

史建磊,熊自立,李 涛,等. 番茄组蛋白乙酰转移酶(*HAT*)的全基因组鉴定与分析[J].江苏农业学报,2020,36(3):666-674.
doi:10.3969/j.issn.1000-4440.2020.03.019

番茄组蛋白乙酰转移酶 (*HAT*) 的全基因组鉴定与分析

史建磊¹, 熊自立¹, 李 涛², 张海利¹, 宰文珊¹

(1.温州科技职业学院,浙江 温州 325006; 2.广东省农业科学院蔬菜研究所,广东 广州 510640)

摘要: 组蛋白乙酰化修饰在基因表达调控方面扮演着重要角色。为深入挖掘番茄组蛋白乙酰转移酶(*HAT*)基因,本研究运用生物信息学方法,共鉴定出 26 个番茄 *HAT*。聚类分析发现,这些 *HAT* 可以分为 7 组,分别为 *HAG*、*HAG1*、*HAG2*、*MCC1*、*HAM*、*HAF* 和 *HAC*。其中,*GNAT* 家族(*GCN5*、*HAT1*、*MCC1*) 18 个、*p300/CBP* 家族 4 个、*MYST* 和 *TAFII250* 家族各 1 个。6 对同源基因 *Ka/Ks* 值均小于 1,说明其进化中经历了纯化选择。不同家族 *HAT* 理化性质存在差异,但均为亲水蛋白质,主要定位于细胞核,蛋白质二级结构主要包括 α 螺旋和无规则卷曲,具有典型的保守基序和特征结构域。染色体定位发现,番茄 *HAT* 基因不均匀地散布于 12 条染色体上,且呈两端分布,其中 1 号和 9 号染色体最多,部分基因经串联重复形成基因簇。基因表达分析发现,*GNAT* 家族番茄特有的 *HAT* 和 *HAG2* 组基因在病菌、盐和热胁迫下具有较高的表达水平,一些基因具有病菌诱导特异性和耐盐(热)与盐(热)敏材料上表达差异性。可见,番茄 *HAT* 在结构、性质和功能上具有多样性,其在全基因组水平上的鉴定与分析为相关基因克隆利用提供了依据。

关键词: 番茄; 组蛋白乙酰转移酶; 生物信息学

中图分类号: Q786 文献标识码: A 文章编号: 1000-4440(2020)03-0666-09

Genome-wide identification and analysis of histone acetyltransferase (*HAT*) in tomato

SHI Jian-lei¹, XIONG Zi-li¹, LI Tao², ZHANG Hai-li¹, ZAI Wen-shan¹

(1. Wenzhou Vocational College of Science and Technology, Wenzhou 325006, China; 2. Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China)

Abstract: Histone acetylation plays an important role in the regulation of gene expression. In order to further explore histone acetyltransferase (*HAT*) gene in tomato, 26 *HAT* genes were identified by using bioinformatics method in this study. The results of cluster analysis showed that these *HAT* could be divided into seven groups, namely *HAG*, *HAG1*, *HAG2*, *MCC1*, *HAM*, *HAF* and *HAC*. Among them, 18 *HAT* genes were *GNAT* family members (*GCN5*, *HAT1*, *MCC1*), four *HAT* genes were *p300/CBP* family members, and one was *MYST* or *TAFII250* family member. The *Ka/Ks* values of six pairs of paralogous genes were all less than one, indicating that purifying selection had occurred in evolution. The physicochemical properties of *HAT* in different families were different, but all of them were hydrophilic proteins, mainly located in the nucle-

us. Furthermore, the secondary structure of proteins mainly included alpha helix and random coil, and had typical conserved motifs and domains. Chromosome localization results showed that tomato *HAT* gene were unevenly distributed on 12 chromosomes and at both ends. In addition, the number of chromosome 1 and 9 was the most, and some genes formed gene clusters by tandem duplication. Results of gene expression analysis indicated that the

收稿日期:2019-11-20

基金项目:浙江省高等学校访问学者项目(FX2018147);浙江省农业(蔬菜)新品种选育重大科技专项子课题(2016C02051-1-2);温州市科技计划项目(2019ZX007)

作者简介:史建磊(1982-),男,河北张家口人,硕士,副教授,研究方向为蔬菜遗传育种与生物技术。(E-mail) sjlhebau@163.com

通讯作者:宰文珊, (E-mail) t000606@wzvcst.edu.cn

HAT genes, which were unique in tomato, and *HAG2* gene, had higher expression levels under pathogen, salt and heat stress. Also, some genes had pathogen-inducible specificity and different expression levels between resistant and sensitive materials. It can be seen that *HAT* gene in tomato has diversity in structure, character and function, and the genome-wide identification and analysis results provide a basis for cloning and utilization of related genes.

Key words: tomato; histone acetyltransferase (*HAT*); bioinformatics

植物在生长发育过程中面临着各种生物与非生物胁迫,同时,在长期历史进化中植物本身也形成了多种调控与防御机制,来保障个体正常生长发育与群体繁衍。组蛋白修饰在基因表达调控方面扮演着重要角色,通过组蛋白乙酰转移酶(*HAT*)促进基因表达和组蛋白去乙酰化酶(*HDAC*)抑制基因表达。组蛋白乙酰化是一种重要的翻译后修饰,通常与基因的转录激活有关,根据 *HAT* 在细胞内分布及诱导乙酰化后的效应分为 A 型(核内)和 B 型(胞质)两类^[1]。目前已知的 A 型 *HAT* 种类较多,如 *GNAT* (*Gcn5* 相关 *N*-末端乙酰转移酶)、*MYST* (*MOZ*, *Ybf2/Sas3*, *Sas2* 和 *Tip60*)、*p300/CBP* (cAMP 响应原件结合蛋白)、*TAF-II250* (TATA 结合蛋白相关因子)等^[2-4]。

前人已对拟南芥(*Arabidopsis thaliana*)^[5]、水稻(*Oryza sativa*)^[6]和葡萄(*Vitis vinifera*)^[7]组蛋白乙酰化基因家族进行了生物信息和基因功能研究。组蛋白乙酰化和去乙酰化修饰除参与植物生长发育外,也参与多种生物与非生物胁迫响应^[8-10]。番茄(*Solanum lycopersicum*)是世界上重要的园艺和经济作物,也是重要的科研模式作物,尽管 Cigliano 等^[11] 2013 年在番茄中发现了 32 个 *HAT*,但与 *HDAC* 相比,相关研究仍非常有限。同时,番茄全基因组测序的完成^[12],为在全基因组水平上鉴定番茄 *HAT* 提供了条件。鉴于 *HAT* 基因家族的重要生理功能,本研究利用生物信息学方法对番茄 *HAT* 进行鉴定与分析,旨在为其进一步功能研究和分子育种应用提供基础信息。

