基因组学方法用于水稻种质资源实质派生的检测结果和应用讨论

张上都^{1,2,3†}, 袁定阳^{1†}, 路洪凤⁴, 简燕¹, 李秀欣⁴, 黄安平¹, 罗正良¹, 吕启明¹, 谭炎宁¹, 张勇飞^{1*}, 袁隆平^{1*}, 柏连阳^{1,2*}

1. 湖南省农业科学院, 长沙 410125;

2. 中南大学研究生院隆平分院, 长沙 410125;

3. 贵州省农业科学院水稻研究所, 贵阳 550006;

4. 北京诺禾致源科技股份有限公司, 天津 301700

† 同等贡献

* 联系人: E-mail: yongfeizhang@hunaas.cn; lpyuan@hhrc.ac.cn; lybai@hunaas.cn

收稿日期: 2020-03-03; 接受日期: 2020-04-23; 网络版发表日期: 2020-06-08

湖南省农业委员会资助的农业重大科研项目“低镉积累水稻新品种选育”子课题资助

摘要 我国水稻育种中存在亲本品种高度雷同的问题, 一定程度上挫伤了育种创新的积极性. 本研究探讨了利用针对动物群体遗传学开发的相同起源基因比较的特性分值(identity score, IS)方法, 对6个主流的两系法水稻杂交种雄性不育系亲本和“农垦58”常规稻种子(农垦58S天然突变的出发材料)进行基因组相似度精确定量分析. 研究表明, 比起现行的基于简单重复序列(simple sequence repeat, SSR)分子标记的水稻品种真实性鉴定的国家行业标准NYT 1433-2007, IS方法作为精准鉴定水稻实质派生品种的技术手段, 在覆盖度和准确度方面具有明显优势. 本方法鉴定出的一对实质派生不育系, 在使用SSR标记检验时, 明显具有高度的多态性, 表明SSR标记对水稻品种鉴定具有局限性, 无法实现精准区分. 同时, 本文对将IS分析作为推进水稻育种技术提高的潜在应用方向也进行了讨论.

关键词 水稻, 相似度, 实质派生品种, 特性分值, 阈值

农业部门的管理者早就注意到我国水稻育种存在遗传基础狭窄、种质资源利用水平不高、育成品种大多在同一水平上重复、原始育种创新积极性不够、企业参与育种创新动力不足等问题^[1,2]. 企业代表也认为, 目前我国植物新品种保护制度存在缺陷, 特别是缺乏限制实质性派生品种(essentially derived variety, EDV)

的保护制度. 北京某知名种业公司分析认为, 现行条例法律位阶低, 不利于品种权全面、有效保护, 不适应国际形势, 缺乏保护措施和执法依据, 建议将条例调整为法, 以助于防止发达国家控制发展中国家的农业遗传资源, 遏制国外育种家对国内育种家的侵权^[3].

2019年2月1日, 农业农村部发布通知, 启动了《中

引用格式: 张上都, 袁定阳, 路洪凤, 等. 基因组学方法用于水稻种质资源实质派生的检测结果和应用讨论. 中国科学: 生命科学, 2020, 50
Zhang S D, Yuan D Y, Lu H F, et al. The results of rice germplasm EDV test by genomic analysis and related discussions (in Chinese). Sci Sin
Vitae, 2020, 50, doi: [10.1360/SSV-2020-0068](https://doi.org/10.1360/SSV-2020-0068)

华人民共和国植物新品种保护条例》修订工作, 明确提出要建立实质性派生品种制度, 限制修饰性育种的商业行为, 征求意见稿鼓励原始创新, 正式面向全国征集对《条例修订草案》和植物新品种的实质派生特性鉴定的反馈意见。

此决定对于我国种业无疑是一件大事, 定会推动重构水稻育种和水稻种业的版图。我国现阶段玉米、马铃薯和花卉的大多数推广品种都含有来自国外种质的血缘, 而根据实质性派生品种制度的要求, 我国在引入国外优良育种资源时要购买相应专利或品种使用权。对于玉米等作物的育种创新无疑会增加运营成本和难度。

显然, 不推进派生品种的管理对育种和作物种植业的伤害非常严重。以水稻为例, 唐力等人^[4]分析了原始品种遗传资源对中国水稻生产的经济影响。结果表明, 1999~2009年中国原始品种遗传资源对水稻生产平均遗传贡献仅为0.25, 原始品种遗传单一性并未呈逐渐增强趋势, 并且原始品种遗传资源对水稻改良的遗传贡献每增加1%, 水稻单产将提高1.6%。因此本文呼吁重视原始创新与有效运用。

鉴于我国的杂交水稻在国际种业处于技术和资源输出的有利位置, 如何提升技术手段, 防止和阻退技术进口国的EDV, 也是我国必须要从战略层面重视的问题。同时, 也必须认识到, 限制水稻EDV可能会暂时影响目前经营利润主要来自生产和出售EDV品种的利益团体, 但是却更加有利地提振我国的水稻育种总体创新能力和技术含量, 有利于将我国占据优势的水稻种业做大做强。因此, 本文认为, 国家管理部门可以对不同的作物区别对待, 以水稻作为试点, 先行开展水稻EDV鉴定工作。

实质性派生品种EDV, 也叫作实质派生品种。国际种子联盟(International Seed Federation, ISF)的1991年“国际植物新品种保护联盟(International Union for the Protection of New Varieties of Plants, UPOV)”公约版本第14条规定: 当一个品种符合下面三个条件, 即可认定为EDV。这三个条件是: (i) 该品种主要是从一个原始品种派生出来, 或者从一个本身也是派生品种的品种派生而来; (ii) 该品种与原始品种具有可以分辨的某些特性; (iii) 除了这些刻意获得的派生过程中的区别, 派生品种与原始品种其余特性的基因型和各种性状基因型的组合本质上是一致的。UPOV1991的另一个新

规定就是实质性派生品种保护规则, 即“EDV保护规则”, 强调对原始育种人的利益保护, 从而阻断或减少植物育种创新中剽窃、投机取巧培育EDV的行为^[5,6], 但是实质上并没有带来遗传增益(genetic gain)的年度累积增加等。该方向文献丰富, 但限于篇幅本文只列中国和国外的该方向的两篇综述^[7,8]。

对于我国水稻育种而言, EDV是一个全新的概念。UPOV的定义多关注于法律文字方面的表述, 而关键的细节, 如不同作物判定是否为EDV的阈值标准是什么? 用什么样的技术作EDV鉴定等都没有明确界定。国际种业界在具体操作方面面临着商讨和制定不同作物的阈值标准时, 大公司要求标准从严(比如玉米EDV的鉴定阈值被建议设定为高于或等于90%的基因组相似度就可判定为EDV^[9]), 小公司要求标准放宽的争议。而且具体到鉴定技术上, 不同作物使用的技术不同, 从简单重复序列(simple sequence repeat, SSR)、随机扩增多态性DNA(random amplified polymorphic DNA, RAPD)到基因芯片等都在使用。目前在鉴定技术方面还不是很完善^[10]。

我国有近7万份水稻种质资源^[11], 由于缺少准确的品种相似度定量手段, 材料的独特性仍缺乏分子和基因组学方面研究, 研究利用深度有待加强。本文利用6个水稻两系法不育系和1个作为本研究方法质量判断内控参考的、已知实质派生品种来源的原始品种作为实验材料, 针对水稻品种EDV的精确鉴定技术作了创新性质的探讨。本文的研究表明, 基于动物遗传开发的特性分值(identity scores, IS)分析, 利用全基因组单核苷酸多态性SNP(single nucleotide polymorphism)作为比对水稻品种间相似度的思路是可行的, 比现行水稻品种的真实性鉴定的国家标准具有更加广泛的基因组覆盖度, 而且没有SSR倾向于非基因编码区的特点。本文还对与之相关的判定涉嫌的水稻是否属于EDV的基因组相似度阈值提出了建议标准。

