

· 园艺 ·

DOI: 10.15886/j.cnki.rdswwb.20250024



主持人: 朱国鹏

Open Access



睡莲 KNOX 基因家族鉴定及生物信息学分析

许慧娴[#], 黎 洁, 李梅儿, 张汝鑫, 王童欣, 李霆格, 赵 莹, 王 健^{*}

(海南大学 热带农林学院/海南省热带特色花木资源生物学重点实验室/热带特色花木花卉遗传与种质创新教育部重点实验室, 海南 儋州 571737 中国)

摘 要: KNOX 基因家族是编码同源异型盒蛋白的转录因子, 在植物生长发育过程中起着重要的调控作用。为了解 KNOX 基因家族的生物信息特征及其在睡莲 (*Nymphaea*) 胎生苗发生过程中可能发挥的作用, 本研究通过生物信息学方法从睡莲基因组中鉴定到 15 个 KNOX 基因, 发现其分布于 1、2、3、4、5、6、7、8、9、11、13 号染色体上, 并对其理化性质、顺式作用元件及物种共线性等进行了分析。理化性质分析结果表明, 睡莲 KNOX 编码 249~2 270 个氨基酸, 预测均为亲水性蛋白, 分子质量在 2.84~24.95 kDa 之间, 其成员基因主要位于细胞核。大多数睡莲 KNOX 基因含有典型 KNOX1、KNOX2、ELK、Homeobox_KN 4 种蛋白保守结构域; 根据睡莲 KNOX 基因家族的结构特征和系统发育分析, 将其分为 Class I 和 Class II 2 类, 其中, Class II-B 可能是睡莲特有的类别。种间共线性分析发现, 睡莲 KNOX 基因家族成员中有 13 个与拟南芥 (*Arabidopsis thaliana*)、番茄 (*Solanum lycopersicum*)、水稻 (*Oryza sativa*) 和玉米 (*Zea mays*) 的 KNOX 基因均存在共线性关系, 且与番茄的同源性最高。启动子顺式作用元件分析发现, 睡莲 KNOX 基因家族含有较多与生长发育、激素和胁迫响应相关的顺式作用元件。通过不同时期叶片的表达模式分析发现, I 类 NcKNOX 基因极有可能对睡莲叶片胎生芽的产生和发育起到关键作用。

关键词: 睡莲; KNOX 基因家族; 叶片胎生

中图分类号: S682.32 文献标志码: A 文章编号: 1674-7054(2025)06-0842-13

许慧娴, 黎洁, 李梅儿, 等. 睡莲 KNOX 基因家族鉴定及生物信息学分析 [J]. 热带生物学报, 2025, 16(6): 842-854. doi: 10.15886/j.cnki.rdswwb.20250024

KNOX(KNOTTED1-like homeobox)基因家族属于同源异形盒基因 TALE(Three Amino-acid Loop Extension)超家族中的一类亚家族, 在植物生长发育中发挥重要作用。KNOX 基因家族根据保守结构域的不同可将其分为 4 种类型: KNOX1、KNOX2、ELK 和 Homeobox_KN^[1]。在结构域 N 端的 KNOX1 和 KNOX2 协同作用, 不仅调控靶基因的表达, 还促进蛋白质间的同源二聚化。在结构域上游的 ELK 则具有转录抑制功能, 影响基因的表达。而位于结构域 C 末端的 Homeobox_KN 能够识别并锁定下游靶基因中的启动子序列, 从而启动或调控特定的基因表达过程^[2]。根据 KNOX 基因家族的结构特征和系统发育分析, 将

该家族分为 Class I、Class II 和 Class M 3 类, 不同类别的 KNOX 基因的表达模式及生物学功能表现出明显差异^[3-4]。I 类亚家族对于植物分生组织的形成和维持起着核心作用。这类基因主要在植物的顶端分生组织中特异性表达, 通过调控与植物器官形成密切相关的细胞分化过程, 最终影响侧生器官的形态建成^[5]。相对 I 类亚家族而言, II 类亚家族的表达部位分布较为广泛, 在植物的各个组织器官中均有所表达。II 类亚家族主要参与调控植物器官的分化过程, 并且在次级细胞壁的生物合成中发挥重要作用^[6]。M 类亚家族主要表达在植物器官原基的近端域和成熟器官的边缘, 可能参与感知或产生开花信号^[7-8]。

收稿日期: 2025-01-21

修回日期: 2025-03-20

基金项目: 海口市重点项目(2022-13); 海南省重点研发项目(ZDYF2020152, ZDYF2022XDNY179); 海南省研究生创新科研课题(Qhys2022-272)

[#]第一作者: 许慧娴(1998—), 女, 海南大学热带农林学院 2022 级硕士研究生。E-mail: huixian.xu@139.com

^{*}通信作者: 王健(1977—), 男, 博士, 教授。研究方向: 园林植物种质资源与利用。E-mail: wjhainu@hainanu.edu.cn

在拟南芥(*Arabidopsis thaliana*)KNOX 基因家族中, I类亚家族包含 *STM*(*SHOOT MERISTEMLESS*)、*KNAT1*、*KNAT2* 和 *KNAT6*, 而 II类亚家族则涵盖 *KNAT3*、*KNAT4*、*KNAT5* 和 *KNAT7*, 这些基因在植物体内广泛分布和表达^[9]。*STM* 和 *KNAT1* 在 SAM(Shoot Apical Meristem, 茎尖分生组织)的分化和维持中发挥着重要作用, *KNAT6* 也被证实参与 SAM 和胚胎发生过程中参与维持组织边界的完整性, *KNAT2* 则着重于调节花型特征^[10]。I类亚家族的基因不仅参与侧向器官的形态发生过程, 还维持 SAM 功能以及塑造叶片形态的多样性^[11]。II类亚家族的 *KNAT3* 基因与拟南芥种子萌发和幼苗生长发育过程中的脱落酸响应过程有关^[12], *KNAT7* 基因与植物次生细胞壁合成有关^[13]。对 M类亚家族研究较少, 在拟南芥的研究中发现其与叶片的发育有关^[7]。

睡莲属(*Nymphaea*)是睡莲科(Nymphaeaceae)的多年生水生草本植物, 具有极其丰富的遗传多样性^[14]。睡莲各部位含有丰富的类黄酮和多糖等活性物质, 具有极高的药用价值、食用价值和观赏价值^[15]。睡莲的繁殖方式一般分为有性繁殖和无性繁殖, 其中, 无性繁殖方式较为多样, 包括块茎繁殖、球茎繁殖、根状茎繁殖、营养体繁殖和胎生繁殖^[16]。睡莲的叶片胎生尤为独特, 表现为在叶脐的部位形成胎生点, 之后发育形成胎生芽, 发育成熟后脱离母体形成新的植株, 常见于热带睡莲。‘蓝鸟’睡莲(*Nymphaea* ‘Blue Bird’)是睡莲中栽培最广泛、商品价值最高的热带品种之一, 同时, 也是胎生睡莲的代表性品种。本实验室初步的研究表明, KNOX 基因家族也可能在‘蓝鸟’睡莲叶片胎生繁殖中发挥重要作用。随着各个物种全基因组数据的发布, 目前, 在拟南芥、番茄(*Solanum lycopersicum*)、水稻(*Oryza sativa*)和玉米(*Zea mays*)等植物中已完成了 KNOX 基因家族的鉴定和分析, 而睡莲虽然已经公布了代表性种蓝星睡莲(*Nymphaea caerulea*)的基因组^[17], 但关于 KNOX 基因家族的报道尚属空白。因此, 本研究首先对蓝星睡莲基因组中的 KNOX 基因家族成员进行了系统性的分析, 成功确定了睡莲 15 个 KNOX 基因, 并进一步探究了它们的理化性质、亚细胞定位特征, 构建了系统进化树以揭示其进化关系, 然后分析了这些基因在染色体上的定位情况和种间共线性情况, 最后通过转录组数据分析了这些基因在‘蓝鸟’睡莲不同发育时期胎生和非胎生叶片中

的表达情况。这些研究成果不仅可以为了解 KNOX 基因家族的进化特征提供基础的理论支持, 同时也为挖掘该家族在睡莲中的潜在功能奠定了坚实的基础。

1 材料与方法

首先从 Ensemble Plants 数据库(<https://plants.ensembl.org/index.html>)下载蓝星睡莲基因组序列、蛋白质序列和注释文件。同时为进行跨物种比较, 还需下载拟南芥、番茄、水稻和玉米的 KNOX 家族成员蛋白序列。为了识别蓝星睡莲中可能含有 KNOX 家族相关结构域的蛋白, 本研究从 Pfam 数据库(<http://pfam-legacy.xfam.org/>)下载了 KNOX1(PF03790)、KNOX2(PF03791)、ELK(PF03789)及 Homeobox_KN(PF05920)的结构域信息和隐马尔可夫模型(HMM)文件。

