



葡萄GH5基因家族全基因组鉴定及生物信息学分析

黎霖¹, 聂鸿飞², 叶文秀³, 赵旗峰^{4,*}, 纪薇^{1,2,3,*}

¹山西农业大学园艺学院, 山西晋中030801

²运城市县农业农村局, 山西运城044500

³北京大学现代农业研究院, 潍坊市葡萄种质创新与利用重点实验室, 潍坊现代农业山东省实验室, 山东潍坊261000

⁴山西农业大学果树研究所, 山西晋中030801

*共同通信作者: 纪薇(jiweiputao@163.com)、赵旗峰(gssqfzhao@163.com)

摘要: 纤维素酶是一种能将纤维素分解成寡糖或单糖的蛋白质。在糖苷水解酶家族中, *GH5*基因家族是植物当中功能最多样化且数量最多的植物纤维素酶。本研究鉴定到分布在葡萄5条染色体上的18个*GH5*家族基因。进化树分析显示葡萄的*GH5*家族基因分为3个亚家族, 同在一个进化分类的成员其基因结构相似。通过基因结构分析发现葡萄*GH5*家族成员含1~6个内含子, 内含子与外显子数量差异较小。亚细胞定位分析发现葡萄*GH5*家族成员主要分布在细胞核。蛋白结构分析发现葡萄*GH5*家族成员以 α -螺旋和无规则卷曲为主。顺式作用元件分析发现葡萄*VvGH5*基因上游启动子序列包含大量的激素响应元件, 暗示这些基因家族成员在响应植物内源激素调控中有重要作用。本研究将为深入研究葡萄*GH5*家族成员的基因功能和分子育种工作提供理论基础。

关键词: 葡萄; *GH5*基因家族; 生物信息学

Whole genome identification and bioinformatics analysis of grape *GH5* gene family

LI Lin¹, NIE Hongfei², YE Wenxiu³, ZHAO Qifeng^{4,*}, JI Wei^{1,2,3,*}

¹College of Horticulture, Shanxi Agricultural University, Jinzhong, Shanxi 030801, China

²Yongji County Agricultural and Rural Bureau, Yuncheng, Shanxi 044500, China

³Institute of Modern Agriculture, Peking University, Weifang Modern Agriculture Shandong Province Laboratory, Weifang Key Laboratory of Grapes Germplasm Innovation and Utilization, Weifang, Shandong 261000, China

⁴Institute of Fruit Tree Research, Shanxi Agricultural University, Jinzhong, Shanxi 030801, China

*Co-corresponding authors: Ji W (jiweiputao@163.com), Zhao QF (gssqfzhao@163.com)

Abstract: Cellulase is a kind of protein that can decompose cellulose into oligosaccharides or monosaccharides. Among the glycoside hydrolase families, *GH5* gene family is the plant cellulase with the most diverse functions and the largest number in plants. In this study, 18 *GH5* family genes distributed on 5 chromosomes of grape were identified. Phylogenetic tree analysis showed that the *GH5* family genes of grapes were divided into three subfamilies, which were similar to the members of the same evolutionary classifi-

收稿 2022-02-26 修定 2023-04-08

资助 山西省科技成果转化引导专项项目(202204021301028)、山西省高等学校优秀成果(科学技术)培育项目(2020KJ037)、山西省研究生教育创新计划项目(2022Y335)和山西省果树产业技术体系(2023CYJSTX07-01)。

cation. Through gene structure analysis, it was found that grape GH5 family members contain 1–6 introns, and there is little difference in the number of introns and exons. Subcellular localization analysis showed that grape GH5 family members were mainly distributed in the nucleus. Protein structure analysis showed that grape GH5 family members α -spiral and irregular curl are dominant. Through homeopathic element analysis, it was found that the upstream promoter sequence of grape *VvGH5* gene contained a large number of hormone response elements, it is suggested that these gene family members play an important role in response to plant endogenous hormone regulation. This study will provide a theoretical basis for in-depth study of gene function and molecular breeding of grape GH5 family members.

