

doi:10.3969/j.issn.1672-5565.2016.03.02

# 基于不同物种的热休克蛋白 90 的生物信息学分析

叶 静, 陈 伟\*, 金殿川

(华北理工大学理学院, 河北 唐山 063009)

**摘要:**热休克蛋白 90 (Heat shock protein 90, Hsp90) 是生物体受到刺激时发生应激反应而产生的一类应激蛋白。Hsp90 包含 Hsp90A, Hsp90B, Hsp90C, TRAP 和 HtpG5 个亚家族。本文采用生物信息学方法对所选 11 个物种的 Hsp90 基因进行了分析。统计 Hsp90 亚家族在物种间的分布情况, 验证了 Hsp90 亚家族在物种间的分布规律, 即 Hsp90A 亚家族分布于除细菌外的其他所有物种中, Hsp90B 和 TRAP1 亚家族在物种间的分布无明显规律, Hsp90C 亚家族只存在于植物中, HtpG 亚家族大部分存在于细菌中。通过构建系统发育树, 发现 Hsp90 家族在进化过程中具有保守性。使用 Cell-PLoc, SubLoc v1.0, PSORT II 和 MultiLoc 四种亚细胞定位软件对所选的 11 个物种的 Hsp90 进行亚细胞定位分析, 发现 Hsp90A, HtpG 亚家族偏好出现在细胞质中, Hsp90B 亚家族除存在于细胞质外还存在于内质网中, Hsp90C 亚家族则集中于细胞质和线粒体中, TRAP1 亚家族基本位于线粒体中。

**关键词:**热休克蛋白 90; 系统发育树; 基因结构; 亚细胞定位

**中图分类号:** Q811      **文献标志码:** A      **文章编号:** 1672-5565(2016)03-134-05

## Bioinformatic analysis of Heat shock protein 90 from multiple species

YE Jing, CHEN Wei\*, JIN Dianchuan

(School of Science, North China University of Science and Technology, Tangshan Hebei 063009, China)

**Abstract:** Heat Shock Protein 90 (Hsp90) is a kind of proteins which are produced under stresses. Hsp90 includes five subfamilies, namely Hsp90A, Hsp90B, Hsp90C, TRAP and HtpG. In this paper, the Hsp90 genes from 11 species were analyzed using bioinformatics methods. The distributions of Hsp90 across species were studied. Hsp90A is found in all the selected species except for bacteria. Hsp90C subfamily is only present in plants. HtpG subfamily mostly presents in bacteria. Phylogenetic analysis revealed the evolutionary conservation of Hsp90 across species. Finally, by using Cell-PLoc, SubLoc v1.0, PSORT II and MultiLoc, the subcellular location of Hsp90 was predicted. Hsp90A and HtpG are bias to cytoplasm. Hsp90B prefers to both endoplasmic reticulum and cytoplasm. Hsp90C are located in cytoplasm and mitochondria. TRAP1 mainly located in mitochondria.

**Keywords:** Hsp90; Phylogenetic tree; Gene Structure; Subcellular localization

热休克蛋白(Heat shock protein, Hsp)是生物体受到环境中物理、化学、生物等因素刺激时发生应激反应而在细胞中产生的一类应激蛋白<sup>[1]</sup>。按照其分子量大小的不同,热休克蛋白可以分为 Hsp20 (sHsp), Hsp40 (J-class protein), Hsp60, Hsp70, Hsp90 和 Hsp100 等 6 个家族<sup>[2]</sup>,它们在细胞中发挥着不同的生物学功能。

自 1962 年首次在果蝇体内发现以来<sup>[3]</sup>, 研究人

员已从细菌,酵母,线虫,大鼠,小鼠,和人类等生物体内分别发现了热休克蛋白的存在。近年来,热休克蛋白的应用更是十分广泛,从农业到抗癌试剂的制备,热休克蛋白都扮演着重要的角色。

然而我们对热休克蛋白进化关系的认识仍旧有限,因此通过比较基因组学和生物信息学方法对热休克蛋白进行研究,将帮助我们阐明热休克蛋白的进化规律,对理解热休克蛋白的功能也具有重要的

收稿日期:2016-06-06;修回日期:2016-06-23.

基金项目:河北省自然科学基金项目(C2013209105)。

作者简介:叶静,女,硕士研究生,研究方向:生物信息学;E-mail: yejing@gmail.com.

