

大豆天冬氨酸代谢途径关键酶基因电子定位与结构分析

于妍^{1,2}, 姜威^{1,2}, 唐敬仙³, 刘春燕^{1,2}, 陈庆山^{1,2}, 胡国华^{1,4}

(1. 东北农业大学 农学院, 黑龙江 哈尔滨 150030; 2. 黑龙江省农垦科研育种中心, 黑龙江 哈尔滨 150090; 3. 吉林农业大学 发展学院, 吉林 长春 130600; 4. 国家大豆工程技术研究中心, 黑龙江 哈尔滨 150050)

摘要:运用电子定位的方法,利用大豆基因组物理图和遗传图的整合图谱,完成了17个大豆天冬氨酸代谢途径中关键酶基因的定位以及结构分析。结果表明:这些酶基因分别定位在A1、B2、D1a、D1b、D2、F、G、H、J、L和N等11个连锁群上,并获得了相应连锁群区间两侧的标记;在基因结构分析上,AHAS、DHDPS、HK与TS没有内含子,DHAD基因的内含子数目最多,为13个,AK基因有12个内含子。通过定位获得的对应分子标记可以为基因分子辅助育种奠定基础,而基因结构信息能更好的实现基因功能分析。

关键词:大豆;天冬氨酸;关键酶基因;电子定位;基因结构

中图分类号:S565.1

文献标识码:A

文章编号:1000-9841(2010)01-0022-06

Electronic Mapping and Structure Analysis of Key Enzyme Genes on Aspartic Acid Metabolic Pathway in Soybean

YU Yan^{1,2}, JIANG Wei^{1,2}, TANG Jing-xian³, LIU Chun-yan^{1,2}, CHEN Qing-shan^{1,2}, HU Guo-hua^{1,4}

(1. Agriculture College, Northeast Agricultural University, Harbin 150030, Heilongjiang; 2. The Crop Research and Breeding Center of Land-Reclamation, Harbin 150090, Heilongjiang; 3. Development Academy, Jilin Agricultural University, Changchun 130600, Jilin; 4. National Research Center of Soybean Engineering and Technology, Harbin 150050, Heilongjiang, China)

Abstract: The nutritive value of soybean mainly focused on the aspartate-derived amino acids, and the content of these amino acids were affected by the key enzymes gene from aspartic acid metabolic pathway. In this research, the integrated map between soybean public genetic maps and physical maps was used, and the key enzyme gene sequences were blasted with soybean genome database by local Blast software. Seventeen key enzyme genes in aspartic acid metabolic pathway of soybean were mapped on soybean genetic map, and gene structures were analyzed firstly. The result indicated that these enzyme genes were mapped on 11 linkages group, including A2, B2, C2, D1b, D2, G, I, L, and M, respectively, and the flanking markers of these genes on the linkage group were obtained. The number of introns was from 0 to 13. There was no intron in AHAS, DHDPS, HK, and TS, otherwise there were 13 introns in DHAD, and 12 introns in AK. The corresponding markers obtained from mapping were the base of molecular assisted selection, while the structure information can make gene function analysis easily.

Key words: Soybean; Aspartic acid; Key enzyme genes; Electronic mapping; Gene structure

随着现代分子生物学理论与实验分析技术的发展,以及各种生物基因组计划的相继启动,尤其是人类基因组计划的实施和顺利发展^[1-2],为其它生物基因组的研究提供了参考和先例。美国作为一个大豆生产大国,近年来,在各界组织合作下,启动了大豆基因组研究系列计划,已经取得了很大进展。大豆基因组计划包括结构基因组(structural genomics)计划和功能基因(functional genomics)计划,随着功

能基因组研究工作的开展,包括大豆基因组的测序与生物信息学分析软件技术的不断完善,因此,可以从基因组水平对植物氨基酸代谢涉及的基因、蛋白及其调控机制进行全面而深入的研究。

大豆是重要的粮食作物,其蛋白质含量在豆科作物中最高,是谷类作物的4~5倍。大豆蛋白质的氨基酸组成与牛奶蛋白质相近,是一种比较完全的蛋白质。随着人们需求的改变,提高大豆营养价值

