

韩国赤芝全基因组重测序分析*

朱风丽¹, 徐晓兰², 陈体强³, 石林春⁴, 缪晓青^{2**}, 兰 进⁴

(1. 福建农林大学食品科学学院 福州 350002; 2. 福建农林大学蜂学学院 福州 350002;
3. 福建省农业科学院食用菌研究所 福州 350014;
4. 中国医学科学院北京协和医学院药用植物研究所 北京 100193)

摘要:目的:探索韩国赤芝与中国赤芝CGMCC5.0026间在基因组水平上存在的差异,为赤芝的遗传育种提供研究基础。方法:采用高通量重测序技术,对韩国赤芝进行全基因组重测序,并对单核苷酸多态位点突变(SNP)、小片段插入缺失变异(InDel)、结构变异(SV)进行检测和注释。对DNA水平变异基因进行KEGG、GO、COG、NR、SwissProt数据库注释。结果:韩国赤芝重测序覆盖深度为44×,与中国赤芝CGMCC5.0026相比,共发现10607个基因发生非同义SNP,4774个InDel和1428个SV,共导致9469个基因发生变异。这些变异基因中有转录因子86个,细胞色素P450 195个。结论:本研究通过对中国赤芝和韩国赤芝基因组序列进行比较,为赤芝的分子标记育种和对完善灵芝模式真菌研究体系提供了基础。

关键词:菌丝生长 全基因组重测序 单核苷酸多态性(SNP) 小片段插入缺失(InDel) 结构变异(SV) 转录因子 细胞色素P450

doi:10.11842/wst.2019.04.032 中图分类号:R917 文献标识码:A

灵芝(*Ganoderma*)也被称为“不朽的蘑菇”、“传统中药的象征”,是著名的药用真菌之一。《中华人民共和国药典》(2015年版)中收录赤芝(*Ganoderma lucidum*)和紫芝(*G. sinense*)为灵芝的药材来源。灵芝能够产生大量的生物活性物质,现有400多种化合物被鉴定,主要是三萜和多糖类物质。灵芝种类较多,栽培种主要为赤芝^[1]。近年来,由于野生灵芝资源的破坏较为严重,使得具有重要药用价值的野生灵芝种质资源濒临灭绝。我国灵芝的主栽品种部分来自韩国和日本等地,由于与引进品种产地的气候环境相差较大,菌种的长时间栽培导致品种衰退,出现产量下降、品质降低、抗病虫及杂菌能力较差等缺点。因此应大力推动自主选育的适于我国生长的优良品种^[2]。目前选育技术方法传统、育种周期长、遗传改良效率偏低。2011年,赤芝CGMCC5.0026基因组测序的完成推进了赤芝基因

组学的研究进程^[3],使赤芝在基因组范围内进行遗传分析成为可能,为分子标记辅助育种提供技术手段,这为赤芝新品种的选育打下基础。

近年来,随着测序技术的发展及许多物种基因组测序的不断完成,重测序技术也得到了广范的应用,利用全基因组重测序技术可以在全基因组范围内挖掘单核苷酸多态性(SNP)、插入缺失标记(InDel)、结构变异(SV)和基因拷贝数变异(CNV),并广泛应用于变异检测、遗传图谱构建、性状定位和群体进化研究等^[4]。Shin等^[5]将韩牛重测序数据分别与UMD 3.1和Btau 4.6.1两个牛的参考基因组进行比对,分别发现了46 301和28 613个SNP,这些SNP是韩牛特有的可能与产奶机制、多汁性以及黄色的牛毛有关,同时发现了大多数的SNP位于启动子区域,这表明SNP调控了与性状相关基因的表达。Zhang等人^[6]通过全基因组重测序技术,将芥菜中与花色相关的基因*BjPC2*定位到了染色体B04上。Varshney等人^[7]将包含选育品系、地方品

收稿日期:2018-08-07

修回日期:2018-12-26

* 国家自然科学基金委青年基金项目(81503189):灵芝三萜MVA途径关键酶基因的转录因子的筛选及功能验证,负责人:徐晓兰;国家自然科学基金委面上基金项目(81573703):内源生长素对光的响应及在灵芝形态建成中的调控机理,负责人:兰进;福建省科技厅区域发展项目(2018N3014):赤芝优良品种的引种和选育,负责人:徐晓兰。

** 通讯作者:缪晓青,教授,主要研究方向:蜂疗学。

种、野生种等在内的292种木豆进行基因组重测序,在基因组上发现了几个可能与驯化、育种相关的区域。并通过基因组关联分析,确定了一些后选基因与农艺重要性状之间的相关性。Lee等人^[8]将起源于韩国的早花大豆突变株系进行全基因组重测序共筛选出了30个与开花相关的SNPs和25个Indels。

为了研究韩国赤芝菌株的基因组序列情况,应用二代测序的方法对韩国赤芝菌株进行了全基因组重测序。本文将韩国赤芝基因组重测序数据与赤芝CGMCC5.0026(参考基因组)进行比较以揭示其在SNP、InDel、SV存在的变异,旨在发现韩国赤芝菌株特有的序列特征,为研究不同来源的菌株之间的差异提供数据。并且初步探索了与菌丝生长速度和灵芝三萜合成相关的基因变异,为赤芝的分子标记育种及完善药用真菌模式体系提供基础。

1 材料与方法

1.1 材料

赤芝CGMCC5.0026^[3]保存于本实验室,韩国赤芝(韩芝)由中国医学科学院药用植物研究所兰进教授馈赠。两种菌株均为二倍体。

1.2 方法

1.2.1 赤芝菌丝的生物量

将2个赤芝菌株在28℃恒温培养箱中进行活化,将活化好的菌株转接至PDA培养基中,当菌丝生长至直径为5~6 cm时,用打孔器在菌丝边缘处打孔,并将10个直径为1 cm的菌丝块转接至PDA液体培养基中,28℃,180 rpm黑暗下培养8 d,然后用超纯水冲洗至洗脱液为无色。45℃烘干至恒重。每个实验平行3次。

1.2.2 灵芝三萜含量的测定

菌丝的PDA液体培养方法同“1.2.1”项下。将清洗干净的菌丝液氮研磨成细粉后烘干至恒重。根据实验室前期得到的提取灵芝三萜的条件,采用香草醛-高氯酸显色法测定灵芝三萜含量。称取0.05 g左右干燥的灵芝菌丝细粉,加入3 mL 80%的乙醇浸提4 h,然后在温度为60℃下超声提取20 min。将提取液定容至10 mL,取1 mL水浴挥干后加入0.2 mL 5%的香草醛-冰醋酸和0.8 mL的高氯酸溶液,70℃水浴15 min,冷却至室温,最后加入4 mL的冰醋酸,摇匀后在560 nm处测定吸光度值。

