

油茶 IDD 基因家族鉴定与生物信息学分析

陈健鑫¹, 吴峰婧琳¹, 魏玉倩¹, 杨娅琳¹, 马焕成², 杨红玉^{3*}, 伍建榕^{1,2*}

1. 西南林业大学生物多样性保护学院/云南省高校森林灾害预警控制重点实验室, 云南昆明 650224; 2. 西南林业大学林学院/西南地区生物多样性保育国家林业局重点实验室, 云南昆明 650224; 3. 昆明学院农学与生命科学学院, 云南昆明 650214

摘要: 油茶 (*Camellia oleifera*) 是山茶科 (Theaceae) 的一种特有的油料植物, 具有重要的观赏价值和经济价值, 随着油茶的集约化种植, 病虫害大量发生, 加之良种选育滞后, 严重制约着油茶产业的发展。INDETERMINATE DOMAIN (IDD) 家族是高等植物中一类保守的转录因子, 通过介导植物内源激素进而调控植物的基础免疫反应。为鉴定油茶中 IDD 基因家族成员, 本研究以油茶全基因组信息为参考, 通过 IDD 基因家族保守结构域鉴定了 29 个油茶 CoIDD 基因家族成员, 进一步明确了各基因的结构、理化性质、系统进化关系、亚细胞定位。通过 RT-PCR 克隆到 1 个与 *AtIDD4/5/6* 同源的 *CoIDD4* 基因, 明确了其理化性质、表达模式及核定位信号, 同时分析了其启动区顺式作用元件的类型。研究结果表明, 29 个 CoIDD 基因家族成员分为 4 个亚组, 平均亲水系数均小于 0, 属于亲水性蛋白, 且亚细胞定位均位于细胞核内; 克隆到 *CoIDD4* 基因的全长 ORF 框包含 1620 bp 碱基, 编码 539 个氨基酸, 主要在叶片和花器官中表达; 该基因启动子区顺式作用元件主要是参与光响应和植物激素响应元件两大类。综上所述, 本研究在油茶中鉴定了 29 个 IDD 基因家族成员, 其中 *CoIDD4* 基因与 *AtIDD4/5/6* 亲缘关系较近, 推测其可能参与油茶激素信号转导和基础免疫反应调节等进程, 为油茶 IDD 转录因子功能解析提供理论基础, 也为油茶的抗性优良品种选育及分子育种提供参考依据。

关键词: 油茶; IDD 基因家族; 生物信息学分析; 植物激素; 抗性调控

中图分类号: S432 **文献标识码:** A

Identification and Bioinformatics Analysis of IDD Gene Family in *Camellia oleifera*

CHEN Jianxin¹, WU Fengjinglin¹, WEI Yuqian¹, YANG Yalin¹, MA Huancheng², YANG Hongyu^{3*}, WU Jianrong^{1,2*}

1. College of Biodiversity Conservation, Southwest Forestry University / Key Laboratory of Forest Disaster Warning and Control in Universities of Yunnan Province, Kunming, Yunnan 650224, China; 2. College of Biodiversity Forestry, Southwest Forestry University / Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Kunming, Yunnan 650224, China; 3. College of Agronomy and Life Sciences, Kunming University, Kunming, Yunnan 650214, China

Abstract: *Camellia oleifera* is a unique oil plant of Theaceae, which has important ornamental value and economic value. With the intensive planting of *Camellia*, many diseases and pests occur. In addition, the delayed breeding of improved varieties seriously restricts the development of the *Camellia* industry. INDETERMINATE DOMAIN (IDD) family is a kind of conservative transcription factors in higher plants, which regulate the basic immune response of plants by mediating plant endogenous hormones. In order to identify the members of IDD gene family in *C. oleifera*, 29 members of CoIDD gene family were identified through the conserved domain of IDD gene family with reference to the whole

收稿日期 2022-12-26; **修回日期** 2023-02-02

基金项目 国家自然科学基金项目 (No. 31860208); 国家重点研发计划项目 (No. 2019YFD100200X); 云南省教育厅科学研究基金项目 (No. 2023Y0720)。

作者简介 陈健鑫 (1996—), 男, 博士研究生, 研究方向: 森林病理学与分子生物学。*通信作者 (Corresponding author): 伍建榕 (WU Jianrong), E-mail: 1176279044@qq.com; 杨红玉 (YANG Hongyu), E-mail: yanghongyukm@126.com。

genome information of *C. oleifera*, and further clarified the structure, physical and chemical properties, phylogenetic relationship, and sub cell localization of each gene. The *CoIDD4* gene homologous to *AtIDD4/5/6* was cloned by RT-PCR, and its physicochemical properties, expression mode and nuclear localization signal were clarified. At the same time, the types of cis acting elements in its promoter region were analyzed. The results showed that 29 *CoIDD* gene family members were divided into 4 subgroups, with average hydrophilicity coefficients less than 0, belonging to hydrophilic proteins, and subcellular localization was located in the nucleus; The full-length ORF frame cloned into *CoIDD4* gene contained 1620 bp base and encoded 539 amino acids, which were mainly expressed in leaf and flower organs; The cis acting elements in the promoter region of the gene were mainly involved in light response and plant hormone response elements. In conclusion, this study identified 29 IDD gene family members in *C. oleifera*, among which *CoIDD4* gene was closely related to *AtIDD4/5/6*. It is speculated that *CoIDD4* gene may be involved in the process of hormone signal transduction and basic immune response regulation of *C. oleifera*, which would provide a theoretical basis for the functional analysis of IDD transcription factors in *C. oleifera* and provide a reference basis for the selection and molecular breeding of superior resistant varieties of *C. oleifera*.

Keywords: *Camellia oleifera*; IDD gene family; bioinformatics analysis; plant hormone; resistance regulation

DOI: 10.3969/j.issn.1000-2561.2024.02.003

油茶(*Camellia oleifera*)是山茶科(Theaceae)山茶属(*Camellia*)的一种灌木或中乔木,是我国一种特有的油料植物。油茶在我国有悠久的种植和栽培历史,从长江流域到华南地区等 14 个省份被广泛的栽培^[1]。鲜榨茶油中富含不饱和脂肪酸以及多酚、黄酮、原花青素、总三萜和木脂素等活性物质,是一种高质量、耐储存的健康食用油^[2]。油茶在云南有悠久的种植历史,主栽的品种是白花油茶和腾冲红花油茶,截至 2019 年底,云南省共发展油茶 23 万 hm^2 ,综合产值 7.9 亿元,位列全国第 11 位^[3-4]。随着油茶集约化种植和油茶产业的快速发展,加之良种选育滞后及栽培管护技术落后导致油茶病虫害发生严重,目前,油茶病虫害的防控仍是以化学防治为主的单一防治措施,一方面病原物的抗药性逐渐增强,另一方面也给生态环境和人类健康带来了负面影响。因此,油茶抗逆相关基因亟需解析。

INDETERMINATE DOMAIN (IDD) 基因家族是 C2H2 型锌指蛋白转录因子的一类,IDD 基因家族的保守 N 端以 DNA 结合域(不定域)为特征,包含 2 个 C2H2 和 2 个 C2HC 锌指结构域,植物中的 IDD 基因通过与多种蛋白质相互作用,与激素级联重叠,控制或调节植物的生长发育与抗逆反应应答^[5]。ID-domain 基因最初由玉米中 1 个典型的开花基因 INDETERMINATE1 (*ID1*) 而被报道^[6],*ID1* 基因编码 1 个包含 4 个锌指单元的核蛋白,具有 C2H2 型锌指蛋白的结构特点,在植物中通过 *ID1* 基因的 ID-domain 所定义的基因家族成为了 C2H2 型锌指蛋白的一个亚族,即 IDD 锌指蛋白家族。在拟南芥(*Arabidopsis thaliana*)

