杨树热激转录因子 *HsfA1d* 的克隆、表达及单核苷酸多态性分析*

赵 杏 郭 琦 陈清清 李百炼 张德强
（北京林业大学 林木育种国家工程实验室 林木花卉遗传育种教育部重点实验室 北京 100083）

摘 要： 利用 RT-PCR 方法，首次由小叶杨受热激胁迫后构建的 cDNA 文库中扩增获得一热激转录因子 *HsfA1d* cDNA 克隆，测序结果表明克隆的 *PsHsfA1d* cDNA 片段总长为 2 036 bp，基因内部含有完整的开放阅读框架，大小为 1 449 bp，可编码长度为 482 个氨基酸残基的蛋白质，所推导的蛋白质氨基酸序列在 HSF DBD 结构功能域与拟南芥 AtHsfA1d 和水稻 OsHsfA1 蛋白的相似性分别为 86.3% 和 87.4%。组织特异性 Realtime-PCR 结果显示，*PsHsfA1d* 主要在杨树叶片和根部组织中高丰度表达。非生物胁迫、激素及糖诱导表达表明，*PsHsfA1d* 不仅受高温、干旱与盐诱导表达，还受到激素中 GA_3 与 ABA，及葡萄糖与蔗糖信号的上调表达。组合利用 MEGA4.0 和 DnaSP4.50.7 软件对小叶杨 36 株基因型个体的 *PsHsfA1d* 序列进行比对和分析，共检测到 207 个单核苷酸多态性(SNP)位点，SNP 频率为 1/16 bp，多样性指数 π 为 0.007 72。在编码区域，共检测到 69 个 SNP 位点，其中 37 个为同义突变，31 个为错义突变，1 个为无义突变；非同义突变与同义突变的比率 0.132 < 1，推测在小叶杨物种演化过程中，纯化选择是该基因内同义 SNP 位点主要的进化驱动力。

关键词： 小叶杨；*HsfA1d* 基因；热胁迫；诱导表达；单核苷酸多样性
中图分类号： S718.46；Q943.2 **文献标识码：** A **文章编号：** 1001-7488(2011)07-0082-09

Isolation, Expression and Single Nucleotide Polymorphisms Analysis of Heat Shock Transcription Factor *HsfA1d* in *Populus*

Zhao Xing Guo Qi Chen Qingqing Li Bailian Zhang Deqiang
(*National Engineering Laboratory for Tree Breeding Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education Beijing Forestry University Beijing* 100083)

Abstract: A cDNA clone encoding *HsfA1d* was isolated from a cDNA library prepared under heat shock stress in *Populus simonii* by the RT-PCR method. The cDNA of *PsHsfA1d* is 2 036 bp in length with an open reading frame (ORF, 1 449 bp in length) which is capable of encoding a protein of 482 amino acids. The deduced protein sequence of the *PsHsfA1d* shares 86.3%, and 87.4% identity with functional domain of HSF DBD of *Arabidopsis thaliana AtHsfA1d*, and *Oryza sativa OsHsfA1*, respectively. Tissue differential expression with realtime-PCR indicated that *PsHsfA1d* was expressed predominantly in the leaves and roots. We analyzed the expression patterns of *PsHsfA1d* under abiotic stress conditions and phytohormone treatment, and the result revealed that expression of *PsHsfA1d* was induced not only by heat-shock, drought and high-salt stress, but also up regulated by GA_3, and ABA, and glucose and sucrose signals. The genomic sequences of *PsHsfA1d* in 36 individuals were aligned, compared and analyzed using the software MEGA4.0 and DnaSP4.50.7. A total of 207 single nucleotide polymorphisms (SNPs) were detected and the frequency and diversity of SNPs were 1/16 bp and 0.007 72, respectively. In total, 69 SNPs were detected in the coding regions of *PsHsfA1d*, of which 37, 31 and 1 were silent, missense, and nonsense mutations, respectively; the nonsynonymous nucleotide substitutions (π_{nonsyn}) was markedly lower than π_{syn}, with the π_{nonsyn}/π_{syn} ratio 0.132 < 1, suggesting that diversity at the synonymous sites of exon regions would have resulted from strong purifying selection.

Key words: *Populus simonii*; *HsfA1d* gene; heat shock stress; inducible expression; single nucleotide polymorphisms

干旱、高盐及极端温度胁迫是影响森林生产力 和制约生态环境建设的主要非生物胁迫因素，也是

收稿日期：2011-02-14；修回日期：2011-05-06。
基金项目：国家林业公益性行业科研专项经费项目(201004009)；教育部新世纪优秀人才支持计划专项基金(NCET-07-0084)。
* 张德强为通讯作者。

实施困难立地造林的主要限制因子。随着全球气候迅速变暖,温室效应日益加剧,研究植物抵御高温热害的分子遗传调控机制就成为国际研究热点。在热胁迫下,生物体会产生一系列的应激反应,启动热激基因表达,导致热激蛋白(heat shock protein,HSP)在体内迅速积累,并以分子伴侣的形式帮助相关蛋白重新折叠、稳定、组装、胞内运输与降解,从而使植物在热逆境中得以存活与生长(Hartl *et al.*, 2002)。而热激蛋白的表达主要受一类热激转录因子(heat shock transcription factors, Hsfs)特异性结合其启动子区域的热激元件,在转录水平上调控热激蛋白基因的开启与关闭,最终促使热激蛋白高效表达(Nover *et al.*, 2001)。

Hsfs 是在真菌、植物和动物等真核细胞体内普遍存在的一类转录因子基因家族。第 1 个 *Hsf* 基因由 Wiederrecht 等(1988)在酵母(*Saccharomyces cerevisiae*)中克隆得到,随后又相继在果蝇(*Drosophila melanogaster*)和哺乳动物中分离出相应的 *Hsfs* 基因(Clos *et al.*, 1990; Rabindran *et al.*, 1991; Sarge *et al.*, 1991)。借助酵母的 *Hsf* 序列,首先在秘鲁番茄(*Lycopersicon peruvianum*)中分离出植物 *Hsfs* 基因成员(Scharf *et al.*, 1990)。与酵母、果蝇及动物相比,植物 *Hsfs* 是一数量庞大的多基因家族,全基因组检测表明,模式植物拟南芥(*Arabidopsis thaliana*)及水稻(*Oryza sativa*)基因组分别含有 25 及 21 个 *Hsf* 成员,且分属于 3 个亚类(Guo *et al.*, 2008; Nover *et al.*, 2001)。通过将 *Hsfs* 基因分别转入模式植物如拟南芥以及主要农作物如水稻、番茄和小麦(*Triticum aestivum*)等的大量研究充分证明了 *Hsfs* 的过量表达可有效激活并影响热激蛋白基因的表达水平,并能显著提高转化植株对高温及其他逆境的忍耐程度,为植物抗逆性状的分子育种奠定了重要的基础(Koskull-Döring *et al.*, 2007)。

