ICS 65.020

B 20

团 体 标 准

**T/HXCY XXX-2024**

**紫花苜蓿高通量基因型鉴定分析技术规程**

**Technical Regulations for High Throughput Genotype Identification and Analysis of Alfalfa**

（征求意见稿）

2024-XX-XX发布 2024-XX-XX实施

北京华夏草业产业技术创新战略联盟发布

目 次

[前 言 I](#_Toc176469602)

[1 范围 1](#_Toc176469604)

[2 规范性引用文件 1](#_Toc176469605)

[3 术语和定义 1](#_Toc176469606)

[4 仪器与设备 3](#_Toc176469607)

[5 测定 3](#_Toc176469608)

[6 生物信息分析 4](#_Toc176469608)

[7 检测报告 6](#_Toc176469609)

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。

本标准由北京华夏草业产业技术创新战略联盟提出并归口。

本文件起草单位：中国林业科学研究院生态保护与修复研究所、华智生物技术有限公司、内蒙古草业技术创新中心有限公司、内蒙古大学、中国农业科学院北京畜牧兽医研究所。

本文件主要起草人：刘希强、李晓霞、黄刚、唐顺学、刘亚玲、任卫波、马琳、仪登霞。

本文件为首次发布。

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这些专利的责任。

紫花苜蓿高通量基因型鉴定分析技术规程

# 1 范围

本标准规定了紫花苜蓿（*Medicago sativa* L.）基因型高通量鉴定的主要技术内容，包括紫花苜蓿gDNA提取与检测、*c*GPS测序文库构建、生物分析流程等。

紫花苜蓿高通量基因型鉴定分析基于紫花苜蓿60K液相芯片开展靶向捕获测序，能够对大规模紫花苜蓿群体开展高效、精准、低成本地基因型分型检测，基于分型结果，可在紫花苜蓿育种或辅助育种中开展应用，主要包括群体遗传主效基因选择、分子标记辅助育种定向改良、全基因组选择育种、紫花苜蓿品种真实性鉴定、遗传图谱构建、基因定位和种质资源遗传多样性评价等。

本标准适用于紫花苜蓿高通量基因型鉴定分析。

# 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，仅该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 30989—2014高通量基因测序技术规程

GB/T 40171—2021磁珠法DNA提取纯化试剂盒检测通则

DB32/T 3860—2020玉米品种及纯度鉴定技术规程SNP标记法

T/JSF T/JSF001—2024美洲黑杨性别苗期鉴定技术规程——SSR分子标记法

# 3术语和定义

下列术语和定义适用于本文件。

**3.1 基因 gene**

位于染色体上编码一个特定功能产物（如蛋白质或RNA分子）的一段核苷酸序列，是遗传信息的基本单位。

**3.2 文库 library**

通过生物来源、人工合成或克隆技术等得到的重建分子群，如基因组文库。

**3.3 探针 probe**

能识别特定碱基序列的、经过人工标记的一小段单链核酸分子。

**3.4 高通量基因测序 high-throughput gene sequencing**

区别于传统双脱Sanger测序，高通量测序能够一次并行对大量核酸片段序列进行大规模平行测定的技术，通常一次测序反应能产出不低于100Mb的测序数据。

**3.5** **基因型 genotype**

指一个生物体内的DNA所包含的基因，也就是说该生物的细胞内所包含的、它所特有的那组基因。

**3.6 引物 primer**

在DNA复制过程中，结合于模板链上并作为复制延伸的起始位点和/或终止位点的，具有一定长度和顺序的寡核苷酸链。

**3.7** **DNA聚合酶 DNA polymerase**

以一条DNA链作为模板合成一条新的DNA链时所需要的酶。

**3.8 DNA连接酶 DNA ligase**

在DNA复制、修复和重组过程中，由发挥作用的催化磷酸二酯键合成的酶。

**3.9 SNP标记 SNP marker**

是指在基因组上单个核苷酸的变异，形成的遗传标记，其数量很多，多态性丰富。

**3.10 液相基因芯片 liquid-phase gene chip**

‌是通过挑选目标区域多态性高、缺失率低、分布均匀的标记位点信息作为参考设计探针，并合成特异性液相芯片，在液相环境下对目标区域序列的定向捕获富集和高通量测序。

**3.11 *c*GPS Genotyping by Pinpoint Sequencing of liquid captured targets**

基于优化的热力学稳定性算法模型对目标区间序列进行特异性探针设计，利用合成的特异性探针对位于不同基因组位置的多个不同目标序列进行液相杂交捕获富集，然后对捕获富集的目标区间进行文库构建和二代高通量测序，从而获得目标区间内的所有SNP/InDel标记位点的基因型。

# 4仪器与设备

4.1 涡旋震荡器

4.2 移液器

4.3 离心机

4.4 磁力架

4.5 高通量自动化核酸提取仪或工作站

4.6 微量紫外分光光度计或酶标仪

4.7 PCR仪

4.8 荧光定量PCR仪

4.9 电泳仪

4.10 凝胶成像系统

4.11 真空浓缩仪

4.12 Array Tape基因分型系统

4.13 SNP line SNP分型系统

4.14 Illumina测序平台

4.15 服务器

# 5 测定

为保证测序数据的质量，在样品提取、文库构建、杂交捕获和测序各环节均设置质控点，对每一个步骤都进行严格质控，从根本上确保基因型分型数据的准确性。

**5.1 样品DNA提取与质控**

采集待检测紫花苜蓿材料的新鲜叶片，利用磁珠法从组织中提取gDNA，按照GB/T 30989-2014的规定执行。用Qubit荧光定量仪检测DNA样品的浓度；用1%琼脂糖凝胶电泳检测DNA样品的完整性，DNA体积≥50µL，浓度≥10ng/µL；OD260/280值=1.8-2.0，OD260/230值≥1.8的样品用于文库制备。

