

黄瓜 R2R3-MYB 亚家族鉴定及生物信息学分析

郭玉婷¹, 杜长霞^{1,2}

(1. 浙江农林大学 园艺科学学院 浙江省山区农业高效绿色生产协同创新中心, 浙江 杭州 311300; 2. 浙江农林大学 农业农村部亚热带果蔬质量安全控制重点实验室, 浙江 杭州 311300)

摘要: 【目的】深入研究黄瓜 *Cucumis sativus* R2R3-MYB 亚家族成员的相关功能。【方法】利用生物信息学手段分析黄瓜全基因组, 鉴定 R2R3-MYB 亚家族成员, 对其系统进化关系、蛋白理化性质、染色体定位、基因结构、保守基序、顺式作用元件、蛋白质互作进行分析。【结果】黄瓜全基因组中含 99 个具有典型结构域的 R2R3-MYB 转录因子, 蛋白序列含 195~552 个氨基酸, 有保守基序及氨基酸位点; 基因在染色体上分布不均匀; 大部分亚家族成员蛋白质的不稳定指数大于 40, 属于不稳定蛋白。顺式作用调控元件分析发现: 大部分基因启动子区所含元件与激素调节、MYB 结合位点、胁迫密切相关。【结论】通过黄瓜全基因组鉴定, 获得黄瓜基因组 99 个 R2R3-MYB 家族成员, 分为 30 个亚组, 映射于 7 条染色体上, 该家族成员的上游启动子区含逆境相关作用元件。图 7 表 1 参 36

关键词: 黄瓜; R2R3-MYB; 转录因子; 全基因组鉴定

中图分类号: S642.2

文献标志码: A

文章编号: 2095-0756(2024)02-0286-11

Identification and bioinformatics analysis of R2R3-MYB subfamily in cucumber

GUO Yuting¹, DU Changxia^{1,2}

(1. Collaborative Innovation Center for Efficient and Green Production of Agriculture in Mountainous Areas of Zhejiang Province, College of Horticulture Science, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China;

2. Key Laboratory of Quality and Safety Control for Subtropical Fruit and Vegetable, Ministry of Agriculture and Rural Affairs, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China)

Abstract: [Objective] This study is aimed to conduct an in-depth study of the functions associated with members of the R2R3-MYB subfamily in cucumber (*Cucumis sativus*). [Method] First, an analysis was conducted of the whole cucumber genome so as to identify R2R3-MYB subfamily members. Then, bioinformatics was employed to analyze their phylogenetic relationships, physicochemical properties of proteins, chromosomal localization, gene structure, conserved motifs, and cis-acting elements, as well as the protein interactions. [Result] The whole cucumber genome contained 99 R2R3-MYB transcription factors with typical structural domains, and the protein sequences contained amino acids ranging from 195 to 552, with conserved motifs and amino acid sites. They were unevenly distributed on chromosomes. They were unstable proteins with the instability index of most subfamily members being greater than 40. As was shown by the analysis of cis-acting regulatory elements, most of the elements contained in the promoter regions of genes were closely related to hormone regulation, MYB binding sites, and stress. [Conclusion] The R2R3-MYB family members were obtained through whole cucumber genome identification, which were divided into 30 subgroups

收稿日期: 2023-05-05; 修回日期: 2023-11-08

基金项目: 浙江省自然科学基金资助项目 (LY23C150005)

作者简介: 郭玉婷 (ORCID: 0009-0009-3497-8158), 从事蔬菜分子生物学研究。E-mail: 623163988@qq.com。通信作者: 杜长霞 (ORCID: 0000-0002-0366-2224), 副教授, 从事蔬菜逆境生物学研究。E-mail: changxiadu@zafu.edu.cn

and mapped on 7 chromosomes and the upstream promoter region of this family members contained stress-related action elements. [Ch, 7 fig. 1 tab. 36 ref.]

Key words: *Cucumis sativus*; R2R3-MYB; transcription factor; whole genome identification

MYB 蛋白质家族成员数量庞大, 功能多样, 存在于绝大多数真核生物中。大多数 MYB 蛋白作为转录因子发挥作用, MYB 结合域有结合 DNA 的能力^[1-2], 在植物生长发育、次生代谢及抗逆胁迫等方面具有重要作用^[3]。R2R3-MYB 作为家族中成员最多的一类, 对植物生长发育意义重大^[4-7]。

MYB 基因家族根据 MYB 结构域的数量和类型分为 4 类: 1R-MYB (或 MYB 相关)、2R-MYB (R2R3-MYB)、3R-MYB (R1R2R3-MYB) 和 4R-MYB^[8]。2R-MYB 类是植物特有的, 并且是最大的亚家族, 已在许多植物中进行了全基因组注释。R2R3-MYB 家族成员数量各不相同, 拟南芥 *Arabidopsis thaliana*^[9] 有 126 个, 水稻 *Oryza sativa*^[10] 有 102 个, 葡萄 *Vitis vinifera*^[11] 有 108 个, 甜橙 *Citrus sinensis*^[12] 有 100 个, 苹果 *Malus domestica*^[13] 有 222 个, 毛果杨 *Populus trichocarpa*^[14] 有 192 个, 玉米 *Zea mays*^[15] 有 200 多个。LI 等^[16] 在黄瓜 *Cucumis sativus* 中仅鉴定出 55 个 R2R3-MYB 基因, 而黄瓜基因组 V2 版本中鉴定出 69 个 R2R3-MYB 基因^[17]。通过最新版黄瓜基因组数据库可知: 黄瓜基因组 V3 版本基因组数量为 24 317 个, 较 V2 版本 (23 248 个) 增多。鉴于此, 本研究对黄瓜基因组数据库 V3 版本中的 R2R3-MYB 亚家族进行鉴定, 分析黄瓜 CsMYB 基因家族成员, 为研究黄瓜 CsMYB 家族功能提供参考。

1 材料与方法

1.1 数据来源

全基因组序列来源于黄瓜基因组数据库 (<http://www.cucurbitgenomics.org/organism/20>)。序列中的 MYB 结合域 (PF00249) 使用 PFam 数据库 (<http://pfam.xfam.org/>) 的隐马尔可夫模型 (HMM) 配置文件进行识别和确认, 该配置文件用作对先前注释的黄瓜 MYB 序列的查询, 通过 HMMER SEARCH 进一步识别并使用默认参数手动筛选 (目标序列的期望值 $E \leq 1.2 \times 10^{-8}$)^[17]。应用 HMMER SEARCH 建立黄瓜 MYB (CsMYB) 基因家族特异性模型^[18], 使用黄瓜 MYB 特异性 HMM 选择 E 低于 0.001 的所有黄瓜 MYB 蛋白。CsMYB 家族的候选基因根据黄瓜特异性 MYB HMM 确定。以上最初获得的所有候选 CsMYB 基因均通过保守域数据库 (conserved domain database, CDD)^[19] 和简单模块化架构研究工具 (simplemodular architecture research tool, SMART) 使用默认参数 $E \leq 0.010$ ^[20]。根据 MYB 序列内的特征性保守结构域, 分离黄瓜 R2R3-MYB 基因。

1.2 黄瓜 R2R3-MYB 基因家族系统进化分析

通过 ClustalW 对模式植物拟南芥、木本植物杨梅 *Morella rubra*^[21] 和草本植物黄瓜^[1] 的 R2R3-MYB 序列进行比对, 采用 MEGA-11 软件的邻接法 (neighbor-joining, NJ), 设置自展法系数 (bootstrap) 为 1 000 次, 其余参数按系统默认构建进化树^[22]。

