

大豆 NAC 基因家族生物信息学分析

王 洋, 柏 锡

(东北农业大学 生命科学院, 黑龙江 哈尔滨 150030)

摘要: NAC 转录因子是植物特有的具有多种生物功能的一类重要转录因子, 广泛参与植物生长发育以及生物与非生物逆境应答等。本文利用生物信息学手段, 结合大豆基因组数据库及 NCBI 数据库, 识别、筛选并获得了大豆 NAC 家族基因的蛋白序列, 并对其进化关系、在基因组上的定位分布以及理化特性进行了预测和分析。结果显示具有 NAC 结构域的 152 条大豆氨基酸序列认为是假定 NAC 蛋白质, 共分为 10 个亚家族, 连锁群定位结果发现大豆的 20 条染色体上均有 NAC 转录因子的分布, 其中第 12 号染色体上分布最多。不同亚族间的大豆假定 NAC 蛋白质理化特性存在一定的差异, 无规则卷曲和 α -螺旋是蛋白质二级结构最大量的结构元件, 并散布于整个蛋白。本研究结果将为大豆 NAC 基因家族的进一步功能分析奠定重要的研究基础。

关键词: 大豆; 生物信息学; NAC 基因家族

中图分类号: S565.1

文献标识码: A

DOI: 10.11861/j.issn.1000-9841.2014.03.0325

Bioinformatics Analysis of NAC Gene Family in *Glycine max* L.

WANG Yang, BAI Xi

(College of Life Sciences, Northeast Agricultural University, Harbin 150030, China)

Abstract: Genes containing the NAC domain (NAC family genes) are plant-specific transcriptional regulators and expressed in various developmental stages and tissues. We performed a bioinformatics analysis of NAC family genes in soybean. Based on bioinformatics methods, taking advantage of soybean database combined with public database (NCBI), we identified 152 NAC proteins from soybean genome. NAC domains from both predicted and known NAC family proteins were classified into ten subgroups by sequence similarity. Further genetic mapping of NAC genome localization found that they distributed on 20 chromosomes which the twelfth chromosome exist the most. We also predicted and analyzed their amino acid composition, physical and chemical characteristics, as well as secondary structures. The research found that the number of amino acid and hydrophobic of amino acid sequences in different subfamilies presented some differences. Meanwhile, the results of secondary structure predicted that the main composition of 152 predicted NAC proteins among them were random coil and α helix. The results of this investigation could definitely provide a significant foundation for further research on the function analysis of soybean NAC gene family.

Key words: Soybean; Bioinformatics; NAC gene family

在植物识别病原物后激活防卫反应的过程中, 转录因子起着重要的作用, 它通过激活或者抑制防卫基因的表达来实现防卫反应的调控。目前, 植物中已知的转录因子根据其 DNA 结构域的不同主要被分为 WRKY、bZIP、MYB、SBP、HB、DREB、AP2/EREBP、C2H2 和 NAC 等若干个家族^[1-2]。

NAC 转录因子是植物特有的一个庞大家族, 在植物生长发育、激素应答及抗逆反应中起着重要作用^[3-4]。研究表明, 该家族基因广泛参与花成熟^[5]、枝顶端分生组织发育^[6]、侧根发育^[3,7]、细胞次生壁合成^[8-9]、逆境诱导开花^[10]以及激素信号传导^[11-13]等生物与非生物逆境应答过程。

关于 NAC 基因家族的研究主要集中在少数几

种模式植物上, 如拟南芥^[14-15]、水稻^[16-17]等。全面了解大豆 NAC 基因家族的功能, 将对理解大豆植物发育和抗逆境的分子机理起到重要作用。本研究结合公共数据库资源, 从基因组水平上对大豆 NAC 基因家族进行了生物信息挖掘, 初步分析了 NAC 基因家族染色体定位、分类、系统发育以及理化特性等, 为 NAC 转录因子的表达调控、深入探讨 NAC 基因家族成员的功能以及耐逆相关研究等奠定了基础。

1 材料与方法

1.1 大豆 NAC 蛋白质序列的获得

根据已知的拟南芥 NAC 家族蛋白质序列, 利用

收稿日期: 2013-08-05

基金项目: 教育部科学技术研究重点项目 (212049); 教育部高等学校博士学科点专项科研基金 (20102325120002); 转基因生物新品种培育重大专项 (2013ZX08004-003)。

第一作者简介: 王洋 (1987-), 女, 在读硕士, 主要从事应用生物信息学研究。E-mail: dongtian_601@126.com。

通讯作者: 柏锡 (1975-), 男, 博士, 硕士生导师, 主要从事植物分子生物学和应用生物信息学研究。E-mail: baixi@neau.edu.cn。

Hammer 3.0 软件建立 Profile HMM 模型^[18], 并利用此模型检索从大豆全基因组数据库中 (<http://www.phytozome.net/index.php>) 下载的蛋白质序列, 去掉冗余后得到的序列被认为是候选蛋白质序列。将这些候选序列利用 SMATR (<http://smart.embl-heidelberg.de/>) 进行在线比对, 预测这些蛋白质有无保守的 NAC 结构域, 保留存在 NAC 保守区域的序列用于大豆 NAC 相关基因的后续分析。

1.2 大豆 NAC 蛋白家族系统发生分析

用分子进化遗传分析软件 MEGA 5.2 内置的 Cluastal W 程序对获得的大豆 NAC 的蛋白质进行氨基酸比对分析。根据多序列连配的结果, 使用 MEGA5.2 软件采用相邻连接法 (neighbor-joining, NJ) 构建系统发生树, 并进行自举检验, 随机抽样重复 (bootstrap) 次数设为 1 000, 其他参数使用默认值。系统发生树的显示同样采用 MEGA5.2 软件。参考拟南芥和水稻^[19] 的亚族分类情况进行大豆的 NAC 蛋白亚族分类。

