

闫学敏,吴英华,史艳,等. 胡萝卜 *YUCCA* 基因家族鉴定及生物信息学分析[J]. 江苏农业科学,2021,49(23):52-57.
doi:10.15889/j.issn.1002-1302.2021.23.009

胡萝卜 *YUCCA* 基因家族鉴定及生物信息学分析

闫学敏,吴英华,史艳,武喆,侯雷平,李梅兰

(山西农业大学园艺学院,山西太谷 030801)

摘要:胡萝卜是重要的根菜类蔬菜之一,根部性状影响最终的产量和品质。*YUCCA* 基因通过介导生长素调控植物根部发育,是生长素合成途径中的关键基因。利用生物信息学方法对胡萝卜 *YUCCA* 基因家族成员进行鉴定,对其理化性质、染色体定位、系统进化树、蛋白质的二级和三级结构以及保守基序进行分析。结果表明,在胡萝卜的 5 条染色体上共鉴定到了 14 个 *YUCCA* 基因家族成员,其大部分基因含有 3~4 个外显子;该家族编码的蛋白质为富含碱性氨基酸的亲水性蛋白质,氨基酸数量为 290~424 个,其蛋白质结构主要以 α -螺旋和无规则卷曲构成,通过构建系统进化树可将其分为 3 个亚族,且亚细胞定位结果显示,*YUCCA* 基因家族蛋白质大部分被定位在细胞质中。这为 *YUCCA* 基因家族在调控胡萝卜根发育和膨大机制研究奠定了理论基础。

关键词:胡萝卜;生长素;*YUCCA* 基因家族;生物信息学;蛋白质结构;染色体定位

中图分类号:S631.201 **文献标志码:**A **文章编号:**1002-1302(2021)23-0052-06

植物根系的生长发育存在着复杂的调控网络,受到环境、生理、激素等多种因素的影响^[1]。近年来,植物激素作为调控生长发育的一大类物质而被研究者广泛关注,生长素是最早发现的植物生长类激素,其代谢调控和信号转导机制研究较为广泛^[2]。其中,*YUCCA* 基因家族已经被证明调控植物的根系发育,在生长素生物合成的吲哚丙酮酸途径中 *YUCCA* 基因家族是重要酶之一,其基因编码黄素单加氧酶(flavin-containing monooxygenase,简称 FMOs)^[3-4]。*YUCCA* 基因首次被鉴定是在拟南芥生长素过量的突变体研究中^[5],目前该基因家族已经在烟草^[6]、拟南芥^[7]、水稻^[8]、草莓^[9]等作物中进行了全基因组鉴定分析,但在胡萝卜中还未见类似

的报道。

胡萝卜(*Daucus carota* var. *sativa* DC.)是以肉质根为食用器官的二年生草本植物,是全球性十大蔬菜中的一种。根是植物在进化过程中适应陆地生活而发展起来的营养器官,胡萝卜作为重要的根菜类蔬菜之一,研究其根部的生长发育对提高胡萝卜的品质和产量至关重要。有研究表明,生长素在植物根尖形成的梯度浓度调控其合成和运输,从而影响植物的根际发育^[10]。因此,本试验利用生物信息学方法鉴定胡萝卜 *YUCCA* 基因家族,对其理化性质、染色体定位、系统进化树、蛋白质的二级结构和三级结构以及保守基序进行分析,进一步为胡萝卜 *YUCCA* 基因功能验证提供一定的理论基础。

1 材料与方法

1.1 胡萝卜 *YUCCA* 基因家族数据获取与鉴定

通过拟南芥官网 TAIR (<https://www.arabidopsis.org/>) 获取已报道过的 *YUCCA* 基因家族成员的蛋白质序列,在 NCBI 网站 (<https://www.ncbi.nlm.nih.gov/>) 使用 BlastP 检索,设置 E-value

收稿日期:2021-07-16

基金项目:山西省重点研发计划重点项目子课题(编号:201703D211001-04-01);山西省重点研发计划(编号:201903D221063)。

作者简介:闫学敏(1997—),女,山西吕梁人,硕士研究生,主要从事蔬菜育种及生物技术应用研究。E-mail:464679520@qq.com。

通信作者:李梅兰,博士,教授,主要从事蔬菜育种及生物技术应用研究。E-mail:15935485975@163.com。

[36]徐建军,赵强,汤在祥,等. 利用重测序的染色体片段代换系群体定位水稻粒型 QTL[J]. 中国水稻科学,2011,25(4):365-369.