中发现了 16 个 IDD 成员,命名为 *AtIDD1-16*,构建基于氨基酸的系统发育树发现,*AtIDD4*、*AtIDD5* 和 *AtIDD6* 位于同一进化支内,且 *IDD4* 与 *IDD6* 关系较近,二者氨基酸的一致性为 64%,3 个 IDD 基因氨基酸序列的一致性为 53%,3 个蛋白的氨基酸序列在 ID-domain 区域的一致性高达 96%^[7-8]。研究发现 *AtIDD4* 基因与根系的发育和叶片极性有关并参与 DELLA 相互作用,*AtIDD5* 基因是一类与 DELLA 相互作用的蛋白;*AtIDD6* 基因与根系的发育有关^[9],*idd4* 拟南芥缺失突变体表现了对丁香假单胞的抗性,揭示了 *AtIDD4* 作为水杨酸(Salicylic acid, SA)信号通路的负调控因子,参与植物激素信号网络转导途径和基础免疫反应^[10]。随着基因组测序技术和分子生物学技术的发展,已有多物种中的 IDD 家族成员被鉴定并报道了其重要的生物学功能。通过 IDD 家族的保守结构域在毛竹(*Phyllostachys edulis*)中鉴定了 32 个具有完整 IDD-domain 的成员,证实 *PheIDD21* 受蔗糖诱导后是侧调控芽发育的关键因子^[11]。甘蓝型油菜(*Brassica napus*)在我国食用油产业中具有举足轻重的地位,在油菜中鉴定了 58 个候选 IDD 蛋白,其中 *Bna.A08IDD7a* 基因通过介导赤霉素(Gibberellin, GA)和生长素(auxin, IAA)的信号转导过程而参与根系生长的调控^[12]。

IDD 基因家族在植物生长发育、植物激素信号转导通路及基础免疫调控中具有重要作用,但在油茶中鲜有研究报道。本研究基于油茶基因组数据鉴定 IDD 基因家族成员,明确各个基因间的系统发育关系,并利用生物信息学方法分析 IDD

基因家族蛋白的理化性质、Motif 预测、亚细胞定位等情况,同时鉴定并克隆 *AtIDD4/5/6* 的同源基因 *CoIDD4* 及分析其表达特征,为后续油茶分子育种和抗病品种的选育提供重要理论依据。

1 材料与方法

1.1 材料

供试植物:本研究所用油茶采集自云南省德宏州梁河县油茶种植基地。选取长势一致的1年生油茶实生苗的叶片组织用于 *CoIDD* 基因的克隆,选取梁河县翁冷村12~15年生油茶,分别收集叶、花、果实和根系组织用于 *CoIDD* 基因表达模式分析。每个组织至少收集3套生物学模板,收集后立即用液氮速冻,于-80℃长期保存。

参考基因组及参考序列:油茶全基因组(ASM2231669v1)信息下载自NCBI数据库(https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_022316695.1/)^[13],拟南芥IDD基因家族氨基酸序列及核苷酸序列下载自TAIR数据库(<https://www.arabidopsis.org>)。

供试试剂:TRIZOL、Ex Taq、10×Ex Taq Buffer、dNTP Mixture、DH5α感受态细胞、pMD™19-T Vector Cloning Kit, PrimeScript™ IV 1st strand cDNA Synthesis Mix 购买自TaKaRa宝生物工程(大连)有限公司。

1.2 方法

1.2.1 油茶IDD基因家族成员鉴定 在TAIR(<https://www.arabidopsis.org>)数据库中下载拟南芥IDD基因家族16个成员的氨基酸序列,建立本地种子序列数据库,利用BLAST软件将拟南芥IDD基因家族氨基酸序列与油茶蛋白质组序列数据库进行比对。筛选后的油茶蛋白质序列在NCBI数据库中进行blastp并分析其结构域,将未含有C2H2结构域的蛋白质剔除。同时将获得的氨基酸序列比对到油茶的转录组数据库(https://github.com/Hengfu-Yin/CON_genome_data)和基因组数据库,获得转录本序列和基因组序列用于后续分析。

候选的家族成员通过ExPASy(<https://web.expasy.org/protparam/>)数据库中的ProtParam工具,对CoIDD4蛋白的氨基酸大小、分子质量、等电点基本信息进行分析,并通过Plant-mPLoc(<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>)进行亚细胞定位预测,若蛋白定位于细胞核内,

则通过NoD(http://www.compbio.dundee.ac.uk/www-nod/form_action_seqInput.jsp)预测分析其核定位序列。

1.2.2 油茶IDD基因家族系统发育树构建 利用MEGA9软件将拟南芥和油茶IDD基因家族成员cDNA序列和氨基酸序列进行多重比对,并通过邻接法(neighbor-Joining, NJ)构建系统发育树,设置Bootstrap值为1000。

1.2.3 油茶IDD基因家族保守结构域(motif)分析 保守结构域是蛋白质中能够折叠成特定三维结构的一段相对区域,其结构亚单位称为基序(Motif),蕴含了各自的遗传进化信息。利用MEGA9软件将拟南芥和油茶IDD基因家族氨基酸序列进行多重比对,通过MEME(<http://meme-suite.org/tools/meme>)在线软件对拟南芥与油茶IDD基因家族序列进行motif搜索,设置分析motif数值为10,并绘制相应保守域结构的logo图。

1.2.4 油茶IDD4蛋白结构预测分析及系统发育分析 通过Phyre2(<http://www.sbg.bio.ic.ac.uk/phyre2/html/help.cgi?id=help/faq>)在线软件预测分析拟南芥AtIDD4/5/6蛋白和油茶CoIDD4蛋白的三级结构,以此分析其蛋白质的三维结构。将CoIDD4基因在NCBI数据库中进行BLAST分析,选取同源性较高的序列进行系统发育分析,利用MEGA中的NJ法构建系统发育树,Bootstrap值设为1000。

1.2.5 油茶IDD4基因启动子顺式作用元件分析及功能预测 根据油茶全基因组序列信息,利用TBtools提取CoIDD4基因序列上游长度为2 kb的启动子序列,并将其上传到PlantCARE(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)进行预测分析。

1.2.6 油茶IDD4基因ORF克隆及表达模式分析 将筛选得到的拟南芥AtIDD4/5/6基因的同源基因CoIDD4的氨基酸比对到油茶的转录组数据中,通过NCBI数据库在线工具ORF finder(<https://www.ncbi.nlm.nih.gov/orffinder/>)对该序列进行开放阅读框(ORF)预测,并通过Primer5软件设计扩增全长的引物(表1)。利用TRIZOL提取油茶叶片总RNA,通过NanoDrop 2000测定RNA浓度和纯度,合格样品利用PrimeScript™ IV 1st strand cDNA Synthesis Mix反转录合成第一链cDNA,以此为模板扩增CoIDD4基因的全长片段。扩增产物连接到pMD19-T载体后转化大肠

杆菌 DH5α感受态细胞，挑取阳性克隆子送昆明擎科生物有限公司测序，将插入片段正确的阳性克隆保存于-80 ℃超低温冰箱。

提取油茶植株的叶、花、果实和根系组织总 RNA 并反转录合成 cDNA，以油茶 3-磷酸甘油醛脱氢酶基因（*GAPDH*）为内参^[14]，通过 qPCR 检测 *CoIDD4* 基因在油茶不同组织内的表达模式。

表 1 本研究所用引物
Tab. 1 Primers used in this study

基因 Gene	引物名称 Primer name	引物序列（5'-3'） Primer sequence（5'-3'）	用途 Uage
<i>CoIDD4</i>	CoIDD4F1	ATGGCAACAAACCGGTCATT TGT	克隆
	CoIDD4R1	CTAGGGAGCACCAAATACATG GC	
<i>CoIDD4</i>	CoIDD4qF1	AACACGAAAGCCTCCCACCA	荧光 定量
	CoIDD4qR1	GCAGGAGAAGTGGCATTGGC	
<i>CoGAPDH</i>	CoGAPDHF1	CTACTGGAGTTTTCACCGA	
	CoGAPDHR2	TAAGACCCTCAACAATGCC	

