

日本落叶松纤维素合酶基因片段的克隆及单核苷酸多态性分析

易敏, 张守攻, 谢允慧, 孙晓梅

(林木遗传育种国家重点实验室, 中国林业科学研究院林业研究所, 北京 100091)

摘要:【目的】纤维素合酶(celulose synthase, *CesA*)在植物纤维素合成途径中发挥主要调节作用,是控制木材纤维品质和产量的重要基因。从日本落叶松中分离克隆与纤维素合成相关的 *LkCesA* 基因,并对其进行核苷酸多样性以及连锁不平衡分析,为在日本落叶松中开展基于 *LkCesA* 基因的连锁不平衡作图及其辅助日本落叶松木材纤维性状的分子育种提供理论依据。【方法】依据日本落叶松转录组数据库检测到的纤维素合酶(*CesA*)基因 ESTs 序列设计引物,从日本落叶松中分离获得 *LkCesA* 基因片段。在此基础上,利用 DnaSP5.0 软件对日本落叶松 40 株基因型个体的 *LkCesA* 序列进行核苷酸多样性和连锁不平衡分析。【结果】从日本落叶松中成功克隆了 *CesA* 基因片段:该片段长 1 209 bp,包含部分开放阅读框,长度为 1 053 bp,可编码 350 个氨基酸,所推导的蛋白质氨基酸序列与火炬松 *Pl-CesA2* 的蛋白质氨基酸序列同源性为 95.4%。在日本落叶松 40 株基因型个体的 *LkCesA* 序列中共检测到 83 个 SNP 位点,SNP 发生频率为 1/21 bp,多样性指数 π_T 为 0.006 05。在这些 SNPs 中,69 个属于转换,14 个属于颠换,其中 19 个为常见 SNPs,64 个为罕见 SNPs。在外显子区域,共检测到 54 个 SNP 位点,其中 34 个为错义突变,20 个为同义突变。进一步的连锁不平衡分析显示,随着核苷酸序列长度的增加,SNP 连锁不平衡程度逐渐减弱。【结论】克隆到的 *LkCesA* 为植物 *CesA* 基因家族中的一员。*LkCesA* 基因的连锁不平衡在基因内部就已衰退,说明选择该基因作为候选基因,在日本落叶松中开展连锁不平衡作图用于指导日本落叶松的定向培育及木材品质改良是可行的。此外,在 *LkCesA* 基因中检测到多个常见 SNP 位点,为进一步开展该基因的连锁不平衡作图提供了材料。

关键词: 日本落叶松;纤维素合酶基因;基因克隆;单核苷酸多态性

中图分类号:S791.223

文献标识码:A

Isolation and Single Nucleotide Polymorphisms Analysis of Cellulose Synthase Gene Fragment in *Larix kaempferi*

YI Min, ZHANG Shou-gong, XIE Yun-hui, SUN Xiao-mei

(State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: Objective Cellulose synthase (*CesA*) plays a key role in the biosynthesis pathway of plant cellulose, which controls the quality and yield of wood fiber. This work aims at cloning the homologous genes cellulose-related from *Larix kaempferi*, and further analyzing the nucleotide diversity and linkage disequilibrium, which may provide important genetic foundation associated with *LkCesA* gene and gene-assisted breeding of new germplasms with desirable wood fiber traits in *L. kaempferi*. **Method** According to the expressed sequence tags (ESTs) of cellulose synthase from *L. kaempferi* transcriptome database, a cDNA fragment encoding *CesA* was isolated from *L. kaempferi* by gene-specific PCR amplification. The genomic sequences of *LkCesA* in 40 individuals were cloned and sequenced, then the single nucleotide polymorphisms (SNPs) diversity and linkage disequilibrium of *LkCesA* were analyzed u-

收稿日期:2014-05-26

基金项目:国家“973”计划项目(2012CB114506)

作者简介:易敏(1985—),女,在读博士研究生,主要从事落叶松遗传改良研究;电话:010-62889685;E-mail:yimin6104@163.com

* 通讯作者:博士,研究员,主要从事落叶松良种选育及定向培育技术研究;电话:010-62889685;E-mail:xmsun@caf.ac.cn

sing DnaSP5.0 software. **Result** The *CesA* fragment was 1 209 bp in length with a partial open reading frame (ORF, 1 053 bp) which would be capable to encode a predicted protein of 350 AA. The sequence indicated that the deduced amino acids shared 95.4% identity with *PtCesA2* from *Pinus taeda*. A total of 83 SNPs was detected, and the frequency and diversity of SNPs (π_T) were 1/21 bp and 0.006 05, respectively. There were 69 SNPs belonging to transition type and 24 belonging to transversion type, including 19 common SNPs and 64 rare SNPs. In total, 54 SNPs were detected in the coding regions of *LkCesA*, among which 20 were synonymous mutation and 34 were missense mutation. The results of linkage disequilibrium analysis showed that the LD declined rapidly with the nucleotide length. **Conclusion** The results of this study indicate that *LkCesA* is a member of *CesA* gene family. The linkage disequilibrium declined rapidly within the gene regions of *LkCesA* suggests that the gene could be used as the candidate gene for LD mapping, which contribute to the directional cultivation and wood quality improvement in Japanese larch. In addition, many common SNPs were detected in *LkCesA* which could be used in further LD mapping.