1 材料与方法

1.1 番茄组蛋白乙酰转移酶(*HAT*)全基因组鉴定

番茄全基因组序列及注释信息(ITAG3.2 release)下载自茄科基因组网站 SGN (<https://sol-genomics.net/>),并利用 BioEdit 7.2.5 构建本地数据库。采用 2 种方法鉴定番茄组蛋白乙酰转移酶(*HAT*):从 Pfam 32.0 (<http://pfam.xfam.org/>)数据库中下载 *HAT* 隐马尔科夫模型(Hidden Markov

Model, HMM)特征序列,结合拟南芥 *HAT*,作为问询序列 Blastp ($E\text{-value} < 1$) 构建本地数据库;同时以“histone acetyltransferase”为关键词在 SGN 中进行检索,去除冗余基因,即获得候选基因。然后通过 Pfam 和 SMART (<http://smart.embl-heidelberg.de/smart/batch.pl>)鉴定 *HAT* 特征结构域,将不含目的结构域的候选基因排除。

1.2 番茄组蛋白乙酰转移酶(*HAT*)系统树的构建

提取番茄 *HAT* 氨基酸序列,使用 ClustalX 1.83 进行序列联配,然后利用 MEGA 7.0 的邻接法(Neighbor-Joining, NJ) (参数设置:Bootstrap method 1 000, p-distance and Pairwise deletion)进行系统发育树构建,去除 Bootstrap 支持率低于 50% 的节点,显示各分支。

1.3 番茄组蛋白乙酰转移酶(*HAT*)理化参数分析

利用在线工具 ExPASy 的 ProtParam 程序(<https://web.expasy.org/protparam/>)进行蛋白质理化参数预测, SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)预测蛋白质二级结构,利用 SWISS-MODEL (<https://swissmodel.expasy.org/>)进行蛋白质三维结构同源建模,利用 WoLF PSORT (<https://wolfsort.hgc.jp/>)进行蛋白质亚细胞定位预测。

1.4 番茄组蛋白乙酰转移酶(*HAT*)结构特征分析

利用在线工具 GSDS 2.0 (<http://gsds.cbi.pku.edu.cn/>)绘制基因的外显子-内含子结构模式图。利用数据库 Pfam 32.0 和在线工具 MEME 5.1.0 (<http://meme-suite.org/tools/meme>)分别分析蛋白质保守结构域和基序(Motifs),并使用 TBtools 软件^[13]对其进行可视化。

1.5 番茄组蛋白乙酰转移酶(*HAT*)的染色体定位

结合 SGN 中基因位置信息(Tomato SL3.0),利用 MapDraw 2.1 对番茄 *HAT* 进行染色体定位,绘制其染色体分布图。同时进行基因簇和串联重复分析,其中划分基因簇的原则为:2 个相邻目的基因间距小于 200 kb 且其他基因不得多于 8 个;判断串联

重复的原则为:相邻目的基因间距小于 100 kb 且序列相似度大于 70%^[14]。用 KaKs_Calculator 2.0 软件^[15]计算基因间碱基替换率(Ka/Ks)。

1.6 番茄组蛋白乙酰转移酶(*HAT*)基因表达分析

基于已发表的 Affymetrix 番茄基因组和 TOM2 芯片数据,下载 8 张芯片表达结果,分别为 E023 (真菌激发子)、E024 (番茄细菌性溃疡病菌)、E031 (灰霉菌)、E048 (马铃薯纺锤块茎类病毒)、E046 (番茄斑萎病毒)、E047 (丛枝菌根真菌球霉菌)、E051 (盐胁迫)和 E053 (干旱和热胁迫),利用 MeV 4.9.0 绘制基因表达谱。

