



王盼乔,虞夏清,翟于菲,等. 甜瓜属异源多倍体不同倍性材料内参基因的筛选及评估[J]. 南京农业大学学报,2024,47(2):205-212.
WANG Panqiao, YU Xiaqing, ZHAI Yufei, et al. Reference gene selection and evaluation for *Cucumis* allopolyploids with different ploidy levels[J].
Journal of Nanjing Agricultural University, 2024, 47(2): 205-212.

甜瓜属异源多倍体不同倍性材料内参基因的筛选及评估

王盼乔^{1,2}, 虞夏清¹, 翟于菲¹, 赵勤政¹, 孟雅¹, 朱早兵¹, 李季¹, 陈劲枫^{1*}

(1.南京农业大学园艺学院/作物遗传与种质创新利用全国重点实验室,江苏 南京 210095;
2.河南农业大学园艺学院,河南 郑州 450002)

摘要: [目的]多倍化是植物界普遍存在的现象,但是内参基因的缺乏限制了多倍化相关基因表达研究的进程。本文旨在建立一组稳定的内参基因,以提高多倍体目标基因定量的准确性和重复性。[方法]以甜瓜属人工异源四倍体、其二倍体双亲和三倍体后代为材料,通过实时荧光定量反应(qPCR)比较了10个候选内参基因(*UBI-ep*、*ACT*、*ACT3*、*TUA*、*EF-1 α* 、*CACS*、*TIP41*、*F-box*、*CYP*和*UBQ*)的表达丰度,并利用geNorm和NormFinder软件对其表达稳定性进行分析。同时,通过转录组测序(RNA-seq)的方法对候选内参基因进行定量,并统计与荧光定量数据的相关性。[结果]荧光定量结果表明,二倍体黄瓜的*CYP*表达丰度最高, C_T 值为15.8;四倍体材料的*F-box*表达量最低, C_T 值为30.5。结合geNorm和NormFinder软件结果,一共筛选出4个稳定的参考基因*F-box*、*TIP41*、*ACT3*和*ACT*。其中*F-box*和*TIP41*在多倍性水平下稳定性值*M*比*ACT3*和*ACT*基因的小,但表达丰度低。以*ACT3*为内参时,qPCR与RNA-seq定量结果极显著相关($R^2=0.844, P<0.01$),*F-box*为内参时相关性最小。[结论]当比较不同倍性水平或跨物种的转录本丰度时,需要注意内参基因的选择。针对甜瓜属多种倍性材料的低丰度表达基因,可以选择*TIP41*作为内参基因;对于高表达基因,可采用*ACT3*作为内参基因。

关键词:甜瓜属;多倍体;内参基因;实时荧光定量PCR;稳定性

中图分类号:S652

文献标志码:A

文章编号:1000-2030(2024)02-0205-08

Reference gene selection and evaluation for *Cucumis* allopolyploids with different ploidy levels

WANG Panqiao^{1,2}, YU Xiaqing¹, ZHAI Yufei¹, ZHAO Qinzhen¹, MENG Ya¹,
ZHU Zaobing¹, LI Ji¹, CHEN Jinfeng^{1*}

(1.College of Horticulture/National Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization,
Nanjing Agricultural University, Nanjing 210095, China;
2.College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China)

Abstract: [Objectives] Polyploidy is widespread in plants kingdom. However, the lack of reference genes has severely limited the research progress of polyploidy-related gene expression. The present study aimed to improve the accuracy and reproducibility of gene expression analyses in polyploids by establishing a set of stable reference genes. [Methods] The abundance of 10 candidate reference genes (*UBI-ep*, *ACT*, *ACT3*, *TUA*, *EF-1 α* , *CACS*, *TIP41*, *F-box*, *CYP* and *UBQ*) were compared by real-time fluorescence quantitative analysis (qPCR) at different ploidy levels of *Cucumis* including artificial allotetraploid, their diploid parents and triploid offsprings. The stability of these candidate reference genes expression was analyzed by geNorm and NormFinder software. Furthermore, the correlation of relative transcript abundance estimated by qPCR analysis and transcriptome sequencing (RNA-seq) data was calculated to evaluate the stability of candidate reference genes. [Results] The qPCR results showed that the reference gene *CYP* in cucumber had the highest expression with a C_T value of 15.8, and the reference gene *F-box* in tetraploid material had the lowest expression with a C_T value of 30.5. Four stable reference genes were selected by combining the results of geNorm and NormFinder, which were *F-box*, *TIP41*, *ACT3* and *ACT*. Among them, *F-box* and *TIP41* had a lower stability value *M* than *ACT3* and *ACT* genes at the polyploid level, but their expression abundance were low. When *ACT3* was used as the internal reference, the relative transcript abundance correlation coefficient between qPCR and RNA-seq was extremely significant ($R^2=0.844, P<0.01$). But the correlation was minimum when *F-box* was an internal parameter. [Conclusions] Attention should be paid to the selection of reference genes when

收稿日期:2023-01-01

基金项目:国家自然科学基金项目(32102388, 31430075);江苏省自然科学基金项目(BK2022148);河南农业大学博士启动基金(30501121)

*通信作者:陈劲枫,教授,博导,主要从事黄瓜遗传育种与分子生物学研究, E-mail: jfchen@njau.edu.cn.

comparing transcript abundance across different ploidy levels. *TIP41* could be selected as the internal reference gene for the low abundance expression genes of multiple ploidy materials in melon. For highly expressed genes, *ACT3* could be used as an internal reference gene.

