越大,连锁图上的标记数从 1987 年的 57 个增至 1999 年的 1423 个。遗传图谱在许多研究领域,从植物改良中的标记辅助选择育种到分子遗传研究中的图位克隆(Map-based clone)都发挥了重要作用。因此,对于得到广泛研究的重要经济作物大豆来讲,开发出一张高密度精密的遗传图谱,其上包括经典的形态标记、同工酶标记和大量的高信息度的 DNA 分子标记,将具有重要意义。在近二十年来,中国和美国先后共研制了 18 张遗传图谱,其中美国 16 张(Soybase, ACEDB V4.3),中国 2 张,随着研究的逐步深入,图谱密度越来越高。所用的作图群体包括种间和种内杂交得到的群体, F_2 群体和重组自交系群体。回顾遗传图谱的研究过程,大豆遗传图谱的发展大致可分为三个阶段:第一阶段,农艺性状和同工酶为基础的连锁图;第二阶段,以 RFLP 标记为基础的连锁图;第三阶段,以 SSR 标记为主,各种标记整合在一起的高密度连锁图。下面分别对三个阶段作一简单回顾并对存在的问题进行初步分析。

第一阶段: 常规标记为主的遗传图谱

大豆的常规标记包括形态、色素和同工酶等标记,其遗传图谱(Classical Linkage Group, CLG)发展速度较慢,研究也不系统,到 1973 年才仅有 7 个连锁群(Bernard & Weiss, 1973),1985 年达到 13 个(Palmer & Kilen, 1985),在大豆分子图谱建成以前,尽管已有 250 个形态和同工酶标记,但大部分没有被定位。其中只有 57 个标记被定位于 17 个连锁图上,长度为 420 cM(Palmer & Keim, 1987)。六年后,标记数增加到 63 个,分布于 20 个连锁图谱上(Palmer, 1993),截止到 1999 年,大豆的常规标记仍仅有 68 个定位于 20 个连锁群上,平均每一个连锁图上有 3.4 个位点(Palmer & Shoemaker, 1998)。大豆常规标记遗传图谱发展慢的原因有二:一是单个分离群体内的性状分离少。一个分离群体的亲本不可能包括所要研究的各个性状,结果一个分离群体仅有少数几个形态标记被定位。二是不同作图群体间较难进行整合,不同群体间缺乏桥梁进行比较。如果有多态性高,而且不同群体间可以进行横向比较的标记,这将为遗传图谱的绘制及利用提供极大的方便。1980 年 RFLP 分子标记问世,科学家随即应用于分子标记遗传图谱的绘制,从此大豆遗传图谱的绘制进入第二个阶段。

第二阶段: 以 RFLP 标记为主的遗传图谱

RFLP 标记作为第一代 DNA 分子标记,因其重复性好,共显性,因此被广泛用于基因作图、多样性分析和进化研究等。1988 年,Apuya 等应用大豆两个栽培品种 Minsoy 和 Noir 1 杂交得到的 F_2 群体,

^{*} 收稿日期:2000-03-20

作者简介:许占友(1968-),男,副研究员,研究方向作物品种资源。

构建了世界上的第一张 RFLP 图谱, 图谱包括 4 个连锁群, 其上共有 11 个 RFLP 标记。

1990 年, Keim 等应用栽培大豆(*G. Max*) 品系 A81—356022 和野生大豆(*G. Soja*)PI468916 杂交的 F₂ 代作为作图群体, 用 RFLP 标记构建了第二张大豆分子连锁图谱, 该图谱包括 130 个 RFLP 标记组成的 26 个连锁群, 覆盖的大豆基因组为 1200cM。随后蛋白质、油分、铁营养等的数量性状位点(QTL)加到该连锁图(Diers, 1992), 到 1993 年, 这张图已包括了 490 个 RFLP 和 RAPD 标记, 分布在 20 个连锁群上, 覆盖基因组达 3000cM。

1993 年 Shoemaker 和 Olson 同样使用栽培大豆 A81—356022×野生大豆 PI468916 杂交得到的 F₂ 群体, 构建了包括 25 个连锁图的遗传图谱, 该图谱共有 383 个标记, 其中 RFLP 标记有 365 个, RAPD 标记 11 个, 同工酶标记 4 个, 经典性状标记 3 个。