1.1 睡莲 KNOX 家族成员鉴定及蛋白理化性质分析 首先将已获取的保守结构域信息和隐马尔可夫模型(HMM)文件导入 TBtools(v2.127)软件^[18], 并执行 HMMER 命令, 筛选条件设置为 e 值小于 $1e^{-5}$, 在蓝星睡莲的基因组数据库中筛选包含特定结构域的蛋白序列。本研究利用拟南芥、水稻、番茄和玉米的 KNOX 蛋白序列作为参照, 通过 TBtools 软件的 BLASTp 命令在蓝星睡莲的蛋白数据库中搜索同源序列, 同样以 e 值小于 $1e^{-5}$ 为筛选标准, 并对 HMMER 和 BLASTp 两种方法得到的序列进行交集分析, 并利用 SMART(<https://smart.embl.de/>)和 NCBI CDD 数据库(<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>)进行了保守结构域的鉴定。通过 MEGA7.0 软件对睡莲 KNOX 家族成员进行多序列比对, 再利用 TBtools 软件的 TrimAL Wrapper 功能进行蛋白序列修剪和可视化处理。

利用 ProtParam(<https://web.expasy.org/protparam/>)工具和 WoLFPSORT(<https://wolfpsort.hgc.jp/>)网站预测每个 KNOX 蛋白的氨基酸数(aa)、分子量、等电点(pI)以及亚细胞定位。

1.2 睡莲 KNOX 基因家族系统发育树构建与分析 为了解拟南芥、水稻、番茄、玉米及睡莲中 KNOX 基因家族成员之间的进化关系, 本研究采用 MEGA7.0 软件的 Muscle 比对方法对 KNOX 蛋白序列进行多重比对, 选用邻接法(Neighbor-joining)来构建系统发育树, 同时运用 ITOL(<https://>

itol.embl.de/)在线网站来对系统发育树进行美化。

1.3 睡莲 KNOX 基因结构分析 基于睡莲 KNOX 家族成员的蛋白序列和注释信息,本研究通过 NCBI Batch CD-search 工具(<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)批量检索并下载保守结构域信息。同时,利用 MEME 在线工具(<https://meme-suite.org/meme/tools/meme>)对序列进行 motif 预测,设定 motif 数量为 10,其余保持默认参数,并下载预测结果。通过 TBtools 软件的 Gene Structure View 功能,对睡莲 KNOX 家族成员的外显子和内含子结构进行分析。

1.4 睡莲 KNOX 基因启动子顺式作用元件分析 利用 TBtools 软件从睡莲基因组数据库中提取 15 个 *NcKNOX* 基因的 CDS 序列上游 2 000 bp 的启动子区域,将获得的启动子区域序列提交到 Plant Care 数据库(http://bioinformatics.psb.ugent.be/webtools/plant_care/html/)来进行顺式作用元件分析。

1.5 睡莲 KNOX 基因染色体定位分析 基于睡莲基因组的注释文件获取 KNOX 基因家族成员的染色体位置信息,利用 TBtools 软件中的 Gene Location Visualize 功能进行了可视化处理。

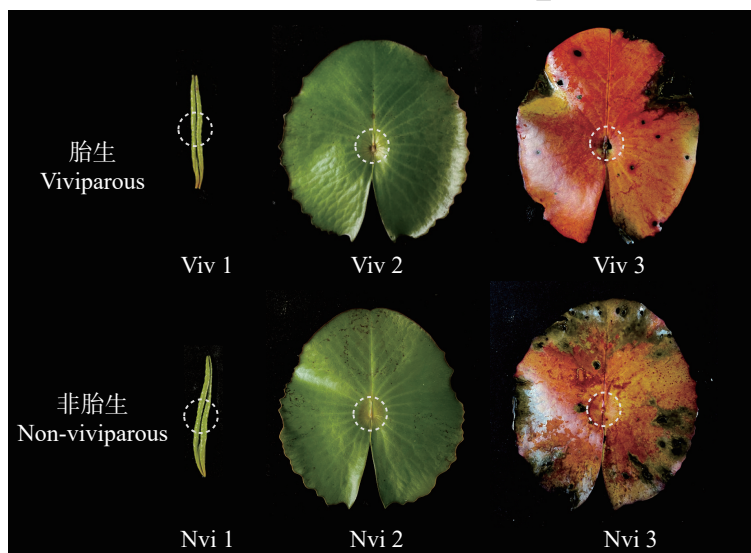
1.6 多物种 KNOX 基因家族共线性分析 基于拟南芥、水稻、番茄、玉米及睡莲的 KNOX 蛋白序列以及注释文件,用 TBtools 软件的 One Step MC Scan X 和 Circle Gene View 功能来进行多物

种间和睡莲种内的共线性分析,获得这几个物种间及睡莲种内 KNOX 基因的共线关系图。

1.7 KNOX 基因家族成员在不同发育时期睡莲叶片的表达分析 为了解 KNOX 基因家族成员在睡莲胎生叶片中的表达模式,探究 *NcKNOX* 基因在睡莲叶片胎生苗发生过程中的可能作用,选用胎生睡莲的代表性品种‘蓝鸟’睡莲的不同时期胎生叶片和非胎生叶片作为实验材料,研究其表达规律。根据‘蓝鸟’睡莲的胎生(Viviparous, Viv)及非胎生(Non-viviparous, Nvi)叶片的生长状态,将其分为胎生和非胎生 1~3 期(Viv 1~3 和 Nvi 1~3)的叶片,并将其胎生点(直径约为 1 cm,图 1)切下置于-80 ℃ 保存,送迈维代谢公司进行转录组测序^[19]。根据转录组数据,使用 TBtools 软件中 HeatMap 对 *NcKNOX* 基因的 FPKM 值进行 Log scale 处理并进行表达模式可视化分析。

2 结果与分析

2.1 睡莲 KNOX 家族成员鉴定及蛋白理化性质分析 结合 BLASTp 和 HMMER 得到的数据取交集并进行人工校正,最终鉴定得到 15 个睡莲 KNOX 基因。蛋白序列比对分析(图 2)可知, KNOX 基因家族的 4 种保守结构域的保守情况,分别是典型保守序列 KNOX1、KNOX2、ELK 和 Homeobox_KN,其中,Homeobox_KN 结构域相较

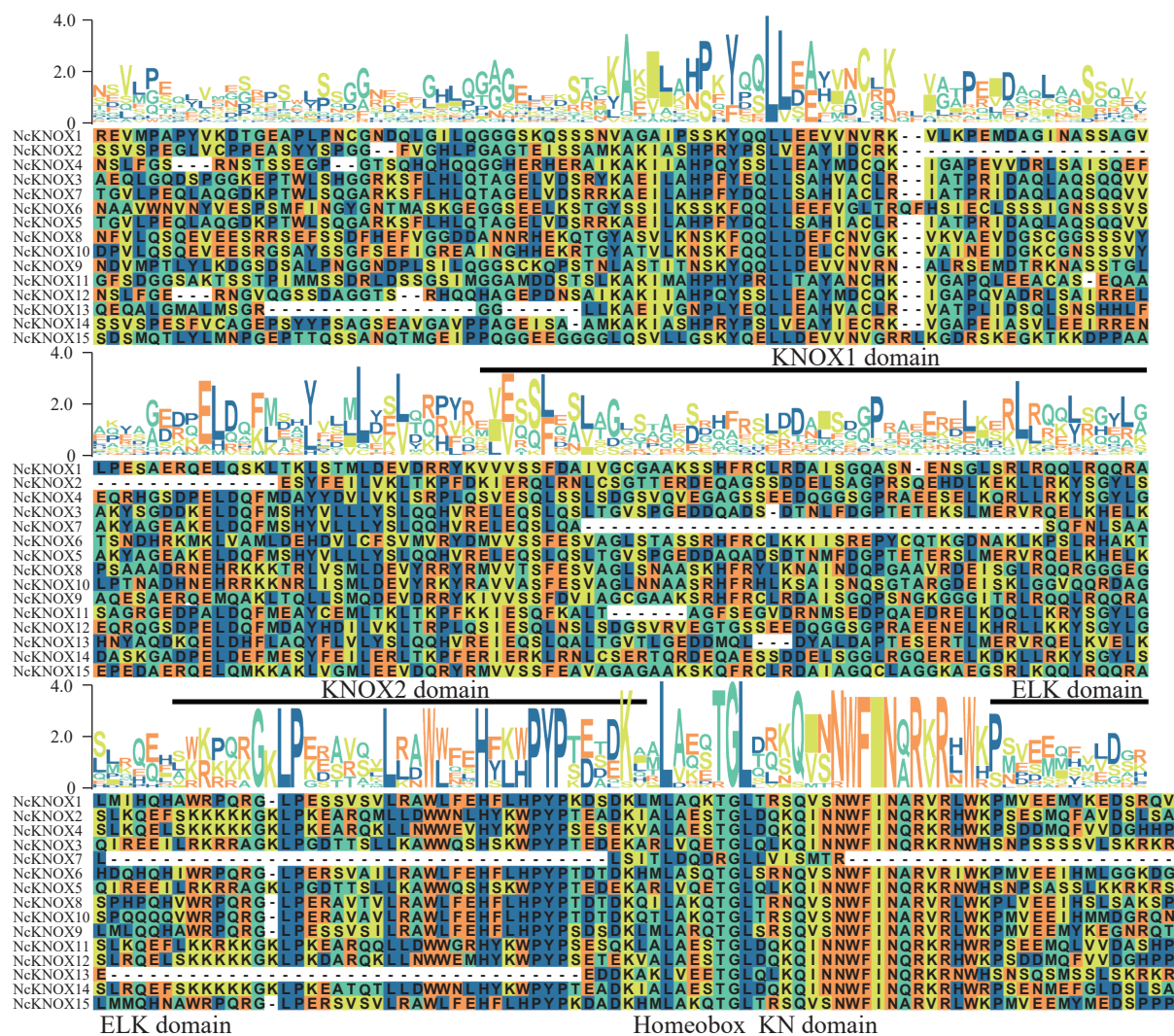


Viv 1.胎生 1 期叶片; Viv 2.胎生 2 期叶片; Viv 3.胎生 3 期叶片; Nvi 1.非胎生 1 期叶片; Nvi 2.非胎生 2 期叶片; Nvi 3.非胎生 3 期叶片。

Viv 1. Leaves at viviparous stage 1; Viv 2. Leaves at viviparous stage 2; Viv 3. Leaves at viviparous stage 3; Nvi 1. Leaves at non viviparous stage 1; Nvi 2. Leaves at non viviparous stage 2; Nvi 3. Leaves at non viviparous stage 3.