1.3 大豆 NAC 蛋白一级结构理化特性分析

采用 ExPaSy 提供的在线 ProtParam 软件^[20] 对获得的大豆 NAC 蛋白序列进行氨基酸数目、分子量、理论等电点、脂肪族氨基酸数和蛋白质疏水性分析。

1.4 大豆 NAC 蛋白二级结构域分析

采用 ExPaSy 提供的在线 SOPMA 程序^[21] 预测分析 α -螺旋、 β -折叠和无规则卷曲等在整体结构中所占的比例。

1.5 大豆 NAC 家族基因的表达模式分析

在 NCBI 的 EST 数据库下载大豆表达序列标签 (EST) 序列 (共下载 262 个 EST 序列), 以大豆 NAC 基因家族预测成员为靶序列进行比对, 评估值 (E-value) 设置为常规默认值 10, 得到的序列去除冗余

后, 筛选保留大豆 NAC 预测基因对应的 EST 序列。根据 EST 来源文库信息分析 NAC 家族基因在不同组织和器官中的表达情况。

2 结果与分析

2.1 大豆 NAC 蛋白质形成一个大家族

根据拟南芥 NAC 保守区域的蛋白序列在 NCBI 及大豆基因组数据库 (Phytozome) 中共检索获得大豆 NAC 蛋白序列 202 条, 去冗余后再用 Conserved Domain Search 软件对其结构域进行检测, 最后共得到 152 条具有 NAC 结构域的蛋白序列, 即获得 152 个假定 NAC 蛋白质 (表 1)。针对鉴定得到的大豆 NAC 基因, 选择如下标准进行分类: 若 NAC 基因位于不同基因座位时, 该基因被认为是 1 个成员; 同一基因座位的多个剪接体被认为是 1 个成员, 保留 NCBI 中记载的与 Phytozome 中一致的剪接体作为代表。

基于已知大豆 NAC 基因的蛋白质序列, 利用 Phytozome 提供的基因搜索浏览器确定预测的大豆 NAC 基因所在染色体。通过染色体定位分析可以发现, NAC 基因在大豆的 20 条染色体上均有分布。其中, 分布在第 3 条染色体上的基因数量最少, 只有 3 条, 分别为 03g27580.1、03g27570.1、03g35570.1; 第 12 条染色体上的 NAC 基因数量最多, 为 16 条, 分别为 12g00540.1、12g02540.1、12g35000.1、12g00760.1、12g22880.1、12g31211.1、12g04383.1、12g22790.1、12g09670.1、12g13710.1、12g33460.1、12g34990.1、12g31150.1、12g29360.1、12g21272.1、12g18996.1。其余的基因分别在染色体上数量为 4~13 条不等。

表 1 大豆 NAC 家族基因数据库信息

Table 1 Information of soybean NAC family gene database

基因登录号 Gene model ID	氨基酸数目 No. of AA	分子量 Molecular weight/Da	理论等电点 Theoretical pI	脂肪族氨基酸指数 A Lipathic index	蛋白质疏水性 GRAVY
Glyma20g31210.1	549	61694.6	4.75	71.89	-0.489
Glyma20g33390.1	604	66939.2	4.82	69.12	-0.537
Glyma08g03590.1	452	50960.8	6.80	62.35	-0.861
Glyma08g04610.1	313	35222.4	8.84	59.81	-0.695
Glyma06g17480.1	269	30522.1	8.89	65.95	-0.657
Glyma10g36360.1	560	62759.8	4.70	71.04	-0.455
Glyma12g00540.1	356	40969.6	7.69	55.56	-0.981
Glyma14g39080.1	644	73440.7	5.08	74.11	-0.515
Glyma15g40510.1	303	34335.0	8.51	61.49	-0.599
Glyma16g04740.1	353	39812.0	8.87	69.78	-0.640
Glyma16g05620.1	216	24521.6	8.53	61.39	-0.699
Glyma16g26810.1	410	46655.2	6.33	57.05	-0.774
Glyma17g10970.1	350	39490.6	7.62	65.17	-0.572
Glyma18g05020.1	678	76538.1	5.52	71.30	-0.505

