

doi:10.3969/j.issn.1672-5565.2015.01.01

2个银叶真藓 HSP70 序列的生物信息学相关分析

王明强^{1,2}, 高 贝¹, 张道远^{1*}

(1.中国科学院干旱区生物地理与生物资源重点实验室,中国科学院新疆生态与地理研究所,乌鲁木齐 830011;

2.中国科学院大学,北京 100049)

摘要:从银叶真藓(*Bryum argenteum*)转录组数据库出发,使用 Pfam 数据库提供的 HMM 模型共得到 33 条长度大于 200aa,注释的热休克蛋白 BaHSP70;其中 2 条(BaHSP70-1, BaHSP70-2)具有完整 ORF,NCBI 核酸数据库登录号为 KP087877 和 KP087878。使用生物信息学在线分析工具和软件,对真藓 HSP70 的两条蛋白质序列从氨基酸组成、保守结构域、理化性质、疏水性/亲水性、信号肽、蛋白质结构、模体的识别及同源性分析等方面进行了预测和分析。结果表明:2 条 BaHSP70s 基因序列 ORF 全长分别为 2 396 bp 和 2 356 bp,分别编码 649aa 和 650aa。序列模体分析表明 BaHSP70s 和其它报道的植物 HSP70 均含有 4 个相同的模体,并且各模体在蛋白质序列上顺序一致。通过对 2 条 BaHSP70s 进行氨基酸多序列比对及基因树分析,发现 BaHSP70-1 和 BaHSP70-2 雪莲相似度最高,分别是 91.2%和 86.6%。本研究为进一步研究 HSP70 基因的克隆和功能验证奠定了基础。

关键词:真藓;热激蛋白 70;RNA-Seq;生物信息学

中图分类号:Q811.4 **文献标志码:**A **文章编号:**1672-5565(2015)-01-001-08

Bioinformatics relevant analysis of two heat shock protein 70 sequence from *Bryum argenteum*

WANG Mingqiang^{1,2}, GAO Bei¹, ZHANG Daoyuan^{1*}

(1.Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and

Geography, Chinese Academy of Sciences, Urumqi 830011, China ;

2.University of Chinese Academy of Sciences, Beijing 10049, China)

Abstract:Our study utilizes the RNA-seq technology which is based on the Hi-seq 2000 sequencing platform from BGI to acquire the RNA sequences data from *Bryum argenteum* dealt with dehydration and rehydration stress. Using the Hidden Markov Model provided by the Pfam database, we have identified 33 HSP70 sequences with the length longer than 200aa. Two of them (BaHSP70-1, BaHSP70-2) contain Open Reading Frame, the Accession No. in NCBI nucleotide database are KP087877 and KP087878, respectively. Applying web-based bioinformatics tools and softwares, we predicted and analyzed *B. argenteum* Heat Shock Protein 70 sequences from the following aspects: amino acid component, conserved domain, physicochemical property, hydrophobicity and hydrophilicity, signal peptide, protein structure, motif recognition, homologous analysis. The result showed the length of two possessed complete ORF HSP70s are 649aa and 650aa respectively. The length of corresponding transcripts in the transcriptome database are 2 396 bp and 2 356 bp. Sequence motif analysis indicates *B. argenteum* and other plant species have 4 identical motifs. Meanwhile, the order of the motifs arranged in protein sequence is the same in all selected species. Multiple sequence alignment and homologous analysis demonstrated that HSP70-1 and HSP70-2 are most similar to *Saussurea* up to the similarity of 91.2% and 86.6% respectively. This study laid foundation for the *B. argenteum* further research in gene clone and functional confirmation.

Keywords:*Bryum argenteum*; Heat shock protein 70; RNA-Seq; Bioinformatics

收稿日期:2014-11-03;修回日期:2015-01-26.

基金项目:NSFC-新疆青年人才联合项目(U1403302)。

作者简介:王明强,男,硕士研究生,研究方向:基因组学技术和生物信息学;E-mail:ant_genome@yahoo.com.

*通信作者:张道远,女,博士,研究员,研究方向:抗逆分子生物学与荒漠植物生物多样性保育;E-mail:zhangdy@ms.xjb.ac.cn.

热激蛋白 70 (Heat Shock Protein 70, HSP70) 是生物机体应对高温及其他胁迫环境时所产生的的一类特定的应激蛋白,在细胞内的大量表达可以明显改善细胞的生存能力,提高生物机体对环境胁迫或伤害的耐受性^[1]。HSP70 在原核生物和真核生物体内都存在,并存在于所有的细胞区室和器官中,序列非常保守;在细胞内参与多肽的折叠^[2]、寡聚蛋白的组装和跨膜运输、蛋白质活性的调节,影响植物的生长发育,是十分重要且基本的细胞功能蛋白^[3-4]。

银叶真藓 (*Bryum argenteum*) 隶属于真藓科真藓属,是藓类植物的典型代表种^[5]。它的形态结构相对简单,但是可以生存在极端干旱、高温、高盐的环境中,是中国荒漠区生物土壤结皮层的优势成分和群落演替过程中的主要先锋种^[6],也是生态系统稳定和退化生态系统恢复的重要指示植物之一^[7]。近年来,关于银叶真藓的形态发育^[8],系统分类^[9],生理生化的研究已有报道^[10],但有关其 HSP70 蛋白生物信息学分析未见报道。

