

番茄NAC转录因子家族的鉴定及生物信息学分析

陈秀玲¹, 王傲雪^{1,2,*}, 张珍珠^{1,2}, 李景富^{1,*}

东北农业大学¹园艺学院, ²生命科学学院, 哈尔滨 150030

摘要: NAC转录因子家族是植物特有的一类转录因子, 在植物的生长发育、器官建成及逆境胁迫和激素信号应答中均发挥重要作用。本研究在基因组范围内, 利用生物信息学方法对番茄的NAC转录因子家族成员、分布及结构和功能等进行分析。预测结果显示番茄NAC家族包含102个蛋白质, 分为12亚族, 其中茄属特有的TNAC亚族中成员最多, 具有25个, 其他NAC转录因子与拟南芥NAC家族具有相似分类。保守基序分析, 在番茄NAC结构域中包含7个保守的NAM基序, 主要分布在序列的N端, 表明这些基序的存在对NAC蛋白质功能的执行是必需的。理化性质和结构分析表明, 番茄NAC蛋白质绝大多数是亲水蛋白质, 主要以无规则卷曲构成, 而 α -螺旋、 β -折叠和 β -转角则散布于整个蛋白质中, 在各亚族中没有规律。

关键词: 番茄; NAC转录因子; 生物信息学

Genome-Wide Identification and Bioinformatics Analysis of NAC Gene Family in Tomato

CHEN Xiu-Ling¹, WANG Ao-Xue^{1,2,*}, ZHANG Zhen-Zhu^{1,2}, LI Jing-Fu^{1,*}

¹School of Horticulture, ²School of Life Science, Northeast Agricultural University, Harbin 150030, China

Abstract: NAC proteins belong to one of the largest plant-specific transcription factor families and play diverse roles in plant development processes, organ formation, response to biotic or abiotic cues, and hormone signaling. NAC proteins of tomato were genome-wide identified, the distribution on the chromosome, the structure and function of protein were analyzed. 102 predicted proteins constitute the tomato NAC transcription factor family, which can be clustered into 12 groups. 25 of them belong to TNAC subfamily (a novel NAC subfamily which was first identified in tobacco), and other members have the similar classification with *Arabidopsis thaliana*. Conservative motif analysis indicates that there are seven conserve NAM motif in N terminal. The physical and chemical properties and structure analysis showed that most of the tomato NAC proteins are hydrophilic proteins, the secondary structure of 102 predicted NAC proteins was mainly random coil, and alpha helix, beta sheet and beta turn are scattered to the whole protein sequence, there is no rule in each subfamily.

Key words: tomato; NAC transcription factor; bioinformatics

NAC (NAM/ATAF/CUC)转录因子家族是植物特有的一类转录因子, 也是目前发现的最大的转录因子家族之一。1996年Souer等在矮牵牛(*petunia hybrida* Vilm.)中克隆得到第一个NAC基因NAM, nam缺失突变体的幼胚缺少根尖分生组织, 不能正常分化出根和叶。随后, Aida等(1997)在拟南芥(*Arabidopsis thaliana*)中发现了具有类似功能的CUC2基因, 并研究发现在矮牵牛NAM、拟南芥ATAT1/2和CUC2编码蛋白的N端都有一段保守的氨基酸序列, 被命名为NAC结构域, 包含该结构域的蛋白组成了NAC转录因子家族。

NAC转录因子家族在植物的生长发育调控中发挥了重要作用, 例如调控顶端分生组织发育(Ooka等2003; Kim等2007; Fang等2008; Zhang等

2011)、细胞分化(Kim等2006)、子叶形成(Aida等1997)、侧根发生(He等2005)、花发育(Sablowski和Meyerowitz 1998)、激素信号传导(Greve等2003; Jensen等2008)、植物器官衰老(Liu等2009)、胚胎发育(Duval等2002)及次级细胞壁形成(Zhong等2008)和纤维的形成(Ko等2007)等。此外, NAC转录因子还具有响应病原菌病原菌感染(Xie等1999;

收稿 2013-12-17 修定 2014-03-05

资助 国家自然科学基金项目(31301780)、黑龙江省教育厅海外学人项目(1252HQ011)、黑龙江省博士后基金(LBH-Z12044)和东北农业大学“寒地蔬菜生物学重点实验室”开放基金(CVB2012-05)。

* 共同通讯作者(E-mail: axwang@neau.edu.cn, Tel: 0451-55191671; E-mail: lijf_2005@126.com, Tel: 0451-55191671)。

Ren等2000; Nakashima等2007)和逆境胁迫应答(He等2005; Hu等2006; Bhatnagar-Mathur等2007; Yoo等2007; Tran等2009)的作用,当植物受到生物胁迫和非生物胁迫时,会诱导大量NAC转录因子成员的产生,从而调控植物进行胁迫响应(Seki等2003; Tran等2004; Delessert等2005)。

随着分子生物学和生物信息学的不断发展,越来越多的NAC转录因子家族成员被鉴定或预测出来,特别是近几年全基因组测序工作的开展和新一代测序技术的应用。目前拟南芥、水稻(*Oryza sativa*)、玉米(*Zea mays*)、小麦(*Triticum aestivum*)、毛果杨(*Populus trichocarpa*)、高粱(*Sorghum bicolor*)、马铃薯(*Solanum tuberosum*)等作物中的NAC转录因子家族均被预测或鉴定出来(Ooka等2003; Fang等2008; Hu等2010; Nuruzzaman等2010; Zhang等2011; Singh等2013),并且不断更新完善中。

番茄是一种极具营养价值和经济价值的世界性蔬菜作物,而且也逐渐成为进化、遗传等方面研究的模式植物。番茄基因组测序完成和数据的释放,使在基因组水平对番茄重要基因家族及重要功能基因的挖掘、分析成为可能,为功能基因组学提供重要的研究数据。本研究利用生物信息学方法对番茄NAC转录因子家族进行预测、挖掘及定位分析,为番茄NAC转录因子家族的功能分析提供基础信息,从而对品种改良提供候选基因。