Hsfs 基因成员虽已从真菌、植物和动物中分离出来并进行了功能性检测,但对于高度杂合、世代周期较长、树体高大,对经济建设和环境保护具有重要作用的用材及生态型树种来说还未见关于该基因分离和检测的报道。因此,在林木,特别是在抗逆树种中分离 *Hsfs* 基因家族成员以及深入研究它们对高温及其他非生物胁迫的分子遗传学调控机制,就显得尤为必要。为此,本研究以我国北方小叶杨(*Populus simonii*)为材料,从经热激胁迫后构建的 cDNA 文库中分离得到一 *PsHsfA1d* 基因的 cDNA 序列,并对其进行了组织特异性及调控表达分析。在此基础上,检测了该基因在小叶杨自然群体中的单核苷酸(single nucleotide polymorphisms,简称 SNPs)多样性及连锁不平衡(linkage disequilibrium, LD)结构。本研究为分离其他树种 *Hsfs* 成员以及阐明该基因的功能奠定了重要的基础。

1　材料与方法

1.1　植物材料

2007 年秋季,课题组由吉林、辽宁、内蒙古、河北、北京市、山东、河南、山西、陕西、青海、甘肃、四川等 12 个省(市、自治区)的小叶杨自然分布区域,收集保存了共 560 个基因型个体。本研究为分析小叶杨 *PsHsfA1d* 基因的单核苷酸多态性,从建立的基因库中按省份、种源选取了 36 株个体的新鲜叶片用于基因组 DNA 提取;而用于不同组织 RNA 提取的材料取自 1 年生小叶杨无性系。

1.2　杨树苗木胁迫处理及植物激素和糖信号诱导

在前期大量预试验的基础上,确定了较为合理的非生物胁迫及激素诱导试验系统,即,对于非生物胁迫处理,每一处理均将经过无性系化的苗龄 4 个月的小叶杨盆栽苗 3 株,分别经过高温(42 ℃,6 h)、干旱(1 天)与盐(250 $mmol \cdot L^{-1}$,5 h);对于激素诱导则分别将质量浓度为 10 $mg \cdot L^{-1}$, 0.175 $mg \cdot L^{-1}$与 50 $mg \cdot L^{-1}$的脱落酸(abscissic acid,ABA)、吲哚乙酸(indole-3-acetic acid,IAA)及赤霉素(gibberellin,GA_3)溶液喷施后 6 h 进行叶片采集。而对于糖信号诱导处理则分别用 20 $g \cdot L^{-1}$的蔗糖与 60 $g \cdot L^{-1}$的葡萄糖溶液喷施后 2 天进行叶片采集。而对照则为相同苗龄的小叶杨盆栽苗置于培养箱中进行正常的光照培养。同时将上述所有处理后及对照植株上采集的嫩叶组织置于液氮中冻存备用。

1.3　总 DNA 的提取

总 DNA 的提取按 DNeasy Plant Mini Kits (Qiagen, Inc., Valencia, CA, USA)描述的方法进行。

1.4　RNA 提取和 cDNA 合成

经胁迫处理、植物激素与糖信号诱导和其他组织材料总 RNA 的提取及 cDNA 合成依照 Zhang 等(2010a)描述的方法进行。

1.5　*PsHsfA1d* 基因检测

为了分离在热激条件下差异表达的基因,作者所在课题组构建了小叶杨热激诱导的 cDNA 文库,并对 ESTs 克隆进行了随机测序与比对,检测到了一编码区全长的 *PsHsfA1d* 基因的 cDNA 序列。

1.6 引物设计和 PCR 扩增

根据检测的 *PsHsfA1d* cDNA 序列,设计 1 对寡聚核苷酸引物。*PsHsfA1d*F: 5′-CCTGTTTCCCTCCGTCGCGTCTGCCTT-3′; *PsHsfA1d*R: 5′-GGACAATCGGATCTATGGACCTAAGGACAA-3′。

应用25 μL DNA 聚合酶链式反应(PCR)体系,以小叶杨热激诱导的 cDNA 文库为模板,加入 2.5 μL 10 × buffer, 1.8 μL 25 $mmol \cdot L^{-1}$ $MgCl_2$, 1 μL 10 $mmol \cdot L^{-1}$ dNTP,*Taq* DNA 聚合酶 1.0 U(以上试剂购自 Promega 公司), 10 $\mu mol \cdot L^{-1}$正向和反向引物各1 μL, 加适量双蒸水至25 μL。PCR 扩增程序如下:94 ℃ 3 min; 94 ℃ 30 s, 68 ℃ 30 s, 72 ℃ 60 s, 35 个循环; 72 ℃ 5 min, 4 ℃保存,最后将反应产物由 PCR 仪取出置于冰箱中冻存。

1.7 PCR 产物的克隆、测序及计算机分析

将 PCR 扩增得到的目的基因片段回收后连接于 pGEM-T Easy 上。连接产物转化大肠杆菌(*Escherichia coli*) DH5α,筛选阳性克隆,送公司测序。应用 DNAMAN6.0 软件和 CLUSTALX 软件包程序推导出氨基酸序列并进行 NCBI 检索分析, 然后分别估算和预测推导蛋白质的分子量和等电点。

1.8 Realtime PCR 检测 *PsHsfA1d* 基因的组织器官及调控表达模式

采用 ABI Primer Express 3.0 软件在 *PsHsfA1d* 基因内设计能够扩增长度约为 63 bp cDNA 片段的 Realtime PCR 引物。*PsHsfA1d*RTF: 5′-AAATGATGTGCAGCCGAATG-3′; *PsHsfA1d*RTR: 5′-TCTGTCAGCTGATCCACATGCT-3′。

以肌动蛋白基因 *Actin* 作为内参,引物序列为 *Act*F: 5′-CTC, CAT, CAT, GAA, ATG, CGA, TG-3′; *Act*R: 5′-TTG, GGG, CTA, GTG, CTG, AGA, TT-3′。定量 PCR 反应按照 Zhang 等(2010b)描述的方法进行,最后用该基因在待测样品中的扩增量与内参 *Actin* 基因扩增量的比值,即相对表达量分析其表达模式。

1.9 *PsHsfA1d* 基因的 SNP 多样性分析

依据已克隆的 *PsHsfA1d* 基因的核苷酸序列,设计能扩增出该基因的3 对特异的引物序列如下:
F1: 5′-CCTGTTTCCCTCCGTCGCGTCTGCCTTT-3′,
R1: 5′-TCTTGGAGTTTGTCAAGGAGAGCATAATGT-3′;
F2: 5′-CCTGGTATTGCTAGCTTACTGTCTACATT-3′,
R2: 5′-CTGGTATATGTGCGGATTGCCCTGAAGTT-3′;
F3: 5′-CGAGCCACACATCAGGAGTGACTCTTCAA-3′,
R3: 5′-GGACAATCGGATCTATGGACCTAAGGACAA-3′。

以选取的36 株基因型个体的总 DNA 分别为模板进行 PCR 扩增; 将 PCR 扩增产物进行琼脂糖凝胶电泳分离,回收、纯化目的片段后与 PGEM-T Easy 载体连接,转化后挑取阳性单克隆进行 3 次序列测定,然后将每一基因片段的核苷酸序列拼接成完整的基因序列。在此基础上,组合利用 MEGA4.0 和 DnaSP4.50.7 软件对每一基因的 36 个序列进行分析,标出 SNP 位点、计算 SNP 频率及多样性指数; 分析同义突变、错义突变和无义突变情况。