Keywords: *Cucumis*; polyploidy; reference gene; quantitative real-time PCR; stability

随着大量植物基因组数据的公布和基因功能注释的完善,目前植物相关研究无论是从解释生理现象出发,还是针对目标性状的特殊试验处理或遗传改造,都会考虑对相关基因进行转录丰度的检测^[1]。实时荧光定量技术 PCR(简称 qPCR)具有低成本、高灵敏度和灵活性以及可扩展性等优势,已经成为检测基因丰度最普遍的方法。利用此方法检测基因表达的准确性在很大程度上取决于用于转录本丰度标准化的内参基因是否合适。理想的参考基因必须在任何时空和不同处理条件下都能保持表达水平稳定一致^[2]。然而,目前许多已被用作内参的基因其稳定性明显无法满足 qPCR 要求^[3]。因此,在通过 qPCR 确定目的基因的表达模式之前,有必要筛选出符合相关试验条件下稳定表达的合适参考基因。目前有多个内参评价算法被开发,其中 geNorm^[4]和 NormFinder^[5]已被应用于许多特定试验条件中来筛选最佳内参基因^[6-9]。

多倍体在被子植物中非常普遍,几乎所有的被子植物基因组都经历过多倍化事件^[10]。1 亿 700 万年到 1 亿 1 800 万年前,葫芦科植物祖先经历了一次全基因组复制(whole genome duplications, WGD),这个古四倍体经过长时间分化最终形成西瓜、甜瓜和黄瓜等重要水果蔬菜^[11]。本试验前期利用普通栽培黄瓜($2n=2x=14$)花粉对甜瓜属野生种酸黄瓜($2n=2x=24$)进行人工远缘杂交,经过胚胎拯救获得种间 F_1 杂种。进而通过体细胞无性系变异加倍杂种 F_1 ,获得人工异源四倍体(S0)^[12],连续自交至第 14 代(S14),异源四倍体的花粉育性逐年提高。利用该异源四倍体与不同基因型黄瓜进行杂交,获得了不同基因型的正反交甜瓜属异源三倍体材料(Ha×N 和 S13×CC3 等)^[13-14],极大丰富了葫芦科物种遗传多样性。研究表明,多倍体形成通常伴随着基因组结构、基因表达和表观遗传的大规模变化,建立一组稳定的内参基因是进行植物多倍化基因丰度分析和高通量表达谱验证的必要前提^[15]。尽管目前广泛使用的一些参考基因已经证明在物种内的不同发育和环境条件下相对稳定,但是很少有研究能关注到它们是否也可以用于物种不同倍性水平间比较^[16]。目前,在甜瓜属内进行系统比较鉴定不同倍性的合适内参基因尚未有报道。因此,本试验利用 geNorm^[4]和 NormFinder^[5]分析 10 个候选内参基因在甜瓜属二倍体、异源三倍体和异源四倍体材料叶片的实时荧光定量(qPCR)变异程度,明确候选内参基因的表达稳定性,为物种不同倍性间转录组研究和目标基因定量分析中内参基因的选择提供借鉴与参考。

1 材料与方法

1.1 植物材料和生长条件

二倍体材料甜瓜属近缘野生种酸黄瓜(*Cucumis hystris*, HH, $2n=2x=24$)和中国北方栽培黄瓜‘北京截头’(*Cucumis sativus*, CC3, $2n=2x=14$),异源四倍体材料(*Cucumis hytivus*, S14, $2n=4x=38$)以及三倍体甜瓜属反交材料 Ha×N 和正交材料 S13×CC3($2n=3x=26$),均由南京农业大学葫芦科作物遗传与种质创新实验室收集保存。2017 年秋季,在南京农业大学江浦农场塑料大棚种植植物材料,设置 3 个生物学重复,自然光照条件,气温 20~30 °C,水肥管理按田间生产进行。

1.2 RNA 提取和 cDNA 合成

在植株长至 11 片真叶的成株期,于晴天 08:00 采集第 4 片真叶(自顶端向下数),将剪下的叶片用锡纸包裹后在液氮中快速冷冻,并在 -80 °C 储存备用。使用 TRIzol 试剂(Invitrogen, USA)提取总 RNA。将总 RNA 用不含 RNA 酶(RNase-free)的 DNA 酶 I (TaKaRa)在 37 °C 处理 30 min,去除基因组 DNA 污染,并采用反转录试剂盒(PrimeScript™ RT reagent Kit with gDNA Eraser, TaKaRa)进行逆转录。通过琼脂糖凝胶电泳评估 RNA 的完整性,并使用 NanoDrop ND-1000 分光光度计(NanoDrop Technologies, USA)测定每份样品的浓度。选择 A_{260}/A_{280} 值为 1.9~2.0 和 A_{260}/A_{230} 值为 2.0 的样品用于进一步分析。将上述总 RNA 制备合成 cDNA。按照使用说明,用 ultraclean Oligo (dT) 和 SuperScript III Reverse Transcriptase (TaKaRa) 处理 1 mg 等份的总 RNA。为验证 RNA 中不存在基因组 DNA 污染以及引物的特异性扩增,以水为空白对照,总 RNA 与引物混合为阴性对照,以 cDNA 为模板对各对引物进行 PCR 扩增。PCR 程序:94 °C 预变性 4 min;94 °C 变性 40 s,55~60 °C 退火 40 s,72 °C 延伸 40 s,共循环 35 次;最后 72 °C 延伸 7 min;4 °C

10 min。产物进行 $6 \text{ g} \cdot \text{L}^{-1}$ 聚丙烯酰胺凝胶电泳 (PAGE) 和银染显带。

1.3 PCR 引物和 qPCR

基于对不同品种、生长阶段、器官组织以及处理条件下的黄瓜多个内参基因综合比较,挑选出 10 个候选内参基因(表 1)。引物设计条件:引物跨越内含子,无发夹和二聚体结构,退火温度(T_m)为 $55 \sim 61 \text{ }^\circ\text{C}$,长度为 $18 \sim 25 \text{ bp}$,GC 含量为 $40\% \sim 60\%$,PCR 产物预期长度适宜($80 \sim 300 \text{ bp}$)。