前期研究过程中,项目组构建了银叶真藓干燥-复苏过程的转录组数据库。本研究从转录组数据库出发,使用 Pfam 数据库中提供 HSP70 的 HMM 模型,从已注释的真藓 BaHSP70s 中得到了 2 条全长的 BaHSP70 序列,分别命名为 BaHSP70-1 (GenBank No. KP087877) 和 BaHSP70-2 (GenBank No. KP087878)。利用生物信息学相关软件分析了这 2 条基因的序列特征、编码蛋白质的理化性质、蛋白质结构、相关的生物学功能预测等,并与其他报道的 HSP70 进行了多序列比对与基因树分析^[11]。研究有助于进一步理解 BaHSP70s 蛋白抗逆调控机制,并为后期 BaHSP70s 基因实验克隆和功能验证奠定基础。

1 材料与方法

1.1 数据收集与预处理

本研究实验材料银叶真藓取自于新疆古尔班通古特沙漠,通过无菌操作,在人工控制条件下进行组织培养获得再生完整植株,提取银叶真藓在干旱-复水过程中的 RNA 构建 RNA 测序文库^[12];利用华大基因提供的 Illumina 二代测序 RNA-seq 技术^[13],使用 Trinity 软件对序列进行组装^[14],同时利用相关蛋白质数据库对组装的 Contig 进行注释,从而得到银叶真藓转录组 Unigene 数据库和蛋白质数据库。绿藻植物、小立碗藓、拟南芥、水稻、玉米等植物的 HSP70 基因从 NCBI 的 PROTEIN 数据库和 UniProtKB 数据库中下载得到。从 Pfam 数据库中

下载 HSP70 的 HMM 概率模型^[15],文件命名为 hsp70.hmm。

1.2 真藓 HSP70 基因的鉴定

使用 HMMER-3.1 软件搜索 HSP70 的同源序列^[16],其中 e-value 值设置为 5,输出格式为 pfamtblout,输入命令: hmmsearch-E5-pfamtblout HSP1 hsp70.hmm Protein > result,获得具有 HSP70 结构域的序列;根据输出的比对信息,提取比对结果具有全长的片段,再次输入 Pfam 提供的在线工具进行验证,如图 1 所示,具有完整的结构域。同时使用 NCBI 提供的在线分析工具 ORF Finder 确认序列的起始密码子、终止密码子和长度;Blast 两两比对结果证明序列间的一致性达 83%,因此分别命名为 BaHSP70-1 和 BaHSP70-2。

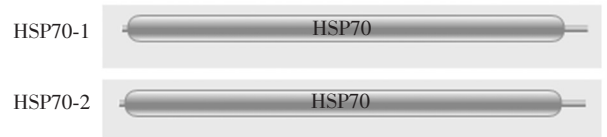


图 1 基于 Pfam 数据库确认 HSP70 蛋白质结构域
Fig.1 Conformation of HSP70 domain based on Pfam database

1.3 HSP70 的生物信息学分析

使用 ProtParam 在线分析工具检索蛋白质序列的理化参数,包括分子质量,氨基酸组成,消光系数等;使用 CDD 分析保守结构域;使用 ProtScale 程序绘制亲疏水性序列谱,反映蛋白质在水环境下的折叠情况;使用 NetNES 分析信号肽位点;使用 NPSA 在线分析工具提供的 MRLC 方法预测真藓 HSP70 蛋白的二级结构;使用瑞士模型工作区 (Swiss-Model Workspace) 网络综合服务器对 HSP70 蛋白质结构做同源建模;使用 ClustalW 对获取的 HSP70 蛋白序列进行多序列比对;除去比对后序列两侧的冗余序列,提取序列间相似性较高的序列区域,使用 MEGA5.2 构建 NJ 进化树并进行 Bootstrap 检测^[17],使用工具和地址如表 1 所示。

2 结果与分析

2.1 银叶真藓 HSP70 基因序列分析

基于 Pfam 数据库提供的 HMM 模型方法,从银叶真藓转录组注释蛋白质库中鉴定出 2 条具有全长的 HSP70 序列,分别命名为 BaHSP70-1 和 BaHSP70-2,它们在 Unigene 数据库中对应的的转录本核酸序列长度分别为 2 396 bp 和 2 356 bp。用 NCBI 提供的在线工具 ORF Finder 预测其编码蛋白质序列的长度为 649 和 650 个氨基酸,开放式阅读框具有起始密码子和终止密

码子,多序列比对分析表明,BaHSP70 最保守的序列区段位于序列的第 200 位至 250 位,氨基酸序列为 LDVTPLSLGLLETLGGVMTKLIPRNTTIPTKKSQVFSTYAD NQTGVLIQVY,该区段位于核酸结合区(Nucleotide Binding Domain, NBD),具有结合并水解 ATP 的活性中心,它与 ATP 的结合-水解过程调控着它对多肽的亲合力。