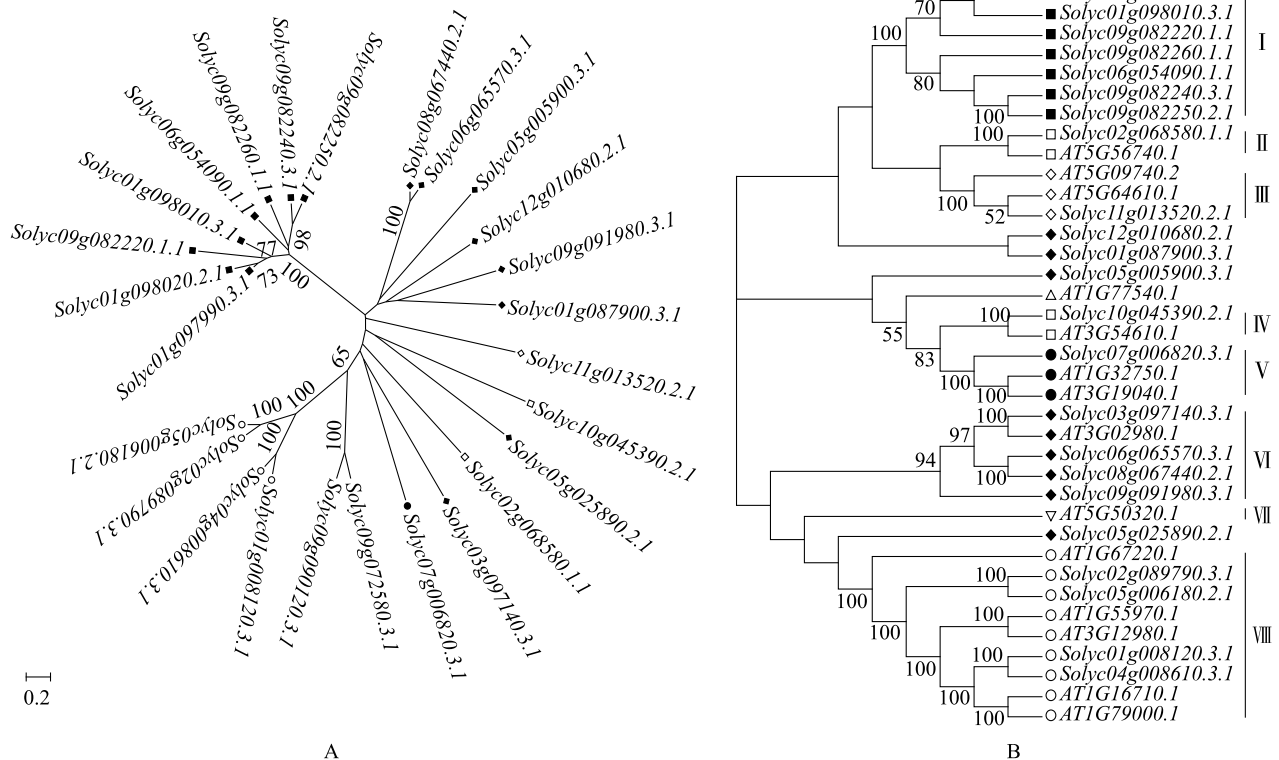
2 结果与分析

2.1 番茄组蛋白乙酰转移酶(*HAT*)鉴定及系统聚类

通过 Blastp 和关键词检索,共鉴定出 26 个番茄组蛋白乙酰转移酶(*HAT*)。经多重序列比对后利用邻接法构建系统发生树(图 1)。38 个番茄和拟南芥 *HAT* 可聚为 8 组(I~VIII)。根据拟南芥 *HAT* 分类标准,第 I 组 8 个 *HAT* 为番茄所特有;第 II 组 1

个番茄 *HAT*,且与拟南芥 *AtHAG2* 同源;第 IV 组 1 个番茄 *HAT*,且与 *AtHAG1* 同源;第 VI 组 4 个番茄 *HAT*,且与拟南芥 *AtMCC1* 同源,另有 4 个 *AtMCC1* 同源 *HAT* 散落于树中,这 8 个 *HAT* 属于 *GNAT* 家族;第 III 组 1 个番茄 *HAT*,且与 *AtHAM* 同源,属于 *MYST* 家族;第 V 组 1 个番茄 *HAT*,且与 *AtHAF* 同源,属于 *TAFII250* 家族;第 VIII 组 4 个 *HAT*,且与 *AtHAC* 同源,属于 *p300/CBP* 家族。整体来看,*GNAT* 家族的 18 个 *HAT* 大部分(77.78%)聚在第 I、II、IV 和 VI 组,但仍有少数 *HAT* 散落于其他组。同时,番茄中未发现与拟南芥 *AtHAG3* 同源的基因(VII)。

结合序列相似度比较,发现 6 对(13 个)同源基因(*Solyc01g008120.3.1* 和 *Solyc04g008610.3.1*, *Solyc02g089790.3.1* 和 *Solyc05g006180.2.1*, *Solyc06g065570.3.1* 和 *Solyc08g067440.2.1*, *Solyc09g082240.3.1* 和 *Solyc09g082250.2.1*, *Solyc09g072580.3.1* 和 *Solyc09g090120.3.1*, *Solyc01g097990.3.1*、*Solyc01g098010.3.1* 和 *Solyc01g098020.2.1*),且对应基因间 Ka/Ks 值均小于 1,说明其在自然进化过程中经历了纯化选择。



A: 26 个番茄 *HAT* 的系统聚类; B: 38 个番茄和拟南芥 *HAT* 的系统聚类。

图 1 番茄和拟南芥组蛋白乙酰转移酶(*HAT*)系统发生树

Fig.1 The phylogenetic tree of histone acetyltransferases (*HAT*) in tomato and *Arabidopsis thaliana*

2.2 番茄组蛋白乙酰转移酶(HAT)的理化特征分析

从表1可见,26个番茄HAT氨基酸数为80~1 856,相对分子质量为9 200~210 630,其中TAFII250和p300/CBP家族成员所含氨基酸较多、相对分子质量较大;理论等电点为4.62~9.58,平均7.40,整体接近中性;TAFII250和p300/CBP为不稳定蛋白质,而多数GNAT和MYST家族成员为稳定蛋白质;所有鉴定成员均为亲水性蛋白质,主要定位于细胞核,其次定位于细胞质和叶绿体;蛋白质二级结构主要包括 α 螺旋和无规则卷曲。结合系统聚类结果(图1),

选取有代表性的8个基因(*G1 Solyc01g098020.2.1*, *G2 Solyc08g067440.2.1*, *G3 Solyc11g013520.2.1*, *G4 Solyc10g045390.2.1*, *G5 Solyc02g068580.1.1*, *G6 Solyc07g006820.3.1*, *G7 Solyc09g072580.3.1*和*G8 Solyc05g006180.2.1*)编码的蛋白质进行蛋白质三维建模(图2),发现GNAT(*G1*、*G2*、*G4*、*G5*)和MYST(*G3*)家族除了 α 螺旋,还存在 β 折叠,但TAFII250(*G6*)和p300/CBP(*G8*)家族无 β 折叠。同时,发现部分GNAT和MYST家族成员具有活性位点。

表1 番茄组蛋白乙酰转移酶(HAT)结构与理化性质

Table 1 The structure and physicochemical properties of histone acetyltransferases (HAT) in tomato