表 2 *CoIDD* 蛋白基本信息
Tab. 2 Basic information of *CoIDD* proteins

基因 Gene	氨基酸数量 Number of amino acids	分子量 Mw/Da	等电点 pI	不稳定系数 Instability index	脂肪系数 Aliphatic index	平均亲水系数 GRAVY	亚细胞定位 Subcellular localization
<i>CoIDD1</i>	550	59 787.96	8.86	52.96	55.75	-0.693	Nucleus
<i>CoIDD2</i>	446	47 892.81	8.87	56.31	63.48	-0.539	Nucleus
<i>CoIDD3</i>	464	51 285.65	8.95	37.23	67.07	-0.625	Nucleus
<i>CoIDD4</i>	539	58 566.99	9.05	41.61	57.77	-0.663	Nucleus
<i>CoIDD5</i>	556	61 738.15	5.53	50.30	68.22	-0.629	Nucleus
<i>CoIDD6</i>	570	63 201.23	9.52	56.75	54.28	-0.748	Nucleus
<i>CoIDD7</i>	530	58 432.15	7.06	57.51	53.00	-0.749	Nucleus
<i>CoIDD8</i>	431	47 867.32	8.80	46.27	57.49	-0.694	Nucleus
<i>CoIDD9</i>	550	60 670.03	8.77	58.27	58.58	-0.707	Nucleus
<i>CoIDD10</i>	514	55 654.88	9.28	44.58	61.11	-0.487	Nucleus
<i>CoIDD11</i>	540	58 869.88	8.78	52.92	55.33	-0.719	Nucleus
<i>CoIDD12</i>	498	55 270.51	9.18	52.47	67.37	-0.651	Nucleus
<i>CoIDD13</i>	401	44 299.85	9.07	51.87	58.13	-0.676	Nucleus
<i>CoIDD14</i>	406	45 289.68	8.39	69.44	62.56	-0.812	Nucleus
<i>CoIDD15</i>	406	45 275.65	8.39	69.44	62.32	-0.813	Nucleus
<i>CoIDD16</i>	520	54 628.32	8.90	54.02	69.50	-0.332	Nucleus
<i>CoIDD17</i>	515	56 557.99	8.99	40.41	70.52	-0.402	Nucleus
<i>CoIDD18</i>	535	59 058.99	8.94	53.01	54.54	-0.702	Nucleus
<i>CoIDD19</i>	399	44 438.03	8.99	47.74	62.78	-0.686	Nucleus
<i>CoIDD20</i>	493	53 828.44	8.86	43.35	63.75	-0.604	Nucleus
<i>CoIDD21</i>	413	45 890.80	5.73	51.04	62.40	-0.563	Nucleus
<i>CoIDD22</i>	411	45 618.31	5.44	53.39	63.41	-0.585	Nucleus
<i>CoIDD23</i>	344	38 894.66	6.56	59.63	53.63	-0.721	Nucleus
<i>CoIDD24</i>	325	37 385.46	8.83	59.09	52.25	-0.851	Nucleus
<i>CoIDD25</i>	415	45 113.57	9.04	56.16	56.29	-0.684	Nucleus
<i>CoIDD26</i>	288	32 949.78	9.02	46.12	50.17	-0.725	Nucleus
<i>CoIDD27</i>	280	32 027.26	9.25	42.40	54.68	-0.895	Nucleus
<i>CoIDD28</i>	350	39 019.00	8.97	59.14	59.34	-0.680	Nucleus
<i>CoIDD29</i>	332	37 118.75	8.56	60.68	55.81	-0.704	Nucleus

2 结果与分析

2.1 油茶 *IDD* 基因家族成员鉴定及系统发育分析

利用拟南芥 *IDDs* 基因的氨基酸序列在油茶蛋白组数据库初步筛选出 35 条候选序列，通过 NCBI 数据库进行 blastp 分析去除不含 C2H2 结构域的序列后，最终获得 29 个油茶 *IDD* 基因家族成员，暂命名为 *CoIDD1*~*29*（表 2）。*CoIDD* 基因家族 29 个成员的理化性质分析表明，29 个 *IDD* 家族成员编码氨基酸长度在 280~570 个之间，氨基酸长度存在较大的变化；相对分子质量介于 32027.26~63201.23 Da 之间；等电点（pI）介于 5.44~9.52 之间，其中 pI<7 的酸性氨基酸有 4 个；pI>7 的碱性氨基酸有 25 个；不稳定系数介于 37.23~69.44 之间，有 1 个成员不稳定系数小于 40，属于稳定蛋白，其余成员均属于不稳定蛋白；29 个 *CoIDD* 家族成员的平均亲水系数均小于 0，

属于亲水性蛋白,且亚细胞定位均位于细胞核内。

采用邻接构建的拟南芥和油茶的 IDD 基因家族氨基酸序列的系统发育树共分为 4 个进化支 (图 1)。第 I 进化支由 CoIDD23/24/26/28/29 五个成员组成, 第 II 进化支由 CoIDD5/19/21/22/27 五个成员组成, 前 2 个分支首先从 IDD 基因家族中分离出来, 主要由油茶 IDD 家族成员组成, 不含有拟南芥 IDD 家族成员, 因此属于油茶中特有的 IDD 蛋白。由拟南芥 AtIDD14/15/16 和油茶

CoIDD14/15 五个成员共同组成了第Ⅲ进化支, 进化距离较近。第Ⅳ分支由众多拟南芥 AtIDDs 蛋白和油茶 CoIDDs 蛋白共同组成, 其中 CoIDD6/7/17/18 在该分支中更早的被分离出来, CoIDD4 与 AtIDD4/5/6 处于同一进化亚支内, 亲缘关系较近。由此可见, 油茶中存在与拟南芥 IDD 蛋白进化距离较近的同源 IDD 蛋白, 同时 CoIDD 蛋白与 AtIDD 蛋白也显示了较为明显的分化。

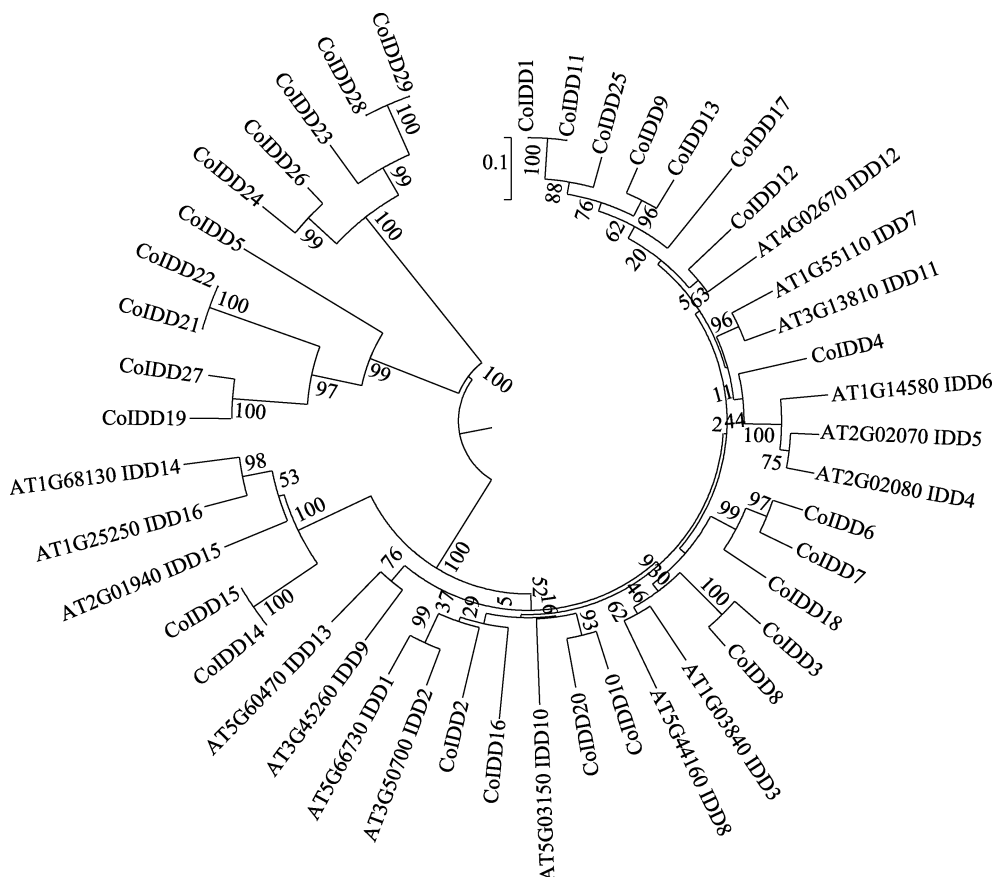


图 1 油茶和拟南芥 IDD 基因家族系统发育树

Fig. 1 Phylogenetic tree of IDD gene family in *C. oleifera* and *A. thaliana*

2.2 油茶 IDD 基因家族保守结构域分析

通过 MEME 软件分析 IDD 家族蛋白质的保守结构域的结果表明,油茶 CoIDD 蛋白与拟南芥 AtIDD 蛋白存在相似的保守结构域。AtIDD 基因所包含保守的 ID-domain 区域,含有一组 4 个串联的锌指结构 (ZF1~4)^[15]。在 CoIDD 蛋白的 10 个保守 motif 中, motif 1 (RFVCEICNKGFRDQNLQLHRRGHNLPWK) 最为保守 (每个 Motif 中不同位置处出现的氨基酸种类越少,说明该位置处保守性越高),该 motif 为 IDD 家族第 1 个锌指结构 ZF1; motif 2 (YAVOSDWKAHSKTCGTR