Key words: *Larix kaempferi*; cellulose synthase; gene cloning; single nucleotide polymorphism

纤维素是植物细胞壁的主要成分之一,也是地球上最丰富的天然有机物和重要的可再生资源^[1]。在树木中,纤维素含量为 40 ~ 50%^[2],是造纸和纸浆业最重要的天然原料,可以创造巨大的商业价值。由于日本落叶松(*Larix kaempferi* (Lamb.) Carr.) 具有生长迅速、适应性强、木材密度较其他针叶树种高、造纸工业利用率高等优点,被认为是有广阔市场前景的短周期纸浆用材树种^[3-4]。因此,深入研究纤维素的生物合成途径,挖掘控制木材纤维品质和产量的重要基因,对于日本落叶松的定向培育及木材品质改良具有重要意义。

纤维素合成酶 (cellulose synthase, *CesA*) 是植物纤维素合成途径中的关键酶,通常以一个六边玫瑰型复合体的形式存在于细胞质膜上^[5]。1996 年 Pear 等^[6] 从棉花 (*Gossypium hirsutum* Linn) 中克隆了第 1 个植物纤维素合酶基因,随后又相继在拟南芥 (*Arabidopsis thaliana* (L.) Heynh.)^[7]、玉米 (*Zea mays* L.)^[8]、水稻 (*Oryza sativa* L.)^[9] 等植物中分离克隆出多个 *CesA* 基因。林木中首次克隆的纤维素合酶基因是从美洲山杨 (*Populus tremuloides* Michx) 中得到的 *PtrCesA1* 基因,表达分析表明该基因在木质部特异表达,与次生壁的形成有关^[10]。随后,又在毛果杨 (*P. trichocarpa* Torr. & Gray) 中鉴定出了 18 个 *CesA* 基因,其中 *PtrCesA4*, *PtrCesA5*, *PtrCesA7*, *PtrCesA8*, *PtrCesA17* 和 *PtrCesA18* 都在木质部特异表达^[11-12]。而针叶树纤维素合酶的研究起步较晚,直到 2005 年才从新西兰辐射松 (*Pinus radiata* D. Don) 中克隆到 1 个全长的 *PrCesA10* 基因,该基因在次生木质部中表达水平较高^[13]。之后,又在火炬松 (*Pinus taeda* L.) 中克隆到 3 个与次生细胞壁合成有

关的 *CesA* 基因^[14]。上述研究显示,虽然纤维素合酶基因已在部分树种中分离克隆,且这些基因都不同程度的参与了植物初生、次生细胞壁的形成,但不同植物的纤维素合酶基因的成员数量、结构以及表达模式不尽相同,而且目前关于日本落叶松纤维素合酶基因的克隆研究也罕有报道。所以,在日本落叶松中克隆 *CesA* 基因成员具有重要的应用价值。

为此,本研究以选自日本落叶松原生种源/家系试验林能最大程度地反映日本落叶松分布范围的 40 株基因型个体为材料,克隆 *LkCesA* 基因片段,分析该基因的结构和系统进化,为后续的 *LkCesA* 基因全长的克隆及功能分析提供了试验依据。另外还在在此基础上对该基因片的单核苷酸多态性 (single nucleotide polymorphisms, SNPs) 和连锁不平衡 (linkage disequilibrium, LD) 进行了分析,为在日本落叶松中开展 *LkCesA* 基因的连锁不平衡作图奠定基础。

1 材料和方法

1.1 植物材料

以 1998 年建于湖北长岭岗林场的日本落叶松种源/家系试验林为材料,该试验林包括日本落叶松自然分布区内 6 个种源 (产地) 的 84 个天然林分 (草津林分 20 个、浅间林分 6 个、富士林分 16 个、日光林分 18 个、伊那林分 15 个、松本林分 9 个) 和日本国内各种子园的 127 个家系 (北海道美呗种子园 4 个、栗山种子园 45 个、岩手县盛冈种子园 7 个、长野县小诸种子园 67 个、江刺种子园 2 个以及北海道和长野县 2 个混和), 共 211 个林分/家系,同时以辽宁丹东、山东崂山和湖北建始的 3 个母树林及 1 个辽宁种子园做为对照。从这片试验林中随机选取了

40 株最能反映日本落叶松天然分布范围的基因型个体(表 1),采集完全展开的新叶置硅胶中带回实

验室放于超低温冰箱(-80℃)中保存备用。

表 1 供试材料来源

| 产地 | 种子来源 | 林分/ 家系数 | 入选 株数 | 采种地情况 | | | | |
|-------|--------------|------------|----------|--------|---------|-------|---------|---------|
| | | | | 纬度 | 经度 | 海拔/m | 年平均气温/℃ | 年降水量/mm |
| 日本 | 草津国有林天然林 | 20 | 6 | 36°35' | 138°30' | 1 374 | 6.9 | 1 747 |
| | 浅间国有林天然林 | 6 | 2 | 36°20' | 138°30' | 927 | 9.3 | 1 360 |
| | 富士国有林天然林 | 16 | 6 | 35°30' | 138°45' | 844 | 10.5 | 1 595 |
| | 伊那国有林天然林 | 15 | 5 | 35°45' | 138°10' | 1 726 | 5.9 | 1 795 |
| | 日光国有林天然林 | 18 | 7 | 36°45' | 139°25' | 1 444 | 5.9 | 2 003 |
| | 松本国有林天然林 | 9 | 3 | 36°20' | 139°45' | 1 620 | 14.0 | 1 266 |
| 日本北海道 | 北海道美呗种子园 | 4 | 1 | 43°20' | 141°50' | 20 | 7.0 | 1 305 |
| | 北海道栗山种子园 | 45 | 1 | 43°08' | 141°45' | 26 | 7.2 | 1 238 |
| | 北海道远轻种子园(混合) | 1 | 1 | 44°00' | 143°20' | 236 | 5.3 | 929 |
| 日本岩手县 | 岩手县盛冈种子园 | 7 | 1 | 39°42' | 141°10' | 135 | 10.3 | 1 310 |
| | 岩手县江刺种子园(混合) | 2 | 1 | 39°10' | 141°12' | 40 | 10.7 | 1 203 |
| 日本长野县 | 长野县小诸种子园 | 67 | 1 | 36°20' | 138°40' | 843 | 9.9 | 1 578 |
| | 长野县安云种子园(混合) | 1 | 1 | 36°10' | 137°45' | 1 315 | 7.1 | 1 913 |
| 中国 | 辽宁丹东母树林(混合) | 1 | 1 | 40°30' | 124°05' | 320 | 8.0 | 750 |
| | 山东崂山母树林(混合) | 1 | 1 | 36°05' | 120°25' | 110 | 12.0 | 1 000 |
| | 辽宁清源种子园(混合) | 1 | 1 | 42°45' | 125°48' | 550 | 6.0 | 650 |
| | 湖北建始母树林(混合) | 1 | 1 | 30°45' | 109°45' | 1 700 | 15.0 | 1 600 |
| 总计 | | 211 | 40 | | | | | |