收稿日期:2009-04-07

基金项目:引进国际先进农业科学技术计划资助项目(2006-G1(A));国家高技术研究发展计划资助项目(2006AA10Z1F4);黑龙江省博士后科研启动基金资助项目(LHK-04014)。

第一作者简介:于妍(1984-),女,硕士,研究方向为大豆基因克隆与定位。E-mail:yuyan_i@163.com。

通讯作者:陈庆山,教授,博士。E-mail:qshchen@sohu.com;胡国华,研究员,博士。E-mail:Hugh757@vip.163.com。

成为大豆研究者的重要课题。大豆的营养价值主要体现在以天冬氨酸为前体的必需氨基酸含量上。天冬氨酸代谢途径是植物氨基酸合成的重要途径^[3],这个途径生成的氨基酸都是重要必须氨基酸。该途径可以分为二个分支,一条是赖氨酸的生物合成,而另一条分支又分为二个亚类,一个是苏氨酸和异亮氨酸的生物合成,另一个是蛋氨酸的生物合成。生物化学研究资料表明天冬氨酸途径中关键酶基因控制合成的氨基酸含量。因此,要改良大豆的品质,研究这些关键酶的基因是十分必要的。

该研究运用电子定位的方法,利用大豆基因组物理图和遗传图的整合图谱^[4],首次完成了 17 个大豆天冬氨酸代谢途径中关键酶基因的定位,并进行了基因的结构分析。通过定位获得对应分子标记可以为基因分子辅助选择奠定基础,而基因结构信息能更好的实现基因功能分析。

1 材料与方法

1.1 供试材料

从 NCBI 网站下载 17 个大豆天冬氨酸代谢途径的关键酶基因(表 1)。从美国能源部网站下载初

表 1 大豆天冬氨酸代谢途径关键酶基因信息

Table 1 Information of the key gene of key enzyme genes on aspartic acid metabolic pathway of soybean

基因	登陆号	基因长度	参考文献
Gene	Accession No.	Length/bp	References
AHRI	FJ594399	1764	于妍 ^[5] ,2009
AHAS	FJ581423	1962	于妍 ^[5] ,2009
BCAT	FJ581426	1050	于妍 ^[5] ,2009
ASADH	FJ581425	1131	于妍 ^[5] ,2009
CBL	FJ581427	1314	于妍 ^[5] ,2009
DAPD	FJ594400	1467	于妍 ^[5] ,2009
DAPE	FJ594401	1080	于妍 ^[5] ,2009
DHDPR	FJ594402	1035	于妍 ^[5] ,2009
DHAD	FJ594403	1806	于妍 ^[5] ,2009
HK	FJ594404	1083	于妍 ^[5] ,2009
TD	FJ594405	1731	于妍 ^[5] ,2009
CGS	AF141602	1611	Hughes et al ^[6] ,1999
MS	AF518566	2695	Wang et al,Unpublished
DHDPS	I36436	945	Silk et al ^[7] ,1994
TS	DQ275353	3166	Yanamadala et al,Unpublished
AK	AF135862	1695	Esau et al,Unpublished
HSDH	DQ172918	1472	Yanamadala et al,Unpublished

步的大豆基因组序列(网址:ftp://ftp.jgi-psf.org/pub/JGI_data/Glycine_max)该部分数据为 7 倍测序结果。并从 NCBI 的 FTP,下载用于进行本地比对的 blast2.2.16 软件包(网址:ftp://ftp.ncbi.nih.gov/blast/executables/release/2.2.16/)。

1.2 试验方法

1.2.1 格式化数据库 在开始运行中打 cmd 进入 DOS 命令行,用 cd 命令进入 blast 安装目录。formatdb 命令格式:formatdb-i c:\ * * * * \ * * * *.txt -p F-o T 其中-i 是输入文件的地址,-p 指示要格式化的数据是蛋白还是核酸,如果是蛋白,则输入-p T,如果是核酸,则输入-p F,-o 指是否要输出,若是则输入 T,参数输入使用大写。

1.2.2 用基因序列进行本地比对(local blast) 在开始运行中打 cmd 进入 DOS 命令行,用 cd 命令进入 blast 安装目录。

blastall-p blastn-d c:\ * * * * \ * * * *. * * *-i c:\ * * * * \ * * * *.txt-e 2e-10-m8-o c:\ * * * * \ * * * *.txt