\* 通信作者:陈伟,男,副教授,研究方向:生物信息学;E-mail: greatchen@ncst.edu.cn.

生物学意义。

Hsp90 作为热休克蛋白家族中的一类应激蛋白,广泛存在于各类机体中,古老且具有高度保守性<sup>[4]</sup>。其家族之下又分为五个亚家族,分别是 Hsp90A, Hsp90B, Hsp90C, TRAP 和 HtpG 亚家族,其中 Hsp90A 亚家族在 Hsp90 家族中数量是最多的,分布于除细菌外的其他所有物种中, Hsp90C 亚家族只存在于植物中,而 HtpG 亚家族则绝大多数存在于细菌中<sup>[5]</sup>。2006 年,Chen<sup>[5]</sup>等人通过对 32 个物种的 Hsp90 家族的研究,建立了包含所有 Hsp90 家族信息及进化的框架。

作为一种重要的分子伴侣, Hsp90 参与多种生理过程,如信号传导、蛋白质折叠降解和多种应激反应的调节<sup>[6]</sup>。研究表明 Hsp90 的表达与多种癌症是密切相关的,如肺癌,胰腺癌等,因此可以通过检测 Hsp90 的含量进而有效地监控癌症<sup>[7]</sup>。然而,随着科技的飞速发展,我们需要快速简洁的从生物信息学的角度去分析 Hsp90 家族的进化情况,进而研究 Hsp90 的功能作用,确定其进化及基因结构情况,验证 Hsp90 五个亚家族在物种间的分布情况以及亚细胞定位情况。

## 1 材料与方法

### 1.1 材料

从 HSP1R (<http://pdslab.biochem.iisc.ernet.in/>)

数据库,收集百日咳博代氏杆菌 (*Bordetella pertussis*),破伤风杆菌 (*Clostridium tetani*),致病疫霉 (*Phytophthora infestans*),四膜虫 (*Tetrahymena thermophile*),构巢曲霉 (*Aspergillus nidulans*),新型隐球菌 (*Cryptococcus neoformans*),莱茵衣藻 (*Chlamydomonas reinhardtii*),埃及斑蚊 (*Aedes aegypti*),大豆 (*Glycine max*),高粱 (*Sorghum bicolor*) 和 水稻 (*Oryza sativa*) 11 个物种的 Hsp90 基因家族成员的 CDS (或基因) 序列和蛋白质序列。

由此得到来自上述 11 个物种的 37 条 Hsp90 序列,它们在物种间分布情况和 Hsp90 亚家族的分布情况见表 1。其中属于 Hsp90A 亚家族的有 20 条, 约占了所选物种亚家族的 54.05%, 在除百日咳博代氏杆菌和破伤风杆菌外的其余物种中都存在; Hsp90B 亚家族的数量较少但分布较广泛, 在致病疫霉、新型隐球菌、莱茵衣藻、高粱、水稻和埃及斑蚊中都有发现; Hsp90C 亚家族存在于莱茵衣藻、高粱和 水稻中; TRAP1 亚家族存在于致病疫霉、四膜虫、水稻和埃及斑蚊四个物种中; HtpG 亚家族只存在于百日咳博代氏杆菌和破伤风杆菌中, 其他物种中没有发现。进一步验证了 Chen<sup>[5]</sup> 对于 Hsp90 亚家族在物种间的分布情况即 Hsp90A 亚家族分布于除细菌外的其他所有物种中, Hsp90B 和 TRAP1 亚家族在物种间的分布情况没有规律, Hsp90C 亚家族只存在于植物中, HtpG 亚家族则绝大多数存在于细菌中。

表 1 Hsp90 亚家族在 11 个物种间分布情况

Table 1 The distribution of Hsp90 family in 11 species

物种名称	Abbreviation	Hsp90A	Hsp90B	Hsp90C	TRAP1	HtpG	All
<i>Bordetella pertussis</i>	Bp	0	0	0	0	1	1
<i>Clostridium tetani</i>	Ct	0	0	0	0	1	1
<i>Phytophthora infestans</i>	Pi	1	1	0	1	0	3
<i>Tetrahymena thermophile</i>	Tt	1	0	0	1	0	2
<i>Aspergillus nidulan</i>	An	1	0	0	0	0	1
<i>Cryptococcus neoformans</i>	Cn	1	1	0	0	0	2
<i>Chlamydomonas reinhardtii</i>	Cr	1	1	1	0	0	3
<i>Glycine max</i>	Gm	4	0	0	0	0	4
<i>Sorghum bicolor</i>	Sb	4	1	2	0	0	7
<i>Oryza sativa</i>	Os	4	1	2	1	0	8
<i>Aedes aegypti</i>	Aa	3	1	0	1	0	5

