

灵芝优良品种全基因组分子标记的建立

刘逸婷, 蒋晓涵, 杨春艳, 陈健辉, 王车昭, 吕晓萌, 杨治康, 邓优锦*, 吴小平*

福建农林大学生命科学学院, 福建 福州 350012

摘要: 灵芝 *Ganoderma lingzhi* 的优良品种灵芝 13 具有农艺性状较好、子实体产量较高、孢子粉产量较高的优良性状。本研究首先以其单核菌株灵芝 13-5 为材料, 通过三代 HiFi 测序及 Hi-C 测序组装成完整的基因组; 然后选取 5 个重复的双核菌株通过 Illumina 高通量测序后, 以灵芝 13-5 基因组作为参考进行 SNP calling, 建立了基于灵芝 13 全基因组的分子标记。基因组包含 13 条染色体, 共 45.73 Mb。5 个重复菌株 SNP calling 去重合并后, 获得 319 074 个共有的异核等位差异位点, 作为用于灵芝 13 菌株鉴别的分子标记库。从实验室菌种库中随机抽取 7 个灵芝菌株, 与灵芝 13 菌株共有的等位差异位点占比介于 22.20%–45.18% 之间, 远低于阈值, 说明这些菌株与灵芝 13 差异较大, 不是相同菌株。本研究建立了基于灵芝 13 全基因组及其异核等位差异位点的分子标记, 能够准确区分灵芝 13 这一优良品种。

关键词: 三代 HiFi 测序; SNP calling; 品种鉴别; 品种保护

[引用本文]

刘逸婷, 蒋晓涵, 杨春艳, 陈健辉, 王车昭, 吕晓萌, 杨治康, 邓优锦, 吴小平, 2025. 灵芝优良品种全基因组分子标记的建立. 菌物学报, 44(3): 240141

Liu YT, Jiang XH, Yang CY, Chen JH, Wang CZ, Lü XM, Yang ZK, Deng YJ, Wu XP, 2025. Establishment of molecular markers in the whole genome of *Ganoderma lingzhi* elite varieties. Mycosystema, 44(3): 240141

资助项目: 福建省科技重大专项(2022NZ029015)

This work was supported by the Fujian Provincial Major National Research and Development Project (2022NZ029015).

*Corresponding authors. E-mail: DENG Youjin, dengyoujin1980@163.com; WU Xiaoping, fjwpx@126.com

ORCID: LIU Yiting (0009-0003-3097-9643)

Received: 2024-05-10; Accepted: 2024-10-25

Establishment of molecular markers in the whole genome of *Ganoderma lingzhi* elite varieties

LIU Yiting, JIANG Xiaohan, YANG Chunyan, CHEN Jianhui, WANG Chezhaohao, LÜ Xiaomeng, YANG Zhikang, DENG Youjin*, WU Xiaoping*

College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350012, Fujian, China

Abstract: *Ganoderma lingzhi* 13 is an excellent cultivated variety with superior agronomic character. The mononucleate strain, *G. lingzhi* 13-5, as assembled into a complete genome by three-generation HiFi sequencing and Hi-C sequencing, and five replicated bi-nucleate strains were subjected to high throughput sequencing by Illumina, and then the genome of *G. lingzhi* 13-5 was used as a reference for SNP calling to establish the molecular markers based on the whole genome of *G. lingzhi* 13. The genome contains 13 chromosomes with a total of 45.73 Mb. After the SNP calling of five replicate strains were de-emphasized and merged, 319 074 shared heterokaryotic allelic difference sites were obtained as a molecular marker library for the identification of *G. lingzhi* 13 strains. Seven post-fruiting strains were randomly selected from the strain library. It was found that the percentage of allelic difference sites of the post-fruiting strains shared with *G. lingzhi* 13 ranged from 22.20% to 45.18%, being far inferior to threshold value, indicating that these strains were different from *G. lingzhi* 13 and they were not the same strains. In this study, molecular markers based on the whole genome of *G. lingzhi* 13 and its heterokaryotic allelic differentiation sites were established, which can accurately differentiate *G. lingzhi* 13, a superior cultivated variety.

Keywords: three generations of HiFi sequencing; SNP calling; variety identification; variety protection

灵芝 *Ganoderma lingzhi* Sheng H. Wu, Y. Cao & Y.C. Dai 俗称赤芝，是我国著名的药食同源经济真菌(才晓玲等 2016; 戴玉成等 2021)。研究表明，目前已从灵芝中分离出 400 多种活性物质(Tan *et al.* 2018)，灵芝多糖、氨基酸、多肽和三萜等为主要有效成分，赋予了灵芝多样的药理作用，如扶正固本、抗炎、抗癌、免疫调节、控制血糖、抗氧化、降血脂、延缓衰老等作用(Dan *et al.* 2016; Soccol *et al.* 2016; Wasser 2017; Chen *et al.* 2018; 戚爱华等 2018)。

但野生灵芝资源数量有限，远不能满足市场日益增长的需求，现阶段我国人工栽培的灵芝菌株主要是从韩国(如韩芝)和日本(如日芝)引

进，随着多年的连续性种植，菌株已产生一定程度的退化，导致灵芝整体质量及产量下降(左洪波等 2021)。因此灵芝的育种工作具有紧迫性，然而由于新品种权未能得到较好识别和保护，阻碍了灵芝菌株的自主研发进程。

知识产权是对创新成果最有效的制度保障，随着经济的快速发展，各国对知识产权越来越重视(孙缘缘 2019)。随着食用菌产业的不断发展，越来越多的新菌株被选育出来，食用菌知识产权的保护也逐渐受到重视。日本是较早将食用菌列为保护名单的国家(李绩 2007)，我国对食用菌品种保护起步较晚(张清洋等 2020)，受到保护的种属至今只有 15 个，仅占总的农业保护菌株的 5% (刘晓柳等 2022)。目前还存在对新品种保

护不充分, 育种者权益不足等问题。但通过不断的法律修订和制度完善, 正在逐步加强对育种者权益的保护, 激励种业的创新发展。

灵芝 13 是筛选出的农艺性状较好、子实体产量较高、孢子粉产量较高的优良杂交子。为了有效保护育种者的权益, 同时为新品种认定提供技术支持, 本研究通过对新菌株灵芝 13 的单核菌株进行三代测序, 组装出完整的基因组作为参考基因组, 对灵芝 13 的 5 个重复双核菌株的重测序数据进行 SNP calling 后, 构建出基于灵芝 13 全基因组的分子标记库, 并随机抽取 7 个灵芝菌株对该方法进行验证。为后续灵芝 13 进行新品种认定及品种权保护提供技术与理论支持。

1 材料与方法

1.1 材料

1.1.1 实验菌株

灵芝 102 与 119(83)通过双-单杂交得到杂交子灵芝 13。

1.1.2 培养基

麸皮培养基(1 000 mL): 马铃薯 200 g, 麸皮 20 g, 加入适量水与马铃薯一同煮沸后捞出过滤, 定容至 1 000 mL。再次煮沸后加入 20 g 琼脂至其完全煮化, 关火加入 20 g 葡萄糖, 溶解后进行分装, 121 °C 灭菌 30 min。