-p 指定用于比对的程序,核酸和核酸数据库之间比对用 blastn,蛋白与蛋白数据库之间比对用 blastp,蛋白与核酸数据库之间比对用 tblastn,核酸与蛋白数据库之间比对用 blastx。

-e 指定期望值,只有低于这个值的比对才会被显示。

-d 指定数据库所在目录。

-i 指定要比较的序列所在的目录。

-m 指定输出的格式为表格,总共有 11 种输出格式,默认的是经典的 blast 格式,即比对序列上下排列,匹配的核苷酸或氨基酸残基之间用短线连接。

-o 指定输出结果到哪一个目录。

1.2.3 基因的电子定位 将下载的大豆基因组序列进行本地建库,用大豆公共图谱上的标记序列信息与本地大豆基因组序列库进行比对,将得到的结果按照匹配度的大小择优,经过筛选,共有 160 个拼接的 Super(占序列总数据的 72.77%)定位到连锁群上,用 MapChart2.1 画出连锁群,并按照比对的结果将 Super 画在与之相应的标记位置,得到了一套物理图谱与遗传图谱的整合图谱。

由本地比对的结果可以得到基因定位的具体信息,并结合整合的图谱确定该基因定位的连锁群,最后根据比对结果得出定位在 Super 上的位置,确定距离基因较近的标记。

2 结果与分析

2.1 基因的物理图定位

通过基因与基因组的比对,完成基因的物理图定位。表 2 中列出了基因和基因组序列的匹配区段、E 值和匹配率 3 个参数。具有内含子的基因与

基因组比对会出现多片段联配,表中列出匹配区段的 E 值范围,匹配率取平均值。*AHAS*、*HK*、*DHDPS*、*TS* 基因没有内含子,因此 E 值没有范围是具体的数值。从表 2 可以看出,基因与基因组匹配的 E 值均接近于 0,平均匹配率均超过 94%,表明基因与基因组的匹配是可信的。

表 2 大豆天冬氨酸代谢途径关键酶基因的物理图定位

Table 2 Physical mapping of the key gene of key enzyme genes on aspartic acid metabolic pathway of soybean

基因 Gene	基因长度 Length/bp	基因匹配区段 Query	基因组匹配区段 Sbjct	E 值范围 E-value	平均匹配率 Identity/%	Super
<i>AHRI</i>	1764	1-1764	1915371-1918933	$0 \sim 3 \times 10^{-18}$	94.80	105
<i>AHAS</i>	1962	1-1962	4074338-4076299	0	100.00	29
<i>BCAT</i>	1050	1-1050	7404161-7408957	$1 \times 10^{-111} \sim 2 \times 10^{-50}$	98.75	12
<i>ASADH</i>	1131	1-1131	805911-808674	$0 \sim 7 \times 10^{-19}$	99.50	163
<i>CBL</i>	1314	1-1314	244235-250052	$1 \times 10^{-82} \sim 1 \times 10^{-10}$	98.25	127
<i>DAPD</i>	1467	1-1467	314760-319421	$0 \sim 8 \times 10^{-40}$	99.38	40
<i>DAPE</i>	1080	1-1080	2825711-2829438	$1 \times 10^{-148} \sim 8 \times 10^{-26}$	99.78	62
<i>DHDPR</i>	1035	1-1035	1521652-1524809	$1 \times 10^{-132} \sim 3 \times 10^{-31}$	99.75	154
<i>DHAD</i>	1806	1-1806	1123289-1130003	$0 \sim 3 \times 10^{-17}$	99.53	29
<i>HK</i>	1083	1-1083	125160-126242	0	100.00	89
<i>TD</i>	1731	1-1731	509191-513487	$0 \sim 1 \times 10^{-34}$	94.83	199
<i>CGS</i>	1611	1-1611	6601968-6606649	$0 \sim 3 \times 10^{-28}$	98.64	20
<i>MS</i>	2695	1-2695	667054-672182	$0 \sim 5 \times 10^{-29}$	99.91	154
<i>DHDPS</i>	945	1-945	2736499-2737443	0	100.00	20
<i>TS</i>	3166	1-3166	3961023-3964188	0	100.00	24
<i>AK</i>	1695	1-1695	1108997-1114958	$0 \sim 2 \times 10^{-18}$	98.87	110
<i>HSDH</i>	1472	1-1472	2014345-2023593	$0 \sim 1 \times 10^{-15}$	99.17	61