续表 1

基因登录号	氨基酸数目	分子量	理论等电点	脂肪族氨基酸指数 A	蛋白质疏水性
Gene model ID	No. of AA	Molecular weight/Da	Theoretical pI	Liphatic index	GRAVY
Glyma09g31650.1	331	37151.1	8.56	74.80	-0.467
Glyma09g36820.1	359	41277.0	7.70	53.20	-1.006
Glyma10g34130.1	465	52044.4	6.85	58.34	-0.772
Glyma06g15990.1	324	37091.7	4.72	65.86	-0.542
Glyma20g33430.1	465	52099.4	6.81	60.00	-0.749
Glyma06g15840.1	503	54466.8	4.82	71.11	-0.351
Glyma05g35090.1	321	36343.7	7.77	60.72	-0.704
Glyma05g23840.1	345	38904.4	5.66	56.78	-0.617
Glyma05g00930.1	348	39185.1	6.72	62.76	-0.584
Glyma02g40750.1	643	73271.2	5.26	44.13	-0.549
Glyma02g07760.1	410	46762.4	6.33	55.15	-0.785
Glyma01g00880.1	462	52074.1	6.58	61.41	-0.833
Glyma12g02540.1	297	34334.5	8.51	71.52	-0.623
Glyma12g35000.1	345	38981.7	8.15	70.23	-0.612
Glyma02g26480.1	268	30561.7	8.10	64.81	-0.619
Glyma06g38410.1	337	38003.9	6.26	65.31	-0.513
Glyma01g06150.1	279	31988.3	9.01	70.22	-0.643
Glyma06g11970.1	299	33949.4	6.54	63.98	-0.653
Glyma04g42800.1	300	34310.1	8.60	63.07	-0.672
Glyma02g12220.1	279	32040.4	9.00	69.18	-0.622
Glyma14g24220.1	280	31777.0	7.03	66.21	-0.629
Glyma19g44890.1	265	30624.7	6.60	69.47	-0.578
Glyma02g38710.1	131	15205.3	8.44	75.80	-0.560
Glyma08g18470.1	302	34357.1	8.04	62.62	-0.595
Glyma15g05690.1	191	22066.5	5.14	69.32	-0.729
Glyma01g05680.1	438	49723.0	5.95	66.00	-0.568
Glyma05g32850.1	298	3423.7	6.46	64.16	-0.726
Glyma09g36600.1	365	42343.6	6.12	71.81	-0.727
Glyma10g36050.2	350	40296.9	6.27	50.66	-0.791
Glyma12g00760.1	358	41520.5	5.66	72.12	-0.719
Glyma04g13660.2	361	42248.1	5.97	57.84	-0.934
Glyma19g44910.1	265	30588.6	6.09	64.34	-0.591
Glyma11g07990.2	375	42681.8	5.22	69.89	-0.639
Glyma06g44250.2	285	32385.4	8.41	64.21	-0.864
Glyma13g39090.2	370	41819.7	7.10	60.05	-0.699
Glyma18g49620.3	364	41156.2	5.43	53.90	-0.731
Glyma20g04400.2	280	32132.6	8.15	68.61	-0.610
Glyma13g24320.1	329	37350.8	6.40	68.15	-0.800
Glyma07g05660.1	400	45045.5	6.11	53.40	-0.979
Glyma19g00640.2	429	48158.6	6.02	63.40	-0.717
Glyma04g38560.1	291	33654.1	6.41	66.32	-0.758
Glyma06g38440.2	368	40939.2	6.26	62.77	-0.797
Glyma12g22880.1	340	37983.8	6.01	63.62	-0.478
Glyma02g11140.1	436	48534.0	5.14	69.95	-0.732
Glyma01g22510.1	426	47973.3	4.98	68.17	-0.796
Glyma17g08391.1	370	42837.6	6.43	60.27	-0.851
Glyma16g02200.2	443	49410.5	6.35	57.45	-0.833
Glyma05g04250.2	463	42302.8	6.41	58.08	-0.913
Glyma12g31211.1	371	41788.3	8.50	61.97	-0.603
Glyma05g09110.2	393	44336.2	5.14	65.50	-0.738
Glyma06g47680.1	363	42457.4	6.03	59.89	-0.925
Glyma02g05620.1	362	42022.2	6.10	70.28	-0.706
Glyma01g37310.2	344	39585.2	5.14	67.99	-0.726
Glyma11g03340.1	356	41224.2	6.92	61.35	-0.801
Glyma07g05351.2	497	56740.6	5.77	72.68	-0.592
Glyma19g08510.1	314	35214.9	4.95	66.50	-0.793

续表 1

基因登录号 Gene model ID	氨基酸数目 No. of AA	分子量 Molecular weight/Da	理论等电点 Theoretical pI	脂肪族氨基酸指数 A Lipathic index	蛋白质疏水性 GRAVY
Glyma15g40950.1	318	36938.9	6.89	57.89	-0.959
Glyma20g32690.2	439	49492.6	5.25	79.02	-0.596
Glyma06g16440.1	295	34174.4	6.10	67.42	-0.806
Glyma03g27580.1	197	21399.5	4.34	76.24	-0.703
Glyma06g14250.1	223	24262.5	4.27	69.96	-0.795
Glyma19g30550.1	205	22452.8	4.46	74.68	-0.823
Glyma19g30540.1	202	21998.2	4.40	76.78	-0.721
Glyma04g40550.1	223	24245.5	4.26	68.25	-0.840
Glyma04g40540.1	218	23735.2	4.43	73.85	-0.727
Glyma01g18780.1	159	17405.7	6.16	82.83	-0.588
Glyma12g04383.1	161	17519.8	7.91	78.82	-0.624
Glyma03g27570.1	203	22129.3	4.43	73.99	-0.785
Glyma06g14260.1	225	24276.6	4.33	72.44	-0.752
Glyma08g47520.1	224	25619.0	9.36	67.81	-0.680
Glyma13g35550.1	343	38666.4	8.11	68.40	-0.597
Glyma18g53954.1	229	26173.4	8.99	62.10	-0.791
Glyma16g26740.1	455	51901.0	8.39	71.65	-0.591
Glyma17g00650.2	403	45803.4	6.51	63.18	-0.697
Glyma13g35560.1	375	41433.3	7.82	61.92	-0.722
Glyma04g38990.1	201	23157.9	4.94	71.74	-0.530
Glyma06g21020.1	357	40199.3	8.82	66.89	-0.571
Glyma16g04720.1	407	46089.2	6.20	63.73	-0.820
Glyma18g13574.1	304	35308.2	8.74	64.70	-0.681
Glyma13g18620.2	291	33177.2	6.49	67.01	-0.607
Glyma13g40251.1	349	39939.6	6.91	62.29	-0.812
Glyma05g24910.1	189	21956.3	4.77	65.93	-0.831
Glyma05g38380.1	241	27227.6	9.20	54.94	-0.683
Glyma12g22790.1	360	40050.4	8.27	65.75	-0.741
Glyma12g09670.1	248	28438.9	6.71	61.33	-0.623
Glyma04g08320.6	340	37852.6	9.09	58.21	-0.637
Glyma10g09246.1	218	24507.0	9.05	67.16	-0.403
Glyma12g13710.1	284	32321.3	8.75	62.39	-0.910
Glyma07g31220.1	362	40535.1	7.24	56.33	-0.817
Glyma08g19300.1	190	21917.4	5.14	66.63	-0.756
Glyma10g20830.2	133	15387.8	9.65	76.17	-0.516
Glyma13g31660.2	363	40637.1	8.41	55.84	-0.775
Glyma08g08010.1	342	37535.0	5.58	63.68	-0.507
Glyma07g05360.1	405	45872.1	4.88	64.42	-0.515
Glyma19g02850.1	362	41382.0	6.23	49.34	-0.831
Glyma13g39160.1	253	28450.6	8.85	53.95	-0.761
Glyma10g04350.1	296	33840.8	6.41	65.88	-0.613
Glyma13g36980.1	279	31849.0	8.82	68.06	-0.781
Glyma04g34530.1	76	8790.0	9.14	76.97	-0.639
Glyma07g40140.1	400	45500.0	6.71	63.65	-0.693
Glyma15g08480.3	322	36190.7	6.54	71.74	-0.658
Glyma19g26950.1	215	24286.3	5.26	69.35	-0.560
Glyma02g07700.1	354	40900.2	8.51	67.51	-0.724
Glyma15g07620.1	354	39585.0	8.58	61.38	-0.731
Glyma06g08440.1	355	39576.6	8.48	62.65	-0.621
Glyma08g01280.1	241	27331.7	9.20	56.97	-0.700
Glyma03g35570.1	257	29481.9	6.40	54.20	-0.855
Glyma07g32250.1	326	37090.5	6.43	67.55	-0.846
Glyma16g07501.1	399	44665.4	4.85	71.90	-7.030
Glyma13g30800.3	332	37362.2	8.69	67.83	-0.699
Glyma12g33460.1	279	31818.9	8.93	65.27	-0.804
Glyma08g17351.1	348	39940.9	4.89	61.12	-0.711

