

杉木生长性状的杂种优势转录组分析

齐明¹, 何贵平^{1*}, 周建革², 程亚平³, 徐肇友⁴

(1. 中国林业科学研究院亚热带林业研究所, 浙江省林木育种技术研究重点实验室, 浙江 杭州 311400; 2. 浙江省遂昌县林业技术推广总站, 浙江 遂昌 323300; 3. 浙江省庆元县庆元林场, 浙江 庆元 323805; 4. 浙江省龙泉市林业科学研究院, 浙江 龙泉 323700)

摘要: [目的] 研究杉木杂种 F1 的基因表达模式, 从基因表达水平揭示杂种优势形成分子机理, 为杉木杂种深度开发利用提供参考依据。 [方法] 采用 Illumina HiSeq4000 高通量测序技术对不同生长势的杉木杂种(龙 15 × 1339) HF1(3 个超亲杂种)、LF2(3 个低亲杂种) 及其亲本进行转录组测序和差异比较。利用无参转录组分析, 先将测序 reads 进行 de novo 拼接, 组装获得 unigenes 和 transcripts, 然后进行功能注释、差异表达等项目分析。 [结果] 12 个样本, 转录组测序共产生 Clean reads 5.8E+08 条, 总拼接长度 49 803 726 pb, BLASTX 分析, 比对结果产生 80 171 个基因。韦恩图揭示基因在上下代间的传递模式有 5 种: (1) 双亲可表达但杂种中不表达(双亲共沉默型); (2) 只在双亲之一中表达, 不在杂种中表达(亲本特异表达型); (3) 只在杂种中表达, 不在双亲中表达(杂种特异表达型); (4) 在杂种和一个亲本中表达(单亲表达一致型); (5) 在双亲和杂种中都表达。在 HF1VSP1 比较组中, 筛选出 236 个不同差异表达的基因; 在 HF1VSP2 比较组中, 筛选出 505 个差异表达基因。在 LF2VSP1 中, 筛选出 1 483 个差异表达基因; 在 LF2VSP2 中筛选出了 2 335 个差异表达基因。从亲代和子代不同样本组间, 各挑选出 100 个左右差异表达极显著的基因, 采用其表达量进行热聚类分析, 得到不同的聚类块, 聚类块内的基因表达量在父母本间是互补的, 说明杉木杂种优势分子机理是超显性; 聚类块有大小之别, 则说明杉木有些性状是寡基因控制, 有些性状是多基因控制; 杉木生长的超亲优势是龙 15 中下调而在超亲子代中上调的这 14 个基因决定的。 [结论] 杉木杂种优势的分子机理是超显性, 杉木超亲杂种高生产力与 14 个上调基因有关, 环境显著刺激了这 14 个基因的上调表达, 从而促进生长优势的产生。

关键词: 杉木; 生长; 杂种优势; 转录组测序; 分子机理

中图分类号: S718.3

文献标识码: A

文章编号: 1001-1498(2019)03-0113-08

杉木(*Cunninghamia lanceolata* (Lamb.) Hook) 是我国南方林区重要的速生用材造林树种, 经过 40 多年的遗传改良, 选育了一大批速生优质的杉木新品种应用于生产中。杂交育种是杉木遗传改良的主要途径^[1], 主要通过选择遗传互补的优良亲本进行交配, 再选择出速生优质的杂交新品种。在杉木杂交育种研究中, 生长性状的杂种优势现象已有报道^[2], 主要从数量遗传学或表观遗传学等方面进行了探讨^[2-4]。

随着分子生物学技术不断发展, 利用分子标记、RNA 差异显示、基因芯片等技术对杂种优势已有所研究, 如 Birchler 认为基因差异表达模式与杂种 F1

的表型存在一定的联系, 杂种的特定性状可能是父本、母本特定位点的等位基因共同表达的结果^[5-6]。Li 等^[7] 利用基因芯片技术对落叶松(*Larix* spp.) 的杂种优势研究表明, 杂种 F1 与亲本间有 54 个差异显著基因为非加性表达模式, 且这些基因参与生理过程、应激反应及淀粉和蔗糖代谢等多种生化途径。丁昌俊等^[4] 运用转录组测序法研究了不同生长势的美洲黑杨(*Populus deltoides* Marsh) 转录组的差异, 获得了大量差异基因, 发现杨树杂种优势的形成可能是由于相关基因的显著表达, 调节光合作用、物质代谢吸收等与生长紧密联系的代谢活动, 进而促进了杂种生长优势。

收稿日期: 2018-03-25 修回日期: 2019-02-11

基金项目: 浙江省“十三五”林木新品种育种专项(2016C02056-5); 中国林科院中央级公益性科研院所基本科研业务费专项资金资助子项(CAFYBB2017 ZA001-1-2)