2.2 银叶真蕨 HSP70 保守域分析

本研究使用 NCBI 提供的 CDD 数据库对 BaHSP70s 序列进行保守域分析^[18],结果表明,银叶真蕨 HSP70 序列含有 HSP70 结构域,其中核酸结合区结构域较为保守,属于 NBD sugar-kinase HSP70 action superfamily,结果如图 2 所示;研究进一步表明基于 HMM 模型鉴定的完整基因属于热激蛋白 HSP70 家族。

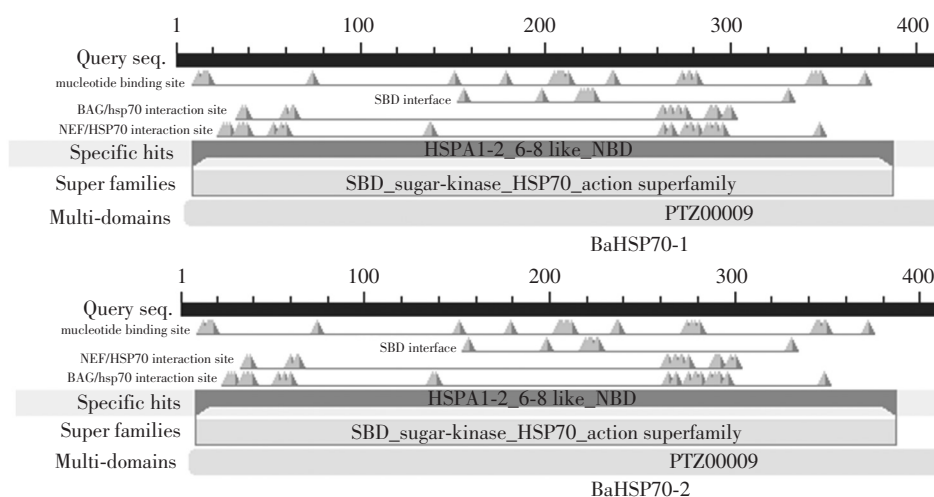


图 2 BaHSP70 保守结构域分析

Fig.2 The analysis of conserved domain prediction of *Bryum argenteum* HSP70

2.3 银叶真蕨 HSP70 蛋白理化性质分析

使用 ProtParam 对 HSP70 蛋白质序列进行理化性质分析,结果如表 2 所示:两个蛋白质序列消光系数相同,BaHSP70-1 不稳定指数高于 BaHSP70-2,说明一般情况下 BaHSP70-2 相对较稳定;但是总体平均亲水性 BaHSP70-1 相对较高。

2.4 银叶真蕨 HSP70 蛋白质亲疏水性分析

蛋白质氨基酸组成是蛋白质折叠,发挥功能的主要驱动力;蛋白质折叠时会形成疏水内核和亲水的表面,同时在潜在的跨膜区出现高疏水值区域。

使用 ProtScale 在线工具对 BaHSP70 蛋白质氨基酸序列的亲水性和疏水性进行预测,选取信号最为显著的 Kyte& Doolittle 氨基酸标度值,其它参数默认。

表 1 研究使用生物信息学在线分析工具

Table 1 The Web-Based tool utilizing for bioinformatics analysis

分析内容	在线分析工具	地址链接
全长鉴定	Pfam search	pfam.xfam.org/search
ORF	ORF Finder	ncbi.nlm.nih.gov/gorf/gorf.html
理化性质	ProtParam	web.expasy.org/protparam/
结构域分析	CDD	www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
疏水性分析	ProtScale	web.expasy.org/protscale/
信号肽预测	NetNES 1.1 server	www.cbs.dtu.dk/services/NetNES/
二级结构	NPSA-MRLC	npsa-prabi.ibcp.fr/
三级结构	Swiss-Model Workspace	swissmodel.expasy.org/
功能预测	Protfun 2.2 Server	www.cbs.dtu.dk/services/ProtFun/
模体分析	MEME	meme.nbcr.net/meme/

表 2 BaHSP70 一级结构理化性质分析

Table 2 Primary structure physicochemical property analysis of *Bryum argenteum* HSP70

理化性质	HSP70-1	HSP70-2
氨基酸数	649	650
分子式	C ₃₁₀₄ H ₄₉₈₄ N ₈₅₆ O ₉₉₈ S ₂₁	C ₃₁₁₁ H ₅₀₀₅ N ₈₇₅ O ₉₉₆ S ₁₆
理论等电点	5.12	5.14
负正电氨基酸残基数	98 & 81	91 & 83
不稳定性指数	33.56	28.59
脂肪指数	80.88	82.68
消光系数(M ⁻¹ cm ⁻¹)	42 080	42 080
分子质量	70 936.1	71 115.1
总平均亲水性	-0.429	-0.459

如图 3 所示,横坐标表示序列位置,纵坐标表示氨基酸的标度值。HSP70-1 有 11 个低分值峰, HSP70-2 有 9 个低分值峰,这些氨基酸的区域如图方框标记,属于高亲水性区域。其表面往往富集亲水性氨基酸,同时也是蛋白质进化中氨基酸插入的主要位点。HSP70-1 和 HSP70-2 疏水性的波形图不一样也反映两种蛋白在理化性质上有所差别。

