DOI: 10.3724/SP.J.1006.2020.93048

兼容型 maizeSNP384 标记筛选与玉米杂交种 DNA 指纹图谱构建

田红丽** 杨 扬** 王 璐 王 蕊 易红梅 许理文 张云龙
葛建镛 王凤格* 赵久然*

北京市农林科学院玉米研究中心 / 玉米 DNA 指纹及分子育种北京市重点实验室, 北京 100097

摘 要: 为加强玉米品种管理和知识产权保护, 评估确定了一套兼容多技术平台, 适于玉米分子鉴定的 SNP 位点组合, 该组合包含 384 个位点; 构建了 335 个玉米杂交种 SNP-DNA 指纹, 并针对位点和样品从各个角度进行分析。384 个位点全部分布在基因内区域, 显示了较好的多态性; MAF、PIC、DP 平均值分别为 0.39、0.36、0.60, 384 个位点中 88% 的位点 MAF 值高于 0.30, 98% 的位点 PIC 值高于 0.30, 98% 的位点 DP 值高于 0.50。对 335 个玉米杂交种进行遗传相似系数两两比较, GD (1 - Nei 遗传距离) 的变异范围为 0.60~0.99, GD 0.98、0.95、0.90 者分别占比 0.10%、0.38%、1.40%; GS (相同等位基因比) 的变异范围为 0.50~0.99, GS 0.98、0.95、0.90 者分别占比 0.03%、0.11%、0.35%。从 384 个位点中抽取最优位点组合, 12 个位点组合时品种识别率为 0.99, 20 个位点组合能够识别 335 个品种。综上所述, 本研究报道的 384 个核心 SNP 位点具有兼容多平台、高稳定性、高重复性、高品种区分能力, 利用核心位点构建了 335 个玉米杂交种 SNP-DNA 指纹, 为玉米品种分子鉴定、指纹数据构建以及分子育种提供了关键数据支撑。

关键词: 玉米; SNP; 位点组合; 品种分子鉴定; DNA 指纹图谱; 芯片

Screening of compatible maizeSNP384 markers and the construction of DNA fingerprints of maize varieties

TIAN Hong-Li**, YANG Yang**, WANG Lu, WANG Rui, YI Hong-Mei, XU Li-Wen, ZHANG Yun-Long, GE Jian-Rong, WANG Feng-Ge*, and ZHAO Jiu-Ran*

Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences / Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Beijing 100097, China

Abstract: In order to improve the management of maize varieties and the protection of intellectual property rights, a compatible multiple platform loci combination maizeSNP384 was developed as a means of identifying varieties of maize. SNP-DNA fingerprints of 335 maize hybrids were constructed. All the 384 SNPs were located in the intra-genic region, showing good polymorphism. The average MAF, PIC, and DP of 384 loci were 0.39, 0.36, and 0.60, respectively. MAF in 88% of the 384 loci was most than 0.30, PIC in 98% was most than 0.30, and DP in 98% was most than 0.50. Pairwise comparative analysis of 335 maize hybrids using genetic similarity showed that the range of variation of GD (1-Nei genetic distance) was 0.60–0.99, and the ratios of GD value greater than or equal to 0.98, 0.95, and 0.90 were 0.10%, 0.38%, and 1.40%, respectively. The range of variation of GS (same allele ratio) was 0.50–0.99, and the ratios of GS value greater than or equal to 0.98, 0.95, and 0.90 were 0.03%, 0.11%, and 0.35%, respectively. The optimal locus combination was extracted from 384 SNPs, with the variety discrimination power of 0.99 when 12 SNPs were extracted; the combination of 20 SNPs was able to identify all 335 varieties. In summary, 384 core SNPs reported in this study are compatible with multiple platforms, highly stable, highly repeatable and having high variety discrimination power. SNP-DNA fingerprints of 335 maize hybrids were constructed based on these core SNPs, which provides key data

本研究由“十三五”国家重点研发计划项目(2017YFD0102001)和北京学者计划项目(BSP041)资助。

This study was supported by the 13th Five-Year National Key Research and Development Program of China (2017YFD0102001) and the Beijing Scholars Program (BSP041).

* 通信作者(Corresponding authors): 王凤格, E-mail: gege0106@163.com, Tel: 86-10-51503558, Fax: 86-10-51503986; 赵久然, E-mail: maizeshao@126.com, Tel: 86-10-51503936, Fax: 86-10-51503404

** 同等贡献(Contributed equally to this work)

第一作者联系方式: 田红丽, E-mail: tianhongli9963@163.com; 杨扬, E-mail: caurwx@163.com

Received (收稿日期): 2019-08-30; Accepted (接受日期): 2020-03-24; Published online (网络出版日期): 2020-04-17.

URL: <http://kns.cnki.net/kcms/detail/11.1809.S.20200417.1110.004.html>

support for molecular identification, fingerprint data construction, and molecular breeding of maize.