1 实验材料与方法

本研究7个样品是农垦58(NK58)、农垦58S(NK58S)、淮S(ZHS)、安农S-1(ANS-1)、Y58S(Y58S)、培矮64S(PA64S)和深08S(SH08S)。材料中的农垦58S, 淮S, 安农S-1, Y58S, 培矮64S, 深08S六份资源是两系法杂交水稻育种过程中比较主流的雄性不育

系. 其中, “农垦58S”是世界上第一个粳型光温敏不育系; “安农S-1”是湖南杂交水稻研究中心选育的世界上第一个籼稻光温敏不育系; “培矮64S”“Y58S”和“准S”由湖南省农业科学院选育, 并且在水稻两系法杂交种生产中发挥重要作用; “深08S”由国家杂交水稻工程技术研究中心清华深圳龙岗研究所选育. “农垦58”常规稻是作为“农垦58S”天然突变来源的、从日本引进的原初供体品种.

1.1 文库构建、测序和质控

各品种种子在常温下催芽, 待芽长5 cm左右时, 剪取叶片组织提取DNA, 分别构建350bp, 2K, 5K文库, 进行Illumina PE150(天津诺禾致源科技有限公司, 天津)测序, 测序深度为>100X, 然后进行基因组组装和变异检测. 选取1.5 μ g的DNA使用Covaris超声波破碎仪(Woburn Massachusetts, 美国)随机打断成长度为350 bp的片段. 采用TruSeq Library Construction Kit(Illumina, 美国)建库, DNA片段经末端修复、加polyA尾、加测序接头、纯化、PCR扩增等步骤完成整个文库制备, 文库合格之后使用Illumina HiSeq进行PE150测序. 检测合格的DNA样品通过Covaris随机打断成片段, 经末端修复、加polyA、加罗氏环化接头、环化、二次打断、末端修复、加polyA、加测序接头、纯化、PCR扩增等步骤完成整个大片段文库(2K, 5K, 10K)的制备, 构建好的文库通过Illumina HiSeq PE150测序.

测序得到的原始测序序列称为“sequenced reads”或者“raw reads”, 里面包含接头(adapter)和低质量的reads. 为了保证信息分析的准确性, 需要对raw reads进行过滤进而得到clean reads: 去掉带接头的paired-end reads, 过滤掉单端测序read中N含量超过该条read长度比例10%的数据, 去掉单端测序read中含有的低质量($Q \leq 5$)碱基数超过该条read长度比例50%的paired-end reads.

1.2 基因组组装

采用SOAPdenovo^[12]软件对7个水稻材料进行拼接, 拼接基本过程如下: (i) 将DNA片段打断成17 bp的*k*-mer序列; (ii) 利用*k*-mer之间的overlap关系构建*de Bruijn*图; (iii) 对*de Bruijn*图进行化简: 对小分枝进行修剪, 去掉低覆盖的连接关系, 解重复结构, 合并杂合泡状结构; (iv) 在重复区域边界位置进行剪切, 得到

contig序列; (v) 根据大片段数据的pair-end关系, 构建scaffold序列; (vi) 用reads对scaffold的gap区域进行填补.

1.3 变异检测单核苷酸多态性的检测

利用LASTZ^[13]软件将组装好的7个水稻基因组分别与珍汕97的基因组序列进行全基因组比对, 获得相对于珍汕97的比对信息, 鉴定出每个材料相对于珍汕97的变异信息(SNP, 插入缺失(insertion-deletion, In-Del), 结构变异(structural variation, SV), 拷贝数变异(copy number variation, CNV)). 此外, 将每个水稻的小片段文库的测序序列分别比对到珍汕97的基因组序列上, 验证这个位点与使用组装序列的方法检测到的是否一致; 对于reads没有比对到的区域, 保留通过组装基因组比对检测到的位点.

1.4 进化树的构建

使用7个材料的SNP信息, 利用邻接法(neighbor-joining methods)进行系统进化树的构建. 两个个体*i*和*j*之间的*p*-距离通过公式(1)计算^[14]:

$$D_{ij} = \frac{1}{L} \sum_{l=1}^L d_{ij}^{(l)} \quad (1)$$

公式中*L*为高质量SNPs区域长度, 在位置*l*的等位基因为A/C, 那么有(2)中的4个情况:

$$d_{ij}^{(l)} = \begin{cases} 0, & \text{如果两个个体的基因型是AA和AA} \\ 0.5, & \text{如果两个个体的基因型是AA和AC} \\ 0.5, & \text{如果两个个体的基因型是AC和AC} \\ 1, & \text{如果两个个体的基因型是AA和CC} \end{cases} \quad (2)$$

运用Treebest-1.9.2(<http://treesoft.sourceforge.net/treebest.shtml>)软件计算距离矩阵. 引导值(bootstrap values)经过达1000次计算获得, 标记在节点前的分枝下方, 用来检验分枝的可靠性, 该值越大表明该节点的置信度越高.

1.5 主成分分析

利用7个材料的SNP信息将这7个个体按主成分聚类成不同的亚群. 主要原理是: 个体*i*, *k*位置的SNP用[0,1,2]表示, 若个体*i*与参考等位基因是纯合, 则=0; 若是杂合, 则=1; 若个体*i*与非参考等位基因是纯合,

则=2.

公式(3)中 $E(d_{ik})$ 是 $d_k d_k$ 的平均值, 个体样本协方差 $n \times n$ 矩阵通过 $X=MMT/S$ 进行计算. 本研究使用GCTA软件进行主成分分析(principal component analysis, PCA)^[15].

$$d_{ik}' = (d_{ik} - E(d_k)) / \sqrt{E(d_k) \times (1 - E(d_k)/2) / 2} \quad (3)$$

1.6 特性分值分析

特性分值, 简称IS分析, 是一种计算不同样品间序列一致性程度的分析方法, 可用于辅助佐证群体结构分析结果. 首先根据群体基因型文件进行打分, 将基因型文件转化成由0, 0.5和1三种遗传距离参数值组成的Dsi矩阵(纯合且与参考基因组相同者Dsi为0, 纯合且不同于参考基因组的Dsi为1, 杂合的Dsi均为0.5); 然后沿染色体按照每个窗口20 kb来计算每个窗口的IS值^[16~18]. IS值越大说明两个个体间差异越小.

为了保证7个水稻材料IS比较的合理性, 本研究选择珍汕97作为统一的参考基因组, 进行比较的每个材料的SNP位点相对于珍汕97都是一样的.

S_i 是在位点 i 与珍汕97相似的位点数, n 是20 kb的窗口中的SNP总数, n' 是在20 kb的窗口中缺失的位点数(公式(4)).

$$IS = \frac{\sum_{i=1}^n S_i}{2(n - n')} \quad (4)$$

1.7 structure分析

利用7个材料的SNP信息进行structure分析, 分析

软件是frappe. 使用最大似然法(maximum likelihood)评估每个样品的混合程度^[19].

1.8 SSR标记分析

将已知的SSR序列与组装好的基因组序列进行blast比对, 鉴定本研究先前发现的两个EDV品种中SSR的多样性. 使用组装好的基因组进行比较, 可以看到更多的变异信息, 进而能够更好地分析SV, CNV, SSR; 也可以看出IS方法和SSR方法各自的特点, 找出潜在的问题.