续表 1

基因登录号 Gene model ID	氨基酸数目 No. of AA	分子量 Molecular weight/Da	理论等电点 Theoretical pI	脂肪族氨基酸指数 A Liphatic index	蛋白质疏水性 GRAVY
Glyma17g35930.3	279	31336.2	5.46	78.24	-0.399
Glyma09g29760.2	265	30871.7	5.64	56.26	-0.837
Glyma07g10240.1	324	36387.2	8.15	75.19	-0.473
Glyma14g09240.1	278	31253.2	5.31	76.08	-0.416
Glyma05g15533.1	223	26261.8	9.81	58.61	-0.890
Glyma12g34990.1	375	41277.2	8.85	59.60	-0.695
Glyma02g45371.1	320	36904.5	8.57	66.09	-0.789
Glyma09g37050.1	363	41119.1	5.75	58.07	-0.735
Glyma12g31150.1	244	27616.6	6.76	54.75	-0.800
Glyma16g01930.1	431	49684.8	4.91	55.13	-0.761
Glyma13g05540.1	358	40711.4	6.02	51.26	-0.765
Glyma11g18770.1	246	28409.7	6.11	58.62	-0.749
Glyma05g32590.2	206	23046.9	4.73	66.21	-0.414
Glyma08g16630.2	316	35000.3	4.85	71.55	-0.426
Glyma19g02580.1	367	41529.8	6.60	56.68	-0.633
Glyma12g29360.1	349	39983.6	6.96	61.46	-0.844
Glyma04g33270.1	357	40474.5	8.68	64.68	-0.643
Glyma14g20276.1	373	41593.1	8.54	60.94	-0.609
Glyma12g21272.1	212	24159.1	9.43	79.62	-0.221
Glyma16g01940.1	400	45291.2	4.60	68.22	-0.503
Glyma12g18996.1	223	25625.9	9.44	84.80	-0.400
Glyma05g32470.2	317	35026.5	4.78	73.82	-0.371
Glyma07g05370.1	308	34914.3	5.37	67.08	-0.433
Glyma11g10230.1	302	34881.0	8.5	70.66	-0.656
Glyma14g03441.1	326	37654.5	8.72	67.52	-0.768
Glyma17g23740.1	217	24741.4	9.24	65.99	-0.503

2.2 大豆 NAC 蛋白质系统发生树的构建

对预测的大豆 NAC 蛋白序列基于 NAC 的结构域进行分类,根据预测到含 NAC 结构域的氨基酸序列的相似性,对大豆的 152 条氨基酸序列利用 MEGA5.2 软件进行关联分析。构建的系统发生树(图 1)结果显示,152 个 NAC 基因被分成了 10 个亚家族,有些分支很长(如第 X 亚族中的基因),说明这些基因的“祖先”基因在很早就发生了分化,相应基因序列已经发生了大的分化;有些 NAC 基因分支较近(如第 I 亚族、第 II 亚族、第 III 亚族、第 IV 亚族),说明这些基因发生序列变异而出现进化的历史较短,相应的基因功能变化较少,可以认为这几类亚族基因所行使的功能相对类似,进而可以推测那些功能未知的基因功能。在所被划分的 10 个亚族中,第 I 亚族中的 NAC 成员数量最多,为 28 个;第 X 亚族中成员的数量最少,为 10 个。

2.3 大豆 NACF 蛋白质一级结构理化特性分析

利用在线分析工具 Protparam^[22-23] 对大豆 NAC 蛋白的理化特性进行分析,结果显示大豆 NAC 家族蛋白质分子量和氨基酸数目成正相关,组成蛋白质

的氨基酸数目在不同亚族之间具有一定的差异,第 IV 亚族的平均氨基酸数目最多(479 个),第 X 亚族的氨基酸数目最少(202 个)。第 I、V、VII 亚组的理论等电点大多数在碱性范围内,其余亚族的理论等电点大多在酸性范围内(表 1,图 1)。总体看大豆 NAC 家族氨基酸序列的等电点大多数在酸性范围内,蛋白质分子中富含酸性氨基酸。