图 1 ‘蓝鸟’睡莲叶片分期图

Fig. 1 Developmental stages of 'Blue Bird' water lily leaves



每个字母的高度与该位置的碱基或氨基酸残基的出现频率、保守性呈正相关。

The height of each letter is proportional to the frequency and conservation of the nucleotide or amino acid residue at that position.

图 2 NcKNOX 蛋白保守序列比对
Fig. 2 Sequence alignment of NcKNOX proteins

于其他结构域, 其保守性更强, 表明其在睡莲 KNOX 基因中更加稳定。

从表 1 可知, 睡莲 KNOX 基因家族成员的 CDS 序列长度为 909 ~ 6 813 bp, 氨基酸序列长度为 249 ~ 2 270 aa, 氨基酸序列分子质量在 2.84 ~ 24.95 kDa 之间, 等电点(pI)在 5.07 ~ 8.8 之间, 其中 pI<7 的酸性氨基酸有 13 个, pI>7 的碱性氨基酸有 2 个, 富含酸性氨基酸, 这与玉米的 KNOX 基因家族相似^[20]。除了 NC6G0258750.1(*NcKNOX10*) 以外, 其余的蛋白亲疏水性均小于 0, 且溶脂指数均小于 100, 则说明睡莲 KNOX 基因家族成员大都具备显著的亲水性特征。亚细胞定位预测结果显示, *NcKNOX* 基因均位于细胞核内, 这也与玉米、谷子等物种的 KNOX 基因家族成员细

胞定位一致, 与 *KNOX* 基因的转录因子属性特征相符合^[21-22]。

2.2 睡莲 KNOX 基因家族系统发育树分析 使用 MEGA7.0 软件利用蛋白序列绘制包含模式植物拟南芥和睡莲等多种植物 KNOX 基因家族成员的系统发育进化树。从图 3 可知, KNOX 基因被分为 3 大类: Class I、Class II 和 Class M。其中 Class I 类中含有 5 个 *NcKNOX* 成员, 分别为 *NcKNOX2*、*NcKNOX4*、*NcKNOX11*、*NcKNOX12*、*NcKNOX14*, 与拟南芥的 *KNAT1*、*STM* 等基因进化关系比较近; Class II 类中含有 10 个成员, 分别为 *NcKNOX1*、*NcKNOX3*、*NcKNOX5*、*NcKNOX6*、*NcKNOX8*、*NcKNOX9*、*NcKNOX10*、*NcKNOX7*、*NcKNOX13* 和 *NcKNOX15*, 与拟南芥的 *KNAT3*、

表 1 睡莲 NcKNOX 蛋白家族的基本信息

Tab. 1 Basic information of the NcKNOX protein family in water lily

基因ID Gene ID	基因名称 Gene Name	氨基酸/aa Amino acid number/aa	分子质量/kDa Molecular weight/kDa	等电点 PI	不稳定指数 Instability index	溶脂指数 Aliphatic index (AI)	亲水性均值 Grand average of hydrophobicity	亚细胞定位预测 Subcellular localization
NC1G0108910.1	<i>NcKNOX1</i>	1 371	15.20	8.30	49.32	74.49	-0.468	Nucleus
NC2G0008460.1	<i>NcKNOX2</i>	296	3.29	5.39	50.65	60.41	-0.698	Nucleus
NC3G0221260.1	<i>NcKNOX3</i>	456	5.02	5.67	55.30	67.04	-0.675	Nucleus
NC3G0227930.1	<i>NcKNOX4</i>	421	4.82	6.34	48.79	62.11	-0.766	Nucleus
NC4G0011790.1	<i>NcKNOX5</i>	443	4.82	5.29	58.34	66.84	-0.667	Nucleus
NC4G0020470.1	<i>NcKNOX6</i>	780	8.46	6.03	46.25	71.01	-0.406	Nucleus
NC4G0153810.1	<i>NcKNOX7</i>	302	3.21	5.09	58.05	77.35	-0.308	Nucleus
NC5G0049120.1	<i>NcKNOX8</i>	800	8.69	6.92	47.53	69.46	-0.554	Nucleus
NC6G0041000.1	<i>NcKNOX9</i>	810	8.91	6.12	45.98	71.53	-0.508	Nucleus
NC6G0258750.1	<i>NcKNOX10</i>	2 270	24.95	8.80	43.15	89.85	0.022	Nucleus
NC7G0308290.1	<i>NcKNOX11</i>	395	4.33	5.67	64.02	57.90	-0.648	Nucleus
NC9G0170310.1	<i>NcKNOX12</i>	367	4.12	6.02	47.00	58.53	-0.860	Nucleus
NC9G0274690.1	<i>NcKNOX13</i>	249	2.84	5.2	62.95	84.26	-0.602	Nucleus
NC11G0123640.1	<i>NcKNOX14</i>	373	4.18	5.17	52.20	68.10	-0.506	Nucleus
NC13G0026850.1	<i>NcKNOX15</i>	714	7.68	6.28	52.37	69.41	-0.477	Nucleus

KNAT7 等基因进化关系比较近。

睡莲目作为被子植物谱系中的基源分支,其分化进程早于单、双子叶植物,既保留了双子叶植

物的某些特征,又兼具单子叶植物的某些属性。

而 Class M 类基因是双子叶植物中一个特殊的 KNOX 基因亚家族^[23],在系统发育树中仅有番茄

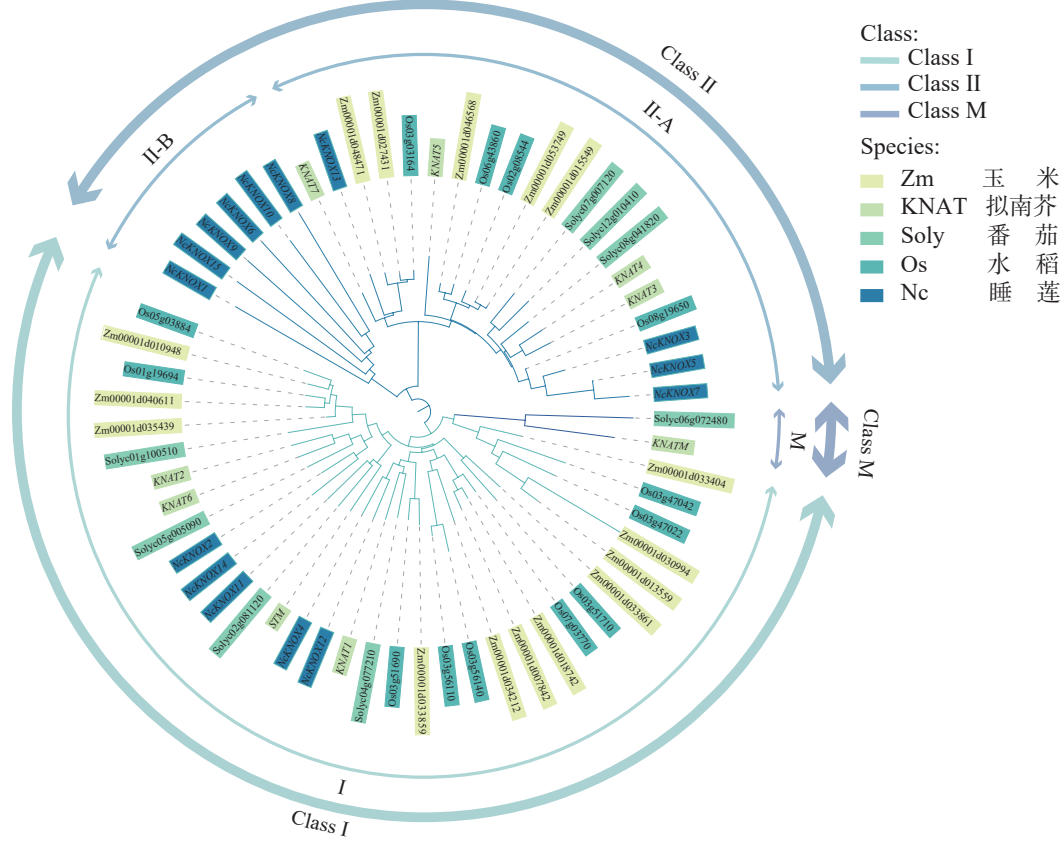


图 3 拟南芥、番茄、水稻、玉米和睡莲 KNOX 蛋白的系统发育树分析
Fig. 3 Phylogenetic tree analysis of KNOX proteins from Arabidopsis, tomato, rice, maize, and water lily