2.2 基因的遗传图定位

分别利用上述整合图谱,对大豆天冬氨酸代谢途径中的 17 个关键酶基因完成了定位,结果见表

3。它们被定位在 A1、B2、D1a、D1b、D2、F、G、H、J、L、N 共 11 个连锁群上。*TD* 与 *HSDH* 基因由于所在 Super 没有对应连锁群上,这 2 个基因没有完成

表 3 大豆天冬氨酸代谢途径关键酶基因遗传图基因定位

Table 3 Genetic mapping result of the key gene of key enzyme genes on aspartic acid metabolic pathway of soybean

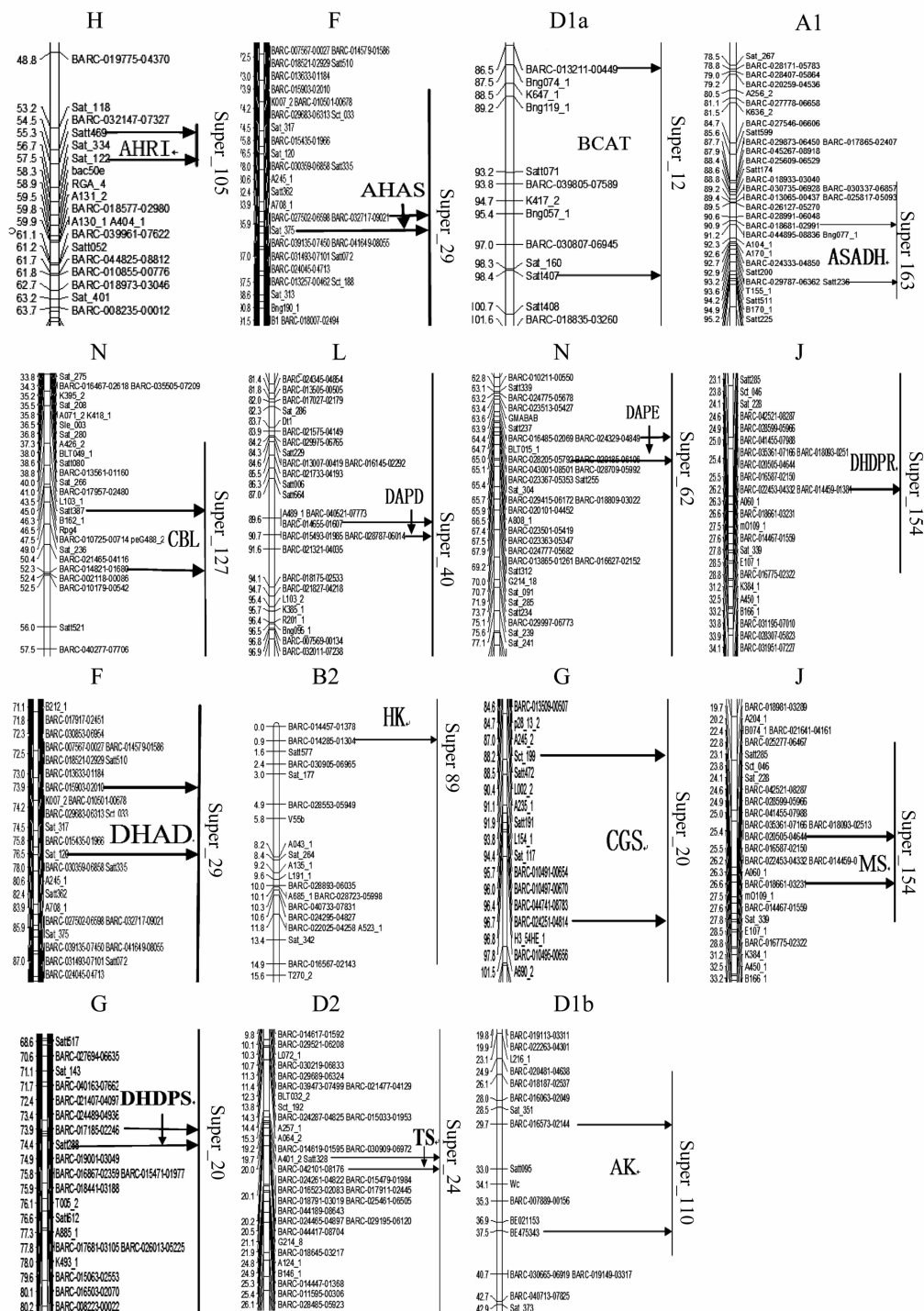
基因 Gene	基因组编号 No. of super	位置 Location	附近标记 Markers	定位区间 Mapping interval/cM	连锁群 LG
<i>AHRI</i>	105	1915371-1918933	Satt469- Sat_122	55.3 ~ 57.5	H
<i>AHAS</i>	29	4074338-4076299	BARC-027502-06598- Sat_375	88.6 ~ 90.8	F
<i>BCAT</i>	12	7404161-7408957	BARC-013211-00449- Satt407	86.5 ~ 98.4	D1a
<i>ASADH</i>	163	805911-808674	BARC-018681-02991- BARC-029787-06362	90.9 ~ 93.2	A1
<i>CBL</i>	127	244235-250052	Satt387- BARC-014821-01680	45.1 ~ 52.3	N
<i>DAPD</i>	40	314760-319421	BARC-014655-01607- BARC-028787-06014	89.6 ~ 90.7	L
<i>DAPE</i>	62	2825711-2829438	BARC-024329-04849- BARC-028205-05792	64.4 ~ 65.0	N
<i>DHDPR</i>	154	1521652-1524809	BARC-022453-04332- BARC-014459-01381	26.3	J
<i>DHAD</i>	29	1123289-1130003	BARC-015903-02010- Sat_120	74.5 ~ 82.4	F
<i>HK</i>	89	125160-126242	? -BARC-014285-01304	0 ~ 0.9	B2
<i>TD</i>	199	509191-513487	-	-	-
<i>CGS</i>	20	6601968-6606649	Satt191- BARC-024251-04814	88.2 ~ 96.7	G
<i>MS</i>	154	667054-672182	BARC-020505-04644- BARC-018661-03231	25.5 ~ 27.5	J
<i>DHDPS</i>	20	2736499-2737443	BARC-017185-02246- Satt288	75.9 ~ 76.1	G
<i>TS</i>	24	3961023-3964188	Satt328- BARC-042101-08176	19.7 ~ 20.0	D2
<i>AK</i>	110	1108997-1114958	BARC-016573-02144- BE475343	29.7 ~ 37.5	D1b
<i>HSDH</i>	61	2014345-2023593	—	—	—