表 1 引物序列和扩增特征

Table 1 Primer sequences and amplicon characteristics

基因 Gene	登录号 Accession number	功能 Annotation	引物序列 Primer sequence(5'→3')	扩增长度/bp Amplification length	退火温度/ $^\circ\text{C}$ Annealing temperature	扩增效率 Amplification efficiency
<i>UBI-ep</i>	AY372537	泛素延伸蛋白 Ubiquitin extension protein	CACCAAGCCCAAGAAGATC/ TAAACCTAATCACCACCAGC	220	58	1.923
<i>ACT</i>	AB010922	肌动蛋白 Actin	CCGTTCTGTCCTCTACGCTAGTG/ GGAACCTGCTCTTTGCAGTCTCGAG	290	60	1.905
<i>ACT3</i>	DQ115883	肌动蛋白 3 Actin 3	GCCAGTGGTGGTGAACATG/ TTCTGGTGATGGTGTGAGTC	149	60	1.972
<i>TUA</i>	AJ715498	α -微管蛋白 α -tubulin	CATTCTCTTTGGAACACACTGA/ TCAAACCTGGCAGTTAAAGATGAAA	154	60	1.998
<i>EF-1α</i>	EF446145	延伸因子 EF-1 α Elongation factor 1-alpha	ACTTTATCAAGAACATGATTAC/ TTCCTTCACAATTTTCATCG	230	60	1.959
<i>CACS</i>	XM_004148438.2	网格蛋白衔接蛋白复合 体亚基 Clathrin adaptor complex subunit	GTGCTTTCTTTCTGGAATGC/ TGAACCTCGTCAAATTTACACA	158	60	1.943
<i>TIP41</i>	GW881871	TIP41 样蛋白 TIP41-like family protein	CAACAGGTGATATTGGATTATGATTATAC/ GCCAGCTCATCTCATATAAG	221	60	1.939
<i>F-box</i>	GW881870	F-box/kelch 重复蛋白 F-box/kelch-repeat protein	GGTTCATCTGGTGGTCT/ TCITTTAAACGAACGGTCACTCC	166	60	1.995
<i>CYP</i>	AY942800	肽基脯氨酰顺反异构酶 活性亲环蛋白 Peptidyl-prolyl cis-trans isomerase CYP19-1	GGAATGCTACAGGAGGTG/ CATAACCTCAACGACTTGAC	201	60	1.932
<i>UBQ</i>	XM_004135144.2	泛素连接酶 Ubiquitin-protein ligase	TAACATTTATGCCGATGGAAGC/ CGCACAGAAGCGACTGAATAGA	107	61	1.917

依据实时荧光定量 PCR 标准 (MIQE) 指南^[4-5],进行 qPCR 的设计和统计。每 $20 \mu\text{L}$ qPCR 反应体系中含有 $2 \mu\text{L}$ 稀释的 cDNA, $10 \mu\text{L}$ SYBR *Premix Ex Taq*TM II (TaKaRa), $0.6 \mu\text{L}$ 每种引物 ($10 \text{ mmol} \cdot \text{L}^{-1}$) 和 $6.8 \mu\text{L}$ ddH₂O。为了减少移液次数,先将模板和荧光染料以及双蒸水进行 30 份体系混合,按标准体积 (0.2 mL) 等分到 MicroAmpTM 光学 96 孔反应板 (Thermo) 中。为了降低污染风险,所有 PCR 操作均在避光的超净工作台中进行,并检查引物的 DNA 污染,以添加无模板的水作为对照。在定量仪器 Bio-Rad CFX96 中进行 qPCR 反应,为了降低 PCR 板不同区域孔的条件差异,3 次生物学重复在光学 PCR 板上采取不同的样品排列方式,基于相同引物的反应尽量放同一个板进行。qPCR 反应程序: $95 \text{ }^\circ\text{C}$ 预变性 2 min ; 然后进行 40 个循环, $95 \text{ }^\circ\text{C}$ 30 s , $58 \sim 61 \text{ }^\circ\text{C}$ 30 s , $72 \text{ }^\circ\text{C}$ 30 s ; 最后进行熔解曲线分析 $60 \sim 95 \text{ }^\circ\text{C}$, 以确定每对引物扩增特异性。每个基因型包含 3 个生物学重复,每份样品设置 3 个技术重复。

1.4 RNA-seq

为了比较 qPCR 和 RNA-seq 计算的相对转录物丰度,按照 Illumina (San Diego, CA, USA) HiSeqTM 2000 平台建库标准流程,简单流程:使用 TRIzol 试剂 (Invitrogen) 从二倍体材料 CC3、*C. hystrix* 和四倍体材料 S14 提取总 RNA,然后用无 RNA 酶的 DNA 酶 I (New England Biolabs) 处理以除去可能的基因组 DNA 污染;通过 Dynabeads oligo (dT) (Dyna, Invitrogen) 包被的珠子从总 RNA 中纯化 mRNA;使用反转录酶 (Superscript II, Invitrogen) 双链 cDNA 合成试剂盒 (Invitrogen, USA) 产生 cDNA 第 2 链,加入片段化缓冲液以产生 $100 \sim 400 \text{ nt}$ 的片段,通过 QIAquick PCR 提取试剂盒 (Qiagen, Germany) 纯化;通过掺入测序接头将产物相互连接,在琼脂糖凝胶电泳后,选择合适大小范围的片段。对所得文库进行双末端 (pair end) 测序,去掉引物和接头,过滤低质量 raw reads。使用 Hist2 软件^[17] 将过滤后的 clean reads 比对到 10 对基因序