Key words: grape (*Vitis vinifera*); GH5 gene family; bioinformatics

植物细胞壁是保护封闭细胞的刚性结构,植物在细胞和细胞壁的形状上表现出丰富的多样性,植物细胞壁有少量的结构蛋白(约10%),主要由碳水化合物和酚类化合物组成(Höfte等2017)。细胞壁合成和降解的研究近年来取得了突破性进展,植物纤维素酶的作用机制已基本明确(孟亚轩等2021)。纤维素酶是一种能将纤维素分解成寡糖或单糖的蛋白质(何芳芳等2020)。其分布在至少17个糖苷水解酶家族(glycoside hydrolase, GH)中,而GH5基因家族拥有功能最多样化且数量最多的植物纤维素酶(Chen等2021; 曾佳佳等2021; 袁红梅等2016)。葡萄(*Vitis vinifera*)是一种重要的经济果树(Khan等2021)。2020年葡萄基因组学数据库Vitis-GDB的建立为葡萄基因组学的研究提供了便捷的生物信息学工具(Dong等2020)。而基于生物信息学分析的葡萄GH5基因家族研究却鲜有报道。本研究结合前人的方法,对葡萄GH5基因家族进行挖掘及鉴定,分析其家族成员的结构和功能,以期为进一步深入研究葡萄GH5基因家族提供理论参考。

1 材料与方法

1.1 葡萄GH5基因家族鉴定与染色体定位

葡萄(*Vitis vinifera* L.)基因组测序数据及其注释信息均从VitisGDB (<http://vitisgdb.ynau.edu.cn/>)数据库下载(Dong等2020)。使用UniProt (<https://www.uniprot.org/uploadlists/>)下载到水稻(*Oryza sativa* L.)的GH5基因家族序列作为本地BLAST的参考序列(Alistair等2020; Wintersinger等2015)。使用TBtools (v1.098)软件进行本地BLAST比对,得到葡萄GH5基因家族候选序列ID,在TBtools (v1.089)程序中

绘制葡萄GH5基因在染色体上的分布图(Chen等2020)。

1.2 葡萄GH5基因家族信号肽预测分析

使用在线网站SignalP 5.0 (<https://services.healthtech.dtu.dk>)对葡萄GH5基因家族蛋白进行信号肽预测(Almagro等2019)。

1.3 葡萄GH5基因家族基因结构、系统发育树及保守基序分析

为探究葡萄GH5基因家族的系统发育关系,使用MEGA 11.0软件对获取的葡萄GH5蛋白序列进行比对,选用ML法、泊松校正、Bootstrap=100等参数绘制系统发育树(Tamura等2013; Larkin等2007),并在FigTree v1.4.3中进行美化。运用MEME (<http://memesuite.org/>)在线工具分析葡萄GH5蛋白的氨基酸序列的保守基序,基序个数设置为10个(Bailey等1994)。然后用TBtools (v1.089)软件将预测结果进行可视化;运用TBtools (v1.089)程序将葡萄基因组序列和水稻的GH5基因家族序列分析得到葡萄GH5的DNA序列FASTA格式。利用TBtools (v1.089)程序绘制基因编码区以及非编码区的分布图。并在TBtools (v1.089)程序中将三者进行联合可视化(Chen等2020)。

1.4 葡萄VvGH5基因家族蛋白亚细胞定位和理化性质分析

使用MEME (<http://memesuite.org/>)在线软件分析VvGH5蛋白序列(Bailey 1994)。使用ExPASy提供的蛋白质基本性质在线分析工具ProtParam (<https://web.expasy.org/protparam/>),分析葡萄GH5基因编码的蛋白质基本性质,包括氨基酸残基的个数、相对分子量、等电点和亲水性等(Wang等2012)。利用

在线PSORT (<http://psort.hgc.jp/>)工具进行葡萄GH5基因家族蛋白质亚细胞定位预测(Paul等2007)。

1.5 葡萄GH5基因家族蛋白结构及跨膜结构域预测分析

使用在线工具SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)进行葡萄GH5蛋白序列二级结构的构建(Li等2021; Han等2020)。利用Swiss-Modle (<https://swissmodel.expasy.org/>)在线工具获取三维结构模型,完成蛋白的同源建模(Florian等2009; Chen等2006)。使用TMHMM-2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>)在线工具对葡萄GH5基因家族蛋白进行跨膜结构域预测(Chen等2003; Krogh等2001)。

1.6 葡萄GH5基因家族基因顺式作用元件分析

利用TBtools (v1.089)对葡萄GH5基因家族进行基因顺式作用元件预测,将得到的葡萄GH5基因家族基因组序列,使用gff3序列提取工具提取CDS上游2 000 bp的参数。将得到的CDS序列导入在线网站PlantCare,使用TBtools (v1.089)将结果可视化(Chen等2020)。并使用Adobe Photoshop 20进行美化。