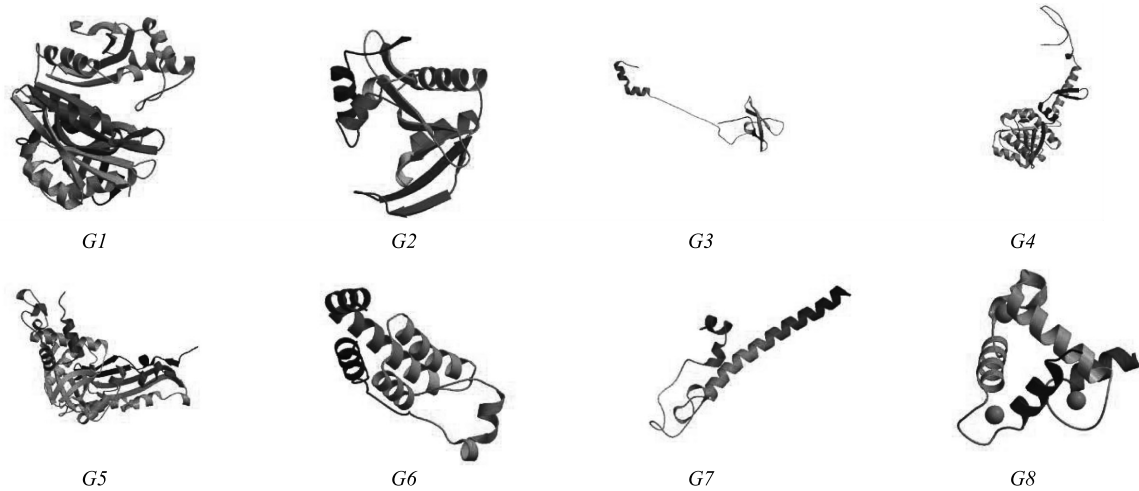
HAT 基因	氨基酸数量	相对分子质量	理论等电点	不稳定指数	总平均亲水性	亚细胞定位	蛋白质二级结构占比(%)				家族
							α 螺旋	延伸链	β 转角	无规则卷曲	
<i>Solyc01g097990.3.1</i>	80	9 200	8.24	32.09	-0.100	Cytoplasm	56.25	12.50	10.00	21.25	GNAT
<i>Solyc01g098010.3.1</i>	221	25 320	6.92	34.79	-0.234	Cytoplasm	51.58	15.84	6.33	26.24	
<i>Solyc01g098020.2.1</i>	166	18 750	6.08	32.46	-0.183	Nucleus	51.20	14.46	5.42	28.92	
<i>Solyc06g054090.1.1</i>	193	22 170	5.34	27.86	-0.272	Cytoplasm	43.52	12.95	5.70	37.82	
<i>Solyc09g082220.1.1</i>	105	11 990	9.58	48.31	-0.050	Chloroplast	31.43	25.71	6.67	36.19	
<i>Solyc09g082240.3.1</i>	207	23 550	4.78	26.87	-0.331	Cytoskeleton	53.14	11.59	3.86	31.40	
<i>Solyc09g082250.2.1</i>	419	47 530	8.89	30.21	-0.118	Cytoskeleton	37.95	20.29	6.92	34.84	
<i>Solyc09g082260.1.1</i>	302	34 490	4.62	34.29	-0.486	Nucleus	32.78	17.55	9.27	40.40	
<i>Solyc01g087900.3.1</i>	189	21 480	7.72	30.86	-0.459	Chloroplast	39.15	16.93	9.52	34.39	GNAT/MCC1
<i>Solyc05g005900.3.1</i>	311	3 470	7.15	57.97	-0.328	Chloroplast	37.30	17.36	3.54	41.80	
<i>Solyc06g065570.3.1</i>	165	18 560	8.96	15.69	-0.097	Cytoplasm	36.97	20.61	3.64	38.79	
<i>Solyc08g067440.2.1</i>	162	18 440	8.78	15.74	-0.182	Cytoplasm	39.51	19.75	5.56	35.19	
<i>Solyc09g091980.3.1</i>	205	23 430	6.39	27.24	-0.233	Cytoplasm	36.10	18.54	6.34	39.02	
<i>Solyc12g010680.2.1</i>	179	20 650	8.64	29.33	-0.355	Cytoplasm	58.66	15.08	7.82	18.44	
<i>Solyc03g097140.3.1</i>	254	29 300	9.30	46.54	-0.291	Nucleus	33.86	20.47	4.33	41.34	
<i>Solyc05g025890.2.1</i>	283	32 410	9.46	46.56	-0.525	Chloroplast	42.05	17.67	3.89	36.40	
<i>Solyc02g068580.1.1</i>	468	52 720	5.74	47.95	-0.258	Cytoplasm	39.53	19.02	4.06	37.39	GNAT/HAT1
<i>Solyc10g045390.2.1</i>	542	60 750	6.02	49.83	-0.607	Nucleus	37.08	10.33	6.46	46.13	
<i>Solyc11g013520.2.1</i>	447	51 540	6.74	35.25	-0.607	Nucleus	34.45	17.45	5.15	42.95	MYST
<i>Solyc07g006820.3.1</i>	1 856	210 630	5.47	50.21	-0.810	Nucleus	38.20	11.64	3.99	46.17	
<i>Solyc01g008120.3.1</i>	1 739	194 830	8.50	51.61	-0.619	Nucleus	31.05	10.18	3.39	55.38	p300/CBP
<i>Solyc02g089790.3.1</i>	1 475	166 770	6.27	55.50	-0.502	Nucleus	38.17	11.39	4.14	46.31	
<i>Solyc04g008610.3.1</i>	1 650	185 820	8.56	53.66	-0.662	Nucleus	30.36	10.12	3.58	55.94	Others
<i>Solyc05g006180.2.1</i>	738	85 660	8.71	50.34	-0.486	Nucleus	42.01	13.01	3.93	41.06	
<i>Solyc09g072580.3.1</i>	144	16 040	6.75	54.79	-0.935	Nucleus	37.50	9.03	7.64	45.83	
<i>Solyc09g090120.3.1</i>	157	17 520	8.86	37.33	-1.039	Nucleus	36.31	8.28	1.91	53.50	

基因结构分析发现,GNAT家族基因整体较小、外显子数较少,*Solyc10g045390.2.1*基因较大但氨基酸数较少,推测主要由非编码区组成;而MYST、

TAFII250和p300/CBP家族基因较大、外显子数较多。GNAT家族基因主要是1/2位内含子,MYST是0/1位内含子,TAFII250和p300/CBP是0/1/2位内

含子,同时,部分基因的部分外显子为非对称外显子(图3)。基序分析发现,20个蛋白质保守基序大小为21~50,不同家族 *HAT* 所含基序种类不同,如 *GNAT* 组 1 主要含 Motif 1、Motif 7、Motif 8 和 Motif 17, *GNAT* 组 2 主要含 Motif 14、Motif 18 和 Motif 19; *MYST* 和 *TAFII250* 家族与 *GNAT* 组 2 相似; *p300*/

CBP 家族主要含 Motif2~ Motif6、Motif 9~ Motif 13、Motif 15~ Motif 16 和 Motif 20,同一家族成员存在基序丢失情况。另外,不同家族 *HAT* 具有典型的特征结构域,如 *GNAT* 家族的 *Acetyltransf_1* 和 *Hat1_N*, *MYST* 家族的 *zf-MYST* 和 *MOZ_SAS*, *p300*/*CBP* 家族的 *HAT_KAT11* 和 *zf-TAZ*。