* 通讯作者: 何贵平. E-mail: guipinghe@126.com

不少学者借助转录组测序或 RNA-seq 高通量测序技术,研究了杉木纤维性状发育的形成机制和木材形成活动的机理^[8-12],但杉木生长性状杂种优势产生的分子机理研究尚未见报道。本研究利用无参转录组测序技术,在基因差异研究的基础上,以基因差异表达分析为切入点,以杉木 1 个优良组合的超亲、低亲两组杂交子代及其亲本为研究材料,展开了杂种子代和亲本两两间的比较分析,以期揭示杉木生长性状杂种优势的分子机理,获得具有重要生物学功能的基因,为杉木杂交育种提供科学依据。

1 材料与方法

1.1 研究材料

研究材料为已选育出的杉木优良杂交组合(龙 15 × 1339)及其亲本。1996 年春,对前期试验中表现突出的杉木杂交组合,在浙江遂昌杉木种子园进行了重复制种,1997 年完成杂交子代育苗。同年,将浙江遂昌多系种子园几个小区改建成龙 15 × 1339 等组合的双系种子园。1997 年嫁接成园。本研究中所用的亲本和杂种(龙 15 × 1339)不仅林龄相同,且都达到成熟阶段。

2017 年 6 月 21 日,对杉木双系种子园中龙 15 和 1339 及同龄区域化试验林中的龙 15 × 1339 子代取样,叶样来自样树顶部当年生的嫩枝叶,3 个生物学重复,龙 15(P1)3 个样株的编号为 P1-1、P1-2 和 P1-3,1339(P2)3 个样株的编号为 P2-1、P2-2 和 P2-3。由于杂交组合内子代会发生分离,杂种子代的取样参照丁昌俊等^[4]的方法。对试验林内该组合抽取超亲子代 3 株:HF1-1、HF1-2、HF1-3,低亲子代 3 株:LF2-1、LF2-2、LF2-3,构成 4 个样本组 HF1、LF2、P1 和 P2,进行测序、序列组装、功能注释,最后亲本与子代共 4 个样本组,两两比较,总共 6 个比较组进行处理与分析。杉木龙 15 × 1339 子代表现和超亲杂种优势见表 1。

取样时超亲子代 HF1 的平均胸径是超低亲子代 LF2 胸径的 134.81%~138.32%;HF1 平均树高是 LF2 的 134.60%~150.00%。亲本龙 15 和 1339 取自双 8 区生长发育正常的植株。

1.2 研究方法

1.2.1 cDNA 文库准备及 RNA-seq 测序 文库构建、无参转录组测序以及随后的 unigene 功能注释、基因表达分析等项目的分析参见有关文献^[12-17]。

表 1 浙江遂昌杉木杂交组合龙 15 × 1339 的超亲优势
Table 1 The superior advantages of the superior combinations Long 15 × 1339 of Zhejiang Suichang

杂交组合 Cross combinations	材积 Volume/dm ³	树高 Height/m	胸径 DBH/cm
龙 15 × 1339	19.69	5.75	8.35
Long 15 × 1339	(20.87%)	(7.48%)	(8.58%)
龙 15 × 闽 33	18.51	5.46	8.21
Long 15 × Min33	(13.63%)	(2.06%)	(6.76%)
龙 15 半同胞	16.29	5.35	7.69
Long 15 family of half siblings			

注:括号中的数值是超亲优势。

Note: The value in parentheses is transgressive advantage.

实验流程按照 Illumina 公司提供的标准步骤执行。文库质检合格后,用 Illumina Hiseq4000 测序仪进行测序,测序策略为双末端测序,即高通量测序数据通常为 2 × 150 bp。测序深度采用施季森等^[8-10]杉木转录组测序的方法,为 6G。原始 RAW reads 过滤后,使用 Trinity 软件对高质量数据[Clean reads]进行从头组装,即进行 de novo 拼接,获得 unigenes 和 transcripts。随后进行功能注释:将 unigenes 序列分别比对到 NCBI 的数据库 Nr、Pfam、KEGG (kyoto encyclopedia of genes and genomes)、Swiss-port、KOG 及 GO (gene ontology),获得与 unigene 相似性最高的序列,继而确定该 unigene 的功能和名称。

1.2.2 序列比对及差异表达基因分析 通过 Illumina Hiseq4000 测序获得的转录组测序数据,需经过以下几个步骤的生物信息学处理与分析^[12-17],方能获得有意义的结果:(1)原始数据处理;(2)序列组装;(3)unigene 序列的功能注释;(4)差异表达的 unigene 的热聚类分析等。除了样本针叶的采集和研究方案的制定外,整个项目中的测序和初步分析委托杭州联川生物技术股份有限公司完成。

unigene 的表达量 FPKM (Fragments per kb per million fragments),其计算公式: $FPKM = 10^6 C / [(NL)/10^3]$,式中: FPKM 是某个基因(A)的表达量,C 是唯一比对到基因 A 的片段数,N 是唯一比对到所有 unigene 的总片段数,L 为 unigene A 的碱基数。

2 结果与分析

2.1 测序数据质量和基因表达概况

杉木亲本龙 15 (P1) 和 1339 (P2)、超亲子代 HF1 和低亲子代 LF2,每个品种有 3 个生物学重复,12 株样树(4 个样本组),测序测得的原始测序序列