Keywords: maize; SNP; loci combination; variety molecular identification; DNA fingerprint; chip

玉米(*Zea mays* L.)是我国和全球第一大作物,是杂交良种应用最早最普及的作物,在中国农业经济结构中具有重要作用。随着种业市场的发展,截至2018年底,我国已有玉米审定品种12,000个以上,申请植物新品种权7000件以上,每年仍有5000多品种组合在参试(<http://www.zys.moa.gov.cn/gsgg/pzsdgs/>, http://www.nybkjffzx.cn/p_pzbh/pzbh.aspx)。与日俱增的品种数量在繁荣市场的同时,也存在品种间同质化严重、套牌侵权泛滥、种质资源难以正常流动的问题,这些问题严重制约着现代种业的发展。同时随着分子标记辅助育种技术的应用增加了派生品种(Establishing Essential Derivation, EDV)、近似品种的数量,给品种确权鉴定带来新的挑战,迫切需要增加标记密度进而准确评估品种之间的遗传背景。

SNP 标记技术被国际植物新品种保护联盟(International Union for the Protection of New Varieties of Plants, UPOV)推荐为农作物品种鉴定和指纹数据库的方法之一。该技术具有共显性,检测内容为DNA序列信息差异,以二等位基因变异为主,易实现数据整合共享,位点在基因组上分布密度更高更均匀,SNP位点在基因内区域分布较多,增加了与功能基因相关的概率,易实现高通量检测,单个数据点成本低^[1-2]。高通量分型平台的快速发展极大地推动了SNP标记技术在农作物品种鉴定领域中的应用,如位点高通量的芯片平台,样本高通量的原位扫描平台KASP(Kompetitive Allele Specific PCR)、Taqman技术,以及位点和样本均高通量的靶向测序技术等。基于上述平台玉米等主要农作物开展了相关探索工作,Mammadov等报道了玉米SNP位点在上述平台的兼容性情况^[3];大豆方面,筛选了14个SNP标记区分599份大豆种质资源材料^[4];甘蓝方面,筛选了50个核心SNP标记并构建了59个品种的指纹图谱^[5];辣椒方面,筛选了40个SNP位点用于品种鉴定^[6];中国陆地棉方面,筛选了23对核心SNP位点用于品种鉴定^[7]。

就芯片技术而言,全球主要应用美国Illumina和Thermo Affymetrix公司推出的GGGT、infinium、axiom 3种平台;前两者属于光纤微珠芯片,后者为激光微刻芯片;3种技术在数据分析时均对采集的荧

光信号进行标准化,获得数据点的坐标值,然后所有数据点被划分成3簇分别代表2种纯合、1种杂合基因型。基于上述主流芯片平台,报道了多款农作物SNP芯片产品^[8-11]。玉米为二倍体杂交作物,具有重组率高、遗传多样性丰富以及较好的基因组学研究基础,其SNP位点开发、评估具有较好的研究基础^[12-14]。同时多款SNP芯片产品也被推出应用,如在全基因组关联分析、QTL定位、分子育种中应用的maizeSNP50、maize55K芯片;在玉米种质资源遗传评估、分子育种中应用包含768、1536个位点的芯片;适于玉米品种DNA指纹分析和分子育种的通用型SNP芯片maizeSNP3072等^[15-19]。尽管有多种不同通量玉米SNP芯片可供使用,但是商业化的芯片评估材料以美国和欧洲为主,定制芯片基本上是以分析应用于自交系、群体材料、种质资源为主;而玉米品种真实性鉴定和指纹数据库构建的样品类型以杂交种为主,筛选位点需考虑杂交种的特性,因此并不能直接采用已报道的位点组合。

本研究基于已有研究基础,利用代表性玉米自交系和杂交种材料,筛选确定玉米分子鉴定位点组合maizeSNP384,此384个核心SNP位点具有兼容多平台、高稳定性、高重复性、高品种区分能力;基于核心位点建立了335个国家审定玉米品种SNP-DNA指纹数据,并进行了相关分析;以期为玉米分子鉴定、指纹数据构建以及分子育种提供支撑。

1 材料与方法

1.1 供试材料

筛选30套包含杂交种及其双亲的玉米样品作为SNP位点试验效果评估材料(表1),这30份杂交种具有较大推广种植面积,或者为区试组的对照品种。选取1984—2013年国家审定玉米品种335个样品评估SNP位点的多态性、品种区分能力,同时建立标准指纹图谱;样品具体信息参考《国家审定玉米品种SSR指纹图谱》^[20]。

1.2 总DNA提取

选用混株方案制备供试样品,每个样品混合30个以上单株,采用改良CTAB法提取总DNA^[21]。用紫外分光光度计(Nanodrop 2000)测定DNA浓度和质量,工作液浓度统一调整为50 ng μL^{-1} 。

表 1 用于评估 SNP 位点的 30 套杂交种及其双亲的样品信息

Table 1 Sample information of 30 sets hybrids and their parents for SNPs evaluation