1.2.3 韩芝的全基因组重测序

用植物基因组DNA提取试剂盒(天根生化科技

(北京)有限公司)进行DNA的提取。将提取的质量较好的DNA送至北京百迈客生物科技有限公司进行全基因组重测序。首先对提取的基因组DNA进行检测,合格后用超声波法将DNA片段化,然后将片段化的DNA进行片段纯化、末端修复、3'端加A、连接测序接头,然后用琼脂糖凝胶电泳进行片段大小选择,进行PCR扩增以形成测序文库,建好的文库先进行文库质检,质检合格的文库用Xten进行测序。对测序得到的原始reads(双端序列)进行质量评估,并过滤得到Clean Reads,用于后续生物信息学分析。用bwa软件^[9]将Clean Reads与参考基因组序列进行比对,基于比对结果进行SNP、Small InDel、SV的检测和注释,并对DNA水平差异基因进行挖掘与注释分析。

1.2.4 韩芝与赤芝CGMCC5.0026基因组间变异的检测与注释

根据韩芝的Clean Reads在赤芝CGMCC5.0026基因组^[3]上的定位结果,用GATK软件工具包^[10]进行SNP和Small InDel的检测,其中SNP的检测是根据Clean Reads在参考基因组的定位结果,使用Picard(<http://sourceforge.net/projects/picard/>.(Picard))进行去重复(Mark Duplicates)、GATK进行局部重比对(Local Realignment)、碱基质量值校正(Base Recalibration)等预处理,以保证检测得到的SNP准确性,再使用GATK进行SNP的检测,过滤,并得到最终的SNP位点集。并用SnpEff软件^[11]进行SNP和Small InDel变异的注释与预测。使用breakDancer检测SV变异,并对SV进行注释。

1.2.5 DNA水平变异基因的挖掘与功能注释

寻找韩国赤芝与赤芝CGMCC5.0026间发生的非同义突变SNP、CDS区发生的InDel与SV的基因,将这些变异基因通过BLAST^[12]与NR^[13]、SwissProt^[13]、GO^[14]、COG^[15]、KEGG^[16]功能数据库进行比对,并对这些基因的功能进行注释。

1.2.6 RT-PCR检测

将“1.2.1”中培养的菌丝使用TranZol UP(北京全式金)提取RNA,然后用TaKaRa PrimeScript™ RT reagent Kit with gDNA Eraser Kit(购自TaKaRa)反转录成cDNA。根据灵芝全基因组测序的结果,找出本实验中注释到的9个与生长相关的转录因子的CDS序列,使用Primer premier 5软件设计引物,引物序列(表1),以灵芝菌丝cDNA为模板,用SYBR® Premix Ex Taq™ II(Tli RNaseH Plus)(购自TakaRa)进行PCR。

表1 RT-PCR引物序列

引物	序列
GL30073-F	GCCCACGACAGAAACCAGA
GL30073-R	GAAGAAATCATCCACGGTGTATC
GL18755-F	GCAAGGCTCCCGAAAGTA
GL18755-R	ATGGCTGTGCGACGTGTAT
GL27476-F	GACAGTGCCTGCGAATG
GL27476-R	CGAGTTGAGCTGGCGTTT
GL21797-F	GGCTGATGGGTTCTACTGC
GL21797-R	CGTTGTGCGTCCGCTTGT
GL24821-F	TCCGCAAGCAAACCAAGC
GL24821-R	CGGGAAGGGAGCATAAGG
GL27323-F	TACTACCAACAGCTCCTCCG
GL27323-R	GCTTCCGTCCTTACTCG

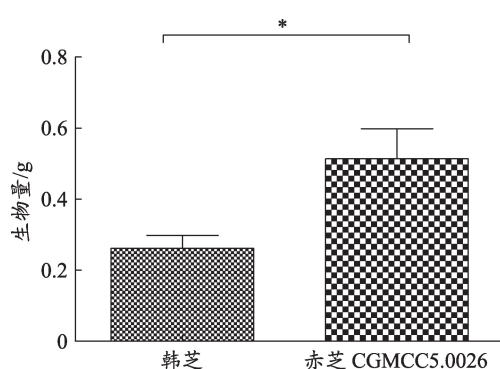


图1 两种菌丝的生物量

注:韩芝与灵芝相比,* $P < 0.05$ 。

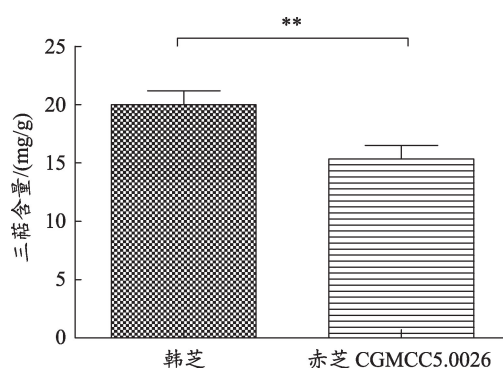


图2 两种菌丝的三萜含量

注:韩芝与灵芝相比,** $P < 0.05$ 。

反应程序为:94℃预变性 30 s,94℃变性 5 s,60℃退火 30 s,共40个循环。荧光定量用甘油醛-3-磷酸脱氢酶(GPD)作为内参基因,基因的相对表达量用 $2^{-\Delta\Delta CT}$ 计算。

1.2.7 数据处理

数据分析使用SPSS17.0,显著性分析采用单因素方差分析和多重比较。 $P < 0.05$ 为差异性显著, $P < 0.01$

为差异性极显著。

2 结果与分析

2.1 两种赤芝菌株的菌丝生物量

液体摇培8天后,两种赤芝的菌丝干重(图1)所示,赤芝CGMCC5.0026的菌丝干重为0.51 g是韩芝的1.97倍,并在统计学上表现为显著性差异。

2.2 三萜含量的比较

对赤芝CGMCC5.0026和韩芝菌丝的三萜含量进行比较,结果表明(图2),韩芝菌丝的三萜含量为20.00 $\text{mg} \cdot \text{g}^{-1}$ 是赤芝CGMCC5.0026的1.30倍,在统计学表现为极显著差异。

2.3 韩芝的全基因组重测序

对重测序的Q30、GC含量、与参考基因组的比对率、平均覆盖深度进行统计,其统计结果(表2)。得到的reads数目有7 679 704个,将重测序结果与赤芝CGMCC5.0026基因组比较,发现与参考基因组的匹配率为87.79%,测序深度为44倍。