### 1.2 系统发育树的简述

根据所处理数据的类型, 可以将系统发育树的构建方法分为两大类。一类是基于距离的构建方法, 利用所有物种或分类单元间的进化距离, 依据一定的原则及算法构建对应的系统发育树。其中包括非加权分组平均法 (Unweighted-pair-group method

with arithmetic mean, UPGMA)、最小进化法 (Minimum evolution) 和邻近归并法 (Neighbor-joining method) 等。其中 UPGMA 法已经较少使用, 因大多数蛋白家族起源较早, 序列分化程度较大, 相互之间较为远源, 因此构建此类系统发育树的时候使用邻近归并法较多。

另一类方法是基于离散特征的构建方法,利用的是具有离散特征状态的数据,如基因序列中特定位点的核苷酸。此类方法中包括最大简约法 (Maximum parsimony method), 最大似然法 (Maximum likelihood approach) 等。

1.3 亚细胞定位的概述及软件介绍

亚细胞定位是某种蛋白或者蛋白的产物在细胞内具体存在的位置,即根据所给出的蛋白质序列来预测蛋白质所在的亚细胞位置<sup>[8]</sup>。蛋白质亚细胞定位的准确预测使得我们可以更快的了解该蛋白质功能。为了准确的给出 Hsp90 的亚细胞定位信息,选取 Cell-PLoc<sup>[9]</sup>, SubLoc v1.0<sup>[10]</sup>, PSORT II<sup>[11]</sup> 和 MultiLoc<sup>[12]</sup> 四个常用的蛋白质亚细胞定位预测软件,采用集成分类器方法对 Hsp90 进行亚细胞定位分析,即通过上述四种软件分别对 Hsp90 进行亚细胞定位预测,以预测结果中所出现的频率最高的亚

细胞位置作为最终的亚细胞定位结果。

2 结果和分析

2.1 Hsp90 基因结构分析

根据 HSP1R 数据库中的基因信息在 NCBI 数据库中搜索本文所选的 37 条 Hsp90 的基因序列,搜索结果显示,在 37 条基因序列中,有 8 条序列无注释信息,使用 GSDS2.0<sup>[13]</sup> (<https://gsds.cbi.pku.edu.cn/>) 软件绘制其余 29 条 Hsp90 基因序列的外显子-内含子结构图(见图 1)。发现在 Hsp90 的 5 个亚家族中,Hsp90C 亚家族的基因结构最为复杂,内含子数目较其余 4 个亚家族的内含子数目要多,其中高粱基因组中的 Hsp90C0251Sb 和 Hsp90C0252Sb 基因的结构最为复杂,含有 18 个内含子。