[37]卫纯洁,陶亚军,范方军,等. 利用重测序染色体片段代换系群体定位水稻籽粒长宽比 QTL[J]. 江苏农业科学,2020,48(6):36-40.

[38]张健,杨靖,王豪,等. 基于高密度遗传图谱定位水稻籽粒大小相关性状 QTL[J]. 中国农业科学,2020,53(2):225-238.

[39]Tanabe S, Ashikari M, Fujioka S, et al. A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length[J]. The Plant Cell,2005,17(3):776-790.

值为 10^{-5} , 获得拟南芥、烟草、玉米和胡萝卜中的 *YUCCA* 基因家族候选基因的蛋白质序列和基因序列的 *fasta* 格式。通过 *CDD* 和 *SMART* 数据库比对验证候选基因的蛋白质序列结构域^[11], 最终获得 11 个拟南芥 *YUCCA* 基因、20 个烟草 *YUCCA* 基因、30 个玉米 *YUCCA* 基因和 14 个胡萝卜 *YUCCA* 基因, 并对其分别命名。

1.2 胡萝卜 *YUCCA* 蛋白质理化性质分析

将筛选出的 14 个胡萝卜 *YUCCA* 基因家族蛋白质序列通过 ExPASy (<https://web.expasy.org/protparam/>) 在线网站分析其理化性质^[12]。

1.3 胡萝卜 *YUCCA* 基因的染色体定位

从 NCBI 上获取基因的染色体位置信息, 使用在线网站 MG2C_V2.0 (http://mg2c.iask.in/mg2c_v2.0/) 对基因进行染色体定位绘图并简单修饰。

1.4 拟南芥、烟草、玉米和胡萝卜 *YUCCA* 基因家族系统进化树分析

在软件 MEGA 7.0 中对比 4 个物种的蛋白质序列, 采用邻接法构建进化树, bootstrap 设为 1 000^[13]。用在线网站 iTOL (<https://itol.embl.de/itol.cgi>) 对进化树进行环化。

1.5 胡萝卜 *YUCCA* 蛋白质二、三级结构预测及亚细胞定位

通过在线网站 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 和 CELLO (<http://cello.life.nctu.edu.tw/>) 分别对胡萝卜 *YUCCA* 基因家族的蛋白质序列

进行二级结构预测和亚细胞定位。通过在线网站 SWISS - MODEL (<https://swissmodel.expasy.org/>) 对胡萝卜 *YUCCA* 基因家族的蛋白质序列进行三级结构模型分析^[14]。

1.6 胡萝卜 *YUCCA* 基因家族的基因结构及蛋白质保守基序分析

将 14 个基因的 CDS 和 FASTA 文件导入在线网站 GSDS 2.0 (<http://gsds.gao-lab.org/index.php>) 进行胡萝卜 *YUCCA* 基因结构分析^[15]。使用 MEME 在线网站 (<https://meme-suite.org/meme/tools/meme>) 和 TBtools 对胡萝卜 *YUCCA* 蛋白质的保守基序进行分析^[16]。

2 结果与分析

2.1 胡萝卜 *YUCCA* 蛋白质理化性质

通过对胡萝卜 *YUCCA* 基因家族的蛋白质进行理化性质分析(表 1)可知, *YUCCA* 蛋白质的氨基酸数为 290 ~ 424 个, 相对分子质量为 32.14 ~ 47.32 ku, 氨基酸数和分子质量最大的均为蛋白质 DcYUCCA3b, 最小的均为蛋白质 DcYUCCA6; 等电点为 6.61 ~ 9.35; 不稳定系数为 31.98 ~ 50.25, 其中蛋白质 DcYUCCA5a、DcYUCCA5b、DcYUCCA5c、DcYUCCA5d、DcYUCCA10a 和 DcYUCCA10b 均为稳定蛋白, 其余的蛋白质 (DcYUCCA2、DcYUCCA3a、DcYUCCA3b、DcYUCCA6、DcYUCCA8a、DcYUCCA8b、DcYUCCA8c 和 DcYUCCA9) 为不稳定蛋白; 亲水性均为负值, 说明胡萝卜中 *YUCCA* 蛋白质均为亲水性蛋白。