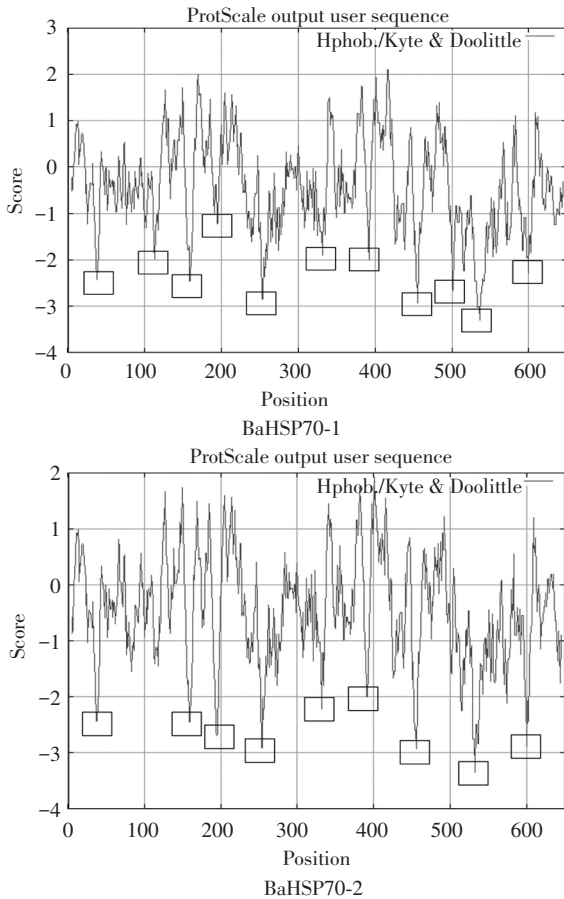


图 3 真藓 HSP70 蛋白质疏水性和亲水性预测

Fig.3 The prediction of hydrophobicity and hydrophilicity of *Bryum argenteum* HSP70

2.5 真藓 HSP70 蛋白信号肽预测和分析

采用基于神经网络模型 (Neural Networks models) 和隐马可夫模型 (Hidden Markov Models) 的 NetNES 1.1 server 预测银叶真藓热激蛋白 70 的富含亮氨酸的核输出信号^[19]; 结果如图 4 所示, BaHSP70-1 在序列第 400 位有富含亮氨酸的核输出信号, BaHSP70-2 在序列第 310 位和第 400 位出现富含亮氨酸的核输出信号。此区段说明 BaHSP70

有核输出信号,可以被某些特定的输出蛋白所结合,引导 HSP70 蛋白经由核孔进入细胞质特定位置执行相关功能。

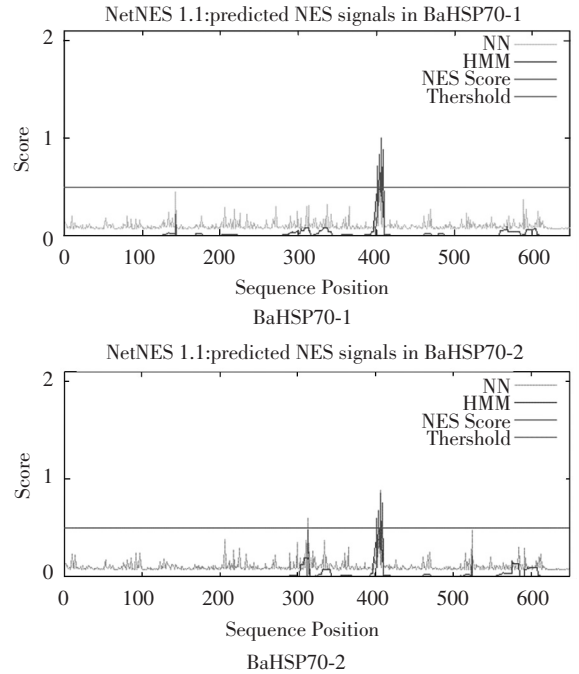


图 4 BaHSP70 核输出信号预测

Fig.4 The prediction of nuclear export signal of *Bryum argenteum* HSP70

2.6 真藓 HSP70 蛋白二级结构分析

使用 NPSA 在线分析工具提供的 MRLC 方法对 BaHSP70-1 和 BaHSP70-2 蛋白质的二级结构预测分析结果表明, HSP70 蛋白二级结构主要由 α 螺旋, β 折叠, 和无规卷曲三种形式构成, 其中 α 螺旋所占比例较高, N-段主要分布无规卷曲结构, C-端主要分布 α 螺旋, 统计的比例如表 3 所示, 二级结构分布如图 5 所示。

表 3 银叶真藓 HSP70 的二级结构预测分析

Table 3 The predicted analysis of secondary structure of HSP70

二级结构类型	氨基酸数目		所占比例/%	
	HSP70-1	HSP70-2	HSP70-1	HSP70-2
α 螺旋	256	245	39.45	37.69
β 折叠	120	121	18.49	18.62
无规卷曲	273	284	42.06	43.69

