

基于最小平方误差的单倍体和二倍体分类方法研究

李浩光^{1,2} 李卫军¹ 覃 鸿¹ 陈绍江³ 刘 金³ 李 伟³

(1. 中国科学院半导体研究所高速电路与神经网络实验室, 北京 100083; 2. 中国石油大学(华东)信息与控制工程学院, 东营 257061; 3. 中国农业大学国家玉米改良中心, 北京 100193)

摘要: 基于玉米油分的花粉直感效应,采用核磁共振鉴别单倍体,快速、准确、低成本地确定单倍体与二倍体之间的含油率阈值,是实现大批量自动鉴别的前提和关键。提出一种基于最小平方误差的单倍体和二倍体分类方法,通过少量籽粒确定含油率阈值的研究,对单倍体和二倍体进行识别。以 10 个玉米组合为研究对象,利用核磁共振分析仪测定各组合一定数量单倍体和二倍体籽粒的含油率,依据最小平方误差方法确定各组合油分阈值,并通过逐级递增训练集样本数量的方法,确定油分阈值所需训练集规模,获得含油率阈值,能够满足单倍体正确识别率及对二倍体正确拒识率的均值达到 90% 以上的要求,为核磁共振全自动单倍体鉴别系统的后期大规模分类提供参数指导,提高分选效率。同时因单倍体生产效率低,运用最小平方误差的方法,通过少量的单倍体及二倍体籽粒来构建油分鉴别单倍体模型,以较小的投入获得更优的分类效果在玉米单倍体自动化分选中具有非常重要的现实意义,将该分类方法应用到全自动核磁分选单倍体系统中,可以达到较好的效果,为加快单倍体工程化育种提供保障。

关键词: 玉米育种; 油分; 单倍体; 二倍体; 最小平方误差; 分类

中图分类号: O657.3; S123 **文献标识码:** A **文章编号:** 1000-1298(2016)06-0259-06

Classifying Method of Haploid and Diploid Based on Least Square Error

Li Haoguang^{1,2} Li Weijun¹ Qin Hong¹ Chen Shaojiang³ Liu Jin³ Li Wei³

(1. Laboratory of High Speed Circuit and Neural Network, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China

2. College of Information and Control Engineering, China University of Petroleum (East China), Dongying 257061, China

3. National Maize Improvement Center of China, China Agricultural University, Beijing 100193, China)

Abstract: In the identification of haploid maize based on pollen xenia effect, it is of vital importance to determine oil content threshold between haploid and diploid rapidly, accurately and efficiently for automatic classification of haploid by automatic magnetic sorting system in large quantities. A new method based on the least square error was proposed to discriminate haploid from diploid maize seed. Aiming at 10 crosses of maize, through incrementing the number of samples contained in each training set, the size of the training set which can get ideal recognition rate was determined. Principle of least square error used to classify haploid from diploid maize seed was described, and then several experiments were conducted to verify the effectiveness of the proposed method. Seed oil content was measured by nuclear magnetic resonance (NMR) analyzer in the experiments. The least square error method can rapidly determine oil content threshold between haploid and diploid with low amount of samples to realize the practical goal that the haploid correct recognition rate and the diploid correct rejection rate reached more than 90%, and it would provide parameters guidance for the haploid automatic nuclear magnetic

收稿日期: 2015-11-27 修回日期: 2016-02-02

基金项目: 国家重大科学仪器设备开发专项(2014YQ470377)、中央高校基本科研业务费专项(15CX02103A)和国家公派访问学者项目(留金发[2014]3012号)

作者简介: 李浩光(1981—),男,博士生,中国石油大学讲师,主要从事智能信息处理和模式识别研究,E-mail: lihaoguang@semi.ac.cn

通信作者: 李卫军(1975—),男,研究员,博士生导师,主要从事智能信息处理、机器视觉和近红外光谱研究,E-mail: wjli@semi.ac.cn

resonance (NMR) verification system in the later large-scale classification and improve the efficiency of separation. It was very practical to obtain good classification result with low cost, and it would provide support for the development of haploid engineering.