G1~G8 为 8 个代表性 *HAT*。

图 2 番茄组蛋白乙酰转移酶(*HAT*) 三级结构

Fig.2 Tertiary structure of histone acetyltransferases (*HAT*) in tomato

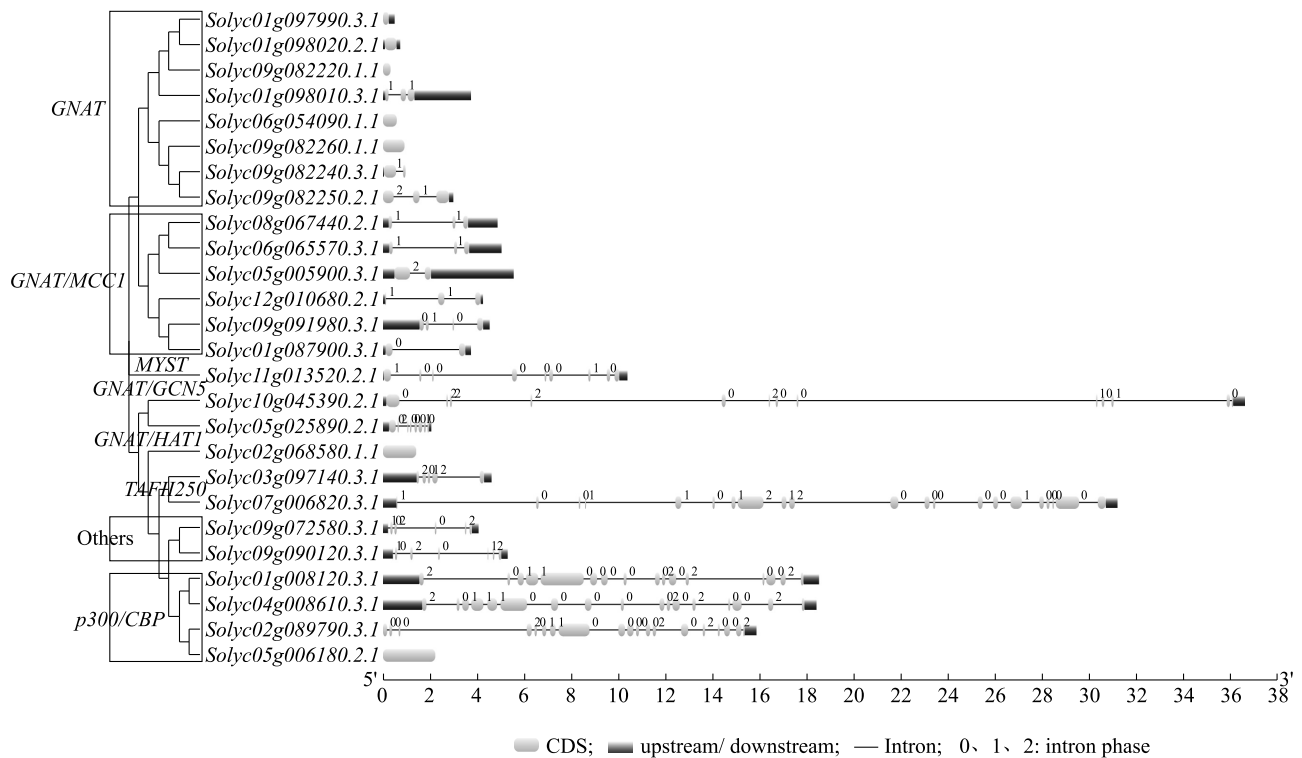


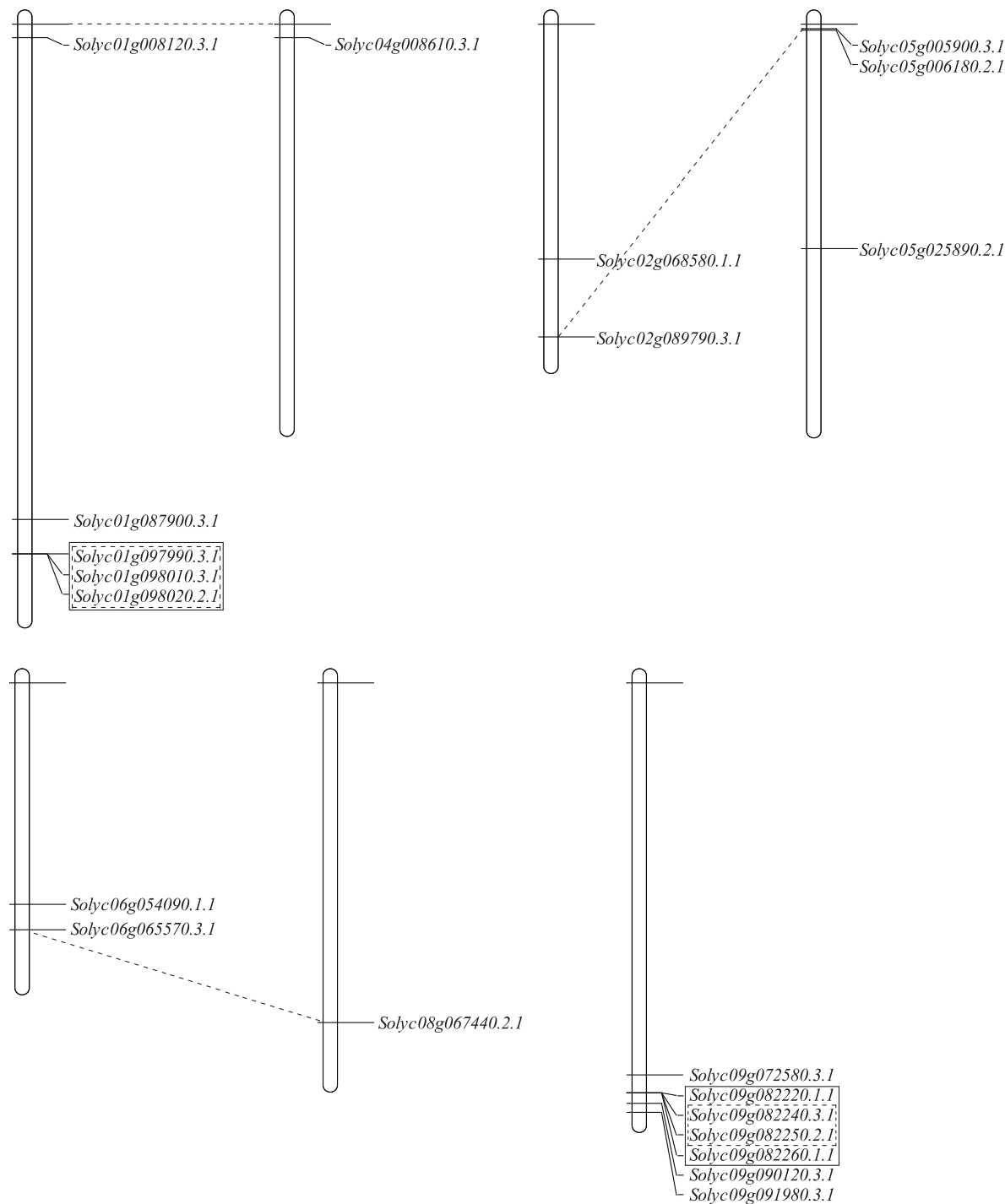
图 3 番茄组蛋白乙酰转移酶(*HAT*) 基因结构

Fig.3 The structure of histone acetyltransferases (*HAT*) gene in tomato

2.3 番茄组蛋白乙酰转移酶基因 (*HAT*) 染色体定位

26 个番茄组蛋白乙酰转移酶基因(*HAT*)不均匀地散布于番茄 12 条染色体上,且存在于两端,其中 9 号染色体分布有 7 个 *HAT*,1 号染色体为 5 个 *HAT*,其

他染色体为 1~3 个 *HAT* (图 4)。其中,9 号染色体和 1 号染色体分别有 4 个和 3 个基因形成单基因簇。聚类分析结合序列相似度比较获得的 8 对同源基因中,3 对位于不同染色体上,可能是经片段重复而来;4 对分别位于同一染色体基因簇内,为串联重复而来。



实线方框表示基因簇,虚线方框表示串联重复,虚线连接表示片段重复。