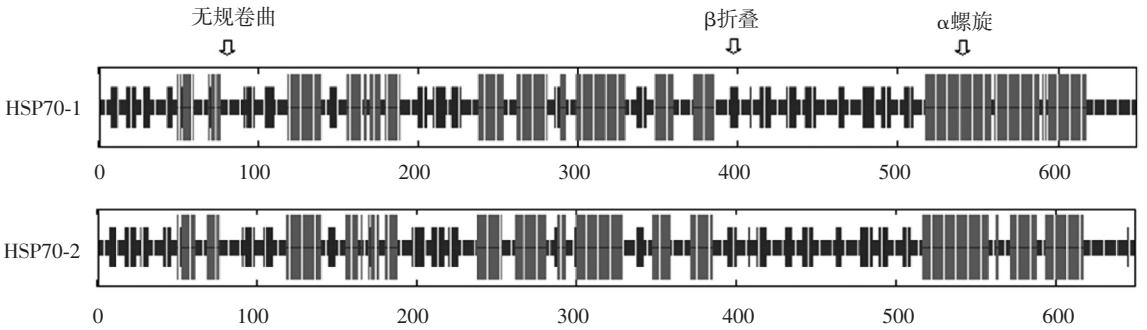


图 5 银叶真蕨 HSP70 二级结构

Fig.5 Predicted secondary structure of *Bryum argenteum* HSP70

2.7 三维结构模型分析

使用瑞士模型工作区 (Swiss-Model Workspace) 网络综合服务器对 HSP70 蛋白质结构做同源建模^[20], 建模方式采用自动方式, 同源模型建立主要分四个步骤: (1) 鉴定模板结构, 即首先将蛋白质序列通过 Blast 比对, 发现数据库中一致性较高的序列; 根据服务器分析结果, HSP70-1 和 HSP70-2 同时与数据库中的 Hsc71 kDa protein 结构最相似, 前者一致性达到 55%, 后者达 54%; (2) 目标序列和模板序列结构的比对, 即已找到的模板序列为参考序列进行比对, 对于氨基酸序列完全一致的区域, 建立相同的三级结构并进行结构叠合; (3) 建立模型, 根据网络综合服务器提供的算法, 构建蛋白质模型, 对于比对出未知的蛋白质, 首先找出主链结构, 然后侧链建模, 最后利用能量计算的方法进行结构优化; (4) 模型质量的评估, 即对构建好的模型用全局法和局部法进行质量评估。

本次模型评估使用全局法中的 QMEANscore4 记分值得估算。构建的蛋白质三维结构如图 6 所示, 对于 HSP70-1, 可信度值为 0.52, 可信度较高, 模板序列方法使用 X-ray, 分辨率为 1.90 Å, 比对的范围在第 7~559 个氨基酸, 覆盖度达到 0.85; 对于 HSP70-2, 可信度值为 0.34, 模型可参考, 模板序列方法使用 X-ray, 分辨率为 1.90 Å, 比对的范围在第 7~559 个氨基酸, 覆盖度达到 0.85。

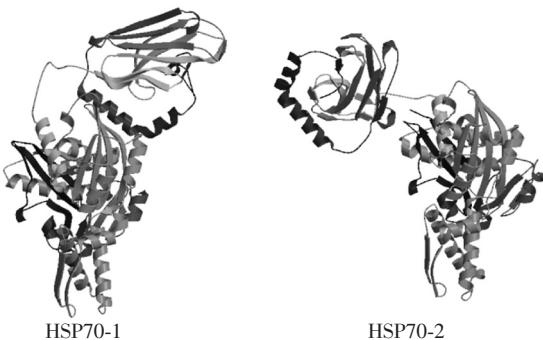


图 6 HSP70-1 和 HSP70-2 蛋白质的三维结构模型

Fig.6 The Three-dimensional structure model of HSP70-1 and HSP70-2

2.8 功能预测

使用 CBS 的 Profun 软件预测真蕨 HSP70 基因编码的功能^[21], 如表 4 所示, 热激蛋白 70 具有转录调控、生长因子、结构蛋白的可能性较高, 在 BaHSP70-1 中可能性依次为 0.442, 0.456, 0.436; 在 BaHSP70-2 分别中可能性依次为 0.396, 0.372, 0.465; 这符合生物学上植物受到外界环境胁迫时, 胁迫诱导产生 HSP70 蛋白, HSP70 作为分子伴侣识别细胞内的信号来引导转运新合成多肽的折叠以使其形成正确的构象发挥生物学功能的功能。

表 4 真蕨 HSP70 GO 功能预测

Table 4 GO functional prediction of *Bryum argenteum* HSP70

基因本体论分类	HSP70-1		HSP70-2	
	概率	可能性	概率	可能性
信号传导	0.065	0.304	0.056	0.26
受体	0.003	0.018	0.003	0.018
激素	0.001	0.206	0.001	0.206
结构蛋白	0.012	0.436	0.013	0.465
运载体	0.025	0.231	0.025	0.231
离子通道	0.013	0.223	0.013	0.219
电压门控离子通道	0.004	0.186	0.004	0.178
阳离子通道	0.01	0.218	0.01	0.216
转录	0.049	0.38	0.04	0.31
转录调控	0.055	0.442	0.05	0.396
胁迫应答	0.019	0.214	0.016	0.183
免疫应答	0.03	0.353	0.032	0.382
生长因子	0.006	0.456	0.005	0.372
金属离子转移	0.018	0.039	0.012	0.025