Key words: maize seed; oil content; haploid; diploid; least square error; classification

引言

单倍体育种技术是一种生物技术与常规育种方法相结合的快速、高效育种新途径,在该技术中获取足量的单倍体籽粒至关重要。自然条件下,玉米产生单倍体的概率极低(0.05% ~ 0.1%),即使利用玉米单倍体诱导系进行生物诱导也不足10%^[1-4]。工程化育种需要从大量单倍体、二倍体混合籽粒中快速、准确地鉴定出少量的单倍体籽粒。目前国内外对玉米单倍体常规鉴定的方法有:形态学鉴定、细胞学和解剖学鉴定、分子生物学鉴定、射线照射、遗传标记法等^[5-7],但以上方法存在耗时、耗力、成本高等缺陷,且多为有损检测,无法实现玉米单倍体籽粒的快速、无损检测。

宋天明和陈绍江教授于2003年率先提出利用高油玉米的花粉直感效应鉴别单倍体,为油分鉴别单倍体提供了理论依据,这种方法相比上述常规方法具有无损、快速、准确的特点。利用高油诱导系作为父本对普通玉米自交系杂交诱导时,单倍体籽粒的胚由母本雌配子单独发育而来^[8-9],不存在花粉直感效应,因此单倍体籽粒含油率(下文亦称“油分”)明显低于杂交二倍体籽粒,若通过核磁共振仪器检测各玉米籽粒含油率,并确定单倍体及二倍体的含油率阈值就可以对两类籽粒进行区分。

利用玉米含油率鉴别单倍体的前提和关键是确定含油率阈值。MELCHINGER等^[10-11]采用贝叶斯分类方法,在假设单倍体和二倍体籽粒含油率分布是高斯分布的基础上,使用基于概率统计的贝叶斯分类器计算误选率(False discovery rate, FDR)、漏选率(False negative rate, FNR)、F-score,并指出单倍体鉴别的阈值 t 应当由以上3个因素来综合确定。2015年MELCHINGER等^[12]对控制单倍体和二倍体籽粒误识率进行了分析,并对高斯分布中的均值和方差的具体确定方法进行了研究。MELCHINGER等^[13]分析了含油率、籽粒质量、单籽粒油分质量,在单倍体和二倍体籽粒的分类中起到不同比重的作用,得出含油率具有最优分类效果的结论。MELCHINGER等^[10]采集了大量的单倍体和二倍体玉米籽粒,为确定4个组合玉米的油分阈值,采集种子数量分别为213、218、919、1 310粒,并逐个测量其油分,进而估算单倍体和二倍体概率密度

函数中的均值和方差,最终确定各组合的玉米油分阈值,由此可见,通过基于贝叶斯分类器的方法来确定油分阈值需花费大量时间与精力。

本文以中国农业大学国家玉米改良中心提供的10份玉米杂交组合(不同杂种优势群内杂交组合)的玉米籽粒为研究对象,提出一种基于最小平方误差的单倍体和二倍体分类方法,研究利用少量籽粒对单倍体和二倍体进行两类样本判别区分,以较小的代价获取最优的识别分类效果,避免利用基于概率统计的贝叶斯分类器确定单倍体与二倍体油分阈值时所需的大批量测试工作,并利用10份玉米杂交组合的全部数据,采用逐级递增训练集所含样本的方法,确定能够获得理想分类效果的训练集籽粒规模,并将所获模型应用到全自动核磁分选单倍体系统中,达到实际应用的效果,为核磁共振全自动单倍体鉴别系统提供参数指导,同时也为开发基于油分花粉直感效应的玉米单倍体核磁共振全自动分选仪提供理论基础,为加快单倍体工程化育种的实现提供保障。