表 1 胡萝卜 *YUCCA* 基因家族蛋白质理化性质分析

基因名称	登录号	氨基酸数 (个)	分子质量 (ku)	等电点	原子总数	不稳定系数	脂肪系数	亲水性
<i>DcYUCCA2</i>	KZM83868.1	400	45.352	8.22	6 409	40.62	89.88	-0.272
<i>DcYUCCA3a</i>	KZM85217.1	422	46.948	9.13	6 654	44.64	90.07	-0.117
<i>DcYUCCA3b</i>	KZM83207.1	424	47.319	9.35	6 686	43.94	88.87	-0.104
<i>DcYUCCA5a</i>	KZN07891.1	410	45.528	9.03	6 445	33.68	85.78	-0.202
<i>DcYUCCA5b</i>	KZN07892.1	414	46.189	8.87	6 537	33.83	85.19	-0.224
<i>DcYUCCA5c</i>	KZN07893.1	414	46.209	8.58	6 525	31.98	84.95	-0.228
<i>DcYUCCA5d</i>	KZN07894.1	410	45.649	8.81	6 453	32.31	86.02	-0.198
<i>DcYUCCA6</i>	KZN08071.1	290	32.142	8.07	4 502	50.25	81.93	-0.231
<i>DcYUCCA8a</i>	KZM95291.1	346	38.931	6.61	5 486	41.70	93.99	-0.038
<i>DcYUCCA8b</i>	KZN12041.1	296	33.241	8.44	4 676	41.25	82.64	-0.219
<i>DcYUCCA8c</i>	KZN03861.1	419	46.971	8.93	6 656	43.02	93.22	-0.121
<i>DcYUCCA9</i>	KZN07881.1	419	47.026	8.61	6 615	42.14	81.36	-0.231
<i>DcYUCCA10a</i>	KZM95122.1	379	42.094	8.48	5 942	32.27	92.30	-0.017
<i>DcYUCCA10b</i>	KZN03673.1	378	42.611	8.64	6 016	35.30	85.63	-0.118

2.2 胡萝卜 YUCCA 基因的染色体定位

胡萝卜共有 9 条染色体。染色体定位结果(图 1)显示,胡萝卜 YUCCA 基因家族中有 7 个基因被定位在 1 号染色体上,分布最多,基因存在聚集现象,

说明这类基因功能具有一定的相似性。其次,有 2 个基因被定位在 3 号染色体上,2 个基因被定位在 5 号染色体上,2 个基因被定位在 8 号染色体上,1 个基因被定位在 9 号染色体上。

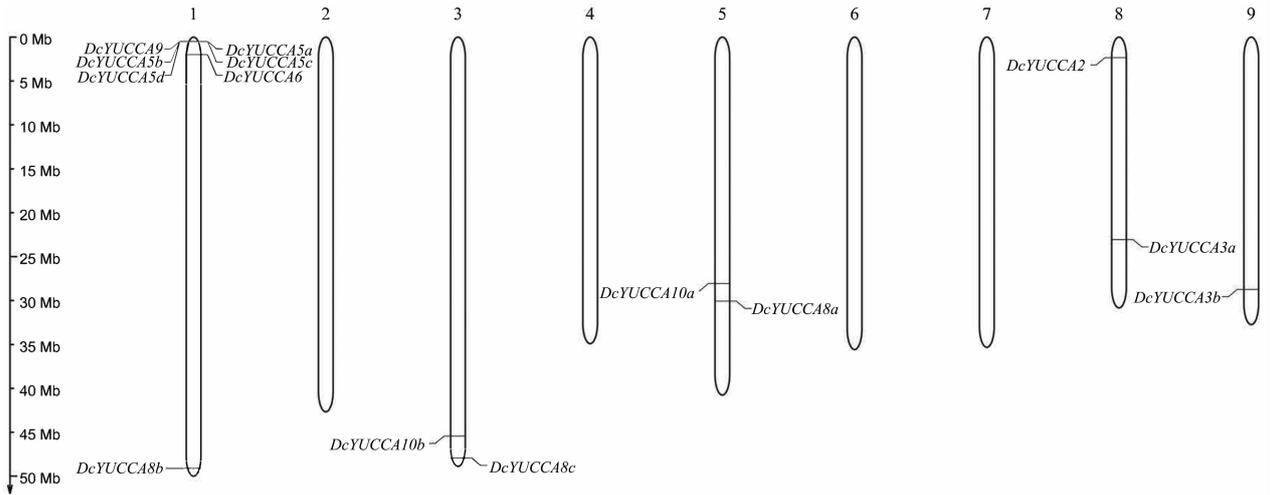


图1 胡萝卜 YUCCA 基因在染色体上的位置

2.3 拟南芥、烟草、玉米和胡萝卜 YUCCA 基因家族系统进化树

为进一步了解胡萝卜 YUCCA 基因家族进化情况,选用拟南芥(11 个)、烟草(20 个)、玉米(30 个)

