

玉米淀粉合成相关基因 *ZmSSIIa* 的序列变异分析

杨泽峰, 张恩盈, 杜灿灿, 马思佳, 胡芸芸, 谢政文, 潘亮, 陈庆, 徐辰武

扬州大学农学院; 教育部植物功能基因组学重点实验室; 江苏省作物遗传生理重点实验室, 扬州 225009

摘要 玉米淀粉合成酶 *ZmSSIIa* 在淀粉生物合成途径中主要作用是将支链淀粉的分支链由 6~10 个延长至 12~25 个, 对于玉米籽粒产量和品质具有重要的影响。本研究在 86 个优异玉米自交系中对该基因进行基因序列定点捕获, 并获得该基因全长为 4836 bp 的核苷酸序列。基因多态性分析发现, 该基因座位上有 78 个 SNP 和 19 个 Indel。玉米 *ZmSSIIa* 基因的编码区具有 24 个 SNP, 将该基因划分成 9 种编码区单倍型, 并编码 7 种 *ZmSSIIa* 蛋白质。供试材料的群体中, *ZmSSIIa* 基因至少经历 8 次重组, 可能对单倍型的分化和连锁不平衡具有重要作用。中性进化测试表明该基因在供试群体中没有明显的人工选择。

关键词 玉米; *ZmSSIIa* 基因; 序列多态性; 单倍型; 连锁不平衡

中图分类号 Q943.2

文献标志码 A

doi 10.3981/j.issn.1000-7857.2014.36.008

Analysis of Sequence Polymorphism for Maize Starch Biosynthesis Related Gene *ZmSSIIa*

YANG Zefeng, ZHANG Enying, DU Cancan, MA Sijia, HU Yunyun, XIE Zhengwen, PAN Liang, CHEN Qing, XU Chenwu

Key Laboratory of the Ministry of Education for Plant Functional Genomics; Jiangsu Key Laboratory of Crop Genetics and Physiology; Agricultural College, Yangzhou University, Yangzhou 225009, China

Abstract The maize starch synthesis *ZmSSIIa* elongates amylopectin chains from degree of polymerization 6–10 to degree of polymerization 12–25 in the pathway of starch biosynthesis, and plays important roles in kernel yield and quality. The analysis of sequence polymorphism of maize *ZmSSIIa* gene will provide information for further detection of beneficial mutants associated with yield and quality. In order to reveal the sequence polymorphism of the maize *ZmSSIIa* gene, a total of 4836 bp sequence of this locus is captured in 86 elite maize inbred lines through a NimbleGensequence capture array. Analysis of the sequence polymorphism reveals that there are 78 SNPs and 19 Indels in this locus. The coding sequences of the maize *ZmSSIIa* gene contain 24 SNPs, which divide the coding regions into 9 haplotypes, and 7 different *ZmSSIIa* proteins are encoded by the tested inbred lines. At least 8 recombination events are found to contribute to the haplotype diversity and linkage disequilibrium between polymorphic sites in the maize *ZmSSIIa* gene. According to the results of the test of neutral evolution, there is no evidence of artificial selection for this gene in the tested inbred lines.

Keywords maize; *ZmSSIIa* gene; sequence polymorphism; haplotype; linkage disequilibrium

淀粉是植物种子的主要成分,也是最重要的碳能量储存物质。淀粉不仅是人类食物的主要组成成分,还是一种重要

的工业原料,被广泛应用于酿酒、食品、造纸、医药、建筑、黏合剂、纺织及生物降解塑料等行业。在植物的叶绿体与造粉

收稿日期:2014-10-27;修回日期:2014-11-14

基金项目:高等学校博士学科点专项科研项目(20123250110001);江苏省自然科学基金项目(BK2012261);江苏省高校自然科学研究重大项目(14KJA210005);江苏省“青蓝工程”科技创新团队项目

作者简介:杨泽峰,副教授,研究方向为植物功能基因组学与生物信息学,电子信箱:zfyang@yzu.edu.cn;徐辰武(通信作者),教授,研究方向为数量遗传学与生物信息学,电子信箱:qtl@yzu.edu.cn

引用格式:杨泽峰,张恩盈,杜灿灿,等.玉米淀粉合成相关基因 *ZmSSIIa* 的序列变异分析[J].科技导报,2014,32(36):52-58.