编号 No.	品种 Variety	母本 Female parent	父本 Male parent
T01	SC704	B73	Mo17
T02	丹玉 39 Danyu 39	C8605-2	丹 598 Dan 598
T03	登海 11 Denghai 11	DH65232	DH40
T04	登海 9 号 Denghai 9	DH65232	8723
T05	东单 60 Dongdan 60	A801	丹 598 Dan 598
T06	会单 4 号 Huidan 4	SW1611	掖 107 Ye 107
T07	吉单 180 Jidan 180	吉 853 Ji 853	Mo17
T08	吉单 209 Jidan 209	8902	吉 853 Ji 853
T09	京科 968 Jingke 968	京 724 Jing 724	京 92 Jing 92
T10	京科糯 2000 Jingkenuo 2000	京糯 6 Jingnuo 6	白糯 6 Bainuo 6
T11	浚单 20 Xundan 20	浚 9058 Xun 9058	浚 92-8 Xun 92-8
T12	鑫玉 6 号 Liyu 6	543	连 87 Lian 87
T13	龙单 13 Longdan 13	K10	龙抗 11 Longkang 11
T14	鲁单 981 Ludan 981	齐 319 Qi 319	Lx 9801
T15	农大 108 Nongda 108	X178	黄 C Huang C
T16	农大 3138 Nongda 3138	综 31 Zong 31	P138
T17	农华 101 Nonghua 101	NH60	S121
T18	沈单 10 Shendan 10	沈 137 Shen 137	Q1261
T19	沈单 16 Shendan 16	K12	沈 137 Shen 137
T20	四单 19 Sidan 19	444	Mo17
T21	苏玉糯 1 号 Suyunuo 1	通系 5 号 Tongxi 5	衡白 522 Hengbai 522
T22	苏玉糯 5 号 Suyunuo 5	通系 5 号 Tongxi 5	通 366 Tong 366
T23	先玉 335 Xianyu 335	PH6WC	PH4CV
T24	掖单 13 Yedan 13	掖 478 Ye 478	丹 340 Dan 340
T25	渝单 8 号 Yudan 8	掖 478 Ye 478	渝 561 Yu 561
T26	渝糯 7 号 Yunuo 7	S147	S181
T27	豫玉 22 Yuyu 22	综 3 Zong 3	豫自 87-1 Yuzi 87-1
T28	正大 619 Zhengda 619	F06	F19
T29	郑单 958 Zhengdan 958	郑 58 Zheng 58	昌 7-2 Chang 7-2
T30	中单 2 号 Zhongdan 2	Mo17	自 330 Zi 330

1.3 玉米分子鉴定核心 SNP 位点组合的确定, 玉米杂交种 SNP-DNA 指纹图谱构建

将已公布的玉米基因组 3072 个 SNP 位点^[19], 在 Infinium、GGGT、Axiom 不同平台(美国 Illumina 和 Thermo Affymetrix 公司)上进行平行试验, 评估获得兼容多平台、高质量的位点。基于 Illumina GGGT 芯片平台, 利用 335 个玉米杂交种、3072 个位点, 采用 iScan 芯片扫描仪进行荧光信号采集, 获得原始数据, 利用 Genome Studio Genotyping Software v2011 软件分析, 获得基因型数据。同时利用 335 个杂交种进一步评估试验效果, 结合多态性、品种区分能力和均匀分布的原则, 确定候选核心位点; 利

用最优遗传算法^[22]最终确定 1 套适于玉米品种真实性鉴定和标准 DNA 指纹数据构建的核心 SNP 位点组合。同时抽取核心位点组合数据, 形成 335 个玉米杂交种 SNP-DNA 指纹图谱。

1.4 数据分析

利用 python3 语言, 采用 biopython 图形库绘制 SNP 核心位点在染色体上的物理位置示意图(<https://biopython.org/>)。利用 SNP 比对统计工具 (V1.0)(北京市农林科学院玉米研究中心, 软著登记号: 2018SR026743)分析核心位点的系列参数, 包括 MAF (Minor Allele Frequency)、PIC (Polymorphism Information Content)和 DP (Discrimination Power);

DP 是指单个位点的品种识别能力, $DP=1-\sum(p_i)^2$, p_i 是指每种基因型的频率^[23]。

利用 python3 语言, 采用 matplotlib 图形库绘制玉米杂交种的 SNP-DNA 指纹图谱 (<https://matplotlib.org/>)。利用 SNP 比对统计工具(V1.0)(软著登记号为 2018SR026743)分析 335 个杂交种的杂合率, 品种两两成对比较差异位点分布, 两两比较的遗传相似系数值的分布图。遗传相似参数分别采用(1 - Nei1973)遗传距离(GD), 相同等位基因比值(GS), $(1 - \text{差异位点比值}) \times 100\%$ (GL) 3 种方案。GS = $(1-b/a) \times 100\%$, 其中 a 为比较总等位基因数目, 由于 SNP 位点为二等位基因变异, 故 a 为比较总位点数的 2 倍, b 为差异等位基因数目。GL 为 $1 - \text{差异位点数目}/\text{比较总位点数目}$ 。

利用 SNP/INDEL 位点筛选工具(V1.0) (北京市农林科学院玉米研究中心, 软著登记号为 2018SR003573), 采用“基于遗传算法的植物品种真实性鉴定位点筛选方法”^[22], 从本研究确定的 384 个 SNP 核心位点中逐步抽取位点组合, 包含的位点数目从 1 个增加到 20 个, 分析其品种识别率的变化情况。每组位点抽取时主要依据位点和位点组合的品种区分能力, 在 384 个位点中随机抽取。

2 结果与分析

2.1 适于玉米品种真实性鉴定和标准 DNA 指纹数据构建的 SNP 位点筛选

基于 maizeSNP3072 位点组合, 根据试验效果、

品种区分效率、均匀分布原则筛选确定 1 套适于玉米品种真实鉴定和标准 DNA 指纹数据构建的核心位点组合 maizeSNP384^[19]。具体筛选过程分为 3 步: (1)基于 30 套杂交种及其双亲样品, 利用 Infinium、GGGT、Axiom 3 种不同芯片平台进行位点筛选。评估指标主要为, 在 3 种芯片平台上均能成功设计探针序列, 评估分值 0.70, 3 个平台之间数据完全一致; 数据获得率高于 95%, MAF 值 0.05; 获得候选 SNP 位点为 1212 个。(2)根据 335 个杂交种的数据, 进一步做试验质量和多态性评估。要求在不同芯片平台上 3 种基因型簇 AA、BB 和 AB 之间具有明显的界限, 并且各自内聚(图 1); 多态性指标为 MAF 0.2、PIC 0.25、DP 0.4; 获得候选 SNP 位点为 800 个。(3)利用最优遗传算法, 结合在染色体上相对均匀分布的原则, 最终筛选确定了兼容多平台、高稳定性、高重复性、高品种区分能力的 384 个核心 SNP 位点(附表 1)。