EYRCD) 为 IDD 家族第 3 个锌指结构 ZF3; motif 3 (CVHHDPsRALGDLTGIIKKHFCRKHGEKKW) 为 IDD 家族第 2 个锌指结构 ZF2; motif 4 为 IDD 家族第 4 个锌指结构 ZF4 (图 2)。

油茶和拟南芥 IDD 蛋白的保守结构域联合分析结果表明,拟南芥和油茶 IDD 蛋白结构域特征分化与共同构建的系统发育树总体趋势相吻合(图 3)。motif 1~4 代表了 IDD 家族中 4 个串联的锌指结构(ZF1~4)为最保守的结构,在油茶和拟南芥 IDD 蛋白中保守存在(图 4)。

位于系统发育树第Ⅲ进化支的拟南芥 AtIDD14/

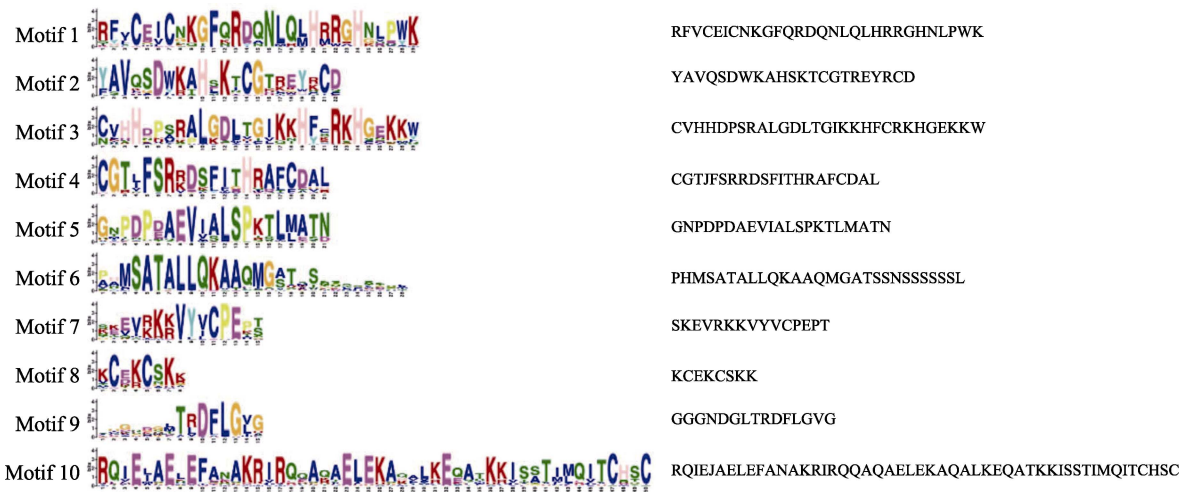


图 2 CoIDD 蛋白保守基序
Fig. 2 Motifs sequence of CoIDD protein

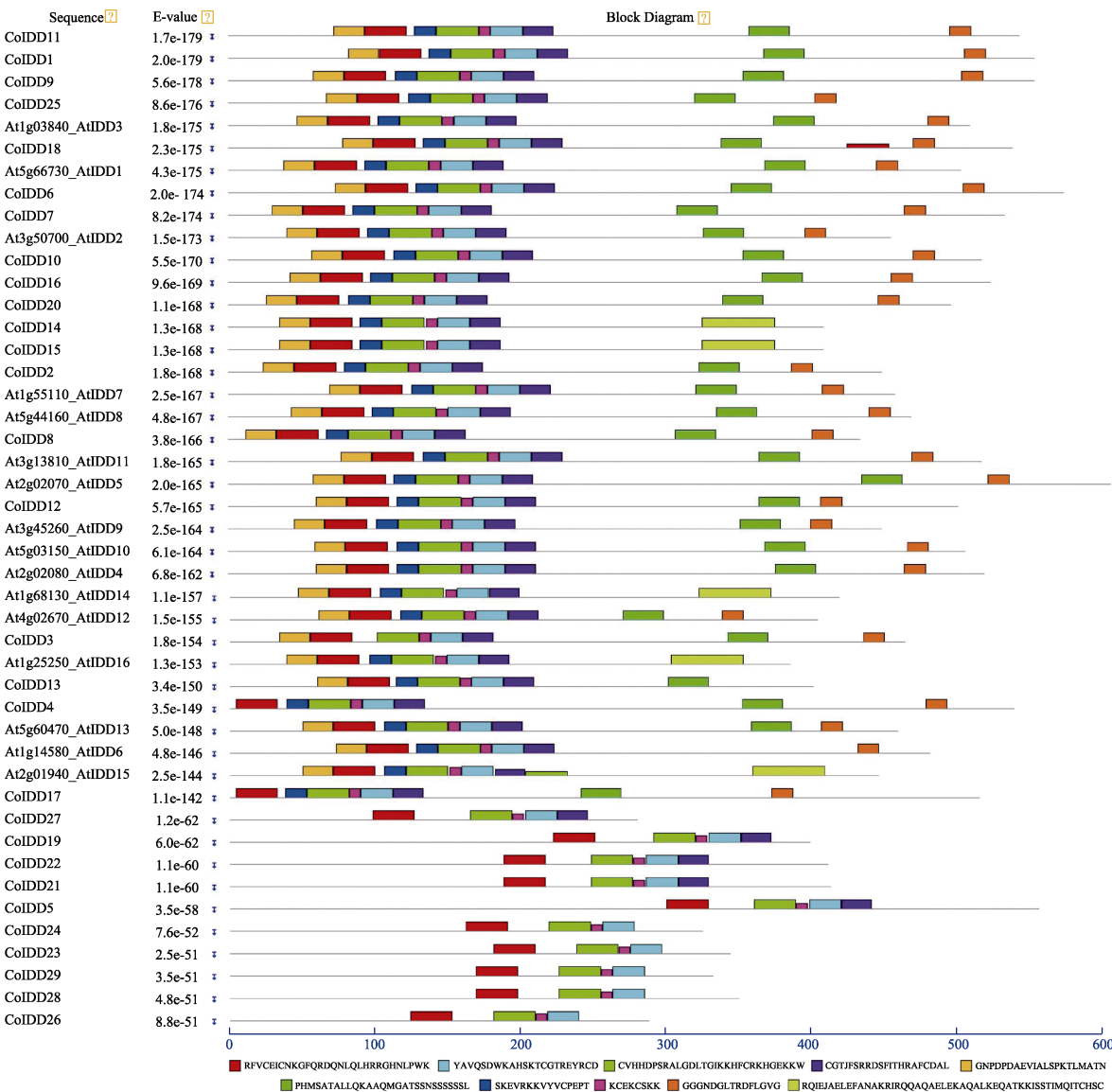


图 3 油茶和拟南芥 IDD 家族成员保守基序分析
Fig. 3 Analysis of motifs of IDD family members in *C. oleifera* and *A. thaliana*