体中,淀粉由ADPG焦磷酸化酶(ADP-Glcpyrophosphorylase, AGPPase)、淀粉合成酶(starch synthase, SS)、淀粉分支酶(Starch branching enzyme, SBE)及去分支酶(debranching enzyme, DBE)等一系列酶的催化而成^[1,2]。

淀粉合成酶是植物淀粉合成的关键酶,与淀粉的含量、直链淀粉与支链淀粉的比例、支链淀粉的链长以及淀粉粒的结构有关。淀粉合成酶以寡聚糖为前提,通过 $\alpha-1,4$ 糖苷键不断增加寡聚糖的葡萄糖单位,最终合成 $\alpha-1,4$ 糖苷键连接的多聚糖。该多聚糖又作为淀粉分支酶的底物合成支链淀粉^[3]。淀粉合成酶分为两类,分别是可溶性淀粉合成酶(soluble starch synthase, SSS)和颗粒结合型淀粉合成酶(granule bound starch synthase, GBSS)。GBSS的主要功能是参与直链淀粉的合成;SSS主要与分支酶共同作用合成支链淀粉^[4]。根据蛋白质中氨基酸结构的不同,禾本科植物SSS被分为4个同工型家族,分别为SSI、SSII、SSIII和SSIV^[5]。在甜玉米育种中起到重要作用的sugary2(su2)突变体是由于编码一个可溶性淀粉合成酶基因*ZmSSIIa*的突变引起的。玉米*ZmSSIIa*基因主要在胚乳中表达,其在水稻中同源基因的序列变异调控支链淀粉结构并对糊化温度具有显著影响^[6]。基于遗传分析发现在功能缺失型突变体su2中,支链淀粉短分支链增加,而中等链和长支链的比例减少,直链淀粉的含量也会明显增加。进一步研究发现玉米淀粉合成酶*ZmSSIIa*对支链淀粉的分支链由6~10个延长至12~25个起到主要作用^[6,7]。

玉米是中国第一大粮食作物,而且是重要的饲料和工业原料,研究与玉米产量和品质相关基因的序列变异将对培育高产优质玉米品种的分子设计育种奠定基础。尽管*ZmSSIIa*基因的突变体在甜玉米育种中已经被广泛应用,但是该基因在中国玉米优异自交系中的核苷酸多态性尚不得而知。鉴定玉米*ZmSSIIa*基因的序列变异,将进一步通过关联分析筛选出与产量和品质相关的优异等位变异奠定基础。本研究在86份优异自交系中对玉米*ZmSSIIa*基因的序列进行定点捕获,并分析该基因的核苷酸多态性、单倍型密度、连锁不平衡和遗传重组。

1 材料和方法

1.1 试验材料

本研究选取了具有广泛代表性的86份玉米优异自交系,包括中国温带玉米的5个杂种优势群的代表性种质、热带/亚热带种质、糯玉米自交系和来源于国外的种质材料(表1)。

1.2 DNA提取和*ZmSSIIa*基因测序

在实验室水培种植供试玉米材料,至出苗15 d左右剪取嫩叶备用。采用改良的CTAB法^[8]提取DNA。用分光光度计检测提取出的DNA浓度;并用琼脂糖凝胶电泳的方法检测所提取DNA的质量。采用B73的*ZmSSIIa*基因(XM_008651039.1)序列为模板,在供试玉米材料中采用NimbleGen平台对该基因进行目标序列定点捕获并测序。基因的目标序列定点捕获和测序均由华大基因公司完成。

表1 研究所用玉米自交系列表

Table 1 Maize inbred lines used in the study

编号	自交系	编号	自交系	编号	自交系
1	QH19612	31	4866	61	KWS456
2	昌7-2	32	3189	62	QF-11
3	LX9801	33	铁9206	63	QF01
4	107	34	本育15	64	爆裂玉米
5	黄518	35	QZ01	65	JND-1
6	k12	36	478选	66	JND-2
7	nx335	37	郑58	67	BJ-2
8	s80	38	7922	68	QF02
9	H21	39	8605-2	69	QF03
10	吉853	40	178	70	QDM01
11	412	41	QP1721	71	M1
12	杂107	42	Ex韩	72	WT262
13	黄早四	43	xy35	73	WT26
14	502	44	P138	74	QZN01F
15	鲁原92	45	6819	75	QKN01M
16	10168	46	丹988	76	QLY01Z
17	丹598	47	319B	77	QJKN20F
18	综3	48	白轴齐319	78	QZN12
19	4CV	49	齐318	79	QBN48
20	E28	50	沈137	80	QBN029
21	丹340	51	11099	81	BJ-4
22	340改	52	suwan	82	QBN02
23	丹黄25	53	11118	83	QBN3186
24	JH3372	54	11200	84	M1132
25	丹99	55	10533-1	85	BJ-5
26	齐232	56	RCML15	86	Y53
27	OH43	57	QS7M		
28	Mo17	58	RBS11		
29	8112	59	FLB01		
30	K8112	60	Q52		

1.3 数据分析

对获得的*ZmSSIIa*基因的全长序列通过Clustal X软件^[9]进行比对,并进行手工校对。根据NCBI网站上对该基因的注释将其划分成不同的区段(含1个启动子区段、9个外显子和8个内含子)。基因序列的核苷酸多态性、单倍型密度和中性进化测验的分析均采用DNA_{sp}软件^[10]进行。在本研究中利用参数 π 和 θ 表示核苷酸多态性,其中前者表示同一位点不