1.2 原生质体单核化及鉴定

将活化好的菌株转接至装有碎玻璃碴的麸皮液体培养基中, 30 °C、150 r/min 摇床培养 3–4 d。新鲜液体菌丝倒入 50 mL 离心管中, 封口膜封口, 4 500 r/min 离心 20 min, 在超净工作台里去除上清液。用 30 mL 0.7 mol/L KCl、10 mmol/L CaCl₂·H₂O 缓冲液冲洗, 4 500 r/min 离心 10 min, 弃上清。将菌丝置于盛有 10 mL 酶液(0.2 g 融壁酶、0.7 mol/L KCl、10 mmol/L CaCl₂·H₂O 缓冲液, 滤头过滤)的 50 mL 离心管中, 28 °C、150 r/min 振荡酶解 3 h。吸取微

量菌液镜检, 有原生质体后取出酶液, 加少许 0.7 mol/L KCl、10 mmol/L CaCl₂·H₂O 缓冲液摇晃, 倒于 3 层灭菌纸过滤, 再用 0.7 mol/L KCl、10 mmol/L CaCl₂·H₂O 缓冲液冲洗 1–2 次, 去残渣, 收集滤液于 50 mL 离心管中, 4 °C、2 500 r/min 离心 10 min, 弃上清。用适量 0.7 mol/L KCl、10 mmol/L CaCl₂·H₂O 缓冲液冲洗, 将沉淀混匀后转移至 2 mL 离心管中, 4 °C、2 500 r/min 离心 10 min, 弃上清。加 300 μL 1×STC 悬浮, 计数, 使原生质体浓度达到 10⁸ 个/mL, 涂布到 90 mm 再生培养基培养皿中, 30 °C 培养 2–3 d, 挑取菌落小、生长速度较慢的单菌落至麸皮平板中。待长出菌丝后, 在显微镜下观察是否存在锁状联合结构, 若无锁状联合结构, 说明原生质体单核化成功。

1.3 样品收集及基因组测序

将活化好的灵芝 13 二倍体菌株及通过原生质体单核化得到的单倍体菌株灵芝 13-5 转接至贴有玻璃纸的 90 mm 麸皮平板中, 30 °C 黑暗培养至菌丝即将长满皿, 收集菌丝至少 0.3 g, –80 °C 保存。单倍体菌株全基因组三代 HiFi 测序及 Hi-C 测序委托安诺优达公司完成, 三代 HiFi 测序深度为 100×, 数据量为 10 G, Hi-C 测序深度为 100×, 数据量为 20 G。二倍体菌株同时转接 5 个培养皿, 分别收集菌丝后作为 5 个重复, 委托北京诺禾致源生物信息科技有限公司进行 Illumina Hiseq 双端测序, 测序深度为 100×, 数据量为 4 G。

1.4 基因组组装

使用 bamtools (Barnett *et al.* 2011) 中的 bam2fastq 子命令从灵芝 13-5 菌株的原始测序数据 bam 文件中提取 fastq 序列。使用 HiFiasm 软件(Cheng *et al.* 2021)对提取获得的高精度 HiFi reads 进行组装, 设置 Hi-C 模式, 其余参数默认。在 Hi-C 模式下, HiFiasm 会进行 all-vs-all 比对, 以找出它们之间的重叠区域。当 reads 之间存在碱基差异时, HiFiasm 会根据支持该差异的 reads

数量来决定是否将其视为 SNP (单核苷酸多态性)。如果有 3 个或更多的 reads 支持这一差异, HiFiasm 将其视为杂合变异并保留为 SNP。如果支持的 reads 数量不足 3 个, 则认为这是一个测序错误, 并将其纠正。SNP 信息会被 HiFiasm 用于定相。

使用 Editplus 程序打开组装后的序列文件, 根据端粒 DNA 的特征人工统计灵芝 13-5 的端粒序列。如果一条 contig 两端都含有端粒序列, 说明该 contig 是一条完整的, 没有 gap 的完整染色体。通过统计所有 contig 中的端粒数量, 再除以 2, 初步认为是该基因组的染色体数量。

对使用 HiFiasm 软件(Cheng *et al.* 2021)组装得到的 contigs 进行自身 Blastn 比对, 识别出长度在 10 kb 左右的由串联重复片段组成的 contigs, 从中截取出串联重复单元; 将其与 NCBI 中的 nt 库进行比对, 根据 rDNA 序列的特征判断其是否为 rDNA 重复单元。将识别出的 rDNA 重复单元再次与基因组中的 contigs 进行比对, 将一端携带 rDNA 序列的 2 个 contigs 直接连接起来, 形成一个结构为“序列-rDNA 区域-序列”的 scaffold。通过 HiFi reads 的测序深度和基因组单拷贝区域的测序深度计算出 rDNA 区域重复单元的串联个数。使用 Hi-C Pro (Servant *et al.* 2015)工具检测剩余长度大于 100 kb 的长片段 contigs 之间的连接关系, 并将认为应该相连的 contigs 连接起来形成 scaffolds。最后利用 HiFi reads 来填平 scaffolds 中的缺口。

1.5 基因组组装完整性和准确性验证

通过基因组组装质量评估软件 BUSCO v5.6.1 (Seppey *et al.* 2019)对灵芝 13-5 基因组组装完整性进行评估, 采用担子菌数据库第十版 (basidiomycota_odb10.2020-09-10)。染色体内和染色体之间的互作都遵循一定的规律: 同一条染色体内的 2 个基因, 距离越近其互作频率越高; 不同染色体之间的 2 个基因的互作频率一定低于同一染色体内的互作频率。根据该性质

使用 Hi-C Pro 软件(Servant *et al.* 2015)获得染色体互作热点图, 验证灵芝 13-5 染色体组装框架的准确性。

1.6 着丝粒分析

大多数真菌细胞的染色体着丝粒具有互作信号强的特点, 如酵母菌 (Varoquaux *et al.* 2015)、小麦契诃夫氏锈菌 (Sperschneider *et al.* 2021)等。使用 Hi-C Pro 软件, 获得染色体互作热点图, 识别出基因组中每个 40 kb Bin 之间的 Hi-C 互作信号, 每条染色体互作信号最强的位置即为着丝粒所在的位置。

1.7 基因预测和分析

使用 Funannotate 软件(<https://funannotate.readthedocs.io/en/latest/>)对灵芝 13-5 基因组进行注释, 用灵芝的蛋白质组作为蛋白证据, 设置 Augustus 参数模型为灰盖鬼伞。将灵芝 13-5 的基因组序列分割成大小为 100 kb 的窗口, 统计每个窗口含有的基因数目, 比较每条染色体的基因密度分布情况。

1.8 重复序列分析

使用包括 RepeatMasker (Chen 2004)、tandem Repeat Finder (Benson 1999)及 TE classify 软件对灵芝 13-5 的基因组的重复序列进行识别查找和分类。将灵芝 13 的基因组序列分割成大小为 100 kb 的窗口, 统计每个窗口含有多少重复序列, 比较每条染色体中重复序列的分布情况。

1.9 重测序数据处理及全基因组分子标记建立

组装好的灵芝 13-5 基因组进行质控, 去除重复序列后作为参考基因组, 使用 GATK 软件给参考基因组建立索引, 将 5 个重复菌株的二代重测序数据比对到参考基因组, 生成 SAM 文件; 排序并生成 BAM 文件, 并对其进行 PCR 重复标记, 生成 MarkDup.bam; 对 MarkDup.bam 建立索引后找到 SNP 差异位点, 进行 SNP calling 后生成 gVCF 文件; 最后合并 5 个重复菌株的 gVCFs, 生成 VCF 文件。在此基础上, 筛