1 材料与方法

1.1 材料获取与选择

选择10份不同杂交组合优势群内杂交组合作为母本,杂交组合父本为中国农业大学选育的高油诱导系CAUHOI,具有R1-nj标记和茎秆深紫色性状,诱导率为3% ~ 6%。该批群内杂交组合种植于2013年春季北京(VC13),每份材料种植3行。所有参试材料均采用随机排列种植,行长2.50 m,株距0.25 m,行距0.60 m,每2.50 m行长前后设置0.50 m的过道。母本雌穗花丝抽齐后,用高油诱导系CAUHOI授粉进行单倍体诱导。实验时对单倍体诱导系杂交诱导产生的玉米穗,经干燥、脱粒后进行粗筛,去除其中的杂质及不规则籽粒,以获取所需实验样本。对于该10份杂交优势群内杂交组合,每份先通过R1-nj颜色识别方法人工肉眼挑选出60粒单倍体、60粒二倍体,为保证实验结果的可靠性,挑选颜色标记清楚、容易辨认的籽粒,对于颜色表达不清、肉眼难于判定的籽粒不予采用,另外,在单倍体挑选过程中直接舍弃败育的籽粒。

将所有按R1-nj颜色标记方法人工挑选出来的玉米籽粒依次按序存放,并进行含油率单粒测试,

籽粒含油率测定采用 minispec mq20 型(布鲁克,德国)核磁共振分析仪^[14-16]。

1.2 最小平方误差方法确定油分阈值的原理

基于玉米油分的花粉直感效应鉴别单倍体,即要利用单倍体与杂交二倍体籽粒的含油率差异,将这两类籽粒正确识别区分,实际中,对于大多数组合而言,单倍体与二倍体籽粒的含油率分布存在交叉,不能线性可分,因此在设计分类器时,应把握以下原则:搜寻最优油分阈值 t ,使分类器能够尽可能多地识别单倍体、二倍体混合籽粒中的单倍体,同时能够对混合籽粒中的二倍体具有足够的拒识能力^[17-21]。

评价分类器对单倍体与二倍体分类性能的指标有:正确识别率(Correct acceptance rate, R_{ca})、正确拒识率(Correct rejection rate, R_{cr}),即

$$R_{ca} = \frac{N_{hca}}{N_h} \times 100\% \tag{1}$$

$$R_{cr} = \frac{N_{der}}{N_d} \times 100\% \tag{2}$$

式中 N_{hca} ——正确识别的单倍体数
 N_h ——总的单倍体数
 N_{der} ——正确拒识的二倍体数
 N_d ——总的二倍体数

R_{ca} 说明分类器将单倍体从混合群体中识别出来的能力,而 R_{cr} 说明分类器将二倍体排除出单倍体这一类的能力。对于一个分类器来说,希望 R_{ca} 和 R_{cr} 越高越好,但这又是一对相互矛盾的指标,所以一个完善的分类器,应根据要求权衡考虑这两个指标,选择合适的分类阈值,避免采用单一指标评价的片面性。综合考虑 R_{ca} 及 R_{cr} ,结合工程化育种的实际需求,以二者平均值达到 90% 为宜。

具体算法实现考虑以下 2 种情况,第 1 种是单倍体和二倍体通过油分线性可分,如图 1 所示,图中圆圈表示二倍体,五角星表示单倍体,为了直观地表示单倍体及二倍体的在油分上的分布,在画图时将每粒种子的质量作为第 2 维变量。图中横坐标表示单粒种子质量,纵坐标表示玉米种子油分,该情况下,单倍体(五角星)与二倍体(圆圈)之间可通过一个分类阈值分开,即单倍体与二倍体分布没有交错部分。通过搜索距离最近的两个单倍体及二倍体油分值,假设分别为 x_h 、 x_d ,则分类阈值为

$$t = \frac{1}{2}(x_h + x_d) \tag{3}$$

通过式(3)可获得最优的分类阈值 t ,且具有一定的泛化能力^[17]。

第 2 种情形下,单倍体和二倍体通过油分线性不可分,如图 2 所示,单倍体与二倍体分布存在交错

部分。若以图中横线对应的油分作为分类阈值,图中黑色实心圆圈和五角星分别表示错分到对方区域的二倍体和单倍体,此时最小平方误差方法的目标是使分错的单倍体和二倍体到分类阈值之间的距离平方和最小,以保证分类器对训练集的分类能力,且具有一定的泛化推广能力。