2 实验结果

2.1 葡萄VvGH5基因家族鉴定与染色体定位

通过TBtools (v1.098)软件将分别从VitisGDB与

UniProt 数据库获得的葡萄基因组数据和水稻GH5基因家族序列进行BLAST比对,筛选鉴定共得到18个葡萄VvGH5,依次将它们命名为VvGH5.1~VvGH5.18 (图1)。这18个葡萄VvGH5基因不均匀分布于5条染色体上,其中在4号染色体上分布最多,共有7个,其中有6条串联在一起;在2号和3号两条染色体上分别分布有4个,均串联在一起;1号和5号染色体分别分布2个(图2)。

2.2 葡萄VvGH5基因家族信号肽预测分析

经葡萄VvGH5基因家族信号肽预测分析可知,仅有葡萄VvGH5.6、VvGH5.9、VvGH5.10、VvGH5.11、VvGH5.12、VvGH5.13、VvGH5.18基因具有信号肽,且葡萄VvGH5.10基因剪切位点位于第30和31号氨基酸STA—VG之间;VvGH5.6、VvGH5.11、VvGH5.12基因剪切位点位于第23和24号氨基酸CEA—RV之间;VvGH5.13基因的剪切位点位于第23和24号氨基酸CQA—RV之间;VvGH5.9基因剪切位点位于第25和26号氨基酸VEA—AD之间;VvGH5.18基因剪切位点位于第25和26号氨基酸VEA—GD之间(Rosa等2020) (图3)。