选出<10 bp、双等位基因的 indel, 使用 GATK 软件进行过滤。将最终得到的 SNP 结果去重合并后构建成分子标记库。

以稳定出现在 5 个重复菌株中的等位差异位点作为基准, 计算出每个重复菌株与基准相差的等位差异位点数, 设定基准/与基准相差最多的重复菌株的等位差异位点数的比值作为判断阈值。若两菌株间的 $GS \leq$ 阈值, 则认定为“不同品种”; 若两菌株间的 $GS >$ 阈值时, 则认定为“相同品种”(GS=待测菌株与灵芝 13 共有的等位差异位点数/全基因组分子标记库的等位差异位点数)。

1.10 全基因组分子标记验证

从菌种库中随机抽取 7 个已出过菇的灵芝菌株进行二代重测序, 使用参考基因组(灵芝 13-5)对 7 个菌株的重测序数据进行 SNP calling 后, 将待测菌株 SNP 位点与灵芝 13 异核等位差

异位点一致的数量作为分子, 分子标记库中等位差异位点的总数作为分母, 计算出比值。验证基于全基因组构建的分子标记库的可行性。

2 结果与分析

2.1 基因组组装结果

通过 PacBio 测序平台对通过原生质体单核化得到的单倍体菌株灵芝 13-5 进行基因组测序。灵芝 13-5 菌株测序获得 62.91 万个 HiFi reads, 其 N50 长度为 19 179 bp, GC 含量为 53%, 共获得 11.54 G 测序数据。

使用 HiFiasm 软件对灵芝 13-5 的 HiFi reads 进行组装, 模式设置为 Hi-C 模式, 并挂载灵芝 13-5 的 Hi-C 数据, 组装获得基因组数据。基因组含有 789 个 contigs, 组装大小总共为 77.33 Mb (表1)。在这些 contigs 中, tig01、tig02、tig03、tig04、tig05、tig06、tig07、tig08、tig09、

表 1 灵芝 13-5 基因组信息

Table 1 Genomic information of *Ganoderma lingzhi* 13-5

| 拼接后的 序列 Contig | 大小 Size (Mb) | 5'端端粒 Telomere at 5' end | 3'端端粒 Telomere at 3' end | 完整染色体 Complete chromosome | 注释 Notes |
|----------------------|--------------------|--------------------------------|--------------------------------|---------------------------------|--|
| tig01 | 4.76 | + | + | + | |
| tig02 | 3.34 | + | + | + | |
| tig03 | 4.14 | + | + | + | |
| tig04 | 3.14 | + | + | + | |
| tig05 | 4.55 | + | + | + | |
| tig06 | 3.41 | + | + | + | |
| tig07 | 2.60 | + | + | + | |
| tig08 | 3.19 | + | + | + | |
| tig09 | 3.39 | + | + | + | |
| tig10 | 2.53 | | + | | 在另一端有连接的 rDNA 区域 Linked rDNA region on the other end |
| tig11 | 5.31 | + | + | + | |
| tig12 | 2.50 | + | + | + | |
| tig13 | 3.32 | | + | | |
| tig14 | 0.30 | | + | | |
| Others | 32.67 | | | | 共 775 个小序列, 114 个属于 rDNA 区域, 500 个属于 线粒体基因组 Small contigs totaled 775, including 114 belonging to rDNA region and 500 belonging to mitochondrial genome |

注: +表示有端粒, 是完整的染色体

Note: + indicates the presence of telomeres and is the complete chromosome.

tig11、tig12 共 11 条序列的两端含有端粒序列，因此认为它们是完整的染色体；tig10、tig13、tig14 共 3 条序列，其中一端带有端粒序列。

每个基因组只含有一个 rDNA 区，通过 Blastn 比对后发现 tig10 的 3'端携带有 rDNA 区域的序列，将 tig10 的 5'端序列与灵芝 13 的 HiFi reads 库进行比对，将其延长 7 168 bp，并携带端粒。rDNA 的重复单元为 10 150 bp，包含完整的 18S rRNA 基因、ITS1 序列、5.8S rRNA 基因、ITS2 序列和 25S rRNA 基因。使用 HiC-Pro 对剩余 2 个长片段序列进行分析，发现 tig13 的 5'末端与 tig14 反向互补序列的 3'末端有一段 21 486 bp 的重叠区域，这一连接获得 HiFi 长片段 reads 支持。

从基因组的交联信号热图可以看出，其任何一条染色体内部的交联信号不存在明显的断点(图 1A)，说明基因组的框架组装是正确的。使用软件 BUSCO 对灵芝 13-5 基因组组装完整性进行评估。在担子菌 1 764 个直系同源物中，发现 hap1 完整同源基因为 1 675 个(图 1B，94.9%)，其中单拷贝同源基因为 1 664 个(94.3%)，多拷贝同源基因为 11 个(0.6%)，这一结果证明了 hap1 基因组组装的高准确性和完整性。

2.2 着丝粒的特征

Hi-C 分析发现(图 1A)，灵芝 13-5 基因组中各条染色体的着丝粒之间具有较强的交联信号。根据此特征，找出基因组中各条染色体着丝粒的具体位置(图 2)：Chr04、Chr11、Chr12 和 Chr13 4 条染色体的着丝粒接近端粒区；除 Chr05 的着丝粒位于染色体中间区域，其余 8 条染色体的着丝粒均偏向染色体一端。

高通量染色体三维构象捕获技术无法确定染色体着丝粒的起始和终止位置。为了分析着丝粒的共有特征，截取着丝粒交联信号最强的 40 kb bin 及其前后各 40 kb 序列。13 条染色体着丝粒的重复序列所占比例差异较大，介于 29.50%–91.55%之间；除 Chr04 和 Chr09 外，其余

染色体着丝粒重复序列比例均超过 50% (图 3)。Chr06 着丝粒包含重复序列基序最少，为 51 个，Chr01 最多，高达 108 个(表 2)。灵芝 13-5 基因组中共预测到 272 个重复序列 rnd-1_family-0 家族，其中 116 个落在着丝粒区域中，占比

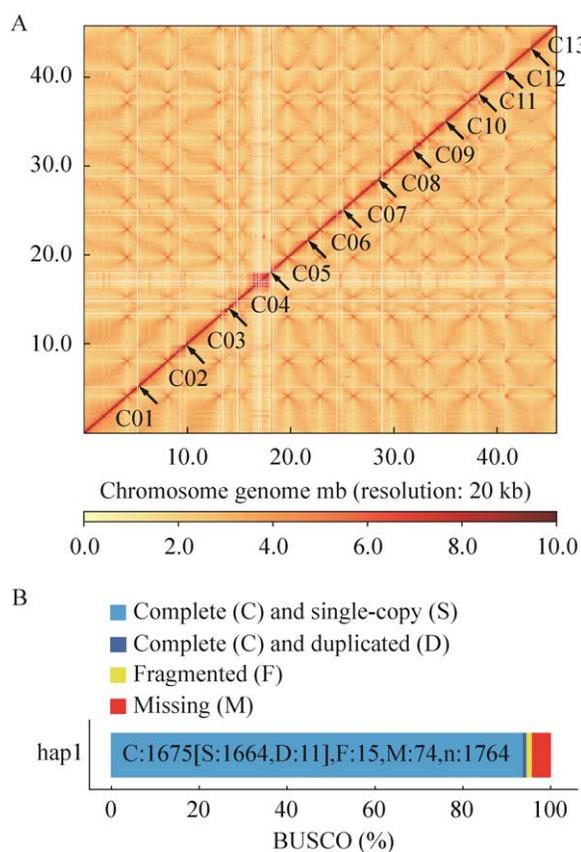


图 1 灵芝 13-5 基因组完整性和准确性评估结果 A: Hi-C 交联热图；图中颜色从浅到深表示互动程度的增加，颜色越深互动越强；横坐标及纵坐标表示其在基因组上的 N * bin 位置；B: BUSCO 评估结果；C: 完整同源基因；S: 单拷贝同源基因；D: 多拷贝同源基因；F: 片段基因；M: 缺失的基因