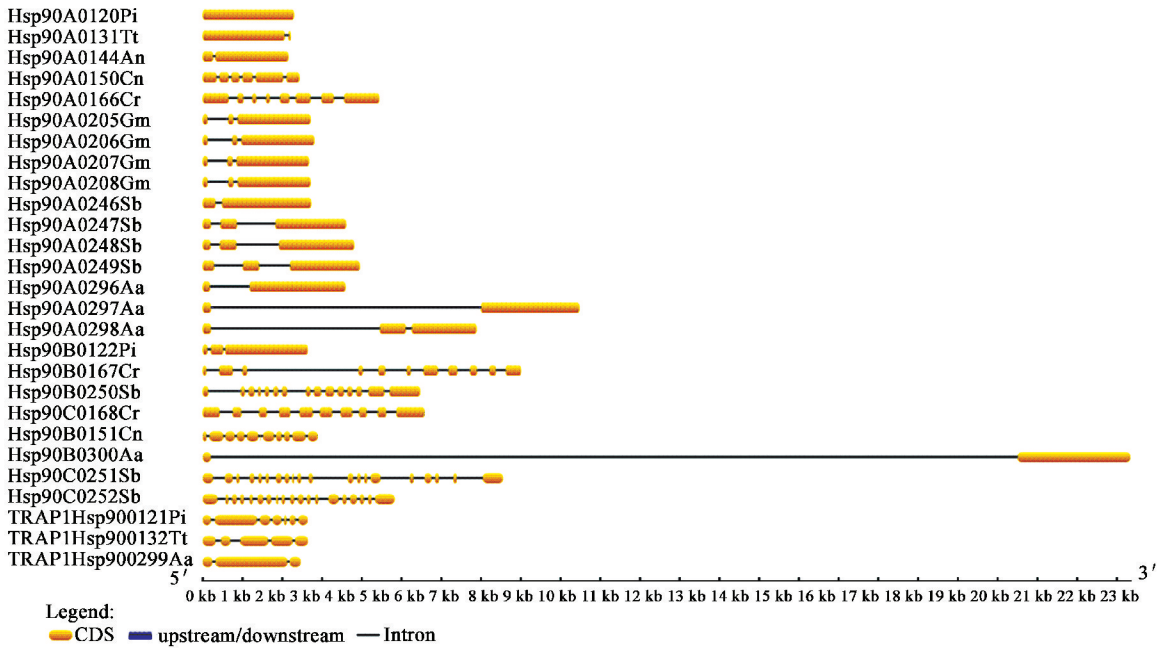


图 1 Hsp90 家族的基因结构分析

Fig. 1 The genetic structure analysis of Hsp90 family

2.2 Hsp90 家族系统发育树的构建和分析

利用 MEGA 软件对所研究的 11 个物种的 Hsp90 家族的基因做 CLUSTAL 多序列比对,之后使用邻近归并法构建对应的系统发育树,设置自举检验 (Bootstrap) 的值为 1 000,并对系统发育树中自举检验值小于 50 的树枝做归并处理,结果如图 2 所示。从构建出的系统发育树可以看出,Hsp90 家族形成了四个分支,分别对应亚家族 Hsp90A,亚家族 Hsp90C,和亚家族 TRAP1,而 Hsp90B 和 HtpG 两个亚家族区分不明显,两者聚集到了同一分枝上,这可能与 HtpG 亚家族样本数量少有关。

2.3 Hsp90 家族的亚细胞定位的结果

分别使用 Cell-PLoc、SubLoc v1.0、PSORT II 和 MultiLoc 对 Hsp90 家族的蛋白质进行定位预测,综合四个分类器得到的结果最后确定每个蛋白质在亚细胞中的位置,如表 2 所示。可以看出,Hsp90 家族在细胞质、线粒体、内质网和细胞核等细胞器中都有发现。其中 Hsp90A 和 HtpG 两个亚家族在细胞质中存在的比例较高,Hsp90B 亚家族集中于细胞质和内质网中,Hsp90C 亚家族则集中于细胞质和线粒体中,TPAP1 亚家族则更偏好定位于线粒体中。