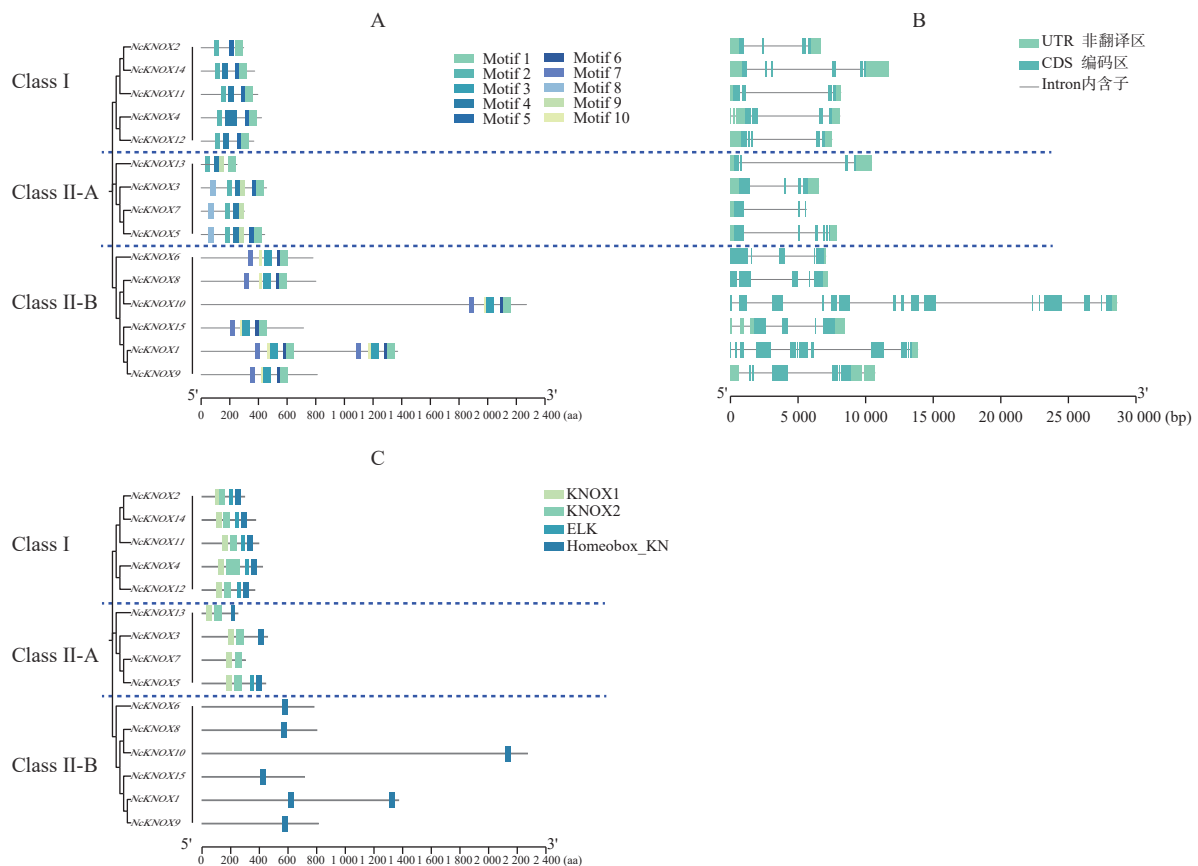
的 *Solyc06g072480* 基因与拟南芥的 *KNATM* 基因所属同一分支, 与前人研究相符^[24], 而在睡莲中没有发现与 *KNATM* 基因相近的基因。因此, 睡莲 *NcKNOX* 家族被分为 Class I 类和 Class II 类, 没有 Class M 类基因。值得关注的是, 睡莲 *NcKNOX1*、*NcKNOX6*、*NcKNOX8*、*NcKNOX9*、*NcKNOX10* 和 *NcKNOX15* 基因虽属于 Class II 类基因, 但在进化关系上与其他物种的 Class II 基因相隔较远, 独属于一个比较独特的分支, 因此, 将其命名为 Class II-B 分支, 对应的将其他 Class II 基因命名为 Class II-A。

2.3 睡莲 KNOX 基因结构分析 从图 4-A 可知, 睡莲 KNOX 蛋白家族的大多数 motif 在蛋白质序列的 C 末端区域高度富集。某些基序在大多数 KNOX 蛋白中都有, 如 motif 1; 也有一些基序的分布具有特异性, 如 Class I 类亚族没有 motif 6、7、8、9、10, 而 Class II 类的 II-B 分支则都有 motif 2, 说明这两类基因在进化上可能保留了某些共同的功能基础但又具有本亚族的独特性。

基因结构分析显示(图 4-B), *NcKNOX* 基因家族成员的序列长度相差较大, 导致不同成员之间的内含子与外显子数目存在较大差异。*NcKNOX* 基因家族主要由 3~16 个外显子和 2~15 个内含子组成, 其中, 序列最长的 *NcKNOX10* 含有 16 个外显子和 15 个内含子, 序列最短的 *NcKNOX7* 含有 3 个外显子和 2 个内含子。

保守结构域分析显示(图 4-C), 除了 *NcKNOX7* 之外, 所有的 *NcKNOX* 基因均编码了 Homeobox_KN 保守结构域, 其中, Class II-B 分支上的 *NcKNOX1*、*NcKNOX6*、*NcKNOX8*、*NcKNOX9*、*NcKNOX10* 和 *NcKNOX15* 等基因比较特殊, 仅包含 Homeobox_KN 结构域, 而 Class I 与 Class II-A 的 *NcKNOX* 基因还普遍含有 KNOX1 和 KNOX2 这两个核心结构域, Class I 类的全部成员以及 Class II-A 类中的 *NcKNOX5* 基因还具有 ELK 结构域。

进一步将睡莲与拟南芥、水稻等植物的 Class II KNOX 基因进行比较分析(图 5), 结果发现, 睡莲 Class II-A 基因同其他植物的 Class II KNOX



A. Conserved motifs of *NcKNOX* proteins; B. Conserved domains of *NcKNOX* genes; C. Structure of *NcKNOX* genes.

图 4 睡莲 *NcKNOX* 基因家族基序、结构及保守结构域分析

Fig. 4 Motif, structure and conserved domain of the *NcKNOX* gene family in water lily

基因一样,均含有 KNOX1 和 KNOX2 结构域,但睡莲 Class II-B 基因只含有 Homeobox_KN。

进一步说明 Class II-B 可能是睡莲 KNOX 基因家族特有的一个进化分支。

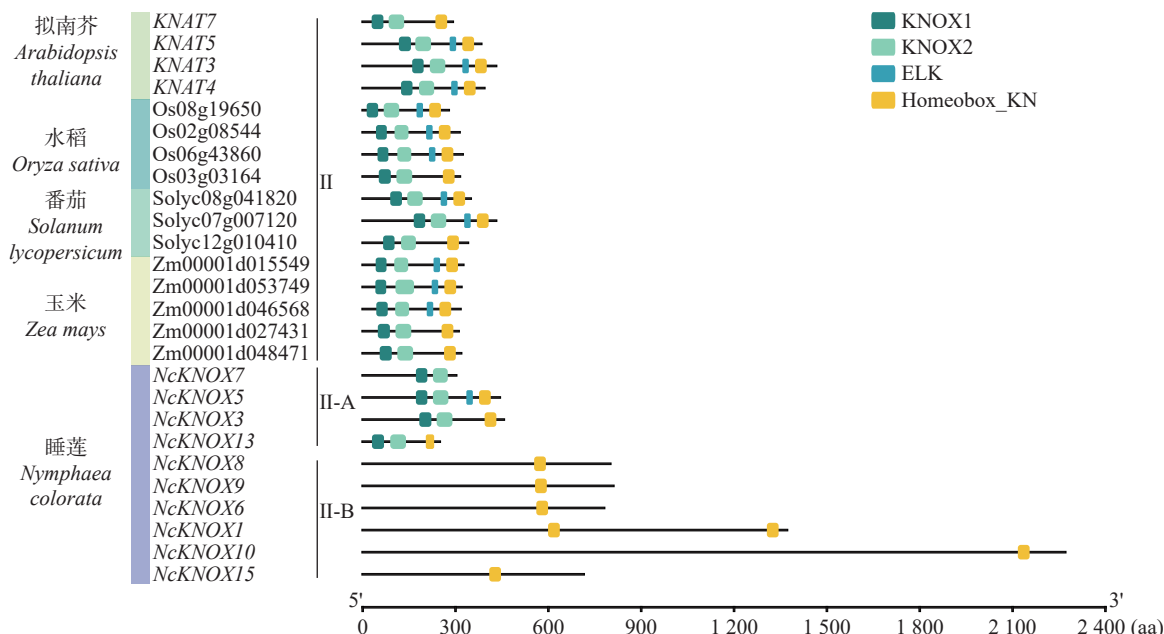


图 5 各物种 Class II KNOX 基因保守结构域分析

Fig. 5 Conserved domain analysis of Class II KNOX genes in different species

2.4 睡莲 KNOX 基因启动子顺式作用元件分析 结果表明,睡莲 *NcKNOX* 除了真核生物中的基本启动子元件外,还包括植物激素响应 (phytohormone response)、植物生长 (plant growth and development)、非生物和生物胁迫 (abiotic and biotic stress response) 和光响应 (light response) 相关的作用元件。