Fig. 1 Evaluation results of genome completeness and accuracy of *Ganoderma lingzhi* 13-5. A: Heatmap of the Hi-C crosslinking. The color in the figure increases with the intensity of interaction; The coordinate and ordinate indicate the N * bin position on the genome; B: Results of the BUSCO assessment. C: Intact orthologous genes; S: Single copy orthologous genes; D: Multiple copy orthologous genes; F: Fragmented genes; M: Missing genes.

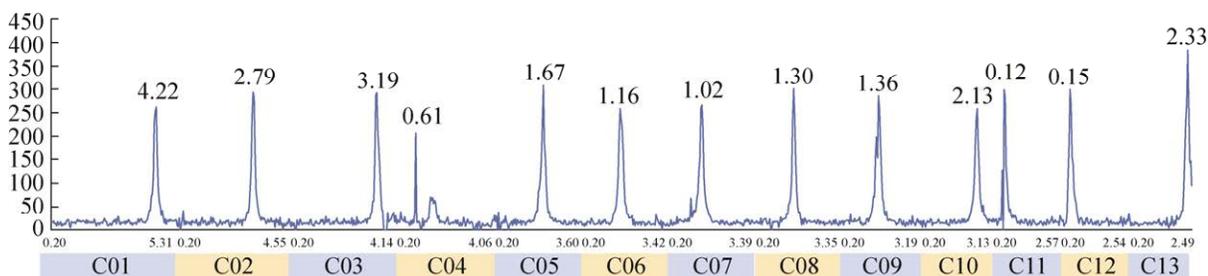


图 2 基因组各条染色体着丝粒位置信息 横坐标：每条染色体的长度；纵坐标：Hi-C 互作值，数值越高表示交联信号越强

Fig. 2 Location information of centromere of each chromosome in the genome. Abscissa: Length of each chromosome; Ordinate: Hi-C interaction value, higher value indicates stronger cross-link signal.

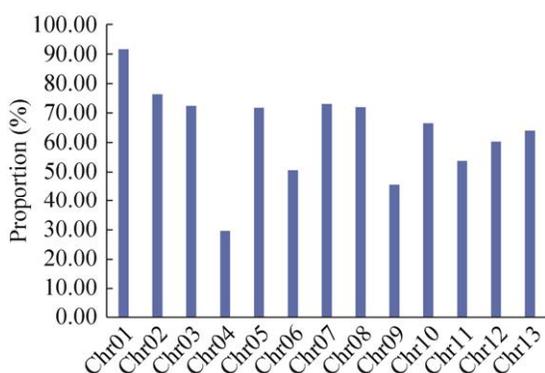


图 3 染色体着丝粒重复序列占比
Fig. 3 Proportion of centromere repeats.

42.65%。rnd-1_family-0、rnd-3_family-15、rnd-1_family-8 家族除了在 Chr04 着丝粒中不存在，成簇坐落于其他每一个着丝粒中，数量为 1–17 个。此外，携带有 rnd-1_family-2 家族的着丝粒有 11 个；携带 rnd-4_family-1322 家族的着丝粒有 10 个；携带有 rnd-1_family-20、rnd-4_family-276 家族的着丝粒均有 9 个；携带有 rnd-1_family-3、GA-rich、rnd-1_family-146、rnd-1_family-1 家族的着丝粒均有 8 个；携带有 (CCTCAT)_n 家族的着丝粒均有 7 个。

2.3 基因预测

基因组共预测 13 383 个基因，其中蛋白表达基因 13 179 个，tRNA 基因 204 个。蛋白表达基因平均长度 1 652.64 bp，平均 CDS 长度 1 323.79 bp，内含子长度为 327.61；平均每个

基因包含 4.81 个外显子，每个外显子平均长度 274.84 bp (表 3)。

基因组各条染色体携带基因密度见图 4，Chr04 每 100 kb 序列平均携带基因数量为 41.47 个，基因密度最大；其余 12 条染色体的基因密度相差不大，介于 30.90–37.07 个，其中 Chr02 每 100 kb 序列平均携带基因数量为 30.90 个，基因密度最小。

2.4 重复序列

基因组中检测到 14 213 个重复序列片段，合计 5.63 Mb，占基因组序列的 12.32% (表 4)。在重复序列中，含有反转座子 4 297 个，合计 4.45 Mb，占比 78.90%；含有 DNA 转座子 1 149 个，合计 0.56 Mb，占比 9.97%；含有串联重复序列 7 881 个，合计 0.55 Mb，占比 9.74%。反转座子 Gypsy 和 LINE 所占比例最大，分别为 31.41%和 12.72%。

灵芝 13-5 基因组各条染色体重复序列含量 (图 5)，Chr04 重复序列占比最大，为 23.89%，其次为 Chr09，占比为 17.62%；Chr03 的重复序列占比最少，为 7.60%；其余 10 条染色体的重复序列占比相差不大，在 8.50%–12.47% 之间。

2.5 SNP calling 结果分析

以单核化菌株灵芝 13-5 基因组 (灵芝 13 的 A 核基因组) 为参考基因组，对 5 个重复菌株的二代重测序数据进行 SNP calling，去重合并后获

表 2 基因组各条染色体着丝粒特征

Table 2 Centromere characteristics of each chromosome in the genome

| 重复序列家族 The family of repeat sequences | 1号 染色 体 C01 | 2号 染色 体 C02 | 3号 染色 体 C03 | 4号 染色 体 C04 | 5号 染色 体 C05 | 6号 染色 体 C06 | 7号 染色 体 C07 | 8号 染色 体 C08 | 9号 染色 体 C09 | 10号 染色 体 C10 | 11号 染色 体 C11 | 12号 染色 体 C12 | 13号 染色 体 C13 | 统计 Total |
|--|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------|
| rnd-1_family-0 | 13 | 11 | 14 | 0 | 13 | 8 | 8 | 14 | 6 | 8 | 3 | 7 | 11 | 116 |
| rnd-3_family-15 | 14 | 9 | 8 | 0 | 10 | 1 | 9 | 17 | 6 | 6 | 8 | 10 | 13 | 111 |
| rnd-1_family-8 | 17 | 3 | 7 | 0 | 9 | 2 | 2 | 8 | 10 | 2 | 4 | 14 | 7 | 85 |
| rnd-4_family-276 | 5 | 0 | 7 | 0 | 0 | 0 | 7 | 6 | 3 | 6 | 2 | 6 | 6 | 48 |
| rnd-1_family-2 | 16 | 4 | 2 | 0 | 1 | 3 | 2 | 3 | 0 | 3 | 3 | 2 | 7 | 46 |
| rnd-4_family-1322 | 3 | 0 | 2 | 0 | 1 | 0 | 4 | 4 | 1 | 2 | 2 | 3 | 3 | 25 |
| rnd-1_family-20 | 3 | 1 | 2 | 0 | 0 | 0 | 1 | 3 | 2 | 3 | 4 | 1 | 0 | 20 |
| rnd-1_family-3 | 3 | 1 | 3 | 0 | 1 | 0 | 0 | 3 | 0 | 1 | 2 | 0 | 3 | 17 |
| rnd-1_family-146 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 4 | 0 | 0 | 2 | 3 | 14 |
| rnd-1_family-1 | 0 | 2 | 1 | 0 | 0 | 1 | 2 | 2 | 3 | 0 | 0 | 1 | 1 | 13 |
| (CCTCAT) _n | 2 | 3 | 0 | 0 | 1 | 2 | 0 | 2 | 0 | 1 | 0 | 2 | 0 | 13 |
| GA-rich | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 1 | 11 |
| Total number | 108 | 60 | 57 | 65 | 60 | 51 | 58 | 81 | 62 | 71 | 81 | 70 | 74 | 898 |
| Percent (%) | 2.07 | 2.01 | 2.10 | 0.87 | 2.39 | 1.77 | 2.58 | 2.58 | 1.71 | 2.55 | 2.51 | 2.84 | 3.07 | |