2.9 序列模体识别

使用 MEME 程序包对绿藻植物 (*Helicosporidium* sp)、银叶真蕨、小立碗蕨 (*Physcomitrella patens*)、水稻 (*Oryza sativa*)、拟南芥 (*Arabidopsis thaliana*)、玉米 (*Zea mays*) HSP70 蛋白进行序列模体识别分析, 模体分布参数设置为 "Any number of repetitions", 其

它参数默认。结果如图7所示,从序列中发现了5个 motif,其中以上物种含有公共 Motif5, Motif1, Motif2, Motif4,并且模体在 HSP70 基因上的分布顺序是一致的,间隔区长度也相近;其中,银叶真藓的 HSP70-1 和 HSP70-2 含有相同的模体结构;真藓 HSP70 蛋白的保守基序串联形式与水稻所具有的保守基序最相近,包括所有发现的5种模体,并且模体间的顺序完全一致。与绿藻和小立碗藓比较发现,绿藻 HSP70 序列的 C 端和小立碗藓序列的 N 端

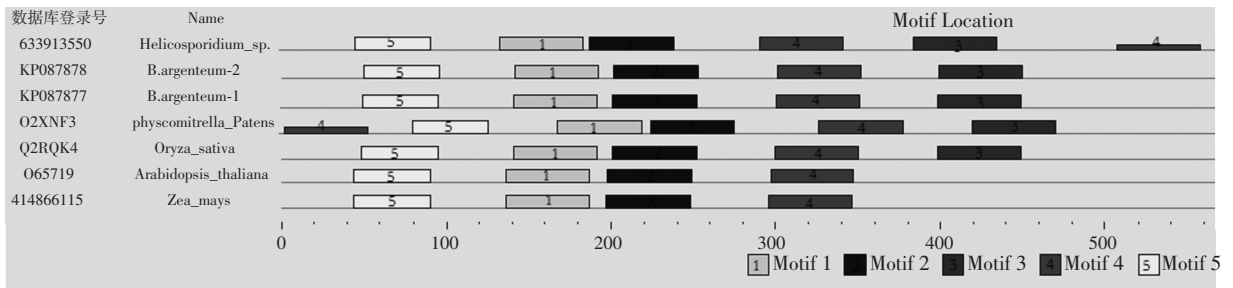


图7 HSP70 序列模体分析

Fig.7 The motif analysis of HSP70 sequence in plant

2.10 同源性分析及基因树构建

将银叶真藓 HSP70 蛋白质序列与胶球藻、团藻、扁藻等低等绿藻类植物和小立碗藓,拟南芥、水稻、玉米、小麦等高等植物的 HSP70 蛋白进行多序列比对发现,从低等绿藻类的植物到高等维管植物的蛋白质序列有较高的同源性。使用 ClustalW 软件进行多序列比对,选取比对结果中保守区段的蛋白质序列,使用 MEGA5.2 软件 Phylogeny 提供的邻位连接法 (Neighbor-Joining) 对同源的序列构建基因树,对于参数选项设置:检验方法选择“Bootstrap method”,重复次数为 1 000 次进行检测,以确保基因树分枝的稳定性,其它参数选择默认值。

构建的基因树如图8所示:基因树大致可以分为5支,高等植物节节草、小麦、菠菜、玉米等聚在一支(分支 I);低等植物螺旋藻、胶球藻、团藻、原藻、植物病原藻、扁藻、爱尔兰海藻、小立碗藓等聚在一支(分支 II);高等植物苜蓿、桑、玉米、节节草、小麦聚在一支(分支 III);低等植物爱尔兰海藻、团藻、植物病原藻、等较低等植物聚在一支(分支 IV);高等植物苹果、菠菜、桑、雪莲、铁皮石斛、银叶真藓聚在一支(分支 V)。选取的5条拟南芥 HSP70 序列分别位于基因树不同的分支上,在高等植物和低等植物类群分支中都有分布,说明不同 HSP 家族内部成员序列差别较大,分布广泛;而根据下载到 HSP70 蛋白质序列的注释信息,分析发现定位于不同亚细胞器的 HSP70 蛋白距离较近,其中,小立碗藓、拟南

多含有 Motif4,但是末端出现的 Motif4 保守度相对中间区段的 Motif 较低;与拟南芥和玉米比较发现,银叶真藓 HSP70 序列在 C 端多含有一个 Motif3。从植物由低等进化到高等的进化的角度来看,说明了 HSP70 序列家族在关键的功能结构域上保持着较高的一致性,可能这些保守的、顺序一致的模体在 HSP70 功能中发挥关键作用,但同时根据两端的较低模体的保守度,又丰富了 HSP70 在进化上的复杂性和多样性。

芥、黄瓜的叶绿体 HSP70 在基因树上位置较近(菱形标识),而小麦、节节草、原藻、拟南芥的线粒体 HSP70 在基因树上位置较近(三角形标识)。此外,银叶真藓的两条 HSP70 序列与雪莲的亲缘关系较近,聚为一支(圆形标识),同源性分别是 91.2% 和 86.6%,其次是轮藻,同源性分别是 76.5% 和 72.4%。