疏水性是决定蛋白质三维空间构想的重要因素之一,许多转录因子通常会利用其疏水性区与其他蛋白质发生相互作用,而这种互作对于植物的生长调控具有重要的作用^[24]。正值越大表示越疏水,负值越大表示越亲水,介于 0.5 ~ -0.5 的主要为两性氨基酸。152 条氨基酸的疏水性平均值(grand average of hydropathicity, GRAVY)均为负值,表明大豆蛋白均属于亲水蛋白,但它们在亲水程度方面存在差异,其中亲水性最强的(GRAVY 在 -1.0 临近)的基因有 Glyma07g05660.1、Glyma09g36820.1、Glyma12g00540.1、Glyma06g47680.1、Glyma04g13660.2、Glyma15g40950.1、Glyma05g04250.2 和 Glyma12g13710.1,其中 Glyma09g36820.1 的亲水性最强,为 -1.006。

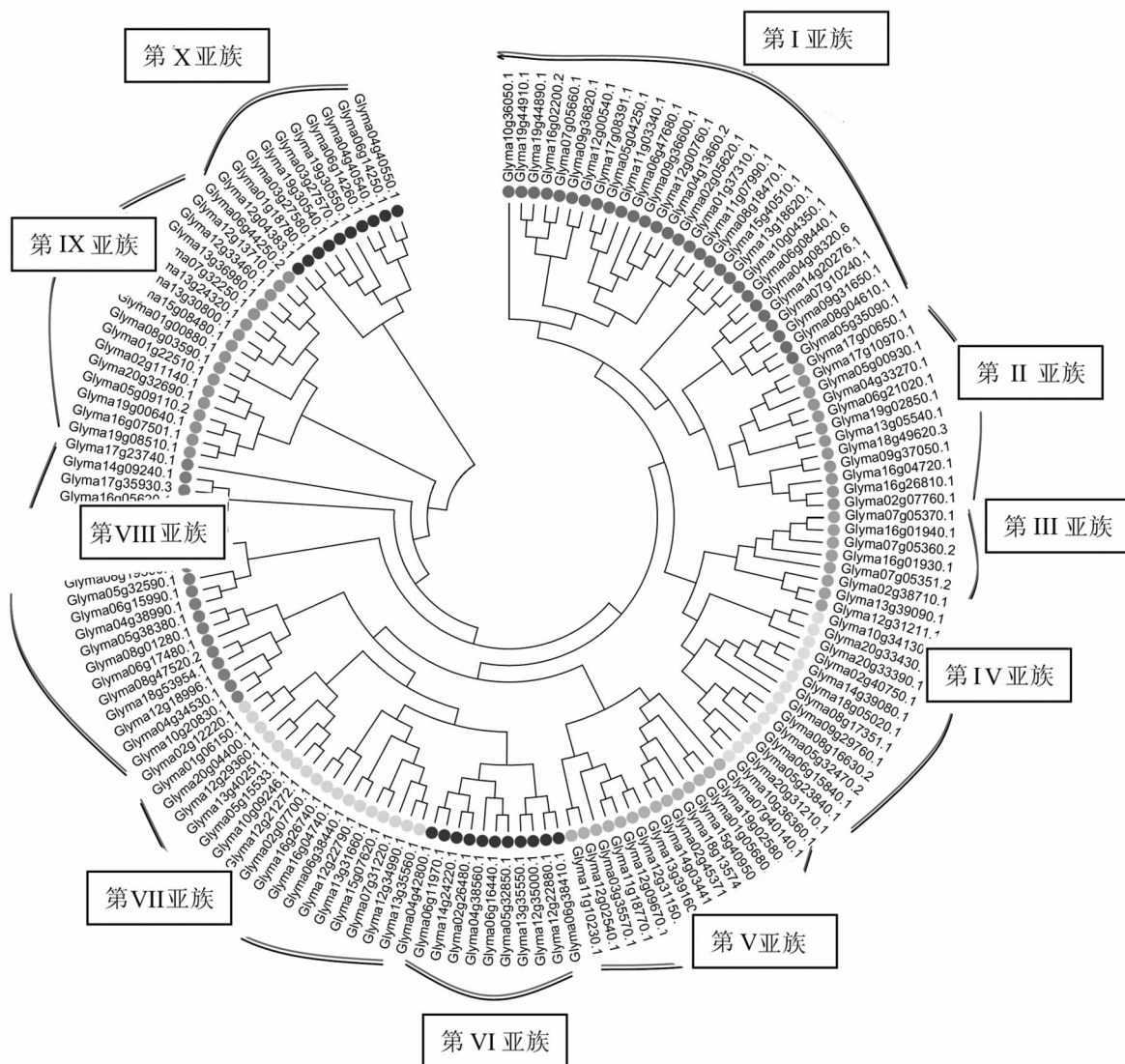


图 1 大豆 NAC 基因家族蛋白质系统进化树

Fig. 1 Phylogenetic tree of NAC gene family in *Glycine max*

2.4 大豆 NAC 蛋白质二级结构预测分析

蛋白质分子的生物学功能由其高级结构决定,其多肽链通常折叠和盘曲呈比较稳定的空间结构,以形成特有的生物学活性和理化性质,大豆 NAC 转录因子 N 端所含的由 151 ~ 159 个氨基酸组成的 NAC 结构域是其最主要的结构特点。因此,蛋白质二级结构分析与预测,对了解其空间结构及功能具有重要意义。蛋白质二级结构有 α -螺旋、 β -转角、无规则卷曲、延伸链等。利用 SOPMA 程序对 152 个大豆 NAC 家族蛋白质的二级结构预测结果显示所有蛋白质的二级结构中以无规则卷曲为主要构成元件,比例均占到 50% 以上,而 α -螺旋为第二主要构成部分, β -转角和延伸链则散布于整个蛋白质中。