注: 统计每条染色体着丝粒区域 120 kb 序列信息; The family of repeat sequences: 显示同时出现在 6 条染色体着丝粒区域的 family; Total number: 统计 120 kb 区域的 family 总数; Percent: 着丝粒重复序列长度占对应染色体长度比值

Note: 120 kb sequence information of centromere region of each chromosome was collected. The family of repeat sequences: Family that appear simultaneously in the centromere region of 6 chromosomes. Total number: The total number of family in the 120 kb area. Percent: The ratio of centromere repeat sequence length to corresponding chromosome length.

表 3 基因预测结果

Table 3 Results of genome prediction

| 基因预测结果 Results of genome prediction | 基因组大小 Genome size (bp) |
|---|---------------------------|
| 预测基因 Predicted genes | 13 383 |
| 蛋白质编码基因 Protein coding genes | 13 179 |
| tRNA 基因 tRNA genes | 204 |
| 平均基因长度 Average gene length | 1 652.64 |
| 平均 CDS 长度 Average CDS length | 1 323.79 |
| 每个基因的平均外显子数 Average exons per gene (个) | 4.81 |
| 平均外显子长度 Average exon length | 274.84 |
| 平均内含子长度 Average intron length | 327.61 |

得 379 887 个 SNP 位点, 其中稳定出现在 5 个菌株中的 SNP 位点有 319 074 个(图 6)。这些 SNP 位点以灵芝 13 的 A 核基因组为参考序列, 采用异核菌株的重测序分析获得, 即为 B 核基因组与 A 核基因组存在差异的单碱基多态性位点, 称为异核等位差异位点。我们将灵芝 13 的 319 074 个异核等位差异位点作为一个分子标

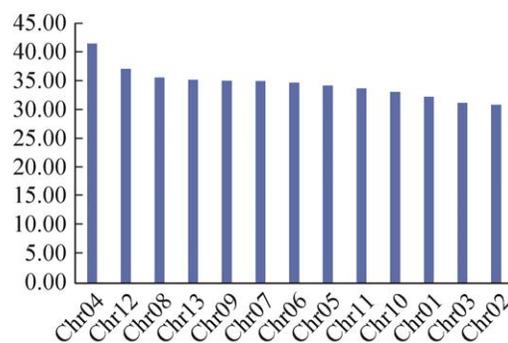


图 4 基因组染色体平均基因密度分布图

Fig. 4 Genome chromosome average gene density distribution map.

记库, 作为评判其他菌株与灵芝 13 是否为相同菌株的工具。

将稳定出现在 5 个菌株中的 319 074 个 SNP 位点作为基准, LZ1 (重复菌株 1) 有 361 003 个 SNP 位点, 与基准相差 41 929 个 SNP 位点; LZ2 (重复菌株 2) 有 347 865 个 SNP 位点, 与基准相差 28 791 个位点; LZ3 (重复菌株 3) 有

表 4 基因组重复序列信息

Table 4 Information about genome repeat sequence

| 分类 | | 数量 | 长度 | 占比 | |
|--------------------------|---------------------------|-------------|-------------|----------|------|
| Classification | | Number | Length (Mb) | Per. (%) | |
| 重复序列片段 | Total repeat fraction | 14 213 | 5.63 | 12.32 | |
| 分类 I: 反转录因子 | 统计 Total | 4 297 | 4.45 | 9.72 | |
| Class I: Retroelement | LTR 反转录转座子 | 统计 Total | 2 106 | 3.40 | 7.44 |
| | LTR Retrotransposon | Ty1/Copia | 276 | 0.58 | 1.28 |
| | | Ty3/Gypsy | 670 | 1.77 | 3.87 |
| | | Others | 1 160 | 1.05 | 2.29 |
| | | 统计 Total | 1 283 | 0.75 | 1.64 |
| | 非 LTR 反转录转座子 | 统计 Total | 1 039 | 0.72 | 1.57 |
| | non-LTR Retrotransposon | LINE | 244 | 0.03 | 0.07 |
| | SINE | 908 | 0.30 | 0.65 | |
| 未分类的反转座子 | Unclassified retroelement | 908 | 0.30 | 0.65 | |
| 分类 II: DNA 转座子 | 统计 Total | 1 149 | 0.56 | 1.23 | |
| Class II: DNA Transposon | TIR | CMC | 57 | 0.06 | 0.12 |
| | | Tc1/Mariner | 63 | 0.05 | 0.12 |
| | | Others | 966 | 0.40 | 0.87 |
| | | 统计 Total | 1 149 | 0.56 | 1.23 |
| 串联重复序列 | Tandem repeats | 7 881 | 0.55 | 1.20 | |
| 未知 | Unkown | 169 | 0.14 | 0.31 | |

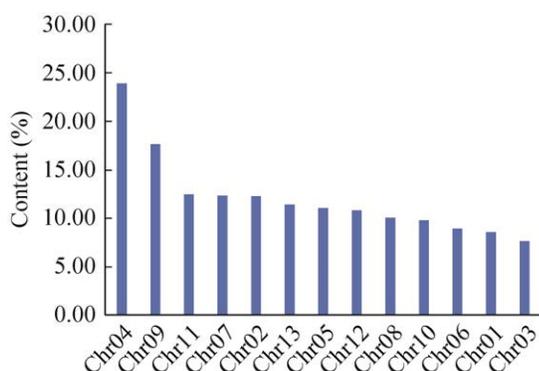


图 5 基因组染色体重复序列含量分布图

Fig. 5 Distribution of chromosome repeat content in the genome.