2.4 韩芝SNP的检测与注释

韩芝的SNP检测结果表明,总共检测到的SNP数量为291212个,其中SNP的Ti/Tv(转换/颠换)的比值为2.81,其中杂合类型的SNP比例为62.87%。SNP变异的基因数量为13 914,占灵芝预测总基因的86.35%。对SNP的检测结果进行注释,发现,SNP的变异主要发生在CDS区(37.04%),在CDS区中主要的突变类型为同义编码突变(58.57%)和非同义编码突变(40.53%)(表3)。

2.5 韩芝的Small InDel检测与注释

根据样品的Clean Reads在参考基因组上的定位结果,检测韩芝与参考基因组间在CDS区的Small InDel变异,其变异总数为9 458个(表4)。根据检测得到的Small InDel位点在参考基因组上的位置信息,对比参考基因组的基因、CDS位置等信息,对InDel位点进行功能或类型的分类,结果(表5)所示。从表中可以看出,InDel变异多发生在基因的上游(37.31%)、下游(33.43%)区域和CDS区域(12.71%),在CDS区中,发生的变异大多为移码突变(50.58%)。

2.6 韩芝的SV的检测与注释

使用breakDancer对SV进行检测的结果表明,韩芝的结构变异总数为3925,其中,变异最多的为缺失类型变异,其数量为1437(表6)。根据检测到的SV变异在参考基因组上的位置信息,对比参考基因组的基

表2 赤芝5.644 测序数据统计

过滤后的 reads(Clean_Reads)	Q30(%)	GC(%)	定位比/Mapped(%)	平均覆盖深度(Ave_depth)
7679704	89.1	55.98	87.79	44

表3 SNP 注释结果

类型	SNP数目	区域
基因间区 INTERGENIC	826	
基因内 INTRAGENIC	3	
内含子 INTRON	25 877	
基因上游区域(5K 以内)UPSTREAM	77 845	
基因下游区域(5K 以内)DOWNSTREAM	68 487	
基因的 5'UTR 内 UTR_5_PRIME	1 625	
基因的 3'UTR 内 UTR_3_PRIME	3 733	
剪切受体突变(exon 前 2bp 内)SPLICE_SITE_ACCEPTOR	191	
剪切供体突变(exon 后 2bp 内)SPLICE_SITE_DONOR	156	
剪切位点区域突变 SPLICE_SITE_REGION	4 098	
起始密码子获得(非编码区)START_GAINED	257	
无法得到准确判断的突变 Other	248	CDS
同义编码突变 SYNONYMOUS_CODING	63 178	CDS
非同义的起始密码子突变 NON_SYNONYMOUS_START	11	CDS
非同义编码突变 NON_SYNONYMOUS_CODING	43 722	CDS
同义终止密码子突变 SYNONYMOUS_STOP	78	CDS
终止密码子获得 STOP_GAINED	562	CDS
终止密码子丢失 STOP_LOST	168	CDS
起始密码子丢失 START_LOST	147	CDS

表4 Small Indels 的CDS区统计

CDS 插入(CDS-Insertion)	CDS 缺失(CDS-Deletion)	CDS 杂合(CDS-Het)	CDS 纯合(CDS-Homo)	CDS Indel 总数(CDS-Total)
4 625	4 833	3 154	6 304	9 458

表5 Small Indel 注释结果

类型	SNP数目	区域
基因上游区域(5K 以内)UPSTREAM	27 755	
基因内 INTRAGENIC	53	
基因下游区域(5K 以内)DOWNSTREAM	24 870	
剪切供体突变 SPLICE_SITE_DONOR	87	
剪切受体突变 SPLICE_SITE_ACCEPTOR	130	
内含子 INTRON	8 860	
基因间区 INTERGENIC	295	
基因的 3'UTR 内 UTR_3_PRIME	1 206	
基因的 5'UTR 内 UTR_5_PRIME	606	
剪切位点区域突变 SPLICE_SITE_REGION	1 039	
无法得到准确判断的突变 Other	39	
移码突变(非 3 的整数倍插入或删除)FRAME_SHIFT	4 784	CDS
起始密码子丢失 START_LOST	43	CDS
密码子插入(3 的整数倍)CODON_INSERTION	1 636	CDS
终止密码子丢失 STOP_LOST	65	CDS
密码子删除(3 的整数倍)CODON_DELETION	1 277	CDS
终止密码子获得 STOP_GAINED	119	CDS
密码子边界上的 3 的整数倍删除 CODON_CHANGE_PLUS_CODON_DELETION	929	CDS
非密码子边界上的 3 的整数倍的删除 CODON_CHANGE_PLUS_CODON_INSERTION	605	CDS

表6 SV统计

SV总数	插入(INS)	缺失(DEL)	反转(INV)	染色体内异位(ITX)	染色体间异位(CTX)	复杂SV(UN)
3 925	768	1 437	699	217	772	32

表7 SV注释

SV变异类型	外显子区的变异	内含子区的变异	基因间区的变异
DEL	709	56	672
INS	441	56	271
INV	651	1	47
Total	1 801	113	990

因、CDS位置等信息,对韩芝SV变异进行位置的注释,并对缺失(DEL)、插入(INS)、反转(INV)三种类型的结构变异进行注释,其结果(表7)所示,从表中可以看出,SV的变异主要发生在外显子区域。

2.7 韩芝DNA水平变异的分析

在CDS区发生的变异可能会引起基因功能的变化,检测到韩芝与参考基因组间发生的非同义突变SNP、CDS区发生的InDel与SV的基因数量分别为10 607、4 774、1 428。共有9 649个变异基因得到了注释,COG共注释到了4061个变异基因,其中存在较多变异基因的功能类为氨基酸运输和代谢(361个)、次生代谢物的生物合成、运输和分解代谢(352个)、碳水化合物的运输和代谢(315个)(图3)。将得到的9649个变异基

因进行KEGG分析,这些变异基因主要注释到的代谢通路为碳代谢(79个)、氨基酸的生物合成(72个)和RNA转运(70个)(图4)。

这些变异基因中有86个为转录因子,目前对于灵芝中转录因子的研究,分析与灵芝菌丝生长及灵芝三萜含量相关的转录因子 *GlSwi6*^[17]和 *GlPacC*^[18]发现,两个基因CDS区发生的变异类型分别为密码子删除和非同义编码突变。除此之外,从这些变异的转录因子中还发现了4个可能与韩芝菌丝生长相关的候选基因 *PacC/Rim101*、*Zap1*、*Spt8*和 *Bfr2*。这些转录因子的CDS区均发生了非同义编码突变(表8)。