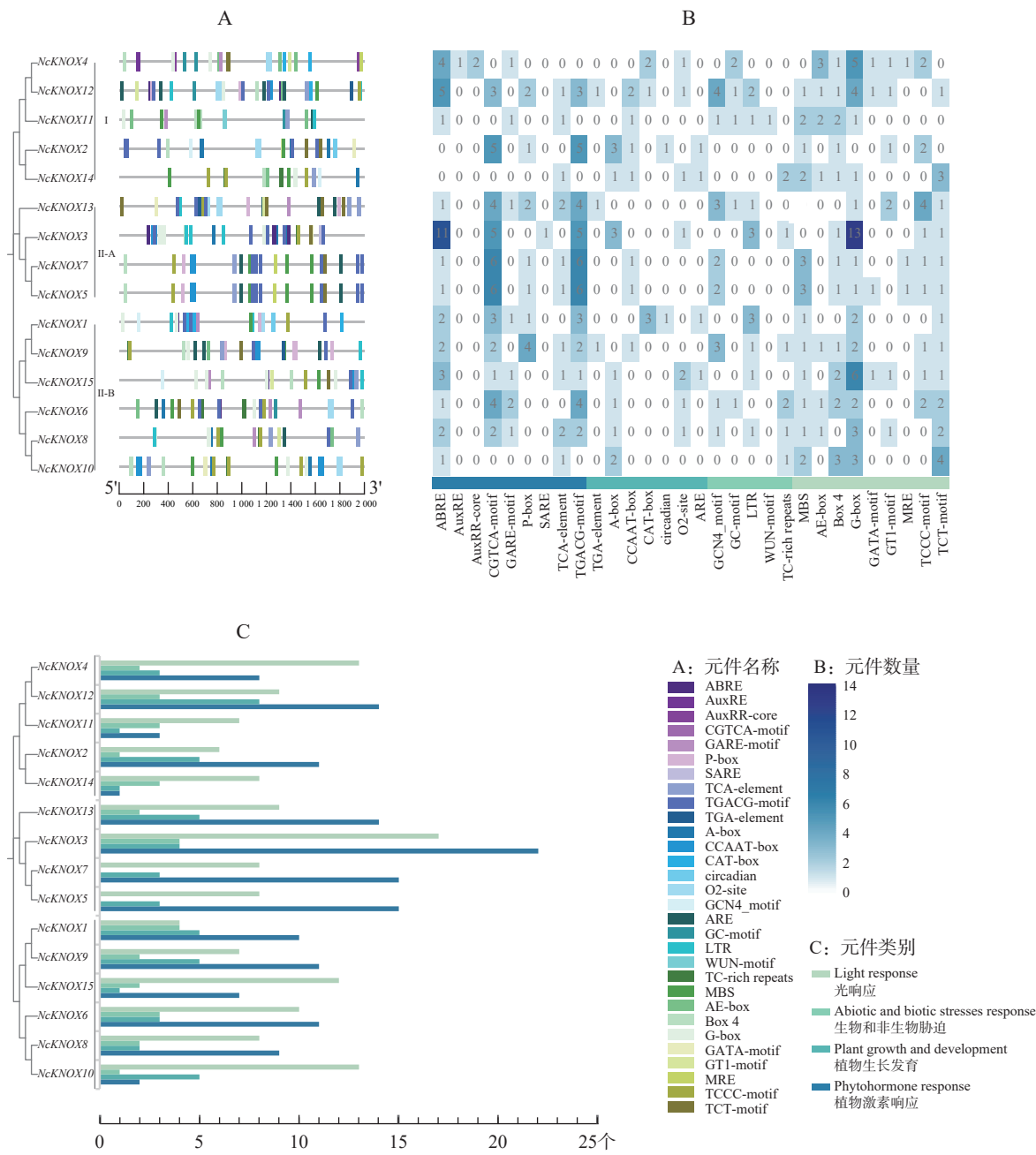
如图 6-A 所示,植物激素响应相关的元件有 ABRE、AuxRE 和 AuxRR-core 等;植物生长相关的元件有 TGA-element、A-box 和 CCAAT-box 等;胁迫反应相关的元件有 ARE、GC-motif 和 LTR 等;与光响应相关的元件有 MBS、AE-box 和 Box4 等。结合图 6-B 和图 6-C 分析,结果发现,在睡莲 *NcKNOX* 基因中植物激素响应和光响应元件数量占比更多,其中, *NcKNOX3* 基因较为突出,含有 11 个 ABRE 元件(脱落酸响应元件)和 13 个 G-box 元件(光响应顺式调节元件)。

2.5 睡莲 KNOX 基因染色体定位分析 染色体定位结果表明(图 7),睡莲 *NcKNOX* 基因家族的 15 个家族成员分布于 10 条不同染色体上。其中,1 号(*NcKNOX1*)、2 号(*NcKNOX2*)、5 号(*NcKNOX8*)、7 号(*NcKNOX11*)、11 号(*NcKNOX14*)和 13 号(*NcKNOX15*)染色体上仅分布 1 个家族成员,3 号

(*NcKNOX3* 和 *NcKNOX4*)、6 号(*NcKNOX9* 和 *NcKNOX10*)和 9 号染色体分别有 2 个家族成员。4 号染色体有 3 个家族成员(*NcKNOX5*、*NcKNOX6* 和 *NcKNOX7*),其中 *NcKNOX5* 与 *NcKNOX7* 呈成簇排列,表明两者有着十分紧密的连锁关系。

2.6 睡莲 KNOX 基因的共线性分析 为进一步了解睡莲 *NcKNOX* 基因家族进化关系,本研究对其进行种内共线性分析和其他物种的种间共线性分析(图 8)。种内共线性分析结果(图 8-A)显示,睡莲染色体上存在 5 组明确的片段复制事件和一组串联重复基因,分别为 *NcKNOX1* 与 *NcKNOX9*、*NcKNOX2* 与 *NcKNOX14*、*NcKNOX3* 与 *NcKNOX5*、*NcKNOX4* 与 *NcKNOX12* 及 *NcKNOX8* 与 *NcKNOX10* 之间的复制对,而 *NcKNOX5* 和 *NcKNOX7* 为串联重复基因。这些发生片段复制的 *NcKNOX* 基因对在构建的系统发育树中均归属于同一进化分支,与睡莲 KNOX 蛋白系统发育树的构建结果相互印证(图 3)。

种间共线性分析结果(图 8-B)表明,除 *NcKNOX7* 和 *NcKNOX11* 基因,其余 13 个 *NcKNOX* 基因与拟南芥等模式植物存在一定的共线性关系。如表 2 所示,睡莲分别有 4 个、6 个、10 个和 2 个 *NcKNOX* 基因与拟南芥、水稻、番茄和玉米基因存在共线性关系。



A. 顺式元件在基因的具体位置; B. 顺式元件在每个基因上的数量; C. 每个 *NcKNOX* 基因中植物激素响应、植物生长、胁迫响应及光响应元件数量。

A. Specific locations of cis-elements in the genes; B. Number of cis-elements in each gene; C. Number of plant hormone response, plant growth, stress response, and light response elements in each *NcKNOX* gene.

图 6 睡莲 KNOX 基因的启动子顺式作用元件分析

Fig. 6 Cis-regulatory element analysis of the promoters of KNOX genes in water lily

2.7 KNOX 基因家族成员在睡莲胎生叶片不同发育时期的表达分析 本研究基于蓝鸟睡莲不同发育时期胎生叶片的转录组数据, 分析了 *NcKNOX* 基因家族成员在不同发育时期叶片的表达模式 (图 9)。结果显示, I 类 KNOX 基因多数在胎生叶片 (Viv) 时期中高表达, 而在非胎生 (Nvi) 叶片中低表达, 如 *NcKNOX11*、*NcKNOX12* 和 *NcKNOX14* 等, 且这几个基因的表达量均随着胎生苗的发育

明显上调; 而 II-A 类 KNOX 基因在所有胎生和非胎生叶片均高表达, 如 *NcKNOX13*、*NcKNOX7* 和 *NcKNOX3* 等; 而 II-B 类 KNOX 基因中, 除了 *NcKNOX8* 和 *NcKNOX15* 在胎生和非胎生叶片中高表达, 其余基因都呈现低表达的趋势。总体上来看, 无论是胎生叶片还是非胎生叶片, 大多数 *NcKNOX* 基因从第 1 期到第 3 期叶片的表达量均呈上调的趋势, 这就说明 *NcKNOX* 在睡莲叶片发育过程中