YUCCA 基因家族基因为参考序列构建系统进化树(图 2),结果分为 6 个亚族。其中胡萝卜 YUCCA 基因家族主要分布在 VI 亚族中,有 10 个基因,其次为 III 和 V 亚族中。

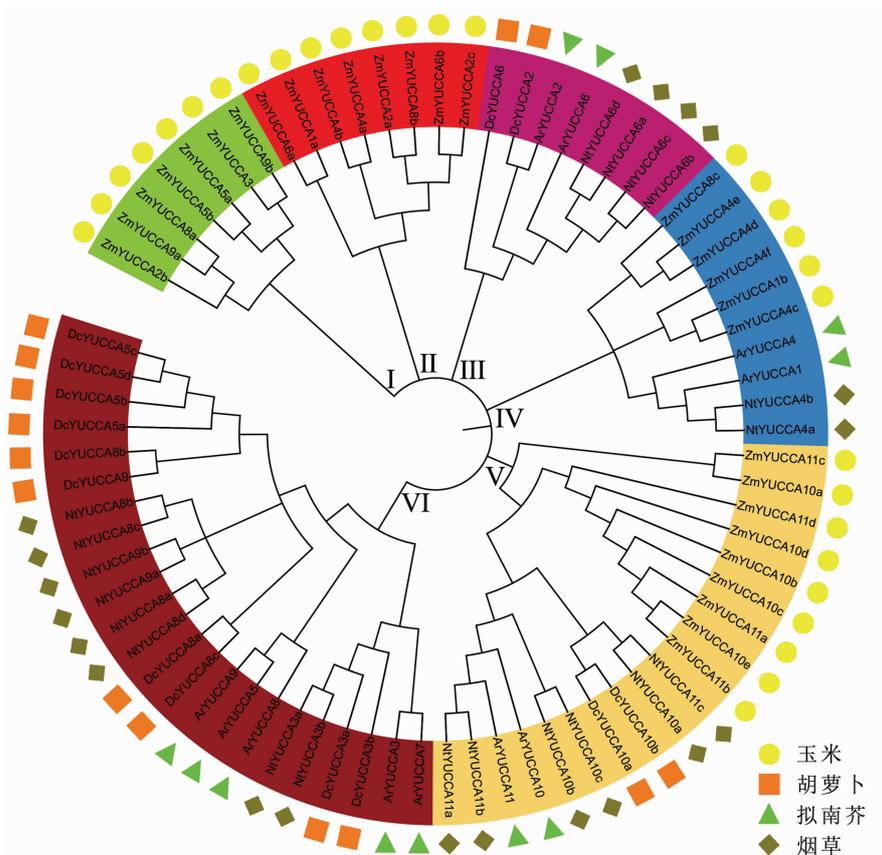


图2 拟南芥、烟草、玉米和胡萝卜 YUCCA 基因家族系统进化树

2.4 胡萝卜 YUCCA 蛋白质二、三级结构预测及亚细胞定位

对胡萝卜 YUCCA 蛋白质的二级结构预测结果(表 2)显示,14 个蛋白质主要以 α -螺旋和无规则卷曲构成。为了验证二级结构的结果准确性,对 14 个蛋白质的三级结构进行模型预测(图 3),结果显示三级结构与二级结构基本一致。蛋白质的亚细胞定位结果显示大多数的蛋白质均被定位在细胞质中(13 个),还有少部分的蛋白质被定位在周质和外膜上。

2.5 胡萝卜 YUCCA 基因结构

为了分析胡萝卜 YUCCA 基因家族的功能,对 14 个基因进行基因结构分析(图 4),结果得出基因的长度不同,外显子和内含子也存在差别。基因 *DcYUCCA2*、*DcYUCCA6*、*DcYUCCA10a* 和 *DcYUCCA10b* 均含有 4 个外显子和 3 个内含子,其中基因 *DcYUCCA10a* 无上下游;基因 *DcYUCCA3a*、*DcYUCCA3b*、*DcYUCCA5a*、*DcYUCCA5b*、*DcYUCCA5c*、*DcYUCCA5d*、*DcYUCCA8a*、*DcYUCCA8b* 和 *DcYUCCA8c* 含有 3 个外显子和 2 个内含子,其中基因 *DcYUCCA5a*、