注释为细胞色素P450的变异基因有195个,与已报道的^[3]和羊毛甾醇合酶(LSS)高度相关的78个CYP基因相比,有70个相同的基因,其中有12个CYP512和1个CYP5144,这13个基因在CDS区均发生了变异,变异类型为非同义编码突变、移码突变、终止密码子获得、密码子改变+密码子删除及密码子(3的整数倍)插入(表9)。对基因GL15605的变异进行图示,该基因DNA序列2186核苷酸处发生单碱基突变(图5A),

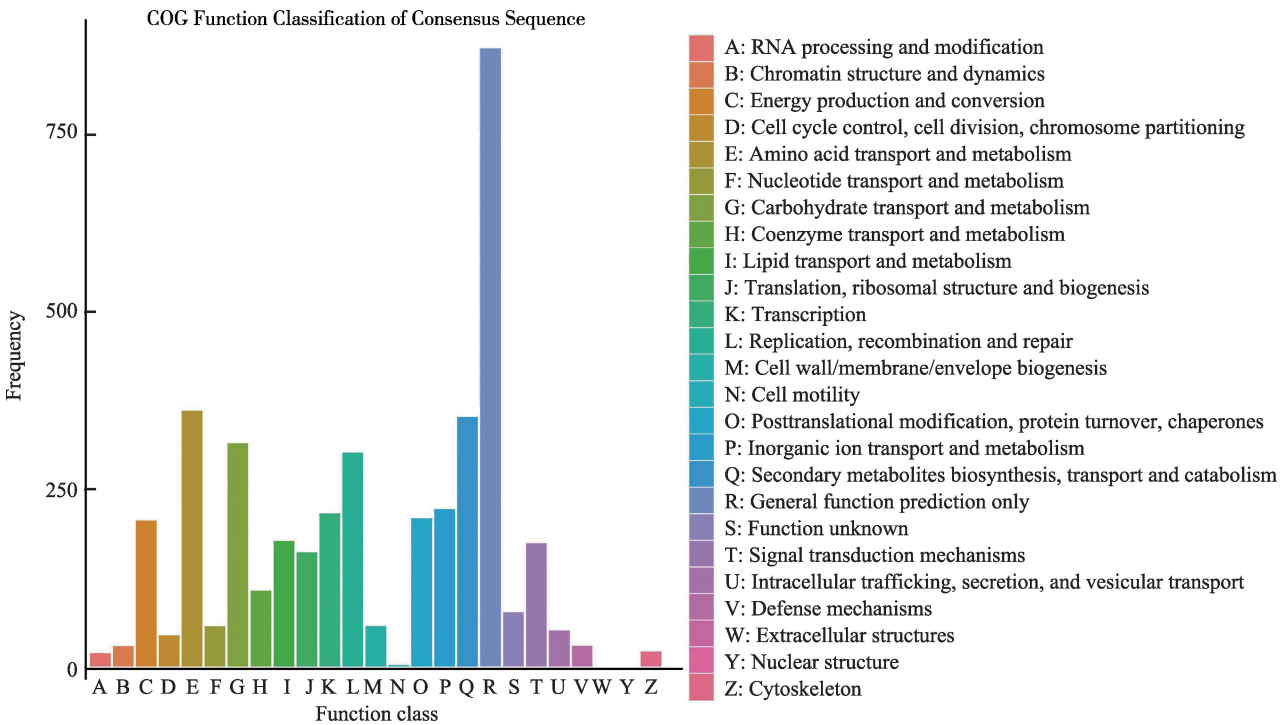


图3 变异基因COG注释图

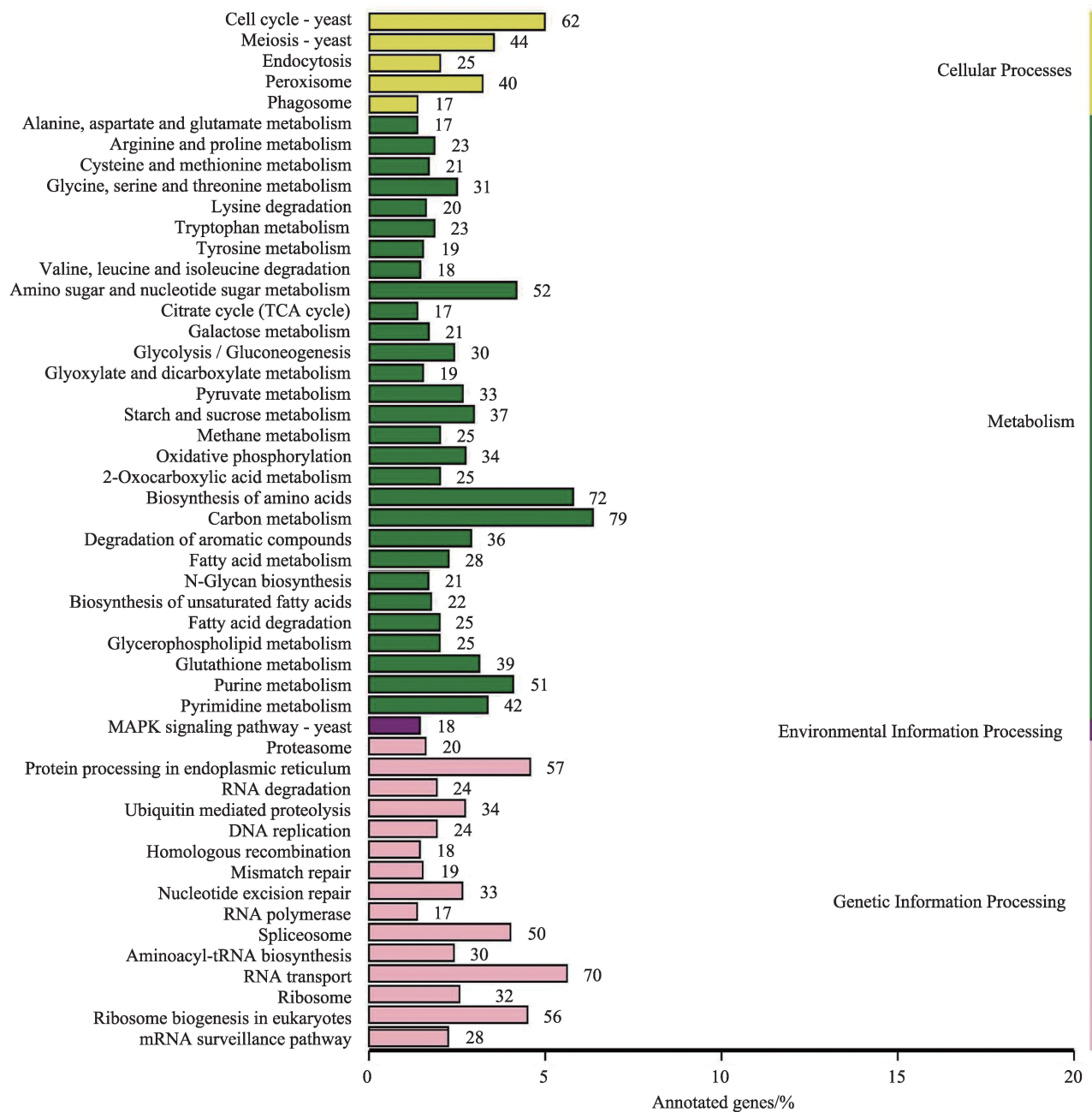


图4 变异基因的KEGG注释图

表8 与生长相关的转录因子的变异

基因名称	注释	变异类型	变异信息
GL30073	GlPacC	密码子删除(3的整数倍)	GaLu96scf_1,3657743 bp, ACGCCCCCGC/A
GL18755	GLSwi6	非同义编码突变	GaLu96scf_17,416178 bp, C/T; GaLu96scf_17,416438 bp, A/G; GaLu96scf_17,416574 bp, T/A
GL27476	PacC/Rim101	非同义编码突变	GaLu96scf_19,777035 bp, C/G; GaLu96scf_19,777106 bp, C/T; GaLu96scf_19,777127 bp, C/T, GaLu96scf_19,777154 bp, A/G; GaLu96scf_19,777177 bp, A/G; GaLu96scf_19,777199 bp, A/G; GaLu96scf_19,777241 bp, G/A; GaLu96scf_19,777336 bp, T/C
GL21797	Zap1	非同义编码突变	GaLu96scf_16,879014 bp, T/C; GaLu96scf_16,879065 bp, C/T
GL24821	Spt8	非同义编码突变	GaLu96scf_4,278409 bp, C/T
GL27323	Bfr2	非同义编码突变	GaLu96scf_17,52616 bp, C/G, T; GaLu96scf_17,52980 bp, T/C; GaLu96scf_17,52987 bp, A/C

表9 细胞色素 P512 和 P5144 家族的变异

基因名称	家族	变异类型	变异信息
GL31772	CYP512A4	非同义编码突变	GaLu96scf_1,3908878 bp, G/A; GaLu96scf_1,3908620 bp, T/C; GaLu96scf_1,3908244 bp, T/C; GaLu96scf_1,3907450 bp, T/A; GaLu96scf_1,3907567 bp, G/A; GaLu96scf_1,3907028 bp, A/G
GL31771	CYP512A3	非同义编码突变	GaLu96scf_1,3903515 bp, C/T; GaLu96scf_1,3902923 bp, A/G; GaLu96scf_1,3903682 bp, G/A; GaLu96scf_1,3903069 bp, G/C; GaLu96scf_1,3902986 bp, A/G; GaLu96scf_1,3903344 bp, T/G; GaLu96scf_1,3903688 bp, A/G; GaLu96scf_1,3904364 bp, G/A; GaLu96scf_1,3903351 bp, G/A; GaLu96scf_1,3902975 bp, C/T; GaLu96scf_1,3903866 bp, A/G; GaLu96scf_1,3904455 bp, A/T; GaLu96scf_1,3902998 bp, A/T
GL31761	CYP512U6	非同义编码突变	GaLu96scf_22,105409 bp, G/T; GaLu96scf_22,103773 bp, T/C; GaLu96scf_22,105132 bp, C/T; GaLu96scf_22,103414 bp, T/C; GaLu96scf_22,105208 bp, C/G
GL23338	CYP512V2	非同义编码突变	GaLu96scf_26,358866 bp, G/A; GaLu96scf_26,357738 bp, C/T; GaLu96scf_26,358260 bp, T/C; GaLu96scf_26,358931 bp, G/T; GaLu96scf_26,358887 bp, C/T; GaLu96scf_26,357808 bp, C/A