图 4 番茄部分组蛋白乙酰转移酶基因 (*HAT*) 的染色体定位

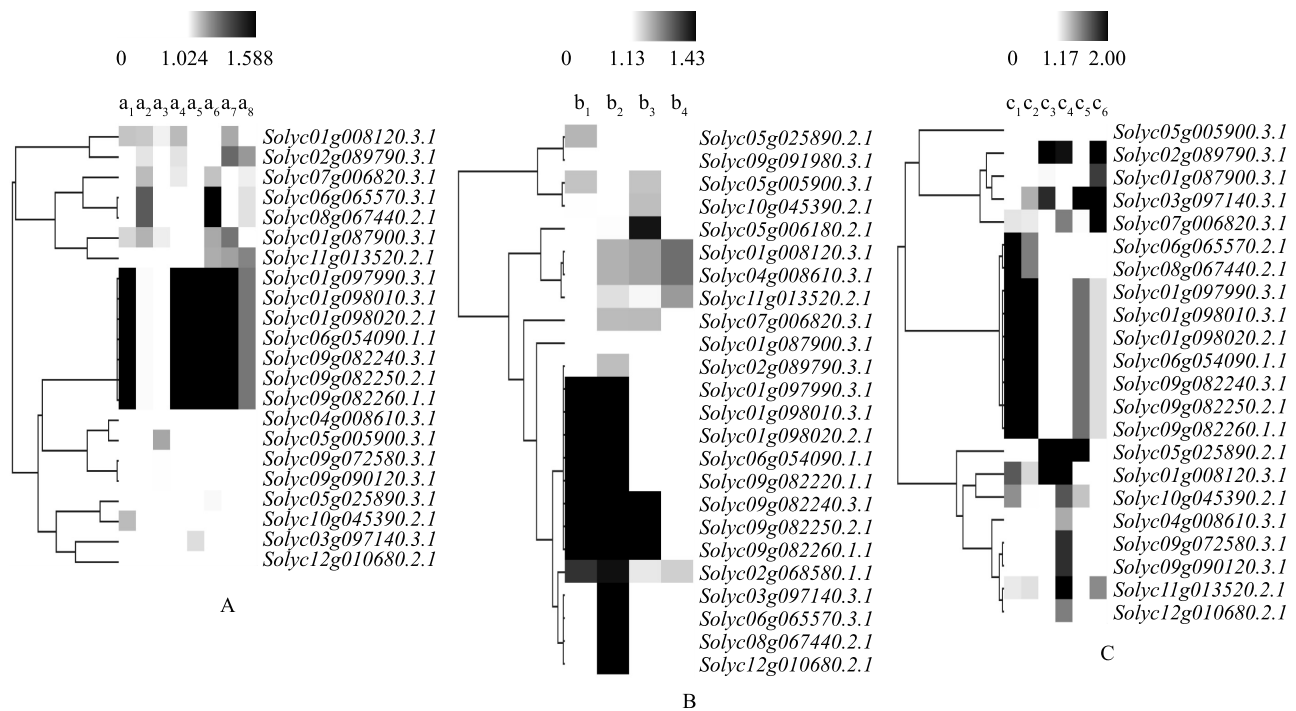
Fig.4 Chromosome localization of histone acetyltransferases (*HAT*) gene in tomato

2.4 番茄组蛋白乙酰转移酶基因(*HAT*)表达谱分析

番茄生物和非生物胁迫(盐、干旱和热)芯片数据表达谱见图5(部分基因未筛选到相应探针)。由图5A和5B可知,*GNAT*家族番茄特有*HAT*(I)和*AtHAG2*同源基因*Solyc02g068580.1.1*(II)在所有病菌侵染下,都具有较高的表达水平,部分*AtMCC1*同源基因在灰霉菌和番茄斑萎病毒(TSWV)侵染下具有较高的表达水平,*p300/CBP*家族番茄*HAT*(VIII)在TSWV、丛枝菌根真菌球霉菌(MYC)和马铃薯纺锤块茎类病毒(PSTVd)侵染下具有较高的表达水平,*MYST*家族的*Solyc11g013520.2.1*(III)在灰霉菌和PSTVd侵染下具有较高的表达水平,*TAFII250*家族的*Solyc07g006820.3.1*(V)在TSWV和MYC侵染下具有较高的表达水平,其他*AtMCC1*同源基因和*AtHAG1*同源基因*Solyc10g045390.2.1*(IV)表