(Raw reads)介于 $4.3E+07 \sim 5.2E+07$; 12个样本的 Clean reads 分别介于 $4.3E+07 \sim 5.2E+07$ 。Clean reads 有效数据占原始 Raw reads 的比例在 98% 以上。

12个样本的 Phred 数值大于 Q20 和 Q30 的碱基占总体碱基的百分比,分别介于 98.20% ~ 98.84% 和 95.07% ~ 96.4%。原始测序序列中,碱基 G 和 C 的数量总和占总碱基数的百分比介于 44.20% ~ 45.08% 之间。

表2 拼接结果的统计

Table 2 Statistics of splicing results

项目 Item	总数 All	平均 GC Mean GC/%	最小基因长度 Min length	最大基因长度 Max length	总组装基座 Total assembled bases	读长数达 50% 时 读长长度 N50/pb
基因 Gene	80 171	40.78	201	14 718	49 803 726	1 214
转录本 Transcript	131 241	40.89	201	14 718	113 134 804	1 660

2.3 基因注释结果

对 Clean reads 在 6 个数据库进行 BLASTX 分析。将获得的基因分别在 Swiss-prot、Nr、Pfam、KEGG、KOG 和 GO 中注释,结果杉木亲代和子代共注释有 80 171 个基因(表 3)。Nr 数据库的注释结果为 32.50%,这远远低于通常报道的结果(56%)。Nr 数据库的注释结果与注释软件、阈值等有关,也与基因的组装结果有关。

杉木为无参测序,测得的数据通常采用 BLASTX 序列比对,观察表 3 可以发现,各数据库所注释的基因比率大于 100%,这表明有的基因同时不同的数据库得到重复注释。由于针叶树已进行全基因组测序的树种太少,因而与农作物相比,针叶树种中挖掘出的基因也少。为了获得良好的测序结果,今后要加快针叶树尤其是杉木的全基因组测序研究。

表3 不同数据库 BLAST 注释结果统计

Table 3 Statistics of BLAST annotation of different database

项目 Item	Swiss-prot	Nr	Pfam	KEGG	KOG	GO
基因数 Gene_number	16 837	26 055	20 239	9 162	21 059	15 088
比例 Percentage/%	21.00	32.50	25.24	11.43	26.27	18.82

2.4 亲代与其杂种子代 4 组样本间的基因差异模式分析

基于转录组测序结果,绘制亲本和子代基因传递与表达韦恩图(图 1)。可以发现,将杂种和双亲

12 株样树测得的基因表达量在不同区间的分布接近正态分布。综合以上几个测序质量评价指标,说明 12 个样株的测序质量较高,能满足后续研究和数据分析的要求。

2.2 测序数据的拼接结果

从测序数据的拼接结果可见(表 2),平均 GC% 达 40.78%;当读长数达 50% 时,该读长的长度为 1 214 pb。综合其它各项结果,可以得出测序数据的组装拼接十分成功。

相比,基因差异表达呈现以下 5 种模式:(1)双亲中可表达但杂种中不表达(双亲共沉默型);(2)只在双亲之一中表达,不在杂种中表达(亲本特异表达型);(3)只在杂种中表达,不在双亲中表达(杂种特异表达型);(4)在杂种和一个亲本中表达(单亲表达一致型);(5)在双亲和杂种中都表达。前 4 种模式属基因表达质的差异,即存在与缺失变异(Presence/absence variation,简称 PAV),而第 5 种模式则属于基因表达量的差异。以上结果与徐进等^[18]对鹅掌楸(*Liriodendron Chinense* (Hemsl) Sarg.)、张小蒙等^[19]对水稻(*Oryza sativa* L.)及王章奎等^[20]对小麦(*Triticum aestivum* L.)的研究结果基本一致,与张君等^[21]对大豆(*Glycine max* (L.) Merr.)的研究结果完全一致。这表明亲本间存在差异基因与杂种优势有关。

2.5 亲代与其杂种子代 4 组样本间的差异基因表达分析

从杉木 4 个样本组组间两两比较的差异基因表达分析结果(图 2)可以看出,P2VSP1、HF1VSP1、HF1VSP2、LF2VSHF1 4 组差异表达基因不多,上调表达差异基因少于下调基因数;LF2VSP1、LF2VSP2 两组间的表达差异基因多,但上调基因占多数;每组内的上调/下调基因数略有差别,但两者都有基因表达。这一结果与丁昌俊等^[4]在美洲黑杨中的研究结果基本一致。龙 15 × 1339 这一杂交组合内,HF1 是超亲子代,LF2 是低亲子代,HF1 较 LF2 表达差异基因的数量少,上调/下调基因比率不均匀、不平衡,这导致了 HF1 生长快,LF2 生长慢。