2 结果与讨论

2.1 小叶杨 *PsHsfA1d* 同源 cDNA 的克隆及序列分析

在前期获得小叶杨 *HsfA1d* ESTs 序列的基础上设计基因特异的引物,以热激诱导后构建的 cDNA 文库为模板进行 PCR 扩增,测序结果表明,克隆的 *PsHsfA1d* cDNA 片段总长度为 2 036 bp,分别包含长度为 319 bp 与 268 bp 的 5′未翻译区域(5′ untranslated region, 5′ UTR)及 3′ UTR(图 1); 在 *PsHsfA1d* 序列中,基因内部含有完整的开放阅读框架,大小为 1 449 bp,可编码长度为 482 个氨基酸残基的蛋白质(基因注册号为 JF905565)。估算推导的该蛋白质的分子量约为 53.4 ku,等电点为 4.76。鉴于植物 *Hsf* 是一结构复杂、数量庞大的基因家族,模式植物拟南芥及水稻全基因组分别含有 25 及 21 个 *Hsf* 成员,且分属于 3 个亚类(Guo *et al.*, 2008; Nover *et al.*, 2001)。为了确定克隆的小叶杨 *PsHsfA1d* 属于哪一亚类 *Hsf* 成员,对拟南芥及水稻中所有 *Hsf* 基因共 36 个成员与小叶杨 *PsHsfA1d* 蛋白进行了同源性进化分析(图 2)。由图 2 可见: 克隆的 *PsHsfA1d* 与拟南芥 *AtHsfA1d* 和水稻 *OsHsfA1* 成员属于同一亚类, 即 *HsfA1* 型,故命名为小叶杨 *PsHsfA1d*。

为了进行木本植物与草本植物 *HsfA1d* 基因组结构的比较基因组学研究,以小叶杨基因组总 DNA 为模板进行了 PCR 扩增和序列测定,并将其与拟南芥和水稻的直源同源基因(orthologous gene), 即 *AtHsfA1d* 和 *OsHsfA1* 基因的5′ UTR、外显子、内含子和3′ UTR 进行了同源性比较(图 3 与表 1)。由图 3 可见: *HsfA1d* 基因结构与序列长度在杨树、拟南芥和水稻中显示了不同的模式,如对于杨树与水稻,该基因的编码区均由 1 个内含子所分割,其长度分别为 1216 bp 和 1 386 bp,而在拟南芥中则包含长度分别为 1 420 bp 和 92 bp 的 2 个内含子。与 *PsHsfA1d* 相比,内含子在草本植物 *AtHsfA1d* 和 *OsHsfA1* 基因内的位置相同(位于 +288 核苷酸

处)，长度也相近。但比较 *HsfA1* 编码区的长度，则显示 *PsHsfA1d* 与 *AtHsfA1d* 相差较小(9 bp)，而比水稻该基因的编码区短了 72 bp(图 3)。在此基础上，将 *PsHsfA1d* 与 *AtHsfA1d* 和 *OsHsfA1* 基因的 5′ UTR、外显子、内含子和 3′ UTR 进行了同源性比较(表 1)。由表 1 可见：*PsHsfA1d* 与 *AtHsfA1d* 和 *OsHsfA1* 基因不同区域的同源性均不相同，其同源性由大到小的顺序为外显子 > 5′ UTR > 内含子 > 3′ UTR。这一结果显示，*HsfA1d* 基因在进化过程中，基因不同区域所受的选择压不同，与基因的其他区域相比，外显子区域则较为保守，同源性均高于 57.3%。

```
   1    CCTGTTTCCCTCCGTCGCGTCTGCCTTTTATAGAATCCCACTCTTCATTTCATTTCCCGTGACCGTACAACAACAACC
  79    ACCTGTCTCCTCTTTACCCCCTTTCTCCTTCCTCCCTATACAATCTCGATTCGTTCTCTCTCCTCTCCTTCAATCTAG
 157    CTTTTTTGCTCTCACAAATGCCTGCTCTTCCAAAACCCTAAAAAGTTCCAATCAATAAACATTTCCTTTATTAATCGC
 235    TACTCTGATGGATGGAACAATCAACTCTGCAAGCGGTGGCGACTCCTCGATCAGCGGCGGTGGGGCACAAGCGGCGCC
 313    GGCGCCGATGGCGTCACAGTCGAATGCGCCGCCGCCATTTCTAAGTAAGACTTATGACATGGTAGACGATCCGGCGAC
   1              M  A  S  Q  S  N  A  P  P  P  F  L  S  K  T  Y  D  M  V  D  D  P  A  T
 391    AGACGCGGTGGTGTCTTGGAGTTCGACGAATAATAGTTTTGTGGTTTGGAATCCGCCGGAGTTCGCTCGCGATCTTTT
  25     D  A  V  V  S  W  S  S  T  N  N  S  F  V  V  W  N  P  P  E  F  A  R  D  L  L
 469    GCCGAAGTATTTCAAGCATAATAATTTCTCCAGCTTTGTTAGGCAACTGAACACTTACGGTTTTCGAAAGGTGAATCC
  51     P  K  Y  F  K  H  N  N  F  S  S  F  V  R  Q  L  N  T  Y  G  F  R  K  V  N  P
 547    AGACCGTTGGGAATTTGCCAACGAAGGATTTTTGAGGGGTCAGAAGCACCTGCTTAGGATTATCAGTCGGCGAAAACC
  77     D  R  W  E  F  A  N  E  G  F  L  R  G  Q  K  H  L  L  R  I  I  S  R  R  K  P
 625    TGCCCACGGGCATGCTAATCAACAGCCACAACAACCACATGGACAGAACTCATCGGTTGCCGCCTGTGTTGAGGTTGG
 103     A  H  G  H  A  N  Q  Q  P  Q  Q  P  H  G  Q  N  S  S  V  A  A  C  V  E  V  G
 703    GAAATTTGGGCTCAAGGAAGAGGTGGAGAGGCTTAAGAGGGACAAGAATGTCCTTATGCAAGAACTTGTTAGGTTGAG
 129     K  F  G  L  K  E  E  V  E  R  L  K  R  D  K  N  V  L  M  Q  E  L  V  R  L  R
 781    GCAGCAGCAACAGTCCACAGATAGCCAGTTGCAAACCATGGTACAGCGACTTCAAGGAATGGAGCAGCGTCAGCAACA
 155     Q  Q  Q  Q  S  T  D  S  Q  L  Q  T  M  V  Q  R  L  Q  G  M  E  Q  R  Q  Q  Q
 859    GATGATGTCATTCCTGGCCAAGGCGATGAACAGCCCAGGCTTCTTGGCTCAGTTTGTACAGCAGCAAAATGACAACAA
 181     M  M  S  F  L  A  K  A  M  N  S  P  G  F  L  A  Q  F  V  Q  Q  Q  N  D  N  N
 937    TAGGCGCATAACGGAAGCCAATAAAAAACGAAGGCTCAAGCAGGAGGATGTTCCTGAGAATGAGGGTTCTGGTTCTCA
 207     R  R  I  T  E  A  N  K  K  R  R  L  K  Q  E  D  V  P  E  N  E  G  S  G  S  H
1015    TGATGGGCAGATTGTGAAGTATCAGCCTCTGATGAATGAGGCAGCCCAAGCAATGCTTAGGCAGATTATGAAAATGGA
 233     D  G  Q  I  V  K  Y  Q  P  L  M  N  E  A  A  Q  A  M  L  R  Q  I  M  K  M  D
1093    TGCTTCTTCTAAGTTGGAATCTTATGACAATAATCTTGGTGCAAAAGACAGTGGAAGCTCTTCGAGCCACACATCAGG
 259     A  S  S  K  L  E  S  Y  D  N  N  L  G  A  K  D  S  G  S  S  S  S  H  T  S  G
1171    AGTGACTCTTCAAGAGGTTCTGCCAACTTCAGGGCAATCCGCACATATACCAGCAGCTTTTGGGATTTCCAGGCAGGG
 285     V  T  L  Q  E  V  L  P  T  S  G  Q  S  A  H  I  P  A  A  F  G  I  S  R  Q  G
1249    CCCATCGGCTCCAATATCTGAAATACAATCTTCCCTGCAAATTGCTAGTTCTGAAAAGGTTACAGCTTCTCAGTTTCC
 311     P  S  A  P  I  S  E  I  Q  S  S  L  Q  I  A  S  S  E  K  V  T  A  S  Q  F  P
1327    GGATATAAGTATGCTGGTTGGAGCACAGGGGGCAACCTCCATACCTATTCCTCAGGCGGACGTGATCATGCCCCATGT
 337     D  I  S  M  L  V  G  A  Q  G  A  T  S  I  P  I  P  Q  A  D  V  I  M  P  H  V
1405    TTCTCAAATGCCTGAAATGGTACCAGAAAATGTTGCTGATATCACTTGTGAAGATTACATGGAACCAGAGACATGCAA
 363     S  Q  M  P  E  M  V  P  E  N  V  A  D  I  T  C  E  D  Y  M  E  P  E  T  C  N
1483    TGATGGATTCATAGATCCAGCTTCATTAGGAATTAATGGAACAATACCCATAGATATTGACAACATCTCCCCTGATCC
 389     D  G  F  I  D  P  A  S  L  G  I  N  G  T  I  P  I  D  I  D  N  I  S  P  D  P
1561    TGATATTGATGCTTTGCTTGATAATTCTAGCTTCTGGGATGACCTTCTTGTGCAGAGTCCAGATCCTGAGGATATTGA
 415     D  I  D  A  L  L  D  N  S  S  F  W  D  D  L  L  V  Q  S  P  D  P  E  D  I  E
1639    ATCAAGTTCTGTAGAGGCTAAAGCCAATGGAAATGATGTGCAGCCGAATGCAAACGGGTGGGACAAAGCTCAGCATGT
 441     S  S  S  V  E  A  K  A  N  G  N  D  V  Q  P  N  A  N  G  W  D  K  A  Q  H  V
1717    GGATCAGCTGACAGAACAAATGGAGCTTCTGACATCAGACAAGAAGCTTTAATCACACGAAATGTACCTTTGTCCCAC
 467     D  Q  L  T  E  Q  M  E  L  L  T  S  D  K  K  L终止密码子
1795    ACTGTATGAGGTAAACACCTACTAAACTTTTAAATCAGCGGGTCCTCCTCCTTTGGCCATTGGAACAAGCAATGCCTT
1873    TGGTATCTGTTCAGCGTAGCTTTATTTTTGCACATATGGAGACATAAAATTCAGTTTTCTTTTTCCCAAATTTTTATT
1951    CAGAACTTTTCTCTTCATCATCTTATGAGAGGTTCAAGATCTCAAGAAACATTATGTTGTCCTTAGGTCCATAGATCC
2029    GATTGTCC
```