1.3 黄瓜 R2R3-MYB 家族成员的基本性质及定位分析

通过筛选结果得到的黄瓜 R2R3-MYB 基因的序列信息, 利用 TBtools 将其定位于染色体上。采用 ExPASy-ProtParam (<https://www.expasy.org/resources/protparam>) 在线预测黄瓜 R2R3-MYB 蛋白的氨基酸数量、分子量、理论等电点、不稳定指数及脂肪系数; 使用 Plant-mPLoc 2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>) 预测亚细胞定位。

1.4 黄瓜 R2R3-MYB 家族成员的染色体定位及共线性分析

利用 TBtools 中的 Gene Location Visualize 功能绘制黄瓜 R2R3-MYB 家族成员的染色体定位图, 并通过 MCscanX 将黄瓜 R2R3-MYB 亚家族成员基因与拟南芥同源基因进行共线性分析。

1.5 黄瓜 R2R3-MYB 亚家族成员的基因结构及基序分析

利用 TBtools 对黄瓜 R2R3-MYB 亚家族成员的基因结构进行可视化分析, 并对其蛋白的保守基序进行分析, 限值为 15, 绘制结果图。

1.6 黄瓜 R2R3-MYB 亚家族上游的顺式元件预测

利用 TBtools 提取黄瓜 R2R3-MYB 亚家族成员上游 1 500 bp 的序列, 通过 PlantCARE 数据库

(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 获得启动子顺式作用元件，利用 Excel 筛选结果，再利用 TBtools 对其最终结果进行可视化。

1.7 黄瓜 R2R3-MYB 亚家族蛋白成员的互作预测

通过网站 STRING (<https://www.string-db.org/>) 对黄瓜 R2R3-MYB 蛋白成员之间进行互作预测分析，研究黄瓜 R2R3-MYB 蛋白成员之间的作用关系。

2 结果与分析

2.1 黄瓜 R2R3-MYB 亚家族全基因组鉴定及进化树构建

将黄瓜基因组数据与拟南芥进行比对，根据 HMM 模型进行搜索，取其交集，筛选鉴定其具有特殊结构域的序列，最终获得黄瓜 99 个 R2R3-MYB 基因家族成员。参照 DUBOS 等^[23] 的分类方法，将鉴定出的成员进一步分为 30 个亚组。为了探索系统发育关系，通过 ClustalW 进行序列比对，使用黄瓜 99 个 CsMYB 蛋白、拟南芥 125 个 AtMYB 蛋白和杨梅 122 个 MrMYB 蛋白构建了系统发育树 (图 1)。按照染色体定位结果，分别命名为 Cs2RMYB1~Cs2RMYB99 (表 1)。根据蛋白质序列的相似性和聚类结果的高自举性，将 CsMYB 进一步分为 30 个进化枝 (指定为 C1~C30)，黄瓜 R2R3-MYB 基因家族成员被分成 30 个亚组。其中，C16、C20 和 C23 亚组的家族成员数量最多，有 21 个成员；C18 亚组成员数为 16 个；C14、C24 和 C27 亚组家族成员数均为 15 个；C10 和 C28 亚组家族成员数量等同，为 4 个；C9 亚组成员数最少，仅 3 个。黄瓜 R2R3-MYB 亚家族的进化树聚类方式与拟南芥有一定的相似性。

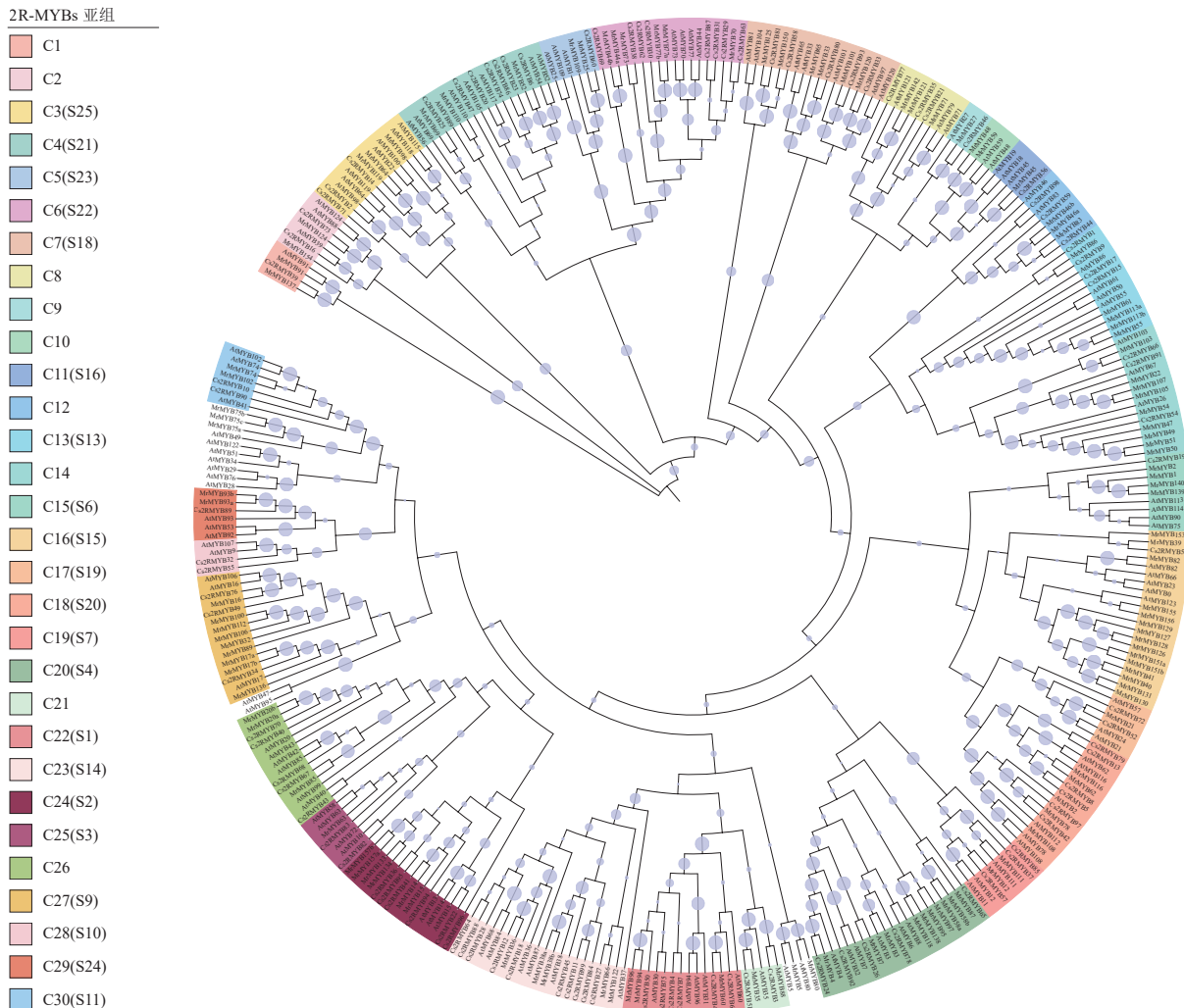


图 1 黄瓜与拟南芥、杨梅 R2R3-MYB 亚家族成员的系统进化树构建

Figure 1 Phylogenetic tree construction of *C. sativus*, *A. thaliana* and *M. rubra* R2R3-MYB subfamily members