1.2 总 DNA、总 RNA 提取及 cDNA 合成

总 DNA 的提取参照 DNAsure Plant Kit(TIAN-GEN)说明书描述的方法进行。总 RNA 的提取使用 TAKARA 公司试剂盒(RNAiso Plus),cDNA 第一链的反转录合成使用 TAKARA 公司试剂盒(Reverse Transcriptase M-MLV),按试剂盒的说明进行操作。

1.3 *LkCesA* 基因片段的克隆

对本研究前期测序获得的日本落叶松转录组数据库^[15]进行比对分析后获得 *CesA* 基因片段的 EST 序列,以该序列为模板,使用 Primer Premier 5.0 软件设计一对特异性引物,上游引物:5'-CATCAACGCTTATG-GAAAACGG-3';下游引物:5'-TTGTGTTTTATCCTTCG-GTTCA-3'。由深圳华大基因公司合成。

扩增的 PCR 反应体系为 30 μL:LA 10 × buffer 3 μL、dNTP 4 μL(10 mmol · L⁻¹)、LA Taq DNA 聚合酶 0.3 μL(5 U/μL)(以上试剂购自 TAKARA 公司),上游和下游引物各 1.2 μL(10 μmol · L⁻¹)、cDNA 1 μL、ddH₂O 19.3 μL。扩增程序为:94℃ 预变性 5 min,94℃ 变性 30 s,56℃ 退火 30 s,72℃ 延伸 60 s,反应共 35 个循环,再经 72℃ 延伸 7 min 后于 4℃ 保存。PCR 扩增产物经胶回收与 pMD 19-T Simple Vector 载体连接转化大肠杆菌 DH5α,经菌落检测的阳性克隆送至上海美吉公司进行测序。

1.4 目的基因的序列分析

应用 DNAMAN6.0 软件对目的基因的 cDNA 编码氨基酸序列进行分析;并用 BioEdit version 7.0 软件和 BLAST 检索 GenBank 进行多重序列比较和同源性分析;用 MEGA5.05 软件的 Neighbor-Joining (NJ) 算法分析同源基因间的进化关系,并构建该蛋白的系统进化树。

1.5 *LkCesA* 基因的 SNP 多样性分析

利用 BioEdit 的 Clustal W 程序对同一基因片段的 40 条拼接完整的序列进行比对,获得基因片段的一致序列,结合 Dnasp5.0 软件进行分析,标出 SNP 位点,计算 SNP 频率、转换和颠换的 SNP 数量,分析同义突变和非同义突变的碱基变异情况及对相应氨基酸的影响,估算基因的核苷酸多样性水平,并进一步分析连锁不平衡的延伸模式。

2 结果与分析

2.1 日本落叶松 *LkCesA* 基因片段的克隆及其序列分析

分别以日本落叶松 cDNA 和总 DNA 为模板,依据日本落叶松转录组数据序列检索到的 *CesA* EST 序列设计特异性引物进行 PCR 扩增,经 1% 的琼脂糖电泳检测,发现特异性扩增出 1 条约 1 200 bp(图

1 A) 和 1 条约 1 700 bp (图 1 B) 的基因组 DNA 片段。

经回收、连接、转化、测序之后发现,克隆的 cDNA 片段长为 1 209 bp,包括 156 bp 的 3 端非编码区 (3'Un-translated region,3'UTR) 和部分开放阅读框,大小为 1 053 bp,可编码 350 个氨基酸。克隆的基因组 DNA 片段长为 1 715 bp,与 cDNA 序列对比发现该片段包含 3 个外显子和 2 个内含子,其中第一外显子不完全。外显子长度分别为 117 bp、351 bp 和 585 bp,符合典型的 GT/AG 规则(图 2)。