达水平较低。

由图5C可知,*GNAT*家族番茄特有*HAT*(I)在盐和热胁迫下均有较高的表达水平,但耐盐(热)和盐(热)敏材料区分不明显;*p300/CBP*家族的*Solyc01g008120.3.1*、*Solyc02g089790.3.1*(VIII)和*AtMCC1*同源基因*Solyc05g025890.2.1*在干旱胁迫下具有较高的表达水平,同样与对照材料区分不明显,但后两者能够区分耐热和热敏材料;*AtMCC1*同源基因*Solyc03g097140.3.1*(VI)在热胁迫下具有较高的表达水平,与对照材料区分不明显,但可以区分耐盐/干旱和不耐盐/干旱材料;除*AtMCC1*同源基因*Solyc06g065570.3.1*、*Solyc08g067440.2.1*(VI,盐胁迫下表达水平较高)和*Solyc05g005900.3*(表达水平较低)外,其他基因在至少一种非生物胁迫下与对照材料间表达差异明显。



A和B:番茄*HAT*在6种生物胁迫下的基因表达;C:番茄*HAT*在3种非生物胁迫下的基因表达。*a*₁:真菌激发子诱导4 h;*a*₂:真菌激发子诱导8 h;*a*₃:番茄细菌性溃疡病菌接种4 d;*a*₄:番茄细菌性溃疡病菌接种8 d;*a*₅:接种灰霉菌的成熟绿果;*a*₆:接种灰霉菌的成熟红果;*a*₇:接种马铃薯纺锤块茎类病毒的栽培种Rutgers;*a*₈:接种马铃薯纺锤块茎类病毒的栽培种Money maker。*b*₁:番茄斑萎病毒侵染51 d的根;*b*₂:番茄斑萎病毒侵染51 d的叶;*b*₃:丛枝菌根真菌球霉菌处理51 d的根;*b*₄:丛枝菌根真菌球霉菌处理51 d的叶。*c*₁:盐胁迫下的栽培种Money maker;*c*₂:盐胁迫下的野生醋栗番茄PI365967;*c*₃:干旱胁迫下的敏感材料;*c*₄:干旱胁迫下的耐旱材料;*c*₅:热胁迫下的敏感材料;*c*₆:热胁迫的耐热材料。

图5 番茄组蛋白乙酰转移酶基因(*HAT*)表达谱

Fig.5 Expression profile of histone acetyltransferases (*HAT*) gene in tomato

3 讨论

植物表观遗传学是当今研究热点之一,而组蛋白修饰调控是表观遗传学的重要研究领域。组蛋白乙酰化作为组蛋白修饰的一种,由组蛋白乙酰化和去乙酰化共同完成^[16]。在真核生物中,组蛋白乙酰化是染色质中的可逆生物学过程,是促进染色质松弛和基因转录调控的主要因素之一^[17]。

3.1 番茄组蛋白乙酰转移酶(*HAT*)鉴定与聚类

本研究结合 Blastp 和关键词检索,共鉴定出 26 个番茄组蛋白乙酰转移酶(*HAT*)。通过与拟南芥 *HAT* 共同构建系统发生树,可将番茄 *HAT* 分为 7 组,第 I 组为番茄所特有,第 II 组与拟南芥 *AtHAG2* 同源,第 IV 组与 *AtHAG1* 同源,第 VI 组与 *AtMCC1* 同源,这 4 组属于 *GNAT* 家族;第 III 组与 *AtHAM* 同源,属于 *MYST* 家族;第 V 组与 *AtHAF* 同源,属于 *TAF-II250* 家族;第 VIII 组与 *AtHAC* 同源,属于 *p300/CBP* 家族。*GNAT* 家族扩张程度较大,散布于多个组中,但番茄中未发现 *AtHAG3* 同源基因。所建发生树包含基因染色体来源不同的混合进化枝,反映出其在染色体上的扩张情况,可能是产生具有新功能基因的源泉^[18]。同一进化枝包含不同物种基因,说明其可能按照物种特异性的方式进行了扩张。A 型 *HAT* 仅存在于核内,并与基因转录密切相关;B 型 *HAT* 通常只能使新合成的 H3 和 H4 组蛋白 N 端尾部特定赖氨酸残基乙酰化,并影响其在核小体组装中的定位^[1]。本研究发现多数番茄 *HAT* 为 A 型,但也有 B 型存在,如第 II 组的 *AtHAG2* 同源基因 *Solyc02g068580.1.1*。

3.2 番茄组蛋白乙酰转移酶(*HAT*)的理化特征

26 个番茄 *HAT* 中,8 个 *GNAT* 家族 *HAT* 为番茄所特有,其他 3 个家族 *HAT* 均比拟南芥要少^[5],也比 Cigliano 等^[11]发现的番茄 *HAT* 要少,可能与物种本身、所用数据库和鉴定标准有关,今后需进一步完善数据库和标准,使鉴定种类和数量更科学。这 26 个 *HAT* 基因大小及其编码的蛋白质稳定性等存在明显差异,且具有家族特异性,但均为亲水蛋白质,整体接近中性,主要定位于细胞核,包括 α 螺旋和无规则卷曲,*GNAT*/*MYST* 家族成员还存在 β 折叠。基因结构分析发现不同家族基因内含子相位不同,可能发生了内含子单独获取或丢失^[19],且部分基因的部分外显子为非对称外显子,容易造成阅读框的

推移,可能是造成基因多样性的原因之一。不同家族蛋白质具有典型的保守基序和特征结构域,是彼此区分的重要依据。

3.3 番茄组蛋白乙酰转移酶基因(*HAT*)染色体定位

26 个番茄组蛋白乙酰转移酶基因(*HAT*)不均匀地散布于番茄 12 条染色体上,且主要(46.15%)位于 1 号和 9 号染色体上,呈两端分布,这种基因在进化过程中可能更易变异。基因簇和串联重复分析发现,1 号和 9 号染色体上各有一个基因簇,簇内分别有 3 个和 2 个基因形成串联重复,说明串联重复是基因扩张和形成基因簇的重要机制^[20],在其他家族基因中同样如此^[19,21]。基因同源性分析发现,有 3 对基因来自不同染色体,可能经片段重复而来,其中 1 对基因在染色体上的排列顺序相反,暗示这些区段可能发生了倒位。同时,同源基因间 *Ka/Ks* 均小于 1,说明其在自然进化过程中经历了纯化选择。