材料与方法

1 番茄NAC转录因子家族序列检索

根据拟南芥中已经鉴定出来的NAC基因及其编码的蛋白质序列,构建隐马氏模型序列,利用Hmmer3.1b1软件对从SOL (SGN, <http://solgenomics.net>, release v2.40)下载的番茄(*Lycopersicon esculentum*)蛋白数据进行检索和去冗余,得到候选蛋白序列。将这些候选序列利用Pfam (<http://pfam.janelia.org>) (Punta等2012)和SMART (<http://smart.embl-heidelberg.de/>) (Letunic等2012)在线分析NAM结构域有无,同时利用NCBI (<http://www.ncbi.nlm.nih.gov/>)、UniProt (www.uniprot.org)和PlantTFDB v3.0 (<http://planttfdb.cbi.edu.cn/>)数据库

对候选序列进行进一步鉴定分析。此外,分析候选序列的蛋白长度、染色体定位、内含子个数和分子量等基本信息,并利用在线工具ProtParam (<http://expasy.org/tools/protparam.html>) (Gasteiger等2005)进行等电点分析,Protcomp 9.0 (<http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc>)和CELLO v2.5 (<http://cello.life.nctu.edu.tw/>) (Yu等2006)进行亚细胞定位分析。

2 染色体定位分析

根据检索到的番茄NAC转录因子基因组信息,利用番茄基因组数据(SGN, <http://solgenomics.net>, release v2.40),对番茄NAC转录因子进行染色体定位分析。

3 番茄NAC蛋白质系统发生分析

从TAIR10 (<http://www.arabidopsis.org>)数据库中下载拟南芥的NAC蛋白全长序列,将候选番茄NAC家族蛋白序列与拟南芥蛋白序列用clustalX2软件进行多重匹配分析,利用MEGA5.05 (Tamura等2011)采用相邻连接法(neighbor-joining, NJ)构建系统发育树,对构建的树进行自检(bootstrap),重复设定为1 000。

4 保守基序的鉴定及结构分析

利用Meme 4.6.1 (<http://meme.nbcr.net/meme/intro.html>)在线平台进行番茄NAC蛋白中保守基序(motif)分析,最大motif检索数值定为12。利用Ex-PaSy提供的在线工具Protparam软件分析番茄NAC的蛋白质一级结构(<http://web.expasy.org/prot-param/>);二级结构预测采用SOPMA 2.0软件(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html),预测分析 α -螺旋、 β -折叠和无规则卷曲;三级结构采用Phyre 2.0 (<http://www.sbg.bio.ic.ac.uk/~phyre/>)进行同源建模分析,并利用在线工具CombFunc (<http://www.sbg.bio.ic.ac.uk/~mwass/combfunc/>)中的GO分析对NAC蛋白的功能进行预测。

实验结果

1 番茄NAC转录因子家族鉴定

利用番茄基因组数据库SGN v2.40对番茄NAC转录因子家族进行鉴定,结合Pfam和SMART

在线分析结构域, 共得到102个含有NAM结构域的 NAC蛋白质序列(表1)。其中番茄NAC蛋白长度为 109~1 029 aa, 平均长度为330 aa; 分子量为 12.52~116.99 kDa, 平均分子量为37.25 kDa。

表1 番茄NAC转录因子家族分布及理化性质分析

Table 1 The distribution of NAC transcription factor family on the scaffolds of genome and physico-chemical analysis in tomato