图 1　小叶杨 *PsHsfA1d* 基因核苷酸及其推测的蛋白质序列

Fig. 1　Nucleotide and deduced amino acid sequences of *PsHsfA1d*

2.2　预测的蛋白质一级结构及其氨基酸序列保守性分析

为了分析 PsHsfA1d 蛋白的一级结构特征，将 *PsHsfA1d* 推导的氨基酸序列与研究较为广泛的番茄 LpHsfA1 及模式植物拟南芥 AtHsfA1d 及水稻 OsHsfA1 蛋白质序列进行了同源比较和结构分析(图 4)。由图 4 可见：与报道的其他草本植物 HsfA1 相似，PsHsfA1d 含有典型植物所具有的一切热激转录因子保守结构域，如在 PsHsfA1d 氨基酸序列 N 端第 10—104 氨基酸残基位置含有高度保守的 HSF DNA 结合结构域(DNA-binding domain，DBD)，与其他草本植物氨基酸序列的同源性均超过 86.3%。DBD 通过结合于热激蛋白基因 Hsps 的热激元件上，来募集其他转录因子结合于这些 Hsps 基因，形成转录复合体，从而促进 Hsps 基因的表达(Nover *et al.*，2001)；除了高度保守的 DBD 结构域外，在 N 端还存在另一保守的寡聚化结构域(位于氨基酸序列的第 124—189 位)，其主要功能被认为是在热逆境条件下，通过寡聚化作用使得 Hsf 单体形成三聚体，提高 DBD 的结合能力(Nover *et al.*，