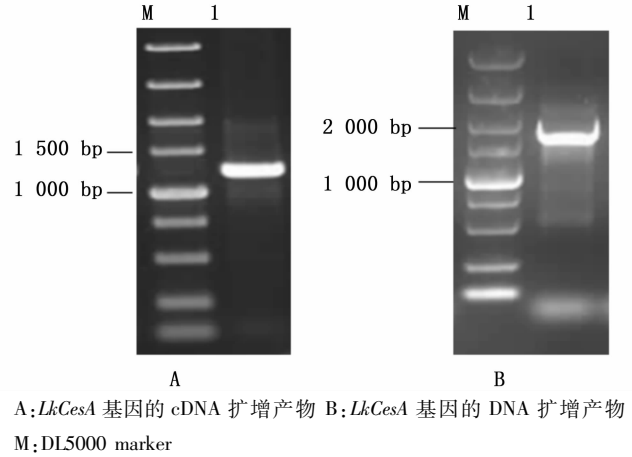


图 1 *LkCesA* 基因的 PCR 扩增结果

| | | | | | | |
|------|--------------------|-------------------|-------------------|--------------------|--------------------|-------------------|
| 1 | TCAAGCCTTA | TGGAAAACGG | TGGTGTCCA | GAGTCTGTA | ACTCACCTTC | TCTGATAAAG |
| 61 | GAAGCCATTC | ATGTCATAAG | CTGTGGATAT | GAGGAGAAAA | CCGAGTGGGG | AAAAGAGGTA |
| 121 | TGAATCTATT | GAAGCTAAGC | ATTTGGTGCC | TGTATGATTA | TATAAAGAAT | TACAAAAGGA |
| 181 | TTGATAACCA | CAGTTTTCAA | TTTGCAATAT | AGTCAATGTT | TATGATGTCT | CTGAAACATC |
| 241 | TTGGTTGCCA | AACTTATAAA | TTACTCAAAT | TTTCATGGCC | AATCAATAC | CACCTTATACA |
| 301 | GAATAACTGA | TTTCTTCAC | TGTTGACAGA | TCGGATGGAT | ATATGGTTCA | GTGACAGAGG |
| 361 | ATATCTGAC | AGGCTTCAA | ATGCACTGCC | GTGGTTGGCG | ATCGGTATAC | TGCATGCCCA |
| 421 | ACAGGCCAGC | ATTCAAAGGA | TCGGCTCCTA | TTAATTTATC | GGATCGTTTG | CACCAGGTTT |
| 481 | TTCGTTGGGC | AITGGGTTCC | ATCGAAATTT | TCATGAGTAG | ACATTCGCCA | CTCTGGTATG |
| 541 | CCTATGGAGG | AAAGCTAAAA | TGGCTTGAAA | GATTCGCCTA | TGTCATATCA | ATCGTCTACC |
| 601 | CATTCACCTC | AATCCCGCTC | GTGCTTACT | GTACTCTCCC | TGCCATATGT | CTTCTTACAG |
| 661 | GAAAATTCAAT | CATCCCAACG | GTATGGAATT | ACATGAAGAA | TCAAAAATTAC | TCAACCTTTT |
| 721 | ATAAATCAIT | TGAGTGTCTG | TCTTACTGGT | TGGATAAATT | AGTGTTCTA | GCTTGTGAA |
| 781 | TGAATTAGCT | TCTGCTGCAT | CTCATTTTCA | ATATCTTATA | GTTTACAAAA | TAGGGATAAA |
| 841 | CTTTCATGG | GTTTATATTT | TTCCATGTGT | TTCACTGTTG | AGATAGTTAG | ATATGATTAG |
| 901 | ACATGAANAAC | TAACTTTAA | TTGACAAATT | TTGTCAAATT | TCTCATGTAA | GTAAGTGTG |
| 961 | TTTATGTATT | GCAGCTTGCA | AGCCTTGCAA | GCGTCTGGTT | CATGGGACTT | TTCTTATCCA |
| 1021 | TCATAGCGAC | AGGAGTCTA | GAGCTGAGAT | GGAGCGGAGT | GAGCATCGAA | GAGTTCGGA |
| 1081 | GAAACGAACA | GTTCTGGGTG | ATCGGAGGAG | TTTCTGCTCA | CCTTTTTCGA | GTTTTCAGG |
| 1141 | GCTGTCTCAA | GGTGTGGCG | GGTGTGGACA | CCAATTTTAC | AGTCACAGCC | AAGGGTAGCG |
| 1201 | ATGAAGAAGA | CCAATTCGGC | GAATATACT | TGTTCAAATG | GACGACCCGT | CTCATCCCCC |
| 1261 | CAACCACCTT | GCTGATAATA | AACTCTGGTA | GCCTCGTTGC | AGGTGTTTCG | TCTGCTGTAA |
| 1321 | ACAATAAATA | CCAATCATGG | GGACCGCTCT | TTGGAAAAGCT | GTCTCTCGCC | TTCTGGGTGA |
| 1381 | TATTCATCT | GTATCCCTTC | CTCAAGGGTC | TTCTGGGAAG | GCAGAACCGA | ACGCCCACTA |
| 1441 | TAGTTGTAT | GTGGTCTAAT | CTGCCGGCTT | CCATCTTCTC | GCTTATCTGG | GTCAGGATCG |
| 1501 | ACCCGTTCT | ACCCAAGGTG | GAAGGCCCAA | TTCTTCAGCA | ATGTGGCGTA | GATGCTATG |
| 1561 | TAGCTCCTCT | CTTCTGATCA | TTTATGAATT | CTATTCATTT | GTATTGAGAT | TTTGTAAATGG |
| 1621 | ACTGAAGGAT | AAAACCTGAAG | GCAATCGAAT | AGATAGTTTA | TGAATGCTAT | TCATTCGGAT |
| 1681 | GCAAATTCCTG | TAATGAACCG | AAGGATAAAA | CACAA | | |

图 2 日本落叶松 *LkCesA* 基因片段的核苷酸序列(阴影部分为外显子区域)

将日本落叶松 *CesA* 基因与 GenBank 中 4 个 *CesA* 基因所编码蛋白的氨基酸序列进行同源比较和结构分析,结果表明(图 3):*LkCesA* 与火炬松 *PrCesA2* 的氨基酸同源性最高,达到 95.4%;与拟南芥 *AtCesA2* 的同源性最低,为 69.9%;而与巨桉 *EgCesA2* 和美洲山杨 *PtrCesA2* 的蛋白质氨基酸序列同源性分别为 84.6% 和 78.6%。典型的植物纤维素合酶基因含有 2 个高度保守的亚结构域(Subdomains A 和 B),在 Subdomains A 中含有 2 个保守的天冬氨酸残基(D),主要为纤维素合成所需底物(UDP-G)

的结合域;Subdomains B 中除含有 1 个保守的天冬氨酸残基外,还有另一个 QxxRW(谷氨酰胺-x-x-精氨酸-色氨酸)保守序列,其功能与纤维素合成酶的催化活性有关^[16-17]。另外还含有 8 个高度保守的跨膜结构域,其中 2 个位于 N-端,6 个位于 C-端,是 β -1,4-葡萄糖苷链穿过质膜进入细胞壁的重要通道^[18]。从图 3 可以看出,*LkCesA* 基因片段的氨基酸序列也含有位于 C 端的 6 个跨膜结构域区域,同时还含有位于 Subdomains B 中的一个保守 D 残基和 QxxRW 保守序列。