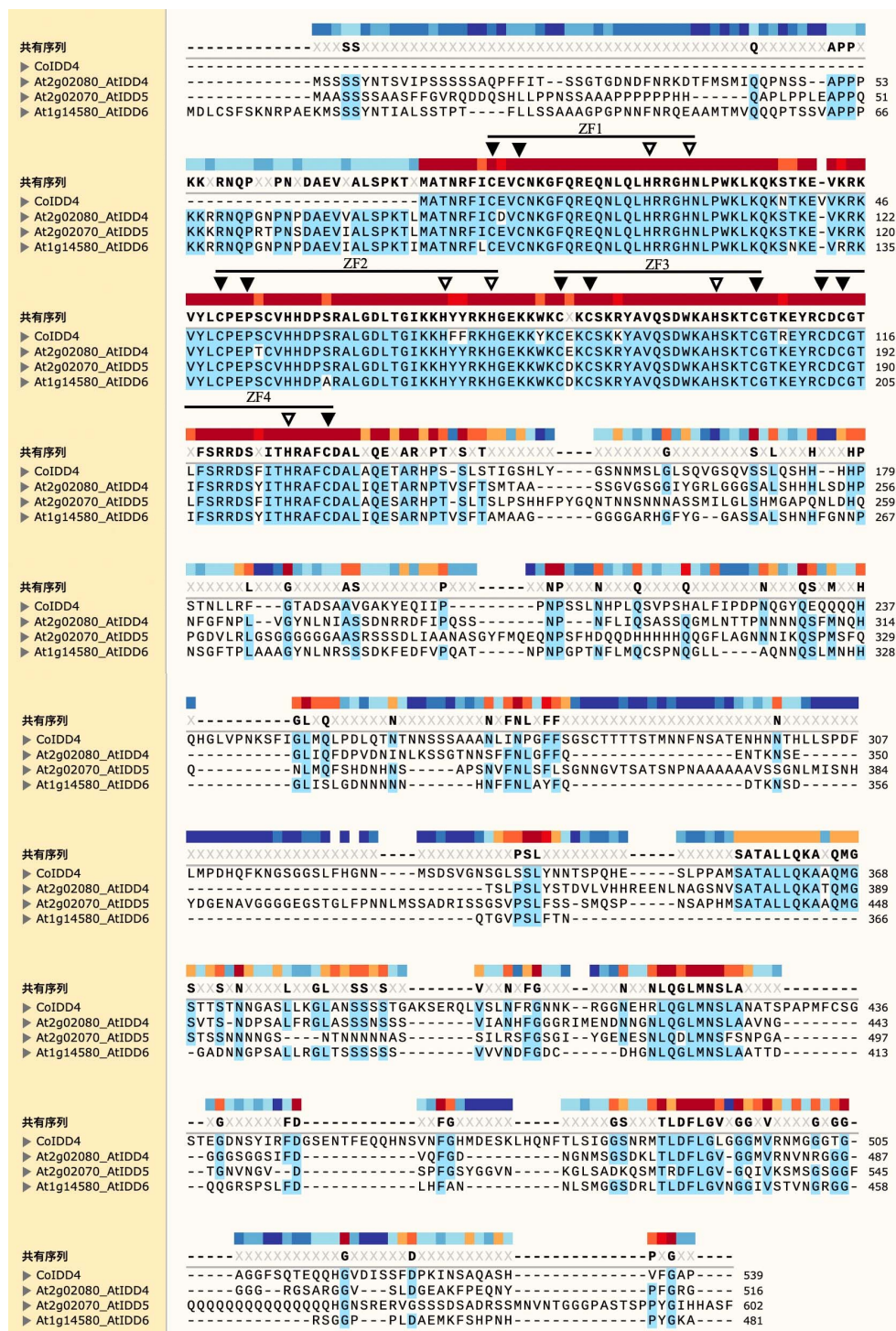


图 4 CoIDD4 和 AtIDD4/5/6 多序列比对

Fig. 4 Multiple alignments of CoIDD4 and AtIDD4/5/6

15/16 和油茶 CoIDD14/15 特异存在 motif 10 (RQ IEJAELEFANAKRIRQQAQAELEKAQALKEQAT KKISSTIMQITCHSO); 位于第 I ~ II 进化支内的成员缺少了 motif 5 (GNPDPDAEVIALSPKTLMA TN) 和 motif 9 (GGGNDGLTRDF LGVG), 这些亲缘关系较近的蛋白中特异性存在或缺失的区段可

能行使着特殊且重要的生物学功能。

2.3 CoIDD4 理化性质及结构分析

拟南芥 *AtIDD4/5/6* 在拟南芥生长发育中扮演着重要的角色, 通过系统发育和 motif 分析, 在油茶中鉴定到 1 个 *AtIDD4/5/6* 的同源基因, 命名为 *CoIDD4*。CoIDD4 与 AtIDD4/5/6 存在高度相

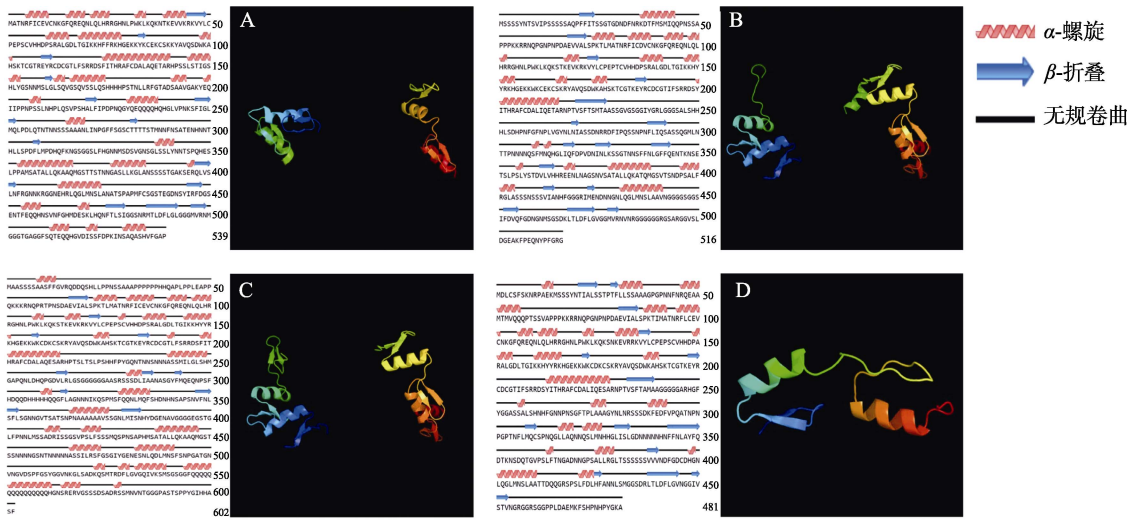
似的保守结构域（图 3，图 4）。通过 ORF finder 搜索发现 *CoIDD4* 的全长 ORF 框包含 1620 bp 碱基，编码 539 个氨基酸，与 *AtIDD4/5/6* 相似，该蛋白属于碱性不稳定亲水性蛋白质，且亚细胞定位均位于细胞核内（表 3）。

通过 Phyre2 软件分析 *AtIDD4/5/6* 和 *CoIDD4* 的二级结构，结果表明 *CoIDD4* 和 *AtIDD4/5/6* 拥有相似的二级结构，主要由无规则卷曲（random coil）、 α -螺旋（ α -helix）和 β -折叠（ β -strand）构成，无转角结构。Phyre2 预测的三级结构模型与二级结构相吻合（图 5）。

通过 Plant-mPLOC 软件分析显示 *CoIDD4* 和 *AtIDD4/5/6* 亚细胞定位均位于细胞核内。通过 NoD 分析核定位序列，结果表明，*CoIDD4* 的 N 端具有核定位信号（nuclear localization signal, NLS），其序列预测可信度最高的 2 条序列分别为 HRRGHNLPPWKLKQKNTKEVVKRKVYLC（可信度为 0.85）和 DLTGIKKHFFRKHGEKKYKCEKCSKKYAVQSDWK（可信度为 0.90）。而 *AtIDD4/5/6* 在靠近 N 端的一个核定位为 KKKR 或者 KRKR，但该核定位信号在 *CoIDD4* 中并未完全吻合（图 6）。