基因	SGN 位点	染色体	位置	长度/ aa	内含 子数	亚细胞定位	分子量/ kDa	等电点	脂肪族 氨基酸 指数	蛋白质 疏水性
<i>SINAC01</i>	Solyc01g009860	I	4384348..4382336	272	3	1	31.34	9.47	66.55	-0.803
<i>SINAC02</i>	Solyc01g021730	I	32778063..32776998	264	0	1、5、6	30.30	6.36	59.05	-0.759
<i>SINAC03</i>	Solyc01g094490	I	77709650..77710279	209	0	2、6、1	24.58	5.45	76.89	-0.857
<i>SINAC04</i>	Solyc01g102740	I	83236708..83238761	413	2	1	46.93	6.15	68.72	-0.811
<i>SINAC05</i>	Solyc01g104900	I	84979448..84978258	198	2	1	22.85	4.91	68.84	-0.713
<i>SINAC06</i>	Solyc02g036430	II	21468369..21472137	565	5	1	24.58	5.45	76.89	-0.857
<i>SINAC07</i>	Solyc02g061780	II	27876587..27874513	244	2	1	28.23	7.03	71.56	-0.645
<i>SINAC08</i>	Solyc02g061870	II	28013091..28012432	219	0	1、6、2	25.21	6.60	72.88	-0.560
<i>SINAC09</i>	Solyc02g061900	II	28028950..28028036	304	0	1、5、2	34.60	5.25	74.97	-0.554
<i>SINAC10</i>	Solyc02g061910	II	28046382..28045588	264	0	7、6	29.91	5.99	80.76	-0.429
<i>SINAC11</i>	--	II	28055471..28053934	451	1	1、2、6	50.69	5.06	83.17	-0.336
<i>SINAC12</i>	Solyc02g062060	II	28223809..28224429	206	0	1、6、3、2	23.93	6.45	69.47	-0.666
<i>SINAC13</i>	Solyc02g062070	II	28231047..28232074	305	1	4、8、1、5	34.52	5.07	79.77	-0.366
<i>SINAC14</i>	Solyc02g062090	II	28249515..28250305	182	1	4、2、	20.71	4.87	84.01	-0.418
<i>SINAC15</i>	Solyc02g062210	II	28380350..28379709	214	0	1、2、5	24.41	5.56	65.56	-0.680
<i>SINAC16</i>	Solyc02g069960	II	34318868..34315619	310	2	1	35.72	5.95	65.97	-0.795
<i>SINAC17</i>	Solyc02g070140	II	34533661..34534248	195	0	1、6、3	23.35	9.20	62.41	-0.982
<i>SINAC18</i>	Solyc02g077610	II	37077632..37073622	292	3	1	33.10	8.81	76.99	-0.635
<i>SINAC19</i>	Solyc02g081270	II	39877970..39887403	1 029	16	1、4、2	116.99	5.45	74.46	-0.448
<i>SINAC20</i>	Solyc02g084350	II	42043859..42041409	349	2	1	40.13	5.01	66.50	-0.840
<i>SINAC21</i>	Solyc02g087920	II	44750151..44751604	383	2	1	43.37	6.48	60.39	-0.663
<i>SINAC22</i>	Solyc02g088180	II	44915098..44916628	319	2	1	36.31	6.51	61.72	-0.691
<i>SINAC23</i>	Solyc02g093420	II	48873416..48876572	393	5	1	43.86	5.28	62.75	-0.789
<i>SINAC24</i>	Solyc03g059300	III	25888463..25887780	227	0	1、8、5、6	26.57	9.18	72.11	-0.648
<i>SINAC25</i>	Solyc03g062670	III	29295085..29295933	282	0	1	33.03	4.81	75.60	-0.630
<i>SINAC26</i>	Solyc03g062750	III	29437208..29436840	122	0	4、2	14.43	9.90	47.19	-0.770
<i>SINAC27</i>	Solyc03g078120	III	43642171..43637430	150	3	1	17.85	9.30	75.33	-0.442
<i>SINAC28</i>	Solyc03g080090	III	45541347..45546345	629	5	2、1、4	69.85	4.87	72.35	-0.591
<i>SINAC29</i>	Solyc03g083880	III	47361452..47364301	312	2	1	36.45	6.35	67.40	-0.850
<i>SINAC30</i>	Solyc03g097650	III	53428478..53430021	256	2	1	29.33	8.33	70.47	-0.673
<i>SINAC31</i>	Solyc03g098190	III	53982305..53977592	249	2	1	28.96	5.10	63.41	-0.785
<i>SINAC32</i>	Solyc03g114220	III	58265021..58264293	242	0	1、7	28.14	5.12	65.99	-0.743
<i>SINAC33</i>	Solyc03g114260	III	58344558..58345355	265	0	1、2、7	31.14	5.41	69.47	-0.707
<i>SINAC34</i>	Solyc03g115850	III	59504979..59503805	326	2	1	37.55	8.53	66.15	-0.683
<i>SINAC35</i>	Solyc04g005610	IV	405228..406656	275	2	1、8	31.88	8.56	62.69	-0.730
<i>SINAC36</i>	Solyc04g009440	IV	2856469..2858639	301	3	1	34.84	7.62	58.70	-0.837
<i>SINAC37</i>	Solyc04g015960	IV	6506971..6509066	282	1	1、2	32.59	9.53	71.81	-0.822
<i>SINAC38</i>	Solyc04g025760	IV	22330393..22329710	227	0	2、9	26.88	5.57	73.79	-0.800
<i>SINAC39</i>	Solyc04g072220	IV	56810161..56813229	599	2	1	67.09	4.80	67.21	-0.549
<i>SINAC40</i>	Solyc04g078670	IV	60958460..60960829	272	2	6、3、5	30.07	6.24	82.39	-0.236
<i>SINAC41</i>	Solyc04g079940	IV	61857677..61853024	390	5	1	43.50	5.26	65.77	-0.682
<i>SINAC42</i>	Solyc05g007550	V	2096116..2098418	390	2	1	44.01	6.60	66.54	-0.569
<i>SINAC43</i>	Solyc05g007770	V	2229720..2231644	282	2	1	32.50	6.84	70.11	-0.631