表 1 黄瓜 R2R3-MYB 亚家族成员信息

Table 1 Cucumber R2R3-MYB subfamily member information

| 基因命名 | 氨基酸 数量/个 | 分子量/kDa | 等电点 | 不稳定 指数 | 脂肪指数 | 染色体 位置 | 基因命名 | 氨基酸 数量/个 | 分子量/ kDa | 等电点 | 不稳定 指数 | 脂肪指数 | 染色体 位置 |
|-----------|-------------|---------|-------|-----------|-------|-----------|-----------|-------------|-------------|------|-----------|-------|-------------|
| Cs2RMYB1 | 360 | 40.91 | 6.44 | 52.41 | 57.17 | Chr 1 | Cs2RMYB51 | 268 | 30.01 | 8.59 | 54.65 | 64.40 | Chr 3 |
| Cs2RMYB2 | 471 | 54.88 | 5.91 | 42.84 | 66.43 | Chr 1 | Cs2RMYB52 | 217 | 25.13 | 7.21 | 55.75 | 65.21 | Chr 3 |
| Cs2RMYB3 | 367 | 41.94 | 5.83 | 59.22 | 75.50 | Chr 1 | Cs2RMYB53 | 195 | 22.67 | 8.37 | 72.59 | 73.54 | Chr 3 |
| Cs2RMYB4 | 331 | 37.60 | 9.02 | 65.15 | 61.90 | Chr 1 | Cs2RMYB54 | 355 | 39.21 | 6.25 | 54.46 | 82.96 | Chr 3 |
| Cs2RMYB5 | 277 | 31.19 | 5.90 | 78.88 | 66.21 | Chr 1 | Cs2RMYB55 | 305 | 34.19 | 6.19 | 45.79 | 74.52 | Chr 3 |
| Cs2RMYB6 | 232 | 26.59 | 8.83 | 57.26 | 66.51 | Chr 1 | Cs2RMYB56 | 253 | 28.52 | 9.10 | 49.09 | 65.53 | Chr 3 |
| Cs2RMYB7 | 326 | 36.76 | 6.42 | 53.91 | 67.94 | Chr 1 | Cs2RMYB57 | 313 | 35.05 | 4.97 | 68.24 | 75.97 | Chr 4 |
| Cs2RMYB8 | 288 | 32.50 | 6.31 | 71.30 | 64.03 | Chr 1 | Cs2RMYB58 | 314 | 36.06 | 6.10 | 62.98 | 58.06 | Chr 4 |
| Cs2RMYB9 | 377 | 42.09 | 6.46 | 50.23 | 66.71 | Chr 1 | Cs2RMYB59 | 322 | 35.68 | 6.14 | 54.46 | 67.55 | Chr 4 |
| Cs2RMYB10 | 337 | 37.92 | 5.74 | 45.24 | 71.19 | Chr 1 | Cs2RMYB60 | 377 | 40.50 | 5.41 | 39.90 | 73.24 | Chr 4 |
| Cs2RMYB11 | 299 | 33.99 | 9.17 | 58.17 | 61.64 | Chr 1 | Cs2RMYB61 | 353 | 39.81 | 6.22 | 60.28 | 59.92 | Chr 4 |
| Cs2RMYB12 | 372 | 42.33 | 6.12 | 54.32 | 67.63 | Chr 1 | Cs2RMYB62 | 229 | 25.19 | 9.29 | 58.90 | 61.70 | Chr 4 |
| Cs2RMYB13 | 277 | 32.11 | 5.83 | 63.46 | 79.96 | Chr 1 | Cs2RMYB63 | 263 | 29.14 | 8.34 | 61.14 | 67.15 | Chr 4 |
| Cs2RMYB14 | 337 | 38.24 | 8.66 | 52.38 | 59.08 | Chr 1 | Cs2RMYB64 | 280 | 31.51 | 8.92 | 59.53 | 72.11 | Chr 4 |
| Cs2RMYB15 | 370 | 41.15 | 6.13 | 51.97 | 61.95 | Chr 1 | Cs2RMYB65 | 260 | 28.84 | 5.00 | 38.86 | 76.54 | Chr 5 |
| Cs2RMYB16 | 364 | 41.75 | 9.54 | 55.58 | 69.81 | Chr 1 | Cs2RMYB66 | 323 | 36.41 | 6.00 | 53.24 | 72.45 | Chr 5 |
| Cs2RMYB17 | 327 | 37.06 | 7.68 | 62.31 | 69.48 | Chr 2 | Cs2RMYB67 | 258 | 29.46 | 5.84 | 45.28 | 67.67 | Chr 5 |
| Cs2RMYB18 | 295 | 32.99 | 9.07 | 45.41 | 70.10 | Chr 2 | Cs2RMYB68 | 236 | 26.98 | 6.71 | 53.48 | 74.83 | Chr 5 |
| Cs2RMYB19 | 222 | 25.74 | 6.00 | 53.86 | 72.43 | Chr 2 | Cs2RMYB69 | 284 | 31.23 | 6.41 | 56.18 | 72.46 | Chr 5 |
| Cs2RMYB20 | 264 | 29.98 | 8.97 | 58.24 | 63.14 | Chr 2 | Cs2RMYB70 | 300 | 33.77 | 5.29 | 55.23 | 71.50 | Chr 5 |
| Cs2RMYB21 | 265 | 31.42 | 9.64 | 54.52 | 67.02 | Chr 2 | Cs2RMYB71 | 337 | 38.25 | 7.69 | 55.10 | 66.59 | Chr 5 |
| Cs2RMYB22 | 286 | 32.59 | 5.60 | 58.72 | 59.69 | Chr 2 | Cs2RMYB72 | 210 | 24.43 | 5.19 | 51.25 | 60.38 | Chr 5 |
| Cs2RMYB23 | 228 | 26.87 | 10.27 | 58.08 | 65.44 | Chr 2 | Cs2RMYB73 | 464 | 52.29 | 6.82 | 66.09 | 66.85 | Chr 5 |
| Cs2RMYB24 | 256 | 29.15 | 8.24 | 48.25 | 79.22 | Chr 2 | Cs2RMYB74 | 369 | 41.40 | 6.61 | 50.68 | 56.34 | Chr 5 |
| Cs2RMYB25 | 210 | 23.54 | 10.31 | 65.32 | 59.48 | Chr 2 | Cs2RMYB75 | 307 | 34.25 | 7.51 | 48.64 | 64.59 | Chr 5 |
| Cs2RMYB26 | 230 | 26.32 | 8.83 | 57.64 | 63.65 | Chr 2 | Cs2RMYB76 | 398 | 44.52 | 6.49 | 53.56 | 62.29 | Chr 5 |
| Cs2RMYB27 | 204 | 23.04 | 8.44 | 43.78 | 67.94 | Chr 2 | Cs2RMYB77 | 202 | 23.93 | 8.25 | 43.61 | 58.37 | Chr 5 |
| Cs2RMYB28 | 338 | 38.46 | 8.56 | 58.00 | 68.99 | Chr 2 | Cs2RMYB78 | 294 | 33.13 | 8.72 | 52.07 | 70.41 | Chr 5 |
| Cs2RMYB29 | 290 | 31.80 | 9.23 | 59.33 | 65.93 | Chr 2 | Cs2RMYB79 | 286 | 32.77 | 6.40 | 64.03 | 65.17 | Chr 6 |
| Cs2RMYB30 | 297 | 33.70 | 9.97 | 55.65 | 69.97 | Chr 2 | Cs2RMYB80 | 552 | 61.13 | 5.50 | 52.51 | 61.54 | Chr 6 |
| Cs2RMYB31 | 301 | 32.30 | 8.19 | 55.90 | 64.15 | Chr 2 | Cs2RMYB81 | 301 | 34.40 | 5.56 | 70.97 | 65.08 | Chr 6 |
| Cs2RMYB32 | 308 | 34.86 | 5.93 | 48.47 | 84.90 | Chr 2 | Cs2RMYB82 | 269 | 31.33 | 6.45 | 45.95 | 63.42 | Chr 6 |
| Cs2RMYB33 | 519 | 57.03 | 7.46 | 59.70 | 57.53 | Chr 3 | Cs2RMYB83 | 334 | 38.24 | 4.96 | 46.95 | 59.85 | Chr 6 |
| Cs2RMYB34 | 305 | 34.64 | 5.70 | 57.58 | 70.72 | Chr 3 | Cs2RMYB84 | 286 | 32.92 | 5.80 | 56.50 | 54.58 | Chr 6 |
| Cs2RMYB35 | 248 | 29.14 | 9.10 | 56.61 | 66.49 | Chr 3 | Cs2RMYB85 | 291 | 33.53 | 6.63 | 61.70 | 46.94 | Chr 6 |
| Cs2RMYB36 | 259 | 29.98 | 6.50 | 48.74 | 59.54 | Chr 3 | Cs2RMYB86 | 277 | 32.26 | 9.35 | 51.45 | 59.49 | Chr 6 |
| Cs2RMYB37 | 296 | 33.70 | 5.35 | 56.27 | 67.20 | Chr 3 | Cs2RMYB87 | 301 | 32.45 | 9.32 | 50.91 | 64.82 | Chr 6 |
| Cs2RMYB38 | 219 | 25.25 | 6.32 | 59.79 | 70.32 | Chr 3 | Cs2RMYB88 | 319 | 35.85 | 6.53 | 48.51 | 65.49 | Chr 6 |
| Cs2RMYB39 | 354 | 40.50 | 9.20 | 58.22 | 77.40 | Chr 3 | Cs2RMYB89 | 344 | 38.44 | 6.12 | 57.72 | 80.00 | Chr 6 |
| Cs2RMYB40 | 267 | 30.10 | 5.59 | 55.70 | 69.33 | Chr 3 | Cs2RMYB90 | 387 | 44.36 | 5.81 | 54.93 | 61.01 | Chr 6 |
| Cs2RMYB41 | 245 | 28.13 | 5.75 | 48.70 | 62.57 | Chr 3 | Cs2RMYB91 | 289 | 33.19 | 6.45 | 53.15 | 70.17 | Chr 6 |
| Cs2RMYB42 | 224 | 25.60 | 7.64 | 65.43 | 74.87 | Chr 3 | Cs2RMYB92 | 248 | 27.74 | 9.11 | 46.24 | 76.73 | Chr 7 |
| Cs2RMYB43 | 275 | 31.61 | 8.71 | 39.67 | 79.75 | Chr 3 | Cs2RMYB93 | 500 | 54.62 | 5.89 | 55.46 | 57.82 | Chr 7 |
| Cs2RMYB44 | 316 | 36.22 | 5.63 | 58.87 | 56.52 | Chr 3 | Cs2RMYB94 | 254 | 29.45 | 5.36 | 74.13 | 75.63 | Chr 7 |
| Cs2RMYB45 | 312 | 34.71 | 6.99 | 55.70 | 62.79 | Chr 3 | Cs2RMYB95 | 271 | 31.90 | 5.56 | 53.17 | 62.66 | Chr 7 |
| Cs2RMYB46 | 248 | 29.70 | 8.46 | 52.73 | 69.60 | Chr 3 | Cs2RMYB96 | 254 | 29.40 | 5.02 | 43.86 | 79.49 | Chr 7 |
| Cs2RMYB47 | 348 | 39.13 | 8.86 | 46.82 | 64.77 | Chr 3 | Cs2RMYB97 | 280 | 32.36 | 5.20 | 59.21 | 68.86 | Chr 7 |
| Cs2RMYB48 | 294 | 32.66 | 6.17 | 57.40 | 64.42 | Chr 3 | Cs2RMYB98 | 285 | 32.82 | 5.32 | 58.47 | 64.70 | Chr 7 |
| Cs2RMYB49 | 233 | 26.13 | 9.33 | 51.33 | 65.41 | Chr 3 | Cs2RMYB99 | 248 | 27.78 | 6.24 | 44.80 | 65.28 | Scaffold 72 |
| Cs2RMYB50 | 313 | 35.13 | 5.95 | 54.63 | 65.50 | Chr 3 | | | | | | | |