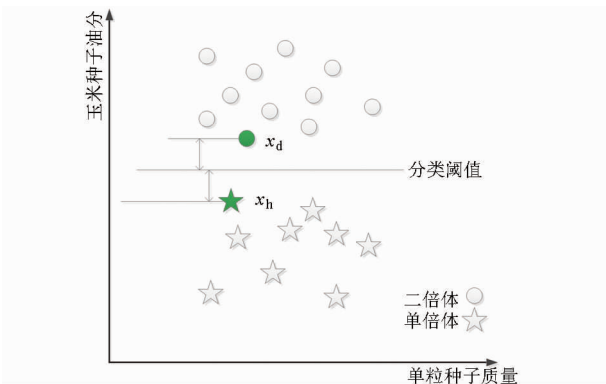


图 1 单倍体和二倍体分布图(线性可分)

Fig. 1 Haploid and diploid scatter plot (linearly separable)

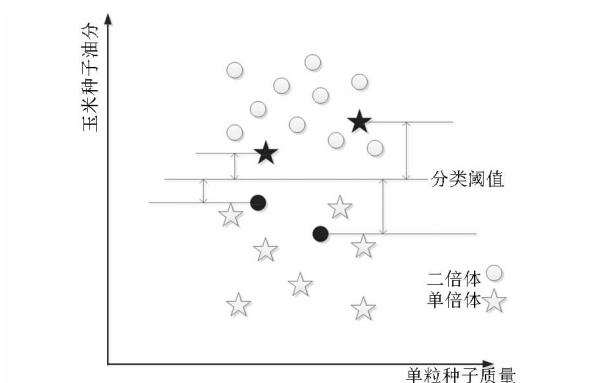


图 2 单倍体和二倍体分布图(线性不可分)

Fig. 2 Haploid and diploid scatter plot (linearly inseparable)

以某组合的单倍体、二倍体识别为例,说明基于平方误差分类方法的原理。①假设有 n 粒已知类别的单倍体及二倍体玉米籽粒,其油分分别为 $x_1, x_2, x_3, \dots, x_{n-2}, x_{n-1}, x_n$,所有玉米籽粒标明类别标签:单倍体为 1,二倍体为 2。②搜索所有某组合单倍体、二倍体籽粒油分最小值 $\min(x_i)$ 与最大值 $\max(x_i)$ 。③阈值 t 从最小值 $\min(x_i)$ 逐次增加至最大值 $\max(x_i)$,并进行迭代循环,迭代步长为 0.02,在每一步循环中,查找用当前阈值 t 对已知分类标签的单倍体及二倍体进行分类时,分错的单倍体及二倍体籽粒。④在每一步循环中,计算平方误差函数 $\text{cost}(t)$, $\text{cost}(t)$ 是一个关于分类阈值 t 的函数,用来表示被分错的籽粒油分 x_i 与油分分类阈值 t 差值的平方和,即

$$\text{cost}(t) = \sum_{i=1}^m (x_i - t)^2 \tag{4}$$

$$\text{s. t. } \min(x_i) \leq t \leq \max(x_i) \quad (i = 1, 2, \dots, m)$$

⑤搜索所有循环步骤中平方误差函数的最小值,并求出平方误差函数最小值所对应的阈值,并将该最小值所对应的油分确定为最优分类阈值。

2 实验结果与分析

选定 10 个组合的群内杂交组合 2013 年北京的数据为研究对象,并在通过颜色标记方法挑选出的单倍体和二倍体籽粒中各随机抽取 60 粒作为实验样本集。实验中使用 Kennard - Stone 算法^[22]从 60 个样本中选择 30 个作为训练集,建立预测模型,计算单倍体和二倍体籽粒含油率分类阈值,剩余的 30 个样本作为测试集,测试所选阈值对未训练样本的预测能力^[23-25]。

从训练集中每次递增 5 粒抽样,分别随机抽取 5 粒、10 粒、15 粒、20 粒、25 粒、30 粒来构成训练集,共进行 6 次抽样,直到训练集最终的样本量为 30 粒时为止,并计算不同杂交组合的油分分类阈值。所有计算程序均基于 Matlab(The Mathworks,美国)平台编写。