2.2 maizeSNP384 位点组合的基本特征

将 384 个 SNP 位点的侧翼序列各 35 bp 比对到 B73 参考基因组(AGP-V3 版本)获得每个位点所在的染色体和物理位置信息。图 2-A 展示了位点在玉米 10 对染色体上的分布情况, 384 个位点基本均匀分布在长臂和短臂上, 在端粒和着丝粒区域分布极少, 第 2 染色体位点数目相对少。384 个核心位点全部分布在基因内区域, 其中 52%分布在 exon 区, 20%分布在 3'UTR 区, 16%分布在 promoter 区, 7%分布在 5'UTR 区, 5%分布在 intron 区(图 2-B)。

图 1 SNP 位点(MG004)在 3 种芯片平台上的分型图

Fig. 1 Genotyping patterns of MG004 SNP locus on three different chip platforms

A: illumina GGGT 芯片平台; B: illumina infinium 芯片平台; C: affymetrix-axiom 芯片平台。图中每个点为 1 个基因型数据, 相同基因型的数据点聚集在一起; 两侧红色和蓝色的 2 组数据点为纯合基因型, 中间 1 组数据点为杂合基因型。

A: illumina GGGT platform; B: illumina infinium platform; C: affymetrix-axiom platform. Each point is an actual genotype call, the data points of the same genotype gather together; the red and blue data points on both sides are homozygous genotypes, the data point on the middle group are heterozygous genotypes.

图 2 maizeSNP384 位点组合的分布

Fig. 2 Distribution of maizeSNP384 combinations

A: 384 个 SNP 位点在玉米 10 对染色体上分布图; B: 384 个 SNP 位点在基因上的分布情况。

A: the distribution of 384 SNPs on maize 10 chromosomes; B: the distribution of SNPs in different gene regions.

基于 335 个玉米杂交种的数据评估 SNP 位点的各遗传参数, 384 个位点显示了较好的多态性、品种区分能力(图 3)。384 个核心 SNP 位点的 MAF、PIC、DP 值变异范围分别为 0.17~0.50、0.25~0.40、0.44~0.66, 平均值分别为 0.39、0.36、0.60。384 个位点中 88% 的位点 MAF 值高于 0.30, 98% 的位点 PIC 值高于 0.30, 98% 的位点 DP 值高于 0.50 (图 3, 附表 1)。

2.3 335 个国家审定玉米品种指纹图谱构建及分析

图 4 为利用 384 个 SNP 标记组合构建了 335 个玉米杂交种的基因型指纹图谱, 该图显示每个品种的基因型组合均不同, 即 384 个 SNP 位点能够有效区分所测试的杂交种。基于 384 个 SNP 位点的基因数据, 335 个国家审定玉米品种基因型数据获得率为 97.33%; 样品杂合基因型(父母本为不同的纯合基因

型情况下, 杂交种为杂合基因型)频率变异范围为 0.17~0.63, 平均值为 0.47, 97% 的样品杂合基因型频率值介于 0.35~0.60 之间(图 5-A)。335 个玉米品种两两比较, 总共比较的对数有 55,945 对; 结果显示差异位点数目范围为 1~283 个, 差异位点数目多集中在 100~270 个, 差异位点数目在 20 个以内的有 12 对品种(图 5-B)。基于 384 个 SNP 位点基因型数据, 对 335 个玉米杂交种进行遗传相似系数两两比较, GD (1 - Nei 遗传距离)、GS (相同等位基因比值)、GL (1 - 差异位点的比值) 3 种参数的值依次降低。GD 的变异范围为 0.60~0.99, 平均值为 0.77, 相似性数值多集中在 0.65~0.90, GD 0.98、0.95、0.90 者分别占比 0.10%、0.38%、1.40%。GS 的变异范围为 0.50~0.99, 平均值为 0.65, 相似性数值多集中在 0.55~0.80, GS 0.98、0.95、0.90 者分别占比 0.03%、0.11%、0.35%。GL 的变异范围为 0.22~0.99, 平均

图 3 384 个 SNP 位点的 MAF 和 PIC 多态性指数变化分布图

Fig. 3 Distribution of MAF and PIC values of 384 SNP loci

横坐标为 MAF、PIC 的区间值, 纵坐标为 SNP 位点数目。

The abscissa is the interval value of MAF and PIC, and the ordinate is the number of SNP loci.

图 4 335 个玉米杂交种 SNP-DNA 指纹总体展示图

Fig. 4 The overall display of SNP-DNA fingerprints of 335 maize hybrids

每行代表 1 个品种, 按照国家审定编号排序; 每列代表 1 个 SNP 位点, 按照在基因组上物理位置排序; DNA 指纹分别用不同的颜色表示。

Each row represents one variety sorted by national approval number; each column represents one SNP locus, sorted by physical location in genome; DNA fingerprints are displayed in different colors.