续表

基因	SGN 位点	染色体	位置	长度/ aa	内含 子数	亚细胞定位	分子量/ kDa	等电点	脂肪族 氨基酸 指数	蛋白质 疏水性
<i>SINAC44</i>	Solyc05g009840	V	4063872..4058802	435	5	1	48.99	4.86	62.97	-0.891
<i>SINAC45</i>	Solyc05g010740	V	4936601..4938374	305	2	1	34.39	9.42	69.31	-0.587
<i>SINAC46</i>	Solyc05g021090	V	26343121..26349617	279	2	1	32.57	7.11	54.84	-0.862
<i>SINAC47</i>	Solyc05g026180	V	40024726..40030220	493	6	1	55.36	6.38	60.55	-0.841
<i>SINAC48</i>	Solyc05g055470	V	64254800..64249833	544	5	1	60.76	4.68	68.14	-0.526
<i>SINAC49</i>	Solyc05g055480	V	64259429..64255495	386	3	1	43.47	5.43	69.46	-0.682
<i>SINAC50</i>	--	V	64900437..64903343	221	3	1、8	25.69	9.18	82.94	-0.454
<i>SINAC51</i>	Solyc06g008360	VI	2253973..2260316	282	2	1	32.41	5.66	67.06	-0.601
<i>SINAC52</i>	Solyc06g034340	VI	20776461..20774519	340	2	1、9、8	39.50	6.23	62.82	-0.868
<i>SINAC53</i>	Solyc06g060230	VI	34578909..34577474	296	2	1	33.92	6.76	62.30	-0.712
<i>SINAC54</i>	Solyc06g061080	VI	35499147..35501008	337	3	1	38.11	5.70	58.07	-0.630
<i>SINAC55</i>	Solyc06g063380	VI	36444430..36442937	497	0	2、1	56.99	4.58	73.76	-0.626
<i>SINAC56</i>	Solyc06g063430	VI	36500202..36497407	318	2	1	36.86	7.15	54.62	-0.996
<i>SINAC57</i>	Solyc06g065410	VI	37220630..37222822	322	2	1	37.57	6.01	59.35	-0.894
<i>SINAC58</i>	Solyc06g068580	VI	38899104..38900356	350	2	2、4	39.95	4.82	68.74	-0.709
<i>SINAC59</i>	Solyc06g069100	VI	39281353..39276715	497	5	1	56.36	6.13	62.17	-0.799
<i>SINAC60</i>	Solyc06g069710	VI	39810628..39809375	339	2	1	38.05	8.07	73.01	-0.404
<i>SINAC61</i>	Solyc06g073050	VI	41395982..41399877	595	5	4、8、5	65.88	5.17	73.53	-0.505
<i>SINAC62</i>	Solyc06g074170	VI	42257940..42254209	409	3	5、4	46.25	8.86	71.69	-0.552
<i>SINAC63</i>	--	VI	45438598..45439877	153	1	1	17.87	8.40	78.37	-0.393
<i>SINAC64</i>	Solyc06g083850	VI	45439537..45440528	109	1	7、1	12.52	9.84	59.91	-0.812
<i>SINAC65</i>	Solyc07g006840	VII	1707455..1710527	483	2	2、1	55.51	4.97	54.84	-0.975
<i>SINAC66</i>	Solyc07g045030	VII	55436220..55441033	331	2	1	38.45	9.05	65.32	-0.763
<i>SINAC67</i>	Solyc07g053590	VII	59335327..59332888	271	2	1、7、4、6	30.78	9.38	73.65	-0.689
<i>SINAC68</i>	Solyc07g053680	VII	59444996..59444350	179	1	7、1	20.86	5.13	67.43	-1.116
<i>SINAC69</i>	Solyc07g062240	VII	62252409..62253224	213	1	2、1、5、4	24.92	9.06	80.52	-0.740
<i>SINAC70</i>	Solyc07g062840	VII	62708889..62711033	352	2	1	39.51	8.40	60.71	-0.514
<i>SINAC71</i>	Solyc07g063410	VII	63107439..63109016	356	2	1	40.09	7.02	63.51	-0.623
<i>SINAC72</i>	Solyc07g063420	VII	63124354..63127164	329	2	1	37.01	8.56	61.95	-0.783
<i>SINAC73</i>	Solyc07g066330	VII	65024400..65019493	342	2	1	38.83	7.60	67.28	-0.625
<i>SINAC74</i>	Solyc08g006020	VIII	784730..788036	302	2	1	34.86	9.16	49.70	-0.998
<i>SINAC75</i>	Solyc08g007020	VIII	1589496..1592028	168	3	1	19.21	5.28	47.56	-0.856
<i>SINAC76</i>	Solyc08g008660	VIII	3026965..3025338	349	2	1	40.18	5.88	65.36	-0.838
<i>SINAC77</i>	Solyc08g028850	VIII	21634238..21634564	226	0	6、5、4、7	25.81	4.89	70.71	-0.454
<i>SINAC78</i>	Solyc08g068380	VIII	54692200..54696753	278	2	2、5	32.05	9.35	63.09	-0.610
<i>SINAC79</i>	Solyc08g074300	VIII	55580998..55577743	341	2	1	38.43	6.64	57.71	-0.749
<i>SINAC80</i>	Solyc08g077110	VIII	58202853..58200673	340	3	1	38.19	5.27	55.59	-0.608
<i>SINAC81</i>	Solyc08g079120	VIII	59930002..59926058	307	3	1	35.95	8.61	69.25	-0.796
<i>SINAC82</i>	Solyc09g010160	IX	3540687..3545497	249	2	1	28.94	8.86	64.18	-0.873
<i>SINAC83</i>	Solyc09g025310	IX	19328080..19323700	270	2	1	31.37	5.43	68.22	-0.656
<i>SINAC84</i>	Solyc10g005010	X	12389..11300	287	1	1	33.54	6.36	64.84	-0.885
<i>SINAC85</i>	Solyc10g006880	X	1308510..1305620	355	2	1	40.27	8.34	67.52	-0.790
<i>SINAC86</i>	Solyc10g047060	X	34778791..34780357	212	2	1	24.65	5.75	62.92	-0.833
<i>SINAC87</i>	Solyc10g055760	X	52871391..52872402	283	2	1	32.56	6.72	73.36	-0.583
<i>SINAC88</i>	Solyc10g079220	X	60126209..60125137	327	1	7、3	37.03	5.15	69.63	-0.588
<i>SINAC89</i>	Solyc10g083450	X	62573756..62570813	289	2	1	34.37	5.73	60.73	-0.954
<i>SINAC90</i>	Solyc11g005920	XI	743287..748004	687	5	2、5	76.25	4.87	72.82	-0.375
<i>SINAC91</i>	Solyc11g008000	XI	2208572..2205263	528	5	1	59.47	4.75	67.25	-0.601
<i>SINAC92</i>	Solyc11g008010	XI	2208572..2205263	465	5	4、5、7	52.47	5.81	77.35	-0.420

续表

基因	SGN 位点	染色体	位置	长度/ aa	内含 子数	亚细胞定位	分子量/ kDa	等电点	脂肪族 氨基酸 指数	蛋白质 疏水性
<i>SINAC93</i>	Solyc11g017470	XI	8423170..8422146	291	2	1	33.46	6.67	62.37	-0.691
<i>SINAC94</i>	Solyc11g018660	XI	8887768..8890395	311	2	1	36.06	5.93	60.19	-0.886
<i>SINAC95</i>	Solyc11g022480	XI	14024569..14024183	128	0	7、3、4	14.90	6.50	61.64	-0.571
<i>SINAC96</i>	Solyc11g065540	XI	48127492..48129178	309	2	1、7	35.99	5.20	66.57	-0.820
<i>SINAC97</i>	Solyc11g068620	XI	50352818..50350576	260	2	1	30.39	8.75	71.19	-0.621
<i>SINAC98</i>	Solyc11g068750	XI	50465592..50470925	452	3	3、1	51.63	9.03	77.83	-0.537
<i>SINAC99</i>	Solyc12g013620	XII	4464637..4462011	349	2	1	39.48	8.53	65.64	-0.729
<i>SINAC100</i>	Solyc12g017400	XII	6652079..6650295	319	2	1	36.60	7.22	65.11	-0.722
<i>SINAC101</i>	Solyc12g036480	XII	31383643..31380780	326	2	1	37.08	6.25	67.02	-0.668
<i>SINAC102</i>	Solyc12g056790	XII	48227553..48230632	590	2	1、7	66.50	4.90	70.08	-0.511

--: 无对应SGN位点; 数字代表亚细胞定位, 1: 细胞核; 2: 细胞质; 3: 质膜; 4: 叶绿体; 5: 高尔基体; 6: 线粒体; 7: 细胞外; 8: 液泡; 9: 内质网。