图 3 和图 4 分别是 10 个组合玉米籽粒的单倍体正确识别率和二倍体正确拒识率随抽样粒数变化的趋势曲线。由图 3、4 可以看出,对于铁 7922/B73 等 10 个组合的玉米籽粒,随着训练集样本所含粒数由 5 粒增加至 30 粒,所建模型对单倍体正确识别率有上升且稳定的趋势,图 4 中只能观察到 4 条曲线,因其中有 7 个组合的二倍体正确拒识率一直保持在 100%,导致该 7 个组合在图 4 中相互重叠;而铁 7922/B73 等 3 个组合二倍体的正确拒识率曲线未发生重叠,抽样粒数变化过程中识别率略有波动,但也保持在 90% 左右。当训练集抽样粒数达到 20 粒以上时,单倍体正确识别率与二倍体正确拒识率基本稳定,正确识别率均保持在 90% 以上,而正确拒识率也稳定且保持在 90% 左右或以上。由上可见,当训练集数量达到 20 粒时,对测试集的未学习过的 30 颗籽粒进行识别,能够达到一个较好的效果,除个别组合外,最小平方误差方法获得的分类阈值对未训练样本具有较好的推广能力。

表 1 为各组合平方误差函数最小值,将表 1 与图 3、4 对照可知,对于 R_{ca} 较低的 B73/8112 及 R_{cr} 较低的铁 7922/B73 及丹 3130/豫 87-1 的 3 个组合来说,其平方误差函数最小值相对于其余 7 个组合较大,说明在获得最优的分类阈值时,上述 4 个组合分错的籽粒数目较多,这种现象与各组合的单倍体及二倍体含油率分布发生交错密切相关。

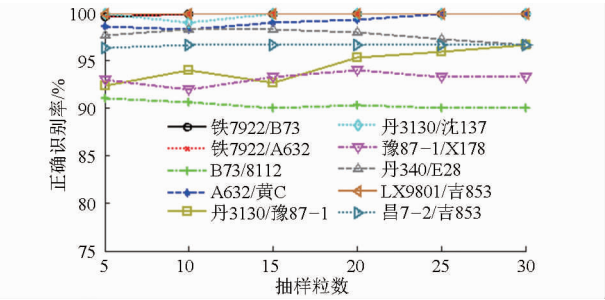


图 3 单倍体正确识别率随抽样粒数的变化曲线
Fig. 3 Changing curves of correct acceptance rate for haploid with sampled numbers

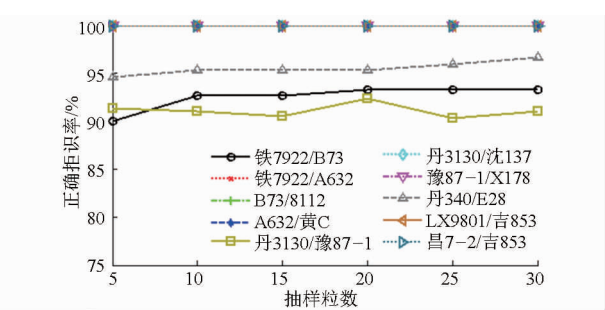


图 4 二倍体正确拒识率随抽样粒数的变化曲线
Fig. 4 Changing curves of correct rejection rate for diploid with sampled numbers

表 1 各杂交组合平方误差函数最小值(采样粒数 30 粒)
Tab.1 Minimum value of square error cost function of each variety(30 seeds sampled)

杂交组合	平方误差函数最小值
铁 7922/B73	0.10
铁 7922/A632	0.02
B73/8112	0.15
A632/黄 C	0.06
丹 3130/豫 87-1	0.20
丹 3130/沈 137	0.01
豫 87-1/X178	0.05
丹 340/E28	0.02
LX9801/吉 853	0
昌 7-2/吉 853	0.01

因篇幅所限,挑选含油率分布具有代表性的丹 3130/豫 87-1 及 LX9801/吉 853 2 个组合,并绘制其含油率分布如图 5 所示。其中丹 3130/豫 87-1 具有明显的线性不可分现象,而 LX9801/吉 853 属于线性可分的情况,丹 3130/豫 87-1 组合的单倍体及二倍体在含油率分布上出现了交错,若以单一阈值进行分类,势必会出现若干籽粒错分的情形,因此在图 4 中,丹 3130/豫 87-1 出现正确拒识率贴近 90% 的现象。而 R_{ca} 及 R_{cr} 的高低取决于含油率分布交错的程度,此时误识在所难免,分类器应着力于降低误识率,综合考虑 R_{ca} 及 R_{cr} ,若二者平均值达到 90%,则所设计分类器适用于该组合。图 5 中