3 结论

本研究以真藓转录组数据库为基础,通过 Pfam 数据库提供的 HSP70 隐马可夫模型从数据库中鉴定出 2 条含有完整 ORF 的 HSP70 序列:HSP70-1 和 HSP70-2,二者一致性达 83%,全长含有 649aa 和 650aa,其编码的基因在 Unigene 数据库中的长度分别为 2 396 bp 和 2 356 bp。序列模体识别和解析研究发现真藓和其它植物含有 4 个相同的模体,并且各模体在蛋白质序列上分布顺序一致,可能在生命活动中发挥关键作用,N 端和 C 端也发现有保守度比较低的模体,丰富了 HSP70 在进化上的复杂度和多样性。真藓 HSP70 蛋白属于亲水性蛋白,HSP70-1 有 11 个亲水区,HSP70-2 有 9 个亲水区。二级结构、三级结构预测发现包括 α 螺旋, β 折叠, β 转角和卷曲螺旋结构。三级结构可信度高,在以后的研究中模型可参考。银叶真藓的两条热激蛋白序列 BaHSP70-1 和 BaHSP70-2 位于高等植物的进化支上,与雪莲同源性最高,同源性分别是 91.2% 和

86.6%。综上所述,该研究获得的真蕨两条 HSP70 蛋白的序列和基因序列,并且认识了 HSP70 蛋白质

的生物化学特性,蛋白质结构功能,以及在植物同源关系中的位置。

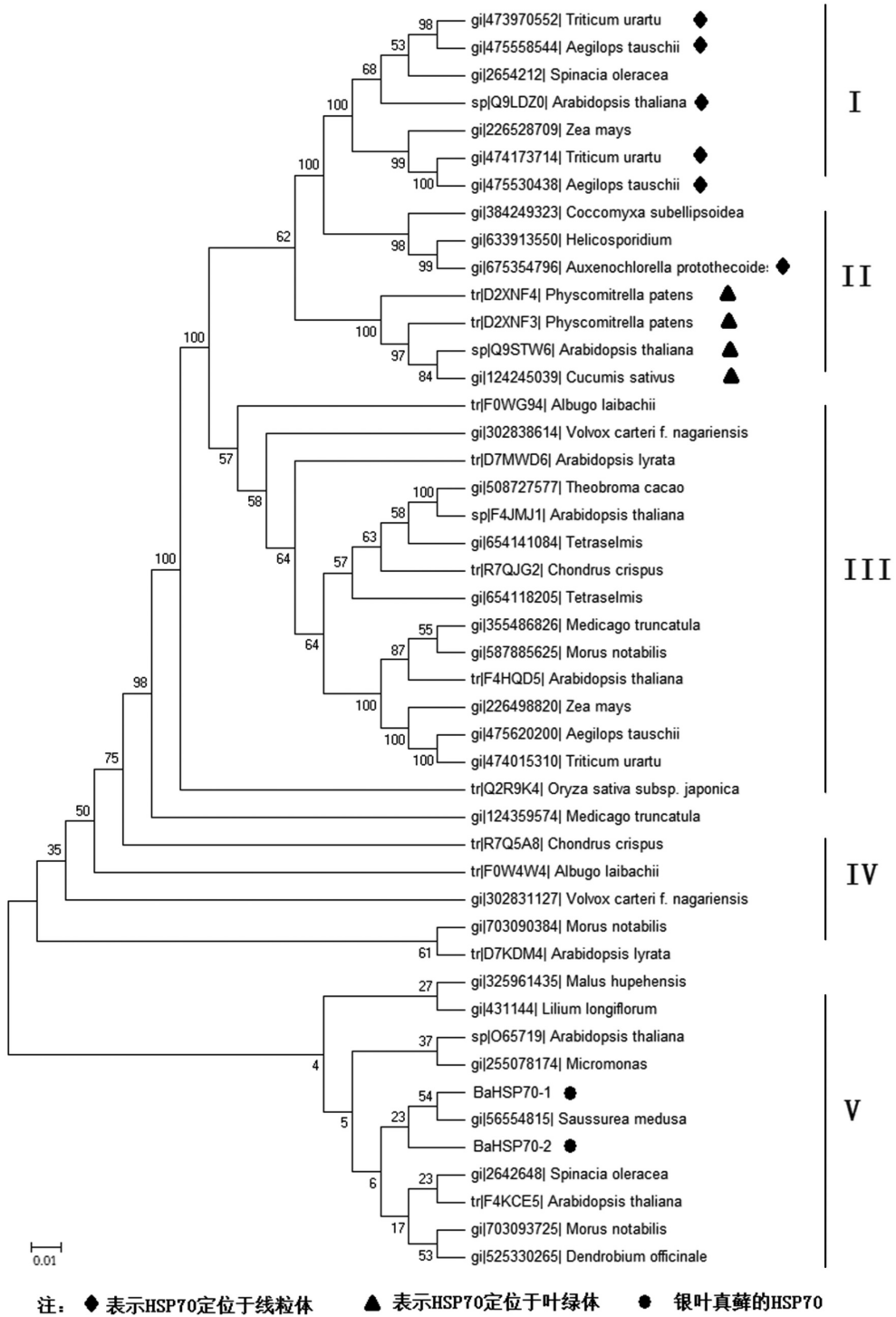


图 8 不同植物 HSP70 蛋白质序列基因树

Fig.8 Phylogenetic tree constructed by HSP70 protein sequence in plant

参考文献(References)

[1] BALLINGER D G, PARDUE M L. The control of protein-synthesis during heat-shock in drosophila cells involves

altered polypeptide elongation rates [J]. Cell, 1983, 33 (1):103-113.