说明: Cs2RMYB48基因定位于叶绿体, 其他基因定位于细胞核。

2.2 黄瓜 R2R3-MYB 亚家族的理化性质分析

表 1 表明：黄瓜 R2R3-MYB 亚家族成员编码 195~552 个氨基酸，分子量为 22.67~61.13 kDa，理论等电点为 4.96~10.31，59 个 R2R3-MYB 蛋白的等电点小于 7，40 个 R2R3-MYB 蛋白的等电点大于 7，超过半数的蛋白为酸性。除 Cs2RMYB43、Cs2RMYB60、Cs2RMYB65 外，其余 CsR2R3-MYB 亚家族成员蛋白质不稳定指数为 42.84~78.88，大于 40.00，为不稳定蛋白。所有成员的脂肪系数相对较高，说明黄瓜 R2R3-MYB 亚家族成员适应环境的多样性。Plant-mPLoc 2.0 网站预测结果显示：除 Cs2RMYB48 定位于叶绿体，其他成员均定位于细胞核。

2.3 黄瓜 R2R3-MYB 亚家族染色体定位及共线性分析

黄瓜 R2R3-MYB 亚家族成员映射于 1~7 号染色体上，Cs2RMYB99 的染色体定位结果未知 (图 2)。每条染色体上的基因分布相对不均匀，1、4、5、7 号染色体上基因大多分布在上下两端，而剩下的染色体上基因分布较为均匀。3 号染色体上数量最多，可以定位到 24 个 R2R3-MYB 亚家族成员；1、2 号染色体各有 16 个 R2R3-MYB 亚家族成员；5、6 号染色体分别为 14 和 13 个 R2R3-MYB 亚家族成员；7 号染色体上最少，为 7 个 R2R3-MYB 亚家族成员。此外，2 号染色体上的 Cs2RMYB27 和 Cs2RMYB28、3 号染色体上的 Cs2RMYB45 和 Cs2RMYB46、5 号染色体上的 Cs2RMYB71 和 Cs2RMYB72、7 号染色体上的 Cs2RMYB94 和 Cs2RMYB95 均在各自染色体上形成基因簇。根据进化树分析可知：形成基因簇的成员之间同源性较高，推测成员间功能相对保守。