LX9801/吉 853 组合单倍体及二倍体含油率分布具有一定的间隔,若按前述方法将阈值设置在间隔中点,理论上该组合 R_{ca} 及 R_{cr} 可达 100%,而图 3、图 4 中,LX9801/吉 853 的 R_{ca} 及 R_{cr} 始终保持在 100%,且表 1 中该组合最小平方误差值为零,验证了前述推论。

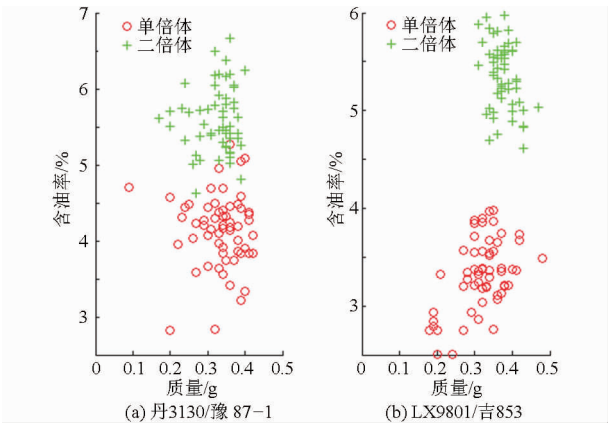


图 5 典型组合单倍体和二倍体分布图

Fig. 5 Haploid and diploid scatter plots of two typical varieties

图 6 是逐级递增训练集样本规模所获得的油分阈值随抽样粒数变化的曲线。从图 6 可以看出,各组合玉米阈值分布在 4.1% ~ 5.1% 之间。当抽样粒数增至 20 粒时以上时,油分阈值基本固定,均未发现大幅变动,说明抽样粒数大于 20 粒时得到的阈值具有一定的群体代表性,能够对未学习的 30 个样本具有稳定的识别效果,无需像贝叶斯方法^[10-13]那

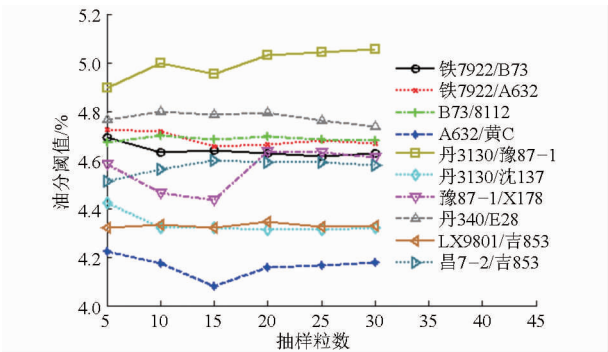


图 6 油分阈值随抽样粒数的变化曲线

Fig. 6 Changing curves of oil content threshold with sampled numbers

样采集大量的籽粒油分。且其中有 5 个组合的油分阈值集中在 4.5% ~ 4.8% 之间,而 LX9801/吉 853 与丹 3130/沈 137 组合的油分阈值比较接近且在 4.3% 左右,若对油分阈值接近的若干组合采用同一阈值,可进一步提高效率。在分选时,将依据本文所提方法获得的各组合最优分类阈值作为参数输入核磁共振全自动单倍体鉴别系统,可为后期大批量分选单倍体提供参数依据,提高分选的效率。

3 结论

(1) 采用更加广泛遗传背景的种质资源作为被诱导母本,高油诱导系 CAUHOI 作父本,获得 10 个组合的若干玉米单倍体二倍体混合籽粒,通过核磁共振分析仪测定经过类别标定的各籽粒油分。利用最小平方误差的方法,通过少量籽粒确定单倍体和二倍体的油分阈值,对单倍体和二倍体进行了两类样本的判别区分,并通过逐级递增训练集所含样本的方法,推算确定某玉米组合的单倍体识别的阈值所需训练集规模。实验表明:利用少量的单倍体二倍体籽粒构建油分鉴别单倍体模型,具有较强的泛化能力,对未训练样本也具有较好的分类效果,避免了利用贝叶斯分类方法确定玉米单倍体及二倍体油分的分类阈值时所需大量测试工作。根据实验结果可以得到推论:在生产实践中,每个玉米组合各挑选颜色标记清楚的 30 粒单倍体及 30 粒二倍体,并测定其油分,利用最小平方误差方法,即可迅速确定该组合区分单倍体和二倍体的油分阈值,并使 R_{cr} 及 R_{ca} 均值达到 90% 以上。