354 173 个 SNP 位点，与基准相差 35 099 个位点；LZ4 (重复菌株 4) 有 355 723 个 SNP 位点，与基准相差 36 649 个位点；LZ5 (重复菌株 5) 有 332 000 个 SNP 位点，与基准相差最少，为 12 926 个位点。设定与基准相差最多的 41 929 个 SNP 位点($41\ 929/361\ 003 \times 100\% \approx 11.61\%$)，也就是以 LZ1 与基准共有的等位差异位点数作为判断阈值($319\ 074/361\ 003 \times 100\% \approx 88.39\%$)，若两菌株间的 $GS \leq 88.39\%$ ，则认定为“不同品种”；若两菌株间的 $GS > 88.39\%$ 时，则认定为

“相同品种”。将 5 个菌株 SNP calling 的结果绘制成密度曲线图，以第一条染色体为例(图 7)，可以看出这 5 个菌株几乎完全一致，说明比较 SNP 密度曲线走势也是判断两菌株是否相同的方法之一。

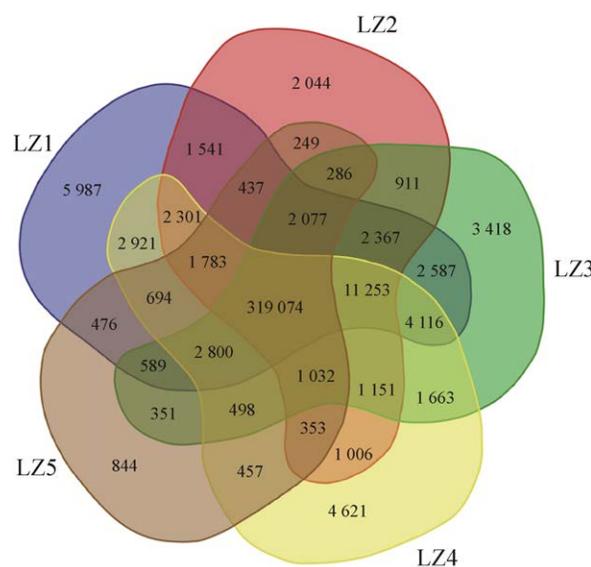


图 6 5 个重复菌株之间的韦恩图 LZ1-5: 重复菌株 1-5

Fig. 6 Venn diagram between five repeating strains. LZ1-5: Duplicate strain 1-5.

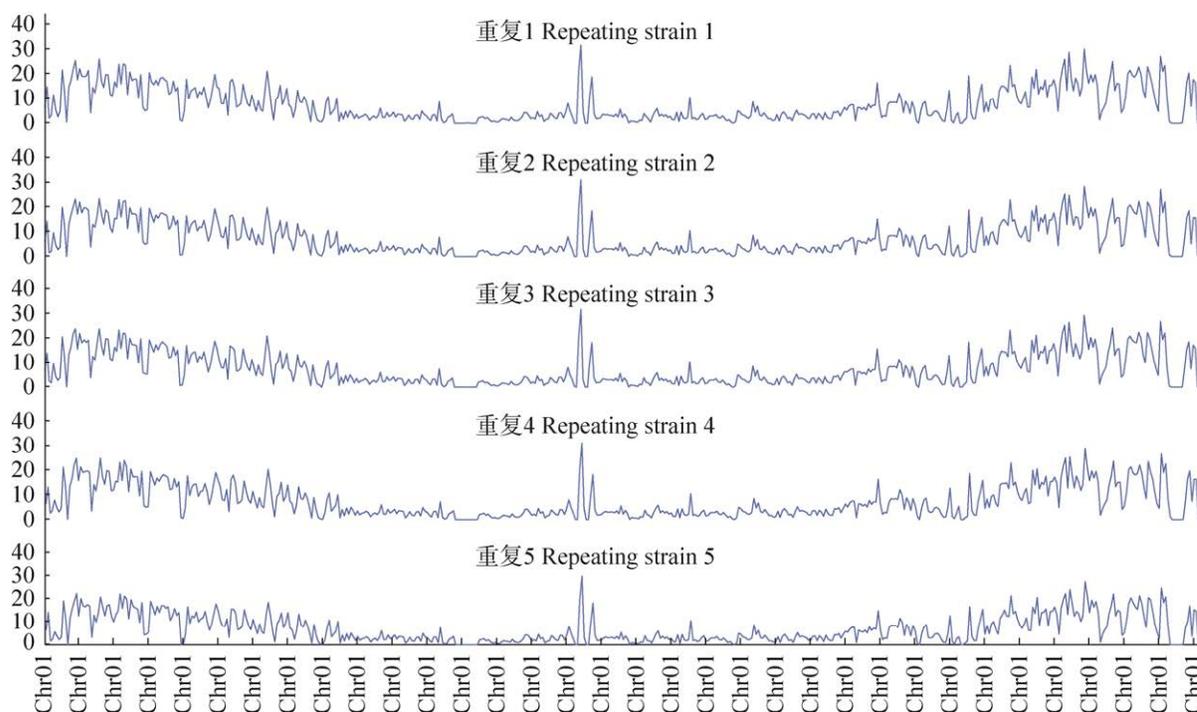


图 7 5 个重复菌株第一条染色体的 SNP calling 结果图

Fig. 7 SNP calling results of the first chromosome of five repeating strains.

2.6 全基因组分子标记的验证

从实验室菌种库中随机抽取 7 个灵芝菌株并进行二代基因组测序，以单核化菌株灵芝 13-5 基因组为参考进行 SNP calling，将获得的 SNP 位点信息逐一与灵芝 13 的异核等位差异位点库进行比对，计算出各个菌株与灵芝 13 共有等位差异位点数占库容比值(表 5)。

表 5 灵芝 13 与 7 个待测菌株共有的等位差异位点数

Table 5 Number of allelic difference sites shared between *Ganoderma lingzhi* 13 and seven post-fruiting strains to be tested

| 待测菌株编号 The strain to be tested | 共有等位差异位点数 Number of shared allelic difference loci | 占比 Proportion (%) |
|-----------------------------------|---|----------------------|
| SL9 | 75 174 | 23.56 |
| G.10069 | 75 429 | 23.64 |
| 日芝 <i>G. japonicum</i> | 144 157 | 45.18 |
| G.10051 | 76 450 | 23.96 |
| G.10064 | 70 834 | 22.20 |
| G.10081 | 76 865 | 24.09 |
| 杂交子 12 Hybridons 12 | 89 500 | 28.05 |

G.10064 与灵芝 13 的共有等位差异位点数最少，为 70 834 个，占比 22.20%，日芝与灵芝 13 的共有等位差异位点数最多，为 144 157 个，占比 45.18%，远低于同一品种(菌株)的阈值(88.39%)，说明它们均不属于灵芝 13 这一菌种。这 7 个菌株的 SNP 密度曲线走势与灵芝 13 均具有较大差异(图 8)，也说明灵芝 13 与这 7 个菌株不是同一个菌株。

3 讨论

本研究通过对优良单核菌株灵芝 13-5 进行三代 HiFi 测序，辅助 Hi-C 数据进行基因组组装。基因组大小为 45.73 Mb，共包含 13 条染色体，大小在 2.49–5.31 Mb 之间。对组装好的基因组进一步分析发现，着丝粒可以分布在近端粒区，或染色体中间，但大多偏向染色体的一端。着丝粒中含有大量重复序列是其特征之一，13 条染色体着丝粒中的重复序列占比介于 29.50%–91.55%之间。整个基因组中共检测到

14 213 个重复序列片段, 合计 5.63 Mb, 占基因组序列的 12.32%。在重复序列中, 含有反转座子 4 297 个(78.90%); 含有 DNA 转座子 1 149 个(9.97%); 含有串联重复序列 7 881 个(9.74%)。使用 funannotate 软件对基因组进行预测, 共预测到 13 383 个基因, 其中蛋白表达基因 13 179 个, tRNA 基因 204 个。