表 3 *CoIDD4* 和 *AtIDD4/5/6* 蛋白基本信息
Tab. 3 Basic information of *CoIDD4* and *AtIDD4/5/6*

基因 Gene	氨基酸数量 Number of amino acids	分子量 Mw/Da	等电点 pI	不稳定系数 Instability index	脂肪系数 Aliphatic index	平均亲水系数 GRAVY	亚细胞定位 Subcellular localization
<i>CoIDD4</i>	539	58 566.99	9.05	41.61	57.77	-0.663	Nucleus
<i>AtIDD4</i>	516	55 793.86	9.35	42.92	57.46	-0.653	Nucleus
<i>AtIDD5</i>	602	64 463.70	9.18	56.44	50.65	-0.798	Nucleus
<i>AtIDD6</i>	481	52 244.22	9.31	42.72	56.05	-0.682	Nucleus



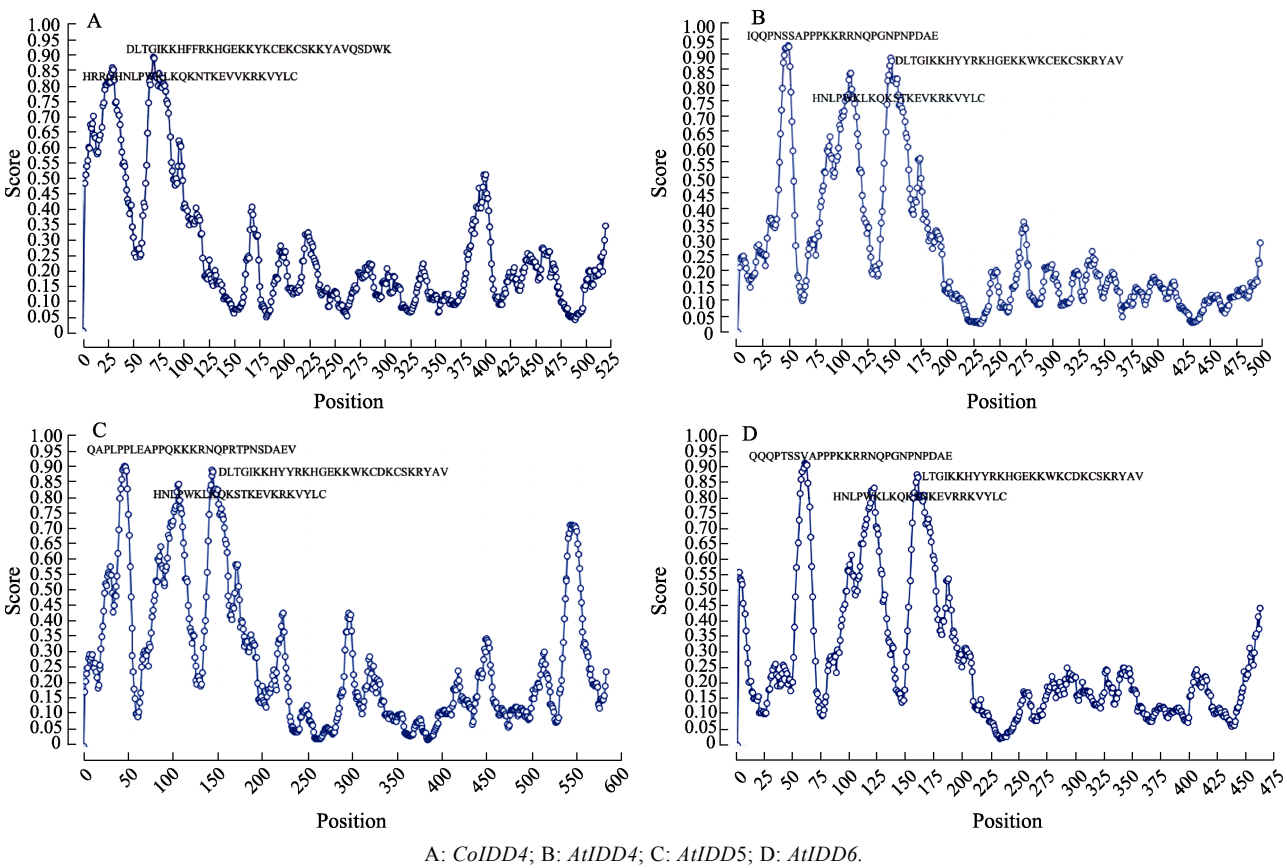
A: *CoIDD4*; B: *AtIDD4*; C: *AtIDD5*; D: *AtIDD6*.
图 5 *CoIDD4* 和 *AtIDD4/5/6* 的二级结构和三级结构预测

Fig. 5 Prediction of secondary structure and tertiary structure of *CoIDD4* and *AtIDD4/5/6*

2.4 *CoIDD4* 系统发育分析

CoIDD4 基因及其同源性较高的序列的系统发育分析结果显示（图 7），用于构建系统发育树的物种氨基酸序列共分为两大进化支，*CoIDD4* 与山茶（*Camellia sinensis*）*IDD5* 以 100 的自展支持率具为第一进化支内的一个小分支内，证实 *CoIDD4* 与 *CsIDD5* 具有高度的同源性。*CoIDD4* 与毛果杨（*Populus trichocarpa*）、可可树（*Theobroma ca-*

cao）、麻风树（*Jatropha curcas*）、橡胶树（*Hevea brasiliensis*）、木薯（*Manihot esculenta*）和荷花（*Nelumbo nucifera*）的 *IDD5* 基因共聚为第一进化支内，具有一定的亲缘关系。另外，芒果（*Mangifera indica*）、开心果（*Pistacia vera*）、蓖麻（*Ricinus communis*）和榴莲（*Durio zibethinus*）的 *IDD5* 基因与花生（*Arachis hypogaea*）的 *IDD4* 基因共同聚为第二进化支，与 *CoIDD4* 亲缘关系较远。



A: *CoIDD4*; B: *AtIDD4*; C: *AtIDD5*; D: *AtIDD6*.

图 6 *CoIDD4* 和 *AtIDD4/5/6* 的核定位信号预测

Fig. 6 Prediction of nuclear localization signal of *CoIDD4* and *AtIDD4/5/6*

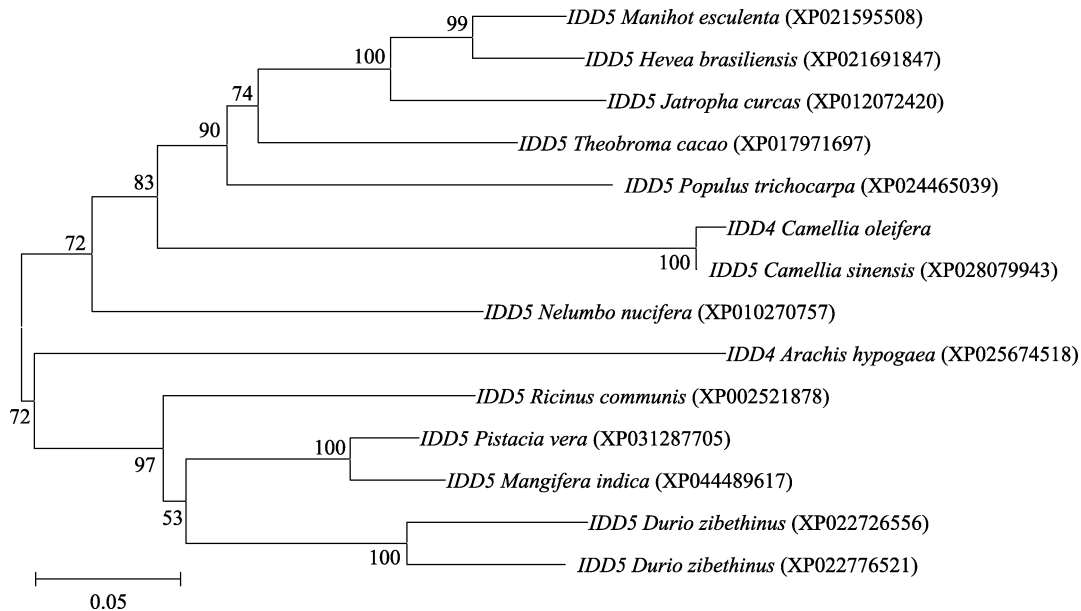


图 7 *CoIDD4* 及其近缘序列的系统发育树

Fig. 7 Phylogenetic tree of *CoIDD4* and related sequences

2.5 *CoIDD4* 基因启动子顺式作用元件分析

对 *CoIDD4* 起始密码子上游 2000 bp 的启动子区域序列进行顺式作用元件分析，结果发现，*CoIDD4* 启动子区包括 13 种类型的顺式作用元

件，其中 CAAT-box 和 TATA-box 两类核心顺式作用元件所占比例最大，其余顺式作用元件按功能可分为植物激素响应元件、光响应元件和防御响应元件几大类。其中光响应元件所占比例最大，