GL23109	CYP512A6	非同义编码突变	GaLu96scf_23,392317 bp, C/T; GaLu96scf_23,393543 bp, C/T; GaLu96scf_23,393129 bp, C/G; GaLu96scf_23,393469 bp, A/C; GaLu96scf_23,391533 bp, G/C
GL22911	CYP512U2	非同义编码突变	GaLu96scf_22,132423 bp, A/T; GaLu96scf_22,132416 bp, C/A; GaLu96scf_22,133445 bp, G/C; GaLu96scf_22,132569 bp, C/T; GaLu96scf_22,132736 bp, C/G; GaLu96scf_22,133515 bp, G/T
GL22909	CYP512U5	非同义编码突变	GaLu96scf_22,92097 bp, A/G; GaLu96scf_22,93468 bp, G/A; GaLu96scf_22,93163 bp, G/A; GaLu96scf_22,91702 bp, A/G; GaLu96scf_22,91568 bp, G/A; GaLu96scf_22,93115 bp, G/A
		终止密码子获得	GaLu96scf_22,93531 bp, G/GTTTGAA
GL22480	CYP512Y1	非同义编码突变	GaLu96scf_2,1738987 bp, C/T
GL20660	CYP512T1	非同义编码突变	GaLu96scf_11,774317 bp, A/T; GaLu96scf_11,99461 bp, A/G; GaLu96scf_11,99561 bp, G/A; GaLu96scf_11,101414 bp, A/C
GL20623	CYP512U7	非同义编码突变	GaLu96scf_10,601813 bp, C/T; GaLu96scf_10,602127 bp, C/T; GaLu96scf_10,602928 bp, C/T; GaLu96scf_10,602192 bp, C/T; GaLu96scf_10,601225 bp, T/C; GaLu96scf_10,602666 bp, T/A; GaLu96scf_10,602865 bp, T/C; GaLu96scf_10,602043 bp, A/T; GaLu96scf_10,600881 bp, C/G; GaLu96scf_10,602723 bp, A/G; GaLu96scf_10,602237 bp, C/A; GaLu96scf_10,601591 bp, A/G; GaLu96scf_10,602261 bp, G/C
		密码子插入(3 的整数倍)	GaLu96scf_10,603100 bp, A/ATCGTGAAGCTGT, ATCGTGAAGTTGT; GaLu96scf_10,603068 bp, C/CGCT
		密码子边界上的 3 的整数倍删除	GaLu96scf_10,602830 bp, GGGA/G
		移码突变	GaLu96scf_10,602227 bp, GA/C; GaLu96scf_10,602848bp, A/AC
GL19231	CYP512U4	非同义编码突变	GaLu96scf_24,206806 bp, T/C; GaLu96scf_24,205343 bp, T/C; GaLu96scf_24,205894 bp, A/G; GaLu96scf_24,205579 bp, A/G
		终止密码子获得	GaLu96scf_24,206843 bp, C/A
		移码突变	GaLu96scf_24,205603 bp, A/AGG; GaLu96scf_24,206633 bp, GC/G
		密码子边界上的 3 的整数倍删除	GaLu96scf_24,205249 bp, AGTG/A; GaLu96scf_24,207388 bp, GGCC/G
GL15605	CYP512A2	非同义编码突变	GaLu96scf_15,590612 bp, G/A
GL23174	CYP5144N1	非同义编码突变	GaLu96scf_24,88458 bp, C/G; GaLu96scf_24,88786 bp, G/T; GaLu96scf_24,87416 bp, C/T; GaLu96scf_24,88305 bp, C/T; GaLu96scf_24,89209 bp, G/A; GaLu96scf_24,88437 bp, C/T; GaLu96scf_24,88350 bp, C/G; GaLu96scf_24,87897 bp, A/G; GaLu96scf_24,87274 bp, T/G