Lark 等(1993)与 Apuya 一样, 应用 Minsoy 和 Noir 1 杂交组合的 F₂; F₃ 群体构建了包括 31 个连锁图的遗传图谱, 其中分布有 132 个标记(RFLP, 同工酶, 形态和化学标记), 覆盖基因组 1550 cM。

以 RFLP 标记为基础的 DNA 分子标记遗传图谱绘制发展速度很快, 但如果它不整合尽可能多的形态标记, 同工酶标记等常规标记, 将限制其在分子标记辅助选择育种和遗传研究中的利用。大豆的上述几个分子图谱也包括少数几个常规标记, 因为这些标记在用于作图的群体中幸运地表现了分离, 致

使其与分子标记同时被定位(Keim 1990; Dier 1992; Lark 1993; Shoemaker 1993)。另外, 为了在分子遗传图谱上加入更多的常规标记, Landan 等(1991)通过超级结瘤基因的 F₂ 分离群体, Diers 等(1992)对大豆蛋白质含量、油份含量以及铁营养建立了 F₂ 分离群体, 进行了 QTL 分析, 并检测这些位点与已定位的分子标记间的连锁, 并由此推断这些标记所应属的分子连锁群。但是这些 F₂ 群体中表现分离的标记数目是有限的, 往往不足以进一步确定常规标记在某个分子连锁群上的特定位置。为了实现将较多的常规标记整合到分子连锁图谱上, Nikell 等(1994), Polzin 等(1994)先后分别构建了特殊群体, 想把尽量多的常规标记加到分子连锁图上, 使人感到不足的是, 绝大多数的常规标记仍未被整合到分子连锁图谱中。直到 Shoemaker 和 Specht(1995)通过近等基因系 Clark×Harosoy 杂交得到 F₂ 分离群体, 在该群体中共计有 20 个形态和同工酶标记表现分离, 据此, 通过 JoinMap 作图软件将这些标记与已定位 RFLP 座位连锁的常规标记整合到一张遗传图谱上, 这张图是当时被公认的第一张“公共图谱”该图包括 13 个经典标记, 7 个同工酶标记, 110 个 RFLP 标记和 8 个 RAPD 标记, 同时通过使用一套锚定的 RFLP 探针, 根据其在 Clark×Harosoy 和 A81—356022×PI468916 这两个群体中的分离, 确定了分子和经典图谱的同源连锁群。

表 1 大豆遗传图谱的发展过程

Table 1 Progress of the genetic linkage mapping in soybean

时间 Time	作图亲本 Parents of the population	作图群体 Population	标记数 No. marker	连锁群 Linkages	覆盖长度 Length	参考文献 References
1987	/	/	57	17	420 cM	Palmer
1988	Minsoy 和 Noir 1	F ₂	11	4	197cM	Apuya
1990	A81—356022 和 PI468916	F ₂	130	26	1200 cM	Keim
1993	Minsoy 和 Noir 1	F ₂	132	31	1550 cM	Lark
1993	A81—356022 和 PI468916	F ₂	383	25	3000 cM	Shoemaker
1995	A81—356022 和 PI468916	F ₂	138	26	/	Shoemaker
1995	Clark &Harosoy	F ₂	172	29	1486 cM	Akkaya
1996	Minsoy 和 Noir 1	RIL(重组自交系)196		30	1981 cM	Mansur
1997	BSR—101 和 PI437654	RIL	840	28	3441 cM	Keim
1997	长农 4×新民 6 号	F ₂	71	20	1446.8 cM	张德水等
1998	上述三个群体	F ₂ & RIL	68	20	/	Palmer
1999	上述三个群体	F ₂ & RIL	1423	20		Cregan
1999	长农 4×新民 6 号	RIL	240	22	3713.5 cM	刘峰等