DcYUCCA5b、*DcYUCCA5c*、*DcYUCCA5d* 和 *DcYUCCA8b* 无上下游;基因 *DcYUCCA9* 含有 2 个外显子和 1 个内含子。

表 2 胡萝卜 YUCCA 蛋白质二级结构预测及亚细胞定位

基因名称	占比(%)				亚细胞定位
	α -螺旋	β -转角	延伸链	无规则卷曲	
<i>DcYUCCA2</i>	31.75	5.75	15.75	46.75	细胞质
<i>DcYUCCA3a</i>	33.89	6.87	16.82	42.42	细胞质
<i>DcYUCCA3b</i>	28.77	6.60	20.05	44.58	周质
<i>DcYUCCA5a</i>	30.24	7.07	17.32	45.37	周质、细胞质
<i>DcYUCCA5b</i>	29.95	6.76	17.63	45.65	细胞质
<i>DcYUCCA5c</i>	29.95	7.00	17.15	45.89	细胞质
<i>DcYUCCA5d</i>	31.71	8.29	16.83	43.17	细胞质
<i>DcYUCCA6</i>	30.34	6.55	20.34	42.76	周质、细胞质
<i>DcYUCCA8a</i>	25.72	6.36	21.10	46.82	细胞质
<i>DcYUCCA8b</i>	24.32	3.72	16.89	55.07	细胞质
<i>DcYUCCA8c</i>	32.94	7.16	17.42	42.48	细胞质
<i>DcYUCCA9</i>	30.55	8.35	17.66	43.44	细胞质
<i>DcYUCCA10a</i>	34.04	8.18	17.94	39.84	周质、细胞质、外膜
<i>DcYUCCA10b</i>	31.22	7.14	19.58	42.06	细胞质

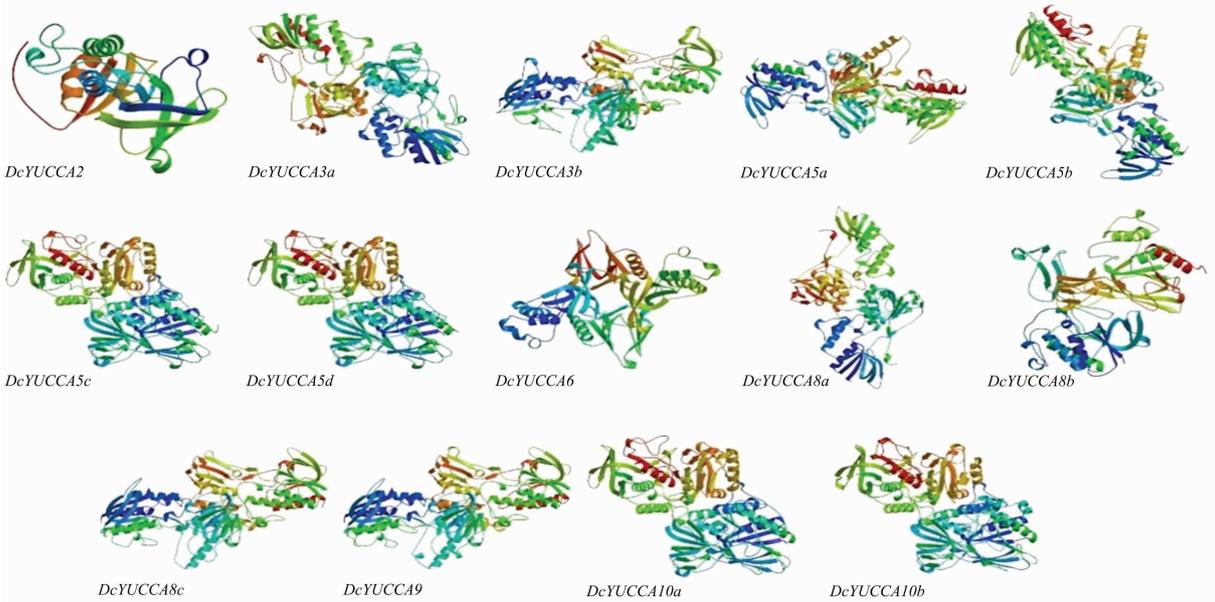


图 3 胡萝卜 YUCCA 蛋白质三级结构模型

2.6 胡萝卜 YUCCA 基因家族蛋白质保守基序

对胡萝卜 YUCCA 基因家族蛋白质序列进行保守基序分析(图 5),发现 14 个蛋白质的保守结构域为 6 ~ 10 个。蛋白质 *DcYUCCA3a*、*DcYUCCA3b*、*DcYUCCA5a*、*DcYUCCA5b*、*DcYUCCA5c*、*DcYUCCA5d*、*DcYUCCA8c* 和 *DcYUCCA9* 含有 10 个 motif,结构和功能最为相似,14 个 YUCCA 蛋白质中均含有 motif