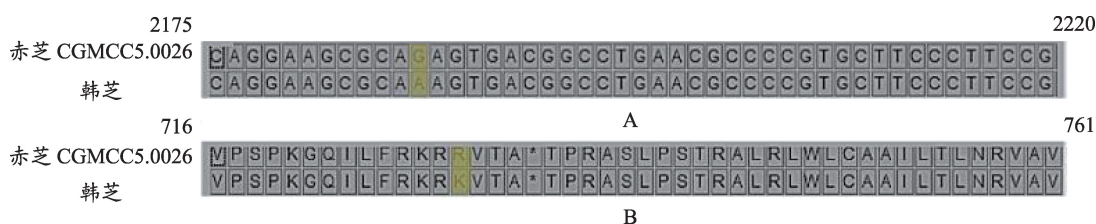


图5 基因GL15605序列比对

注:A. 基因GL15605 DNA序列比对, B. 基因GL15605蛋白序列比对。

使得氨基酸从精氨酸转变为赖氨酸(图5 B)。

2.8 对注释的与生长相关的转录因子的RT-PCR检测结果

对注释的与生长相关的6个转录因子进行RT-PCR分析,结果表明(图6),韩芝的基因表达量高于赤芝 CGMCC5.0026,注释为 *PacC/Rim101* (GL27476) 和 *Zap1* (GL21797) 的基因,在两种菌丝中的差异具有统计学意义。其它4个基因的表达量差异不显著。

3 讨论

随着越来越多生物的基因组被测序,其基因组重测序技术已成为分子育种研究的重要手段。本文对韩芝进行基因组重测序,得到了大量的变异基因,在COG注释中,这些变异基因主要注释到了氨基酸的运输和代谢,次生代谢物的生物合成、运输和分解代谢,碳水化合物的运输和代谢。

变异基因中,注释到的转录因子共有86个,主要的类型为bZIP、C2H2、CBF、HTH、MADS、TFII_related和一些真菌特有的转录因子。其中,C2H2类型转录因子 *PacC* 的缺失对真菌菌丝的生长和分生孢子的产量具有一定的抑制作用^[19,20]。赤芝转录因子 *GLPacC* 沉默,对菌丝生长速度、子实体发育和灵芝酸合成具有抑制作用。韩芝中该基因CDS区发生了密码子删除(3的整数倍),CGCCCCCGC碱基缺失。在韩芝中还发现了另一个pH响应转录因子 *PacC/Rim101* (GL27476) 在CDS区也发生了非同义编码突变。此外, *Sui6* 转录因子属于APSES家族,是真菌特有的一个蛋白家族^[21]。有研究表明,赤芝中转录因子 *GLSui6* 基因沉默菌株降低了菌丝生长速度,增加了菌丝分支,并且抑制了原基和子实体的形成,降低了灵芝酸的含量^[17],本研究发现韩芝中该基因CDS区发生了非同义编码突变。除此之外,还发现了一些与生长相关的转录因子 *Zap1*、*Spt8* 和 *Bfr2* 在CDS区也发生了非同义编码突变。有研究表明,酵母菌在锌受限型培养基中生长受到限制,这与

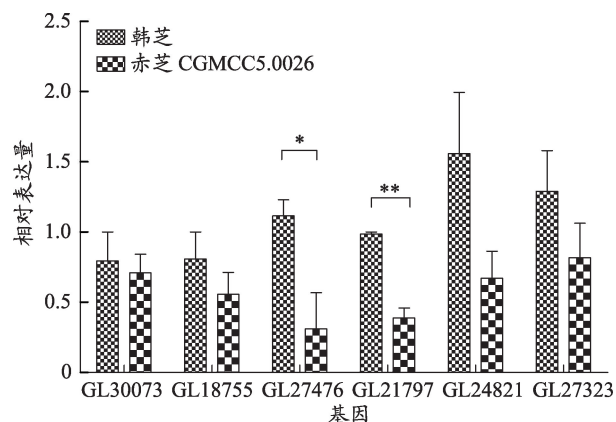


图6 不同菌丝的基因相对表达量

注:不同菌丝相比较,* $P < 0.05$,** $P < 0.01$ 。

Zap1 转录因子的靶基因有关^[22]。SPT包括SPT3、SPT7、SPT8和SPT20,是SAGA复合体的重要成分蛋白。禾谷镰刀菌中,与野生型相比, *Fgspt3* 和 *Fgspt8* 缺失突变株的菌丝生长明显受到抑制,并且不能产生分生孢子,同时与孢子形成相关的基因 *FgFlbC* 和 *FgRen1* 的表达量也下调^[23]。 *Bfr2* 基因对细胞生长周期具有调控作用^[24]。菌丝生长受很多基因和通路的调控,调控过程比较复杂。转录因子在结合区域产生的SNP可能会影响转录因子的结合特性,影响转录因子在等位基因上的结合强度,从而影响SNP所在基因的表达情况^[25]。在研究中我们发现,赤芝 CGMCC5.0026的生物量高于韩芝。对灵芝中注释的与生长相关的6个转录因子,通过RT-PCR检测发现赤芝 CGMCC5.0026的表达量均低于韩芝,由于在RT-PCR中使用甘油醛-3-磷酸脱氢酶(GPD)作为内参基因,该内参基因在不同灵芝中的表达量有所差异,本实验中发现韩芝中的GPD表达量(C_q 值为24.5)要高于赤芝 CGMCC5.0026(C_q 值为22.17),因此在计算基因的相对表达时,韩芝中基因的表达量要高于赤芝 CGMCC5.0026。这6个转录因子的变异是否会参与灵芝菌丝的生物量以及如何参与菌丝生长的调控还需要后续实验的进一步验证。