“—”表示 super 没有对应合适的连锁群;“?”表示在连锁群没有找到合适的标记

“—”represents there were no suitable linkage groups matched with supers; “?” represents there were no suitable markers in the linkage group.

很好的定位。其余基因均被标注在相应连锁群的标记之间(图1)。其中, *HK* 基因所定位 B2 连锁群上的位置标记不够完全, 没有具体定位在标记之间。 *DHDPR* 基因定位的距离最近, 所定位的标记位于连

锁群的另一位置; 除此之外 *DHDPS* 因距离最近为 0.2 cM, *TS* 基因距离为 0.3 cM, *DAPE* 基因距离为 0.6 cM, *BCAT* 基因距离最远为 11.9 cM。



箭头之间部分为基因定位区间

The part between two arrows in the map represents the mapping interval

图1 大豆天冬氨酸代谢途径关键酶基因定位图

Fig. 1 Location sketch - map of mapped gene of key enzyme genes on aspartic acid metabolic pathway of soybean

2.3 基因结构分析

根据比对结果确定了这 17 个大豆天冬氨酸代谢途径关键酶基因的 cDNA 和 gDNA 序列长度、外显子与内含子的个数(表 4)。这些基因 cDNA 的长度在 945 bp ~2 695 bp 之间,gDNA 的长度在 945 bp ~9 249 bp 之间,外显子数目在 1 ~14 个之间,内含