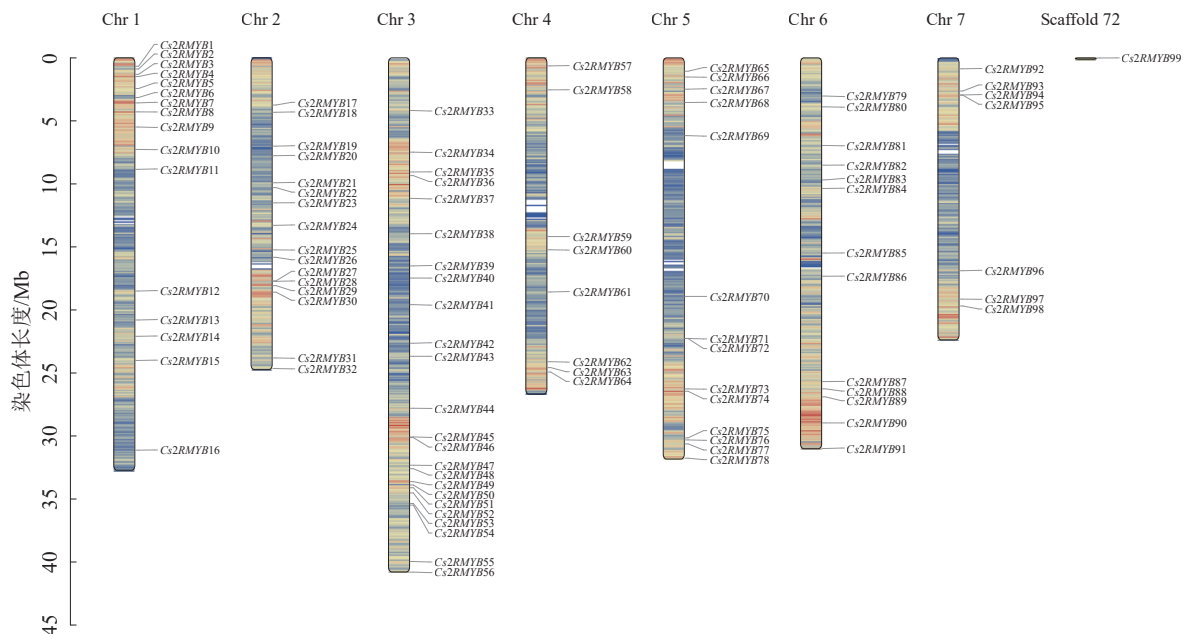


图 2 黄瓜 R2R3-MYB 亚家族染色体定位图

Figure 2 Chromosome location map of *C. sativus* R2R3-MYB subfamily

为揭示黄瓜 R2R3-MYB 亚家族成员在整个黄瓜基因组中的连锁关系，对黄瓜与拟南芥 MYB 基因进行共线性分析 (图 3)，结果表明：25 对基因表现出共线性和保守的连锁关系，即表明 2 个物种之间 MYB 基因家族存在高同源性。

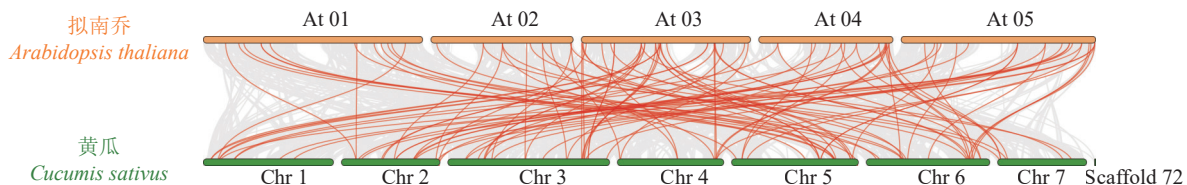


图 3 黄瓜 R2R3-MYB 亚家族与拟南芥共线性分析图

Figure 3 Analysis of the collinearity between *C. sativus* R2R3-MYB subfamily and *A. thaliana*

2.4 黄瓜 R2R3-MYB 亚家族基因结构及保守基序分析

利用 TBtools 对黄瓜 R2R3-MYB 亚家族进行保守基序分析 (图 4)。结果显示：在 99 个 R2R3-

MYB 蛋白中鉴定出 10 个不同的基序，其中 Motif 3、Motif 7、Motif 2 和 Motif 4 是高度保守的基序，所有成员均含有 Motif 3，导致成员间功能相似。表明这些黄瓜 R2R3-MYB 亚家族成员间拥有特有或相同的基序，导致了它们之间具有特定的或者相似的功能。

通过 Jalview 对黄瓜 R2R3-MYB 亚家族进行序列比对 (图 5)。结果显示：2R 重复序列中有 3 个色氨酸

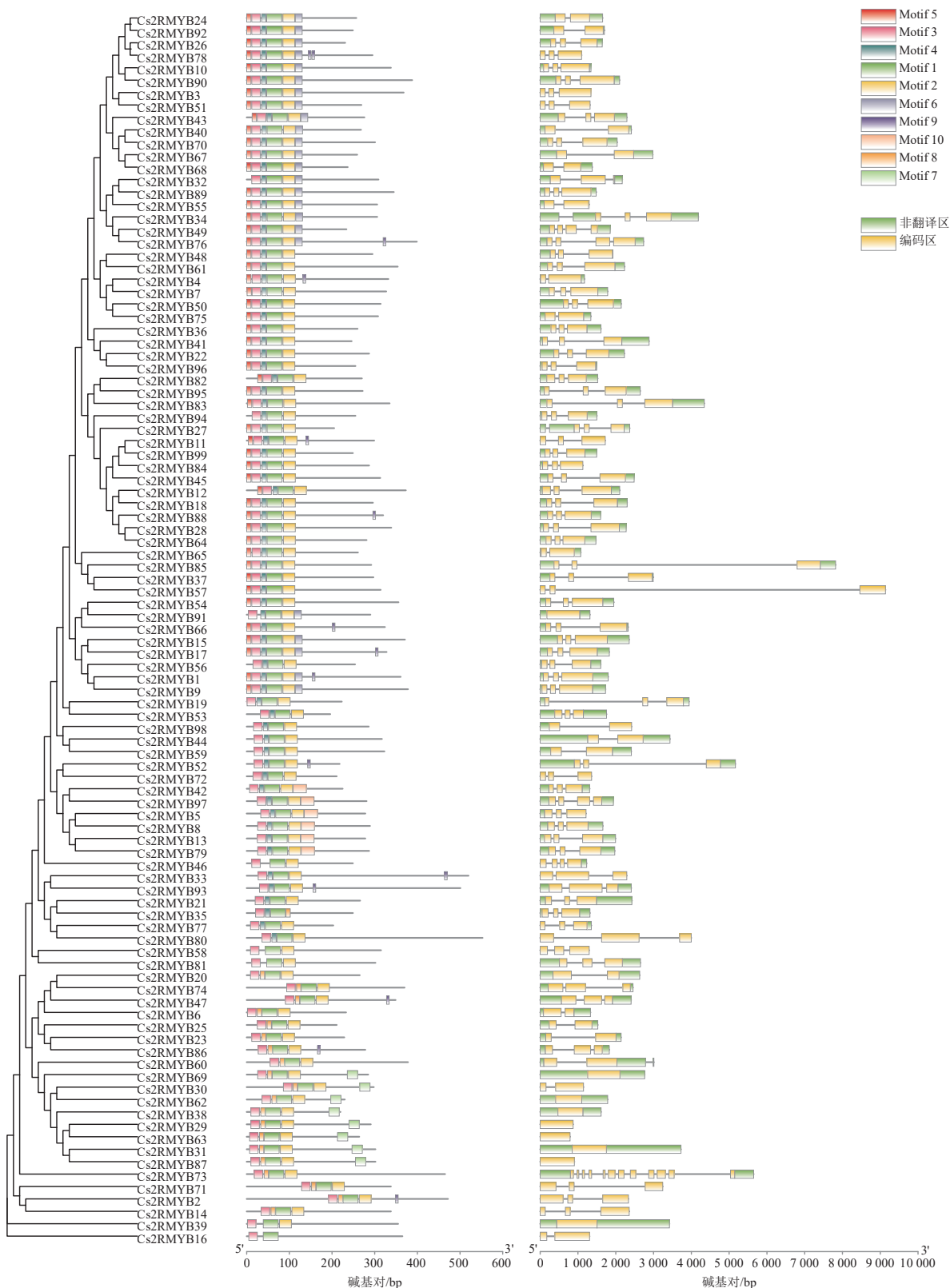


图 4 黄瓜 R2R3-MYB 亚家族蛋白保守基序和基因结构

Figure 4 Protein conservative motifs and gene structure of R2R3-MYB subfamily in *C. sativus*