(2) 在玉米单倍体挑选工作中,采用模式识别中的最小平方误差方法,通过少量单倍体和二倍体建立油分鉴别单倍体的判别模型,可以快速精准地确定单倍体和二倍体分类阈值,获得理想的分类效果。该方法为核磁共振全自动单倍体鉴别系统提供参数指导;同时也为开发基于油分花粉直感效应的玉米单倍体核磁共振全自动分选仪提供理论基础。借此可以极大地减少测试工作量,加快鉴别速度,为加快单倍体工程化育种的实现提供技术保障。

参 考 文 献

1 才卓,徐国良. 玉米单倍体育种研究进展[J]. 玉米科学,2008,16(1):1-5.
CAI Zhuo, XU Guoliang. The advances in haploid breeding of maize [J]. Journal of Maize Science, 2008,16(1): 1-5. (in Chinese)

2 马骏,姜敏,刘欣芳,等. 玉米单倍体育种研究技术的探讨[J]. 东北农业大学学报,2011,42(10):135-139.
MA Jun, JIANG Min, LIU Xinfang, et al. Discussions on haploid breeding in maize [J]. Journal of Northeast Agricultural University, 2011, 42(10): 135-139. (in Chinese)

3 陈绍江,黎亮,李浩川,等. 玉米单倍体育种技术[M]. 2 版. 北京: 中国农业大学出版社, 2012.

4 陈绍江,宋同明. 利用高油分的花粉直感效应鉴别玉米单倍体[J]. 作物学报,2003,29(4):587-590.