列,利用 Samtools^[18] 和 Linux 脚本获取 reads depth,通过简化 RPKM^[19] (reads per Kilobase per million mapped reads,每百万数据量序列测序深度)的方法计算转录本丰度,公式如下:

$$RPKM = \frac{(L \times N) \times MD}{l \times n}; MD = \frac{TD}{l}$$

式中:MD 为序列平均测序深度;TD 是序列单碱基深度总和; $L = 700$ bp,标准化序列长度; $N = 50\ 000$,标准化的 mapped reads 数据量; l 为目标基因长度; n 为比对到所有参考序列的总 reads 数。

1.5 统计分析

利用 LinRegPCR 程序^[20] 计算每个引物对的阈值循环数 (cycle threshold, C_T) 和 PCR 效率和相关系数,利用 R 语言进行不同因素水平间 C_T 的方差分析 (ANOVA)。使用 geNorm 和 NormFinder 软件评估内参基因稳定性。使用 geNorm^[4] 计算每个基因的稳定性值 (M),以 M 值等于 0.7 时作为稳定转录的标准^[21]。NormFinder 是基于 ANOVA 模型直接测量变异程度,并相应地对候选基因进行排序^[5]。为了确定参考基因选择对标准化结果的影响,分别以 *ACT3*、*ACT*、*F-box* 和 *TIP41* 为内参,使用 $2^{-\Delta C_T}$ 方法^[22] 计算剩余 9 个候选基因的表达量,同时对 9 个基因的 RPKM 进行归一化。通过 R 程序计算经过 \log_2 标准化后 qPCR 和 RNA-seq 表达量的相关系数,并绘图。

2 结果与分析

2.1 甜瓜属叶片 RNA 检测与定量引物特征

采集甜瓜属二倍体、三倍体和四倍体植株茎尖第 4 片真叶 (图 1-A),分别提取总 RNA,琼脂糖凝胶检测结果如图 1-B。利用综合条件下黄瓜稳定表达基因来开发引物,10 个候选内参基因的引物序列和扩增特征信息见表 1。熔解曲线 (图 2-A) 和 PAGE (图 2-B) 分析结果显示,每个引物 PCR 产物长度与预计扩增片段大小一致,并且每对引物结合位点呈现特异性,以水和总 RNA 为模板均不能扩增出条带 (未展示)。使用 LinRegPCR 程序分析显示扩增效率为 1.905~1.998,并且相关系数 R^2 为 0.996~0.998。



图 1 甜瓜属不同倍性水平植株叶片及对应总 RNA 电泳检测

Fig. 1 The leaves of *Cucumis* in different ploidy levels and electrophoresis detection of the corresponding total RNA

A. 植株叶片;B. 总 RNA 电泳 (M. DNA 标准品)。1~5 分别为 CC3、HH、Ha×N、S13×CC3、S14。

A. The plant leaf; B. Total RNA electrophoresis (M. DNA marker). 1-5 are CC3, HH, Ha×N, S13×CC3, S14, respectively.

2.2 甜瓜属内参基因的转录丰度

荧光定量结果显示,每个候选基因的 C_T 值在 3 个生物重复间和 3 个技术重复中变化很小 ($F = 1.145$, P -value = 0.332 2),差异主要来源于候选基因和基因型 (倍性) 之间的差异 ($F = 103.675$, P -value = 0.000 1)。最低 C_T 值为 15.8 (CC3, *CYP*),最高为 30.5 (S14, $2n = 4x$, *F-box*)。在所有模板中, *CYP* 是转录水平最高的基因,其平均 C_T 值最低 (17.9),其次是 *UBI-ep* (18.3)、*UBQ* (19.5)、*ACT3* (19.9) 和 *EF-1 α* (22.2); *F-box* 的转录水平最低,其平均 C_T 值相应最高 (27.6),其次是 *TIP41* (26.0)、*CACS* (25.7)、*TUA* (23.1) 和 *ACT* (22.5) (括号中的数值为 C_T 值) (图 3)。10 个候选参考基因在 3 种倍性材料中没有均匀转录,因此有必要对不同倍性水平下标准化参考基因的稳定性进行评估。

2.3 甜瓜属内参基因的稳定性分析

基于 geNorm 分析结果,根据 M 值排名来确定参考基因的稳定性。如图 4 所示:从左到右 M 值越小说明越稳定,当 M 值小于 0.7 表明具有稳定性^[21]。在 10 个候选内参基因中, *F-box*、*TIP41*、*ACT3* 和 *ACT* 的 M 值均小于 0.7,可认为是稳定的,而 *TUA* 表现最不稳定。由 geNorm 计算的成对变化参数 (V) 预测所需参考基因的最佳数量,当 V_n/V_{n+1} 比值小于 0.15 时,表明不需要加入额外的参考基因^[23]。结果 (图 4) 表明,最佳基因数为 4,预计 *F-box*/*TIP41*/*ACT3*/*ACT* 组合可提供最可靠的标准化水平。利用 NormFinder 软件对