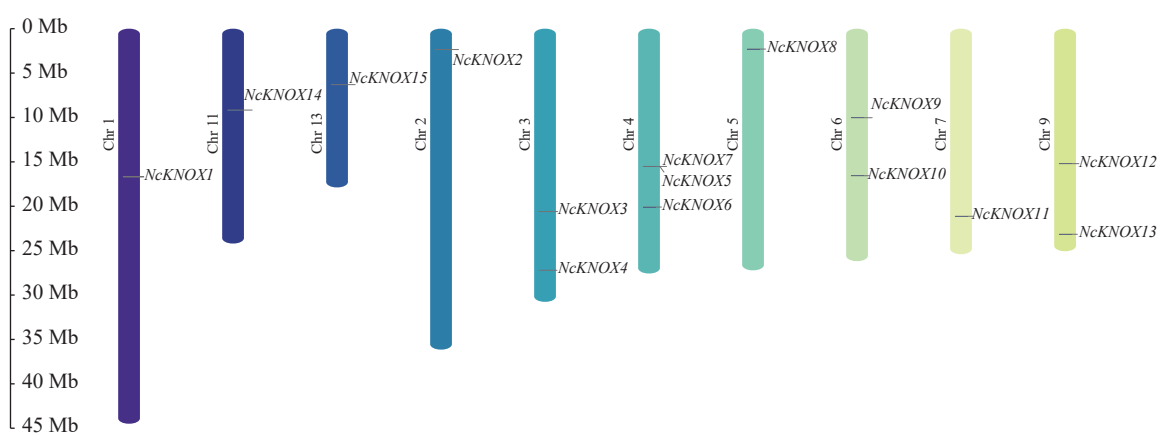
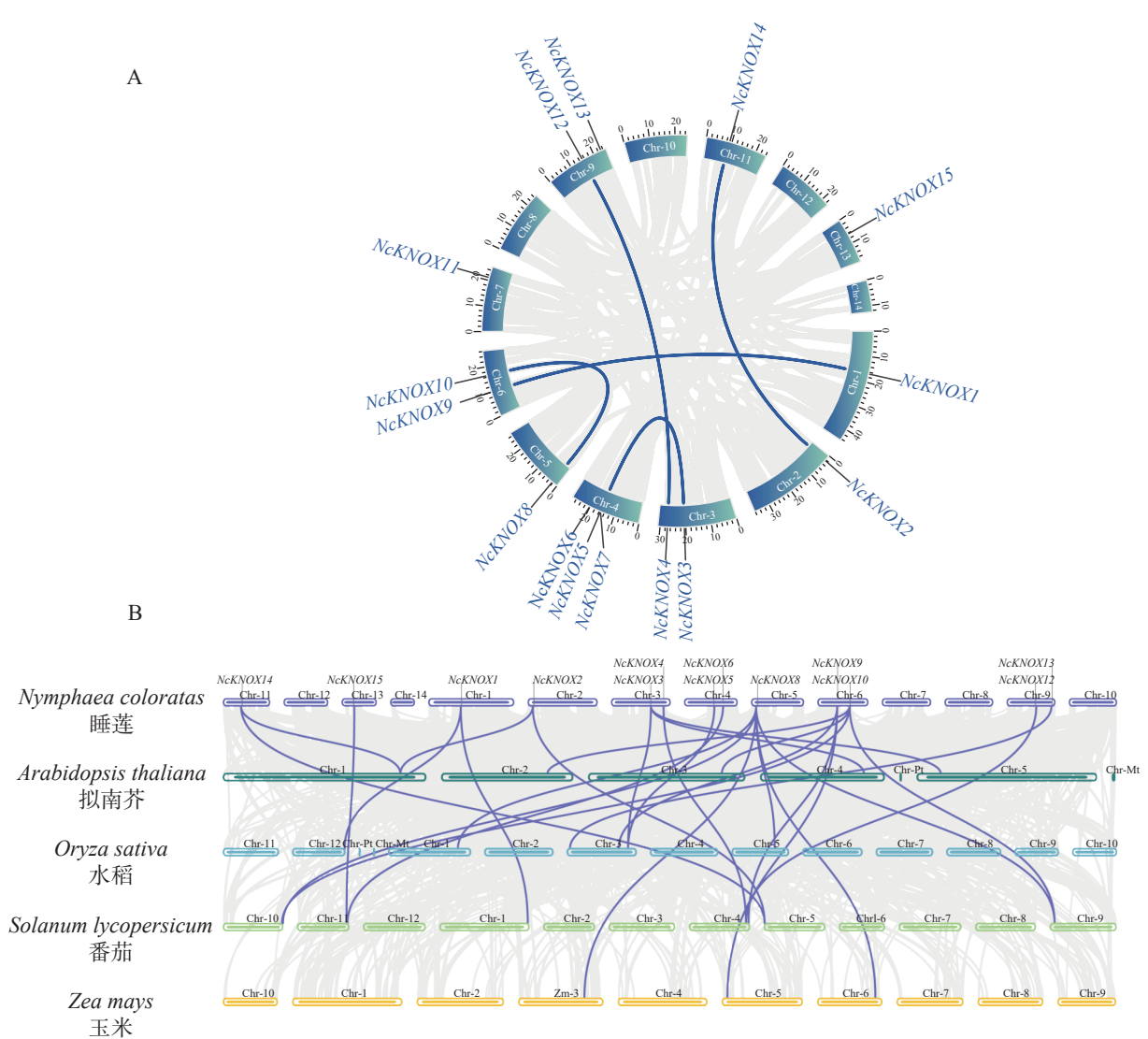


图 7 睡莲 *NcKNOX* 基因染色体定位分析
Fig. 7 Chromosomal localization analysis of *NcKNOX* genes in water lily



A.睡莲 *NcKNOX* 基因种内共线性分析; B.睡莲 *NcKNOX* 基因与拟南芥、水稻、番茄及玉米 *KNOX* 基因的种间共线性分析。
A. Intraspecific synteny analysis of *NcKNOX* genes in water lily; B. Interspecific synteny analysis of *NcKNOX* genes in water lily with *KNOX* genes from Arabidopsis, rice, tomato, and maize.

图 8 *KNOX* 基因共线性分析
Fig. 8 Synteny analysis of *KNOX* genes

表 2 睡莲 *NcKNOX* 基因与拟南芥等物种基因组产生共线性的基因汇总

Tab. 2 Summary of *NcKNOX* genes in water wily with synteny to the genomes of *Arabidopsis* and other species

基因ID Gene ID				
<i>Nymphaea colorata</i> 睡莲	<i>Arabidopsis thaliana</i> 拟南芥	<i>Oryza sativa</i> 水稻	<i>Solanum lycopersicum</i> 番茄	<i>Zea mays</i> 玉米
<i>NcKNOX1</i>	—	Os12t0636200	Solyc01g109980	—
<i>NcKNOX2</i>	<i>KNAT2</i>	—	Solyc05g005090	—
<i>NcKNOX3</i>	<i>KNAT4, KNAT5</i>	Os03t0771500	—	—
<i>NcKNOX4</i>	—	—	Solyc04g077210	—
<i>NcKNOX5</i>	—	Os03t0771500	—	—
<i>NcKNOX6</i>	—	—	Solyc11g069890	—
<i>NcKNOX7</i>	—	—	—	—
<i>NcKNOX8</i>	—	Os01t0848400	Solyc09g011380	Zm00001eb147970
<i>NcKNOX9</i>	—	Os05t0455200	Solyc10g086640	Zm00001eb289480
<i>NcKNOX10</i>	AT2G37900	Os03t0680800	Solyc04g080790	Zm00001eb217350
<i>NcKNOX11</i>	AT3G53960	—	Solyc09g011390	—
<i>NcKNOX12</i>	—	—	Solyc10g084950	—
<i>NcKNOX13</i>	—	—	—	—
<i>NcKNOX14</i>	—	Os03t0123500	Solyc04g077210	—
<i>NcKNOX15</i>	<i>KNAT2</i>	—	Solyc05g005090	—
	—	—	Solyc11g068950	—