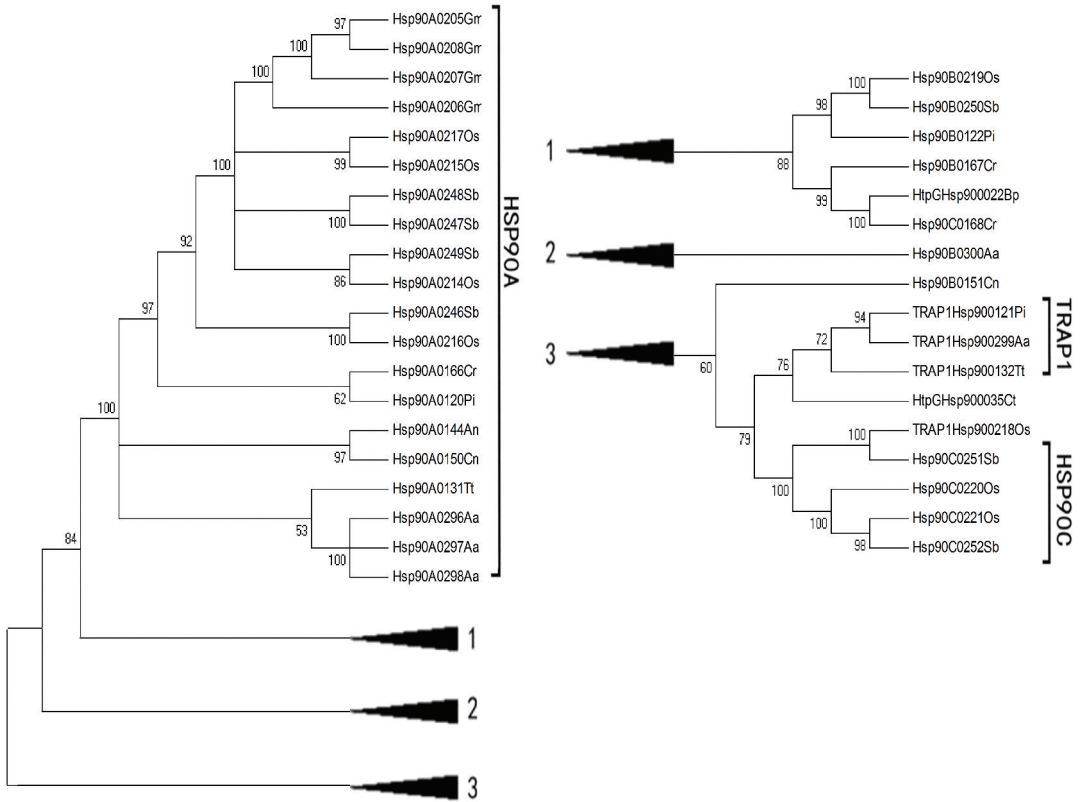


图 2 11 个物种 Hsp90 家族的系统发育树

Fig. 2 The Hsp90 family phylogenetic tree in 11 species

注:图中 Bp 代表百日咳博代氏杆菌、Ct 代表破伤风杆菌、Pi 代表致病疫霉、Ti 代表四膜虫、An 代表构巢曲霉、Cn 代表新型隐球菌、Cr 代表莱茵衣藻、Gm 代表大豆、Sb 代表高粱、Os 代表水稻、Aa 代表埃及斑蚊。

表 2 Hsp90 家族亚细胞预测结果

Table 2 Hsp90 family subcellular prediction results

亚家族	序列名称	Genbank 登录号	亚细胞位点
Hsp90A	Hsp90A0120Pi	XM_002998495	Cytoplasm
	Hsp90A0131Ti	XM_001009780	Cytoplasm
	Hsp90A0144An	XM_676446	Cytoplasm
	Hsp90A0150Cn	XM_766999	Cytoplasm
	Hsp90A0166Cr	XM_001695212	Cytoplasm
	Hsp90A0205Gm	ACI31552	Cytoplasm
	Hsp90A0206Gm	ACI31551	Cytoplasm
	Hsp90A0207Gm	ADC45396	Cytoplasm
	Hsp90A0208Gm	ADC45395	Cytoplasm
	Hsp90A0246Sb	XM_002447368	Cytoplasm
	Hsp90A0247Sb	XM_002462509	Cytoplasm
	Hsp90A0248Sb	XM_002460388	Cytoplasm
	Hsp90A0249Sb	XM_002444759	Cytoplasm
	Hsp90A0214Os	CAA77978	Cytoplasm
	Hsp90A0215Os	EAZ09522	Cytoplasm
	Hsp90A0216Os	EAY92778	Cytoplasm
	Hsp90A0217Os	EEC84949	Nucleus
	Hsp90A0296Aa	XM_001649702	Cytoplasm
	Hsp90A0297Aa	XM_001655591	Cytoplasm
	Hsp90A0298Aa	XM_001649701	Cytoplasm

续(表2)

亚家族	序列名称	Genbank 登录号	亚细胞位点
Hsp90B	Hsp90B0122Pi	XM_002999089	Cytoplasm
	Hsp90B0151Cn	XM_770303	Cytoplasm
	Hsp90B0167Cr	XM_001701833	Cytoplasm
	Hsp90B0250Sb	XM_002438985	Endoplasmic reticulum
	Hsp90B0219Os	EAZ02369	Endoplasmic reticulum
	Hsp90B0300Aa	XM_001662901	Endoplasmic reticulum
	Hsp90C0168Cr	XM_001702932	Mitochondrion
Hsp90C	Hsp90C0251Sb	XM_002443217	Mitochondrion
	Hsp90C0252Sb	XM_002444790	Cytoplasm
	Hsp90C0220Os	EEC84746	Cytoplasm
	Hsp90C0221Os	EEC83781	Cytoplasm
	TRAP1Hsp900121Pi	XM_002999136	Mitochondrion
TRAP1	TRAP1Hsp900132Tt	XM_001015843	Nuclear
	TRAP1Hsp10218Os	EEC69377	Nuclear
	TRAP1Hsp900299Aa	XM_001654708	Mitochondrion
HtpG	HtpGHsp900022Bp	CAE40453	Cytoplasm
	HtpGHsp900035Ct	AA036046	Cytoplasm