虽然 Shoemaker 等构建了第一张整合有较多常规标记的“公共图谱”, 图谱上的标记数目与其它作

物相比仍少的多, 仅有 138 个标记。当时最多的遗传图谱上的标记数目也仅有 383 个, 按大豆的染色

体组有 20 条染色体计算, 平均每个染色体上仅有 19 个。原因在于 RFLP 标记在大豆上多态性较低。而恰在此时, Vos 等(1995)发明了扩增片段长度多态性(AFLP)标记, 其多态性比 RFLP 高。因此, 为了解决遗传图谱的密度问题, Keim 等(1997)应用 BSR101×PI437654 构建了含有 300 个单株的重组自交系群体, 通过该群体, 构建了 RFLP 框架图, 然后选取其中 42 个重组自交系的单株进行 AFLP 标记多态性比较, 结合 RFLP 标记构建了包括 28 个连锁群的遗传图谱, 其上分布有 650 个 AFLP 标记, 165 个 RFLP 标记, 25 个 RAPD 标记, 覆盖基因组达 3441 cM, 平均每 4cM 便有一个标记。其结果表明, 虽然 AFLP 标记经常成簇排列, 但相对于 RFLP 和 RAPD 标记, AFLP 更均匀地分布在所有连锁群中。

张德水等(1997)应用栽培大豆长农 4 号与半野生大豆新民 6 号杂交, 通过 84 个 F₂ 单株, 以 RFLP 标记为主进行连锁群分析, 构建了 20 个连锁群, 共分布有 71 个标记, 其中 RFLP 标记 63 个, RAPD 标记 8 个, 标记间的平均距离为 20.4 cM, 覆盖长度为 1446.8 cM, 是我国构建的首张大豆分子连锁图谱。

RFLP 标记自 1980 年发明以来, 便因其重复性好, 共显性而被广泛应用于生物学的各个领域。尤其是用于生物遗传作图。但用于研究大豆的遗传图谱, 特别是把已公开发表的各个遗传图谱整合到一套公用的图谱时, 发现 RFLP 用以绘制大豆的遗传图谱时存在两大缺点(Cregan, 1999): 一是 RFLP 标记在检测大豆基因组时, 很少检测到多于两个的等位基因, 因为这些等位基因通常有不对称的频率, e.g. $P > 0.9, q < 0.1$ (Keim 等, 1989, 1992)。在一个特定的 RFLP 位点任意两种基因型有多态性的可能性相对较低。当作图群体的亲本都来自改良种质库时, 这一点尤为明显。例如, Muehlbauer(1991)等也观察到在供体亲本、近等基因系或轮回亲本检测时, 仅有 1/3 的 RFLP 探针在亲本间有多态性, Shoemaker 等(1995)报道在 A81-356022×PI468916 的 F₂ 群体中 365 个多态性的 RFLP 标记, 仅 118 个在 Clark×Harosoy 作图群体中分离, 因此在一个群体中定位的多态片段在另一个群体中可能不分离。二是大多数 RFLP 标记在大豆基因组中存在着多于 2 个拷贝的序列。Shoemaker 等(1996)对 280 个随机挑选的 *pst*I 基因组克隆与完全酶切的大豆基因组 DNA 进行 Southern 杂交, 结果显示, 绝大多数基因

组克隆得到的平均杂交条带数目不是一条带, 有 90% 以上的探针检测到 2 条以上的条带, 近 60% 的探针检测到 3 条或以上的条带。证明大豆基因组中存在着大量的重复序列, 基因组中的单拷贝序列不到 10%, 其余大部分可能不仅基因组发生了四倍化, 而且经历了基因组的自复制。Shoemaker 等(1996)对来自栽培大豆×栽培大豆和栽培大豆×野生大豆的 9 个不同群体的 800 多个标记综合分析得到一张整合后的图谱, 通过 RFLP 探针杂交, 检测到不同群体间存在着广泛的同源关系。同源的片段从 1.5cM 到 106.4cM, 平均长度为 45.3cM。每个区段平均复制了 2.55 次, 有的达到 6 次。另外, 还发现“巢式”(nested)复制, 说明其中大豆的一个原始基因组可能在进化过程中进行了额外的四倍体化。由于基因组序列的复制, 在整合不同的 RFLP 遗传图谱至一套图谱时, 产生了位点是否相同的疑点。因为一个 RFLP 探针在基因组中可检测到多个位点。在一个群体中用某一个探针检测到的多态性片段代表一个位点, 而相同的探针在另外群体中检测到的多态性片段可能代表了另一个完全不同且不连锁的位点。这与不同群体中哪一个位点具有多态性有关。虽然通过使用特定的探针/酶组合确定了相同的多态性位点, 从而在图谱上定位相同的位置而解决这一困难, 然而最终解决此问题, 只有发展多态性更高且只有一个位点的分子标记—简单重复序列(Simple Sequence Repeat, SSR)或微卫星标记。从而使大豆的遗传图谱研究进入第三个阶段。