子的数目在 0 ~13 个之间。其中 *AHAS*、*HK*、*DHDPS* 与 *TS* 是单外显子基因,它们的 cDNA 长度都小于 2 000 bp。*DHAD* 基因的内含子数目最多为 13 个,*AK* 基因有 12 个内含子,*HSDH*、*MS* 和 *CBL* 基因有 11 个内含子,其余基因的内含子数目都在 10 以内。

表 4 大豆天冬氨酸代谢途径关键酶基因结构比对信息

Table 4 Blast information of the gene structure of key enzyme genes on aspartic acid metabolic pathway of soybean

基因 Gene	登陆号 Accession No.	cDNA 基因长度 cDNA length/bp	gDNA 基因长度 gDNA gene length/bp	外显子数 No. of exon	内含子数 No. of intron
<i>AHRI</i>	FJ594399	1764	3563	8	9
<i>AHAS</i>	FJ581423	1962	1962	1	0
<i>BCAT</i>	FJ581426	1131	4797	8	7
<i>ASADH</i>	FJ581425	1050	2764	7	6
<i>CBL</i>	FJ581427	1314	5818	12	11
<i>DAPD</i>	FJ594400	1467	4662	8	7
<i>DAPE</i>	FJ594401	1080	3728	9	8
<i>DHDPR</i>	FJ594402	1035	3158	8	7
<i>DHAD</i>	FJ594403	1806	6715	14	13
<i>HK</i>	FJ594404	1083	1083	1	0
<i>TD</i>	FJ594405	1731	4297	9	8
<i>CGS</i>	AAD34548	1608	4682	11	10
<i>MS</i>	AF518566	2695	5129	12	11
<i>DHDPS</i>	L36436	945	945	1	0
<i>TS</i>	DQ275353	3166	3166	1	0
<i>AK</i>	AF135862	1695	5962	13	12
<i>HSDH</i>	DQ172918	1472	9249	12	11

3 讨论

3.1 基因电子定位与分子标记定位

基因定位是对基因在染色体上所在位置的测定,是遗传学研究的重要环节。它对研究基因的结构、功能和相互作用有着重要意义,并可应用于基因工程中的重组 DNA 操作。

目前,基于大豆遗传图谱定位与标记相连锁的性状有很多,但基因方面的不多。有关低亚油酸含量^[8]、氨基酸含量^[9]、大豆花叶病毒抗性^[10]、种子蛋白和产量^[11]以及胞囊线虫抗性^[12]等性状的基因定位研究均有报道。但是即使这些标记与控制性状的基因位点遗传距离很近,在物理距离上也会很远,因

此这些标记都是间接的分子标记。从基因自身序列出发而获得染色体或连锁群的定位可以将遗传距离和物理距离均降至零,实现基因的准确精细定位。目前这方面在大豆上的研究较少,究其根源就是因为没有序列与遗传图的对应关系。

传统的分子标记完成定位在拥有相应定位群体的前提下,还需要经过大量的试验操作,耗费一定的时间和精力。而基因电子定位是借助一定的生物信息数据库和生物学软件,结合互联网上的基因信息实现基因的快速、准确定位。因此运用电子方法完成基因定位与分子标记定位相比具有投入少、见效快、高通量等特点,可以收到事半功倍的效果。随着生物信息学的快速发展、基因组研究的深入以及网

络资源不断丰富和完善,基因电子定位一定会具有良好的发展前景。

3.2 遗传图谱与物理图谱的整合

经过近 10 a 的大豆生物技术研究 and 遗传作图,目前已有几十套大豆分子遗传图谱构建起来。2004 年,利用已经构建好 5 张大豆遗传图谱^[13],整合了 1 张新的公共图谱 soymap2。2007 年,在此基础上又开发了 5 551 个 SNP 标记^[14],并将 1 141 个 SNP 标记整合到原有图谱中,进一步丰富了图谱。2008 年,随着大豆基因组测序的完成,Qi 等利用美国能源部释放的大豆基因组序列,运用本地 Blast 程序首次完成了物理图谱和遗传图谱的初步整合,这为快速完成标记与基因的穿梭提供了前提条件。一方面,通过电子定位能够更加方便的找到对应标记相邻的序列信息或对应序列相邻的标记信息。另一方面,通过标记区间序列信息挖掘功能基因,完成 QTL 到 QTG 的转换,使得遗传作图与基因挖掘紧密联系起来。

