ICS 65.020.99
CCS B 20

中 国 作 物 学 会 团 体 标 准

T/CROPSSC 010—2024

野生稻基因型（SNP）鉴定规范

Specification for identification (SNP) of wild rice genotypes

2024-04-15 发布　　2024-04-15 实施

中 国 作 物 学 会　发 布

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国作物学会提出并归口。

本文件起草单位：中国农业科学院作物科学研究所、崖州湾国家实验室、三亚中国农业科学院国家南繁研究院。

本文件主要起草人：郑晓明、钱前、乔卫华、杨庆文、周雷娜、郭文龙、王银婷、李梓萱、王恺、温思钰。

野生稻基因型（SNP）鉴定规范

1 范围

本文件规定了野生稻基因型（SNP）鉴定的鉴定前准备、基于高通量测序的基因型鉴定、数据分析和构建分子身份证的要求。

本文件适用于二倍体野生稻基因型的鉴定。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 18284 快速响应矩阵码

GB/T 18347 128条码

NY/T 2594 植物品种鉴定 DNA分子标记法 总则

3 术语和定义

3.1

测序深度 sequencing depth

测序得到的碱基（bp）总量与基因组（Genome）大小的比值。

3.2

微效等位基因频率 minor allele frequency

在给定群体中的不常见的等位基因的发生频率。

3.3

杂合度 heterozygosity

个体的基因组中存在不同等位基因的程度。

3.4

滑窗 sliding window

在序列或数据集上以一定的步幅滑动的窗口。

3.5

缺失率 deletion rate

某一基因或染色体上缺失突变的发生频率。

3.6

质量值　genotype quality

在DNA测序过程中为每个测序碱基分配的质量分数。这个分数用于表示每个测序碱基的测序可靠性和准确性。

4　鉴定前准备

4.1　样品制备

按照样品顺序采集无病虫害的新鲜幼嫩植株叶片，叶片应取自同一个分蘖，并挂牌标记，用无菌水或75%乙醇擦拭或冲洗干净表面，吸水纸吸干表面液体，液氮速冻保存。一部分叶片组织用于提取基因组DNA，其余部分在–80℃冷冻条件下备份保存。

4.2　DNA 提取

4.2.1　按 NY/T 2594 的要求，用 CTAB 法从新鲜幼嫩叶片中提取基因组总 DNA，对提取好的 DNA 样品进行 1%琼脂糖凝胶电泳（5 V/cm），总 DNA 样品应表现为一条迁移率很小的整齐条带，分子量大小约为 22 kb~23 kb，以确认 DNA 降解水平较低，点样孔清晰，无拖尾现象和 RNA 污染。

4.2.2　取均匀混合后的 DNA 样品 2 μL，使用 Thermo Scientific NanoDrop 微量分光光度计检测 DNA 样品纯度，仪器应未提示有严重污染物，$1.8<A_{260}/A_{280}\leq 2.0$，$2.0<A_{260}/A_{230}\leq 2.2$。

4.2.3　使用 Invitrogen Qubit dsDNA 定量试剂盒和 Life Technologies Qubit 2.0 荧光计对 DNA 浓度进行精确定量，并保留浓度大于 100 ng/μL 的样品进行下一步文库构建。

4.3　文库构建

4.3.1　将检验合格的 DNA 样品通过超声波破碎机进行片段化，采用 TruSeq DNA Library Prep Kits（Illumina，FC-121-2003）进行建库。DNA 片段经末端修复、加 ployA 尾、加测序接头、PCR 扩增、纯化等步骤完成整个文库制备。

4.3.2　使用 Qubit2.0 进行初步定量，稀释文库至 1 ng/μL，随后使用 Agilent 2100 对文库的插入片段大小（insert size）进行检测，符合预期后，使用实时荧光定量核酸扩增检测系统（Q-PCR）方法对文库的有效浓度进行准确定量。文库有效浓度应大于 2 nmol/L。

5　基于高通量测序的基因型鉴定

将构建好的文库提交第二代测序平台（Illumine和BGI高通量测序平台）进行双端高通量测序。高通量测序得到的原始图像数据文件，经CASAVA碱基识别（Base Calling）分析转化为原始测序序列（Sequenced Reads），结果以FASTQ（fq）文件格式存储，其中包含测序序列（reads）的序列信息以及其对应的测序质量信息。每个样本平均测序深度达到10×。

6 数据分析

6.1 数据质控

按下列条件对raw reads进行过滤和质控，获得高质量测序数据（clean data）：

——使用 Cutadapt v4.6 软件去除 reads 内的接头（adapter）序列；

——使用 FastQC v0.12.1 软件，当单端测序 read 中含有的 N 的含量超过该条 read 长度比例的 10%时，应去除该对 paired reads；