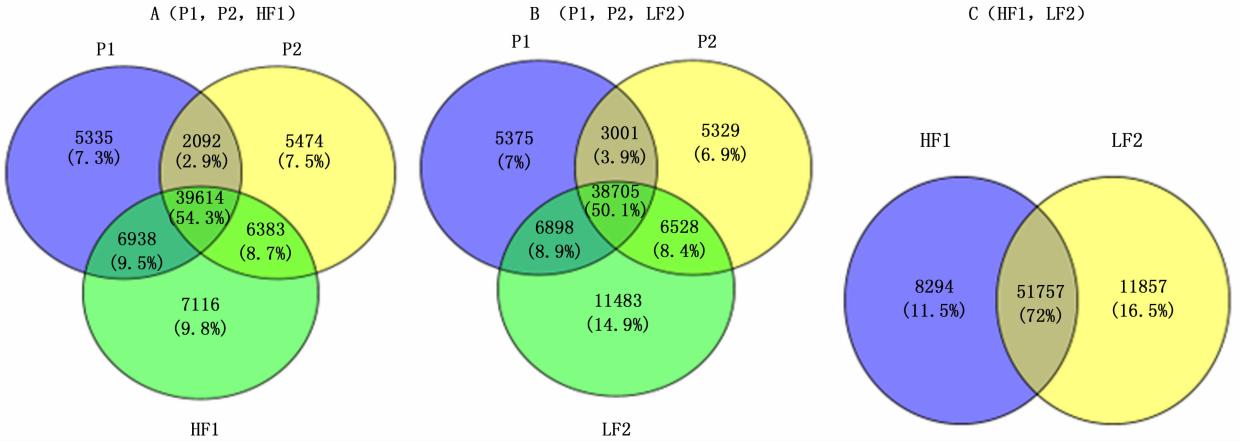


图 1 亲本和子代基因数的韦恩图

Fig. 1 A Venn diagram of the parent and progenies

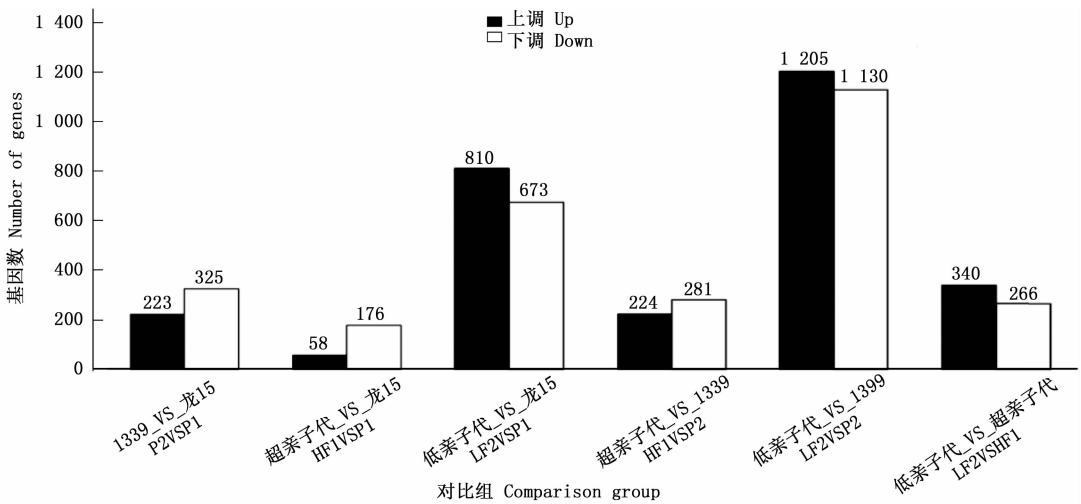


图 2 杉木转录组组内差异基因表达分析结果

Fig. 2 Differentially expressed genes in different groups

2.6 基因表达量的聚类分析与杂种优势的分子机理

从测序组装、比对注释和基因表达量计算到基因的差异分析,最终从各样本组获得的基因中,依据显著富集的 FDR 值,分别抽取表达量极其显著的 100 多个基因。差异基因的富集分析是基于所有的差异基因进行的,并没有人为选择差异基因。根据基因表达量这一性状进行聚类分析,得到基因和亲本的聚类图(图 3)。从聚类图(图 3A)中可以发现:同一亲本不同分株在同一基因上通常表达量是一致的,100 多个基因,按表达量聚类分为不同的聚类块。(1)由于同一聚类块的基因具有类似的功能,并假设它们控制一个性状,由于有些聚类块是由较少基因组成,有些聚类块的基因是多基因组成,这说明杉木数量性状有些是受寡基因控制,有些是受多基因控制的。(2)参与杂交的 2 个亲本,在相同的聚