2 结果

2.1 二代测序结果

通过Illumina平台测序以及数据质控, ZHS, ANS-1, SH08S, NK58四个水稻品种最终获得35~54 G(测序深度为103~159X)高质量的测序数据, 包括350bp, 2K和5K三种插入片段文库.

Y58S, PA64S以及NK58S三个品种在上述三种文库基础上增加了10 kb插入片段文库, 最终获得数据量是73~78 G(测序深度为184~228X). 所有样品的测序数据汇总结果如表1.

2.2 7个水稻的组装结果

利用上述二代测序序列, 经过SOAPdenovo的组装, 7个水稻的基因组的contig N50在8 kb以上, scaffold N50在121 kb以上.

7个水稻的组装结果如表2, 最终获得的基因组总

表 1 7水稻样品的测序数据统计

Table 1 Sequence data of the 7 rice samples

水稻样品	文库类型				
	350bp (bp)	2K (bp)	5K (bp)	10K (bp)	总大小(bp)
ZHS	41525316600	5646216900	7720997400	—	54892530900
ANS-1	33923288700	7560801300	6782547300	—	48266637300
Y58S	25002489300	16293057300	13813428000	17944463100	73053437700
PA64S	25506950700	15006511200	18790728900	17346974700	76651165500
SH08S	25355532300	5944885200	5504082300	—	36804499800
NK58	25043776280	5629485300	4758296400	—	35431557980
NK58S	24791503440	16923415800	16558440300	20566385100	78839744640

表 2 7个水稻样品的基因组组装结果统计^{a)}
Table 2 Genome assembly data of the 7 rice samples^{a)}

分类		总长	数量	最大值	N50	N90
ZHS	Contig (bp)	312792676	162295	70553	8759	1048
	Scaffold (bp)	344579551	91022	914062	128488	10399
ANS-1	Contig (bp)	312268116	159616	74536	8915	1074
	Scaffold (bp)	350511577	85910	1500949	147501	17971
SH08S	Contig (bp)	313638622	157620	74281	8397	1102
	Scaffold (bp)	349219219	83625	800791	121061	15468
Y58S	Contig (bp)	367797861	103795	242160	34960	4237
	Scaffold (bp)	395255494	71852	8373176	1114666	83111
PA64S	Contig (bp)	363713533	87880	244386	40019	5923
	Scaffold (bp)	378387326	66337	4435574	1069516	142391
NK58	Contig (bp)	310218418	147528	77726	8718	1232
	Scaffold (bp)	343495141	74199	1306941	140282	20110
NK58S	Contig (bp)	367719942	83455	2789542	37537	6220
	Scaffold (bp)	383039362	61593	5335554	1342198	148657

a): N50: Reads拼接后会获得一些不同长度的Contigs. 将所有的Contig长度相加, 能获得一个Contig总长度. 然后将所有的Contig按照从长到短进行排序, 当相加的长度达到Contig总长度的50%时, 最后一个加上的Contig长度即为Contig N50. N90: Reads拼接后会获得一些不同长度的Contigs. 将所有的Contig长度相加, 能获得一个Contig总长度. 然后将所有的Contigs按照从长到短进行排序, 当相加的长度达到Contig总长度的90%时, 最后一个加上的Contig长度即为Contig N90

长在343.4~395.2 Mb之间. 其中ZHS, ANS-1, SH08S, NK58品种基因组Contig N50为8.1~8.9 kb, 增加了10K文库的Y58S, PA64S和NK58S三个品种基因组的contig N50在34.9~40.0 kb.

2.3 变异检测结果

组装后, 以籼稻珍汕97为参考基因组对本研究7个品种的SNP, InDel和SV进行检测. 结果见表3. ZHS与ANS-1两个品种在多种变异类型的变异个数相近, 无明显差异, 推测两者可能发生了相同程度的变异; NK58与NK58S发生变异最多, 但两者变异类型不同, NK58发生更多的CNV, 而NK58S发生更多的InDel和SV; Y58S, PA64S与NK58S三个品种发生的缺失和插入类SV变异多于其他四个品种.

2.4 进化树构建结果

SNP检测之后, 得到的个体SNPs首先用于计算种群之间的遗传距离. 运用Treebest-1.9.2软件计算距离矩阵, 以此为基础, 通过邻接法构建系统进化树. 图1是本研究中7个品种在进化树上的聚类情况.

由进化树的分析结果可知, ZHS和ANS-1之间的亲缘关系近, NK58和NK58S之间的亲缘关系近, Y58S和SH08S之间的亲缘关系近; NK58, NK58S和群体材料的亲缘关系较远, 这两个材料从它们的起始种开始发生了不同程度的变异; ZHS和ANS-1从它们的起始种开始发生了类似程度的变异, 但变异程度小于NK58, NK58S与其起始种的变异.

这个判断与籼稻和粳稻较早分枝的情形是吻合的. 与Y58S相比, SH08S相对于它们的起始种发生的变异多. ZHS与ANS-1之间的相似程度超过了NK58和其衍生的自然突变系NK58S, 这两组材料之间的相似程度都高于Y58S和SH08S这一组材料; PA64S材料和其他材料的相似度较低. 碱基替换数量值上体现的是SNP水平的差异度, 并非真实基因组的差异度, 但是值的大小关系可以表示样本差异程度. 通过对比可以看出, NK58和NK58S之间的差异度为2%~3%, 而ZHS与ANS-1的差异度更是低到0.2%左右. 因为进化树是基于每个分枝的差异程度而构建的, 至此可得到暗示, 似乎在6个两系不育系样品中可能存在实质派生现象.

表 3 7个水稻品种的变异检测结果

Table 3 Genomic variation test results of the 7 rice samples

类型		ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
SNP		2128164	2126710	2512963	2414327	2767264	3736203	3548060
InDel		122079	121136	159730	379527	440434	261512	595336
SV	缺失	0	0	1	1075	1334	3	1629
	插入	0	0	2	1287	1427	6	1866
	反转	874	828	1206	823	1188	1452	1054
CNV	缺失	2010	2049	2527	2412	2540	5131	3569
	重复	926	921	995	1074	1157	1340	1034

图 1 7个水稻材料的进化树

Figure 1 Phylogenetic tree of the 7 rice samples

2.5 主成分分析结果

为了进一步明确7个品种在遗传背景方面的近似程度, 本团队又进行了PCA, 可非常明显地看出, PCA也得到类似进化树提供的结论. 结果如图2, 6个不育系和一个突变原初供体材料农垦58可以分为4个聚类, 分别是NK58及其派生系NK58S聚类, ANS-1和ZHS聚类, SH08S和Y58S聚类, 以及PA64S聚类. 从PC2和PC3这个维度来看, 7个被研究对象可以被分成4个类群, 这点可以通过与图1的比较得以印证, PA64S是独立的一个聚类, 这也基本与育种材料的历史吻合.

目前就下结论说PCA可以将育种材料按照遗传差异划分成不同的类群, 并且可以据此指导不育系与恢复系的配组还为时过早. 本团队计划在收集到100个以上的全国稻作区两系不育系时, 将开展延续研究追踪这一现象.