值为 0.40, 相似性数值多集中在 0.30~0.60, GL 0.98、0.95、0.90 者分别占比 0.02%、0.03%、0.09% (图 5-C)。

2.4 针对于 335 个国家审定玉米杂交种, 随着位点组合数目的变化, 品种区分效率变化情况

基于 384 个 SNP 位点、335 个玉米杂交种的基因型数据, 采用“基于遗传算法的植物品种真实性鉴定位点筛选方法”, 从 384 个位点中抽取位点组合, 位点数目为从 1 个增加到 20 个, 分析每套位点组合的品种识别率变化曲线^[22]; 横坐标为最优位点组合包含的位点数目, 纵坐标为品种识别率(图 6)。在本研究中被比较的 2 个品种如果有 1 个位点的基因型不同则视为两者能够被识别。图 6 显示, 2 个位点组合品种识别率为 0.04, 3 个位点的识别率为 0.10, 8 个位点的识别率超过 0.90; 12 个位点或更多位点时品种识别率为 0.99, 其数值趋于稳定; 最终抽取了 20 个位点, 能够识别所有品种; 20 个位点相对均匀分

图 5 335 个玉米品种的杂合基因型频率、遗传差异分析

Fig. 5 Analysis of heterozygous genotype rate and genetic difference of 335 maize hybrid lines

A: 335 个玉米品种的杂合基因型频率分布(杂合基因型的标记数目/384); B: 335 个玉米品种两两成对比较差异位点数分布图, 横坐标为成对比较差异位点数目, 纵坐标为成对比较的数目; C: 335 个玉米品种两两成对比较遗传相似系数分布图, 横坐标为遗传相似系数百分比, 纵坐标为成对比较的数目。

A: the heterozygous genotype rate of the 335 maize hybrid lines (number of loci for heterozygous genotypes/384). B: distribution of different locus number obtained by pairwise comparative analysis of 335 maize hybrid lines, the abscissa is the number of differential loci, and the ordinate is the number of paired comparisons. C: distribution of genetic similarity coefficient obtained by pairwise comparative analysis of 335 maize hybrid lines, the abscissa is percentage of genetic similarity coefficient, and the ordinate is the number of paired comparisons.

图 6 基于最优位点组合的品种识别率变化曲线图

Fig. 6 Variety identity probability curve based on the optimal loci combination

布在 10 条染色体上, 每条染色体上均抽取到了位点, 根据染色体的长度不同包含的位点数目亦不同。

3 讨论

3.1 适于玉米品种分子鉴定的 SNP 标记特征, SNP-DNA 指纹数据表现形式

筛选一套适宜的核心标记位点, 确定指纹数据表现形式对于农作物品种标准 DNA 指纹构建至关重要。农作物品种鉴定内容和目的不同, 筛选标记位点的标准亦不同。适于玉米品种分子鉴定的 SNP 标记组合需在品种间多态性高; 试验评估效果好, 表现为数据信号值高, 基因型数据能够自动化统计; 具有较好的重复性和稳定性; 兼容多平台; 在基因组上物理位置已知; 相对均匀分布, 引物位点间没有紧密连锁。

SNP 标记虽然以二等位基因变异为主, 其基因型数据只出现 A/T/C/G 四种碱基, 但是存在相同位点在不同平台其等位基因的定义不同, DNA 双链中定义哪条链的不同, 2 个等位基因存在定义成 A 或 B 的不同。因此如果不加以规范标准化, 不同平台之间, 相同平台不同的探针设计均会存在数据整合的问题。综上所述, 玉米品种 SNP-DNA 指纹数据格式建议使用 A/T/C/G 碱基, 不推荐采用其他编码转换的方式。为了不同平台、不同实验室、不同批次之间的数据整合, 必须提供至少 2 个参照样品的指纹信息, 要求参照样品在每个位点上均为纯合基因型且无数据缺失, 参照样品之间具有较大的遗传差异。在玉米中, 建议选取不同杂优群的 DH 系作为参照样品。

3.2 基于 SNP 标记方法, 玉米品种鉴定最适位点数目分析

农作物品种鉴定最适位点数目一直是本领域关注的重点和难点, 它与鉴定目的、标记类型、检测材料等相关。一般情况下, 减少位点数目, 能够相对简便、快捷、稳定地反映品种之间的差异。但是如果位点数目太少则遗传距离评估结果不稳定, 尤其是对于亲缘关系较近的材料。这也反映出品种真实性鉴定和派生品种、近似品种鉴定对位点数目的要求不同。

真实性鉴定侧重考察品种间是否存在明显差异, 鉴定的过程是“找差异”, 只要差异位点数目高于阈值即可判定结果。品种派生关系鉴定是基于已经认定具有特异性的、申请品种权保护的品种, 分析待测样品是独立的品种还是与已知品种存在派生关系; 派生关系鉴定侧重考察品种间的遗传背景的相似程度, 鉴定的过程是“找相似”, 真实反映两个品种之间的相似程度。因此品种真实性鉴定和依赖性派生鉴定的最适位点数目级别不同, 前者相对少, 一般在几十至几百级别, 后者相对多, 一般在几百至几千级别^[24-25]。

目前用于农作物品种鉴定标记以 SSR 和 SNP 为主。两种标记由于单个位点多态性的差异, 在品种鉴定时选用的标记数目亦不同。SSR 和 SNP 两种标记位点组合多态性的倍数关系没有定论, 这与筛选位点的原则, 位点组合效果, 评估材料遗传背景等均有关系。ISF (International Seed Federation, 国际种子联盟) 颁布的基于 SNP 标记进行玉米派生品种鉴定的指南中选用了 3072 个 SNP 位点^[26-27]。基于

美国玉米自交系样品的分析, 305 个 SNP 位点组合与 150 个 SSR 位点组合结果一致性最高, 此 150 个 SSR 位点是 ASTA (American Seed Trade Association, 美国种子贸易协会) 发布的用于美国玉米品种鉴定的位点组合^[28-29]。国内玉米品种鉴定 SSR 标记法中公布了 40 对 SSR 引物, 这 40 对 SSR 引物已成熟应用于玉米品种鉴定^[21]。本研究报道的 384 个 SNP 位点, 从中抽取 20 个位点, 能够区分 335 个玉米审定品种; 如果位点数目为 200 个预计能够区分玉米已知品种的 95% 及以上, 与 40 对 SSR 引物区分效果相当; 但是如果应用于玉米派生、近似样品的遗传相似度分析, 需要进一步增加位点到数千个。