类块上,通常龙 15 的基因显著高表达,1339 的同一基因则显著低表达,反之龙 15 的基因显著低表达,1339 的同一基因则显著高表达,即同一基因在父母本间不同亲本基因表达量变化是互补的。既然龙 15 × 1339 这样组配产生了杂种优势,这表明杉木杂种(龙 15 × 1339)生长性状的杂种优势是超显性在起作用(图 3 A)。(3)超亲子代和 1339 间,104 基因按表达量聚类,获得 10 个聚类块(图 3B): I 类有 16 个基因,亲本和超亲子代同上调; II 类有 18 个基因,亲本上调,超亲子代下调; III 类有 2 个基因,亲子同为上调; IV 类有 10 个基因,亲子同上调; V 类有 7 个基因,亲本下调,超亲子代上调; VI 类有 26 个基因,亲本上调,子代下调; VII 类有 9 个基因,亲本下调,子代上调; VIII 类有 9 个基因,亲本下调,子代上调; IX 类有 2 个基因,亲子同为下调; X 类有 5 个基因,亲本下调,子代上调。这 10 个聚类块大致控制 10 个性

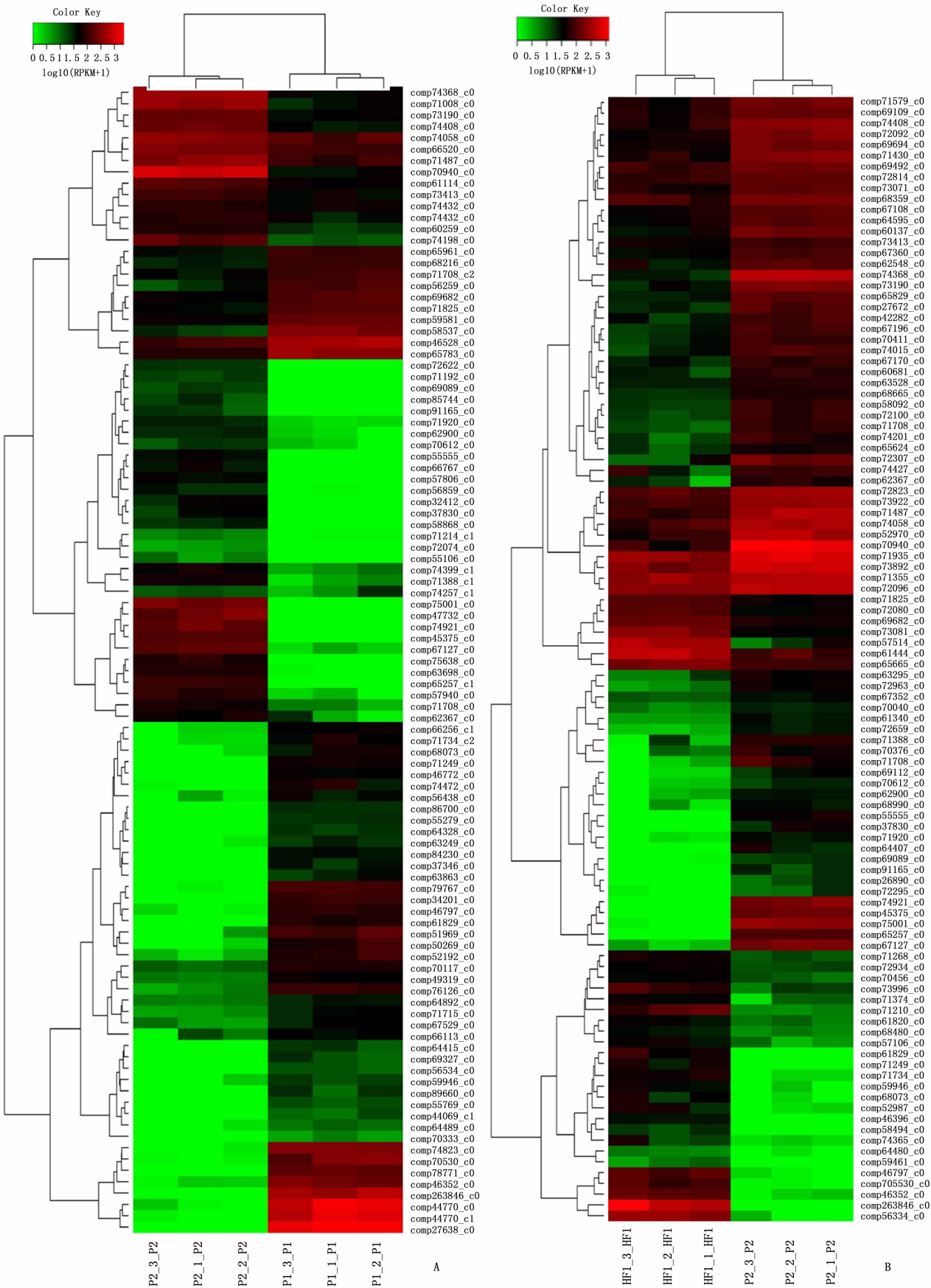


图3 亲本和子代间的差异极显著基因层次聚类分析部分结果

Fig.3 Hierarchical clustering analysis results were shown extremely significant difference between parent and offspring genes