酸残基(W)及3R重复序列中有3个色氨酸残基,形成疏水结构,此为鉴定MYB家族典型的结构域特征。除保守的W外,2R重复序列的末端还存在12个高度保守的氨基酸残基,如精氨酸(R)、甘氨酸(G)、赖氨酸(K)、丝氨酸(S)、半胱氨酸(C)、亮氨酸(L)等,3R重复序列中的谷氨酸(E)、G、丙氨酸(A)、R、苏氨酸(T)等。



图5 黄瓜 R2R3-MYB 亚家族蛋白序列比对图
 Figure 5 Sequence comparison of *C. sativus* R2R3-MYB subfamily proteins

2.5 黄瓜 R2R3-MYB 亚家族顺式作用元件分析

为了进一步研究黄瓜 R2R3-MYB 亚家族成员启动子区的顺式作用元件, 通过 PlantCARE 在线网站分析了该亚家族上游 1 500 bp 区域 (图 6)。其启动子元件多与激素、抗逆及厌氧诱导相关, 大部分黄瓜 R2R3-MYB 亚家族成员含雌激素响应 (ERE) 作用元件。从元件类型来看, ERE 元件最多, 分布于绝大多数成员的启动子区域上。从单个元件的具体分布来看, 约 68% 家族成员启动子区含 ERE 元件, 约 60% 成员启动子区含脱落酸应答元件 (ABRE)。从单个基因家族成员来看, *Cs2RMYB92* 启动子只有 TCA-element 激素调控相关元件; *Cs2RMYB98* 有 MYB 结合位点, 并富含激素响应与厌氧诱导元件, 例如 ABRE、ERE、茉莉酸甲酯相关元件 (CGTCA) 和 MYB 结合位点元件 (MBS)。 *Cs2RMYB5*、*Cs2RMYB7* 和 *Cs2RMYB8* 等基因上有 ABRE 元件, 说明这些基因可能与脱落酸响应有关; *Cs2RMYB2*、*Cs2RMYB6* 和 *Cs2RMYB9* 等基因上含 ERE 元件, 说明这些基因可能响应乙烯的调控; *Cs2RMYB1*、*Cs2RMYB16* 和 *Cs2RMYB25* 等基因上有 CGTCA 元件, 说明这些基因可能与茉莉酸甲酯响应有关; *Cs2RMYB12*、*Cs2RMYB25*、*Cs2RMYB33* 和 *Cs2RMYB34* 等基因启动子元件上都含 MBS 元件, 说明这些基因可能响应干旱胁迫的调控。



图 6 黄瓜 R2R3-MYB 亚家族上游顺式元件预测

Figure 6 Prediction of upstream homeopathic elements of *C. sativus* R2R3-MYB subfamily

2.6 黄瓜 R2R3-MYB 亚家族成员蛋白的互作预测

为深入了解黄瓜 R2R3-MYB 亚家族成员蛋白的互作情况, 对 99 个成员进行蛋白质互作预测 (图 7), 结果表明: *Cs2RMYB25*、*Cs2RMYB3*、*Cs2RMYB84*、*Cs2RMYB98*、*Cs2RMYB6*、*Cs2RMYB6* 均与 *Cs2RMYB44* 存在互作关系; *Cs2RMYB96*、*Cs2RMYB22*、*Cs2RMYB31*、*Cs2RMYB41*、*Cs2RMYB56*、*Cs2RMYB75*、*Cs2RMYB7*、*Cs2RMYB4*、*Cs2RMYB50* 均与 *Cs2RMYB90* 存在互作关系, 且有较强的关联性。此外, 部分成员间虽存在互作相关性, 但各成员之间的具体机制并不明晰, 仍需进一步深入研究。

3 讨论

本研究共鉴定出 99 个黄瓜 R2R3-MYB 亚家族成员, 约占黄瓜总 MYB 数的 60%, 数量多于西瓜 (89 个)^[4], 少于拟南芥 (126 个)^[9]、水稻 (102 个)^[10] 和玉米 (200 个)^[15]。黄瓜的基因组大小为 224.8~251.1 Mb^[24], 而拟南芥、水稻、玉米和西瓜的基因组大小分别为 125.0、430.0、2 300.0 和 365.1 Mb^[25-28], R2R3-MYB 亚家族成员数量与植物的基因组大小并没有直接的相关性。