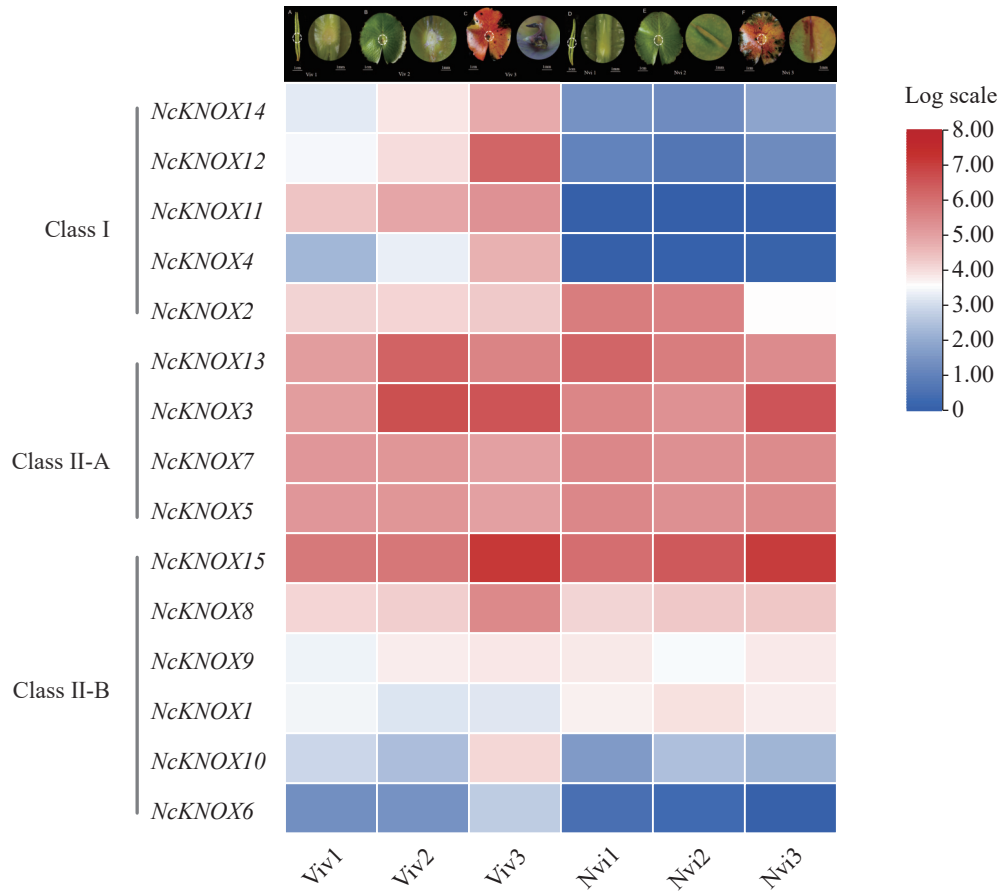


图 9 不同发育时期睡莲胎生叶片中 *NcKNOX* 基因的表达模式

Fig. 9 Expression pattern of *NcKNOX* genes in viviparous water lily leaves at different developmental stages

起到重要作用,与拟南芥、番茄的相关研究类似^[25-26]。

3 讨论

本研究利用 KNOX 家族的保守结构域和多种模式植物的 *KNOX* 基因序列进行比对分析,在睡

莲的基因组中成功鉴定出了 15 个 *NcKNOX* 基因。根据进化关系将睡莲 *NcKNOX* 基因分为 I 和 II 类, II 类亚家族又分为 II -A 和 II -B 类,其中, II -B 类可能是睡莲中特有的类别,其在 KNOX 保守结构域上极具特殊性,仅含有 Homeobox_KN 结构

域。根据前人的研究及睡莲 *NcKNOX* 的序列保守性分析, Homeobox_KN 结构域具有高度保守性^[27], 使其在复杂的进化过程中稳定存在、不易丢失, 这可能是 II-B 类 *NcKNOX* 基因仅含有 Homeobox_KN 结构域的重要因素之一。

KNOX 基因普遍含有 *KNOX1*、*KNOX2*、*ELK* 和 Homeobox_KN 4 种结构域, 但在睡莲 *NcKNOX* 基因家族中, *ELK* 结构域仅存在于 I 类 *NcKNOX* 基因和 II-A 类的 *NcKNOX5*。在拟南芥中, *ELK* 结构域通过调节 *STM* 基因的核定位在其翻译后的调控阶段发挥重要作用^[28], 说明在睡莲中具有该结构域的 Class I 及 Class II-A 的基因成员可能参与 *STM* 转录网络的特殊调控和特异性响应。同时在系统进化树分析中, II 类亚家族基因 *NcKNOX3*、*NcKNOX5* 和 *NcKNOX7* 与番茄的 *LeTl2* (Solyc07g007120)、拟南芥的 *KNAT3* 和 *KNAT4* 在进化关系上较为相近, 且 *LeTl2* 多数在成熟组织中表达^[29], *KNAT3* 和 *KNAT4* 在花序中表达, 调控拟南芥珠被发育^[30]。由此推测, II 类 *NcKNOX* 基因多数会在成熟部位表达, 可能参与叶片发育后期及衰老过程的调控^[25]。

启动子顺式作用元件分析显示, *NcKNOX* 基因家族成员主要受到植物激素和光响应类元件的调控, 与金鱼草的研究结果相似^[31], 可能参与光反应和胁迫反应, 还与植物体内的脱落酸、赤霉素等多种激素存在关联。在苹果中, *MdKNOX19* 基因通过调节脱落酸 (ABA) 的敏感性来抑制种子的萌发和器官的发育过程^[32]。在拟南芥中, *STM* 基因可以激活 SAM 中的细胞分裂素 (CK) 生物合成的相关元件, 促进分生组织的形成^[33]。因此, 睡莲 *NcKNOX* 基因也可能会通过响应不同激素的顺式作用元件, 作为激素信号转导途径中的关键节点, 调节植物生长发育, 参与细胞伸长、分裂和组织分化等过程。同样地, 光响应相关的元件在睡莲 *KNOX* 基因中数量也较高, 说明睡莲 *KNOX* 基因也可能通过光响应元件参与调控植物的光合作用、叶片形态和生长方向等光形态建成过程。

基因复制在物种进化中具有重要的推动作用, 主要包括全基因组复制/片段复制和串联复制 2 种机制^[34]。染色体定位分析发现, *NcKNOX* 基因家族成员分别分布于 10 条染色体, 可能反映了睡莲进化历史上发生的全基因组复制事件^[35], 从而产生大量冗余基因, 随后通过功能分化或亚功能化适

应新环境; 4 号染色体上 *NcKNOX5* 和 *NcKNOX7* 成簇排列, 则说明存在串联重复, 它们可能共享调控元件 (如启动子或增强子), 形成协同表达模块, 在特定发育阶段 (如器官边界形成或逆境响应) 中发挥叠加或互补作用。此外, 15 个 *NcKNOX* 基因在 10 条不同染色体上的分布是不均匀的, 种内共线性分析结果显示睡莲染色体上存在 5 组明确的片段复制对和一组串联重复现象, 说明小规模重复 (small scale duplication, SSD) 可能在睡莲 *KNOX* 基因家族扩增中发挥关键作用^[36]。

从睡莲叶片不同发育时期的 *NcKNOX* 基因的表达热图分析得知, II 类 *NcKNOX* 基因的整体表达水平高于 I 类 *NcKNOX* 基因, 这表明 II 类中的基因可能对睡莲叶片的整个生长周期更为重要。但是多数 I 类 *KNOX* 基因在胎生和非胎生叶片的表达量差异较大, 表现为在非胎生叶片的表达量较低, 在胎生叶片的表达量较高, 且随着胎生芽的发育表达量呈明显上升的趋势。在拟南芥中, I 类 *KNOX* 基因如 *STM*、*KNAT1*、*KNAT6* 对分生组织分化、胚胎发生等过程中发挥关键作用, 还可以维持分生组织功能以及塑造叶片形态的多样性^[10]。因此, 推测睡莲中 I 类 *KNOX* 基因可能对胎生叶片中的胎生芽的发生具有重要的意义, 其中, *NcKNOX4* 和 *NcKNOX11* 的表达差异最大, 可能对睡莲叶片胎生现象的产生起到主要作用。

4 结 论

本研究基于睡莲基因组数据, 通过系统分析鉴定出 15 个 *NcKNOX* 家族成员, 分为 Class I 和 Class II 两类, 其中, Class II-B 可能是睡莲独有的类型。睡莲 *KNOX* 基因存在大量的 CGTCA-motif、AuxRE 及 ABRE 等激素元件, 显示其对生长素、脱落酸等关键植物激素具有高度敏感性, 基因共线性分析显示其染色体上存在五组明确的片段复制事件。结构域分析表明, 睡莲 *KNOX* 家族基因可能具有和拟南芥、水稻同源基因类似的功能, 在睡莲生长发育调控网络中的核心地位, 特别是在响应激素信号、调控分生组织活性以及促进器官分化等关键生物学过程中发挥着不可或缺的作用。睡莲叶片的表达模式表示 I 类和 II 类 *KNOX* 基因对睡莲叶片的整个生长周期起到重要作用, 其中 I 类 *KNOX* 基因可能对睡莲叶片胎生芽的产生和发育起到关键作用。

参考文献:

- [1] GAO J, YANG X, ZHAO W, et al. Evolution, diversification, and expression of KNOX proteins in plants[J]. *Frontiers in Plant Science*, 2015, 6: 882.
- [2] SAKAMOTO T, NISHIMURA A, TAMAOKI M, et al. The conserved KNOX domain mediates specificity of tobacco KNOTTED1-type homeodomain proteins[J]. *The Plant Cell*, 1999, 11(8): 1419 – 1431.
- [3] WANG L, YANG X Y, GAO Y Q, et al. Genome-wide identification and characterization of TALE superfamily genes in soybean (*Glycine max* L.) [J]. *International Journal of Molecular Sciences*, 2021, 22(8): 4117.
- [4] ZHANG D Y, LAN S R, YIN W L, et al. Genome-wide identification and expression pattern analysis of KNOX gene family in Orchidaceae[J]. *Frontiers in Plant Science*, 2022, 13: 901089.
- [5] TSUDA K, ITO Y, SATO Y, et al. Positive autoregulation of a KNOX gene is essential for shoot apical meristem maintenance in rice[J]. *The Plant Cell*, 2011, 23(12): 4368 – 4381.
- [6] WANG S M, YAMAGUCHI M, GRIENENBERGER E, et al. The Class II KNOX genes *KNAT3* and *KNAT7* work cooperatively to influence deposition of secondary cell walls that provide mechanical support to Arabidopsis stems[J]. *The Plant Journal*, 2020, 101(2): 293 – 309.
- [7] MAGNANI E, HAKE S. KNOX lost the OX: the *Arabidopsis* *KNATM* gene defines a novel class of KNOX transcriptional regulators missing the homeodomain[J]. *The Plant Cell*, 2008, 20(4): 875 – 887.
- [8] PENG J L, YU J B, WANG H L, et al. Regulation of compound leaf development in *Medicago truncatula* by *fused compound leaf1*, a class M KNOX gene[J]. *The Plant Cell*, 2011, 23(11): 3929 – 3943.
- [9] JIA P, WANG Y, SHARIF R, et al. KNOTTED1-like homeobox (KNOX) transcription factors-Hubs in a plethora of networks: a review[J]. *International Journal of Biological Macromolecules*, 2023, 253(Pt 3): 126878. doi:10.1016/j.ijbiomac.2023.126878
- [10] RAGNI L, BELLES-BOIX E, GÜNL M, et al. Interaction of *KNAT6* and *KNAT2* with *BREVIPEDICELLUS* and *PENNYWISE* in *Arabidopsis* inflorescences[J]. *The Plant Cell*, 2008, 20(4): 888 – 900.
- [11] HAY A, TSIANTIS M. KNOX genes: versatile regulators of plant development and diversity[J]. *Development*, 2010, 137(19): 3153 – 3165.
- [12] KIM D, CHO Y H, RYU H, et al. BLH1 and *KNAT3* modulate ABA responses during germination and early seedling development in *Arabidopsis* [J]. *The Plant Journal*, 2013, 75(5): 755 – 766.
- [13] LI E Y, BHARGAVA A, QIANG W Y, et al. The Class II KNOX gene *KNAT7* negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus* [J]. *New Phytologist*, 2012, 194(1): 102 – 115.
- [14] WANG S G, YANG H L, MEI J S, et al. Rice homeobox protein *KNAT7* integrates the pathways regulating cell expansion and wall stiffness[J]. *Plant Physiology*, 2019, 181(2): 669 – 682.
- [15] 赵军, 徐芳, 吉腾飞, 等. 睡莲属植物化学成分及生物活性研究进展[J]. *天然产物研究与开发*, 2014, 26(1): 142 – 147.
- [16] 余翠薇, 陈煜初, 余东北, 等. 不同类型睡莲的特点及栽培技术[J]. *浙江农业科学*, 2016, 57(10): 1694 – 1695.
- [17] CHEN F, DONG W, ZHANG J W, et al. The sequenced angiosperm genomes and genome databases[J]. *Frontiers in Plant Science*, 2018, 9: 418.
- [18] CHEN C J, WU Y, LI J W, et al. TBtools- II : a “one for all, all for one” bioinformatics platform for biological big-datamining[J]. *MolecularPlant*, 2023, 16(11): 1733–1742.
- [19] 李清雪, 黎洁, 许慧娴, 等. 热带睡莲叶片胎生苗的生长机理[J/OL]. *热带亚热带植物学报*, 1-11(2024-04-30)[2025-03-16]. <http://kns.cnki.net/kcms/detail/44.1374.q.20240429.1843.002.html>.
- [20] 赵长江, 宋巨奇, 都梦翔, 等. 玉米 KNOX 基因家族鉴定及组织和逆境表达分析[J]. *西北植物学报*, 2021, 41(7): 1109 – 1119.
- [21] 宋宁宁, 梁慧慧, 安义伟, 等. 小麦 KNOX 基因家族鉴定与生物信息学分析[J]. *分子植物育种*, 2022, 20(7): 2105 – 2117.
- [22] 施卫萍, 张欣, 赵媛, 等. 谷子 KNOX 基因家族鉴定及组织表达模式分析[J]. *农业生物技术学报*, 2024, 32(8): 1742 – 1752.
- [23] HAKE S, SMITH H M S, HOLTAN H, et al. The role of *Knox* genes in plant development[J]. *Annual Review of Cell and Developmental Biology*, 2004, 20: 125 – 151.
- [24] ICHIHASHI Y, AGUILAR-MARTÍNEZ J A, FARHI M, et al. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(25): E2616 – E2621.
- [25] LINCOLN C, LONG J, YAMAGUCHI J, et al. A knotted1-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants[J]. *The Plant Cell*, 1994, 6(12): 1859 – 1876.
- [26] KEREN-KEISERMAN A, SHTERN A, LEVY M, et al. CLASS- II KNOX genes coordinate spatial and temporal ripening in tomato[J]. *Plant Physiology*, 2022, 190(1): 657 – 668.
- [27] MUKHERJEE K, BROCCIERI L, BÜRGLIN T R. A comprehensive classification and evolutionary analysis of plant homeobox genes[J]. *Molecular Biology and Evolution*, 2009, 26(12): 2775 – 2794.
- [28] AGUILAR-MARTÍNEZ J A, UCHIDA N, TOWNSELEY B, et al. Transcriptional, posttranscriptional, and posttranslational regulation of *SHOOT MERISTEMLESS* gene expression in *Arabidopsis* determines gene function in the shoot apex[J]. *Plant Physiology*, 2015, 167(2): 424 – 442.
- [29] JANSSEN B J, WILLIAMS A, CHEN J J, et al. Isolation and characterization of two knotted-like homeobox genes from tomato[J]. *Plant Molecular Biology*, 1998, 36(3): 417 – 425.
- [30] CHEN J J, WANG W, QIN W Q, et al. Transcription

- factors KNAT3 and KNAT4 are essential for integument and ovule formation in Arabidopsis[J]. *Plant Physiology*, 2023, 191(1): 463 – 478.
- [31] 顾伟卓, 汪仲毅, 杨洁, 等. 金鱼草 *KNOX* 基因家族鉴定及调控其雄蕊瓣化的候选基因挖掘[J]. *西南农业学报*, 2024, 37(6): 1329 – 1339.
- [32] 贾鹏. MdKNOX 转录因子通过 GA/ABA 途径调控苹果生长发育研究[D]. 杨凌: 西北农林科技大学, 2021. doi:10.27409/d.cnki.gxbnu.2021.000159
- [33] SCOFIELD S, DEWITTE W, NIEUWLAND J, et al. The Arabidopsis homeobox gene *SHOOT MERISTEMLESS* has cellular and meristem-organisational roles with differential requirements for cytokinin and CYCD3 activity[J]. *The Plant Journal*, 2013, 75(1): 53 – 66.
- [34] ZHU Y, WU N N, SONG W L, et al. Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies[J]. *BMC Plant Biology*, 2014, 14: 93.
- [35] ZHANG L S, CHEN F, ZHANG X T, et al. The water lily genome and the early evolution of flowering plants[J]. *Nature*, 2020, 577(7788): 79 – 84.
- [36] TAN S J, MA H J, WANG J B, et al. DNA transposons mediate duplications via transposition-independent and-dependent mechanisms in metazoans[J]. *Nature Communications*, 2021, 12(1): 4280.

Identification and bioinformatics analysis of KNOX gene family in water lily (*Nymphaea*)

XU Huixian[#], LI Jie, LI Meier, ZHANG Ruxin, WANG Tongxin,
LI Tingge, ZHAO Ying, WANG Jian*

(School of Tropical Agriculture and Forestry, Hainan University/Key Laboratory of Germplasm Resources Biology of Tropical Special Ornamental Plants of Hainan/Key Laboratory of Genetics and Germplasm Innovation of Tropical Forest Trees and Ornamental Plants, Ministry of Education, Danzhou, Hainan, 571737, China)

Abstract: The KNOX gene family encodes homeobox proteins that function as transcription factors and are integral to regulating plant growth and development. To investigate the bioinformatic characteristics of the KNOX gene family and its potential roles during viviparous seedling development in water lily (*Nymphaea*), this study identified 15 KNOX genes from the *Nymphaea* genome using bioinformatics approaches. These genes were distributed across chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, and 13. Detailed analyses were performed to assess their physicochemical properties, cis-regulatory elements, and interspecies collinearity. The physicochemical characterization revealed that the KNOX proteins ranged from 249 to 2270 amino acids in length, were hydrophilic, and had molecular weights between 2.84 and 24.95 kDa. Subcellular localization predictions indicated that most KNOX proteins localized in the nucleus. Furthermore, the majority of these proteins contained four conserved domains—KNOX1, KNOX2, ELK, and Homeobox KN. Based on structural features and phylogenetic analysis, the KNOX gene family was classified into Class I and Class II, with Class II-B potentially representing a specific clade unique to *Nymphaea*. Interspecies collinearity analysis revealed that 13 of the KNOX genes shared conserved synteny with KNOX genes from *Arabidopsis thaliana*, *Solanum lycopersicum*, *Oryza sativa*, and *Zea mays*, with the highest level of homology observed with tomato. Additionally, promoter analysis identified numerous cis-regulatory elements related to growth and development, hormone response, and stress response. Expression pattern analysis across different development stages of leaves suggested that Class I genes likely played a pivotal role in the formation and development of leaf viviparous shoots in *Nymphaea*.

Keywords: *Nymphaea*; KNOX gene family; leaf vivipary

(责任编辑: 潘学峰)