如 G-box、GT1-motif、TCCC-motif、GTGGC-motif 和 3-AF1 binding site 等。另外植物激素响应元件也占有较大的比例, 如参与脱落酸 (ABA) 响应的 ABRE 元件, 参与赤霉素 (GA) 响应的 TATC-box 元件, 参与水杨酸 (SA) 响应的 TCA-element 元件和参与生长素 (IAA) 响应的 TGA-element 元件 (图 8)。上述研究结果表明, *CoIDD4* 存在调节光信号通路和植物激素信号通路的潜能, 推测 *CoIDD4* 基因发挥着和 *AtIDD4* 相似的生物学功能, 即通过调控植物内源激素信号转导通路进而调控油茶的基础免疫反应。

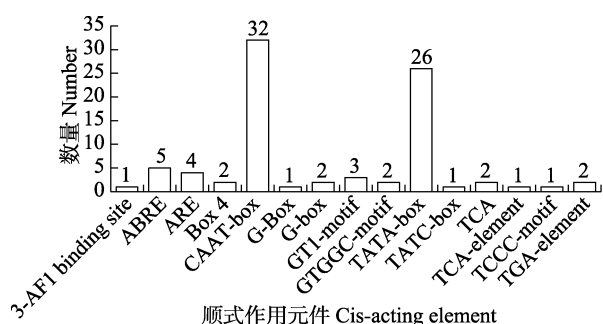


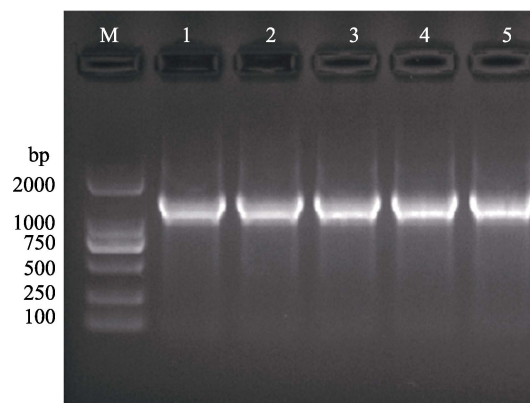
图 8 *CoIDD4* 基因启动子区顺式作用元件分析

Fig. 8 Analysis of cis-acting elements in promoter region of *CoIDD4* gene

2.6 *CoIDD4* 基因克隆及表达分析

通过 RT-PCR 分别从油茶的叶片组织中扩增到 *CoIDD4* 基因 ORF 框的全长编码序列, 经 TA 克隆测序后各获得一条 1620 bp 的特异性片段 (图 9)。将该序列在油茶基因组内进行 BLAST 分析, 结果显示, 该基因为油茶 CON1 个体 12 号染色体的一段基因 (CM039316.1)。在全长转录组中比对到一条 mRNA 链 (47 367 005~47 370 434 bp) (genemark-HiC_scaffold_12-processed-gene-473.55-mRNA-1), 包括 2 个外显子 (exon), 其中 exon1 为 47 370 076~473 704 34 bp, exon2 为 47 367 005~47 368 265 bp; 存在 2 个蛋白编码区 (CDS); 其中 CDS1 包括 47 370 076~47 370 434 bp, CDS2 包括 47 367 005~47 368 265 bp。

采用 RT-qPCR 分析 *CoIDD4* 基因的组织表达情况, 结果表明, 在内参基因表达量较一致的情况下, *CoIDD4* 基因在油茶的各个组织中均有表达, 且在叶片及花组织中表达量较高, 而在根、茎及果组织中表达量较微弱 (图 10)。推测 *CoIDD4* 基因在油茶叶片和花器官中发挥重要的作用。



M: DL2000 DNA marker; 1~5: 叶片组织。

M: DL2000 DNA marker; 1~5: Leaf tissue.

图 9 *CoIDD4* 基因扩增产物

Fig. 9 Amplified products of *CoIDD4* gene

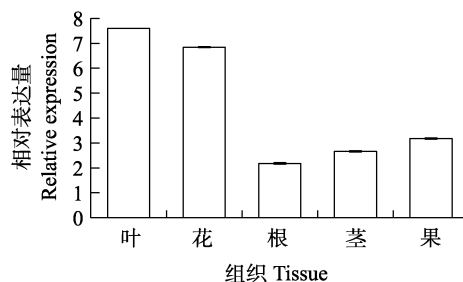
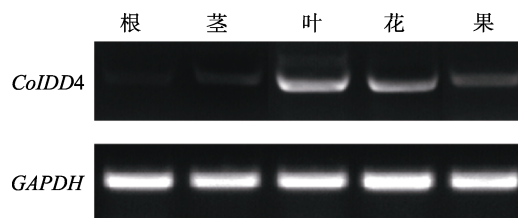


图 10 *CoIDD4* 基因在不同组织中的表达量

Fig. 10 Expression of *CoIDD4* gene in different tissues

3 讨论

本研究通过拟南芥 IDD 基因家族保守结构域, 在油茶全基因组范围内鉴定了 29 个油茶 IDD 基因家族成员, 同时通过 RT-PCR 克隆了与 *AtIDD4/5/6* 亲缘关系较近的 *CoIDD4* 基因, 分析其理化性质、结构和表达模式, 以期为油茶良种选育及抗性研究提供理论基础。

作为植物中 C2H2 家族的一个亚家族成员, IDD 基因家族已经在多个物种中被鉴定, 且启动子区富含与植物激素响应与逆境和抗性响应相关的顺式作用元件, 在调控植物生长发育和基础免疫反应中具有重要的作用^[6, 9, 15-17]。在油茶中还尚未见 IDD 基因家族的报道, 基于油茶全基因组数据和前人对 IDD 基因家族的研究结果, 本研究在

油茶中鉴定了 29 个 IDD 家族成员, 相对于模式植物拟南芥中 16 个 IDD 家族成员, 油茶 *IDD* 基因的数量出现了增加, 这可能与油茶长期人工栽培进行品种杂交和长期适应外界环境有关。对 *CoIDD* 家族成员的鉴定主要依据其高度保守的基因结构域, 通常认为 *AtIDD* 家族所包含保守的 *ID-domain* 区域, 含有一组 4 个串联的锌指结构 (ZF1-4)^[14], 在对拟南芥和油茶 *IDD* 家族成员保守结构域分析时发现, motif1-4 所指代的 ZF1-4 在 *AtIDD* 和 *CoIDD* 均保守存在, 这 4 个 motif 是参与 DNA-蛋白质或蛋白质-蛋白质互作所必需的模块, 能准确地识别靶标并进行转录调控。其他特殊的 motif 在不同的亚组中特异性存在或缺失是导致基因功能出现分化的原因, 这些特殊的基序可能在转录因子调控下游靶标基因的过程中发挥着重要的作用。