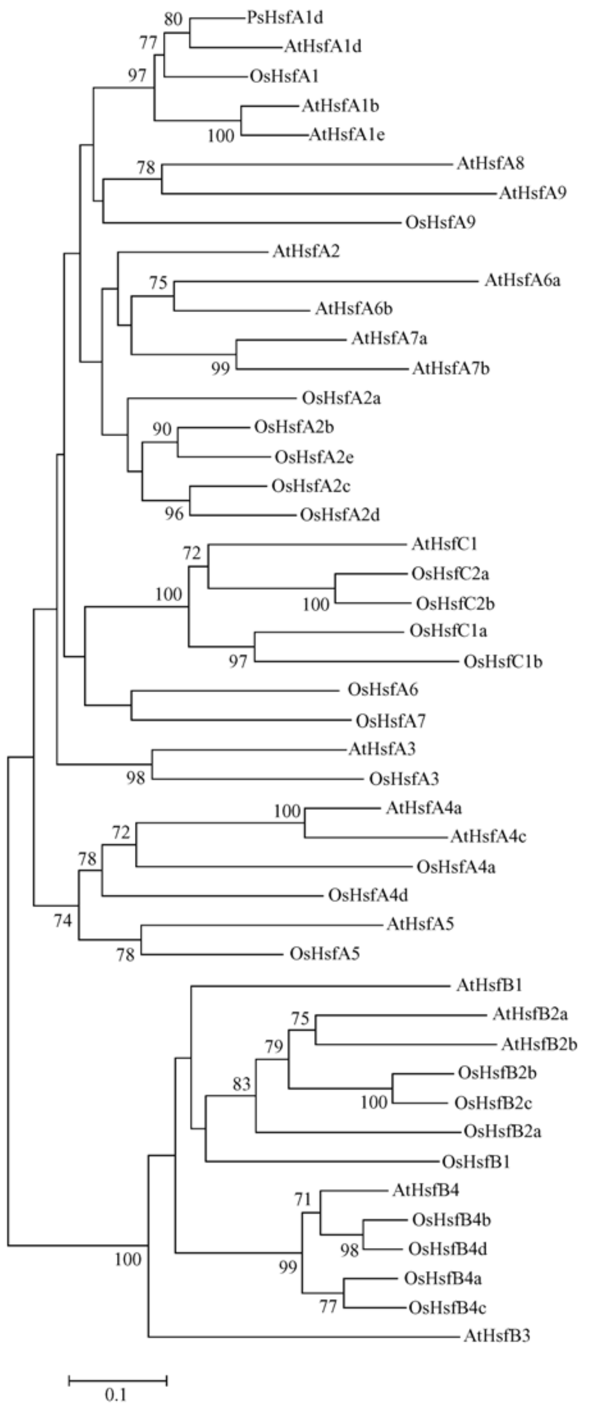

图 2　PsHsfA1d 与拟南芥及水稻 Hsf 蛋白的系统进化关系

Fig. 2　The phylogenetic relationship between PsHsfA1d and Hsfs from *Arabidopsis* and rice

At：拟南芥 *Arabidopsis thaliana*；Os：水稻 *Oryza sativa*. 参考 Refer to：Guo *et al*., 2008；Nover *et al*., 2001.

2001)；在氨基酸序列的中部(氨基酸序列的第 206—220 位)与 C 端(氨基酸序列的第 469—476 位)分别含有核输入信号与核输出信号，主要与 Hsf 蛋白输入核内及由核内运输到细胞质有关(Nover *et al*., 2001)。此外，C 端含有转录活动所必需的 AHA 激活结构域(氨基酸序列的第 458—476 位)(图 4)。

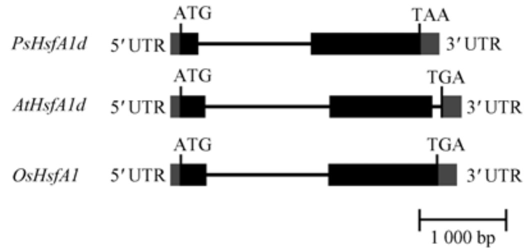

图 3　比较 *PsHsfA1d*，*AtHsfA1d* 与 *OsHsfA1* 基因的结构

Fig. 3　Comparison of schematic structures of *PsHsfA1d*, *AtHsfA1d* and *OsHsfA1*

表 1　*PsHsfA1d* 与 *AtHsfA1d* 和 *OsHsfA1* 基因序列同源性比较

Tab. 1　Comparison of sequence identity between *PsHsfA1d* and *AtHsfA1d* and *OsHsfA1*

PsHsfA1d	序列同源性 Sequence identity(%)							
	AtHsfA1d				*OsHsfA1*			
	5′ UTR	外显子 Exon	内含子 Intron	3′ UTR	5′ UTR	外显子 Exon	内含子 Intron	3′ UTR
5′ UTR	47.2				38.3			
外显子 Exon		67.4				57.3		
内含子 Intron			43.1				37.6	
3′ UTR				36.1				36.1

2.3　*PsHsfA1d* 基因的组织器官及调控表达分析

为了检测 *PsHsfA1d* 基因在杨树不同组织与器官中的表达模式，利用设计的 Realtime-PCR 引物，对该基因进行了杨树组织器官特异性表达分析(图 5)。由图 5 可见：*PsHsfA1d* 基因在杨树根、茎、叶片和顶端分生组织中均有表达，但其表达模式却不同：*PsHsfA1d* 在成熟叶片中表达丰度最高(3.509 ± 0.09)，在根部组织表达量较高，约为内参肌动蛋白表达量的 1.4 倍(1.384 ± 0.04)，而在嫩叶、树干皮层及成熟木质部组织表达量中等，在顶端分生组织及主干韧皮部、形成层与正在发育的木质部组织表达丰度最低。比较分析显示，*PsHsfA1d* 基因在杨树成熟组织如成熟木质部和成熟叶片的表达量约为在幼嫩组织的 5 倍以上。在杨树不同组织与器官中的表达可知，*PsHsfA1d* 基因具有组织器官及发育阶段特异性表达。在逆境状态下，叶片是感知极端温度、脱水及高盐等抗逆信号较为敏感的器官，而根部则是在非生物胁迫下直接遭受逆境伤害的组织，因此由该基因在杨树叶片与根部组织的高丰度表达可推测 *PsHsfA1d* 可能与逆境胁迫有关。

虽然克隆的 *PsHsfA1d* 基因在蛋白质一级结构上具备了植物 *HsfA1* 类型的一切结构特征，但不同物种在进化过程中形成了其特异的抗逆机制，特别

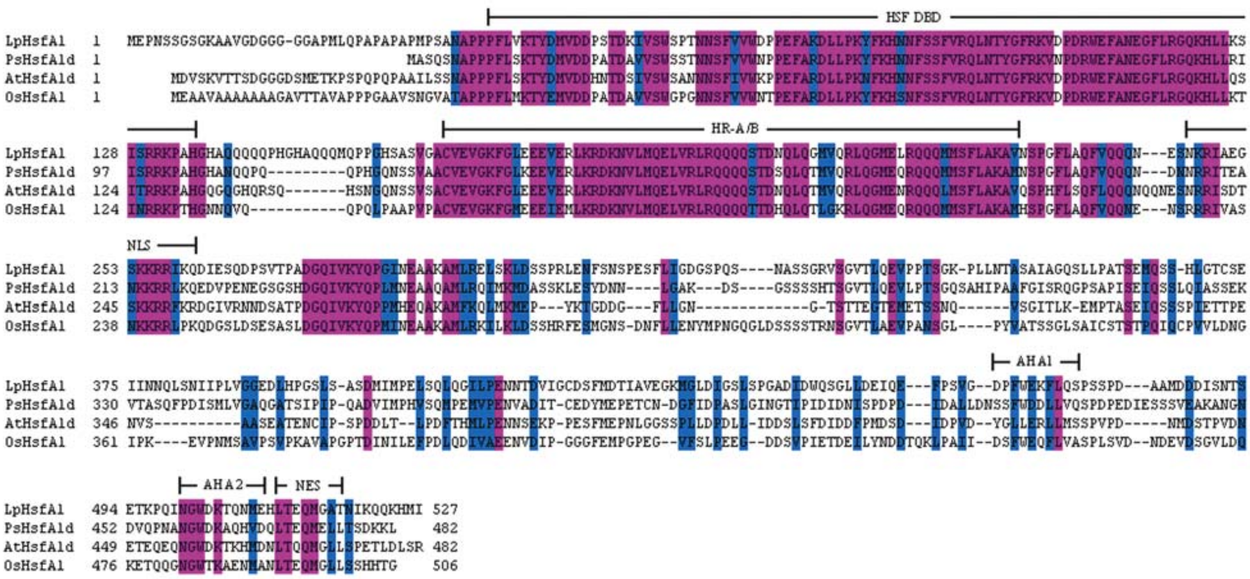

图 4　PsHsfA1d 与其他植物 HsfA1 蛋白质一级结构比较

Fig. 4　Comparison of the primary structure of PsHsfA1d between *P. simonii* and other plants

HSF DBD：热激转录因子 DNA 结合结构域 Heat shock factor DNA-binding domain；HR-A/B：寡聚化结构域 Oligomerization domain；NLS：核定位信号 Nuclear localization signal；AHA：以芳族、大的疏水和酸性氨基酸为特征的激活结构域 Activation domains rich in aromatic, hydrophobic and acidic amino acids；NES：核输出信号 Nuclear export signal.