3.4 番茄组蛋白乙酰转移酶基因(*HAT*)表达

通过番茄生物和非生物胁迫(盐、干旱和热)芯片表达数据分析,发现 *GNAT* 家族番茄特有 *HAT* 和 *AtHAG2* 同源基因 *Solyc02g068580.1.1* 在所有病菌侵染下,都具有较高的表达水平,部分 *AtMCC1* 同源基因和 *AtHAG1* 同源基因 *Solyc10g045390.2.1* 表达水平较低,其他家族基因具有病菌诱导特异性。*GNAT* 家族番茄特有 *HAT* 在盐和热胁迫下均有较高的表达水平,耐盐(热)和盐(热)敏材料区分不明显,但仍有超过 50% 的基因在至少一种非生物胁迫下与对照材料间存在差异表达。同样,*GNAT* 家族基因、尤其是 *AtMCC1* 同源基因具有表达多样性,当然与其数量较多有关,也从侧面反映出其家族特性。基因高表达虽能响应不同生物与非生物胁迫,但同时增加了植物负担,植物通过假基因化、丢失和相关表达调控机制进行自我保护,维持生命活动物质和能量代谢的平衡。

4 结论

本研究运用生物信息学方法,共鉴定出 26 个番茄组蛋白乙酰转移酶(*HAT*),根据序列特征可以分为 7 组,分别为 *HAG*、*HAG1*、*HAG2*、*MCC1*、*HAM*、*HAF* 和 *HAC*。其中,*GNAT* 家族(*GCN5*、*HAT1*、*MCC1*) 18 个,*p300/CBP* 家族 4 个,*MYST* 和 *TAFII250* 家族各 1 个。同时,发现 8 对同源基因且经历了纯化选择。

这些 *HAT* 均为亲水蛋白质,整体接近中性,主要定位于细胞核,包括 α 螺旋和无规则卷曲。不同家族酶具有典型的保守基序和特征结构域。染色体定位发现,番茄 *HAT* 不均匀地散布于 12 条染色体上,且呈两端分布,部分基因经串联重复形成基因簇。*GNAT* 家族番茄特有 *HAT* 和 *HAG2* 组基因在病菌、盐和热胁迫下具有较高的表达水平,一些基因具有病菌诱导特异性和耐盐(热)、盐(热)敏材料表达差异性。

参考文献:

- [1] MIZZEN C A, ALLIS C D. Linking histone acetylation to transcriptional regulation [J]. *Cell Metabolism Life Science*, 1998, 54: 6-20.
- [2] STERNER D E, BERGER S L. Acetylation of histones and transcription-related factors [J]. *Microbiology and Molecular Biology Reviews*, 2000, 64(2): 435-459.
- [3] MIZZEN C A, YANG X J, KOKUBO T, et al. The *TAFII250* subunit of *TFIID* has histone acetyltransferase activity [J]. *Cell*, 1996, 87: 1261-1270.
- [4] BANNISTER A J, KOUZARIDES T. The *CBP* co-activator is a histone acetyltransferase [J]. *Nature*, 1996, 384: 641-643.
- [5] PANDEY R, MULLER A, NAPOLI C A, et al. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes [J]. *Nucleic Acids Research*, 2002, 30(23): 5036-5055.
- [6] LIU X, LUO M, ZHANG W, et al. Histone acetyltransferases in rice (*Oryza sativa* L.): phylogenetic analysis, subcellular localization and expression [J]. *BMC Plant Biology*, 2012, 12: 145.
- [7] AQUEA F, TIMMERMAN T, ARCE-JOHNSON P. Analysis of histone acetyltransferase and deacetylase families of *Vitis vinifera* [J]. *Plant Physiology and Biochemistry*, 2010, 48(2/3): 194-199.
- [8] KORNET N, SCHERES B. Members of the *GCV5* histone acetyltransferase complex regulate *PLETHORA*-mediated root stem cell niche maintenance and transit amplifying cell proliferation in *Arabidopsis* [J]. *The Plant Cell*, 2009, 21(4): 1070-1079.
- [9] LATRASSE D, BENHAMED M, HENRY Y, et al. The *MYST* histone acetyltransferases are essential for gametophyte development in *Arabidopsis* [J]. *BMC Plant Biology*, 2008, 8: 121.
- [10] HAN S K, SONG J D, NOH Y S, et al. Role of plant *CBP/p300*-like genes in the regulation of flowering time [J]. *The Plant Journal*, 2007, 49(1): 103-114.
- [11] CIGLIANO R A, SANSEVERINO W, CREMONA G, et al. Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles [J]. *BMC Genomics*, 2013, 14: 57.
- [12] SATO S, TABATA S, HIRAKAWA H, et al. The tomato genome sequence provides insights into fleshy fruit evolution [J]. *Nature*, 2012, 485(7400): 635-641.
- [13] CHEN C J, XIA R, CHEN H, et al. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface [J]. *BioRxiv*, 2018, 3. doi: <https://doi.org/10.1101/289660>.
- [14] HUANG S, GAO Y, LIU J, et al. Genome-wide analysis of *WRKY* transcription factors in *Solanum lycopersicum* [J]. *Molecular Genetics and Genomics*, 2012, 287(6): 495-513.
- [15] ZHANG Z, LI J, ZHAO X Q, et al. KaKs_calculator: calculating *Ka* and *Ks* through model selection and model averaging [J]. *Genomics Proteomics Bioinformatics*, 2006, 4: 259-263.
- [16] 李涛, 苏慧慧, 李植良, 等. 番茄 *HDACs* 家族基因在胁迫条件下的表达分析 [J]. *热带作物学报*, 2015, 36(11): 1994-2001.
- [17] WATERBORG J H. Plant histone acetylation: in the beginning [J]. *Biochimica et Biophysica Acta*, 2011, 1809: 353-359.
- [18] 刘云飞, 万红建, 韦艳萍, 等. 番茄 *NBS-LRR* 抗病基因家族全基因组分析 [J]. *核农学报*, 2014, 28(5): 790-799.
- [19] 刘潮, 韩利红, 宋培兵, 等. 辣椒类甜蛋白基因家族鉴定及表达分析 [J]. *江苏农业学报*, 2018, 34(1): 122-129.
- [20] YANG Z F, ZHOU Y, WANG X F, et al. Genome-wide comparative phylogenetic and molecular evolutionary analysis of tubby-like protein family in *Arabidopsis*, rice, and poplar [J]. *Genomics*, 2008, 92(4): 246-253.
- [21] 向贵生, 王开锦, 晏慧君, 等. 蔷薇科植物 *MLO* 蛋白家族的生物信息学分析 [J]. *基因组学与应用生物学*, 2018, 37(5): 2043-2059.

(责任编辑:陈海霞)