根据油茶和拟南芥 *IDD* 家族共建的系统发育树, 油茶的 29 个 *IDD* 基因家族成员大致分为 5 个亚组, 除前 2 个亚组中不包含拟南芥 *IDD* 家族基因, 其他每个亚族均含有拟南芥和油茶的 *IDD* 基因, 揭示了大多数 *CoIDD* 基因与拟南芥 *AtIDD* 基因存在一定的进化保守性。通常而言, 处于同一进化支的基因或蛋白, 若相互存在较高的相似性序列, 则可推测它们具有相似的生物学功能, 因此, 通过拟南芥和油茶 *IDD* 蛋白序列相似性预测和探讨油茶 *IDD* 蛋白的功能是基因功能研究的重要切入点。已有研究证明拟南芥 *AtIDD4/5/6* 可能与根系的发育和叶片极性有关, 并参与 DELLA 相互作用从而调控 GA 信号转导通路^[9], 另一方面, VÖLZ 等^[10]学者证实拟南芥 *idd4* 缺失突变对丁香假单胞菌 (*Pseudomonas syringae*) 抗性提高, 转录组及染色质免疫共沉淀 (chromatin immunoprecipitation assay, ChIP) 提示 *IDD4* 作为 SA 的负调控因子, 参与植物激素信号网络转导途径, 作为基础免疫应答和病原相关分子特征激活植物的免疫反应。本研究克隆并鉴定到 1 个与 *AtIDD4/5/6* 亲缘关系较近的 *CoIDD4* 基因, 该基因与 *AtIDD4/5/6* 构成一小进化亚支, 可能与 *AtIDD4/5/6* 具有相似的生物学功能。通过 qRT-PCR 发现 *CoIDD4* 基因主要在油茶的叶片组织和花器官中表达, 这与 *AtIDD4/5/6* 的表达模式存在一定的相似性^[7], 推测 *CoIDD4* 与 *AtIDD4* 功能相似, 通过介导内源植物激素信号转导通路参与油茶的基础免疫进程。

顺式作用元件是参与植物转录调控的重要分子开关, 处于通路的上游转录因子通过靶向作用于下游基因启动子区的顺式作用元件, 激活下游基因的表达, 从而发挥下游基因的作用。本研究预测了 *CoIDD4* 基因起始密码子上游 2 kb 的启动子区域的顺式作用元件, 一方面与光响应相关的作用元件反复出现多次, 显示了 *CoIDD4* 在光响应过程中具有重要的作用; 另一方面在该区域中还有一定数量的植物激素响应元件, 包括响应 GA、IAA、ABA 和 SA 的作用元件, 其中 SA 信号通路常被认为是植物免疫反应调节中重要的信号通路, 植物受病原体攻击后, SA 的合成加速, 并在植物体内迅速积累, 激活下游基因的表达进而引发植物的 SAR 反应^[17], *CoIDD4* 启动子区域中包括了参与 SA 响应的作用元件, 表明 *CoIDD4* 可能参与 SA 信号转导通路, 但 *CoIDD4* 是否参与 SA 积累或通过 SA 信号通路参与油茶免疫调控还有待进一步验证。

参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志[M]. 北京: 科学出版社, 1998, 49(3): 13.
Editorial Committee of Flora of China, Chinese Academy of Sciences. Flora of China[M]. Beijing: Science Press, 1998, 49(3): 13. (in Chinese)
- [2] HE X S, XU L C, PAN C, GONG C, WANG Y J, LIU X L, YU Y C. Drought resistance of *Camellia oleifera* under drought stress: changes in physiology and growth characteristic[J]. PLoS One, 2020, 15(7): e0235795.
- [3] 陈福, 郭晓春, 刘倬志, 张林涛. 云南省油茶产业高质量发展探析[J]. 现代农业科技, 2020(12): 252-255.
CHEN F, GUO X C, LIU Z Z, ZHANG L T. Analysis on high-quality development of *Camellia oleifera* industry in Yunnan province[J]. Modern Agricultural Science and Technology, 2020(12): 252-255. (in Chinese)
- [4] 郑静楠, 鹿杉, 郑进恒. 云南省油茶产业发展现状与对策研究[J]. 林业调查规划, 2021, 46(2): 114-117.
ZHENG J N, LU S, ZHENG J X. Development status and countermeasure of *Camellia oleifera* industry in Yunnan province[J]. Forest Inventory and Planning, 2021, 46(2): 114-117. (in Chinese)
- [5] YOSHIDA H, HIRANO K, SATO T, MITSUDA N, NOMOTO M, MAEO K, KOKETSU E, MITANI R, KAWAMURA M, ISHIGURO S, TADA Y, OHMETAKAGI M, MATSUOKA M, UEGUCHI-TANAKA M. DELLA protein functions as a transcriptional activator

- through the DNA binding of the indeterminate domain family proteins[J]. PNAS, 2014, 111(21): 7861-7866.
- [6] SINGLETON W R. Inheritance of indeterminate growth in maize[J]. Journal of Heredity, 1946, 37(2): 61-64.
- [7] 张喜贤. 拟南芥锌指蛋白 *AtIDD4* 在花丝伸长过程中的作用[D]. 昆明: 云南大学, 2014.
ZHANG X X. The role of *Arabidopsis* zinc finger protein *AtIDD4* in filament elongation[D]. Kunming: Yunnan University, 2014. (in Chinese)
- [8] 乔菊香. 拟南芥转录因子 *AtIDD5* 和 *AtIDD6* 在花丝伸长过程中的作用分析[D]. 昆明: 云南农业大学, 2019.
QIAO J X. Functional analysis of *Arabidopsis* transcription factor *AtIDD5* and *AtIDD6* in filament elongation[D]. Kunming: Yunnan Agricultural University, 2019. (in Chinese)
- [9] COELHO C P, HUANG P, LEE D Y, Brutnell T P. Making roots, shoots, and seeds: IDD gene family diversification in plants[J]. Trends in Plant Science, 2018, 23(1): 66-78.
- [10] VÖLZ R, KIM S K, MI J N, RAWAT A A, VELUCHAMY A, MARIAPPAN K G, RAYAPURAM N, DAVIERE J M, ACHARD P, BLILOU I, AL-BABILI S, BENHAMED M, HIRT H. Indeterminate-domain 4 (*IDD4*) coordinates immune responses with plant-growth in *Arabidopsis thaliana* [J]. PLoS Pathogens, 2019, 15(1): 1007499.
- [11] 陈骄羽. 毛竹 IDD 基因家族与笋芽发育的相关性研究[D]. 杭州: 浙江农林大学, 2020.
CHEN J Y. The correlation between IDD gene family and development of shoot/bud in *Phyllostachys edulis*[D]. Hangzhou: Zhejiang Agriculture and Forestry University, 2020. (in Chinese)
- [12] 段回春. 甘蓝型油菜 IDD 基因家族成员鉴定及 *Bna.A08IDD7a* 基因功能分析[D]. 重庆: 西南大学, 2020.
DUAN H C. Genome-wide identification of the INDETERMINATE DOMAIN (IDD) gene family in rapeseed and gene function analysis of *Bna. A08IDD7a*[D]. Chongqing: Southwest University, 2020. (in Chinese)
- [13] LIN P, WANG K L, WANG Y P, HU Z K, YAN C, HUANG H, MA X J, CAO Y Q, LONG W, LIU W X, LI X L, FAN Z Q, LI J Y, YE N, REN H D, YAO X H, YIN H F. The genome of oil-*Camellia* and population genomics analysis provide insights into seed oil domestication[J]. Genome Biology, 2022, 23(1): 14.
- [14] SEKALSKA B, CIECHANOWICZ A, DOLEGOWSKA B, NARUSZEWICZ M. Optimized RT-PCR method for assaying expression of monocyte chemotactic protein type 1 (*MCP-1*) in rabbit aorta[J]. Biochemical Genetics, 2006, 44(3/4): 133-143.
- [15] ENGLBRECHT C C, SCHOOF H, BÖHM S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome[J]. BMC Genomics, 2004, 5(1): 39.
- [16] 李云富, 王静贤, 杜艳芳, 邹华文, 张祖新. 玉米开花期相关的 Indeterminate domain (IDD)蛋白家族基因的鉴定[J]. 作物学报, 2019, 45(4): 499-507.
LI Y F, WANG J X, DU Y F, ZOU H W, ZHANG Z X. Identification of Indeterminate domain protein family genes associated with flowering time in maize[J]. Acta Agronomica Sinica, 2019, 45(4): 499-507. (in Chinese)
- [17] MOORE J W, LOAKE G J, SPOEL S H. Transcription dynamics in plant immunity[J]. Plant Cell, 2011, 23(8): 2809-2820.