第三阶段: 以 SSR 标记为主的遗传图谱绘制

简单重复序列广泛分布于大豆基因组中。Akkaya 等(1992)首次对大豆基因组中微卫星 DNA 进行了研究, 证明其不仅大量存在并且具有高度的多态性, 大豆中简单重复序列的核心碱基(AT)_n 和(ATT)_n 比例较高, 平均每 40kb 就有一个(AT)_n 微卫星 DNA。在遗传图谱的绘制上, SSR 标记可以弥补 RFLP 标记的两点不足, 首先 SSR 标记多态性比较高, RFLP 标记很少能检测到两个以上的等位基因, 而 SSR 标记可以检测到多个等位基因, Jiang 等(1995)在 96 个样品中最多可检测到 26 个等位基因。其次, SSR 标记是在基因组中仅存在一个拷贝的共显性标记(Akkaya, 1992), 在整合不同的遗传

图谱时, 不会产生不确定因素。

基于 SSR 标记的上述两个优点, Akkaya 等 (1995)通过近等基因系 Clark×Harosoy 杂交得到的 F₂ 代 60 个单株, 用 40 个 SSR 标记对这个分离群体进行研究, 结果 34 个 SSR 标记定位于 29 个连锁中的 18 个, 同 13 个传统性状中的 9 个性状, 7 个同工酶位点中的 2 个连锁。证明 SSR 标记随机分布于大豆基因组中。但在 18 个连锁图中发现 2 个连锁图中有 SSR 标记密集区, 一个是第 13 连锁图中的 5 个 SSR 座位覆盖了 23. 4cM, 第 18 连锁中的 4 个 SSR 座位覆盖了 33. 6 cM, 研究结果进一步证明, SSR 标记在遗传图谱绘制上是比 RFLP 等标记更好的工具。

Mansur 等 (1996)通过两个栽培大豆 Minsoy 与 Noir 1 的杂交, 得到了 284 个 F₇ 代的重组自交系单株, 用 45 个 SSR 标记对其中的 223 个单株进行分析, 把 22 个 SSR 标记加到了这个图谱上。

刘峰等 (1999)应用栽培大豆长农 4 号和半野生大豆新民 6 号杂交得到的 F₈ 代重组自交系, 构建了我国第二张较高密度的遗传图谱, 该图谱共有 240 个标记, 其中包括 2 个形态标记, 100 个 RFLP 标记, 33 个 SSR 标记, 42 个 AFLP 标记, 62 个 RAPD 标记和 1 个 SCAR 标记, 分布在 24 个连锁图上, 总长度为 3713. 5 cM。

Cregan 对于大豆遗传图谱的绘制具有里程碑作用。1999 年, Cregan 等通过 606 对 SSR 引物, 同时对三个大学的被认为是“公共图谱”的三个连锁图谱进行整合。这三个作图群体分别为:

(1) 美国农业部/依阿华州立大学分离群体

(USDA/Iowa state university population); 这是栽培大豆自交系 A81—356022 和野生大豆 PI468916 种间杂交的 F₂ 作图群体, 这一群体由 59 个 F₂ 植株组成。

(2) 犹它大学重组自交群体 (Univ. of Utah RILpopulation); 这是通过两个栽培大豆品种 Minsoy×Noir1 杂交后得到的 F₇ 重组自交系, 由 240 单株组成。