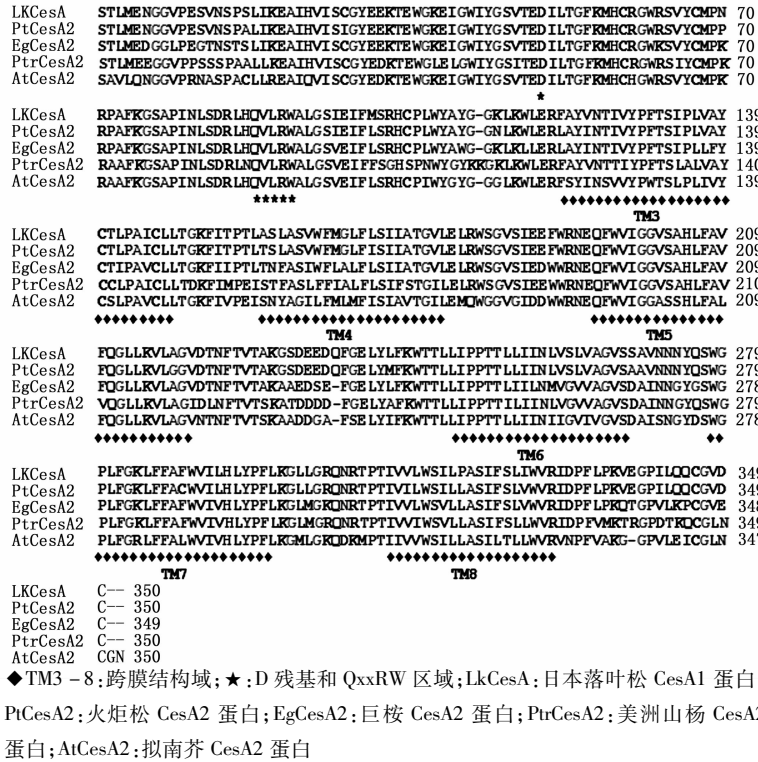
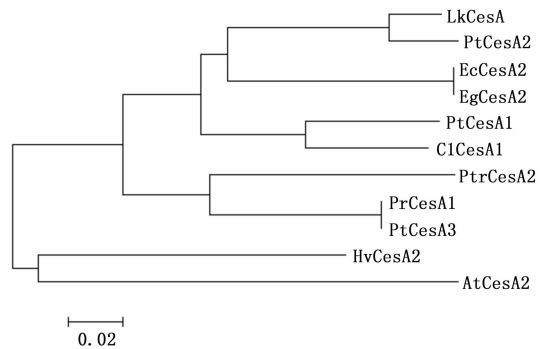


图3 日本落叶松与其它植物 CesA 基因片段的蛋白质一级结构比较

为了分析日本落叶松 LkCesA 与其他植物 CesA 蛋白的系统发育进化关系,利用 MEGA 软件将在 NCBI 注册的拟南芥、大麦、火炬松、海岸松、杉木、赤桉、巨桉、美洲山杨等 8 个物种的氨基酸序列进行序列分析,得到了这些 CesA 蛋白的系统进化树(图 4)。由图 4 可以看出, Ces 蛋白由极点向 2 个方向发生了进化,形成了 2 个分枝。拟南芥 AtCesA2 和大麦 HvCesA2 聚合形成第一分枝。在第 2 分支中又分为 2 个亚分支,分别向裸子植物和被子植物方向发生分歧,其中日本落叶松 LkCesA 与火炬松 PtCesA2 最先聚合,亲缘关系最接近。

2.2 LkCesA 基因的 SNP 多样性分析

对 LkCesA 进行 SNPs 多样性分析可以获得该基因在日本落叶松进化史上所发生的单核苷酸突变位点,是进行连锁不平衡作图的基础。因此,在分析 LkCesA cDNA 序列的基础上,对取自日本落叶松自然分布区、能最大范围覆盖其生长区域的 40 株基因型个体进行克隆测序,然后利用 BioEdit 软件的 Clustal W 程序对 40 条拼接完整的序列进行比对,将出现的单碱基变化的位点确定为 SNP 位点。图 5 是 LkCesA 基因的部分核苷酸序列的比对结果,在第 1 525 bp 处出现的 C-T 变化、第 1 532 bp 处出现的 T-C 变化、第 1 536 bp 处出现的 A-G 变化和第 1 556 bp 处出现的 C-T 变化记为 4 个 SNP 位点。



At: 拟南芥 *Arabidopsis thaliana* (L.) Heynh.; Hv: 大麦 *Hordeum vulgare* L.; Pt: 火炬松 *Pinus taeda* L.; Cl: 杉木 *Cunninghamia lanceolata* (Lamb) Hook.; Pr: 海岸松 *Pinus pinaster* Aiton; Ptr: 美洲山杨 *Populus tremuloides* Michx.; Ec: 赤桉 *Eucalyptus camaldulensis* Dehnh.; Eg: 巨桉 *Eucalyptus grandis* Hill ex Maiden

图4 LkCesA 与其他物种 CesA 的系统发育进化树

由表 2 可知,长度为 1 715 bp 的序列中共检测到 83 个 SNPs 位点,SNP 频率为每 21 bp 基因组序列出现 1 个 SNP 位点,其中外显子、内含子和 3'UTR 的 SNP 位点分别为 54、24 和 5 个,频率分别为 1/20、1/21 和 1/31。另外,在 LkCesA 基因中还发现了 7 个插入/缺失标记,长度在 1~21 bp。这 7 个插入/缺失都发生在基因的非编码区,其中有 4 个为单碱基插入/缺失,1 个双碱基插入/缺失,剩下两个的插入/缺失长度分别为 13 bp 和 21 bp。