同序列两两之间的差异;后者注重衡量群体突变率的指标,与核苷酸变异所占序列位点数的比例有关^[11]。采用Tajima's $D^{[12]}$ 和Fu与Li的 D^* 和 F^* 参数^[13]对 *ZmSSIIa* 基因进行中性测验。玉米 *ZmSSIIa* 基因多态性位点的连锁不平衡(LD)值采用TASSEL软件^[14]计算。在本研究中采用 R^2 值表征LD。LD随距离的衰减曲线参照Remington等^[15]定义的非线性回归方程,并用SAS软件中的NLIN过程拟合该线性回归方程。

2 结果与分析

2.1 玉米 *ZmSSIIa* 基因的序列变异

经多序列比对,本研究共获得玉米 *ZmSSIIa* 基因座位的4836 bp的序列,其中包括该基因自起始密码子至终止密码子之间的全长序列和1094 bp的启动子区段序列。在该基因中,共有9个外显子包括2199 bp的编码序列,此外还有8个内含子含有1543 bp的序列(表2)。

表2 玉米 *ZmSSIIa* 基因的多态性位点

Table 2 Polymorphic sites in the maize *ZmSSIIa* gene

位点	多态性位点	频率/%	多态性类型	位置	位点	多态性位点	频率/%	多态性类型	位置
36	SNP_36	97/3	C/T	promoter	2374	2374SNP_	97/3	T/C	Intron04
42	SNP_42	95/5	G/A	promoter	2400	2400SNP_	97/3	T/C	Intron04
59	SNP_59	93/7	G/T	promoter	2412	SNP_2412	97/3	C/T	Intron04
64	SNP_64	80/20	A/G	promoter	2484	Indel_09	97/3	TGTATTTG/-----	Intron04
78	SNP_78	88/12	C/T	promoter	2694	SNP_2694	81/19	A/T	Intron05
94	SNP_94	97/3	C/T	promoter	2805	Indel_10	79/21	-/T	Intron05
99	SNP_99	92/8	C/T	promoter	2826	SNP_2826	81/19	T/A	Intron05
100	SNP_100	97/3	A/T	promoter	2868	SNP_2868	81/19	G/A	Intron05
101	SNP_101	97/3	A/T	promoter	2875	Indel_11	81/19	AA/--	Intron05
102	SNP_102	97/3	G/A	promoter	2877	Indel_12	93/7	-/A	Intron05
123	SNP_123	84/16	T/C	promoter	2952	Indel_13	81/19	GCTAG/-----	Intron05
132	Indel_01	81/19	-/C	promoter	2985	SNP_2985	81/19	A/G	Intron05
167	SNP_167	88/12	C/T	promoter	3163	SNP_3163	78/22	G/A	Intron06
198	SNP_198	95/5	C/T	promoter	3241	SNP_3241	78/22	A/G	Exon07
201	SNP_201	80/20	T/C	promoter	3289	Indel_14	78/22	----/TGGT	Intron07
202	SNP_202	80/20	A/G	promoter	3356	SNP_3356	78/22	A/T	Intron07
220	SNP_220	97/3	G/T	promoter	3364	Indel_15	78/22	T/-	Intron07
224	SNP_224	97/3	G/A	promoter	3376	SNP_3376	78/22	T/C	Intron07
238	SNP_238	97/3	G/A	promoter	3398	SNP_3398	78/22	C/G	Intron07
278	SNP_278	97/3	G/A	promoter	3401	SNP_3401	97/3	C/T	Intron07
279	SNP_279	87/13	C/A	promoter	3525	SNP_3525	78/22	C/T	Exon08
329	Indel_02	87/13	T/-	promoter	3704	Indel_16	97/3	--/AA	Intron08
330	Indel_03	76/24	T/-	promoter	3707	SNP_3707	78/22	C/T	Intron08
388	Indel_04	97/3	-----/ACTCTGCT	promoter	3715	SNP_3715	97/3	C/T	Intron08
454	SNP_454	97/3	T/G	promoter	3723	Indel_17	78/22	-/A	Intron08
505	Indel_05	86/14	-----/GATTCCTT	promoter	3729	Indel_18	91/9	T/-	Intron08
599	Indel_06	97/3	-----/CCGAA	promoter	3755	SNP_3755	78/22	A/T	Intron08
628	SNP_628	94/6	C/A	promoter	3756	SNP_3756	97/3	C/T	Intron08
698	SNP_698	86/14	C/T	promoter	3787	SNP_3787	78/22	A/C	Intron08
706	SNP_706	97/3	A/G	promoter	3808	SNP_3808	78/22	T/A	Intron08
1074	Indel_07	90/10	-----/CCCCG	promoter	3819	SNP_3819	78/22	C/G	Intron08
1122	SNP_1122	99/1	C/T	Exon01	3820	Indel_19	78/22	ATGCTC/-----	Intron08
1218	SNP_1218	92/8	G/A	Exon02	3838	SNP_3838	78/22	A/C	Intron08
1227	SNP_133	95/5	T/C	Exon03	3849	SNP_3849	78/22	C/T	Intron08
1292	SNP_1227	95/5	G/A	Exon04	3853	SNP_3853	78/22	T/C	Intron08
1384	Indel_08	95/5	AT/--	Intron01	3865	SNP_3865	95/5	C/A	Intron08
1412	SNP_1412	95/5	G/C	Intron01	3959	SNP_3959	81/19	C/G	Exon09
1449	SNP_1449	99/1	T/G	Intron01	3965	SNP_3965	97/3	C/T	Exon09
1542	SNP_1542	97/3	C/T	Exon02	4061	SNP_4061	87/13	C/A	Exon09
1747	SNP_1747	97/3	C/T	Intron02	4134	SNP_4134	99/1	C/G	Exon09
1765	SNP_1765	97/3	G/C	Intron02	4385	SNP_4385	90/10	C/T	Exon09
1792	SNP_1792	78/22	T/C	Intron02	4517	SNP_4517	88/12	C/T	Exon09
1850	SNP_1850	97/3	T/A	Exon03	4520	SNP_4520	90/10	C/T	Exon09
1901	SNP_1901	97/3	A/T	Exon03	4562	SNP_4562	62/38	A/G	Exon09
1937	SNP_1937	97/3	A/C	Exon03	4632	SNP_4632	86/14	A/G	Exon09
1962	SNP_1962	97/3	G/A	Exon03	4719	SNP_4719	99/1	A/G	Exon09
1986	SNP_1986	97/3	C/T	Exon03	4721	SNP_4721	87/13	C/G	Exon09
2056	SNP_2056	97/3	A/G	Intron03	4726	SNP_4726	86/14	A/G	Exon09
2103	SNP_2103	97/3	T/C	Intron03					