2 染色体定位分析

102个NAC基因分布在番茄12条染色体上, 染色体II上含有的NAC基因最多, 达到18个; 其次是染色体VI, 含有14个NAC基因, NAC分布最少的是染色体IX, 只有2个(图1)。番茄特有的NAC基因在除V、IX、XII的染色体上均有分布, 其中, II号染色体上最多, 有8个。在番茄基因组中, 存在8个NAC基因簇, 第II染色体包含2个基因簇, 其他6个基因簇分别位于III、IV、V、VI、VII和XI染色体上。

根据系统发育树及基因定位分析发现, NAC转录因子家族在进化过程中, 存在片段复制和串联复制现象。共发现26对片段复制, 如SINAC29/52、SINAC57/94、SINAC45/79、SINAC34/60、SINAC72/85、SINAC35/43等, 在12条染色体上均有分布, 尤以II、III、V、VI和XI居多。SINAC06-08/SINAC12-19、SINAC21-23/SINAC42-44、SINAC48-49/SINAC91-92、SINAC28-29-34/SINAC61-52-60和SINAC53-57-62/SINAC93-94-98为串联复制, 分别位于II、III、V、VI和XI染色体上。

3 番茄NAC蛋白质系统发生分析

将预测出的番茄NAC蛋白质利用MEGA 5.05软件, 构建系统进化树(图2-A)。将番茄NAC蛋白分成12个亚族, 每个亚族中包含1~25个NAC蛋白序列不等, XII亚族含有的NAC蛋白序列最多, 有25个; 其次是I亚族, 含有15个NAC蛋白序列; 第

VIII、X、XI亚族含有的成员最少, 分别只含有1个NAC蛋白序列。

为了进一步分析番茄NAC蛋白质的多样性, 利用Meme 4.6.1在线工具分析了保守motif(图2-B)。通常聚在一个亚族的蛋白质会有相似的motif组成, 在NAC结构域中包含7个保守的NAM基序, 主要分布在序列的N端, 表明这些基序的存在对NAC蛋白质功能的执行是必需的。Motif 3、5、1、2、4分别代表NAC类转录因子A、B、C、D、E亚结构域。

为了研究番茄中NAC转录因子的进化关系, 对番茄和拟南芥NAC基因家族氨基酸进行系统进化树分析(图3)。番茄XII亚族NAC蛋白质并没有和拟南芥任一NAC蛋白聚到一类, 该亚族中又可以分为4个小族, 分别包含5、5、5和10个成员。番茄XII亚族蛋白质是茄属作物特有的(Singh等2013), 而拟南芥也有一类NAC蛋白质是没有与番茄NAC蛋白质聚到一起的, 表明这些NAC蛋白在拟南芥和番茄进化中具有重要意义。

4 番茄NAC蛋白质结构分析

利用ProtParam软件对番茄NAC蛋白质的理化特性进行了分析(表1), 表明蛋白质分子量与氨基酸个数呈正相关, 等电点为4.58~9.9。蛋白质疏水性在-0.236与-1.116之间, 蛋白质疏水性的负值越大表示越亲水, 介于+0.5与-0.5之间为两性氨基酸, 番茄NAC蛋白质绝大多数是亲水蛋白质。

利用SOPMA 2.0软件对番茄NAC家族转录因

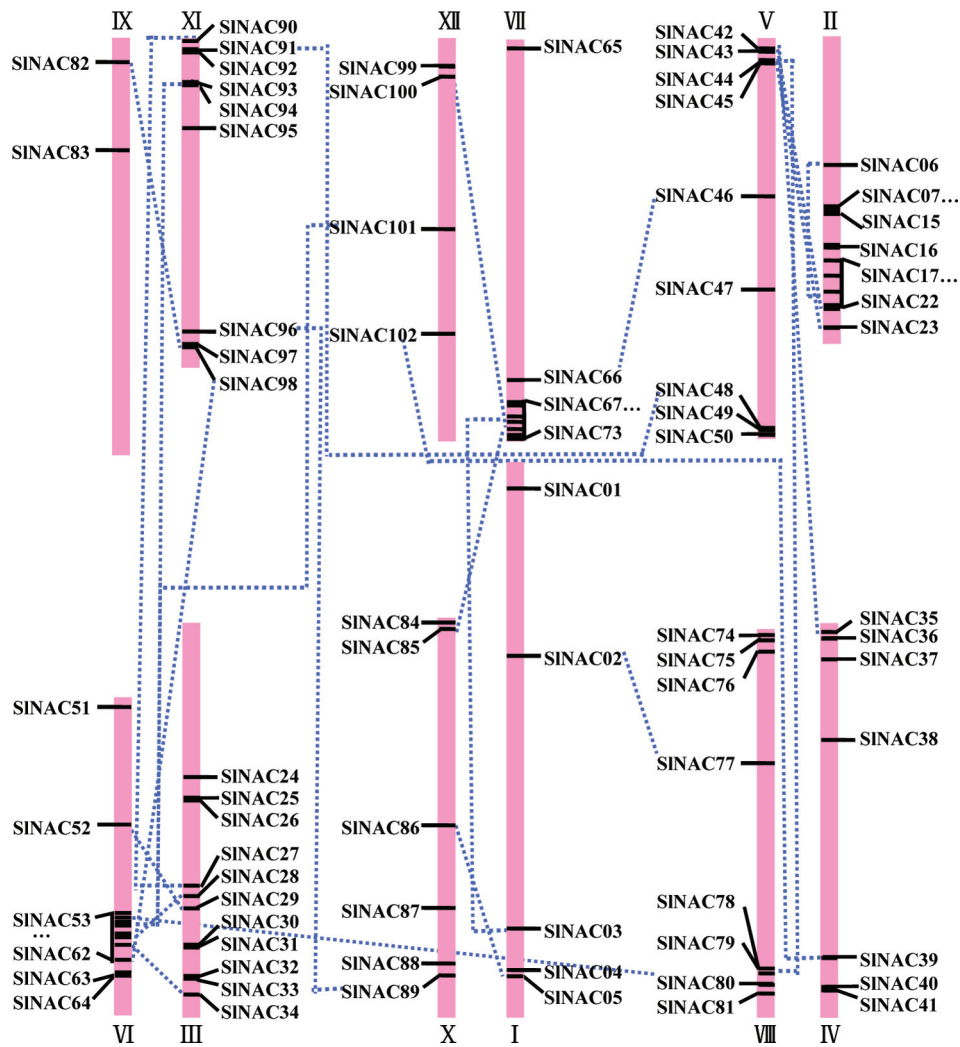


图1 番茄NAC转录因子家族在染色体上的分布

Fig.1 Chromosomal location of tomato NAC transcription factor family on all 12 chromosomes