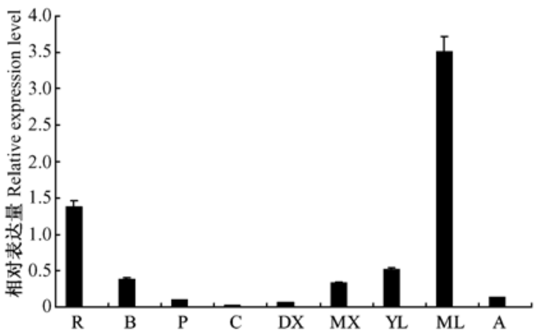

图 5　*PsHsfA1d* 基因的组织表达谱分析

Fig. 5　Expression profile of *PsHsfA1d* in root, stem, leaf and apex tissues

R：根 Root；B：树皮 Bark of stem；P：树干韧皮部 Phloem of stem；C：树干形成层 Cambium of stem；DX：树干正在发育的木质部 Developing xylem of stem；MX：树干成熟木质部 Mature xylem of stem；YL：嫩叶 Young leaf；ML：成熟叶 Mature leaf；A：顶端分生组织 Apex.

是多年生树木与 1 年生草本植物可能具有不同的基因调控机制。因此，很有必要在转录水平上检测 *PsHsfA1d* 基因对各种非生物胁迫是否具有调控表达作用。为此，利用 Realtime-PCR 技术检测了小叶杨 *PsHsfA1d* 在不同逆境条件、激素处理及糖信号诱导下的 RNA 转录水平，并用 *PsHsfA1d* 基因的扩增量与内参 *Actin* 基因扩增量的比值，即相对表达量来表示(图 6)。由图 6 可见：在逆境如高温、干旱及高盐条件下，*PsHsfA1d* 基因在转录水平均有不同程度的上调表达，表达量分别是对照的 1.8 ~ 3.9 倍，其中该基因对干旱胁迫的响应最为明显。因此，Realtime-PCR 检测结果显示，所分离的 *PsHsfA1d* 是受高温、干旱及高盐（NaCl）胁迫上调表达的基因。

植物激素，如 IAA，GA_3 与 ABA 等是植物生长发育及抵御逆境胁迫的重要调节剂，可能参与调控抗逆基因的表达。而本文克隆的 *PsHsfA1d* 基因是否受外源激素诱导表达？为了回答这一问题，作者

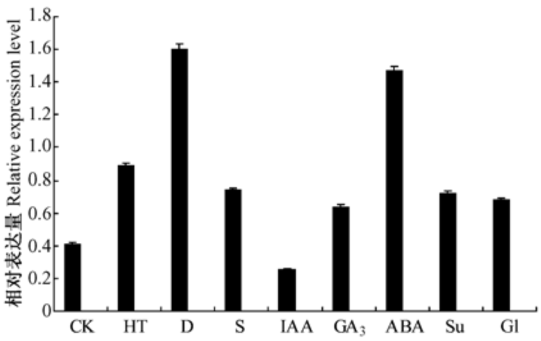

图 6　*PsHsfA1d* 的非生物胁迫、激素及糖诱导差异表达分析

Fig. 6　Regulation of *PsHsfA1d* expression induced by abiotic stress, hormone, and sugars feeding

CK：对照 Control；HT：高温 High temperature；D：干旱 Drought；S：盐 Salt；IAA：吲哚乙酸 Indole-3-acetic acid；GA_3：赤霉素 Gibberellin 3；ABA：脱落酸 Abscissic acid；Su：蔗糖 Sucrose；Gl：葡萄糖 Glucose.

同样采用 Realtime-PCR 技术检测了该基因在这些外源激素诱导下的 mRNA 转录水平(图 6)。由图 6 可见:外源激素如赤霉素 GA_3 及脱落酸 ABA 的施加均显著诱导了 *PsHsfA1d* 基因的上调表达,而在 IAA 的诱导下为下调表达。糖类如葡萄糖、蔗糖不仅是重要的渗透调节剂,也是基因表达的信号,因此,很有必要研究 *PsHsfA1d* 对葡萄糖与蔗糖信号的诱导是否具有响应。研究结果初步表明,蔗糖与葡萄糖信号均能诱导 *PsHsfA1d* 基因的上调表达,表达量至少是对照的 1.7 倍(图 6)。

2.4 *PsHsfA1d* 基因内 SNP 的多样性检测

基因的遗传多样性是伴随物种长期进化过程中形成的种内核苷酸位点的微小变异,而研究基因内 SNP 的多样性是对该基因进行群体遗传学和“基于候选基因内 SNP 连锁不平衡作图”的基础。为此,在分析 *PsHsfA1d* cDNA 序列的基础上,以取自小叶杨自然分布区中能够最大范围覆盖其生长区域的 36 株基因型个体为材料,对包含 315 bp 的 5′ UTR、1 449 bp 的外显子区、1 257 bp 的内含子区和 266 bp 的 3′ UTR 等共 3 287 bp 的核苷酸序列进行了单核苷酸多态性分析(表 2)。由表 2 可见:在小叶杨物种演化过程中 *PsHsfA1d* 基因内分别发生了 4 次插入、8 次缺失及 207 个单核苷酸突变,插入和缺失长度范围为 1 ~ 6 bp。其中,插入和缺失均发生在基因的非编码区。因此,对于 *PsHsfA1d* 基因,在长度为 3 287 bp 的区域内 SNP 发生的频率为 1/16 bp,SNP 多样性,即,任意两序列之间核苷酸差异的平均数 $\pi_T = 0.00772$。由表 2 显示的结果可见:该基因的不同区域在进化过程中形成的 SNP 多样性不同,其在 5′ UTR、外显子区、内含子区和 3′ UTR 的 SNP 多样性分别为 0.006 32, 0.005 99, 0.010 44 和 0.008 26。由此可推测,在小叶杨自然繁衍过程中,该基因不同区域的保守性不同,而在编码区域,其核苷酸变异程度最低,显示了 *PsHsfA1d* 基因在进化过程中编码区域受到了较强的选择压(Suha *et al.*, 2005)。

表 2 *PsHsfA1d* 基因内 SNP 多样性

Tab. 2 SNPs diversity in *PsHsfA1d*

PsHsfA1d	总数 Total	5′未翻译区域 5′ UTR	外显子 Exon	内含子 Intron	3′未翻译区域 3′ UTR
序列长度 Length/bp	3 287	315	1 449	1 257	266
SNP 数量(个) Number	207	19	69	101	18
SNP 频率 Frequency/bp^{-1}	16	17	21	12	15
SNP 多样性 Diversity(π_T)	0.007 72	0.006 32	0.005 99	0.010 44	0.008 26