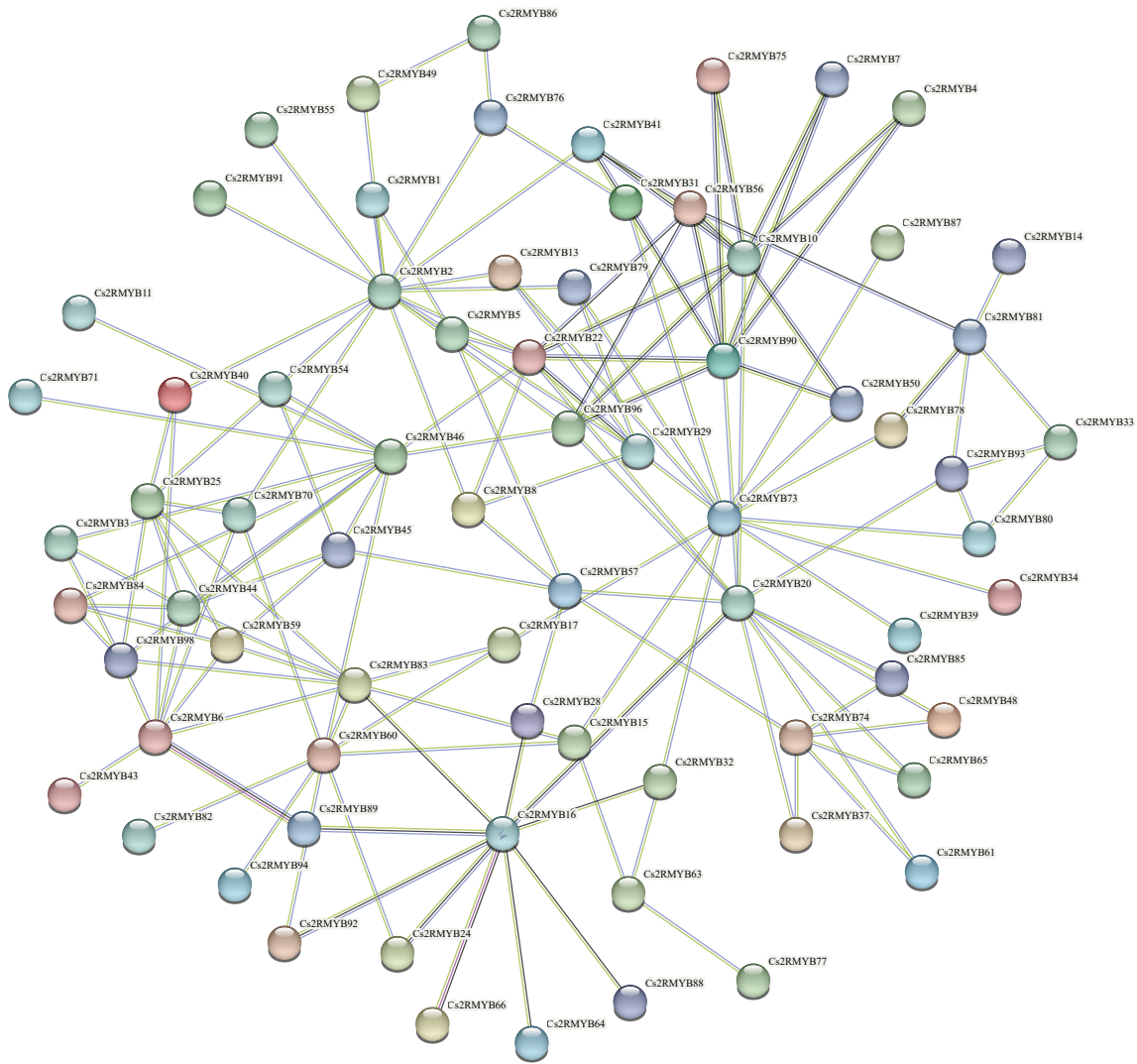


图7 黄瓜 R2R3-MYB 亚家族成员蛋白的互作预测

Figure 7 Prediction of protein interactions among R2R3-MYB subfamily members in *C. sativus*

作为 MYB 家族最大的亚家族，R2R3-MYB 家族成员具有典型的模式结构，在植物发育、新陈代谢、信号转导及响应生物和非生物胁迫中起着非常重要的作用^[29]。为了研究黄瓜基因组进化过程中功能保守的 R2R3-MYB 基因功能，构建其系统进化树并进行分析，结果表明：黄瓜 *Cs2RMYB* 30 个亚组中大多数包含来自拟南芥和杨梅 R2R3-MYB 成员，然而，C28 亚组仅拥有黄瓜和拟南芥 R2R3-MYB 成员，表明它们可能来自共同的祖先。同时，外显子和内含子的结构分析有利于研究基因家族内部的进化关系，通过分组可以初步推测其功能，聚类较近的基因存在功能相似性，如在 C30 亚组中，拟南芥 *AtMYB42* 和 *AtMYB85* 被证明参与木质素生物合成代谢调控^[30]，黄瓜 *Cs2RMYB72* 和 *Cs2RMYB73* 与其表现出较近的亲缘关系，可能与木质素生物合成代谢有关；在 C15 亚组中，*AtMYB75*、*AtMYB90*、*AtMYB113* 和 *AtMYB114* 过量表达会促进植物花青素的积累^[31]，与黄瓜 *Cs2RMYB19* 同属于 C15 亚组；在 C22 亚组中，*AtMYB93* 是其分支上唯一一个影响拟南芥侧根发育的负调控因子^[32]，而黄瓜 *Cs2RMYB89* 与 *AtMYB93* 进化关系最近；宫思宇等^[33] 研究表明：*AtMYB94* 通过调节拟南芥表皮的蜡质含量提高植株耐旱性，而 C22 亚组中的 *Cs2RMYB48* 是其同源基因，可能具有类似功能，同一分支亚家族转录因子可能存在功能相似性^[23, 34]。

基序分析结果发现：所有的 *CsR2R3-MYB* 基因都有共同的保守基序，但也有独特的基序，这为 *CsR2R3-MYB* 亚家族成员的分类提供参考，且可能与 *CsR2R3-MYB* 基因的功能分化相关。此外，启动子序列上的顺式作用元件是转录因子识别并发挥转录调控作用的重要识别位点^[35]。本研究发现：黄瓜 *CsR2R3-MYB* 亚家族成员启动子上的顺式作用元件种类繁多，绝大多数 *CsR2R3-MYB* 基因家族成员基

因启动子区都含有 ERE 元件, 暗示其可能参与黄瓜性别分化等生长发育过程; 富含 ABRE 和 CGTCA 等激素响应顺式作用元件, 表明 CsR2R3-MYB 基因家族在黄瓜应答逆境胁迫中发挥重要作用。由于 CsR2R3-MYB 基因调控元件的不同功能, 启动子区域中这些高度多样化的顺式调控元件也可能反映了转录水平上的功能分歧^[36]。本研究初步预测了 MYB 家族 CsR2R3-MYB 基因与植物生长发育及逆境胁迫响应相关, 可为后续该亚家族成员基因功能深入研究提供参考。