子的二级结构进行了预测, 结果表明番茄NAC蛋白质家族结构主要以无规则卷曲构成, 而 α -螺旋、 β -折叠和 β -转角则散布于整个蛋白质中。

利用Phyre 2.0软件结合在线工具CombFunc对挖掘到的102条番茄NAC转录因子家族氨基酸序列进行三级结构及功能预测分析发现, 三级结构有很大的相似性(图4)。其中SINAC07、SINAC20和SINAC33蛋白中均含有5个 α -螺旋和9个 β -折叠, 但构象却不完全一致, 是由于 α -螺旋、 β -折叠和无规则卷曲的长度不同、分布不同, 导致空间角度的不同, 从而影响其功能。

GO分析表明, 预测得到的102条番茄NAC转录因子均具有转录因子活性, 在功能上以转录正

调控为主, 此外还有调节生物合成、胚后发育器官建成、光刺激应答、防御应答、细胞分裂、细胞大小、信号转导、干旱应答、木质部形成及叶片衰老等方面的功能。

讨 论

转录因子作为调控下游基因表达的重要元件, 调节着各种生理活动的关键环节, NAC转录因子家族是高等植物所特有的转录因子家族。近年来, 随着各种作物基因组测序的完成, 对NAC转录因子家族在基因组范围内进行挖掘成为研究热点。

本研究首次从基因组范围内对番茄NAC转录因子家族进行系统分析, 得到102条番茄NAC蛋白

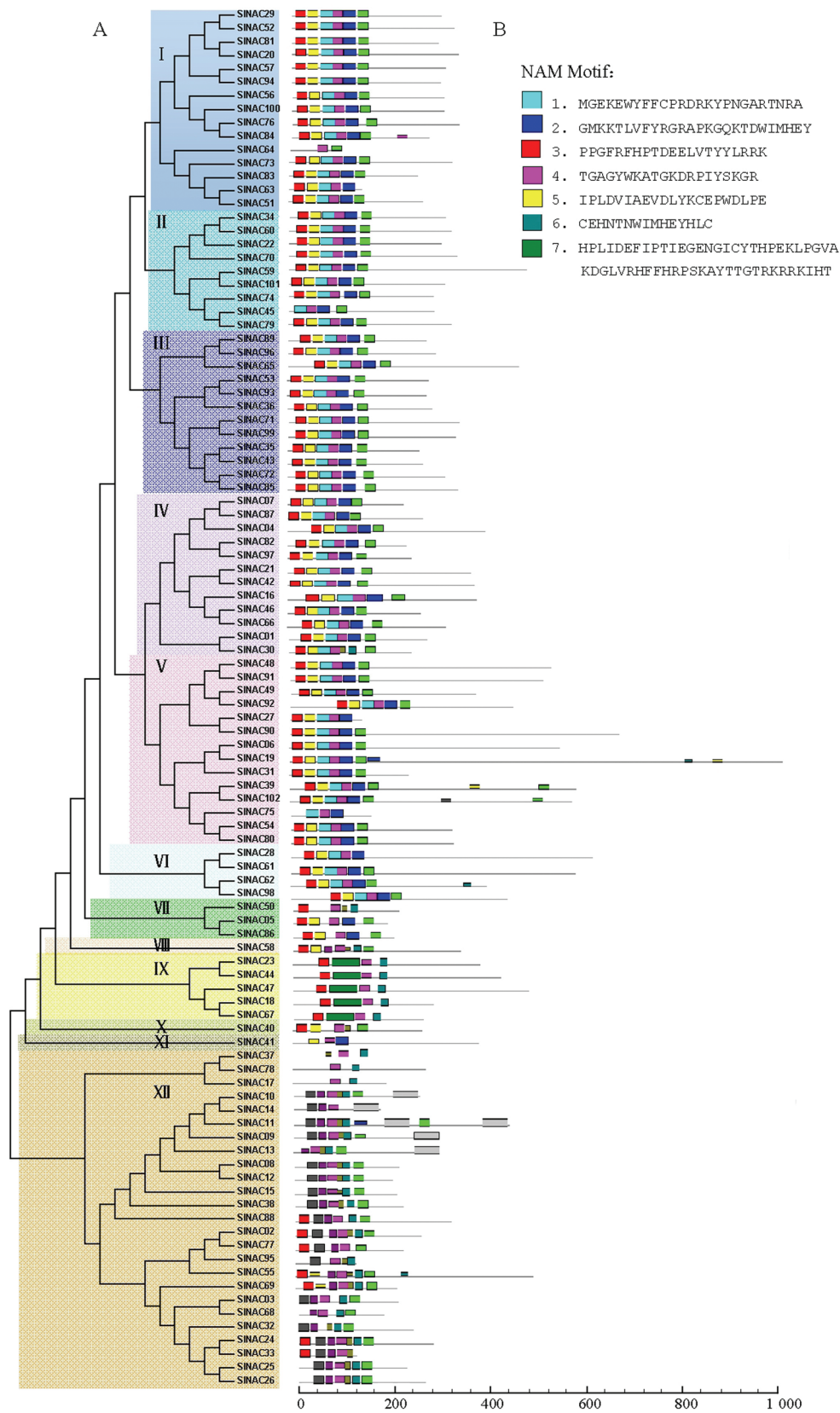


图2 番茄NAC转录因子家族分类及保守基序组成分析

Fig.2 Classification and conserved motif of tomato NAC transcription factor family