2.6 特性分值分析结果

用IS方法比较7个作物材料的基因组相似度尚未见报道. IS分析是针对动物群体遗传学研究开发的方法, 是根据个体基因型差异计算两两个体间的基因序列一致性, 进而体现群体间基因序列相似度的一种分析方法^[14]. 在对7个品种进行较为粗犷的热图比较时, 本团队发现, 农垦58和农垦58S存在很高的相似度, 因为农垦58S是石明松先生利用农垦58这个从日本引进品种的田间自然突变株作为回交父本, 于1973年在单季晚稻成熟时, 根据不育株结粒少、不低头、老来青等特点, 在晚粳稻农垦58的大田中选择3株自然不育株, 次年种成的株行. 其中有1个株行形态上与农垦58十分相似, 整齐一致. 从1974年开始, 在分离出来的可育株中选择单株形成选系, 作为连续回交的父本使用. 1976~1979年间共回交了4次^[20]. 本研究获得的农垦58

图2 7个水稻材料的主成分分析

Figure 2 Principle component analysis results of the 7 rice samples

的种子则不知道是经历了多少年的扩繁，具体的扩繁次数已经无法确定。更令人吃惊的是，准S和安农S-1之间有着比自然突变的农垦58S和变异材料更高的相似度(表4)。结果表明，ZHS和ANS-1之间的SNP相似度达到了99.74%，而已知的派生品种NK58S与出发品种之间的相似度是96.23%。据此可判断，ZHS晚于ANS-1多年才被育成，因而ZHS是ANS-1的派生品种。具体的阈值判断，在本文的讨论部分展开。

NK58和NK58S为两个粳稻的材料，本团队使用日本晴(*Oryza Sativa* L. Ssp. *Japonica*)作为参考基因组，重新进行了SNP检测和IS计算，计算结果表明，NK58与NK58S之间的相似性低于ZHS和ANS-1之间的相似度(表5)，同使用珍汕97作为参考基因组的结果差比顺序和量值范围完全一致(表4)。

除全基因组IS值与使用日本晴作为参考基因组进行比对外，本团队又统计了7个材料SNP的相似程度发现，SNP相似程度的结果与全基因组IS值的结果相似。SNP的结果如表6，ZHS和ANS-1之间的SNP相似度达到了99.48%，而已知的派生品种NK58S与出发品种之间的相似度是94.95%。

本团队为了更全面地比较这7个材料的差异，对每个材料进行了超过100X的基因组重测序和组装。大规模EDV鉴定时可使用重测序的方法。在当前的数据中每个样品随机抽取5X, 10X, 15X和20X的数据，以珍汕97作为参考基因组进行SNP的检测和IS的计算，结果表明，检测趋势同使用组装检测方法的结果一致。不同深度下的IS结果详见表7。

在IS热图比较中，用不同颜色代表不同IS值，IS值越大颜色越红，品种间的相似性越近；IS值越小颜色越浅，品种间的差异越大。由图3可以看出，ZHS与ANS-1间的亲缘关系最近，NK58与NK58S亲缘关系最近，Y58S和NK58、NK58S亲缘关系较远，ZHS和ANS-1与NK58和NK58S关系最远，PA64S、SH08S和其他样本之间没有很明显的亲疏关系。

当比较农垦58与农垦58S的SNP时，可获得结果如图4。虽然目前无法判断农垦58S和农垦58之间的基因组差别是经过多少代积累起来的，但总体而言，全基因组SNP相似度呈现这么多的变异，显然高于人们的想象。本团队判断，水稻基因组中转座子因子起到了主要作用，因为石明松老师选育农垦58S的过程主要

表 4 7个样品使用珍汕97作为参考基因组的IS相似性评估结果

Table 4 IS comparison of 7 samples based on zhenshan97 as the reference genome

水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	99.74%	70.48%	72.54%	64.71%	33.55%	33.45%
ANS-1	99.74%	100.00%	70.48%	72.54%	64.71%	33.55%	33.45%
SH08S	70.48%	70.48%	100.00%	86.74%	78.06%	43.46%	43.82%
Y58S	72.54%	72.54%	86.74%	100.00%	79.66%	44.85%	44.72%
PA64S	64.71%	64.71%	78.06%	79.66%	100.00%	51.93%	52.07%
NK58	33.55%	33.55%	43.46%	44.85%	51.93%	100.00%	96.23%
NK58S	33.45%	33.45%	43.82%	44.72%	52.07%	96.23%	100.00%

表 5 使用日本晴为参考基因组评估的IS结果

Table 5 IS comparison based on Nipponbare as the reference genome

水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	99.74%	69.18%	60.52%	66.86%	25.23%	25.19%
ANS-1	99.74%	100.00%	69.18%	60.52%	66.86%	25.22%	25.18%
SH08S	69.18%	69.18%	100.00%	76.98%	84.90%	37.86%	37.66%
Y58S	60.52%	60.52%	76.98%	100.00%	75.39%	45.78%	45.92%
PA64S	66.86%	66.86%	84.90%	75.39%	100.00%	36.35%	36.70%
NK58	25.23%	25.22%	37.86%	45.78%	36.35%	100.00%	96.23%
NK58S	25.19%	25.18%	37.66%	45.92%	36.70%	96.23%	100.00%

表 6 7个样品的SNP相似性评估结果

Table 6 Similarity evaluation of 7 samples based on SNP comparison

水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	99.48%	69.12%	60.40%	66.78%	26.43%	26.41%
ANS-1	99.48%	100.00%	69.13%	60.40%	66.79%	26.43%	26.41%
SH08S	69.12%	69.13%	100.00%	76.14%	83.96%	38.15%	38.04%
Y58S	60.40%	60.40%	76.14%	100.00%	74.80%	46.84%	47.07%
PA64S	66.78%	66.79%	83.96%	74.80%	100.00%	37.30%	37.78%
NK58	26.43%	26.43%	38.15%	46.84%	37.30%	100.00%	94.95%
NK58S	26.41%	26.41%	38.04%	47.07%	37.78%	94.95%	100.00%

是基于同一起来源姊妹系的杂交后进行株选, 然后和出发材料连续回交^[20]; 而且水稻亲本在迭代繁殖的过程中所积累的变异, 可能根据所经历的环境差异而表现出巨大的差异. 本团队在马铃薯基因组也发现类似的高自主变异的现象(未发表). 这一现象表明, 作物在迭代生长过程中基因组的塑性(plasticity)可能高于一般性认为的百万分之一, 因为转座因子参与和贡献了大

量的变异. 在后续的研究中, 本团队将验证此判断是否正确.

而当比较其余的不育系时, 本团队发现, 有一对不育系之间的差别比农垦58和农垦58S之间的相似度差异还要小. 它们是准S和安农S-1, 两者的相似度达到了99.74%(表4, 图5). 显然, 准S和安农S-1如果不是育种材料中的姊妹系, 就只可能是实质派生品种. 由于我国