3.3 基于 SNP 标记方法, 玉米品种鉴定分析参数和阈值的探讨

农作物品种鉴定分析参数、阈值与鉴定内容、位点数目、以及作物种类等相关。对于品种真实性鉴定, 分析参数一般为差异位点数目; 确定判定阈值的思路为首先基于真实性鉴定的核心位点组合建立已知品种的 DNA 指纹数据库; 根据品种间差异位点分布, 品种内变异情况, 结合实际情况确定与预期相符的差异位点阈值。对于派生品种、近似品种鉴定, 分析参数一般为遗传相似度参数; 确定判定阈值的思路为首先基于全基因组均匀选取一套位点组合, 构建代表性样品的 DNA 指纹数据库; 然后根据样品间的遗传相似度分布情况, 结合表型及系谱信息, 确定合适的遗传相似度阈值。

ISF 颁布的基于 SSR 标记玉米派生品种鉴定指南中, 阈值设定为 82%~90%, 即遗传相似度低于或等于 82% 的判定为非派生品种, 即不同品种; 介于两者之间的为质疑区间, 需提供其他辅助材料进行判定; 高于 90% 的判定为派生品种^[30]。ISF 颁布的基于 SNP 标记玉米派生品种鉴定指南中, 阈值设定为 91% 和 95%, 因为 SNP 单位点区分效率低于 SSR, 所以基于 SNP 标记的阈值做了相应的提升^[27]。国内玉米品种真实性鉴定 SSR 标记法公布了 40 对 SSR 引物, 判定阈值为 2 个位点, 差异位点百分比占据 5%^[21]。基于本研究报道的 384 个位点对 335 个玉米品种进行遗传相似系数的两两比较, GD 0.98、0.95、0.90 者分别占比 0.10%、0.38%、1.40%。根据我国已知玉米品种实际情况和现行的 SSR 行业标准, 基于 SNP 标记遗传相似度阈值的设定建议高于 ISF 颁布的标准, 具体阈值需要进一步深入研究。

3.4 兼容多平台的位点组合 maizeSNP384 的应用前景

核心引物组合法概念早在 2003 年已经提出, 是指在染色体上均匀选取多态性高、稳定性高、重复性好、整体品种区分能力高的一组引物/位点组合; 该方案的提出解决了农作物品种鉴定和数据整合共享的难题^[24,31]。本研究报道的 384 个 SNP 位点为基于最优遗传算法, 结合相对均匀分布的原则, 筛选的一套兼容多平台、高稳定性、高重复性、高品种区分能力的位点组合。从中抽取了 20 个位点作为一组最优位点组合, 能够区分所评估的 335 个杂交种; 如果抽取 200 个位点预计能够区分玉米已知品种的 90% 及以上。故 maizeSNP384 位点组合在未来基于 SNP 标记的玉米品种分子鉴定中能够发挥重要作用, 可作为筛选玉米品种真实性、特异性、纯度鉴定以及类群划分的候选位点组合, 为分子标记辅助育种中背景分析提供研究基础。

SNP 标记可以应用的高通量基因分型技术除了芯片平台之外, 还有样本高通量的原位扫描平台(英国 LGC 公司 KASP 技术, 美国 Life 公司 Taqman 技术), 以及位点和样本均高通量的靶向测序技术等。由于 Illumina GGGT 芯片探针设计原理与 KASP、Taqman 技术相似度极高, 有报道称 GGGT 芯片与 KASP、Taqman 技术的位点转换率高达 98% 和 95%^[3]。本文中的 384 个 SNP 位点在 illumina GGGT、infinium、AFFY axiom 芯片技术中均表现出了较好评估效果。如从 384 个 SNP 位点中选取几十或者上百个最优位点组合, 利用样本高通量的 KASP 或 TaqMan 技术, 可以实现高效率、高灵活性、简便快捷、低成本的检测应用需求。

综上所述, 不管是位点高密度的芯片平台, 还是样本高通量的 KASP、Taqman 技术平台, 亦或是位点样本高通量的测序平台, 本研究报道的 384 个位点均能达到较高成功率, 在未来基于 SNP 标记的玉米品种分子鉴定、分子育种中发挥重要作用。

4 结论

评估确定了玉米分子鉴定位点组合 maizeSNP384, 此 384 个核心 SNP 位点具有兼容多平台、高稳定性、高重复性、高品种区分能力; 基于核心位点建立了 335 个国家审定玉米品种 SNP-DNA 指纹数据, 为玉米品种分子鉴定、指纹数据构建以及分子育种提供了关键数据支撑。

附表 请见网络版: 1) 本刊网站 <http://zwxb.china-crops.org/>; 2) 中国知网 <http://www.cnki.net/>; 3) 万方数据 <http://c.wanfangdata.com.cn/Periodical-zuowxb.aspx>。