2.5 大豆 NAC 家族基因的表达模式分析

对从大豆中预测出的 152 个家族基因在不同组

织和器官中的表达情况进行统计分析,通过文献挖掘共得到 67 个基因的表达组织统计结果。由表 2 可以看出,大豆 NAC 基因在下胚轴、叶、根、茎等营养器官中表达所占比例最大(所占比例依次为 52.2%、65.7%、49.3%、26.9%),在以花为主的生殖器官中(25.8%)、分生组织中(25.4%)、果皮中(37.3%)表达的 NAC 家族基因也占有较大比例;在叶芽和豆荚中表达的 NAC 家族基因数量较少,为 10% 左右。由此进一步证实了大豆 NAC 家族基因在植物生长发育中起调节作用,转录调控植物的根、茎、叶、花等器官与组织的相关表达,在参与植物的营养生长、维持植物生命过程中发挥着至关重要的作用。

表 2 大豆 NAC 基因家族表达组织统计

Table 2 Statistics of expression organs of soybean NAC gene family

转录因子序号 TF ID	基因登录号 Gene model ID	表达 Expression
Gma000102	Glyma09g31650.1	根
Gma000389	Glyma02g26480.1	子叶 下胚轴 叶 豆荚 根 茎
Gma000714	Glyma02g12220.1	子叶 花 下胚轴 叶 根 茎 叶芽
Gma000765	Glyma02g38710.1	子叶 花 叶 分生组织 根 体细胞胚 茎
Gma002354	Glyma06g15840.1	子叶 下胚轴 体细胞胚
Gma002466	Glyma08g16630.1	叶 体细胞胚
Gma004759	Glyma04g38990.1	下胚轴 叶 分生组织 根
Gma004844	Glyma01g06150.1	子叶 花 下胚轴 叶 根
Gma005227	Glyma13g05540.1	种皮
Gma006278	Glyma13g35550.1	子叶 花 下胚轴 叶 分生组织 根 种皮 体细胞胚
Gma007337	Glyma12g02540.1	子叶 花 叶 分生组织 豆荚 根 茎 叶芽
Gma009579	Glyma17g10970.1	子叶 叶 根 种皮
Gma010183	Glyma17g23740.1	花 茎 豆荚
Gma010643	Glyma07g10240.1	根
Gma011496	Glyma08g19300.1	茎
Gma011992	Glyma17g00650.1	叶
Gma012235	Glyma08g47520.1	子叶 下胚轴 叶 根 种皮
Gma012316	Glyma08g18470.1	子叶 花 下胚轴 叶 根
Gma012889	Glyma16g02200.1	下胚轴 叶 根
Gma013288	Glyma13g35560.1	叶 种皮
Gma013795	Glyma11g18770.1	叶
Gma014995	Glyma15g08480.1	子叶 花 下胚轴 根 种皮 体细胞胚 茎
Gma015735	Glyma16g01940.1	下胚轴 叶
Gma016569	Glyma10g34130.1	子叶 上胚轴 下胚轴 叶 分生组织 种皮
Gma017173	Glyma16g04740.1	子叶 花 下胚轴 叶 茎
Gma018414	Glyma10g20830.1	子叶 下胚轴 叶 根 种皮
Gma018485	Glyma13g30800.1	子叶 花 下胚轴 根 种皮 体细胞胚 茎
Gma020219	Glyma12g22880.1	子叶 花 下胚轴 叶 分生组织 种皮
Gma020789	Glyma06g15990.1	下胚轴 叶 分生组织 根
Gma020933	Glyma19g00640.1	体细胞胚
Gma021150	Glyma20g33430.1	子叶 上胚轴 下胚轴 叶 分生组织 种皮
Gma021323	Glyma06g38410.1	子叶 花 下胚轴 叶 分生组织 种皮
Gma022349	Glyma02g05620.1	子叶 花 下胚轴 叶 分生组织 根 种皮 体细胞胚 茎 叶芽
Gma023388	Glyma13g24320.1	叶
Gma023534	Glyma05g38380.1	豆荚
Gma024310	Glyma01g37310.1	子叶 花 下胚轴 叶 分生组织 根 种皮 体细胞胚 茎 叶芽
Gma024440	Glyma12g09670.1	叶
Gma025328	Glyma02g07700.1	花
Gma026036	Glyma20g04400.1	下胚轴 根 体细胞胚
Gma027073	Glyma19g08510.1	体细胞胚
Gma027334	Glyma05g24910.1	茎
Gma027583	Glyma16g26810.1	花 下胚轴 叶 根 体细胞胚
Gma029689	Glyma15g40510.1	下胚轴 叶 根
Gma030330	Glyma10g36360.1	子叶 花 下胚轴 叶 根 种皮 体细胞胚 茎
Gma030461	Glyma07g40140.1	叶
Gma030592	Glyma06g21020.1	种皮
Gma031685	Glyma14g24220.1	子叶 花 叶 分生组织 根 茎
Gma031957	Glyma07g05370.1	下胚轴 叶
Gma031991	Glyma06g11970.1	花 下胚轴 叶 豆荚 根 种皮 叶芽
Gma033114	Glyma01g05680.1	体细胞胚
Gma033412	Glyma20g33390.1	花 叶 分生组织 根 种皮 体细胞胚 茎
Gma033724	Glyma12g35000.1	花 下胚轴 叶 分生组织 种皮
Gma034023	Glyma12g34990.1	叶 种皮
Gma035446	Glyma05g32470.1	叶 体细胞胚
Gma035478	Glyma02g07760.1	花 下胚轴 叶 根 体细胞胚