在该基因中共发现了97个多态性位点,其中包括78个SNP,平均62 bp有1个SNP;共有19个Indel多态性位点,平均254.5 bp有1个Indel。值得注意的是,研究发现多数变异具有较低的最小等位基因频率(Minor Allele Frequency, MAF)。在所有的97个多态性位点中,最小等位基因频率小于5%的有36个,占37.1%;而小于10%的有52个,占53.6%。研究中发现该基因的核苷酸多样性分布并不均衡,外显子区段具有较低的SNP变异频率,平均91.6 bp有1个SNP;而启动子和内含子区则分别是每45.6 bp和51.4 bp便有1个SNP。

此外,玉米*ZmSSIIa*基因座位上19个Indel全部位于启动子区和内含子区,编码区不含有Indel变异。该基因全长序列的核苷酸多样性参数 π 的估计值为0.0030(表3),而内含子区

具有最高的核苷酸多样性, π 的估计值为0.0046,应是编码区受到的选择压力较大所致。

为检验该基因在供试群体中是否偏离中性进化,本研究首先对启动子区、编码区、内含子区和全长序列分别进行Tajima's D测验^[12],结果发现这4个区段的估计值都没有达到显著水平(表3)。此外,研究还对4个区段分别进行Fu与Li的D*测验和F*测验^[13]。除内含子区段之外,其余3个区段的Fu与Li的D*估计值在0.05的显著水平上大于0,表明在这些区段中可能是由于平衡选择或者群体的瓶颈效应而导致较多的中等频率等位基因。此外,本研究中发现Fu与Li的F*估计值都没有达到显著水平。这些结果也进一步说明在供试的玉米群体中,*ZmSSIIa*基因没有经历明显的人工选择。

表3 玉米*ZmSSIIa*基因的核苷酸多态性参数

Table 3 Summary of parameters for analysis of nucleotide polymorphisms of maize *ZmSSIIa* gene

参数	启动子	外显子	内含子	基因全长
序列长度/bp	1094	2199	1543	4836
多态性位点	31	24	42	97
变异频率	0.0302	0.0109	0.0272	0.0201
核苷酸替换数	24	24	30	78
核苷酸替换频率	0.0219	0.0109	0.0194	0.0161
Indel数目	7	0	12	19
Indel位点数	29	0	34	63
Indel平均长度/bp	4.1429	空白	2.8333	3.3158
Indel频率	0.0064	空白	0.0078	0.0039
π	0.0032	0.0019	0.0046	0.003
θ	0.0045	0.0022	0.004	0.0033
Tajima's D	-0.8415	-0.3871	0.492	-0.2118
Fu与Li's D*	1.8189*	0.6895	1.6047*	1.7011*
Fu与Li's F*	0.9787	0.3404	1.4097	1.1041