References

- [1] 唐立群, 肖层林, 王伟平. SNP 分子标记的研究及其应用进展. *中国农学通报*, 2012, 28: 154–158.
Tang L Q, Xiao C L, Wang W P. Research and application progress of SNP markers. *Chin Agric Bull*, 2012, 28:154–158 (in Chinese with English abstract).
- [2] UPOV (International Union for the Protection of New Varieties of Plants). Possible Used of Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS). UPOV, Geneva, Switzerland, 2011.
- [3] Mammadov J, Chen W, Mingus J, Thompson S, Kumpatla S. Development of versatile gene-based SNP assays in maize (*Zea mays* L.). *Mol Breed*, 2012, 29: 779–790.
- [4] 魏中艳, 李慧慧, 李骏, Gamar Y A, 马岩松, 邱丽娟. 应用 SNP 精准鉴定大豆种质及构建可扫描身份证. *作物学报*, 2018, 44: 315–323.
Wei Z Y, Li H H, Li J, Gamar Y A, Ma Y S, Qiu L J. Accurate identification of varieties by nucleotide polymorphisms and establishment of scannable variety IDs for soybean germplasm. *Acta Agron Sin*, 2018, 44: 315–323 (in Chinese with English abstract).
- [5] 李志远, 于海龙, 方智远, 杨丽梅, 刘玉梅, 庄木, 吕红豪, 张扬勇. 甘蓝 SNP 标记开发及主要品种 DNA 指纹图谱构建. *中国农业科学*, 2018, 51: 2771–2787.
Li Z Y, Yu H L, Fang Z Y, Yang L M, Liu Y M, Zhuang M, Lyu H H, Zhang Y Y. Development of SNP markers in cabbage and construction of DNA fingerprinting of main varieties. *Sci Agric Sin*, 2018, 51: 2771–2787 (in Chinese with English abstract).
- [6] Jung J K, Park S W, Liu W Y, Kang B C. Discovery of single nucleotide polymorphism in Capsicum and SNP markers for cultivar identification. *Euphytica*, 2010, 175: 91–107.
- [7] Kuang M, Wei S J, Wang Y Q, Zhou D Y, Ma L, Fang D, Yang W H, Ma Z Y. Development of a core set of SNP markers for the identification of upland cotton cultivars in China. *J Integr Agr*, 2016, 15: 964–962.
- [8] Chen H D, Xie W B, He H, Yu H H, Chen W, Li J, Yu R B, Yao Y, Zhang W H, He Y Q, Tang X Y, Zhou F S, Deng X W, Zhang Q F. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant*, 2014, 7: 541–553.
- [9] Winfield M O, Allen A M, Burrige A J, Barker G L A, Benbow H R, Wilkinson P A, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley A R, Edwards K J. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J*, 2016, 14: 1195–1206.
- [10] Song Q J, Hyten D L, Jia G F, Quigley C V, Fickus E W, Nelson R L, Cregan P B. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One*, 2013, 8: e54985.
- [11] Hulse-Kemp A M, Lemm J, Plieske J, Ashrafi H, Buyyarapu R, Fang D D, Frelichowski J, Giband M, Hague S, Hinze L L, Kochan K J, Riggs P K, Scheffler J A, Udall J A, Ulloa M, Wang S S, Zhu Q H, Bag S K, Bhardwaj A, Burke J J, Byers R L, Claverie M, Gore M A, Harker D B, Islam M S, Jenkins J N, Jones D C, Lacape J M, Llewellyn D J, Percy R G, Pepper A E, Poland J A, Rai K M, Sawant S V, Singh S K, Spriggs A, Taylor J M, Wang F, Yourstone S M, Zheng X T, Lawley C T, Ganai M W, Deynze A V, Wilson I W, Stelly D M. Development of a 63K SNP Array for cotton and high-density mapping of intraspecific and interspecific populations of *Gossypium* spp. *Genes Genomes Genet*, 2015, 5: 1187–1209.
- [12] Lai J S, Li R Q, Xu X, Jin W W, Xu M L, Zhao H N, Xiang X K, Song W B, Ying K, Zhang M, Jiao Y P, Ni P X, Zhang J G, Li D, Guo X S, Ye K X, Jian M, Wang B, Zheng H S, Liang H Q, Zhang X Q, Wang S C, Chen S J, Li J S, Fu Y, Springer N M, Yang H M, Wang J, Dai J R, Schnable P S, Wang J. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet*, 2010, 42: 1027–1030.
- [13] Jiao Y P, Zhao H N, Ren L H, Song W B, Zeng B, Guo J J, Wang B B, Liu Z P, Chen J, Li W, Zhang M, Xie S J, Lai J S. Genome-wide genetic changes during modern breeding of maize. *Nat Genet*, 2012, 44: 812–817.
- [14] Chia J M, Song C, Bradbury P J, Costich D, Leon N D, Doebley J, Elshire R J, Gaut B, Geller L, Glaubitz J C, Gore M, Guill K E, Holland J, Hufford M B, Lai J S, Li M, Liu X, Lu Y L, McCombie R, Nelson R, Poland J, Prasanna B M, Pyhäjärvi T, Rong T Z, Sekhon R S, Sun Q, Tenaillon M I, Tian F, Wang J, Xu X, Zhang Z U, Kaeppler S M, Ross-Ibarra J, McMullen M D, Buckler E S, Zhang G Y, Xu Y B, Ware D. Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet*, 2012, 44: 803–809.
- [15] Ganai M W, Durstewitz G, Polley A, Bérard A, Buckler E S, Charcosset A, Clarke J D, Graner E M, Hansen M, Joets J, Paslier M C, McMullen M D, Montalent P, Rose M, Schön C C, Sun Q, Walter H, Martin O C, Falque M. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping and genetic mapping to compare with the B73 reference genome. *PLoS One*, 2011, 6: e28334.
- [16] Cheng X, Ren Y H, Jian Y Q, Guo Z F, Zhang Y, Xie C X, Fu J J, Wang H W, Wang G Y, Xu Y B, Li P, Zou C. Development of a maize 55K SNP array with improved genome coverage for molecular breeding. *Mol Breed*, 2017, 37: 20.
- [17] Jones E, Chu W C, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski A, Smith O S, McMullen M D, Bezawada C, Warren J, Babayev J, Basu S, Smith S. Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.). *Mol Breed*, 2009, 24: 165–176.
- [18] Yan J B, Yang X H, Shah T, Sánchez-Villeda H, Li J S, Warburton M, Zhou Y, Crouch J H, Xu Y B. High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol Breed*, 2010, 25: 441–451.
- [19] Tian H L, Wang F G, Zhao J R, Yi H M, Wang L, Wang R, Yang Y, Song W. Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. *Mol Breed*, 2015, 35: 136.
- [20] 王凤格, 赵久然, 杨扬. 国家审定玉米品种 SSR 指纹图谱. 北京: 中国农业科学技术出版社, 2015.