(3) 内布拉斯加大学近等基因系群体 (Univ. of Nebraska population); 这是由栽培大豆近等基因系 Clark×Harosoy 杂交的 F₂ 群体, 这一群体由 57 株 F₂ 衍生系组成。

Cregan 等通过 606 对 SSR 引物对这三个分离群体进行多态性研究, 结果表明大多数 SSR 标记在两个或三个作图群体的亲本间都具有多态性。大豆经典图谱共有 68 个位点分布在 20 个小的连锁群上, 每个连锁群上只有几个位点 (Palmer 和 Shoemaker, 1998), 由于 SSR 位点的特异性, 三个遗传图谱中除了经典连锁图 6 (CLG06) 未被整合进这张图谱外, 其它三个群体分别构建的遗传图谱都被整合到一张遗传图谱上, 含有 20 个连锁群, 对应于大豆基因组中的 20 个染色体。这张图谱上的标记数目共有 1423 个, 其中 RFLP 标记 689 个, SSR 标记 606 个, AFLP 标记 10 个, 同工酶标记 10 个, 经典标记 26 个。三张图谱每一个图谱上的各种标记见表 2。至此, 这张图谱被认为是最有代表性的“公共图谱”。当然, 我们期望着标记密度更高, 分存更均匀, 形态标记更多的, 与染色体相对应的遗传图谱的问世。大豆遗传图谱的发展过程汇总于表 1。

表 2 三遗传图谱上的 DNA 标记, 同工酶标记和形态标记分布情况表

Table 2 Number of SSR, RFLP, RAPD, isozyme, and classical genetic markers mapped in three genetic linkage groups

作图群体亲本 Parents of Map population	SSR	RFLP	RAPD	AFLP	同工酶 Isozyme	常规标记 Classical	总计 Total
A81—356022×PI468916	486	501	10	0	4	3	1004
Minsoy×Noir 1	412	209	0	0	2	10	633
Clark×Harosoy	339	95	57	11	7	14	523

大豆遗传图谱研究对应的几个问题

1 豆遗传图谱与染色体的对应问题:

大豆染色体 2n=40, 在数目上比普通小麦少 2 条染色体, 但由于大豆的细胞在有丝分裂时染色体压缩, 使其很难在细胞分裂的中期进行细胞遗传学

分析。然而可以通过对粗线期的染色体进行观察研究, 大豆完整的细胞遗传图谱, 即核形分析已经完成 (Singh & Hymowitz, 1988)。初级三体 (2n=41) 可用于快速把目的基因定位到特定的染色体上, 并且将连锁群与染色体整合。大豆的全套初级三体已经完成 (Xu, 等 1998)。相信在不远的将来, 大豆的 20 个连锁图与染色体的——对应关系将会完成。

2 大豆遗传图谱与新基因的定位:

大豆遗传图谱的绘制,不但可以为分子标记辅助选择育种和基因的图位克隆及基因组结构与功能的研究打下基础,而且也为新基因的定位提供了快速路径。首先利用遗传图谱上的各种标记(SSR 引物序列可在网上查到)来研究携带所要定位基因的两个亲本,从大量的标记中筛选出特异的标记,然后利用这些标记对已有的分离群体(F_2 群体, RIL, DH 等)进行分析,计算出目的基因与这些标记间的距离,根据连锁值便可以直接把新基因定位于某一个连锁图上。然后回过头来再利用该标记周围的其它标记对分离群体进行遗传分析,在目的基因的上游和下游都能找到与基因连锁或共分离的标记,甚至在基因的内部找到标记基因的基因本身的标记,最后根据几个标记的综合数据,对目的基因作出精确定位。

3 遗传图谱作图群体的通用问题:

从遗传图谱的研究过程来看,作图群体的选择对于绘制遗传图谱起到非常重要的作用。美国目前最通用的有三套分离群体,分别为依阿华州立大学 F_2 分离群体,犹它大学重组自交系群体和内布拉斯加大学的近等基因系群体。三个群体在遗传图谱的研究中都发挥了重要作用,但也应看到,由于分离群体的不同,导致了由不同群体构建的遗传图谱间的较难比较问题。为了克服这个问题,科学家应用各种各样的方法,但直到现在,Cregan 的统一图谱也是建立在三张图谱的基础上的,另外,我们也应清楚地看到,目前虽然已有 250 个色素、形态、同工酶和抗病基因等常规标记,但由于绘制遗传图谱时所选群体在这些标记间没有多态,因此仅有 68 个标记定位于 20 个常规连锁图上。这对分子标记辅助选择育种和分子标记为基础的图位克隆产生了不利影响。我国在大豆遗传图谱的研究,现仅有中国科学院遗传研究所的陈受宜教授,利用长农 4 号与新民 6 号杂交组合的 F_2 代分离群体和 F_8 代重组自交系群体进行了遗传图谱的研究,绘制了我国的第一和第二张图谱。到目前还没有见到其它类似的报导。从美国走过的路中,我们是否可以吸取教训,即不要建立太多的分离群体,是否可以仅用这个群体,对原有的图谱进行补加,使这个图谱的密度逐渐饱和。同时再从美国引进三个群体中的一个群体,与我国的群体进行比较研究,使我国的遗传图谱与国际上的

保持一致。

参 考 文 献

- 1 Akkaya M. S., R. C. Shoemaker, J. E. Specht et al. Integration of simple sequence repeat DNA markers into a soybean linkage map[J]. Crop Science. 1995, 35: 1439—1445
- 2 Apuya, N. R. Restriction fragment length polymorphism as genetic marker in soybean[J]. Theor Appl Genet. 1988, 75: 889—901
- 3 Arumanagathan, K. Nuclear DNA content of some important plant species[J]. Plant Mol Biol. Rep. 1991, 9: 229—241
- 4 Cregan, P. B. J. Jarvik. An integrated genetic linkage map of the soybean genome[J]. Crop Science. 1999, 39: 1464—1490
- 5 Diers B. W., RFLP analysis of soybean seed protein and oil content [J]. Theor Appl Genet. 1992, 83: 608—612
- 6 Keim, P. RFLP mapping in soybean; association between marker loci and variation in quantitative traits[J]. Genetics. 1990, 126: 735—742
- 7 Keim, P. Janes M. S. A high-density soybean genetic map based on AFLP markers[J]. Crop Science. 1997, 37: 537—543
- 8 Landau—Ellis D. The genetic locus controlling supemodulation in soybean cosegregates tightly with a cloned molecular marker[J]. Mol Gen Genet. 1991, 228: 221—226
- 9 Lark, K. G. A genetic map of soybean using intraspecific cross of two cultivars, "Minsoy" and "Noir 1"[J]. Theor. Appl. Genet. 1993, 86: 901—906
- 10 Mansur L. M. J. H. Orf, K. Chase. Genetic mapping of agronomic traits using recombinant inbred lines of soybean[J]. Crop Sciences. 1996, 36: 1327—1336
- 11 Morgante M. Genetic mapping and variability of seven soybean simple sequence repeat loci[J]. Genome. 1994, 37: 763—769
- 12 Palmer, R. G. Qualitative genetics and cytogenetics. Soybean; improvement, production, and use[J]. No. 16. American, Inc., Madison, 1987, p135—209
- 13 Palmer R. G [C]. World Soybean Research Conference III. 1985, p337—344
- 14 Shoemaker, R. C., J. E. Specht. Integration of the soybean molecular and classical genetic linkage groups[J]. Crop Science. 1995, 35: 436—446
- 15 Shoemaker R. C. Soybean genomic from 1985 to 2002 [J]. Agbiotech Net. 1999, 1: 14
- 16 Shoemaker R. C., K. Polzin, J. Labate et al. Genome duplication in soybean[J]. Genetics. 1996, 144: 329—338
- 17 张德水, 董伟, 惠东威, 等. 用栽培大豆与半野生大豆间的杂种 F_2 群体构建基因组分子标记连锁框架图[J]. 科学通报, 1997, 42 (12): 13261330
- 18 张德水, 刘峰, 陈受宜. 大豆基因组研究及其进展[J]. 遗传, 1999, 20(1): 2629