注: *表示0.05水平上的差异显著性。

2.2 玉米*ZmSSIIa*基因的单倍型及蛋白质序列分析

本研究所获得的玉米*ZmSSIIa*基因的全长序列中包含78个SNP和19个Indel,这些变异位点可以将该基因划分成20种单倍型,单倍型密度为0.8558(表4)。其中,有9种单倍型只含有1个自交系;频率最高的单倍型为Hap_1,有29个自交系携带该单倍型。虽然在编码区没有Indel变异,但24个SNP变异位点可以将整个编码区划分成9种单倍型,单倍型密度为0.6339(表4)。这9种编码区单倍型中,有3种单倍型仅仅被1个自交系所携带;频率最高的编码区单倍型为CDS_Hap_1,被50个自交系所携带。

编码区所具有的24个SNP中,有12个为同义突变位点,其余的12个为非同义突变位点。然而,研究还发现在这12

个非同义突变位点中,仅仅有4个SNP(SNP_1218/1937/4632/4721)的最小等位基因频率大于0.05。另外还发现,非同义突变位点SNP_4719和SNP_4721属于同一密码子,因此这2个非同义突变相同,从而导致该氨基酸位点存在3种多态类型。由此导致的另外一个结果是12个非同义突变位点导致编码的蛋白质中产生11个氨基酸位点的改变。将编码区的核苷酸序列翻译成氨基酸序列,发现该基因在供试群体中共编码7种不同的*ZmSSIIa*蛋白质(图1)。编码区单倍型CDS_Hap_1、CDS_Hap_8和CDS_Hap_9编码同一种类型的*ZmSSIIa*蛋白质,有59个玉米自交系编码这一种类型,是供试群体中频率最高的类型。另有2种*ZmSSIIa*蛋白质分别被1个自交系所编码。

表4 玉米 *ZmSSIIa* 基因的单倍型

Table 4 Haplotypes of maize *ZmSSIIa* gene

单倍型	自交系数目	编码区单倍型	自交系
Hap_1	29	CDS_Hap_1	LX9801, K8112, 340改, 6819, QH19612, Ex 韩, P138, JND-2, WT26, 8605-2, 178, s80, 502, 白轴齐 319, QJKN20F, 丹黄 25, 319B, 3189, 黄早四, 7922, QZN01F, BJ-5, 铁 9206, 齐 318, 沈 137, suwan, 齐 232, 吉 853, 8112
Hap_2	10	CDS_Hap_2	黄 518, 478 选, 丹 99, 昌 7-2, 郑 58, 10533-1, QBN48, 11118, 鲁原 92, 107
Hap_3	1	CDS_Hap_2	Y53
Hap_4	1	CDS_Hap_3	综 3
Hap_5	3	CDS_Hap_1	QF03, QZ01, 412
Hap_6	5	CDS_Hap_1	丹 988, M1, OH43, 丹 598, 4866
Hap_7	1	CDS_Hap_1	QBN02
Hap_8	6	CDS_Hap_4	QP1721, nx335, Q52, 4CV, BJ-2, Mo17
Hap_9	1	CDS_Hap_4	RCML15
Hap_10	2	CDS_Hap_1	JND-1, 11099
Hap_11	1	CDS_Hap_5	QF-11
Hap_12	5	CDS_Hap_1	k12, RBS11, QKN01M, BJ-4, QS7M
Hap_13	3	CDS_Hap_1	JH3372, 10168, E28
Hap_14	1	CDS_Hap_1	QF01
Hap_15	1	CDS_Hap_1	QF02
Hap_16	1	CDS_Hap_6	H21
Hap_17	3	CDS_Hap_6	QLY01Z, KWS456, 爆裂玉米
Hap_18	3	CDS_Hap_7	QBN029, QBN3186, 丹 340
Hap_19	8	CDS_Hap_8	xy35, WT262, M1132, FLB01, 11200, 本育 15, 杂 107, QDM01
Hap_20	1	CDS_Hap_9	QZN12