### 3 结 论

Hsp90 分布广泛,存在于各个物种间,本文通过对其不同亚家族在物种间分布进行统计,验证了其五个亚家族在物种间分布各有特点,其中 Hsp90A 亚家族数量最多,分布于除细菌外的其他所有物种中,Hsp90C 亚家族只存在于植物中,而 HtpG 亚家族则绝大多数存在于细菌中。Hsp90 家族的基因结构分析说明 Hsp90C 亚家族的结构最为复杂。构建的系统发育树进一步证明 Hsp90 在进化过程中具有保守性,不同物种的同一亚家族在系统发育树上聚集明显,即同一亚家族间的序列有较好的相似性。通过对 Hsp90 亚家族进行亚细胞定位分析,基本确定了 Hsp90A, Hsp90B, Hsp90C, TRAP1 和 HtpG 亚家族在细胞器中的位置。希望这些研究结果能够为 Hsp90 功能的研究提供新线索。

### 参考文献(References)

[1] SAKATANI M, BONILLA L, DOBBS K B, et al. Changes in the transcriptome of morula-stage bovine embryos caused by heat shock: relationship to developmental acquisition of thermotolerance [J]. *Reproductive Biology Endocrinology*, 2013(11):3. DOI: 10.1186/1477-7827-11-3.

[2] RATHEESH K R, NAGARAJAN N S, ARUNRAJ S P, et al. HspIR: a manually annotated heat shock protein information resource [J]. *Bioinformatics*, 2012, 28(21): 2853-2855. DOI: 10.1093/bioinformatics/bts520.

[3] RITOSSA F. A new puffing pattern induced by temperature shock and DNP in drosophila [J]. *Experientia*, 1962, 18(12):571-573. DOI: 10.1007/BF02172188.

[4] CSERMELY P, SCHNAIDER T, SÖTI C, et al. The 90-kDa molecular chaperone family: structure, function, and clinical

applications [J]. *Pharmacology Therapeutics*, 1998, 79(2): 129-168.

[5] CHEN B, ZHONG D, MONTEIRO A. Comparative genomics and evolution of the Hsp90 family of genes across all kingdoms of organisms [J]. *BMC Genomics*, 2006(7):156. DOI:10.1186/1471-2164-7-156.

[6] ALI A B S, O'CARROL R, OVSENEK N. Hsp90 interacts with and regulates the activity of heat shock factor 1 in *Xenopus oocytes* [J]. *Molecular Biology*, 1998, 18(9): 4949-4960.

[7] 宋晓敏. 分泌型热休克蛋白 90a 在肿瘤发生和转移中的作用机理[D]. 北京:清华大学, 2010.  
SONG Xiaomin. Secretory type heat shock protein 90a mechanism in tumorigenesis and metastasis [D]. Beijing: Tsinghua University, 2010.

[8] 吴泽月, 陈月辉. 蛋白质亚细胞定位预测研究进展 [J]. *山东师范大学学报*, 2012, 4(27): 33-37.  
WU Zeyue, CHEN Yuehui. The research progress of protein subcellular localization prediction [J]. *Journal of Shandong Normal University*, 2012, 4(27): 33-37.

[9] CHOU K C, SHEN H B. Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms [J]. *Nature Protocols*, 2008(153): 162. DOI:10.1038/nprot.2007.494.

[10] HUA S, SUN Z. Support vector machine approach for protein subcellular localization prediction [J]. *Bioinformatics*, 2001, 17(8): 721-728.

[11] MITSUTERU C N, KENTA N. Improvement of PSORT II Protein Sorting Prediction for Mammalian Proteins [J]. *Genome Informatics*, 2002(13): 441-442.

[12] HÖGLUNDA, DÖNNES P, BLUM T, et al. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition [J]. *Bioinformatics*, 2006, 22(10): 1158-1165.

[13] BO H, JINPU J, AN-YUAN G, et al. CSDS 2.0: an upgraded gene feature visualization server [J]. *Bioinformatics*, 2015, 31(8): 1296-1297.