[2] HORWICH A L. Molecular chaperones in cellular protein folding: The birth of a field[J]. Cell,2014,157(2):285-288.

[3] BUKAU B, HORWICH A L. The Hsp70 and Hsp60

- chaperone machines[J]. *Cell*, 1998, 92(3):351-366.
- [4] MAYER M P, BUKAU B. Hsp70 chaperones: Cellular functions and molecular mechanism [J]. *Cellular and Molecular Life Sciences*, 2005, 62(6):670-684.
- [5] HUI R, LI X R, CHEN C Y, et al. Responses of photosynthetic properties and chloroplast ultrastructure of *Bryum argenteum* from a desert biological soil crust to elevated ultraviolet-B radiation [J]. *Physiol Plantarum*, 2013, 147(4):489-501.
- [6] 郑云普, 赵建成, 张丙昌, 等. 荒漠生物结皮中藻类和苔藓植物研究进展[J]. *植物学报*, 2009, 3(3):371-380.
- ZHENG Yunpu, ZHAO Jiancheng, ZHANG Bingchang, et al. Advances on ecological studies of algae and mosses in biological soil crust [J]. *Chinese Bulletin of Botany*, 2009, 3(3):371-380.
- [7] BELNAP J. Nitrogen fixation in biological soil crusts from southeast Utah, USA [J]. *Biol Fert Soils*, 2002, 35(2):128-135.
- [8] 李利博, 赵建成. 真藓属 (*Bryum* Hedw.) 蒴齿形态特征及其分类学意义 [J]. *植物研究*, 2009, 29(6):651-658.
- LI Libo, ZHAO Jiancheng. Peristome morphology of *bryum hedw*, (*Musci Bryaceae*) and its taxonomic significance [J]. *Bulletin of Botanical Research*, 2009, 29(6):651-658.
- [9] 李利博, 赵建成, 曹娜. 河北省真藓属 (*Bryum* Hedw.) 植物分类学研究 [J]. *河北师范大学学报 (自然科学版)*, 2009, 29(6):800-809.
- LI Libo, ZHAO Jiancheng, CAO Na. A taxonomical study on genus *bryum hedw* of Hebei, China [J]. *Journal of Hebei Normal University (Natural Science Edition)*, 2009, 29(6):800-809.
- [10] 回嵘, 李新荣, 贾荣亮. 增强 UV-B 辐射对真藓结皮生理特性的影响 [J]. *生态学杂志*, 2012, 31(1):38-43.
- HUI Rong, LI Xinrong, JIA Rongliang. Effects of enhanced UV-B radiation on physiological characteristics of *Bryum argenteum* [J]. *Chinese Journal of Ecology*, 2012, 31(1):38-43.
- [11] SAKUDO A, XUE G A, KAWASHITA N, et al. Structure of the Prion Protein and Its Gene: An Analysis Using Bioinformatics and Computer Simulation [J]. *Curr Protein Pept Sc*, 2010, 11(2):166-79.
- [12] 张道远, 马瑞, 杨红兰, 等. 荒漠生物结皮中齿肋赤藓总 RNA 最佳提取方法的确立 [J]. *生物技术*, 2010, 20(3):52-54.
- ZHANG Daoyuan, MA Rui, YANG Honglan, et al. A method established for total RNA extraction from *syntrichia caninervis* in a temperate desert [J]. *Biotechnology*, 2010, 20(3):52-54.
- [13] WANG Z, GERSTEIN M, SNYDER M. RNA-Seq: a revolutionary tool for transcriptomics [J]. *Nat Rev Genet*, 2009, 10(1):57-63.
- [14] HAAS B J, PAPANICOLAOU A, YASSOUR M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis [J]. *Nature Protocols*, 2013, 8(8):1494-1512.
- [15] FINN R D, BATEMAN A, CLEMENTS J, et al. Pfam: the protein families database [J]. *Nucleic Acids Research*, 2014, 42(1):222-230.
- [16] FINN R D, CLEMENTS J, EDDY S R. HMMER web server: interactive sequence similarity searching [J]. *Nucleic Acids Research*, 2011, 39:29-37.
- [17] TAMURA K, PETERSON D, PETERSON N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Molecular Biology and Evolution*, 2011, 28(10):2731-2739.
- [18] MARCHLER-BAUER A, LU S N, ANDERSON J B, et al. CDD: a conserved domain database for the functional annotation of proteins [J]. *Nucleic Acids Research*, 2011, 39:225-229.
- [19] IA Cour T, KIEMER L, MOLGAARD A, et al. Analysis and prediction of leucine-rich nuclear export signals [J]. *Protein Engineering, Design & Selection: PEDS*, 2004, 17(6):527-536.
- [20] ARNOLD K, BORDOLI L, KOPP J, et al. The Swiss-Model workspace: a web-based environment for protein structure homology modelling [J]. *Bioinformatics*, 2006, 22(2):195-201.
- [21] JENSEN LJ, GUPTA R, STAERFELDT H H, et al. Prediction of human protein function according to Gene Ontology categories [J]. *Bioinformatics*, 2003, 19(5):635-642.