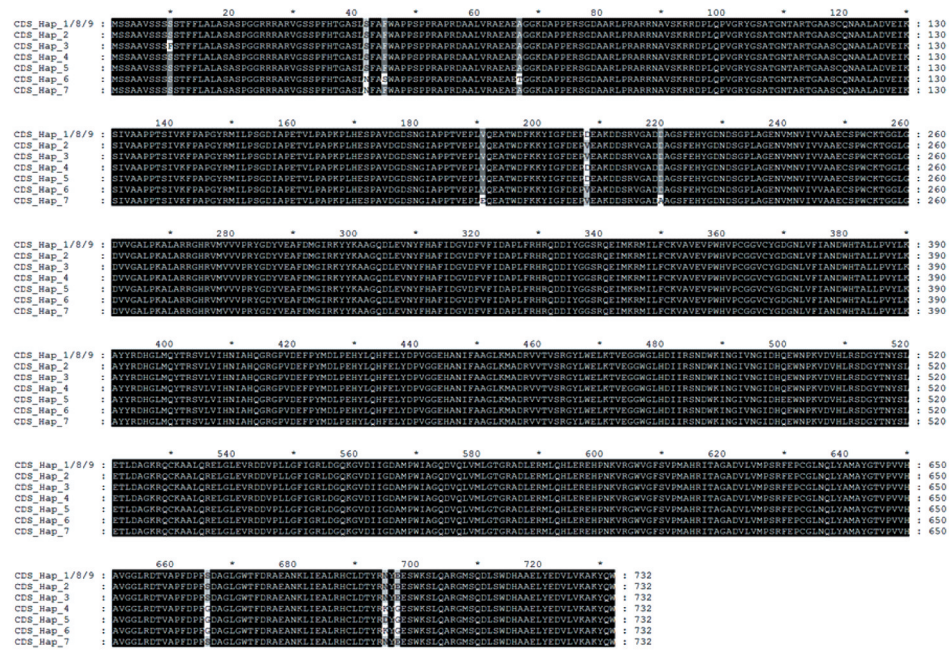


图1 供试玉米自交系编码的 *ZmSSIIa* 蛋白质序列多序列比对

Fig. 1 Alignment of *ZmSSIIa* proteins encoded by tested inbred lines

2.3 连锁不平衡和重组

对所有的SNP变异位点进行连锁不平衡分析,发现玉米 *ZmSSIIa* 基因中存在2个高度连锁不平衡区域,其一为从SNP_1542至SNP_2412的区段内,另一个为从SNP_2604至

SNP_3853的区段内(图2)。这2个高度连锁不平衡区域内的变异位点皆表现为高度连锁,而2个区段之间则连锁较弱。之外,还有一些距离较近的位点也处于高度连锁不平衡,如SNP_1227、SNP_1292和SNP_1412之间也存在完全连锁的关系。

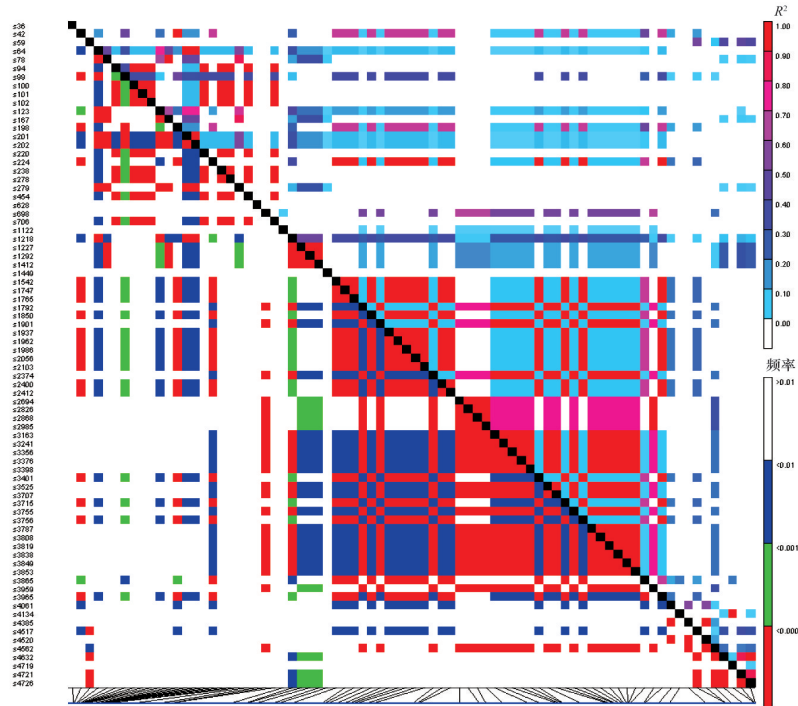


图2 玉米 *ZmSSIIa* 基因的SNP变异位点间的连锁不平衡

Fig. 2 Linkage disequilibrium between SNPs detected in maize *ZmSSIIa* gene

尽管玉米 *ZmSSIIa* 基因的很多变异位点处于高度连锁不平衡状态,但依然能够看出随着多态位点间距离的增加连锁不平衡有衰减的趋势。本研究进一步对玉米 *ZmSSIIa* 基因中所有的SNP多态位点进行连锁不平衡衰减的估计。以多态性位点之间的距离和连锁不平衡参数 R^2 作散点图并以 Remington 等^[15]定义的连锁不平衡衰减曲线来描述其衰减规律(图3)。一般认为,若 R^2 值小于0.1,则位点间不存在连锁不平衡。这主要是因为当 R^2 值小于0.1时,要检测到足够的关联就需要非常大的群体样本,使得关联分析难以操作。