基于组装好的灵芝 13-5 完整基因组序列, 构建了全基因组分子标记。5 个重复菌株通过 SNP calling 去重合并后, 灵芝 13 中 A 核与 B 核不同的 SNP 位点(即异核等位差异位点)共有 319 074 个。将灵芝 13 的 319 074 个异核等位差异位点作为一个分子标记库, 作为评判其他菌株与灵芝 13 是否为相同菌株的工具。将待测菌株与灵芝 13 共有的等位差异位点数作为分子, 分子标记库中等位差异位点的总数作为分母, 计算出比值。因此只要待测菌株有一个核与灵芝 13 的核来源不一致, 两者共有的等位差异位点数就会很少。设定与基准相差最多的 41 929 个 SNP 位点($41\,929/361\,003 \times 100\% \approx 11.61\%$), 即以

LZ1 与基准共有的等位差异位点数作为判断阈值($319\,074/361\,003 \times 100\% \approx 88.39\%$), 若两菌株间的 $GS \leq 88.39\%$, 则认定为“不同品种”; 若两菌株间的 $GS > 88.39\%$ 时, 则认定为“相同品种”。

灵芝 13 整个基因组的大小为 45.73 Mb, SNP calling 出的 319 074 个位点(有差异的部分), 仅占基因组的 0.70%, 即其他 99.30% 的区域都是相似的。因此从整个基因组来看, 11.61% 是在基因组 0.70% 的基础上有差异, 实际上在整个基因组里有差异的部分仅占 0.08%, 菌株间的相似度是很高的。本研究在除去相似部分的基础上进行计算, 因此以 88.39% 作为阈值是可靠的。

目前我国育种者的权益还不能很好地受到保护, 严重打击了育种者的热情。基于全基因组建立的分子标记法在灵芝 13 进行新品种认定时可以提供理论支持, 证明其是新品种; 在后续遇到侵权问题时, 也可将“侵权菌株”进行二代重测序, SNP calling 后与分子标记库进行比对, 说明二者是否为同一菌株(图 9)。

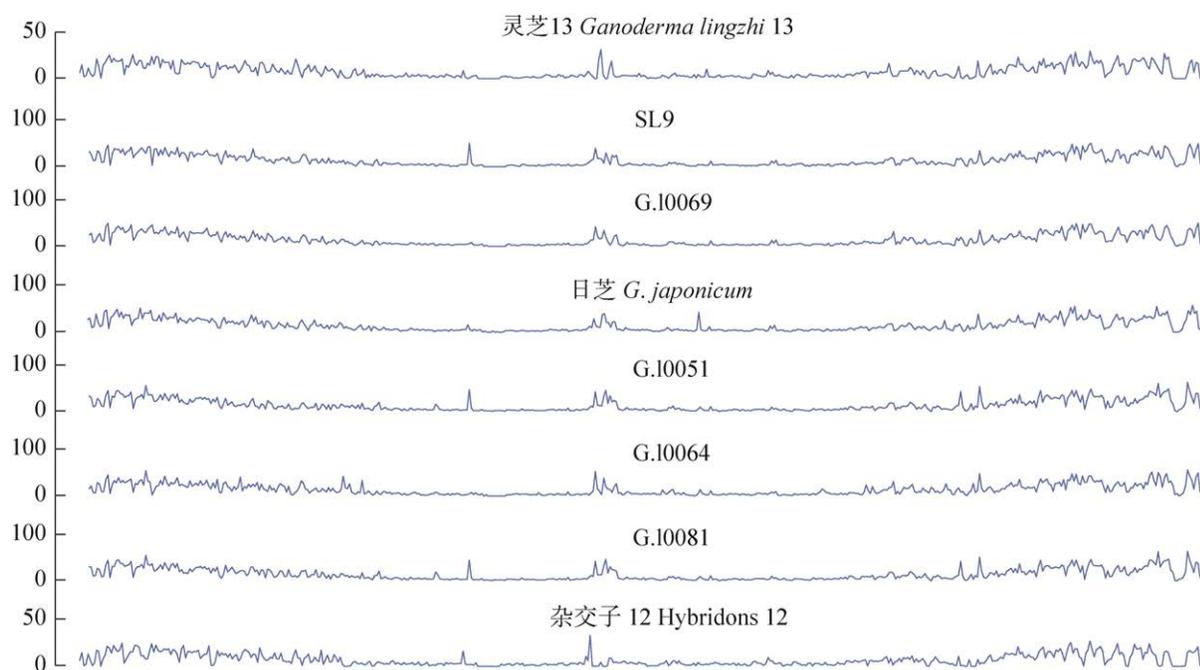


图 8 灵芝 13 与 7 个菌株第一条染色体的 SNP calling 结果图

Fig. 8 SNP calling results of the first chromosome of *Ganoderma lingzhi* 13 and seven post-fruiting strains.

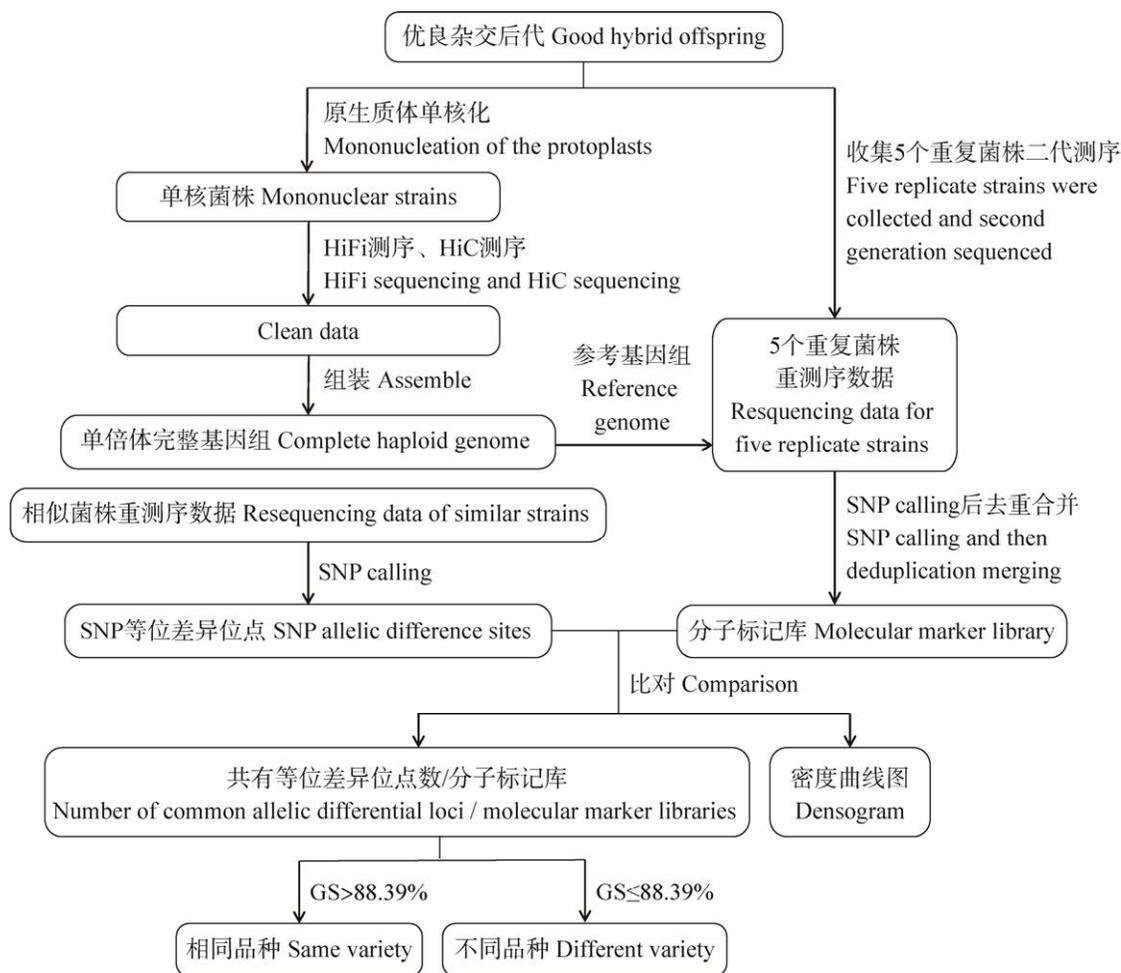


图9 全基因组分子标记法的构建及应用

Fig. 9 Construction and application of genome-wide molecular markers.