——使用 Trimmomatic v0.39 软件对 reads 末端以 4 nt 滑窗计算碱基质量，裁剪质量低于 20 的滑窗；

——利用双端（PE）重叠区域数据，对碱基序列进行矫正；

——使用 fastp v0.23.4 对单端测序 read 中含有的低质量（$Q \leqslant 5$）碱基数超过该条 read 长度比例的 50% 时，该对 paired reads 去除；

——使用 FastQC v0.12.1 软件，修剪末端 polyG，并去除长度小于 30 nt 的片段。

6.2 变异检测

6.2.1 使用 BWA v07.15 软件（参数：bwa mem-T4-K32-M-R）将 clean reads 与水稻日本晴参考基因组（*O. sativa* L. var. Nipponbare，MSU v7.0）比对。

6.2.2 使用 samtools v1.19.2 软件 sort 模块对比对结果文件进行排序。

6.2.3 使用 picard v2.18.7 软件 markduplicate 模块去除 PCR 重复。

6.2.4 使用 GATK v4.5.0 软件中 BaseRecalibrator 模块调整原始碱基质量分数，消除测序产生的系统性误差。

6.2.5 使用 GATK v4.5.0 软件中 HaplotypeCaller 模块进行变异位点检测。

6.2.6 使用 GATK v4.5.0 软件中 CombineGVCFs 和 GenotypeGVCFs 模块合并多样本变异位点检测结果。

6.2.7 使用 GATK v4.5.0 软件中 SelectVariants 和 VariantsToTable 模块提取 SNP 并进行过滤，去除 FS > 60，MQ < 40，QD < 2.0，QUAL < 30.0，SOR > 3.0，MQRankSum < –12.5，ReadPosRankSum < –8.0 的低质量 SNP 位点，得到 PASS 的标记。

6.3 数据过滤

6.3.1 使用 VCFtools v0.1.16 软件对 PASS 的 SNP 标记按下列条件进行过滤，包括：

——测序深度大于 4；

——缺失率小于 0.2；

——微效等位基因频率（MAF）小于 5%；

——杂合度小于 0.05；

——质量值（GQ）大于等于 5。

质量值通常以Phred质量分数的形式表示，Phred质量分数越高表示错误率越低。Phred质量分数*Q*的计算公式如下：

$$Q = -10 \times \log_{10}(P)$$

式中:

Q——XXX；

P——当该碱基未被错误识别时，在该位置上读取到的碱基含量百分比。

6.3.2　删除非双等位基因位点，得到高质量 SNP 数据。

7　构建分子身份证

7.1　利用 CoreSNP 软件筛选出对野生稻品种鉴别力较高（特有）的核心 SNP 标记集，并将每个 SNP 位点基因型分别用 0、1、2 三个数字表征，该 SNP 位点与参考基因组碱基类型相同记为 0，与参考基因组碱基类型不同记为 1，杂合位点记为 2。以具有 30 个核心 SNP 标记集的品种为例，该品种的指纹图谱展示为：10011010010001000101001110001001，以此类推。

7.2　按 NY/T 2594 的要求，将每份资源每个位点的 DNA 指纹（基因型）进行编码和可视化。结合品种基本信息，按 GB/T 18347 和 GB/T 18284 的要求，利用在线条形码生成器和二维码生成软件以条形码和二维码的形式建立和区分水稻品种特有的身份信息，分别生成唯一的条形码和二维码。首先将品种指纹图谱 0/1/2 编码（10012010010001000101001110001001）转化为 58 进制（7VdsWdAMFD8tMeXcKEj6bePpsBY7VP），再加上品种来源国家和省份（以中国福建为例，CHN_FJ）及品种序号（以 1 为例），构成品种的身份信息 CHN_FJ_0001_7VdsWdAMFD8tMeXcKEj6bePpsBY7VP。再用条形码生成器里的 Code128A 进行生成即得到该品种的身份证条形码，最后利用二维码的生成软件生成相对应的身份证二维码。构建一套水稻种质资源分子身份证的数据库。

品种身份证条形码

品种身份证二维码

7.3　进行水稻种质资源的群体结构、主成分、系统发育树分析，明确种质资源中具有杂交优势的亲本类群（即具有遗传异质性类群）。利用 Admixture 软件使用块松弛（block relaxation）方法来交替更新等位基因频率和祖先分数参数，在最佳分群 K 值的指导下区分品种来源及群体结构。利用 EIGENSOFT v8.0.0 软件进行遗传距离测算，通过计算数据矩阵的协方差矩阵，选择特征值最大的特征向量组成矩阵来构建主成分分析图，并绘制 PCoA 主坐标分析图，明确相关种质资源样品之间的遗传关系。利用 tassel v5.0 软件将 SNP VCF 格式转化为 Phylip 格式后，利用 FastTree 软件，基于最大似然法（Maximum likelihood，ML）构建进化树，得到的树文件（.nwk 格式）利用 iTOL（iTOL: Interactive Tree Of Life (embl.de)）网站或者 MEGA 11 软件进行可视化。