4 参考文献

- [1] LIU Yi, ZHANG Zhishuo, FANG Ke, *et al.* Genome-wide analysis of the MYB-related transcription factor family in pepper and functional studies of *CaMYB37* involvement in capsaicin biosynthesis [J/OL]. *International Journal of Molecular Sciences*, 2022, **23**(19): 11667[2023-05-01]. doi: 10.3390/ijms231911667.
- [2] LI Jinlu, HAN Guoliang, SUN Cuifeng, *et al.* Research advances of MYB transcription factors in plant stress resistance and breeding [J/OL]. *Plant Signaling & Behavior*, 2019, **14**(8): e1613131[2023-05-01]. doi: 10.1080/15592324.2019.1613131.
- [3] STRACKE R, HOLTGRÄWE D, SCHNEIDER J, *et al.* Genome-wide identification and characterization of R2R3-MYB genes in sugar beet (*Beta vulgaris*) [J/OL]. *BMC Plant Biology*, 2014, **14**: 249[2023-05-01]. doi:10.1186/s12870-014-0249-8.
- [4] WANG Jin, LIU Yue, CHEN Xueliang, *et al.* Characterization and divergence analysis of duplicated R2R3-MYB genes in watermelon [J]. *Journal of the American Society for Horticultural Science*, 2020, **145**(5): 281 – 288.
- [5] LI Sijia, HUANG Hu, MA Xianjin, *et al.* Characterizations of MYB transcription factors in *Camellia oleifera* reveal the key regulators involved in oil biosynthesis [J]. *Horticulturae*, 2022, **8**(8): 742 – 761.
- [6] ZHAO Kai, CHENG Zihan, GUO Qing, *et al.* Characterization of the poplar R2R3-MYB gene family and over-expression of *PsnMYB108* confers salt tolerance in transgenic tobacco [J/OL]. *Frontiers in Plant Science*, 2020, **11**: 571881[2023-05-01]. doi: 10.3389/fpls.2020.571881.
- [7] SABIR I A, MANZOOR M A, SHAH I H, *et al.* MYB transcription factor family in sweet cherry (*Prunus avium* L.): genome-wide investigation, evolution, structure, characterization and expression patterns [J/OL]. *BMC Plant Biology*, 2022, **22**: 2[2023-05-01]. doi: 10.1186/s12870-021-03374-y.
- [8] 邱文怡, 王诗雨, 李晓芳, 等. MYB转录因子参与植物非生物胁迫响应与植物激素应答的研究进展[J]. *浙江农业学报*, 2020, **32**(7): 1317 – 1328.
QIU Wenyi, WANG Shiyu, LI Xiaofang, *et al.* Progress of MYB transcription factors involved in plant abiotic stress response and plant hormone response [J]. *Zhejiang Journal of Agriculture*, 2020, **32**(7): 1317 – 1328.
- [9] STRACKE R, WERBER M, WEISSHAAR B. The R2R3-MYB gene family in *Arabidopsis thaliana* [J]. *Current Opinion in Plant Biology*, 2001, **4**(5): 447 – 456.
- [10] CHEN Yanhui, YANG Xiaoyuan, HE Kun, *et al.* The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family [J]. *Plant Molecular Biology*, 2006, **60**(1): 107 – 124.
- [11] MATUS J T, AQUEA F, ARCE-JOHNSON P. Analysis of the grape MYB-R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *vitis* and *Arabidopsis* genomes [J/OL]. *BMC Plant Biology*, 2008, **8**: 83[2023-05-01]. doi: 10.1186/1471-2229-8-83.
- [12] LIU Chaoyang, WANG Xia, XU Yuantao, *et al.* Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (*Citrus sinensis*) [J]. *Molecular Biology Reports*, 2014, **41**(10): 6769 – 6785.
- [13] CAO Zhonghui, ZHANG Shizhong, WANG Rongkai, *et al.* Genome wide analysis of the apple MYB transcription factor family allows the identification of *MdoMYB121* gene conferring abiotic stress tolerance in plants [J/OL]. *PLoS One*, 2013, **8**(7): e69955[2023-05-01]. doi:10.1371/journal.pone.0069955.
- [14] WILKINS O, NAHAL H, FOONG J, *et al.* Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors [J]. *Plant Physiology*, 2009, **149**(2): 981 – 993.
- [15] DIAS A P, BRAUN E L, MCMULLEN M D, *et al.* Recently duplicated maize R2R3-MYB genes provide evidence for distinct mechanisms of evolutionary divergence after duplication [J]. *Plant Physiology*, 2003, **131**(2): 610 – 620.
- [16] LI Qiang, ZHANG Cunjia, LI Jing, *et al.* Genome-wide identification and characterization of R2R3-MYB family in *Cucumis sativus* [J]. *PLoS One*, 2012, **7**(10): 1 – 18.
- [17] CHENG Chunyan, LI Qingrong, WANG Xing, *et al.* Identification and expression analysis of the *CsMYB* gene family in

- root knot nematode-resistant and susceptible cucumbers [J/OL]. *Frontiers in Genetics*, 2020, **11**: 550677[2023-05-01]. doi: 10.3389/FGENE.2020.550677.
- [18] LOZANO R, HAMBLIN M T, PROCHNIK S, *et al.* Identification and distribution of the NBS-LRR gene family in the Cassava genome [J]. *BMC Genomics*, 2015, **16**(1): 360 – 374.
- [19] MARCHLER-BAUER A, BO Y, HAN L Y, *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures [J]. *Nucleic Acids Research*, 2017, **45**(D1): D200 – D203.
- [20] MARCHLER-BAUER A, DERBYSHIRE M K, GONZALES N R, *et al.* CDD: NCBI's conserved domain database [J]. *Nucleic Acids Research*, 2015, **43**(D1): D222 – D226.
- [21] CAO Yunlin, JIA Huimin, XING Mengyun, *et al.* Genome-wide analysis of MYB gene family in Chinese bayberry (*Morella rubra*) and identification of members regulating flavonoid biosynthesis [J/OL]. *Frontiers in Plant Science*, 2021, **12**: 691384[2023-05-01]. doi: 10.3389/fpls.2021.691384.
- [22] KUMAR S, STECHER G, TAMURA K. MEGA 7: molecular evolutionary genetics analysis version 7.0 for bigger datasets [J]. *Molecular Biology and Evolution*, 2016, **33**(7): 1870 – 1874.
- [23] DUBOS C, STRACKE R, GROTEWOLD E, *et al.* MYB transcription factors in *Arabidopsis* [J]. *Trends in Plant Science*, 2010, **15**(10): 573 – 581.
- [24] LI Hongbo, WANG Shenhao, CHAI Sen, *et al.* Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber [J]. *Nature Communications*, 2022, **13**(1): 682 – 696.
- [25] KAUL S, KOO H L, JENKINS J, *et al.* Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. *Nature*, 2000, **408**(6814): 796 – 815.
- [26] BURR B. Mapping and sequencing the rice genome [J]. *The Plant Cell*, 2002, **14**(3): 521 – 523.
- [27] SCHNABLE P S, WARE D, FULTON R S, *et al.* The B73 maize genome: complexity, diversity, and dynamics [J]. *Science*, 2009, **326**(5956): 1112 – 1115.
- [28] GUO Shaogui, ZHAO Shengjie, SUN Honghe, *et al.* Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits [J]. *Nature Genetics*, 2019, **51**(11): 1616 – 1623.
- [29] BALDONI E, GENGA A, COMINELLI E. Plant MYB transcription factors: their role in drought response mechanisms [J]. *International Journal of Molecular Sciences*, 2015, **16**(7): 15811 – 15851.
- [30] ZHONG Ruiqin, LEE Chanhui, ZHOU Jianli, *et al.* A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis* [J]. *The Plant Cell*, 2008, **20**(10): 2763 – 2782.
- [31] FENG Shouqian, XU Yongchao, YANG Long, *et al.* Genome-wide identification and characterization of R2R3-MYB transcription factors in pear [J]. *Scientia Horticulturae*, 2015, **197**: 176 – 182.
- [32] GIBBS D J, VOB U, HARDING S A, *et al.* *AtMYB93* is a novel negative regulator of lateral root development in *Arabidopsis* [J]. *New Phytologist*, 2014, **203**(4): 1194 – 1207.
- [33] 宫思宇, 陈海燕, 郭冬雪, 等. 黄瓜抗旱相关基因CsMYB94的克隆与表达分析[J]. 华北农学报, 2020, **35**(增刊1): 18 – 23.
GONG Siyu, CHEN Haiyan, GUO Dongxue, *et al.* Cloning and expression analysis of drought-related gene CsMYB94 in cucumber [J]. *Journal of North China Agriculture*, 2020, **35**(suppl 1): 18 – 23.
- [34] HUANG Wenjun, SUN Wei, LÜ Haiyan, *et al.* Isolation and molecular characterization of thirteen R2R3-MYB transcription factors from *Epimedium sagittatum* [J]. *International Journal of Molecular Sciences*, 2013, **14**(1): 594 – 610.
- [35] LI Junren, CHEN Xiuzhen, ZHOU Xuanxuan, *et al.* Identification of trihelix transcription factors in *Pogostemon cablin* reveals PatGT-1 negatively regulates patchoulol biosynthesis [J/OL]. *Industrial Crops and Products*, 2020, **161**: 113182[2023-05-01]. doi:10.1016/j.indcrop.2020.113182.
- [36] YANG Jiahan, ZHANG Binghui, GU Gang, *et al.* Genome-wide identification and expression analysis of the R2R3-MYB gene family in tobacco (*Nicotiana tabacum* L.) [J]. *BMC Genomics*, 2022, **23**(1): 432 – 453.