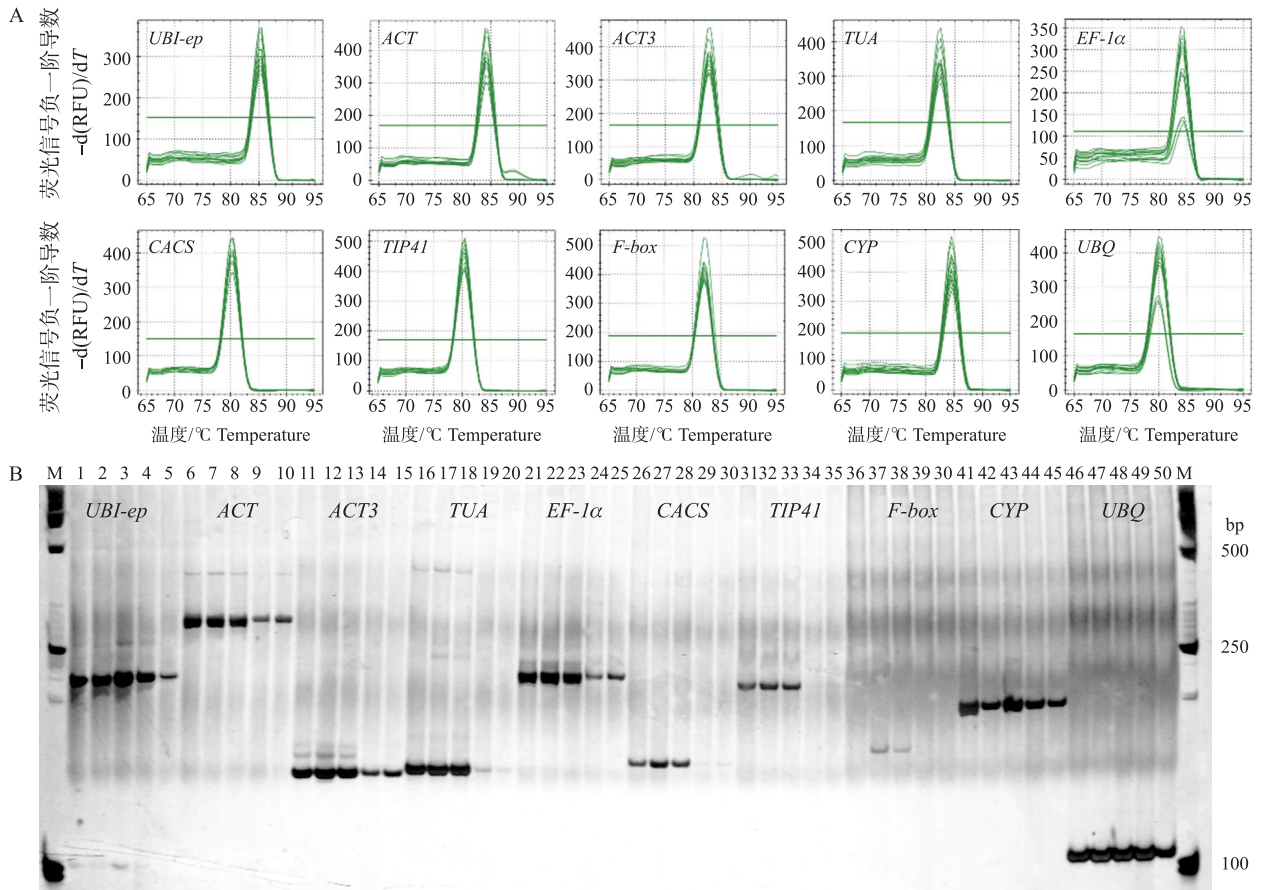


图 2 扩增溶解曲线和 PAGE 电泳分析

Fig. 2 Melting curves and PAGE analysis of PCR amplification

A. 10 个内参基因的解链曲线分析;B. PCR 扩增结果的 PAGE 分析。PAGE 分离泳道中从左到右排序分别为 5 组样品 CC3、HH、Ha×N、S13×CC3、S14 的 *UBI-ep*、*ACT*、*ACT3*、*TUA*、*EF-1α*、*CACS*、*TIP41*、*F-box*、*CYP* 和 *UBQ* 扩增条带。

A. Melting curve analysis of the ten reference genes;B. PAGE analysis of the amplicons. The samples in each PAGE separation are ordered from left to right as *UBI-ep*、*ACT*、*ACT3*、*TUA*、*EF-1α*、*CACS*、*TIP41*、*F-box*、*CYP* 和 *UBQ* in sample CC3, HH, Ha×N, S13×CC3, S14.

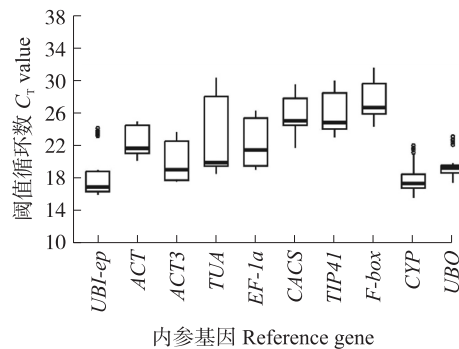


图 3 甜瓜属 15 份 cDNA 样品中 10 个内参基因 qPCR 的阈值循环数 (C_T)

Fig. 3 The cycle threshold (C_T) values for ten reference genes over 15 *Cucumis* samples in qPCR

内参基因 C_T 值由 LinRegPCR 分析,方框代表了第 25 和第 75 百分位之间的数据。框中的线代表中位数,框外线的最高、最低处分别代表最大值和最小值,散点代表异常值。

The raw expression data C_T values of all tested samples for each reference gene are analyzed by LinRegPCR. The box represents the 25th and 75th percentiles of data. A line across the box is depicted as the median. Whiskers represent the maximum and minimum values, the dots represent extreme outliers.