表 7 不同重测序深度下的IS的结果

Table 7 IS comparison based on different sequencing depths of the 7 samples

每个样品随机抽取5X左右的数据							
水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	98.32%	95.20%	93.41%	94.13%	89.02%	88.85%
ANS-1	98.32%	100.00%	95.18%	93.39%	94.13%	88.97%	88.81%
SH08S	95.20%	95.18%	100.00%	94.91%	95.67%	90.30%	90.00%
Y58S	93.41%	93.39%	94.91%	100.00%	94.69%	91.97%	91.72%
PA64S	94.13%	94.13%	95.67%	94.69%	100.00%	90.46%	90.16%
NK58	89.02%	88.97%	90.30%	91.97%	90.46%	100.00%	97.55%
NK58S	88.85%	88.81%	90.00%	91.72%	90.16%	97.55%	100.00%
每个样品随机抽取10X左右的数据							
水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	98.23%	87.60%	82.49%	84.84%	70.51%	70.49%
ANS-1	98.23%	100.00%	87.61%	82.47%	84.87%	70.50%	70.47%
SH08S	87.60%	87.61%	100.00%	88.35%	90.85%	75.26%	75.09%
Y58S	82.49%	82.47%	88.35%	100.00%	87.95%	80.06%	80.06%
PA64S	84.84%	84.87%	90.85%	87.95%	100.00%	75.50%	75.56%
NK58	70.51%	70.50%	75.26%	80.06%	75.50%	100.00%	96.70%
NK58S	70.49%	70.47%	75.09%	80.06%	75.56%	96.70%	100.00%
每个样品随机抽取15X左右的数据							
水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	99.35%	76.50%	70.00%	74.70%	42.46%	42.42%
ANS-1	99.35%	100.00%	76.51%	70.01%	74.72%	42.45%	42.41%
SH08S	76.50%	76.51%	100.00%	82.70%	88.79%	52.09%	51.97%
Y58S	70.00%	70.01%	82.70%	100.00%	80.58%	57.55%	57.61%
PA64S	74.70%	74.72%	88.79%	80.58%	100.00%	50.09%	50.27%
NK58	42.46%	42.45%	52.09%	57.55%	50.09%	100.00%	96.55%
NK58S	42.42%	42.41%	51.97%	57.61%	50.27%	96.55%	100.00%
每个样品随机抽取20X左右的数据							
水稻样品	ZHS	ANS-1	SH08S	Y58S	PA64S	NK58	NK58S
ZHS	100.00%	99.66%	74.03%	66.92%	72.66%	36.47%	36.41%
ANS-1	99.66%	100.00%	74.03%	66.92%	72.66%	36.47%	36.41%
SH08S	74.03%	74.03%	100.00%	80.50%	87.79%	46.43%	46.28%
Y58S	66.92%	66.92%	80.50%	100.00%	79.00%	53.08%	53.16%
PA64S	72.66%	72.66%	87.79%	79.00%	100.00%	44.53%	44.78%
NK58	36.47%	36.47%	46.43%	53.08%	44.53%	100.00%	96.66%
NK58S	36.41%	36.41%	46.28%	53.16%	44.78%	96.66%	100.00%

的水稻育种还没有严格执行育种者亲本种子缴纳和专门储藏库保存原始亲本材料制度，所以不可能像发达国家的育种体系一样，遇到有疑义的问题可以从资源

库调用亲本种子作为分析验证材料。而且，也不排除可能的操作者失误导致的一次混杂或者标签错误放置，进而导致错误被一直传承。

图3 7个水稻雄性不育系基因组比较的热图

Figure 3 A heatmap showing genomics comparison of the 7 male-sterile rice lines

为了获得和验证到底需要使用多少SNP, 才能以尽可能低的成本获得可靠的水稻品种间相似度比较结果. 本团队也利用在基因组均匀分布的不同密度的SNP数量进行IS分析的比较. 本团队将全基因组分别用100, 500, 1000, 10000, 800000个SNP进行IS分析(当然不是在每条染色体上都能找到理想的SNP绝对平均分布以达到上述密度分布的要求, 只是尽量靠近即可). 结果见表8, 随着标记的增加, 样品之间的差异程度更接近于全基因组水平差异程度评估的结果.

可以看出, 当使用平均每条染色体接近10个SNP标记来检测相似度时, 无法区分ZHS和ANS-1. 使用均匀分布的500~1000个SNP进行全基因组相似度分析, 可以获得类似全基因组SNP比较分析的结果. 这对使用24对(或者48对)SSR标记这样稀疏的标记密度来评估品种的差异度是否合乎逻辑, 提出了合理的质疑.

籼稻和粳稻的相似度差异很大, 达到了近40%的SNP差异程度, 这一数据同样支持籼稻和粳稻在很早就分化的、从进化树推论出来的表述.

2.7 structure分析结果

本团队也对7个品种的遗传结构用structure软件进行了分析, 得到图6的结果. 结果表明, 即使在总共只有7个品种的群体中将结构分群的K值上调到4, 两对高度相似的品种仍然没有改变其聚类特征.

结果还显示, 把群体结构分群的K值下调为3的时候, 两个疑是派生品种的特点仍未改变. 当进一步压缩K值到2的时候才会区分为籼稻和粳稻群. 非常有趣的是, 群体结构structure分析可以看出, PA64S这个品种明显带有20%左右的粳稻血缘, 这刚好与培矮64的选育过程吻合^[21].

图4 农垦58与农垦58S的染色体逐一平行比对结果

Figure 4 Comparison results of NK58 and NK58S based on comparing parallel chromosomes site by site

图 5 准S和安农S-1染色体逐一平行比对结果
Figure 5 Comparison results of ZHS and ANS-1 based on comparing parallel chromosomes site by site

图 6 7个水稻的structure结果
Figure 6 Structure analysis results of the 7 rice samples

2.8 SSR标记分析结果

本团队随后对鉴定出的一对EDV品种进行了农业行业标准NYT 1433-2014的水稻品种鉴定技术规程的模拟分析，将每个SSR标记的引物序列定位到各个品种组装获得的基因组上，然后统计有多少对引物匹配，统计两侧引物之间的DNA长度。结果不出意料地看到了本研究鉴定出的一对EDV显现出SSR多态性，而且是非常显著的多态性(表9)。虽然认为这个结果100%准确还过于武断——因为本研究的测序深度还远远达不到*de novo*组装的测序深度的基本要求，所以初步组

装的基因组覆盖度并非完美——但是结果可靠(表9)。

3 讨论

实质派生品种鉴定必须考虑以何种大小相似度作为判定标准。阈值必须由有权威单位和德高望重的育种家在经过充分征求意见和考虑各方的合理利益分配之后提出，由管理部门通过行政审议的方式正式确定下来。

另外一个必须考虑的事是已有的实质派生品种如何处置？本文建议采用国外的授权制度，也就是实质派生品种可以继续使用，但是必须获得原始品种拥有单位或个人的授权，并且利益共享。本文通讯作者之一在美国学习和工作时了解到国外的(美国)做法如下：基于法律的要求，有派生品种的公司或者单位会与原始品种拥有者签订授权使用合同，并且每年向原始品种拥有者支付所获取利益的25%，或者每销售一斤种子向原生品种拥有权益者支付相当于人民币0.3~0.5元的授权费。如果作弊被查实，将面对法律严惩。

也可以参考湖南杂交水稻研究中心的做法，课题组间开放交流使用育种亲本，但是当被审定杂交水稻品种被转让产生效益时，转让获利的课题组必须按照组配杂交种的亲本不同特点，与亲本原初育种课题组

表 8 沿全基因组采用不同密度的SNP数量进行IS分析的结果

Table 8 IS analysis results with different SNP density evenly distributed along the genome