作者贡献

刘逸婷：论文构思、数据处理与分析、实验、撰写；蒋晓涵：实验、审核与编辑写作；杨春艳：实验、审核与编辑写作；陈健辉：实验、数据处理；王车昭：实验、软件分析；吕晓萌：实验、审核与编辑写作；杨治康：实验；邓优锦：数据处理、软件分析、论文构思；吴小平：提供实验材料和菌种、论文构思。

利益冲突

作者声明，该研究不存在任何潜在利益冲突的商业或财务关系。

[REFERENCES]

- Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT, 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12): 1691-1692
- Benson G, 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2): 573-580
- Cai XL, He W, An FQ, 2016. Research progress on *Ganoderma lucidum* germplasm resources. *Modern Agricultural Science and Technology*, 2016(6): 99-100 (in Chinese)
- Chen NS, 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 5(1): 4-10
- Chen TQ, Wu JG, Kan YJ, Yang C, Wu YB, Wu JZ, 2018. Antioxidant and hepatoprotective activities of crude polysaccharide extracts from Lingzhi or Reishi medicinal mushroom, *Ganoderma lucidum* (Agaricomycetes), by ultrasonic-circulating extraction. *International Journal of*

- Medicinal Mushrooms, 20(6): 581-593
- Cheng HY, Concepcion GT, Feng XW, Zhang HW, Li H, 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2): 170-175
- Dai YC, Yang ZL, Cui BK, Wu G, Yuan HS, Zhou LW, He SH, Ge ZW, Wu F, Wei YL, Yuan Y, Si J, 2021. Diversity and systematics of the important macrofungi in Chinese forests. *Mycosystema*, 40: 770-805
- Dan XL, Liu WL, Wong J, Tzi BN, 2016. A ribonuclease isolated from wild *Ganoderma lucidum* suppressed autophagy and triggered apoptosis in colorectal cancer cells. *Frontiers in Pharmacology*, 7: 217
- Li J, 2007. Analysis of intellectual property protection and current situation of edible fungi in China. *Chinese Inventions and Patents*, 2007(8): 48-49 (in Chinese)
- Liu XL, Zhong C, Xie J, Hou FF, Dai JM, Zhang SH, Jin J, 2022. Development status and perspective of new varieties protection and DUS test guide for medicinal and edible fungi. *Chinese Traditional and Herbal Drugs*, 53(4): 1173-1180 (in Chinese)
- Qi AH, Sun YX, Li W, Che XL, Lian F, 2018. The medicinal value of *Ganoderma lucidum*. *Rural Economy and Science and Technology*, 29(10): 141-142 (in Chinese)
- Seppy M, Manni M, Zdobnov EM, 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962: 227-245
- Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E, 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, 16: 259
- Soccol CR, Bissoqui LY, Rodrigues C, Rubel R, Sella SRBR, Leifa F, de Souza VLP, Soccol VT, 2016. Pharmacological properties of biocompounds from spores of the Lingzhi or Reishi medicinal mushroom *Ganoderma lucidum* (Agaricomycetes): a review. *International Journal of Medicinal Mushrooms*, 18(9): 757-767
- Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong CC, Upadhyaya NM, Mago R, Hu YH, Figueroa M, Singh KB, Stone EA, Schwessinger B, Wang MB, Taylor JM, Dodds PN, 2021. The stem rust fungus *Puccinia graminis* f. sp. *tritici* induces centromeric small RNAs during late infection that are associated with genome-wide DNA methylation. *BMC Biology*, 19(1): 203
- Sun YY, 2019. On China's independent intellectual property rights protection along the Belt and Road line. *Guangxi Quality Supervision Guide*, 2019(3): 207-208 (in Chinese)
- Tan XY, Sun JS, Ning HJ, Qin ZF, Miao YX, Sun T, Zhang XQ, 2018. De novo transcriptome sequencing and comprehensive analysis of the heat stress response genes in the basidiomycetes fungus *Ganoderma lucidum*. *Gene*, 661: 139-151
- Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert JP, Noble WS, 2015. Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Research*, 43(11): 5331-5339
- Wasser SP, 2017. Medicinal mushrooms in human clinical studies. Part I. Anticancer, oncoimmunological, and immunomodulatory activities: a review. *International Journal of Medicinal Mushrooms*, 19(4): 279-317
- Zhang QY, Li WX, Yang J, Sheng LZ, Zhu SR, Zhu XK, 2020. Current situation and basic countermeasures for the protection of new varieties of food and medicinal mushrooms in China—taking the regulations for the protection of new varieties of plants as the legal source. *Edible and Medicinal Mushrooms*, 28(2): 98-102 (in Chinese)
- Zuo HB, Xia BY, Wang XL, Li HL, Su XL, Liu HF, 2021. Investigation on germplasm resources and hybrid breeding in *Ganoderma lucidum* region. *Modern Agricultural Research*, 27(3): 109-110 (in Chinese)

[附中文参考文献]

- 才晓玲, 何伟, 安福全, 2016. 灵芝种质资源研究进展. *现代农业科技*, 2016(6): 99-100
- 戴玉成, 杨祝良, 崔宝凯, 吴刚, 袁海生, 周丽伟, 何双辉, 葛再伟, 吴芳, 魏玉莲, 员瑗, 司静, 2021. 中国森林大型真菌重要类群多样性和系统学研究. *菌物学报*, 40: 770-805
- 李绩, 2007. 我国食用菌菌种知识产权保护 and 现状分析. *中国发明与专利*, 2007(8): 48-49
- 刘晓柳, 钟灿, 谢景, 侯凤飞, 戴甲木, 张水寒, 金剑, 2022. 药食用菌新品种保护及 DUS 测试指南研制现状与展望. *中草药*, 53(4): 1173-1180
- 戚爱华, 孙艳霞, 李威, 车晓蕾, 连丰, 2018. 灵芝的药用价值. *农村经济与科技*, 29(10): 141-142
- 孙缘缘, 2019. 论我国在“一带一路”沿线的自主知识产权保护. *广西质量监督导报*, 2019(3): 207-208
- 张清洋, 李文学, 杨军, 盛立柱, 朱姝蕊, 朱星考, 2020. 我国食药用品菌新品种保护现状与基本对策——以植物新品种保护条例为法律渊源. *食药用品菌*, 28(2): 98-102
- 左洪波, 夏伯阳, 王夕亮, 李华利, 苏小林, 刘红粉, 2021. 灵芝产区种质资源调查及杂交育种研究. *现代农业研究*, 27(3): 109-110