内参基因的稳定性 (S) 进行评价, S 值越小表示基因表达越稳定^[5]。NormFinder 分析(表 2) 得到与 geNorm 相似的排名,其中甜瓜属不同倍性材料间候选内参基因 *ACT* 的 S 值最小 ($S=0.120$), 表达稳定性最高,其次为 *F-box* ($S=0.131$) 和 *ACT3* ($S=0.148$), 而 *CYP* 的 S 值最大 ($S=0.391$), 表达稳定性最差。

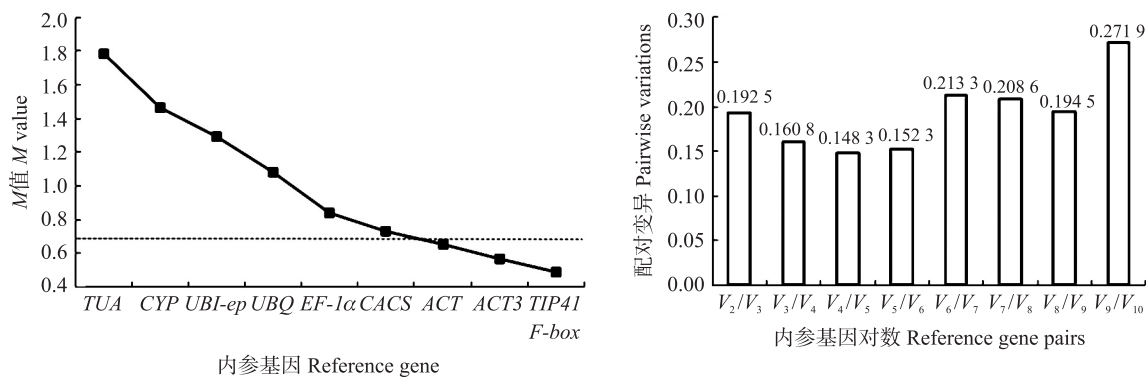


图 4 10 个内参基因的平均表达稳定性值 (M) 和标准化所需内参基因的最佳数目

Fig. 4 Average expression stability values (M) of the ten reference genes and the optimal number of reference genes required for normalization

在内参基因 n 和 $n+1$ 之间分析成对变异 (V_n/V_{n+1}) 以优化定量的标准化。

Pairwise variation (V_n/V_{n+1}) is analyzed between reference genes n and $n+1$ to optimize normalization in the qPCR.

表 2 NorFinder 分析不同倍性水平下 10 个候选内参基因的表达稳定性

Table 2 The expression stability of 10 candidate reference genes in different ploidy levels calculated by NormFinder

基因 Gene	排名 Rank	稳定值 (S) Stability value	基因 Gene	排名 Rank	稳定值 (S) Stability value
ACT	1	0.120	TUA	6	0.202
F-box	2	0.131	UBQ	7	0.225
ACT3	3	0.148	UBI-ep	8	0.268
TIP41	4	0.171	CACS	9	0.284
EF-1α	5	0.190	CYP	10	0.391

2.4 RNA-seq 与 qPCR 转录本定量相关性分析

通过 qPCR 分别以 *F-box*、*TIP41*、*ACT3* 和 *ACT* 计算的其他 9 个内参基因在黄瓜、酸黄瓜和异源四倍体中的转录丰度,与 RNA-seq 预测的相应基因表达量 (RPKM) 进行比较,定义相关系数 0.8~1.0 为极相关,0.6~0.8 为相关。以 *ACT3*、*TIP41*、*ACT* 为参考的 2 种定量方法估计的转录本丰度相关系数达到极相关,分别为 0.844 ($P=3.13E-08$)、0.805 ($P=4.18E-07$) 和 0.805 ($P=4.19E-07$)。以 *F-box* 为内参的 2 种定量结果也达到强相关 $R^2=0.749$ ($P=6.94E-06$)。图 5 为以 *ACT3*、*TIP41*、*ACT* 和 *F-box* 为内参时其他候选基因定量相关性比较结果。可见, C_T 值可以在一定程度上代表相对转录本丰度。

3 讨论

qPCR 已被广泛应用于试验处理或遗传转化后基因的表达分析,选择合适的内参基因是保证荧光定量准确性的前提,目前被选用的内参基因多为不容易受处理影响的看家基因 (house-keeping gene)^[21]。例如,与生物体基本生化代谢过程相关的基因,葡萄糖代谢基因 (*GAPDH*)、转录延伸因子基因 (*EF-1α*)、多聚泛素酶基因 (*UBQ*) 和亲环蛋白基因 (*CYP*);细胞骨架结构成分相关的肌动蛋白基因 (*ACTIN*、*ACT*) 和微管蛋白基因 (*α-tubulin*、*TUA* 和 *β-tubulin*、*TUB*);液泡膜内在蛋白 *TIP41* 基因等^[24-25]。内参基因稳定程度受各种试验条件的影响,例如试验材料组织器官、发育阶段、生物或非生物胁迫等。相同内参基因在不同物种间的同源表达也存在差异,不存在通用性^[26]。对拟南芥属不同倍性种质内参基因比较表明,*F-box* 基因在二倍体和四倍体之间的表达量有较大差异^[16]。而菊花不同倍性间的比较结果显示,*EF-1α* 表现最稳定,其次是 *ACT*,而 *TIP41* 稳定性表现较差^[27]。本研究在甜瓜属不同倍性材料间检测了 10 个普遍使用的参考基因的稳定性,尽管这些基因在二倍体材料整个发育过程中以及在一系列环境条件下呈现良好的稳定性^[7-9],但是在物种之间或不同倍性水平下展现出不同程度的转录变异,需要重新评价稳定性。本研究结果表明,经过 C_T 筛选, NormFinder 与 geNorm 的评价结果基本一致,有 4 个内参基因 *ACT3*、*ACT*、*F-box* 和 *TIP41* 表现较稳定。虽然在排序上稍有差异,推测可能是由于软件算法不同造成的^[4-5]。研究发现, *CYP* (17.9)、*UBI-ep* (18.3) 和 *UBQ* (19.5) 的 C_T 值较小,容易出现较高变异系数和异常值,因此在评价和筛选内参基因时,应该尽量选择表达丰度相似的候选基因。另外,进行 qPCR 试验时尽可能将低表达基因和高表