116 SNPs							
水稻样品	ZHS	ANS-1	Y58S	PA64S	SH08S	NK58	NK58S
ZHS	100.00%	100.00%	75.62%	65.70%	72.73%	33.47%	34.30%
ANS-1	100.00%	100.00%	75.62%	65.70%	72.73%	33.47%	34.30%
Y58S	75.62%	75.62%	100.00%	80.17%	88.02%	42.15%	42.98%
PA64S	65.70%	65.70%	80.17%	100.00%	78.10%	54.55%	55.37%
SH08S	72.73%	72.73%	88.02%	78.10%	100.00%	43.39%	44.21%
NK58	33.47%	33.47%	42.15%	54.55%	43.39%	100.00%	93.39%
NK58S	34.30%	34.30%	42.98%	55.37%	44.21%	93.39%	100.00%
515 SNPs							
水稻样品	ZHS	ANS-1	Y58S	PA64S	SH08S	NK58	NK58S
ZHS	100.00%	99.62%	70.79%	62.45%	68.68%	28.35%	28.64%
ANS-1	99.62%	100.00%	70.79%	62.84%	69.06%	28.35%	28.64%
Y58S	70.79%	70.79%	100.00%	81.71%	86.78%	42.43%	43.10%
PA64S	62.45%	62.84%	81.71%	100.00%	78.83%	49.04%	50.29%
SH08S	68.68%	69.06%	86.78%	78.83%	100.00%	41.28%	42.15%
NK58	28.35%	28.35%	42.43%	49.04%	41.28%	100.00%	97.03%
NK58S	28.64%	28.64%	43.10%	50.29%	42.15%	97.03%	100.00%
1193 SNPs							
水稻样品	ZHS	ANS-1	Y58S	PA64S	SH08S	NK58	NK58S
ZHS	100.00%	99.71%	73.04%	64.13%	72.04%	30.83%	30.92%
ANS-1	99.71%	100.00%	72.83%	64.17%	71.92%	30.79%	30.88%
Y58S	73.04%	72.83%	100.00%	78.00%	86.83%	42.13%	41.96%
PA64S	64.13%	64.17%	78.00%	100.00%	77.33%	50.96%	50.96%
SH08S	72.04%	71.92%	86.83%	77.33%	100.00%	41.04%	41.63%
NK58	30.83%	30.79%	42.13%	50.96%	41.04%	100.00%	96.25%
NK58S	30.92%	30.88%	41.96%	50.96%	41.63%	96.25%	100.00%
9826 SNPs							
水稻样品	ZHS	ANS-1	Y58S	PA64S	SH08S	NK58	NK58S
ZHS	100.00%	99.57%	72.39%	65.03%	70.71%	32.27%	32.23%
ANS-1	99.57%	100.00%	72.37%	65.01%	70.75%	32.29%	32.25%
Y58S	72.39%	72.37%	100.00%	79.51%	87.02%	42.34%	42.49%
PA64S	65.03%	65.01%	79.51%	100.00%	78.53%	49.19%	49.73%
SH08S	70.71%	70.75%	87.02%	78.53%	100.00%	41.05%	41.60%
NK58	32.27%	32.29%	42.34%	49.19%	41.05%	100.00%	95.87%
NK58S	32.23%	32.25%	42.49%	49.73%	41.60%	95.87%	100.00%
80808 SNPs							
水稻样品	ZHS	ANS-1	Y58S	PA64S	SH08S	NK58	NK58S
ZHS	100.00%	99.65%	71.56%	64.78%	70.39%	30.48%	30.32%
ANS-1	99.65%	100.00%	71.54%	64.78%	70.38%	30.48%	30.31%
Y58S	71.56%	71.54%	100.00%	79.55%	87.24%	41.26%	41.21%
PA64S	64.78%	64.78%	79.55%	100.00%	78.48%	47.36%	47.60%
SH08S	70.39%	70.38%	87.24%	78.48%	100.00%	39.81%	40.05%
NK58	30.48%	30.48%	41.26%	47.36%	39.81%	100.00%	96.08%
NK58S	30.32%	30.31%	41.21%	47.60%	40.05%	96.08%	100.00%

表 9 两个不育系的EDV进行全基因组的SSR模拟运行结果
Table 9 Comparison of simulative SSR run results of the 2 putative EDV lines

SSR标记	ANS-1 SSR产物大小(bp)	ZHS SSR产物大小(bp)	多态性与否
RM1195	147, 149	147, 149, 50	是
RM19	236, 249, 180	186, 149, 39, 196, 150, 180	是
RM17	161, 174	164, 174, 184	是
RM18	156, 158	156, 158	否
RM232	159, 149, 157	159, 149, 157	否
RM273	210, 192, 191, 184, 202, 209	210, 192, 191, 184, 202, 209	否
RM190	75, 109	53	是
RM208	173, 179, 169	173, 179, 169	否
RM267	157, 114	157, 150	是
RM274	nil	162, 150	是
RM337	94	94, 99, 96, 77	是
RM278	109, 132	109, 132, 87	是
RM258	141, 133	149, 141, 133	是
RM224	157, 155	157, 155, 94	是
RM297	161, 121	161, 122, 82	是
RM71	150, 130, 113	nil	是
RM253	134	140, 134	是
RM219	161	nil	是
RM311	184, 77, 106	184, 77, 106, 68, 62, 52, 185	是
RM336	nil	101, 37, 152	是
RM209	nil	161, 33	是
RM5414	nil	110, 114	是
RM85	nil	64	是
RM72	311, 276, 273, 266, 235, 225, 205, 184	311, 276, 273, 266, 235, 225, 205, 184加上其他72个不同大小的产物	是

nil指没有SSR扩增产物

分享利益. 如果用的是不育系, 分享40%的转让收益; 如果被使用的是恢复系, 则分享30%转让收益.

同时, 实行水稻派生品种的鉴定, 还有助于规避主栽品种的遗传背景过于接近而带来的粮食安全风险. 人类社会在这个方面已经有过两次深刻教训, 分别是1840~1843年爱尔兰的马铃薯造成的大饥荒和1970~1971年美国的玉米减产造成的动物饲料严重缺乏^[22].

本文所发现和报道的水稻不育系准S是“安农S-1”的实质性派生品种启示人们, 在水稻育种中确实存在个别品种属于实质派生的现象. 不管是出于无意还是有意, 这一现象都应该努力避免. 该现象不仅直接伤

害水稻育种创新的动力, 而且也损害水稻种质资源的多样性. 因此, 这也从另一个角度也提示人们, 对于品种真实性的鉴定要采取更加严格有效的措施, 以确保水稻品种选育的创新性和可追溯性, 推动育种方向朝着不断积累遗传增益的方向发展.

粳型不育系安农S-1的谱系为超40B/H285(♀)6209-3(♂), 是湖南杂交水稻研究中心育成的世界上第1个粳型水稻温敏核不育系, 于1988年7月通过技术鉴定^[23]. 准S是湖南杂交水稻研究中心以安农S-1为不育基因供体选育出的不育系, 由N8S与香2B、怀早4号和早优1号经两轮随机多交混合选择, 再经系统选育而成的水稻品种, 于2003年通过审定^[24].

从准S的选育过程和谱系分析, 不存在如此之高相似度的可能. 当然, 不排除所获得的种子被供种者无意中拿错, 但是这种可能性不会否定IS方法用于水稻品种相似度鉴定的合理性, 因为作为内控的农垦58与农垦58S的相似度比较结果有力地支持了本方法的有效性.

本团队正在进行的200份以上的各地育成的水稻两系法和三系法育成不育系遗传背景分析工作, 相信可以为进一步核实ZHS身份提供更为准确的注解. 下一步的工作, 本团队计划加一些对照, 同一份水稻材料, 检测经过2次、4次、6次和8次自交之后, 各子代相似度变化会在什么水平? 自交过程中的变异主要是由什么因素引发? 转座因子在其中起多大作用? 是否有某些区段特别容易突变? 后期拟针对籼稻或者粳稻设计一套用于品种相似性的标记芯片.

本团队建议, 国家的有关品种审定管理部门实施参审品种的育种者种子管理制度并对育种单位强制提出如下要求.

(1) 育种者应该对进入育种区试阶段的潜在品种保留原始材料, 它们可作为育种者种子, 或者亲本的原始备份, 只要与此亲本有关的杂交品种还没有退市, 就要确保有活力的亲本备份(也叫育种家种子)被一直保留. 这样, 如果后续推广阶段发生亲本混杂或者亲本丢失时, 可以直接调用, 利于提纯复壮和资源保持. 例如, 隆平高科种业研究院院长杨远柱研究员在接受“国科农学院”的采访时就曾提到, 他们团队就有过这种尴尬经历 (<https://mp.weixin.qq.com/s/sgN7uJ2HtXdcMQSBCr4ULw>).