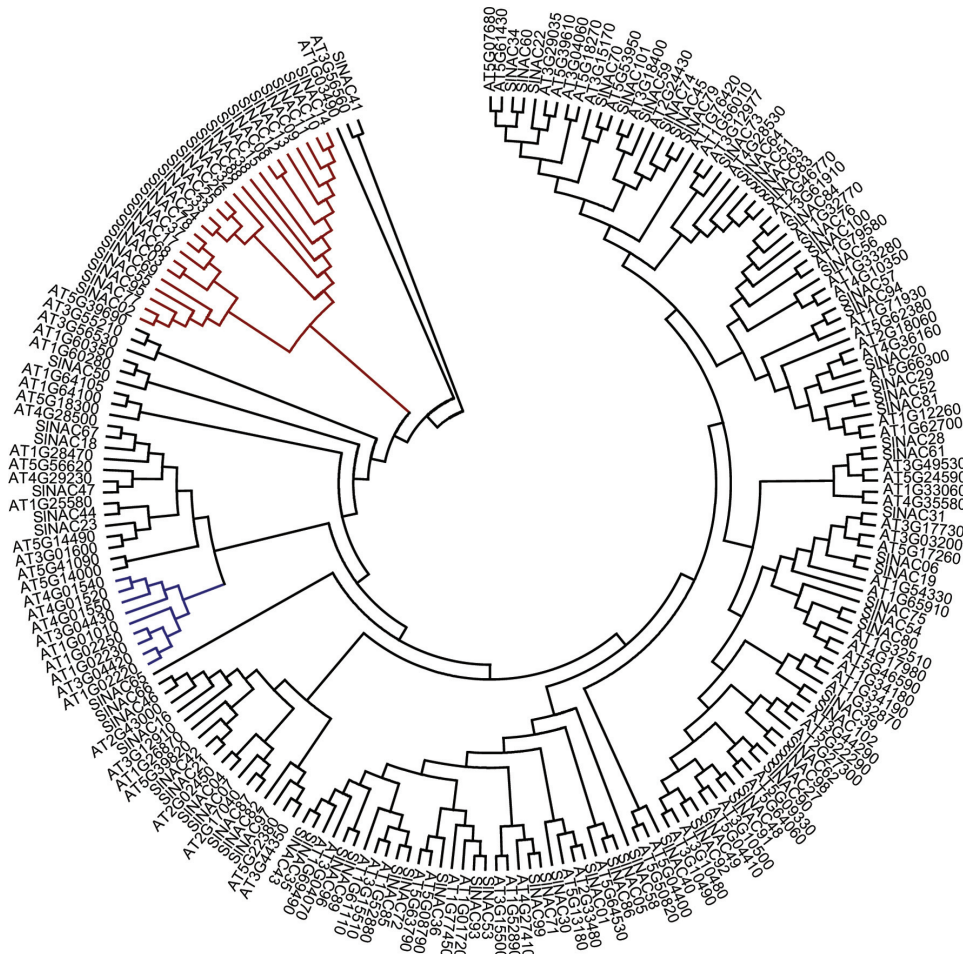


图3 番茄与拟南芥NAC转录因子家族蛋白质系统进化分析

Fig.3 Phylogenetic tree of NAC transcription factor family in tomato and *Arabidopsis thaliana*

红色表示番茄特有的一类NAC蛋白质, 蓝色表示拟南芥特有的一类NAC蛋白质。

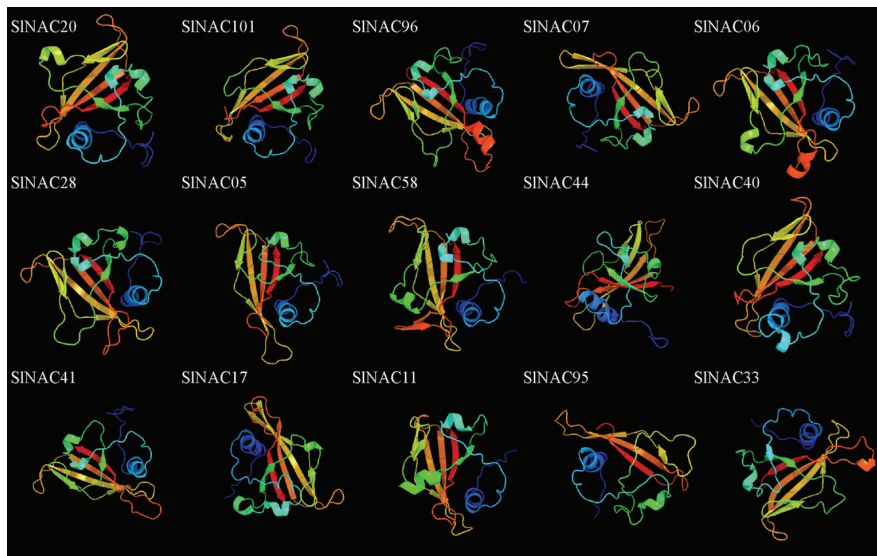


图4 番茄部分代表性NAC蛋白质三级结构

Fig.4 The tertiary structure of representative predicted NAC proteins in tomato

质,与原有已知的番茄NAC转录因子家族相比,得到一条新的NAC序列SINAC50。经过聚类分析发现,与拟南芥相比,番茄有一类特有的NAC转录因子,这与Rushton等(2008)研究一致,这类特有的NAC转录因子被归类为TNAC亚族,也进一步证明了TNAC亚族是茄属作物特有的一类NAC转录因子。

番茄NAC转录因子家族一级机构的理化特性、二级结构中的 α -螺旋、 β -折叠和无规则卷曲,以及三级结构中的蛋白质折叠的空间构象具有较高的保守性,在其他物种的NAC转录因子结构分析中,也发现了这个规律(孙欣等2011;刘更森等2012;张春华等2012)。

功能预测表明预测得到的102条番茄NAC转录因子均具有转录因子活性,并且证明了番茄NAC转录因子家族在调控生长发育、器官建成、响应生物和非生物胁迫中均具有一定作用,NAC转录因子成员具体行使何种功能,需要进一步试验验证。本研究为番茄NAC转录因子的生物学功能分析及其调控提供了重要依据。