本研究通过拟合的非线性方程发现玉米 *ZmSSIIa* 基因的连锁不平衡随距离的增加而快速衰减,并且在大约3140 bp处即可衰减至0.1水平(图3)。

为分析重组事件对连锁不平衡的影响,本研究利用玉米 *ZmSSIIa* 基因序列中的多态位点分析遗传重组事件。结果发现,在供试材料中至少有8次重组事件,其相应的核苷酸位点分别是(100-124)、(199-280)、(280-1219)、(1219-2695)、(3854-3866)、(3866-4062)、(4062-4386)和(4518-4563)。这些重组事件可能对于玉米 *ZmSSIIa* 基因的核苷酸多样性以及多态性位点间连锁不平衡的衰减起到重要作用。

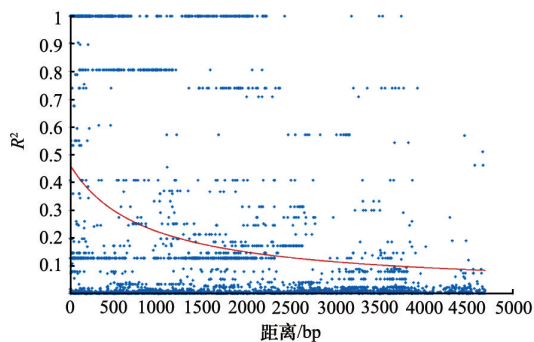


图3 玉米 *ZmSSIIa* 基因的变异位点之间连锁不平衡示意图

Fig. 3 Diagram of the LD between polymorphic sites of the maize *ZmSSIIa* gene

3 讨论

作物农艺性状表型的差异以及DNA序列水平的多态性是作物遗传改良的基础。玉米是典型的异花授粉作物,不仅在表型上不同品种间具有较大的差异,而且在DNA序列水平上也表现出较强的多样性^[16]。SNP和Indel是基因组中最常见的核苷酸多样性变异方式,并且部分核苷酸多样性能引起表型的巨大差异^[17]。基于对27个玉米自交系的序列多样性分析,Gore等^[18]发现玉米在2个自交系之间大约每44 bp便有1个多态性变异位点。玉米自交系之间的序列变异甚至超过了人类和黑猩猩这2个物种间的差异。人类和大猩猩这2个

物种自 350 万年前便已经分化,而玉米是大约在 1 万年左右的时间才由人类自野生近源种大刍草(*Teosinte*)驯化而来^[19]。Ching 等^[20]对 36 份自交系的 18 个基因的核苷酸多样性进行分析,结果发现在非编码区每 47.7 bp 有 1 个 SNP,而编码区则是 1 个 SNP/130.5 bp;非编码区中每 85 bp 有 1 个 Indel,而编码区则是 1 个 Indel/1349 bp。本研究中发现尽管玉米 *ZmSSIIa* 基因的编码区不存在 indel 变异,编码区 SNP 变异频率较 Ching 等^[20]的结果高,而非编码区的 SNP 变异频率与其相似。这一结果表明供试群体具有较好的遗传变异基础。

在作物的驯化和遗传改良(育种)过程中,一些基因由于携带对产量、品质和抗逆具有重要作用的优异等位变异而被人们有目的的选择,从而使得目的基因的核苷酸多样性较低^[21]。比如在玉米的驯化过程中,人们有目的的对 *TGA1* 基因进行筛选。大刍草中的 *TGA1* 基因使得其籽粒被较长的坚硬稃壳包裹,而玉米中的 *TGA1* 则使得玉米籽粒无壳且柔软。尽管两者 *TGA1* 基因之间仅一个核苷酸的差异,人类的选择将该基因的优异等位变异在玉米群体中保留并繁衍下来^[22]。再如玉米的 *Y1* 基因控制籽粒中类胡萝卜素的含量,由于直接的表现型为籽粒的颜色,人们在育种实践中选择类胡萝卜素含量高的种质,所以在该基因位点进行了有目的的人工筛选,导致在黄粒种质中该位点的遗传多样性显著降低^[23]。玉米的 *ZmSSIIa* 基因参与淀粉的合成,可能对玉米籽粒产量和品质具有重要作用。然而本研究中中对 *ZmSSIIa* 基因的中性进化检验中 Tajima's D 测验和 Fu 与 Li 的 D* 和 F* 测验都没有发现人工选择的证据,说明该基因在育种实践中并没有受到人工选择的影响。究其原因,可能是由于 *ZmSSIIa* 基因尽管参与淀粉的合成,可能并不是影响籽粒产量和品质的关键基因;另一种原因则可能是供试群体经历了瓶颈效应,*ZmSSIIa* 基因的变异大多为近期产生的新突变,尚未经过人工选择的影响。由于本研究中发现玉米 *ZmSSIIa* 基因的很多变异均具有较低的最小等位基因频率,所以在今后分析该基因对产量和品质影响时,需要考虑更大的群体。