1、motif 7 和 motif 3。蛋白质 *DcYUCCA2* 不含 motif 9、motif 6 和 motif 10;蛋白质 *DcYUCCA6* 不含 motif 5、motif 4、motif 6 和 motif 10;蛋白质 *DcYUCCA10a* 和 *DcYUCCA10b* 结构基本一致,均不含 motif 8、motif 9 和 motif 10;蛋白质 *DcYUCCA8a* 不含 motif 9、motif 4 和 motif 10;蛋白质 *DcYUCCA8b* 不含 motif 2、motif 5 和 motif 6。

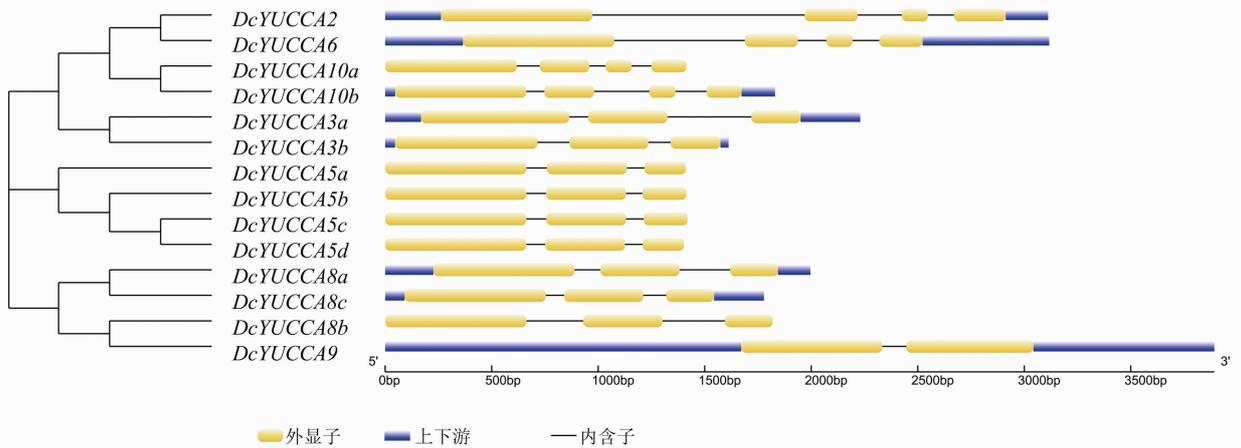


图4 胡萝卜 YUCCA 基因结构

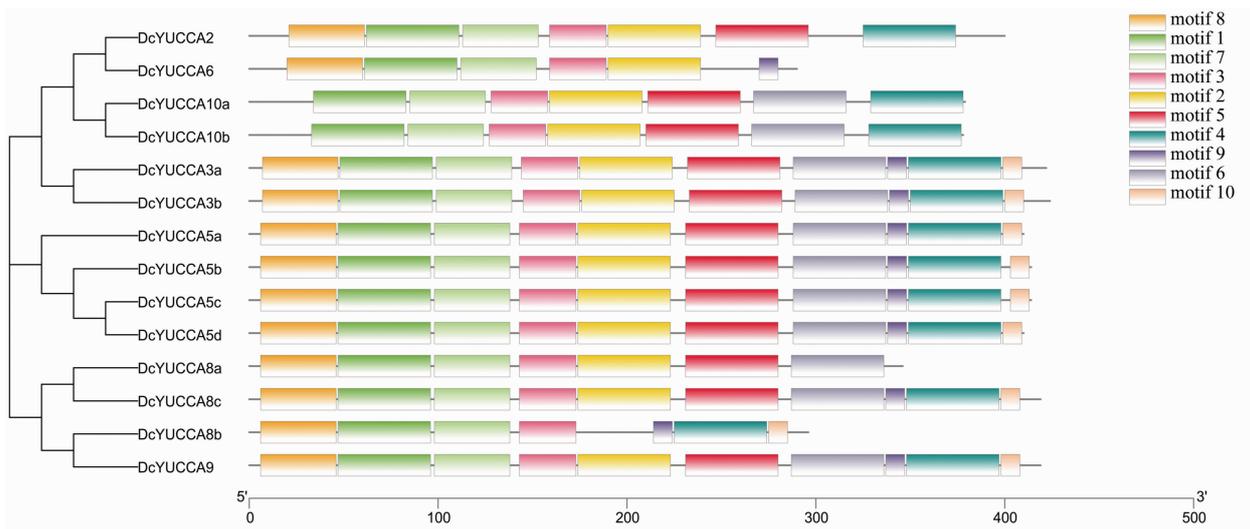


图5 胡萝卜 YUCCA 蛋白质保守基序

3 讨论与结论

植物根部的生长发育是一个动态变化的过程,是由根尖分生区细胞的分裂和伸长区细胞的伸长而生长的,在这个过程中植物激素发挥了重要的作用^[17]。生长素是其中一类生长调节激素,可以促进细胞的伸长、木质部和韧皮部的分化,从而使根部增粗^[18]。在胡萝卜中,研究根部膨大增粗可以提高其品质和产量,提高经济效益。*YUCCA* 基因家族蛋白是植物中第一个被鉴别的 FMO 类超家族,调控生长发育和植物的耐受性^[19]。有研究证实 *YUCCA* 基因介导生长素的生物合成,在拟南芥中抑制 *YUCCA* 表达会抑制生根,调控根器官发育^[20]。

本研究利用生物信息学方法鉴定胡萝卜 *YUCCA* 基因家族成员共 14 个,与拟南芥 *YUCCA* 基因家族的 11 个成员数量较接近。通过理化性质分