- CHEN Shaojiang, SONG Tongming. Identification haploid with high oil xenia effect in maize [J]. *Acta Agronomica Sinica*, 2003, 29(4): 587–590. (in Chinese)
- 5 宋鹏, 吴科斌, 张俊雄, 等. 玉米单倍体籽粒特征提取及识别 [J]. *农业机械学报*, 2012, 43(3): 168–172.
SONG Peng, WU Kebin, ZHANG Junxiong, et al. Cornhaploid grain feature extraction and recognition [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2012, 43(3): 168–172. (in Chinese)
- 6 张俊雄, 武占元, 宋鹏, 等. 玉米单倍体种子胚部特征提取及动态识别方法 [J]. *农业工程学报*, 2013, 29(4): 199–203.
ZHANG Junxiong, WU Zhanyuan, SONG Peng, et al. Embryo feature extraction and dynamic recognition method for maize haploid seeds [J]. *Transactions of the CSAE*, 2013, 29(4): 199–203. (in Chinese)
- 7 韩仲志, 赵友刚, 杨锦忠. 基于籽粒 RGB 图像独立分量的玉米胚部特征检测 [J]. *农业工程学报*, 2010, 26(3): 222–226.
HAN Zhongzhi, ZHAO Yougang, YANG Jinzhong. Detection of embryo based on independent components for kernel RGB images in maize [J]. *Transactions of the CSAE*, 2010, 26(3): 222–226. (in Chinese)
- 8 李浩川, 杨继伟, 曲彦志, 等. 玉米不同基础群体孤雌生殖诱导系诱导能力评价 [J]. *玉米科学*, 2015, 23(1): 73–77.
LI Haochuan, YANG Jiwei, QU Yanzhi, et al. Evaluation for inducing parthenogenesis ability of different inducers from different basic populations in maize [J]. *Journal of Maize Sciences*, 2015, 23(1): 73–77. (in Chinese)
- 9 魏俊杰, 陈梅香, 张晓丽, 等. 玉米单倍体的诱导及其化学加倍方法 [J]. *安徽农业科学*, 2011, 39(2): 667–668.
WEI Junjie, CHEN Meixiang, ZHANG Xiaoli, et al. Induction of maize haploid and its chemical doubling method [J]. *Journal of Anhui Agricultural Sciences*, 2011, 39(2): 667–668. (in Chinese)
- 10 MELCHINGER A E, SCHIPPRACK W, WURSCHUM T, et al. Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize [J]. *Scientific Reports* 3, Article number 2129, 2013.
- 11 MELCHINGER A E, SCHIPPRACK W, FRIEDRICH UTZ H, et al. In vivo haploid induction in maize: identification of haploid seeds by their oil content [J]. *Crop Science*, 2014, 54(4): 1497–1504.
- 12 MELCHINGER A E, WINTER M, MI Xuefei, et al. Controlling misclassification rates in identification of haploid seeds from induction crosses in maize with high-oil inducers [J]. *Crop Science*, 2015, 55(3): 1076–1086.
- 13 MELCHINGER A E, SCHIPPRACK W, MI Xuefei, et al. Oil content is superior to oil mass for identification of haploid seeds in maize produced with high-oil inducers [J]. *Crop Science*, 2015, 55(1): 188–195.
- 14 DONG X, XU X, LI L, et al. Marker assisted selection and evaluation of high oil in vivo haploid inducers in maize [J]. *Molecular Breeding*, 2014, 34(3): 1147–1158.
- 15 DONG X, XU X, MIAO J, et al. Fine mapping of qhirl influencing in vivo haploid induction in maize [J]. *Theoretical and Applied Genetics*, 2013, 126(7): 1713–1720.
- 16 LIU Jin, GUO Tingting, YANG Peiqiang, et al. Development of automatic nuclear magnetic resonance screening system for haploid kernel in maize [J]. *Transactions of the CSAE*, 2012, 28(Supp. 2): 233–238.
- 17 张学工. 模式识别 [M]. 3 版. 北京: 清华大学出版社, 2010.
- 18 张帆, 李绍明, 刘哲, 等. 基于机器视觉的玉米异常果穗筛分方法 [J]. *农业机械学报*, 2015, 46(增刊): 45–49.
ZHANG Fan, LI Shaoming, LIU Zhe, et al. Screening method of abnormal corn ears based on machine vision [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(Supp.): 45–49. (in Chinese)
- 19 李鸿强, 孙红, 李民赞. 基于机器视觉的结球甘蓝形状鉴别方法 [J]. *农业机械学报*, 2015, 46(增刊): 141–146.
LI Hongqiang, SUN Hong, LI Minzan. Identification of cabbage ball shape based on machine vision [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(Supp.): 141–146. (in Chinese)
- 20 陈红, 夏青, 左婷, 等. 基于机器视觉的花菇分选技术 [J]. *农业机械学报*, 2014, 45(1): 281–287.
CHEN Hong, XIA Qing, ZUO Ting, et al. Determination of shiitake mushroom grading based on machine vision [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(1): 281–287. (in Chinese)
- 21 邓继忠, 李山, 张建瓴, 等. 小型农产品分选机设计与试验 [J]. *农业机械学报*, 2015, 46(9): 245–250.
DENG Jizhong, LI Shan, ZHANG Jianling, et al. Design and test of small-scale sorting machine for agricultural products [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(9): 245–250. (in Chinese)
- 22 严衍禄, 陈斌, 朱大洲. 近红外光谱分析的原理、技术与应用 [M]. 北京: 中国轻工业出版社, 2007.
- 23 宋鹏, 吴科斌, 张俊雄, 等. 基于计算机视觉的玉米单倍体自动分选系统 [J]. *农业机械学报*, 2010, 41(增刊): 249–252.
SONG Peng, WU Kebin, ZHANG Junxiong, et al. Sorting system of maize haploid kernels based on computervision [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(Supp.): 249–252. (in Chinese)
- 24 THEODORIDIS Sergios, KOUTROUMBAS Konstantinos. 模式识别 [M]. 北京: 电子工业出版社, 2010.
- 25 WANG Shoujue, LAI Jiangliang. Geometrical learning, descriptive geometry, and biomimetic pattern recognition [J]. *Neurocomputing*, 2005, 67: 9–28.