肌动蛋白骨架在真菌细胞极性生长、产孢、应激和维持细胞形态等方面发挥重要作用。本文中,韩芝中部分肌动蛋白相关蛋白 CDS 区发生了非同义编码突变、移码突变。肌动蛋白相关蛋白(ARP5)缺陷型植物表现出植株矮小和细胞大小异常^[26],本研究注释到的两个肌动蛋白相关蛋白(ARP5)均发生了非同义编码突变。

此外,本研究还发现,注释为细胞色素 P450 的变异基因有 195 个,这些基因在 KOG 分类注释中,有 135 个变异基因参与次生代谢产物的生物合成,转运和分解代谢,4 个变异基因参与了能量的产生和转化。与 Chen 等人研究中 78 个可能和赤芝三萜合成相关的 CYP450 相比,发现有 70 个相同的 CYP450,其中,有 12 个 CYP512 和 1 个 CYP5144,研究表明,在动物中 CYP512 和 CYP5144 对甾体激素和甾酮等激素的结构具有修饰作用,并且这两个 P450 家族与羊毛甾醇合酶(LSS)共表达^[3],由于细胞色素 P450 基因变异较多,因此本实验对不同菌株菌丝的三萜含量进行测定,发现韩芝菌丝三萜含量明显高于赤芝 CGMCC5.0026,这说明韩芝中大量细胞色素 P450 的变异可能会影响赤芝三萜的含量。这将为细胞色素 P450 参与三萜的合成提供新的证据。

赤芝全基因组测序的完成为灵芝成为药用模式真

菌奠定了坚实的基础^[3]。灵芝生长发育和发酵与栽培的研究为灵芝的一般生物学研究奠定了基础^[27]。灵芝的药用成分丰富,目前已分离出 400 多种活性物质,包括多糖、萜类、甾醇类、生物碱等^[28]。灵芝的基因组分析进一步证实了灵芝具有三萜、聚酮、倍半萜等多条次生代谢合成途径^[27],目前研究较多的是合成灵芝三萜的 MVA 途径,其上游途径中的一些关键酶基因如 *AACT*^[29]、*HMGS*^[30]、*HMGR*^[31]、*FPS*^[32]、*MVD*^[33]、*LSS*^[34] 和 *SQS*^[35] 已被克隆及进行基因过表达研究。赤芝基因组测序的完成筛选出 600 多个转录调控蛋白和 219 个 CYP450 编码基因^[3]。目前,灵芝遗传转化的研究也进一步开展,如电击法^[36]、PEG 介导法^[37]以及农杆菌介导法^[38]。这为灵芝功能基因组学的研究奠定了基础。本研究进行韩芝全基因组重测序,筛选出一些潜在的与菌丝生长相关的转录因子,并且对细胞色素 P450 对三萜含量的影响具有指导意义,这对完善灵芝模式真菌研究体系提供了基础。

综上所述,本研究通过全基因组重测序筛选了韩国赤芝与中国赤芝间变异的转录因子和细胞色素 P450,并对韩芝菌丝生长和三萜合成相关基因的变异进行初步挖掘和分析,对揭示赤芝菌丝生长和三萜合成的遗传机制具有重要意义,而这些变异基因的具体功能还有待进一步验证。

参考文献

- 1 池小妹. 我国灵芝人工栽培技术研究现状. 时珍国医国药, 2005, 16(8): 791-792.
- 2 金鑫, 刘宗敏, 黄羽佳, 等. 我国灵芝栽培现状及发展趋势. 食药菌, 2016, 24(1): 33-37.
- 3 Chen S, Xu J, Liu C, et al. Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat Commun*, 2012, 3(2): 913.
- 4 汪文强, 赵生国, 马利青, 等. 动物基因组学重测序的应用研究进展. 畜牧兽医学报, 2016, 47(10): 1947-1953.
- 5 Shin Y, Jung H, Jung M, et al. Discovery of gene sources for economic traits in hanwoo by whole-genome resequencing. *Asian Austral J Anim*, 2016, 29(9): 1353-1362.
- 6 Zhang X, Li R, Chen L, et al. Fine-mapping and candidate gene analysis of the brassica juncea white-flowered mutant Bjpc2 using the whole-genome resequencing. *Mol Genet Genomics*, 2017, 293(2): 359-370.
- 7 Varshney R K, Saxena R K, Upadhyaya H D, et al. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat Genet*, 2017, 49(7): 1082-1088.
- 8 Lee K J, Kim D S, Kim J, et al. Identification of candidate genes for an early-maturing soybean mutant by genome resequencing analysis. *Mol Genet Genomics*, 2016, 291(4): 1561-1571.
- 9 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25(14): 1754-1760.
- 10 McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 2010, 20(9): 1297-1303.
- 11 Cingolani P, Platts A E, Wang L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 2012, 6(2): 80-92.
- 12 Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997, 25(17): 3389-3402.
- 13 Fuchs H E. Integrated nr Database in Protein Annotation System and Its Localization. *Comput Eng*, 2006.