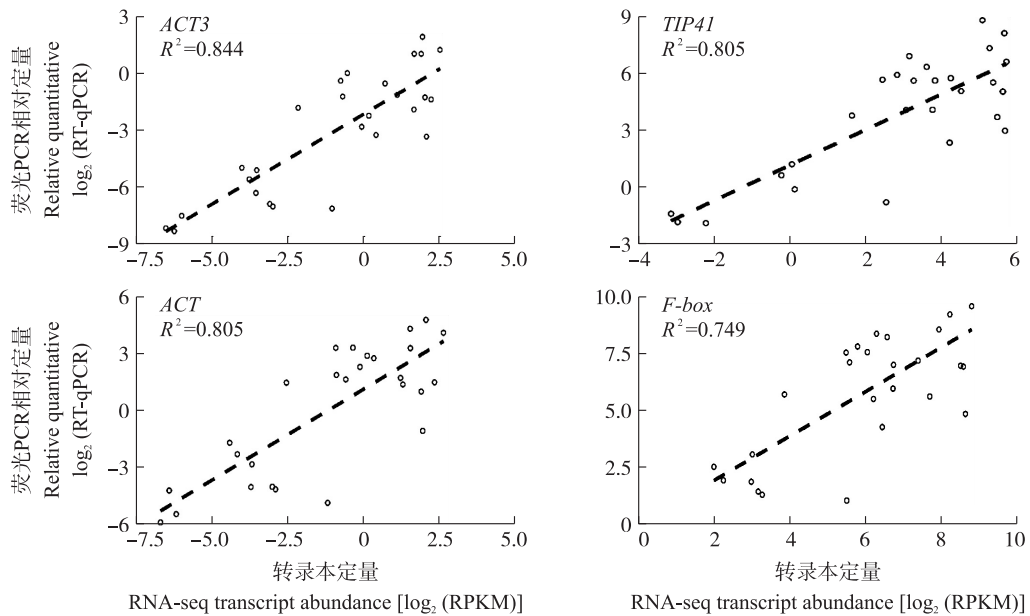


图 5 qPCR 和 RNA-seq 转录本定量相关性

Fig. 5 The correlation of transcript abundance as estimated from qPCR analysis and RNA-seq data

X 轴表示来自 RNA-seq 的相对转录本丰度, Y 轴表示从 qPCR 估计的相对转录本丰度。

X axis represents relative transcript abundance from RNA-seq, and Y axis represents the relative transcript abundance estimated from qPCR.

达基因进行分组检测, 分别选用表达丰度适宜的内参基因。

多倍化是被子植物进化的主要驱动因素, 基因组重组为自然选择提供了大量变异来源, 大多数基因的表达模式也因此发生变化^[15]。不同倍性物种之间的数字基因表达谱 (DGE) 和 RNA-seq 测序的试验已经提供了大量证据, 表明不同倍性水平和组织类型之间的转录水平存在差异^[28-29], 其中也包括部分内参基因。因此直接将常规内参基因应用于跨倍性水平比较时稳定性会受到影响。首先, 基因组加倍引起同源基因数目的增加可能影响内参基因表达的剂量效应 (加性或非加性的表达)^[30]。基因剂量效应常常与染色体数目相关, 而甜瓜属内染色体数目变异丰富, 黄瓜染色体基数为 7, 而酸黄瓜为 12; 人工创制的异源四倍体 ($2n=4x=38$) 和三倍体 ($2n=3x=26$) 后代染色体数目各不相同, 同样引起内参基因同源基因数目的变异。其次, 人工合成拟南芥异源四倍体早期世代中参与能量代谢, 细胞生物发生和植物激素调节的一些基因上调表达, 并导致表型和生态学改变^[30]。在植物生命基础过程中起重要作用的基因, 如蛋白降解相关基因 (*UBQ*)、细胞骨架组分 (*TUA*) 和正常的细胞代谢过程相关基因 (*EF-1a*、*CYP* 和 *UBI-ep*), 虽然在二倍体水平能够保持稳定表达, 但是后期受基因组加倍的影响可能导致其稳定性丧失。本研究结果表明, 10 个候选内参基因中只有 4 个 (*ACT3*、*ACT*、*F-box* 和 *TIP41*) 在不同倍性间表达比较稳定, 而剩余 6 个基因 (*TUA*、*CYP*、*UBI-ep*、*UBQ*、*EF-1 α* 和 *CACS*) 稳定性已经受到严重影响, 进一步验证了以上结论。然而, 内参基因在多倍化中转录调控复杂, 需要更多试验来进一步揭示其稳定表达机制。

综上所述, 在进行多倍化基因表达研究时, 根据具体试验选择合适的内参基因是十分必要的。本研究中候选内参基因的 2 个不同软件评价结果一致, 与转录组定量结果相符, 表明按照标准 qPCR 方法来筛选跨倍性水平的稳定内参基因非常有效。本试验在甜瓜属内筛选到 4 个在不同倍性水平间稳定表达的内参基因, 为进一步分析甜瓜属远缘杂交和多倍化发生过程中基因表达的变化及甜瓜属异源四倍体、三倍体相关分子生物学研究提供参考。

参考文献 References:

- [1] Gachon C, Mingam A, Charrier B. Real-time PCR: what relevance to plant studies? [J]. *Journal of Experimental Botany*, 2004, 55 (402): 1445-1454.
- [2] Udvardi M K, Czechowski T, Scheible W R. Eleven golden rules of quantitative RT-PCR [J]. *The Plant Cell*, 2008, 20 (7): 1736-1737.
- [3] Thellin O, Zorzi W, Lakaye B, et al. Housekeeping genes as internal standards: use and limits [J]. *Journal of Biotechnology*, 1999, 75 (2/3): 291-295.
- [4] Vandesompele J, de Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple

- internal control genes[J]. *Genome Biology*, 2002, 3(7): research0034.1.
- [5] Andersen C L, Jensen J L, Ørntoft T F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets[J]. *Cancer Research*, 2004, 64(15): 5245–5250.
- [6] Radonić A, Thulke S, MacKay I M, et al. Guideline to reference gene selection for quantitative real-time PCR[J]. *Biochemical and Biophysical Research Communications*, 2004, 313(4): 856–862.
- [7] Wang L, Li P, Wang Z, et al. Identification and validation of suitable internal reference genes for SYBR-GREEN RT-qPCR studies during cucumber development[J]. *The Journal of Horticultural Science and Biotechnology*, 2014, 89(3): 312–320.
- [8] Warzybok A, Migocka M. Reliable reference genes for normalization of gene expression in cucumber grown under different nitrogen nutrition[J]. *PLoS One*, 2013, 8(9): e72887.
- [9] Migocka M, Papierniak A. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators[J]. *Molecular Breeding*, 2011, 28(3): 343–357.
- [10] Wendel J F. Genome evolution in polyploids[J]. *Plant Molecular Biology*, 2000, 42(1): 225–249.
- [11] Wang J P, Sun P C, Li Y X, et al. An overlooked paleotetraploidization in Cucurbitaceae[J]. *Molecular Biology and Evolution*, 2018, 35(1): 16–26.
- [12] Chen J F, Staub J E, Tashiro Y, et al. Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr[J]. *Euphytica*, 1997, 96(3): 413–419.
- [13] 陈劲枫, 罗向东, 余纪柱, 等. 异源三倍体黄瓜的离体繁殖和鉴定[J]. *植物生理学通讯*, 2003, 39(2): 109–112.
Chen J F, Luo X D, Yu J Z, et al. Micropropagation and identification of an allotriploid cucumber in vitro[J]. *Plant Physiology Communications*, 2003, 39(2): 109–112 (in Chinese with English abstract).
- [14] 冀刚. 甜瓜属双二倍体与栽培黄瓜杂交亲和性研究[D]. 南京: 南京农业大学, 2012: 21–23.
Ji G. Study on hybridization compatibility between *Cucumis hytivus* and *Cucumis sativus* [D]. Nanjing: Nanjing Agricultural University, 2012: 21–23 (in Chinese with English abstract).
- [15] Jackson S, Chen Z J. Genomic and expression plasticity of polyploidy[J]. *Current Opinion in Plant Biology*, 2010, 13(2): 153–159.
- [16] Wang H B, Wang J J, Jiang J F, et al. Reference genes for normalizing transcription in diploid and tetraploid *Arabidopsis*[J]. *Scientific Reports*, 2014, 4: 6781.
- [17] Kim D, Langmead B, Salzberg S L. HISAT: a fast spliced aligner with low memory requirements[J]. *Nature Methods*, 2015, 12(4): 357–360.
- [18] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools[J]. *Bioinformatics*, 2009, 25(16): 2078–2079.
- [19] Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-seq[J]. *Nature Methods*, 2008, 5(7): 621–628.
- [20] Ramakers C, Ruijter J M, Deprez R H L, et al. Assumption-free analysis of quantitative real-time polymerase chain reaction(PCR) data[J]. *Neuroscience Letters*, 2003, 339(1): 62–66.
- [21] Hellemans J, Mortier G, de Paepe A, et al. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data[J]. *Genome Biology*, 2007, 8(2): R19.
- [22] Schmittgen T D, Livak K J. Analyzing real-time PCR data by the comparative CT method[J]. *Nature Protocols*, 2008, 3(6): 1101–1108.
- [23] Fernandez P, di Rienzo J A, Moschen S, et al. Comparison of predictive methods and biological validation for qPCR reference genes in sunflower leaf senescence transcript analysis[J]. *Plant Cell Reports*, 2011, 30(1): 63–74.
- [24] Huggett J, Dheda K, Bustin S, et al. Real-time RT-PCR normalisation; strategies and considerations[J]. *Genes & Immunity*, 2005, 6(4): 279–284.
- [25] Gutierrez L, Mauriat M, Guénin S, et al. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction(RT-PCR) analysis in plants[J]. *Plant Biotechnology Journal*, 2008, 6(6): 609–618.
- [26] Long X Y, Wang J R, Ouellet T, et al. Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat[J]. *Plant Molecular Biology*, 2010, 74(3): 307–311.
- [27] Wang H B, Chen S M, Jiang J F, et al. Reference gene selection for cross-species and cross-ploidy level comparisons in *Chrysanthemum* spp. [J]. *Scientific Reports*, 2015, 5: 8094.
- [28] Lovén J, Orlando D A, Sigova A A, et al. Revisiting global gene expression analysis[J]. *Cell*, 2012, 151(3): 476–482.
- [29] Gan X C, Stegle O, Behr J, et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana* [J]. *Nature*, 2011, 477(7365): 419–423.
- [30] Chen Z J. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids[J]. *Annual Review of Plant Biology*, 2007, 58: 377–406.