为了检测 *PsHsfA1d* 基因编码区内 SNP 位点变异的发生是否影响了相应密码子编码的氨基酸序列,在分析 SNP 多态性位点的基础上进行了氨基酸的多样性分析。对 *PsHsfA1d* 基因编码区内的 69 个 SNPs 进行了同义突变、错义突变和无义突变分析(表 3)。由表 3 可见:在 *PsHsfA1d* 编码区内的 69 个 SNPs 中,有 37 个属于同义突变、31 个为错义突变,而仅有 1 个属于无义突变。鉴于非同义突变中常见 SNP(在群体中出现频率超过 10.0%)对于候选基因内 SNP 连锁不平衡作图的重要性,需要对这些突变中的常见 SNP 进行详细分析。在 32 个非同义突变中,有 8 个属于常见 SNPs。其中,有 4, 3, 1 个分别位于密码子的第 1, 2, 3 个核苷酸位置上。如,在编码区的第 892 及第 1096 位置上均发生了第 1 位密码子的突变,分别由原来的密码子 GCA 与 CCT 突变成新的密码子 ACA 与 TCT,从而分别导致其编码的氨基酸由原来的丙氨酸(Ala)与脯氨酸(Pro)突变成苏氨酸(Thr)与丝氨酸(Ser)。在编码区中,同义突变与非同义突变的核苷酸多样性分别为 $\pi_{syn} = 0.01416$, $\pi_{nonsyn} = 0.00374$,非同义突变与同义突变的比率为 0.132。因此,对于 *PsHsfA1d* 基因,由其编码区内部发生的非同义突变与同义突变的比率 0.132 < 1,显示了在小叶杨物种演化过程中,是纯化选择,而不是自然选择对 *PsHsfA1d* 基因内同义 SNP 位点起了主要的进化驱动力。

表 3 *PsHsfA1d* 编码区内 SNP 的突变类型

Tab. 3 Mutation types of SNPs located in coding regions in *PsHsfA1d*

基因 Gene	同义突变 Synonymous mutation	非同义突变 Non-synonymous mutation		非同义突变/同义突变 Non-synonymous mutation/synonymous mutation
		错义突变 Missense mutation	无义突变 Nonsense mutation	
SNP 数量(个) Number	37	31	1	0.865
SNP 多样性 Diversity (π_T)	0.014 16	0.003 62	0.000 117	0.132

为了分析小叶杨 *PsHsfA1d* 基因内核苷酸突变的替代类型，对检测到的 207 个 SNPs 进行了变异类型统计（表 4）。由表 4 可见：在这些 SNPs 中，有 153 个属于转换类型，分别包含 73 个 G⇔A 和 80 个 C⇔T。对于颠换替代类型，共检测到 54 个，分别包括 9 个 C⇔A、11 个 G⇔T、26 个 A⇔T 和 8 个 C⇔G。因此，转换/颠换为 2.83 > 2.0，造成这一结果可能是，SNP 在 CG 序列上出现最为频繁，而且多是 C→T，原因是 CG 中 C 即胞嘧啶常为甲基化地、自发地脱氨后即成为胸腺嘧啶（Suha *et al*., 2005）。

表 4　毛白杨 *PsHsfA1d* 基因内部 SNP 的变异类型

Tab. 4　The substitution type of SNPs for *PsHsfA1d*

基因 Gene	数量 Quantity	替代类型 Substitution type					
		转换 Transitions		颠换 Transversions			
		A⇔G	T⇔C	A⇔C	T⇔G	A⇔T	G⇔C
PsHsfA1d	207	73	80	9	11	26	8

2.5　*PsHsfA1d* 基因内 SNPs 的单体型和 LD 分析

在人类遗传学中，对控制疾病性状候选基因内 SNP 的 LD 作图主要利用的是群体中罕见 SNP 位点（Suha *et al*., 2005），与此不同，在树木中则应该选用常见 SNP 位点开展联合遗传学研究。为了提高基于候选基因内 SNP 的连锁不平衡作图的效率就需要获知候选基因内 SNP 的单体型。为此，作者选取了 *PsHsfA1d* 基因内检测到的 48 个常见 SNPs，并利用 DnaSP4.50.7 软件中的 Hap 程序对这些位点进行了单体型分析（图 7）。由图 7 可见：*PsHsfA1d* 内的 48 个 SNPs 在小叶杨自然群体中组成了 21 种单体型，单体型的多样性为 0.947 6，其中 Hap1，Hap2 和 Hap3 分别由 5 个特异的 SNPs 组成，而单体型 Hap5，Hap7 和 Hap15 则分别由 2 个特异的 SNPs 组成，其余单体型均有单个特异的 SNP 组成（图 7）。该结果建议应选取前 3 种单体型进一步在群体中对每一单株开展 SNP 分型，为小叶杨 *PsHsfA1d* 基因的连锁不平衡作图提供理想的 SNP 基因型数据。

Hap_1	GTTGGCTTCTTAGCCACTTGTACCCGTACAGATGAGACACCTCGACCA
Hap_2	GTTGGCTTCTTAGCCACTTGTACCCGCCCGGATGAGACACCTCGACCA
Hap_3	GTTGGCTTCTTAACCACTTGTACCCGCCCGGATGAGACACCTCGACCA
Hap_4	TTTAGCTTCTTAGCCACTTGTACCCGTCCAGATGAGACACCTCGACCA
Hap_5	TTTAGCTTCTTAGCCACTTGTACCCGCCTGGATGAGACACCTCGACCA
Hap_6	GCCATAGGTACGGGGTGGCTATTCCGCCCGATAAGAGTCTTGGAGTTT
Hap_7	GTTAGCTTCTTAGCCACTTGTACCCGCCTGGTTGAGACACCTCGACCA
Hap_8	GCCATAGGTACGGGGTGGCTATTCCGCCTGATAAGAGTCTTGGAGTTT
Hap_9	GTTGGCTTCTTAGCCACTTGTACTTACCCGGATGAGACACCTCGACCA
Hap_10	GTCATAGGTACGGGGTGGCTATTCCGCCCGGATGAGACATTGGAGTTT
Hap_11	GTCATAGGTACGGCCACTTGTACCCGTCCAGATGAGACACCTCGGTTT
Hap_12	GCCATAGGTACGGCCACTTGTACCCGCCCGGATGAGACACCTCGACCA
Hap_13	GCTATAGGTACGGGGTGGCTATTTTACCTGGATGAGACACCTCGACCA
Hap_14	GCTGGCTTCTTAGCCTGGCTATTCCGTCCAGATGAGACACCTCGACCA
Hap_15	GTTGGCTTCTTAGCCACTTGTACCCGTCCAGATGAGACACCTCGACCA
Hap_16	GTTGGCTTCTTAGCCACTTGTACCCGCCTGATAAGAGTCTTGGAGTTT
Hap_17	GTTGGCTTCTTAGCCACTTGTACTTACCTGGATAGAGTCTTGGAATTT
Hap_18	GTTGGCTTCTTAGCCACTTGTACCCGCCCGGATGAGACACTGCGACCA
Hap_19	GCTATAGGTACGGGGTGGCTATTTTACCTGATAAGAGTCTTGGAATTT
Hap_20	GTTGGCTTCTTAGCCACTTGTACCCGCCCAGATGAGACACCTCGACCA
Hap_21	TTTAGCTTCTTAGCCACTTGTATCCGCCTAGATAAGACACCTCGACCA