**5.2 *c*GPS文库构建与质控**

a. 利用酶切试剂将DNA样品酶切成100-500bp大小片段后，加入Taq酶进行末端修复，并在3’端加上A碱基，用琼脂糖凝胶电泳检测片段大小。

b. 使用T4连接酶连接测序接头与DNA片段两端，利用磁珠对连接产物进行纯化。纯化后的产物用Qubit荧光定量仪检测浓度，琼脂糖凝胶电泳检测片段大小。

c. 对纯化后的连接产物进行PCR扩增，利用磁珠对扩增产物进行片段筛选。片段筛选后的产物用Qubit荧光定量仪检测浓度，琼脂糖凝胶电泳检测片段大小。

d. 取200ng质检合格的PCR产物构建杂交文库，使用真空浓缩仪对混合文库进行浓缩。之后加入基因组封阻试剂、RNA酶抑制剂（Rnase Block）与探针液后，50℃孵育16小时-24小时完成杂交反应；将磁珠加入到杂交产物中进行目标区段捕获，室温结合30 min -60 min，吸附探针和目的区间结合体；使用清洗液清洗捕获产物，去除非特异性结合片段；再进行一轮PCR扩增，扩增产物使用片段分选磁珠进行纯化，去除多余引物和引物二聚体，即完成*c*GPS测序文库构建。

**5.3 上机测序**

文库构建完成后，利用Qubit荧光定量仪对文库浓度进行测定，采用Agilent 2100/2200 Bioanalyzer检测文库质量，文库片段大小应在250bp-500bp之间。文库质检合格后，按照测序平台的标准，进行上样及测序。

**6. 生物信息分析**

基于紫花苜蓿液相芯片捕获技术的基因型鉴定分析流程主要分为以下四部分：

（1）测序数据质控：将原始下机数据进行过滤并评估测序质量，并进行污染检测；

（2）比对及质控：将过滤后的测序数据比对到参考基因组，并计算相应质控指标；

（3）变异检测及注释：检测单核苷酸变异（SNP）、小片段插入缺失变异（InDel）及注释；

（4）检测指标计算：根据变异检测结果，统计样本水平及位点水平的质控指标。

**6.1 测序数据质控**

**6.1.1测序数据格式**

高通量测序得到的原始图像数据文件经碱基识别（Base Calling）分析转化为测序序列（Reads），结果以FASTQ（简称为fq）文件格式存储，FASTQ文件包含每条测序序列（Read）的名称、碱基序列以及其对应的测序质量信息。

**6.1.2 原始数据过滤**

测序得到的紫花苜蓿原始测序序列（Raw Reads），使用fastp软件对原始序列进行过滤，去除含有接头污染的Reads和低质量的Reads，得到高质量的Clean Reads，使用Clean Reads进行后续分析。

**6.1.3 测序数据质量评估**

碱基测序质量值反映测序错误率，每个碱基的测序错误率是通过将测序Phred数值（Phred score，Qphred）按照特定公式进行转换后计算得出，而测序质量值是在碱基识别过程通过一种预测碱基判别发生错误概率模型计算得到。

**6.1.4 测序数据碱基含量检测**

碱基分布检查用于检测有无AT、GC分离现象。理论上G和C碱基及A和T碱基含量每个测序循环上应分别相等，且整个测序过程稳定不变，呈水平线。

**6.1.5 测序数据污染检测**

从每个样品的fastq文件中随机选择10,000条序列，利用blastn将序列比对到NCBI NT数据库，进行污染评估。

**6.2 参考基因组比对**

利用比对软件BWA，将过滤后的Clean Reads与紫花苜蓿参考基因组（“新疆大叶”或“中苜4号”基因组）进行比对，定位Clean Reads在参考基因组上的位置，统计比对结果。比对率能反映样本、建库及测序、紫花苜蓿60K液相基因芯片标记序列的质量以及测序数据与紫花苜蓿参考基因组的相似程度。

**6.3 目标位点变异分析**

基因组变异是指在基因组水平上核苷酸的变异所引起的DNA序列多态性，包括SNP、InDel等。基于Clean Reads参考基因组序列比对结果，通过突变分析软件GATK的检测工具HaplotypeCaller进行靶向位点的基因型检测。利用ANNOVAR软件，对检测出的基因变异进行功能注释，得到变异位点在基因组发生的区域（内含子区、基因间区、编码区、5’端UTR区、3’端UTR区等），以及变异产生的影响（同义、非同义突变等）。

**7 检测报告**

测试报告应包括(但不限于)：

—样本信息；

—靶向测序基因型分型技术；

—实验流程；

—生物信息分析流程；

—测序数据质控；

—参考基因组比对；

—目标位点变异分析；

—参考文献；