(2) 申请品种审定的时候应该交纳亲本种子, 由农业农村部指定的区域种质资源中心保管, 农业部门应该每年专门列支经费. 同时, 要对申请材料和其亲本做品种的真实性鉴定, 使用建立在IS基础上的相似度分析. 通过育种者种子和审定单位收藏的种子IS检测, 分析亲本种子与数据库中品种的相似度, 双管齐下. 这样在发生有争议的情况时, 容易鉴定出可能的“挂羊头卖狗肉”现象, 避免个别育种单位找借口掩盖实质派生的育种行为, 促进品种选育的良性发展, 可有效拒止实质派生品种通过审定, 保证水稻品种的生物多样性和粮食安全. 如果进行育种亲本的相似度检测, 在审定的当时就可以判定派生品种亲本或者谱系是否正确, 鉴定出报审品种是否具有实质性派生特性, 避免EDV

材料被登记成品种和亲本品种. 另一方面, 过多的姊妹系进入到育种体系当中, 既浪费国家的育种资源和经费, 又牺牲了水稻品种的多样性, 会导致主推品种高度同质化的粮食安全隐藏风险.

基于重测序的IS分析和本研究获得的用500个均匀分布在基因组的SNP进行的IS分析, 足以获取准确的相似度的研究成果, 对于国家和省级的品种审定管理部门有很高的引用价值. 基于SNP的IS分析不需要除参考基因组外的其他先验信息, 而且可以将IS分析与芯片分析或者“道格拉斯”检测平台结合起来, 值得管理部门重视和采用.

(3) 建议考虑多种检测方法结合使用, 如IS方法检测其相似度, DUS检测(对申请保护的植物新品种进行的特异性(distinctness)、一致性(uniformity)和稳定性(stability)的栽培鉴定实验或室内分析测试的过程)测试其田间表型性状. 如果某个参审品种田间表型性状与数据库内某个品种表现差异小、分子检测相似度高, 就极有可能是实质性派生品种.

实质派生品种的获得一般通过两条途径, 即回交途径和姊妹系途径. 无论何种途径, 只要超过了EDV判断的阈值, 都可认为没有改变EDV与被派生品种遗传结构高度相似的事实, 应予以拒审.

(4) 本研究揭示的个别水稻品种如此高的相似迫使人们不得不考虑和担忧水稻品种的多样性丧失的隐患. 这也提醒管理部门, 应该尽快对湖南省和我国育成的所有水稻品种资源进行EDV鉴定, 摸清水稻遗传资源的家底, 为后续的杂交优势群划分打下基础, 并以此为起点, 建立湖南省和国家的水稻种质基于全基因组相似度比较数据库.

至于具体判断水稻EDV的IS阈值, 在考虑和平衡方方面面的利益之后, 袁隆平院士带领本团队建议, 将品种间基因组相似度达到97.5%作为判断是否为EDV的阈值. 比起ISF认定的玉米EDV相似度阈值90%, 我们的水稻阈值是比较宽松的, 而且是可变的. 其最终设置的权限归属于国家农业管理部门.

在方法还不成熟的时候, 要达到准确追踪判定某个品种的亲本来源, 使该信息和实际育种途径吻合是非常困难的. 推广使用本文的方法, 其育种的流程非常便于追溯. 如果某个亲本特有的单倍型没有进入到申请审定品种的育种过程当中, 而申请资料却报告该亲本的存在, 这就是一种欺骗行为, 应该坚决予以制

止和拒审(终止审定程序)。

本研究表明, 应用动物群体遗传学和基因组学开发的相同起源基因鉴定的IS方法, 其精确度和广度上都比现行的部颁标准使用的24对SSR分子标记比对法, 在定性、定量和基因组覆盖度等方面要显著优越。通过表9的数据可以发现如下事实: (i) SSR标记的电泳图谱远远不能清楚反映全部的PCR产物标记所表现的差异, 基于SSR标记的国标只能捕捉不到1/3的差异, 而且只能反映PCR产物趋于富集的扩增产物; (ii) 在目前重测序价格已经降低到了近40元/1 G的推广成本之后, 水稻品种的真实性和相似度检测已经可以快速升级, 用基于重测序的基因组学手段进行10层以上重测序, 并且可以将成本降低到300元/1个样品。上述两点思考都建议品种审定的管理部门, 重新审议是否应该继续保留使用NYT 1433-2014的SSR为基准的国家行业标准。

本文也对IS方法作为EDV鉴定方法的初步成果和潜在的应用方面做了探讨。本文的研究成果还可以直接用于水稻品种的主成分分析、杂交品种组配过程当中和亲本选育过程的亲本相似度的阈值制定。但是, 基于品种相似度阈值来判断EDV, 对不管何种作物, 一直存在的困难就是到底采用严苛的阈值还是相对宽松的阈值, 即使对EDV阈值研究领先的作物玉米也是如此。阈值的严苛程度在变, 检测是否达到阈值的分子标记手段也在变化, 所以国际种子联盟每5年对每个作物的EDV阈值设定进行再评估。具体到检测手段, 全球各国和地区的做法也不尽一致, 即使使用SSR, 也会出现法国的检测标准优于美国种业联盟(The American Seed Trade Association, ASTA)的情况^[9]。因此, ISF在EDV仲裁中列出的期待专家们回答的问题有10大类,

总量在40个问题以上^[25]。

有经验的育种家都知道, 育种亲本过于相似, 难以获得育种进步; 而自交系选育的亲本间过于不同, 也难组配出很好的遗传背景材料。因此, 本文开发的针对水稻的相似度分析的IS方法也可以成为水稻育种时亲本组配定量把关的手段之一。当然, 水稻杂交育种亲本相似度阈值具体设定为多少才合适, 需要在实践中进一步摸索, 要建立在真实育种测配的田间数据基础之上。但是本文提供的方法对于精确地度量相似度阈值, 为育种者提供决策依据方面会有很大的帮助。

国外在玉米EDV鉴定和相关管理已经远远走在我国前面^[6,26-29], 因为水稻并非其主粮, 所以缺乏推进研究的动力。我们相信, 实质派生品种的管理战略将使我国能够真正地对水稻育种公司进行赋能, 保护和推动育种者获得合理利润, 使得他们能够合理壮大和发展, 尽快形成走向种业深海, 走进国际市场, 获取与国际农业巨头公司, 如拜耳、杜邦陶氏等国际大公司进行比拼的能力。

湖南省农业科学院作为我国创制杂交水稻的先行者, 我们呼吁全社会的水稻育种单位和个人, 共同努力, 推动水稻育种的正规化。同时育种家们也要有紧迫感, 当国家的战略已经在希望我们行走“一带一路”, 彰显国力的时候, 我们的技术水准也应该努力尽快提高, 积极吸收和消化大生物学领域的最新进展, 做到使水稻育种和栽培管理的能力与国家的农业战略相匹配。

综上, 通过相似度检测, 拒止实质派生品种被审定, 从而建立实质派生品种的合理阈值, 推进育种中遗传增益累积和育种创新, 是我国应该提倡和引导的水稻育种发展方向。

致谢 非常感谢湖南省农业科学院邓华凤研究员团队的博士后张先文老师提供6个水稻两系不育系的种子, 也同样非常感谢华中农业大学何予卿教授赠与农垦58常规稻种子用于本文的重测序的常规稻对照, 使我们可以比较从农垦58常规粳稻突变而来的农垦58S和农垦58的相似度, 并且将此作为判断本系统是否合理的一个重要参数。农业农村部植物新品种测试(成都)分中心余毅研究员在行文过程中也给予了帮助, 在此一并致谢。