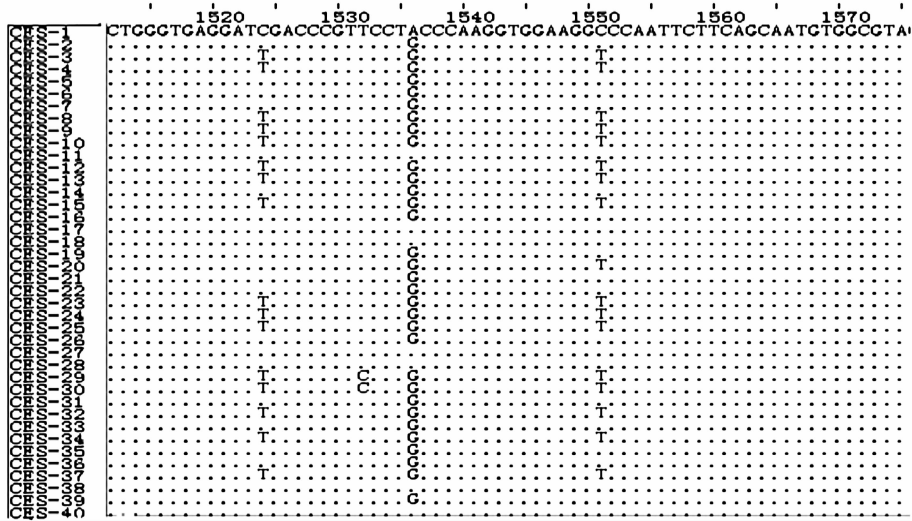


图5 *LkCesA* 基因 SNP 位点筛选示意图

表2 *LkCesA* 基因内 SNP 数量及其频率

| <i>LkCesA</i> | 总计 | 外显子 | 内含子 | 3'UTR |
|---------------|-------|-------|------|-------|
| 序列长度 | 1 715 | 1 053 | 506 | 156 |
| SNP 数量 | 83 | 54 | 24 | 5 |
| SNP 频率 | 1/21 | 1/20 | 1/21 | 1/31 |

进一步分析了 *LkCesA* 基因中的 83 个 SNPs 位点的变异类型(表 3),结果显示:83 个 SNPs 中,有 69 个属于转换类型,分别包含 30 个 A↔G 和 39 个 T↔C;有 14 个属于颠换类型,分别包含 6 个 A↔C、2 个 T↔G、5 个 A↔T 和 1 个 G↔C,转换/颠换的比例为 4.9:1。

表3 *LkCesA* 基因内 SNPs 变异类型

| 基因 | SNP 数量 | 替代类型 | | | | | |
|---------------|--------|------|-----|-----|-----|-----|-----|
| | | 转换 | | 颠换 | | | |
| | | A↔G | T↔C | A↔C | T↔G | A↔T | G↔C |
| <i>LkCesA</i> | 83 | 30 | 39 | 6 | 2 | 5 | 1 |

2.3 SNPs 引起的氨基酸变异

为了检测 *LkCesA* 基因编码区内核苷酸位点的改变是否影响了其编码的氨基酸序列,在分析 SNP 多样性的基础上对 SNP 的突变类型进行了统计,并对其引起的氨基酸变异进行了分析。从表 4 可以看出,在 *LkCesA* 基因内部检测到的 83 个 SNPs 中,有 19 个为常见 SNPs(频率 ≥ 5%),64 个为罕见 SNPs。在编码区内的 54 个 SNPs 位点中,有 20 个属于同义突变,其中有 2 个位于密码子的第 1 个核苷酸上,剩下的 18 个均位于密码子的第 3 个核苷酸上;另外 34 个属于错义突变,没有检测到无义突变。在这 34 个错义突变中,有 12 个位于密码子的第 1 个核苷酸

上,21 个位于密码子的第 2 个核苷酸,此外还有 1 个位于密码子的第 3 个核苷酸上。其中,有 2 个属于常见 SNPs,分别位于密码子的第 1 和第 2 个核苷酸上,其密码子由原来的 ATT 和 TTC 转变为 GTT 和 TCC,氨基酸由原来的异亮氨酸和苯丙氨酸转换为缬氨酸和丝氨酸。

表4 *LkCesA* 基因内 SNPs 的鉴定

| 基因 | 常见 SNP | 罕见 SNP | 非同义突变 | | 同义突变 | 非编码区 |
|---------------|--------|--------|-------|------|------|------|
| | | | 错义突变 | 无义突变 | | |
| <i>LkCesA</i> | 19 | 64 | 34 | 0 | 20 | 29 |

2.4 *LkCesA* 基因的核苷酸多样性分析

使用 Dnasp 软件进行 *LkCesA* 基因的序列多样性分析。以 θ 值和 π 值两个参数衡量候选基因的核苷酸多样性。从表 5 可以看出,日本落叶松 *LkCesA* 基因的总核苷酸多样性参数 θ_w 和 π_T 分别为 0.011 91 和 0.006 05,其中 $Ka/Ks = 0.422 23 < 1$,说明该基因的同义突变遗传变异大于非同义突变,可以推断纯化选择是 *LkCesA* 基因内同义 SNP 位点的主要选择压,核苷酸序列比较保守。采用四配子检测进行最小历史重组事件(R_M)估算结果表明,*LkCesA* 基因的 R_M 值为 5,除以信息位点数后得到每一位点发生的最小重组事件为 0.06。利用 Tajima's D^* 和 Fu and Li's D^* 进行的中性检验结果表明, D^* 值均为负值,而且 Fu and Li's D^* 值还达到显著水平,表明 *LkCesA* 基因可能受到了负选择的作用,而且还存在低频率的等位基因,这与在 *LkCesA* 基因中发现大量的罕见 SNPs 是一致的。