析得出,胡萝卜 *YUCCA* 基因家族蛋白质基本为碱性氨基酸(除了蛋白质 DcYUCCA8a),且 14 个蛋白质均为亲水性蛋白质,基因分别定位在 5 条染色体上(1,3,5,8,9 号),其中在 1 号染色体上较聚集,结构和功能最为相似。系统进化树中聚集在一起的亚族具有相似的进化源和基因功能,通过构建系统进化树将胡萝卜 *YUCCA* 基因家族分为 3 个亚族(Ⅲ、Ⅴ、Ⅵ亚族),与烟草和拟南芥的亲缘关系较近,其中,Ⅲ亚族中的 *DcYUCCA2* 和 *DcYUCCA6* 与拟南芥的 *ArYUCCA2* 和 *ArYUCCA6* 亲缘关系最近,推测可能通过调控生长素信号转导介导植物的生长发育和早期花粉发育^[21]。

蛋白质的结构决定其生物学功能,结构相似其功能相似,在分析蛋白质的二级结构和三级结构模型时发现,*YUCCA* 基因家族蛋白质主要以 α -螺旋和无规则卷曲构成,位于同一家族成员的结构具有

一致性。在亚细胞定位中发现大部分的蛋白质被定位在细胞质中,表明这些基因合成分生长素基本上是在细胞质中完成的,这与王仁汉等的结果^[6,22]一致。通过分析胡萝卜 *YUCCA* 基因家族蛋白质的保守基序发现,14 个蛋白质的 motif 有 6 ~ 10 个不等,说明基因家族功能的多样性,但其中有大部分 *YUCCA* 蛋白质的 motif 数量均为 10 个,说明基因家族的功能具有相似性。总体来看,位于同一基因家族的成员在植物中的结构和功能均具有相似性,为进一步研究基因家族提供参考,为胡萝卜根部膨大形成机制的研究奠定了理论基础。

本研究在胡萝卜的 5 条染色体上共鉴定到了 14 个 *YUCCA* 基因家族成员,其大部分基因含有 3 ~ 4 个外显子;该家族编码的蛋白质为富含碱性氨基酸的亲水性蛋白质,氨基酸数量为 290 ~ 424 个,其蛋白质结构主要以 α - 螺旋和无规则卷曲构成,通过构建系统进化树可将其分为 3 个亚族,且亚细胞定位结果显示,*YUCCA* 基因家族蛋白质大部分被定位在细胞质中。这为 *YUCCA* 基因家族在调控胡萝卜根发育和膨大机制研究奠定了理论基础。

参考文献:

- [1] 马凯,杨凡,蔡毓新,等. 蔬菜不定根形成调控机制研究进展[J]. 中国瓜菜,2017,30(3):1-4.
- [2] Zhao Y D. Auxin biosynthesis and its role in plant development[J]. Annual Review of Plant Biology,2010,61:49-64.
- [3] Brumos J,Alonso J M,Stepanova A N. Genetic aspects of auxin biosynthesis and its regulation[J]. Physiologia Plantarum,2014,151(1):3-12.
- [4] Qin M M,Wang J,Zhang T Y, et al. Genome-wide identification and analysis on *YUCCA* gene family in *Isatis indigotica* Fort. and *lYUCCA6-1* functional exploration[J]. International Journal of Molecular Sciences,2020,21(6):2188.
- [5] Zhao Y,Christensen S K,Fankhauser C, et al. A role for flavin monooxygenase-like enzymes in auxin biosynthesis[J]. Science,2001,291(5502):306-309.
- [6] 王仁汉,宋志美,屈旭,等. 普通烟草 *YUCCA* 基因家族的生物信息学分析[J]. 江苏农业科学,2021,49(3):61-65.
- [7] Cheng Y F,Dai X H,Zhao Y D. Auxin biosynthesis by the *YUCCA* flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*[J]. Genes & Development,2006,20(13):1790-1799.
- [8] Yamamoto Y,Kamiya N,Morinaka Y, et al. Auxin biosynthesis by the *YUCCA* genes in rice[J]. Plant Physiology,2007,143(3):1362-1371.
- [9] Liu H,Xie W F,Zhang L, et al. Auxin biosynthesis by the *YUCCA6* flavin monooxygenase gene in woodland strawberry[J]. Journal of Integrative Plant Biology,2014,56(4):350-363.
- [10] Cao X,Yang H L,Shang C Q, et al. The roles of auxin biosynthesis *YUCCA* gene family in plants[J]. International Journal of Molecular Sciences,2019,20(24):6343.
- [11] Wang R,Liu L,Kong Z S, et al. Identification of *GhLOG* gene family revealed that *GhLOG3* is involved in regulating salinity tolerance in cotton (*Gossypium hirsutum* L.)[J]. Plant Physiology and Biochemistry,2021,166:328-340.
- [12] Jiang S,Li S,Liu X A, et al. Genome-wide identification and characterization of the *MdBZRI* gene family in apple and their roles in improvement of drought tolerance[J]. Scientia Horticulturae,2021,288:110359.
- [13] Zheng L W,Ma S J,Zhou T, et al. Genome-wide identification of Brassicaceae *B-BOX* genes and molecular characterization of their transcriptional responses to various nutrient stresses in allotetraploid rapeseed[J]. BMC Plant Biology,2021,21(1):288.
- [14] 罗康胜,张米欢,翟兆东,等. 生菜 *WOX* 基因家族的鉴定与分析[J/OL]. 分子植物育种:1-18[2021-10-12]. <http://kns.cnki.net/kcms/detail/46.1068.S.20210622.1012.002.html>.
- [15] Haider M S, de Britto S, Nagaraj G, et al. Genome-wide identification, diversification, and expression analysis of lectin receptor-like kinase (*LecRLK*) gene family in cucumber under biotic stress[J]. International Journal of Molecular Sciences,2021,22(12):6585.
- [16] 李明明,晋敏娜,胡海斌,等. 谷子 *MYB-CC* 基因家族的鉴定与表达分析[J]. 激光生物学报,2021,30(3):259-269.
- [17] 王柳萍,覃坤坚,赵立春,等. 植物根膨大的研究进展[J]. 湖北农业科学,2020,59(11):5-9.
- [18] 邹禹,刘园园,张培江,等. 水稻富亮氨酸重复类受体蛋白激酶 OsRPK1 响应外源生长素的作用研究[J]. 江苏农业科学,2019,47(8):46-51.
- [19] Cha J Y, Kim W Y, Kang S B, et al. A novel thiol-reductase activity of *Arabidopsis YUC6* confers drought tolerance independently of auxin biosynthesis[J]. Nature Communications,2015,6:8041.
- [20] Chen L Q, Tong J H, Xiao L T, et al. *YUCCA*-mediated auxin biogenesis is required for cell fate transition occurring during de novo root organogenesis in *Arabidopsis*[J]. Journal of Experimental Botany,2016,67(14):4273-4284.
- [21] Yao X Z, Tian L, Yang J, et al. Auxin production in diploid microsporocytes is necessary and sufficient for early stages of pollen development[J]. PLoS Genetics,2018,14(5):e1007397.
- [22] 綦洋,王束钧,桑园园,等. 大白菜 *YUCCA* 基因家族的鉴定与生物信息学分析[J]. 江苏农业科学,2019,47(3):49-54.