续表 2

转录因子序号 TF ID	基因登录号 Gene model ID	表达 Expression
Gma037240	Glyma06g16440.1	上胚轴 花 下胚轴 叶 根 体细胞胚 茎
Gma037334	Glyma19g02850.1	种皮
Gma037552	Glyma16g26740.1	叶
Gma038855	Glyma17g35930.1	下胚轴 分生组织 根 种皮 体细胞胚 胚芽
Gma041196	Glyma15g05690.1	茎
Gma041620	Glyma04g42800.1	花 下胚轴 叶 豆荚 根 种皮 叶芽
Gma042174	Glyma04g33270.1	种皮
Gma043548	Glyma08g01280.1	豆荚
Gma043799	Glyma07g05360.1	下胚轴 叶
Gma044667	Glyma01g22510.1	子叶 根
Gma045340	Glyma04g38560.1	子叶 花 下胚轴 叶 分生组织 根 茎 叶芽
Gma045604	Glyma05g00930.1	子叶 叶 根 种皮

3 讨 论

以 NAC 蛋白保守区氨基酸序列为信息探针,从大豆数据库中检索到 NAC 转录因子相应的蛋白序列,成功地利用生物信息学数据库和软件工具,对大豆 NAC 转录因子家族基因结构和蛋白质结构等进行了分析,得到该转录因子家族不同成员之间的相互关系和演化历程,至于其生物学功能,初步通过相关文献挖掘发现大部分的大豆 NAC 基因是与胁迫诱导相关的,进一步的深入功能还需要在大豆中进行克隆以及表达分析来验证。

本研究获得的 152 个大豆 NAC 基因成员,数量大量多于已有研究的水稻、拟南芥等作物^[27],因此可以推测 NAC 基因在大豆发育过程中具有更广泛的作用。通过对预测的大豆 NAC 基因进行同源分析,从中得到该转录因子家族不同成员之间的相互关系和演化历程,从所被划分的 10 个亚组结果可以看出,这些基因在结构和表达上都有不同程度的变化,使其功能既有共性又有独特性。对其二级结构进行分析进一步得知大豆蛋白均属于亲水蛋白,但他们在亲水程度方面存在差异;对大豆 NAC 基因家族表达模式分析发现,NAC 基因家族在大豆中表达组织较为广泛,包含花、子叶、根、茎、下胚轴、分生组织以及一些营养器官^[29]。其深入的生物学功能,还需要在大豆中进行克隆以及表达分析来验证。

参考文献

- [1] Birkenbihl R P, Jach G, Saedler H, et al. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains [J]. *Journal of Molecular Biology*, 2005, 352(3): 585-596.
- [2] Yang R C, Deng C T, Ouyang B, et al. Molecular analysis of two salt-responsive NAC-family genes and their expression analysis in tomato [J]. *Molecular Biology Reports*, 2011, 38(2): 857-863.
- [3] He X J, Mu R L, Cao W H, et al. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development [J]. *The Plant Journal*, 2005, 44(6): 903-916.
- [4] Tran L S, Nakashima K, Sakuma Y, et al. Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter [J]. *Plant Cell*, 2004, 16(9): 2481-2498.
- [5] Sablowski R W, Meyerowitz E M. A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA [J]. *Cell*, 1998, 92(1): 93-103.
- [6] Nikovics K, Thomas B, Alexis P, et al. The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis* [J]. *Plant Cell*, 2006, 18(11): 2929-2945.
- [7] Xie Q, Frugis G, Colgan D, et al. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development [J]. *Genes & Development*, 2000, 14(23): 3024-3036.
- [8] Ko J H, Yang S H, Park A H, et al. ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana* [J]. *Plant Journal*, 2007, 50(6): 1035-1048.
- [9] Mitsuda N, Seki M, Shinozaki K, et al. The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence [J]. *Plant Cell*, 2005, 17(11): 2993-3006.
- [10] Yoo S Y, Kin Y, Kin S Y, et al. Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis* [J]. *PLoS One*, 2007, 2(7): 642.
- [11] Jensen M K, Hagedorn P H, de Torres-Zabala M, et al. Transcriptional regulation by an NAC (NAM-ATAF1, 2-CUC2) transcription factor attenuates ABA signalling for efficient basal defence towards *Blumeria graminis* f. sp. *hordei* in *Arabidopsis* [J]. *Plant Journal*, 2008, 56(6): 867-880.
- [12] Bu Q, Jiang H, Li C B, et al. Role of the *Arabidopsis thaliana* NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses [J]. *Cell Research*, 2008, 18(7): 756-767.
- [13] Kim S G, Lee A K, Yoon H K, et al. A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt sig-