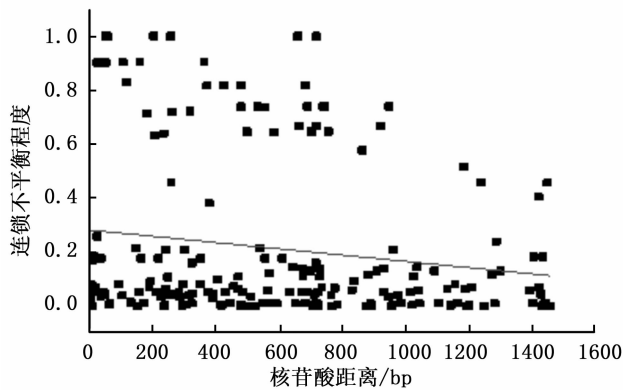
表5 *LkCesA* 基因的核苷酸多样性

| 基因 | S | θ_w | π_T | π_{sil} | π_{syn} | $\pi_{non-syn}$ | Ka/Ks | H | HD | R _M | R _M /S | Tajima' D* | Fu and Li' D* |
|---------------|----|------------|---------|-------------|-------------|-----------------|---------|----|-------|----------------|-------------------|------------|---------------|
| <i>LkCesA</i> | 83 | 0.01191 | 0.00605 | 0.00922 | 0.01837 | 0.00247 | 0.42223 | 38 | 0.997 | 5 | 0.06 | -1.79311 | -4.27481** |

注:S:分离位点数; π_T :整个基因区域单核苷酸多样性; π_{sil} :编码区同义位点及其他区域位点多样性; π_{syn} :同义位点多样性; $\pi_{non-syn}$:非同义位点多样性;Ks:同义突变率;Ka:非同义突变率;H:单体型;HD:单体型多样性;RM:最小历史重组事件;** * $P < 0.01$

2.5 *LkCesA* 基因内 SNPs 的连锁不平衡分析

为了了解 *LkCesA* 基因内 SNP 位点在日本落叶松 40 株基因型个体间连锁不平衡的延伸长度及程度,利用 DnaSP 软件中的 LD 分析程序分析了 *LkCesA* 基因中 SNP 的连锁不平衡水平。由图 6 可以看出,随着 *LkCesA* 基因核苷酸序列长度的增加,SNP 的 LD 程度逐渐减弱,当长度达 680 bp 左右时, r^2 值迅速下降到 0.2 以下,说明 SNP 的连锁不平衡在 *LkCesA* 基因内部就已经衰退了。

图6 日本落叶松 *LkCesA* 基因内 SNPs 的连锁不平衡

3 结论与讨论

通过对日本落叶松转录组数据的查找分析,克隆获得了一个长为 1 209 bp 的 *CesA* 基因片段序列,核苷酸序列分析结果显示,其与多种植物的 *CesA* 基因具有较高的同源性,可初步确定其为植物 *CesA* 基因家族中的一员。多重序列比对结果亦显示其编码的氨基酸序列中含有纤维素合酶所特有的位于 C-端的 6 个跨膜结构域,同时还含有位于 Subdomains B 中的一个保守 D 残基和 QxxRW 区域,从而进一步证明本文克隆到的序列为日本落叶松的纤维素合酶基因片段。目前 GenBank 中尚无落叶松属植物纤维素合酶基因的研究,日本落叶松 *LkCesA* 基因片段的克隆不仅能为 *LkCesA* 基因全长的克隆及功能分析提供信息资源,也为分离其它近源种 *CesA* 基因奠定了基础。

对 *LkCesA* 基因进行 SNP 发生频率的研究,结果

表明,长度为 1 715 bp 的序列中共含有 83 个 SNP 位点,SNP 频率为 1/21 bp,与周亚楠^[19]等对日本落叶松 4CL 基因检测的 1/20.7 bp 的频率相当,高于花旗松 (*Pseudotsuga menziesii* (Mirbel) Franco) 的 1/46 bp^[20],低于毛白杨 (*Populus tomentosa* Carr) *PtUXS1* 基因的 1/18 bp^[21]。这可能是由于不同物种和基因在进化过程中受到的选择压不同导致基因的保守性不同。

对 *LkCesA* 基因的核苷酸多样性的分析表明:*LkCesA* 基因的核苷酸多样性参数 π_T (0.006 05) 与花旗松抗寒/木材质量相关候选基因中发现的多样性程度相类似 ($\pi = 0.006 55$)^[20],高于西伯利亚落叶松 ($\pi = 0.005 36$)^[22]。但与其他针叶树种相比,日本落叶松的多样性水平要比云杉属高出 2~3 倍 ($\pi = 0.002 \sim 0.003$)^[23],略低于毛白杨的 *PtSUS1* ($\pi = 0.00924$) 和 *PtSUS2* ($\pi = 0.010 93$)^[24]。由此推测日本落叶松 *LkCesA* 基因的核苷酸多样性属于中等水平,这主要是因为日本落叶松是一个以异交为主的树种,而且本实验中所选的材料最大程度地反映日本落叶松的天然分布范围,所以核苷酸多样性水平较高。而不同的树种由于分布范围、进化历史以及材料来源、群体大小的不同,会产生不同的遗传基础,从而造成核苷酸多样性的差异。另外,根据所选候选基因的类别、数量以及完整性的不同,核苷酸多样性也会发生相应的变化,所以说在不同树种间进行核苷酸多样性的精确比较还很难。

本研究所得到的 *LkCesA* 基因每一位点发生的最小重组事件数为 0.06。与之前日本柳杉 (*Cryptomeria japonica* (L. f.) D. Don) 内检测到的最小重组事件值 (0.000 0~0.052 6) 相类似^[25],但比欧洲赤松 (*P. sylvestris* L.) 的要低 (0.25)^[26]。造成这一结果的主要原因可能是日本落叶松和日本柳杉的天然分布范围类似,而欧洲赤松则广泛分布在欧亚大陆,有效群体数远大于日本落叶松,从而导致较高的重组率。