2.3 葡萄VvGH5基因家族系统发育树、基因结构及保守基序分析

根据葡萄GH5基因结构间的相似性进行系统发育分析,发现18个葡萄VvGH5基因可分为三类,同一进化分类的成员基因结构相似(图4)。第I类包含

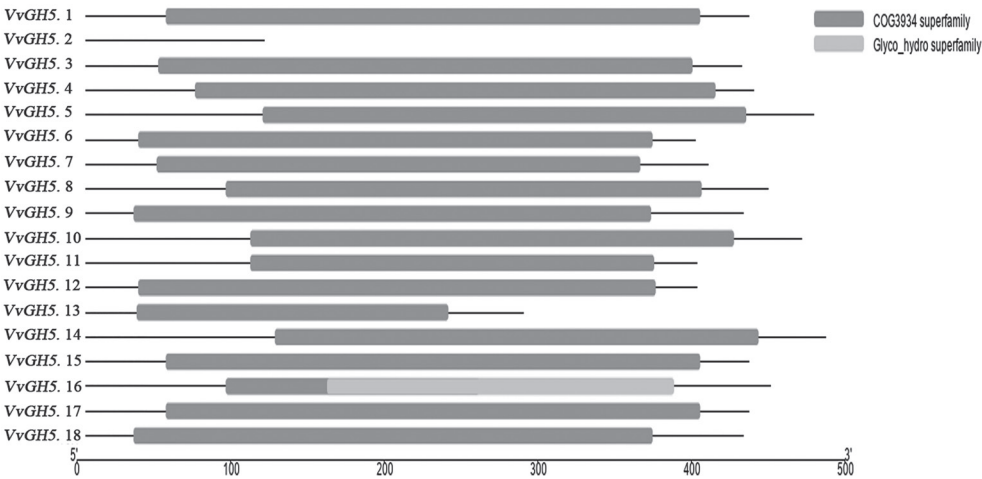


图1 葡萄VvGH5基因家族保守结构域
Fig. 1 Conserved domain of grape VvGH5 gene family

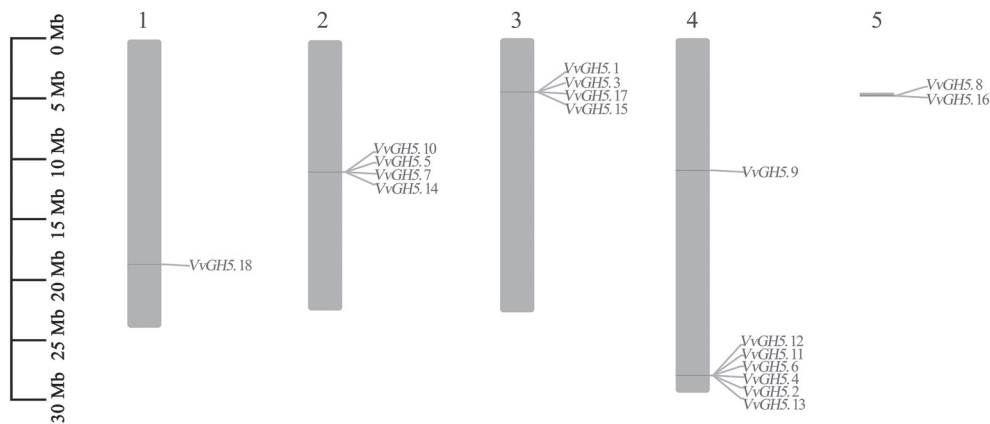


图2 葡萄VvGH5基因家族染色体定位
Fig. 2 Chromosomal location of grape VvGH5 gene family

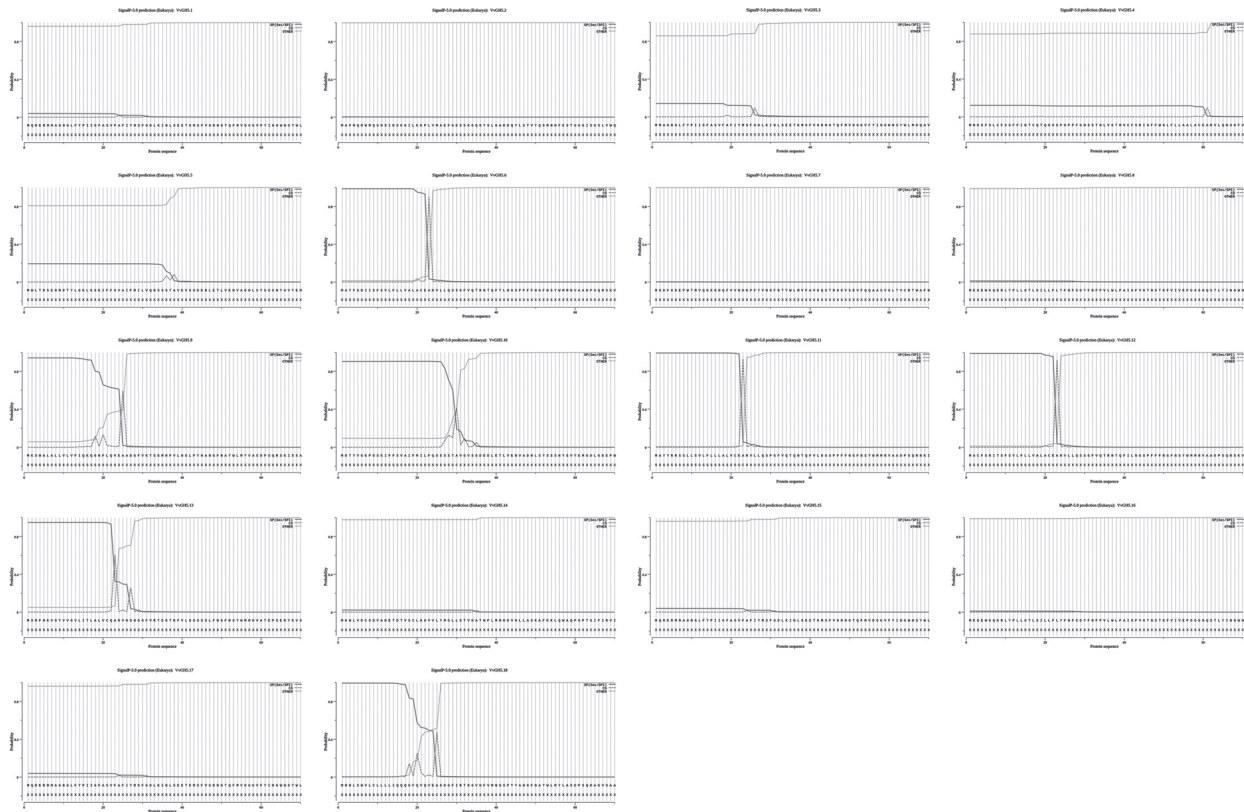


图3 葡萄VvGH5基因家族信号肽预测
Fig. 3 Prediction of signal peptides of grape VvGH5 gene family