4 结论

经基因序列定点捕获和多序列比对,共在全长为 4836 bp 的玉米 *ZmSSIIa* 基因核苷酸序列中发现 78 个 SNP 和 19 个 Indel。该基因的编码区具有 24 个 SNP,其中 12 个为非同义突变位点,可将玉米 *ZmSSIIa* 基因的编码区划分成 9 种单倍型,并编码 7 种蛋白质。中性进化表明该基因在供试玉米自交系中没有经历明显的人工选择。此外,在供试群体中玉米 *ZmSSIIa* 基因至少经历了 8 次遗传重组,而这些遗传重组事件可能对于该基因的遗传多样性及多态性位点间连锁不平衡的快速衰减起到重要作用。

参考文献 (References)

[1] Tian Z, Qian Q, Liu Q, et al. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities[J]. Proceedings of the National Academy of Sciences of the United States of America,

2009, 106(51): 21760–21765.
 [2] Yang Z, Wang Y, Xu S, et al. Molecular evolution and functional divergence of soluble starch synthase genes in cassava (*manihot esculenta* crantz)[J]. Evolutionary Bioinformatics Online, 2013, 9: 239–249.
 [3] Martin C, Smith A M. Starch biosynthesis[J]. The Plant cell, 1995, 7(7): 971–985.
 [4] 谭彩霞,封超年,陈静,等.作物淀粉合成关键酶及其基因表达的研究进展[J].麦类作物学报,2008,28(5):912–919.
 Tan Caixia, Feng Chaonian, Chen Jing, et al. Progress on key enzymes involved in crop starch synthesis and their gene expression[J]. Journal of Triticeae Crops, 2008, 28(5): 912–919.
 [5] Leterrier M, Holappa L D, Broglie K E, et al. Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: Functional and evolutionary implications[J]. BMC Plant Biology, 2008, 8: 98.
 [6] Grimaud F, Rogniaux H, James M G, et al. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis[J]. Journal of Experimental Botany, 2008, 59(12): 3395–3406.
 [7] Zhang X, Colleoni C, Ratushna V, et al. Molecular characterization demonstrates that the *Zea mays* gene *sugary2* codes for the starch synthase isoform SSIIa[J]. Plant Molecular Biology, 2004, 54(6): 865–879.
 [8] Fulton T M, Chunwongse J, Tanksley S D. Microprep protocol for extraction of DNA from tomato and other herbaceous plants[J]. Plant Molecular Biology Reporter, 1995, 13(3): 207–209.
 [9] Aiyar A. The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment[J]. Methods in Molecular Biology, 2000, 132: 221–241.
 [10] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data[J]. Bioinformatics, 2009, 25(11): 1451–1452.
 [11] 朱彩梅,张京.大麦糯性相关基因 *Wx* 单核苷酸多态性分析[J].中国农业科学,2010,43(5):889–898.
 Zhu Caimei, Zhang Jing. Single nucleotide polymorphism of *Wx* gene associated with amylose content in barley germplasm[J]. Scientia Agricultura Sinica, 2010, 43(5): 889–898.
 [12] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism[J]. Genetics, 1989, 123(3): 585–595.
 [13] Fu Y X, Li W H. Statistical tests of neutrality of mutations[J]. Genetics, 1993, 133(3): 693–709.
 [14] Bradbury P J, Zhang Z, Kroon D E, et al. TASSEL: Software for association mapping of complex traits in diverse samples[J]. Bioinformatics, 2007, 23(19): 2633–2635.
 [15] Remington D L, Thornsberry J M, Matsuoka Y, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome[J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98(20): 11479–11484.
 [16] Walbot V. 10 reasons to be tantalized by the B73 maize genome[J]. PLoS Genetics, 2009, 5(11): e1000723.
 [17] Xu S, Yang Z, Zhang E, et al. Nucleotide diversity of Maize *ZmBT1* gene and association with starch physicochemical properties[J]. PloS One, 2014, 9(8): e103627.
 [18] Gore M A, Chia J M, Elshire R J, et al. A first-generation haplotype map of maize[J]. Science, 2009, 326(5956): 1115–1117.
 [19] Yan J, Warburton M, Crouch J. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement[J]. Crop Science, 2011, 51(2): 433–449.
 [20] Ching A, Caldwell K S, Jung M, et al. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines[J]. BMC Genetics, 2002, 3: 19.
 [21] Doebley J F, Gaut B S, Smith B D. The molecular genetics of crop domestication[J]. Cell, 2006, 127(7): 1309–1321.
 [22] Wang H, Nussbaum-Wagler T, Li B, et al. The origin of the naked grains of maize[J]. Nature, 2005, 436(7051): 714–719.
 [23] Palaisa K, Morgante M, Tingey S, et al. Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(26): 9885–9890.

(责任编辑 吴晓丽)