了解连锁不平衡的延伸模式对关联分析来说十分重要,因为 LD 程度不仅影响关联作图的分辨率,而且还决定了在基因组扫描过程中所需的最少标记

数目^[27]。所以究竟是选择全基因关联分析策略还是候选基因关联分析策略主要取决于目标群体内的LD程度。由本研究的结果可以看出,随着 *LkCesA* 基因核苷酸序列长度的增加,SNPs 连锁不平衡程度逐渐减弱,当长度达 680 bp 左右时, r^2 值迅速下降到 0.2 以下。与之前报道的火炬松^[28]、樟子松(*P. sylvestris* var. *mongolica* Litv.)^[29]、花旗松^[20]等的结果一致,SNP 的连锁不平衡同样在候选基因内部就已衰退,也进一步说明在日本落叶松中基于候选基因的 LD 作图是可行的。而在 *LkCesA* 基因中检测到多个 SNP 位点,也将为今后开展日本落叶松材性相关基因的连锁不平衡作图提供理论依据。

参考文献:

- [1] Delmer D P. Cellulose biosynthesis: exciting times for a difficult field of study[J]. *Annu Rev Plant Physiol Mol Biol*,1999,50: 245 - 276
- [2] Fengel D, Wegener G. Wood: chemistry, ultrastructure, reactions [M]. Berlin: Walter de Gruyter, 1984,26 - 65.
- [3] 王军辉,张守攻,石淑兰,等. 日本落叶松纸浆材造纸性能的研究[J]. 北京林业大学学报,2004,26(5):71 - 74.
- [4] 孙晓梅,张守攻,李时元,等. 日本落叶松纸浆材优良家系多性状联合选[J]. 林业科学,2005,41(4):48 - 54.
- [5] Kimura S, Laosinchai W, Itoh T. Immunogold labeling of rosette terminal cellulose synthesizing complexes in the vascular plant *Vigna angularis* [J]. *Plant Cell*,1999,11:2075 - 2085.
- [6] Pear J R, Kawagoe Y, Schreckengost W E, et al. Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase[J]. *Proceedings of the National Academy of Sciences, USA*,1996,93:12637 - 12642.
- [7] Richmond T A, Somerville C R. The cellulose synthase superfamily [J]. *Plant Physiology*,2000,124:495 - 498.
- [8] Appenzeller L, Doblin M, Barreiro R, et al. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family[J]. *Cellulose*,2004,11,287 - 299.
- [9] Tanaka K, Murata K, Yamazaki M, et al. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall[J]. *Plant Physiology*,2003,133,73 - 83.
- [10] Wu L, Joshi C P, Chiang V L. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress[J]. *Plant Journal*,2000,22:495 - 502.
- [11] Soraya D, Mats L, Lars A, et al. The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes[J]. *Planta*,2005,221(5):739 - 746.
- [12] Suzuki S, Li L, Sun Y H, et al. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa* [J]. *Plant Physiology*,2006,142:1233 - 1245.
- [13] Krauskopf E, Harris P J, Putterill J. The cellulose synthase gene *PrCESA10* is involved in cellulose biosynthesis in developing tracheids of the gymnosperm *Pinus radiata* [J]. *Gene*,2005,350(2): 107 - 116.
- [14] Nairn C J, Haselkorn T. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms [J]. *New Phytologist*,2005,166:907 - 915.
- [15] 许晨曦. 落叶松基因挖掘及基因差异表达与杂种优势形成机制探讨[D]. 北京:中国林业科学研究院,2013.
- [16] Joshi C P, Bhandari S, Ranjan P, et al. Genomics of cellulose biosynthesis in poplars [J]. *New Phytologist*,2004,164: 53 - 61.
- [17] Richmond T A, Somerville C R. The cellulose synthase super family [J]. *Plant Physiol*,2000,124:495 - 498.
- [18] Richmond T. Higher plant cellulose synthases [J]. *Genome Biol*, 2000,1(4):3001 - 3006.
- [19] 周亚楠,李 爱,陈成彬,等. 日本落叶松杂种及亲本 *4CL* 基因的克隆和 SNP 多态性分析[J]. 林业科学研究,2013,(专刊): 018 - 024.
- [20] Krutovsky K V, Neale D B. Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas fir [J]. *Genetics*,2005,171:2029 - 2041.
- [21] Du Q, Pan W, Tian J, et al. The UDP - glucuronate decarboxylase gene family in *Populus*: structure, expression, and association genetics [J]. *PLoS One*,2013, 8: e60880.
- [22] Semerikov V L, Semerikova S A, Polezhaeva M A. Nucleotide diversity and linkage disequilibrium of adaptive significant genes in *Larix* (Pinaceae) [J]. *Russian Journal of Genetics*,2013,49:915 - 923.
- [23] Namroud M C, Guillet - Claude C, Mackay J, et al. Molecular evolution of regulatory genes in spruces from different species and continents: heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes [J]. *J Mol Evol*, 2010,70:371 - 386.
- [24] Zhang D, Xu B, Yang X, et al. The sucrose synthase gene family in *Populus*: structure, expression, and evolution [J]. *Tree Genetics & Genomes*,2010,7:443 - 456.
- [25] Kado T, Yoshimaru H, Tsumura Y, et al. DNA variation in a conifer, *Cryptomeria japonica* (Copressaceae sensu lato) [J]. *Genetics*, 2003,164:1547 - 1599.
- [26] Dvornyk V, Sirviö A, Mikkonen M, et al. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris* [J]. *Molecular Biology and Evolution*,2002,19:179 - 188.
- [27] Whitt S R, Buckler E B. Using natural allelic diversity to evaluate gene function [J]. *Methods Mol Biol*,2003;236:123 - 140.
- [28] González-Martínez S C, Ersoz E, Brown G R, et al. DNA sequence variation and selection of tag single nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L [J]. *Genetics* 2006,172:1915 - 1926.
- [29] Dvornyk V, Sirviö A, Mikkonen M, et al. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris* [J]. *Molecular Biology and Evolution*,2002,19:179 - 188.