- 14 Ashburner M, Ball C A, Blake J A, *et al.* Gene ontology: tool for the unification of biology. The Gene ontology Consortium. *Nat Genet*, 2000, 25(1): 25–29.
- 15 Tatusov R L, Galperin M Y, Natale D A, *et al.* The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*, 2000, 28(1): 33–36.
- 16 Kanehisa M, Goto S, Kawashima S, *et al.* The KEGG resource for deciphering the genome. *Nucleic Acids Res*, 2004, 32(90001): 277–280.
- 17 Zhang G, Ren A, Liang S, *et al.* Functional analysis of an APSES transcription factor (GLSwi6) involved in fungal growth, fruiting body development and ganoderic-acid biosynthesis in *Ganoderma lucidum*. *Microbiol Res*, 2018, 207: 280.
- 18 Wu F L, Zhang G, Ren A, *et al.* The pH-responsive transcription factor PacC regulates mycelial growth, fruiting body development, and ganoderic acid biosynthesis in *Ganoderma lucidum*. *Mycologia*, 2016, 108(6): 1104–1113.
- 19 Landraud P, Chuzeville S, Billongrande G, *et al.* Adaptation to pH and Role of PacC in the Rice Blast Fungus *Magnaporthe oryzae*. *Plos One*, 2013, 8(7): e69236.
- 20 Huang W, Shang Y, Chen P, *et al.* MrpacC regulates sporulation, insect cuticle penetration and immune evasion in *Metarhizium robertsii*. *Environ Microbiol*, 2015, 17(4): 994–1008.
- 21 Zhao Y, Su H, Zhou J, *et al.* The APSES family proteins in fungi: Characterizations, evolution and functions. *Fungal Genet Biol*, 2015, 81: 271–280.
- 22 North M, Steffen J, Loguinov A V, *et al.* Genome-wide functional profiling identifies genes and processes important for zinc-limited growth of *Saccharomyces cerevisiae*. *Plos Genet*, 2012, 8(6): e1002699.
- 23 Gao T, Zheng Z, Hou Y, *et al.* Transcription factors spt3 and spt8 are associated with conidiation, mycelium growth, and pathogenicity in *Fusarium graminearum*. *Fems Microbiol Lett*, 2014, 351(1): 42–50.
- 24 Chabane S, Képès F. Expression of the yeast BFR2 gene is regulated at the transcriptional level and through degradation of its product. *Mol Gen Genet*, 1998, 258(3): 215–221.
- 25 董立丽. SNP 对于转录因子结合及基因表达的影响. 2015, 哈尔滨: 哈尔滨工业大学硕士研究生论文, 2015.
- 26 Meagher R B, Kandasamy M K, Smith A P, *et al.* Nuclear actin-related proteins at the core of epigenetic control. *Plant Signaling & Behavior*, 2010, 5(5): 518–522.
- 27 孙超, 胡鸾雷, 徐江, 等. 灵芝: 一种研究天然药物合成的模式真菌. *中国科学: 生命科学*, 2013, 43(6): 447–456.
- 28 Zhou X W, Su K Q, Zhang Y M. Applied modern biotechnology for cultivation of *Ganoderma*, and development of their products. *Appl Microbiol Biot*, 2012, 93(3): 941–963.
- 29 Fang X, Shi L, Ren A, *et al.* The cloning, characterization and functional analysis of a gene encoding an acetyl-CoA acetyltransferase involved in triterpene biosynthesis in *Ganoderma lucidum*. *Mycoscience*, 2013, 54(2): 100–105.
- 30 Ren A, Ouyang X, Shi L, *et al.* Molecular characterization and expression analysis of GHMGS, a gene encoding hydroxymethylglutaryl-CoA synthase from *Ganoderma lucidum* (Ling-zhi) in ganoderic acid biosynthesis pathway. *World J Microb Biot*, 2013, 29(3): 523–531.
- 31 Shang C H, Zhu F, Li N, *et al.* Cloning and Characterization of a Gene Encoding HMG-CoA Reductase from *Ganoderma lucidum*, and Its Functional Identification in Yeast. *Biosci, Biotech Bioch*, 2008, 72(5): 1333–1339.
- 32 Ding Y X, Ou-Yang X, Shang C H, *et al.* Molecular Cloning, Characterization, and Differential Expression of a Farnesyl-Diphosphate Synthase Gene from the Basidiomycetous Fungus *Ganoderma lucidum*. *J Agr Chem Soc JP*, 2008, 72(6): 9.
- 33 Shi L, Qin L, Xu Y, *et al.* Molecular cloning, characterization, and function analysis of a mevalonate pyrophosphate decarboxylase gene from *Ganoderma lucidum*. *Mol Biol Rep*, 2012, 39(5): 6149–6159.
- 34 Zhang D H, Li N, Yu X, *et al.* Overexpression of the homologous lanosterol synthase gene in ganoderic acid biosynthesis in *Ganoderma lingzhi*. *Phytochemistry*, 2017, 134: 46–53.
- 35 Zhou J S, Ji S L, Ren M F, *et al.* Enhanced accumulation of individual ganoderic acids in a submerged culture of *Ganoderma lucidum*, by the overexpression of squalene synthase gene. *Biochem Eng J*, 2014, 90(5): 178–183.
- 36 Li S, Cai H, Xu W, *et al.* Efficient transformation of the medicinal mushroom *Ganoderma lucidum*. *Plant Mol Biol Rep*, 2001, 19(4): 383–384.
- 37 李刚, 王强, 刘秋云, 等. 利用 PEG 法建立药用真菌灵芝的转化系统. *菌物学报*, 2004, 23(2): 255–261.
- 38 Shi L, Fang X, Li M, *et al.* Development of a simple and efficient transformation system for the basidiomycetous medicinal fungus *Ganoderma lucidum*. *World J Microb Biot*, 2012, 28(1): 283–291.

The Analysis of Whole-genome Resequencing of *Ganoderma Lucidum* Originated from South Korea

Zhu Fengli¹, Xu Xiaolan², Chen Tiqiang³, Shi Linchun⁴, Miao Xiaoqing², Lan Jin⁴

(1. College of Food Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. College of Bee science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 3. Fujian Academy of Agricultural Sciences, Fuzhou 350014, China; 4. Institute of Medicinal Plant Development, Chinese Academy of

Medical Sciences and Peking Union Medical College, Beijing 100193, China)

Abstract: The difference in genomic level between *Ganoderma lucidum* (*G. lucidum*) from South Korea and *G. lucidum* CGMCC5.0026 from China were analyzed to provide a research basis for genetic breeding of *G. lucidum*. We used high-throughput sequencing technology to perform whole genome resequencing for *G. lucidum* from South Korea, as well as detected and annotated single nucleotide polymorphisms (SNP), small InDel (InDel) and structure variations (SV). Mutant genes at DNA level were annotated by Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways, Gene Ontology (GO), Clusters of Orthologous Groups of proteins (COG), Non-redundant (NR) and SwissProt. The results showed that the resequencing depth of the *G. lucidum* from South Korea was 44 fold. And compared with the *G. lucidum* CGMCC5.0026 from China, 10607 SNPs, 4774 InDels and 1428 SVs were found, which resulted in 9469 gene mutations. Among the mutant genes, there were 86 transcription factors and 195 cytochrome P450s. This research provides data for molecular marker breeding of *G. lucidum* and the model system for studying medicinal fungi.

Keywords: Mycelial growth, whole genome resequencing, single nucleotide polymorphism (SNP), small indel (InDel), structure variation (SV), transcription factor, cytochrome P450

(责任编辑:刘 宁,责任译审:王 昭)