3.3 基因结构分析

通过电子定位能够同时获得 cDNA 和 gDNA 的基因信息,由此可以根据比对结果获得基因的结构信息,这样便为基因的深入研究奠定了基础。目前大部分对于基因序列的研究,往往都停留在 cDNA 水平上,忽视了基因结构的分析。但随着大豆基因组测序的成功,该研究利用电子克隆的方法结合大豆基因组数据库,对这方面的内容进行了补充。

4 结论

应用物理图和遗传图的整合图谱,首次完成了天冬氨酸合成途径中 17 个关键酶基因的电子定位与结构分析。这些基因分别定位在 A1、B2、D1a、D1b、D2、F、G、H、J、L 和 N 共 11 个连锁群上。并通过序列比对获得了相应连锁群区间两侧的标记。通过基因结构分析结果可以看出,这些基因内含子数目为 0~13 个,其中 AHAS、DHDPS、HK 与 TS 没有内含子;DHAD 基因的内含子数目最多为 13 个,AK 和 HSDH 基因有 12 个内含子。这种定位比以往的定位更加准确、直观、形象;并用遗传标记标出了这些基因所在染色体的大概位置,这对研究基因的结构、功能和相互作用有着重要意义。

参考文献

- [1] Yager T, Nickerson D A, Hood L E. The human genome Project: creating an infrastructure for biology and medicine[J]. Trends in Biochemical Sciences, 1991, 16(12): 454-458.
- [2] Peakall D, Shugart L. The human genome project (HGP)[J]. Ecotoxicology Springer, 2002, 11(1): 7-9.
- [3] 于妍, 宋万坤, 刘春燕, 等. 植物天冬氨酸代谢关键酶基因研究进展[J]. 生物技术通报, 2008, (增刊): 7-11. (Yu Y, Song W K, Liu C Y, et al. Research development of key enzymes gene on aspartic acid metabolic pathway in plants[J]. Biotechnology Bulletin, 2008, (Supplement): 7-11.)
- [4] Qi Z M, Li H, Wu Q, et al. An integrated map of soybean physical map and genetic map[J]. Journal of Northeast Agriculture University, 2009, 16(2): 12-16.
- [5] 于妍. 大豆天冬氨酸代谢相关酶基因克隆与定位分析[D]. 哈尔滨: 东北农业大学, 2009; 26. (Yu Y. Cloning and mapping analysis of related enzymes genes on aspartic acid metabolism soybean[D]. Harbin: Northeast Agricultural University, 2009; 26.)
- [6] Hughes C A, Gebhardt J S, Reuss A, et al. Identification and expression of a cDNA encoding cystathionine- gamma- synthase in soybean[J]. Plant Science, 1999, 146: 69-79.
- [7] Silk G W, Matthews B F, Somers D A, et al. Cloning and expression of the soybean DapA gene encoding dihydrodipicolinate synthase[J]. Plant Molecular Biology, 1994, 26(3): 989-993.
- [8] Brummer E C, Nickell A D, Wilcox J R, et al. Mapping the Fan locus controlling linolenic acid content in soybean oil[J]. Journal of Heredity, 1995, 86: 245-247.
- [9] Panthee D R, Pantalone V R, Saxton A M, et al. Genomic regions associated with amino acid composition in soybean[J]. Molecular Breeding, 2006, 17: 79-89.
- [10] Yu Y G, Saghai Maroof M A, Buss G R, et al. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance[J]. Phytopathology, 1994, 84: 60-64.
- [11] Mansur L, Lark K G, Kross H, et al. Interval mapping of quantitative trait loci for reproductive, morphological and seed traits of soybean (*Glycine max* L)[J]. Theoretical and Applied Genetics, 1993, 86: 907-913.
- [12] Conclbido V C, Denny R L, Boutin S R, et al. DNA Marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycine Ichinohe*)[J]. Crop Science, 1994, 34: 240-246.
- [13] Song Q J, Marek L F, Shoemaker R C, et al. A new integrated genetic linkage map of the soybean[J]. Theoretical and Applied Genetics, 2004, 109: 122-128.
- [14] Choi Y, Hyten D L, Matukumalli L K, et al. A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis[J]. Genetics Society of America, 2007, 176: 685-696.