状。其中, V类、VII类、VIII类和X类这4个性状(聚类块)中,同一基因,亲本1339下调,超亲代上调。

对超亲代与龙15(HF1VSP1)进行热聚类分析,聚类结果与HF1VSP2基本相同(图略),但使用的基因不同。100个基因,按表达量划分为7个聚类块,控制7个性状的表达: I类有30个基因,亲本上调,子代下调; II类有11个基因,亲本上调,子代下调; III类有16个基因,亲子同为上调; IV类有5个基因,亲子同为下调; V类有9个基因,亲本上调,子

代下调; VI类有15个基因,亲子同为下调; VII类有14个基因,亲本下调,子代上调。可见,超亲代与龙15的热聚分析中,只有VII类中的这14个基因可能与杉木杂交子代的高生产力有关,可能是控制生长性状杂种优势的因子。另外,作者还对亲本下调,超亲代上调的基因进行了追踪,试图阐释子代表达量上调的这5个聚类块(性状)中的44个基因归属哪个代谢途径或参与了哪个性状的形成。超亲代与龙15比较组VII类的14个基因追踪结果列于表4。

表4 HF1VSP1中VII类14个基因的功能追踪结果

Table 4 14 Genes function tracking results in HF1VSP1 VII class

基因 Gene	基因的功能 Gene function	基因 Gene	基因的功能 Gene function
Comp73565_c0	与 GO:0005634(细胞核)同源 Nuclear homology	Comp72062_c3	[R]一般的功能预测 [R] Ggeneral functional prediction
Comp61820_c0	脂质运输和代谢 Lipid transport and metabolism	Comp71268_c0	昼夜节律 Circadian rhythms
Comp71030_c0	淀粉和蔗糖代谢 Starch and sucrose metabolism	Comp52987_c0	翻译后修饰;与叶绿体同源 Posttranslational modification;homologous to the chloroplast
Comp61016_c0	(只有上调表达量信息) (Only upregulated expression information)	Comp66767_c0	(只有上调表达量信息) (Only upregulated expression information)
Comp68480_c0	[R]一般的功能预测 [R] General functional prediction	Comp68783_c0	光合作用 Photosynthesis
Comp56601_c0	[P]无机离子运输和代谢 [P] Transport and metabolism of inorganic ions	Comp70456_c1	昼夜节律;与 GO:0030154 同源 Circadian rhythms; same as GO: 0030154
Comp62983_c0	苯丙氨酸代谢 Phenylalanine metabolism	Comp50961_c0	ATP 结合;信号传递机制 ATP binding; signal transfer mechanism

可以看出,这14个在超亲代中上调的基因,多数与生长有关:Comp71030_c0参与淀粉和蔗糖代谢;Comp62983_c0参与苯丙氨酸代谢;Comp71268_c0和Comp70456_c1参与昼夜节律;Comp68783_c0参与光合作用;Comp50961_c0参与ATP结合,并行使信号传递功能。显然这些基因与生长相关,是杉木高生产力的直接原因。

对于HF1VSP2,作者也对亲本下调,超亲代中的上调基因进行了追踪:V类的7个基因多与次生物代谢有关;VII类和VIII类中的18个基因,有8个基因只有表达量信息,其余10个基因涉及不可或缺的膜、假设蛋白、翻译后修饰等功能;X类的5个基因有3个只有表达量信息,另外2个基因涉及叶绿体被膜和一般功能预测;显然这4个聚类块(性状)与生长优势无直接关系,而是参与其它性状的形成。这些推断有待于今后杉木全基因组的测序和功能分析结果的验证。

3 讨论

杂种优势已在许多作物中广泛利用,不少学者

就杂种优势机理提出了许多假说,其中有代表性的假说有Bruce^[22]提出的显性假说与Shull^[23]提出的超显性假说。在杉木遗传育种领域,有关杉木生长性状也存在“加性学说”^[24]和“显性学说”之争。随着分子生物学研究方法和技术的应用,植物遗传学研究不断拓展和深入,但有关杂种优势遗传机制的争议仍然存在,因为分子生物学研究结果仍与研究材料等因素有关^[25-33]。如在水稻研究中,Xiao等^[29]提出显性效应是水稻杂种优势的主要遗传基础,Yu等^[30]则认为上位性效应是水稻杂种优势的重要遗传基础;庄杰云等^[31]进一步考虑了遗传背景的影响,指出基于杂合型遗传背景的超显性效应是水稻杂种优势的重要遗传基础。

本研究得出杉木生长性状杂种优势的分子机理是超显性,这和本项目组关于杉木数量遗传的研究结果是一致的^[2,28]。随着杉木分子遗传学研究的深入,采用不同的杉木研究材料,使用不同的分析方法(如QTL分析),杉木数量性状的分子遗传机制的研究结果有可能与水稻中的情形相同或不同;也可能与本研究相同或不相同。杉木进入了高世代育种阶

段,其亲本材料经过多世代与高强度选择,高世代育种群体在改良性状上多为杂合体,杂合体居多的育种群体,性状的遗传控制方式会由加性方差占主导地位,过渡到显性方差占主导地位^[28]。同时杂种亲本间合理组配,会产生杂种优势。因此,杉木生长性状杂种优势分子机理为超显性这一结果,也适合于杉木高世代育种。