- naling in *Arabidopsis* seed germination[J]. Plant Journal, 2008, 55 (1): 77-88.
- [14] Kato H, Motomura T, Komeda Y, et al. Overexpression of the NAC transcription factor family gene ANAC036 results in a dwarf phenotype in *Arabidopsis thaliana* [J]. Journal of Plant Physiology, 2010, 167(7): 571-577.
- [15] Shahnejat-Bushehri S, Mueller-Roeber B, Balazadeh S. *Arabidopsis* NAC transcription factor JUNGBRUNNEN1 affects thermomemory-associated genes and enhances heat stress tolerance in primed and unprimed conditions [J]. Plant Signaling & Behavior, 2012, 7 (12): 1518-1521.
- [16] Distelfeld A, Pearce S P, Avni R, et al. Divergent functions of orthologous NAC transcription factors in wheat and rice [J]. Plant Molecular Biology, 2012, 78(4-5): 515-524.
- [17] Nuruzzaman M, Manimekalai R, Sharoni A M, et al. Genome-wide analysis of NAC transcription factor family in rice [J]. Gene, 2010, 465(1-2): 30-44.
- [18] 刘洋, 张慧, 辛大伟, 等. 大豆 TCP 转录因子家族结构域分析及功能预测 [J]. 大豆科学, 2012, 31(5): 707-717. (Liu Y, Zhang H, Xin D W, et al. Domain analysis and function prediction of TCP transcription factors family in soybean [J]. Soybean Science, 2012, 31(5): 707-717.)
- [19] Ooka H, Satoh K, Doi K, et al. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana* [J]. DNA Research, 2003, 10(6): 239-247.
- [20] 李乐, 许红亮, 杨兴露, 等. 大豆 LEA 基因家族全基因组鉴定、分类和表达 [J]. 中国农业科学, 2011, 44(19): 3945-3954. (Li L, Xu H L, Yang X L, et al. Genome-wide identification, classification and expression analysis of LEA gene family in soybean [J]. Scientia Agricultura Sinica, 2011, 44(19): 3945-3954.)
- [21] 陈莹, 孙霞, 胡尚连, 等. 拟南芥次生长相关 NAC 转录因子保守结构域分析 [J]. 西北农林科技大学学报, 2009, 37(5): 185-194. (Chen Y, Sun X, Hu S L, et al. Analysis of NAC transcription factor and its domain with secondary growth in *Arabidopsis thaliana* [J]. Journal of Northwest A&F University, 2009, 37(5): 185-194.)
- [22] 江董丽, 才华, 端木慧子, 等. 大豆 GST 基因家族全基因组筛选、分类和表达 [J]. 分子植物育种, 2013, 11(5): 465-475. (Jiang D L, Cai H, Duanmu H Z, et al. Genome-wide filter, classification and expression analysis of GST gene family in soybean [J]. Molecular Plant Breeding, 2013, 11(5): 465-475.)
- [23] Zhang C H, Liu H, Yu M L, et al. Bioinformatics analysis for the NAC gene family in strawberry [J]. Genomics and Applied Biology, 2011, 30(41): 1261-1271.
- [24] 蒋瑶, 陈其兵. 植物 CBF1 转录因子的生物信息学分析 [J]. 林业科学, 2010, 46(6): 43-50. (Jiang Y, Chen Q B. Bioinformatic analysis of CBF1 transcription factors from the plants [J]. Scientia Silvae Sinicae, 2010, 46(6): 43-50.)
- [25] Olsen A N, Ernst H A, Leggio L L, et al. NAC transcription factors: structurally distinct, functionally diverse [J]. Trends in Plant Science, 2005, 10(2): 79-87.
- [26] Strauch M A, Ballar P, Rowshan A J, et al. The DNA-binding specificity of the *Bacillus anthracis* AbrB protein [J]. Microbiology, 2005, 151(6): 1751-1759.
- [27] Peng H, Yu X W, Cheng H Y, et al. A survey of functional studies of the plant-specific NAC transcription factor family [J]. Chinese Bulletin of Botany, 2010, 45(2): 236-248.
- [28] Wang N, Zheng Y, Xin H, et al. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera* [J]. Plant Cell Reports, 2013, 32(1): 61-75.
- [29] Nuruzzaman M, Sharoni A M, Satoh K, et al. Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64 [J]. Molecular Genetics Genomics, 2012, 287(5): 389-410.

(上接第 324 页)

参考文献

- [1] Zhu J K. Active DNA demethylation mediated by DNA glycosylases [J]. Annual Review of Genetics, 2009, 43: 143-166.
- [2] Finnegan E J, Genger R K, Peacock W J, et al. DNA methylation in plants [J]. Annual Review of Plant Biology, 1998, 49(1): 223-247.
- [3] Wada Y, Miyamoto K, Kusano T, et al. Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants [J]. Molecular Genetics and Genomics, 2004, 271: 658-666.
- [4] Pan Y J, Fu B Y, Wang D, et al. Spatial and temporal profiling of DNA methylation induced by drought stress in rice [J]. Scientia Agricultura Sinica, 2009, 42: 3009-3018.
- [5] Finnegan E J, Peacock W J, Dennis E S. DNA methylation, a key regulator of plant development and other processes [J]. Current Opinion in Genetics & Development, 2000, 10(2): 217-223.
- [6] Jost J P, Saluz H P. DNA methylation: Molecular biology and biological significance [M]. Basel: Birkhauser Verlag, 1993: 218-261.
- [7] 方淑梅, 梁喜龙, 纪伟波, 等. 大豆 DNA 甲基化酶生物信息学分析 [J]. 中国油料作物学报, 2013, 35(2): 225-230. (Fang S M, Liang X L, Ji W B, et al. Bioinformatics analysis of DNA methylase from soybean [J]. Chinese Journal of Oil Crop Sciences, 2013, 35(2): 225-230.)
- [8] 范晓军, 宋志芳, 仙笑笑, 等. 金纹细蛾几丁质酶基因生物信息学分析 [J]. 基因组学与应用生物学, 2013, 32(1): 91-96. (Fan X J, Song Z F, Xian X X, et al. Bioinformatic analysis of chitinase gene from *Lithocolletics ringoniella* [J]. Genomics and Applied Biology, 2013, 32(1): 91-96.)
- [9] Agius F, Kapoor A, Zhu J K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation [J]. Proceedings of the National Academy of Sciences, 2006, 103(31): 11796-11801.
- [10] Gehring M, Reik W, Henikoff S. DNA demethylation by DNA repair [J]. Trends in Genetics, 2009, 25(2): 82-90.
- [11] Gong Z, Morales-Ruiz, Ariza R R, et al. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA Glycosylase/Lyase [J]. Cell, 2002, 111(6): 803-814.