参考文献

- 刘更森,张演义,戴洪义,宋伟(2012). 苹果 NAC 转录因子家族生物信息学分析. 江苏农业科学, 40 (6): 14~23
- 孙欣,上官凌飞,房经贵,宋长年,王晨,慕茜(2011). 葡萄NAC转录因子家族生物信息学分析. 基因组学与应用生物学, 30 (2): 229~242
- 张春华,上官凌飞,俞明亮,张彦萍,马瑞娟(2012). 桃NAC基因家族生物信息学分析. 江苏农业学报, 28 (2): 406~414
- Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997). Genes involved in organ separation in *Arabidopsis*: An analysis of the *cup-shaped* cotyledon mutant. *Plant Cell*, 9 (6): 841~857
- Bhatnagar-Mathur P, Devi MJ, Reddy DS, Lavanya M, Vadez V, Serraj R, Yamaguchi-Shinozaki K, Sharma KK (2007). Stress-inducible expression of *AtDREB1A* in transgenic peanut (*Arachis hypogaea* L.) increases transpiration efficiency under water-limiting conditions. *Plant Cell Rep*, 26: 2071~2082
- Delessert C, Kazan K, Wilson IW, Van Der Straeten D, Manners J, Dennis ES, Dolferus R (2005). The transcription factor ATAF2 represses the expression of pathogenesis-related genes in *Arabidopsis*. *Plant J*, 43 (5): 745~757
- Duval M, Hsieh TF, Kim SY, Thomas TL (2002). Molecular characterization of *AtNAM*: A member of the *Arabidopsis* NAC domain superfamily. *Plant Mol Biol*, 50 (2): 237~248
- Fang YJ, You J, Xie KB, Xie WB, Xiong LZ (2008). Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics*, 280 (6): 547~563
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005). Protein identification and analysis tools on the ExPASy Server. In: Walker JM (ed). *The Proteomics Protocols Handbook*. Clifton: Humana Press, 571~607
- Greve K, La Cour T, Jensen MK, Poulsen FM, Skriver K (2003). Interactions between plant RING-H2 and plant specific NAC (*NAM/ATAF1/2/CUC2*) proteins: RING-H2 molecular specificity and cellular localization. *Biochem J*, 371: 97~108
- He XJ, Mu RL, Cao WH, Zhang ZG, Zhang JS, Chen SY (2005). At-NAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J*, 44: 903~916
- Hu HH, Dai MQ, Yao JL, Xiao BZ, Zhang QF, Xiong LZ (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc Nat Acad Sci USA*, 103: 12987~12992
- Hu RB, Qi G, Kong YZ, Kong DJ, Gao Q, Zhou GK (2010). Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biol*, 10: 145~167
- Letunic I, Doerks T, Bork P (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*, 40: 302~305
- Jensen MK, Hagedorn PH, de Torres-Zabala M, Grant MR, Rung JH, Collinge DB, Lyngkjaer MF (2008). 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*. *Plant J*, 56: 867~880
- Kim SG, Kim SY, Park CM (2007). A membrane-associated NAC transcription factor regulates salt-responsive flowering via *FLOWERING LOCUS T* in *Arabidopsis*. *Planta*, 226: 647~654
- Kim YS, Kim SG, Park JE, Park HY, Lim MH, Chua NH, Park CM (2006). A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*. *Plant Cell*, 18: 3132~3144
- Ko JH, Yang SH, Park AH, Lerouxel O, Han KH (2007). ANAC012, a member of the plant specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *Plant J*, 50 (6): 1035~1048
- Liu YZ, Baig MNR, Fan R, Ye JL, Cao CY, Deng XX (2009). Identification and expression pattern of a novel NAM, ATAF, and CUC-like gene from *Citrus sinensis* Osbeck. *Plant Mol Biol Rep*, 27: 292~297
- Nakashima K, Tran LS, VanNguyen D, Fujita M, Maruyama K, Todaka D, Ito Y, Hayashi N, Shinozaki K, Yamaguchi-Shinozaki K (2007). Functional analysis of a NAC-type transcription factor *OsNAC6* involved in abiotic and biotic stress-responsive gene expression in rice. *Plant J*, 51: 617~630
- Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene*, 465: 30~44
- Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Osato N, Kawai J, Carninci P, Hayashizaki Y et al (2003). Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis*

- thaliana*. DNA Res, 10: 239~247
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Bournnell C, Pang N, Forslund K, Ceric G, Clements J et al (2012). The Pfam protein families database. Nucleic Acids Res, 40: 290~301
- Ren T, Qu F, Morris TJ (2000). *HRT* gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. Plant Cell, 12: 1917~1925
- Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, Laudeman TW, Timko MP (2008). Tobacco transcription factors: novel insights into transcriptional regulation in the *Solanaceae*. Plant Physiol, 147: 280~295
- Sablowski RW, Meyerowitz EM (1998). A homolog of no apical meristem is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. Cell, 92: 93~103
- Seki M, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K (2003). Molecular responses to drought, salinity and frost: common and different paths for plant protection. Curr Opin Biotechnol, 14: 194~199
- Singh AK, Sharma V, Pal AK, Acharya V, Ahuja PS (2013). Genome-wide organization and expression profiling of the NAC transcription factor family in potato (*Solanum tuberosum* L.). DNA Res, 20 (4): 403~425
- Souer E, van Houwelingen A, Kloos D, Mol J, Koes R (1996). The no apical meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell, 85 (2): 159~170
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28 (10): 2731~2739
- Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004). 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. Plant Cell, 16: 2481~2498
- Tran LS, Quach TN, Guttikonda SK, Aldrich DL, Kumar R, Neelakandan A, Valliyodan B, Nguyen HT (2009). Molecular characterization of stress-inducible *GmNAC* genes in soybean. Mol Genet Genomics, 281: 647~664
- Xie Q, Sanz-Burgos AP, Guo H, García JA, Gutiérrez C (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. Plant Mol Biol, 39: 647~656
- Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH (2007). Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. PLoS ONE, 2: e642
- Yu CS, Chen YC, Lu CH, Hwang JK (2006). Prediction of protein subcellular localization. Proteins, 64: 643~651
- Zhang H, Jin JP, Tang L, Zhao Y, Gu XC, Gao G, Luo JC (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res, 39: 1114~1117
- Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. Plant Cell, 20 (10): 2763~2782