4 结论

本研究采用转录组测序法研究杉木杂种优势的分子机理,得出杉木生长性状的杂种优势的分子机理是超显性。使用热聚类分析的方法,对基因和研究材料进行了类别划分,基因热聚类分析结果揭示,杉木有些性状通常由2~30个基因控制表达;杉木超亲子代高生产力的原因,是HF1VSP1中Ⅶ类中的14个基因在超亲子代中上调表达的结果。

韦恩图揭示杉木基因在上下代间的传递模式有5种:(1)双亲中可表达但杂种中不表达(双亲共沉默型);(2)只在双亲之一中表达,不在杂种中表达(亲本特异表达型);(3)只在杂种中表达,不在双亲中表达(杂种特异表达型);(4)在杂种和一个亲本中表达(单亲表达一致型);(5)在双亲和杂种中都表达。

通过杉木基因注释和功能分析,特别是热聚类分析,获得了诸如细胞膜、跨膜运输、细胞壁合成和防御反应的相关基因,以及植物昼夜节律、信号传导、淀粉和蔗糖代谢途径、苯丙素的生物合成代谢途径、激素合成与运输代谢等基因,这些上调/下调基因中有些基因参与了杂种优势的形成。

参考文献:

[1] 何贵平,齐明,程亚平,等.杉木杂交育种中亲本选配方法的研究[J].江西农业大学学报,2016,38(4):646-653.

[2] 何贵平,齐明.杉木育种策略及应用[M].北京:中国林业出版社,2017:33-77,149-174.

[3] 洪舟.杉木杂种优势的分子机理研究[D].南京:南京林业大学,2009.

[4] 丁昌俊,张伟溪,高暝,等.不同生长势美洲黑杨转录组差异分析[J].林业科学,2016,52(3):47-58.

[5] Birchler J A. In search of molecular basis of heterosis[J]. Plant Cell,2003,15(10):2236-2239.

[6] Birchler J A. Reflections on studies of gene expression in aneuploids[J]. Biochemical Journal, 2010, 426(2):119-123.

[7] Li Z Y,Zhang T F,Wang S C. Transcriptomic analysis of the highly heterotic maize hybrid Zhengdan 958 and its parents during spikelet and floscule differentiation[J]. Journal of Integrative Agriculture,

2012,11(11):1783-1793.

[8] Zhang Y, Han X, Sang J, et al. Transcriptome analysis of immature xylem in the Chinese fir at different developmental phases[J]. Peer J, 2016, 4(17):e2097. doi:10.7717/peerj.2097.

[9] Huang H H, Xu L L, Tong Z K, et al. De novo characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis[J]. BMC Genomics, 2012, 13(1):648-658.

[10] Wang Z J, Chen J H, Liu W D, et al. Transcriptome characteristics and six alternative expressed genes positively correlated with the phase transition of annual cambial activities in Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook)[J]. PLOS One, 2013, 8(8):1-14.

[11] Qiu Z, Wan L, Chen T, et al. The regulation of activity in Chinese fir (*Cunninghamia lanceolata*) involves extensive transcriptome remodeling[J]. New Phytologist, 2013, 199(3):708-719.

[12] 马智慧. 铝胁迫下杉木幼苗的几种生理过程及转录组序列的研究[D].福州:福建农林大学,2015.

[13] 李阳. 亚硝酸盐对水稻胚性愈伤组织的诱导作用及机制[D].武昌:武汉大学,2016.

[14] 丁健. 沙棘果肉和种子油脂合成积累及转录表达差异研究[D].哈尔滨:东北林业大学,2016.

[15] 冯延芝. 杜仲种仁转录组测序及FAD3基因鉴定与功能研究[D].北京:中国林业科学研究院,2016.

[16] 翟荣荣. 超级稻协优9308根系杂种优势的转录组分析[D].北京:中国农业科学院,2013.

[17] 蒋桂雄. 油桐种子转录组解析及油脂合成重要基因克隆[D].长沙:中南林业科技大学,2014.

[18] 徐进,李帅,李火根,等. 鹅掌楸属植物生长旺盛期叶芽基因差异表达与杂种优势关系的分析[J]. 分子植物育种,2008,6(6):1111-1116.

[19] 张小蒙,肖宁,张洪熙,等. 水稻基因差异表达与杂种优势的关系分析[J]. 中国农业科学,2012,45(7):1235-1245.

[20] 王章奎,倪中福,孟凡荣,等. 小麦杂交种及其亲本拔节期根系差基因表达与杂种优势关系的初步研究[J]. 中国农业科学,2003,36(5):473-479.

[21] 张君,闫冬生,王丕武,等. 大豆杂交种及其亲本籽粒基因差异表达与杂种优势关系[J]. 中国油料作物学报,2010,32(3):354-361.

[22] Bruce A B. The Mendelian theory of heredity and the augmentation of vigor[J]. Science,1910,32(827):627-628.

[23] Shull G H. Duplicate genes for capsule-form in *Bursa bursa-pastoris*[J]. Molecular and General Genetics, 1914,12(1):97-149.