图 7　*PsHsfA1d* 基因内常见 SNP 的单体型分布模式

Fig. 7　Patterns of common SNP based haplotype in *PsHsfA1d*

为了理解 *PsHsfA1d* 基因内 SNP 位点间在小叶杨自然群体内 36 株基因型个体间 LD 的延伸长度及程度，利用 DnaSP4.50.7 软件中的 LD 程序分析了该基因内 SNPs 的 LD 水平（图 8）。由图 8 可见：对于小叶杨 *PsHsfA1d* 基因，在 36 株基因型个体中，随着 *PsHsfA1d* 基因核苷酸序列长度的增加，SNPs 的连锁不平衡程度逐渐消弱，当长度达 1 000 bp 时，即 $R^2 < 0.2$，连锁不平衡迅速消失。这一结果表明：在小叶杨中，SNP 的连锁不平衡在候选基因内部就已衰退。这与在人类、拟南芥和大麦（*Hordeum vulgare*）中 LD 的延伸长度相差很大，如 LD 在人类中可从 5 kb 延伸到 500 kb 这样一个大的范围，这使得在人类中进行全基因组范围的 LD 作图是可行的（Reich *et al*., 2001）。然而，在植物中 LD 的延伸范围变异很大，一般来说，在自交物种如在拟南芥中可延伸至 250 kb（Nordborg *et al*., 2002）。相反，在异交物种如玉米（*Zea mays*）中 LD 延伸仅在 1 kb 之内就已消失（Remington *et al*., 2001）。一般认为，交配系统、强的方向性选择和群体的进化史可严重影响 LD 在目标物种中的延伸长度（Nachman, 2002）。因此，从上述研究结果可知，对于异交物种小叶杨来说，基于候选基因内 SNP 的连锁不平衡作图是可行的。

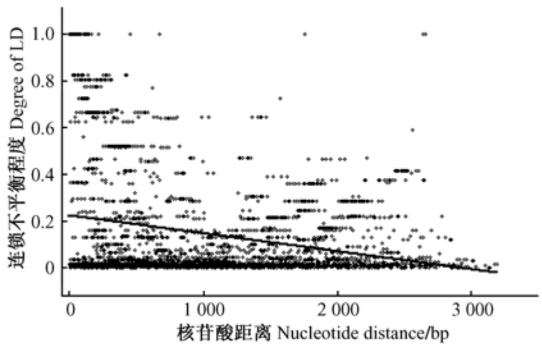

图 8　*PsHsfA1d* 基因内 SNPs 的连锁不平衡

Fig. 8　Linkage disequilibrium (LD) of SNPs in *PsHsfA1d*

3 结论

1）以小叶杨为模式，首次由抗逆林木中分离获得一 *HsfA1d* 型基因，推导的蛋白质一级结构在双子叶植物与单子叶植物中具有较高的保守性，如含有 Hsf DNA 结合结构域、寡聚化结构域、核输入信号与核输出信号，及 AHA 激活结构域等典型植物 *HsfA1* 所具有的一切保守结构域。

2）*PsHsfA1d* 基因的组织器官及调控表达分析表明，该基因在杨树成熟叶片与根部组织中表达丰度最高，且具有发育阶段特异性表达；在逆境包括高温、干旱及高盐条件下，*PsHsfA1d* 基因在转录水平上均有不同程度的上调表达，表达量分别是对照的 1.8～3.9 倍，其中对干旱胁迫的响应最为明显。植物激素诱导表达显示，外源激素如 ABA 及 GA_3 的施加显著提高了 *PsHsfA1d* 基因的表达；蔗糖与葡萄糖信号均能诱导 *PsHsfA1d* 基因的上调表达，表达量是对照的 1.7 倍。

3）在小叶杨自然群体内，*PsHsfA1d* 基因内存在丰富的 SNP 变异，多样性水平 π 高达 0.007 72，SNP 频率为 1/16 bp；对于编码区域，在小叶杨自然群体物种演化过程中，纯化选择是该基因内同义 SNP 位点主要进化驱动力；对该基因内 SNPs 位点进行的连锁不平衡分析表明，随着核苷酸序列长度的增加，SNPs 的连锁不平衡在基因内部就迅速衰退，因此，在小叶杨中，基于候选基因的连锁不平衡作图是可行的。

参 考 文 献

Clos J, Westwood J T, Becker P B, *et al*. 1990. Molecular cloning and expression of a hexameric Drosophila heat shock factor subject to negative regulation. Cell, 63(5): 1085－1097.

Guo Jingkang, Wu Jian, Ji Qian, *et al*. 2008. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. J Genet Genomics, 35(2): 105－118.

Hartl F U, Hayer-Hartl M. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science, 295(5561): 1852－1858.

Koskull-Döring P, Scharf K D, Nover L. 2007. The diversity of plant heat stress transcription factors. Trends in Plant Science, 12(10): 452－457.

Nachman M W. 2002. Variation in recombination rate across the genome: evidence and implications. Curr Opin Genet Dev, 12(6): 657－663.

Nordborg M, Borevitz J O, Bergelson J, *et al*. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet, 30(2): 190－193.

Nover L, Bharti K, Döring P, *et al*. 2001. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones, 6(3): 177－189.

Rabindran S K, Giorgi G, Clos J, *et al*. 1991. Molecular cloning and expression of a human heat stress factor. PNAS, 88(16): 6906－6910.

Reich D E, Cargill M, Bolk S, *et al*. 2001. Linkage disequilibrium in the human genome. Nature, 411(6834): 199－204.

Remington D L, Thornsberry J M, Matsuoka Y, *et al*. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS, 98(20): 11479－11484.

Sarge K D, Zimarino V, Holm K, *et al*. 1991. Cloning and characterization of two mouse heat stress factors with distinct inducible and constitutive DNA binding ability. Genes Dev, 5(10): 1902－1911.

Scharf K D, Rose S, Zott W, *et al*. 1990. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. EMBO J, 9(13): 4495－4501.

Suha Y, Vijg J. 2005. SNP discovery in associating genetic variation with human disease phenotypes. Mutation Research, 573(1/2): 41－53.

Wiederrecht G, Seto D, Parker C S. 1988. Isolation of the gene encoding the *S. cerevisiae* heat shock transcription factor. Cell, 54(6): 841－853.

Zhang Deqiang, Du Qingzhang, Xu Baohua, *et al*. 2010a. The actin multigene family in *Populus*: organization, expression and phylogenetic analysis. Mol Genet Genomics, 284(2): 105－119.

Zhang Deqiang, Yang Xiaohui, Zhang Zhiyi, *et al*. 2010b. Expression and nucleotide diversity of the poplar *COBL* gene. Tree Genetics and Genomes, 6(2): 331－344.

（责任编辑 徐 红）