- Wang F G, Zhao J R, Yang Y. SSR Fingerprinting Data of National Approval of Maize Cultivars. Beijing: China Agricultural Science and Technology Press, 2015 (in Chinese).
- [21] 王凤格, 易红梅, 赵久然, 刘平, 张新明, 田红丽, 堵苑苑. 玉米品种鉴定技术规程 SSR 标记法. 中华人民共和国农业行业标准, NY/T 1432-2014, 2014.
- Wang F G, Yi H M, Zhao J R, Liu P, Zhang X M, Tian H L, Du Y Y. Protocol for the Identification of Maize Varieties—SSR Marker Method. Agricultural Industry Standards of the People's Republic of China, NY/T 1432-2014, 2014 (in Chinese).
- [22] 王凤格, 赵久然, 杨扬. 基于遗传算法的植物品种真实性鉴定位点筛选方法. 中国发明专利, 2013, ZL201310629676.2.
- Wang F G, Zhao J R, Yang Y. The Method of Locus Selecting for Plant Variety Authenticity Identification Based on Genetic Algorithm. Chinese Invention Patent, 2013, ZL201410495417.X (in Chinese).
- [23] Tessier C, David J, This P, Boursiquot J M, Charrier A. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor Appl Genet*, 1999, 98: 171-177.
- [24] 王凤格, 田红丽, 易红梅, 赵涵, 霍永学, 匡猛, 张力科, 吕远大, 丁曼卿, 赵久然. 植物品种 DNA 指纹鉴定原理及其鉴定方案. 分子植物育种, 2018, 16: 4756-4766.
- Wang F G, Tian H L, Yi H M, Zhao H, Huo Y X, Kuang M, Zhang L K, Lyu Y D, Ding M Q, Zhao J R. Principle and strategy of DNA fingerprint identification of plant variety. *Mol Plant Breed*, 2018, 16: 4756-4766 (in Chinese with English abstract).
- [25] 赵久然, 王凤格, 田红丽, 易红梅, 王蕊, 葛建镭. 适于农作物品种分子身份鉴别和确权鉴定的检测方法. 中国发明专利, 2017, 201710527354.5.
- Zhao J R, Wang F G, Tian H L, Yi H M, Wang R, Ge J R. Detection Methods for Identity Distinguish and Intellectual Property Confirmation of Crop Varieties Using Molecular Markers. Chinese Invention Patent, 2017, ZL201710527354.5 (in Chinese).
- [26] Rousselle Y, Jones E, Charcosset A, Moreau P, Robbins K, Stich B, Knaak C, Flament P, Karaman Z, Martinant J R, Fourneau M, Taillardat A, Romestant M, Tabel C, Bertran J, Ranc N, Lespinasse D, Blanchard P, Kahler A, Chen J, Kahler J, Dobrin S, Warner T, Ferris R, Smith S. Study on essential derivation in maize: III. Selection and evaluation of a panel of single nucleotide polymorphism loci for use in European and North American Germplasm. *Crop Sci*, 2015, 55: 1170-1180.
- [27] ISF. Guidelines for Handling Disputes on Essential Derivation of Maize Lines. 2014 (Adopted by the ISF maize and sorghum section in Beijing, in May 2014), (www.worldseed.org).
- [28] Nelson B K, Kahler A L, Kahler J L, Mikel M A, Thompson S A, Ferriss R S, Smith S, Jones E S. Evaluation of the numbers of single nucleotide polymorphisms required to measure genetic distance in maize (*Zea mays* L.). *Crop Sci*, 2011, 51: 1470-1480.
- [29] Kahler A L, Kahler J L, Thompson S A, Ferriss R S, Jones E S, Nelson B K, Mikel M A, Smith S. North American study on essential derivation in maize: II. Selection and evaluation of a panel of simple sequence repeat loci. *Crop Sci*, 2010, 50: 486-503.
- [30] ISF. Guidelines for handling disputes on essential derivation of maize lines. 2008 (Adopted by the ISF maize and sorghum section in Prague, May 2008), (www.worldseed.org).
- [31] 王凤格, 赵久然, 郭景伦, 余花娣, 陈刚. 比较三种 DNA 指纹分析方法在玉米品种纯度及真伪鉴定中的应用. 分子植物育种, 2003, 1: 655-661.
- Wang F G, Zhao J R, Guo J L, She H D, Chen G. Comparison of three DNA fingerprint analyzing methods for maize cultivars' identification. *Mol Plant Breed*, 2003, 1: 655-661 (in Chinese with English abstract).