[24] 叶培忠,陈岳武,刘大林,等. 配合力分析在杉木数量遗传学中的应用[J]. 南京林产工业学院学报,1981,(3):1-21.

[25] 许晨曦,孙晓梅,张守攻. 差异基因表达和杂种优势形成机制[J]. 遗传,2013,35(6):714-726.

[26] 邢俊杰,成志伟,杨剑,等. 利用基因芯片技术分析水稻杂种优势的分子机理[J]. 杂交水稻,2005,20(4):59-61.

[27] Meyer S, Pospisil H, Scholten S. Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern[J]. Plant Molecular Biology,

- 2007, 63(3), 381–391.
- [28] 齐 明. 杉木育种中 GCA 与 SCA 的相对重要性[J]. 林业科学研究, 1996, 9(5): 498–503.
- [29] Xiao J H, Li J M, Yuan L P, *et al.* Dominance is the major genetic basis in rice as revealed by QTL analysis using molecular markers [J]. *Genetics*, 1995, 140(2): 745–754.
- [30] Yu S B, Li J X, Xu C G, *et al.* Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid[J]. *Proc Natl Acad Sci USA*, 1997, 94(17): 9226–9231.
- [31] 庄杰云, 樊叶杨. 超显性效应对水稻杂种优势的重要作用[J]. 中国科学: C 辑, 2001, 31(2): 106–113.
- [32] Meyer S, Pospisil H, Scholten S. Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern[J]. *Plant Molecular Biology*, 2007, 63(3): 381–391.
- [33] Hoecker N, Keller B, Muthreich N, *et al.* Comparison of maize (*Zea mays* L.) F1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of non-additive gene expression and conserved expression trends [J]. *Genetics*, 2008, 179(3): 1275–1283.

Transcriptome Analysis of Heterosis of Growth Traits in Chinese Fir

QI Ming¹, HE Gui-ping¹, ZHOU Jian-ge², CHENG Ya-ping³, XU Zhao-you⁴

(1. Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Key Laboratory of Tree Breeding of Zhejiang Province, Hangzhou 311400, Zhejiang, China; 2. Forestry Technology Extension Station of Suichang County, Zhejiang Province, Suichang 323300, Zhejiang, China; 3. Qingyuan Forest Farm of Qingyuan County, Zhejiang Province, Qingyuan 323805, Zhejiang, China; 4. Forestry Academy of Longquan City, Zhejiang Province, Longquan 323700, Zhejiang, China)

Abstract: [Objective] To study the changes of gene expression patterns in F1 hybrids of Chinese fir, analyze the transcriptomic profiling and to identify the associative genes related to heterosis, which presented a useful reference for the deep development and utilization of Chinese fir (*Cunninghamia lanceolata*). [Method] Using the latest second-generation hybrids (Long 15 × 1339), HF1 (3 super-parent hybrids), LF2 (3 low-parent hybrids) and parents (P1 and P2), all sample groups were compared by transcriptome sequencing. Since Chinese fir transcriptome sequencing is lack of reference genome, so it is needed to do the sequencing reads first for de novo splice to get unigenes and transcripts. Subsequently, the functional annotation of transcripts, differential expression and other bioinformatic analysis approaches were adopted. [Result] In a total of 12 samples, 5.8E + 08 clean reads were generated by transcriptome sequencing, and the total length was 49 803 726 pb. BLASTX analysis is performed on clean reads in six databases (Nr, Swiss-prot, KOG, KEGG, Pfam, GO), through sequence alignment which results in 80 171 unigenes. The Venn diagram analysis revealed 5 transmission modes of DEGs from the parents to the progenies: 1. parental expression but not expressed in hybrid (biparent silent type); 2. expression in one parent only, not in hybrids (parental specific expression); 3. only expressed in hybrids, not expressed in parents (hybrid specific expression); 4. expression in hybrid and one parent (single expression consistent); 5. expression in both parents and hybrids. 236 unigenes with differentially expressed (DEGs) were identified in HF1VSP1 comparison group, in HF1VSP2, LF2VSP1, and LF2VSP2, 1 483, 505, and 2 335 DEGs were revealed respectively. 100 significant different unigenes among parents and progenies are screened out, and the heatmap clustering analysis was carried out. The results revealed the molecular mechanism of the heterosis of Chinese fir was over-dominant. The sizes of clustering block were different, indicating that some traits were controlled by oligogenes and some by polygenes. The superparental dominance of Chinese fir growth was determined by the 14 genes down-regulated in long15 and up-regulated in the superparent progeny. [Conclusion] It is concluded that the molecular mechanisms of Chinese fir heterosis are over-dominant. The high productivity of Chinese fir superparental offsprings is related to 14 up-regulated genes. Environment significantly stimulates the up-regulated expression of these 14 genes, thus promotes the generation of growth advantages.

Keywords: Chinese fir; growth; *Cunninghamia lanceolata*; heterosis; transcriptome sequencing; molecular mechanism