参考文献

- 1 Chen H, Liu P, Lv Bo, et al. Discussion of the necessity of setting up EDV principles in our country (in Chinese). *Manag Agric Sci Technol*, 2009, 28: 10-12 [陈红, 刘平, 吕波, 等. 我国建立实质性派生品种制度的必要性讨论. *农业科技管理*, 2009, 28: 10-12]

- 2 Chen H. The Exploration of Setting Up EDV Protection Rules in Our Country (in Chinese). Beijing: China Agriculture Press, 2012 [陈红. 我国建立实质性派生品种保护制度探究. 北京: 中国农业出版社, 2012]
- 3 Wang H Y, Liu C X, Sun S Z. Strategy of improving plant new variety protection by the involved seed companies (in Chinese). China Seed, 2016, 12: 19–24 [王海阳, 刘彩霞, 孙素珍. 论完善种子企业植物新品种权保护的对策. 中国种业, 2016, 12: 19–24]
- 4 Tang L, Chen C, Zhuang D Y. Case study of original germplasm and rice production—Perspective based on EDV principles (in Chinese). Resour Sci, 2012, 34: 740–748 [唐力, 陈超, 庄道元. 中国原始品种遗传资源与水稻生产的实证研究——基于实质性派生品种制度视角. 资源科学, 2012, 34: 740–748]
- 5 International Seed Federation 2014 ISF Guidelines for Handling Disputes on Essential Derivation of Maize Lines. Available from: URL: https://www.upov.int/upovlex/en/conventions/1991/act1991.html#_7
- 6 UPOV. Act of 1991 international convention for the protection of new varieties of plants. Union Internationale pour la Protection des Obtentions Vegetales. Geneva. 1991
- 7 Xu Y, Li P, Zou C, et al. Enhancing genetic gain in the era of molecular breeding. J Exp Bot, 2017, 68: 2641–2666
- 8 Hayes B J, Bowman P J, Chamberlain A J, et al. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci, 2009, 92: 433–443
- 9 Kahler A L, Kahler J L, Thompson S A, et al. 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
- 10 Noli E, Teriaca M S, and Conti S. Criteria for the definition of similarity thresholds for identifying essentially derived varieties. Plant Breed, 2013, 132: 525–531
- 11 Li Z C. The Study and Utilization of China Rice Germplasm and Founder Lines (in Chinese). Beijing: China Agriculture University Press, 2013 [李自超. 中国稻种资源及其核心种质研究与利用. 北京: 中国农业大学出版社, 2013]
- 12 Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res, 2010, 20: 265–272
- 13 Harris R S. Improved pairwise alignment of genomic DNA. Dissertation for Doctoral Degree. Pennsylvania: The Pennsylvania State University, 2007
- 14 Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol, 1987, 4: 406–425
- 15 Yang J, Lee S H, Goddard M E, et al. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 2011, 88: 76–82
- 16 Rubin C J, Zody M C, Eriksson J, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature, 2010, 464: 587–591
- 17 Carneiro M, Rubin C J, Di Palma F, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science, 2014, 345: 1074–1079
- 18 Chen C, Liu Z, Pan Q, et al. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinixinyuan* n. ssp. Mol Biol Evol, 2016, 33: 1337–1348
- 19 Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res, 2009, 19: 1655–1664
- 20 Shi M S. Discovery and preliminary study of recessive male sterile rice which is sensitive to photoperiod (in Chinese). China Agric Sci, 1985, 2: 44–48 [石明松. 对光照长度敏感的隐性雄性不育水稻的发现与初步研究. 中国农业科学, 1985, 2: 44–48]
- 21 Luo X H, Qiu Z Z, Li R H. Peiai 64S, the male sterile rice line that has lower critical temperature (in Chinese). Hybrid Rice, 1992, 7: 27–29 [罗孝和, 邱趾忠, 李任华. 导致不育临界温度低的两用不育系培矮64S. 杂交水稻, 1992, 7: 27–29]
- 22 Bruns H A. Southern corn leaf blight: A story worth retelling. Agron J, 2017, 109: 1218–1224
- 23 Deng H F, Shu F B, Yuan D Y. Brief introduction of the study and utilization of Annong S-1 (in Chinese). Hybrid Rice, 1999, 14: 1–3 [邓华凤, 舒福北, 袁定阳. 安农S-1的研究及其利用概况. 杂交水稻, 1999, 14: 1–3]
- 24 Wu X J, Yuan L P. The study of employing population improvement on breeding rice thermo-sensitive male sterile line (in Chinese). Crop J, 2004, 30: 589–592 [武小金, 袁隆平. 应用群体改良技术选育水稻温敏核不育系的研究. 作物学报, 2004, 30: 589–592]
- 25 ISF. ISF guidelines for handling dispute on essential derivation of maize lines (Adopted by the ISF Field Crop Section in Beijing in May 2014). International Seed Federation. Nyon. 2014
- 26 Heckenberger M, van der Voort J R, Melchinger A E, et al. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties: II. Genetic and technical sources of variation in AFLP data and comparison with SSR data. Mol Breed, 2003, 12: 97–106

- 27 Heckenberger M, Bohn M, Klein D, et al. Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: II. Morphological distances and heterosis in comparison with simple sequence repeat and amplified fragment length polymorphism data in maize. *Crop Sci*, 2005a, 45: 1132–1140
- 28 Heckenberger M, Bohn M, Melchinger A E. Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: I. Simple sequence repeat data from maize inbreds. *Crop Sci*, 2005b, 45: 1120–1131
- 29 Heckenberger M, Muminović J, van der Voort J R, et al. Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: III. AFLP data from maize inbreds and comparison with SSR data. *Mol Breed*, 2006, 17: 111–125

The results of rice germplasm EDV test by genomic analysis and related discussions

ZHANG ShangDu^{1,2,3}, YUAN DingYang¹, LU HongFeng⁴, JIAN Yan¹, LI XiuXin⁴, HUANG AnPing¹, LUO ZhengLiang¹, LV QiMing¹, TAN YanNing¹, ZHANG YongFei¹, YUAN LongPing¹ & BAI LianYang^{1,2}

¹ Hunan Academy of Agricultural Science, Changsha 410125, China;

² Longping Branch of the Graduate School, Central South University, Changsha 410125, China;

³ Rice Research Institution of Guizhou Agriculture Academy of Science, Guiyang 550006, China;

⁴ Beijing Novogene Bioinformatics Institute, Beijing, 100083, China

It is a matter of fact that essentially derived varieties (EDV) exist in rice breeding in China, which impact breeding innovation negatively. This study explores the potential of using identity score (IS) method, which was originally developed by animal breeding and population genetic society, in determining essentially deviated rice male sterile (MS) lines released by “two-way hybrid rice system”. Six MS lines plus a control line, Nongken 58, which is an autonomous mutation donor of Nongken 58S, are analyzed with IS. Our results show that IS can determine similarity between rice lines to a very good level, with superiority on genome-wide coverage and accuracy when compared with NYT 1433-2007, which is the present rice authenticity standard. It is noteworthy that the 2 EDV lines identified by our study show enough polymorphisms when tested by simple sequence repeat (SSR), which suggests that it should call for challenges for SSR based authenticity test as required by the present national standards. Discussions on how to push forward rice breeding with implementation of IS are also covered.

***Oryza sativa*, similarity, EDV, identity score, threshold**

doi: [10.1360/SSV-2020-0068](https://doi.org/10.1360/SSV-2020-0068)