大部分成员, 共10个; 又可细分为3小组, 分别包括4、2、4个葡萄 *VvGH5* 基因; 第Ⅱ类由5个葡萄 *VvGH5* 基因组成; 第Ⅲ类由3个葡萄 *VvGH5* 基因组成。对葡萄 *VvGH5* 基因家族的编码序列和基因序

列信息进行分析, 结果表明葡萄 *VvGH5* 基因家族的编码序列外显子数量在1~6个之间。此外, 除 *VvGH5.2* 基因以外, 其他基因均含有内含子结构。这体现出葡萄 *VvGH5* 基因家族基因结构简单的特征,

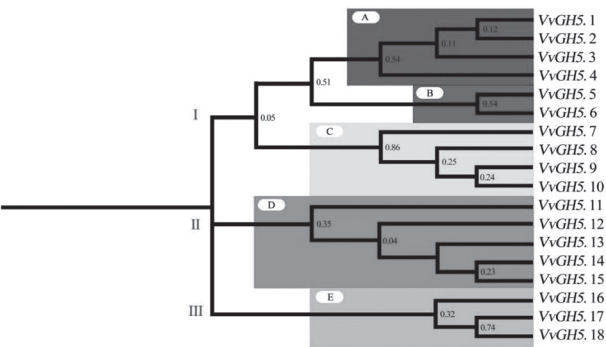


图4 葡萄*VvGH5*基因家族系统发育树
Fig. 4 Phylogenetic tree of grape *VvGH5* gene family

图中不同模块表示18个葡萄*VvGH5*基因进化发育分析不同的分组, 其中, A: 第I类第1小组; B: 第I类第2小组; C: 第I类第3小组; D: 第II类; E: 第III类。

也说明该家族具有一定的保守性(图5)。将基因结构、系统发育树及保守基序三者联合分析可知, 葡萄*VvGH5*基因家族蛋白成员具有1~10个保守基序, 其中Motif 1、Motif 3、Motif 6、Motif 7、Motif 9、Motif 10是*VvGH5.16*、*VvGH5.17*、*VvGH5.18* 3个葡萄*VvGH5*基因成员所共有的。18个葡萄*VvGH5*基因家族中, 除*VvGH5.11*基因外, 其余基因均含有2个以上保守基序。除*VvGH5.16*基因中不含有Motif 5和Motif 8, 其余17个葡萄*VvGH5*基因成员都含有Motif 5和Motif 8, 占比达到94.4%; 说明Motif 5

和Motif 8可能为葡萄*VvGH5*基因家族最保守的基序(图6)。

2.4 葡萄*GH5*基因家族蛋白理化性质分析

*VvGH5*蛋白序列的分析结果表明, 不同*VvGH5*基因家族的序列数目与分布不同(图7和8)。葡萄*VvGH5*基因家族蛋白质亚细胞定位结果显示, 葡萄*VvGH5*基因家族编码蛋白质主要分布在细胞核中, 其次为液泡, 此外, 在高尔基体、叶绿体、细胞质、质膜、细胞外基质、内质网、过氧化物酶体、线粒体以及细胞骨架等亚细胞结构中也有分布(Chou等2010) (表1)。由葡萄*VvGH5*基因蛋白质基本特性分析可知, 葡萄*VvGH5*基因蛋白质氨基酸数目在121~479之间, 其中*VvGH5.5*蛋白质氨基酸数量最多, *VvGH5.2*蛋白氨基酸数量最少; 蛋白质的相对分子质量在13 259.14~55 428.93 kDa之间, 等电点在5.15~9.83之间, 原子数在1 839~7 697之间; 热稳定系数在31.08~40.56之间, 其中*VvGH5.8*蛋白和*VvGH5.16*蛋白不稳定, 寿命较短, 其余多数为寿命较长的蛋白。脂肪系数在32.12~82.94之间; 经分析预测, 葡萄*GH5*基因家族编码蛋白质的亲水指数在-0.46~-0.206之间, 表明葡萄*VvGH5*基因家族蛋白均属于亲水性蛋白(Wang等2012) (表2)。

2.5 葡萄*VvGH5*基因家族蛋白二级结构预测、三维模型建立及跨膜结构域分析

蛋白质二级结构预测结果表明(表3), 葡萄*Vv-*

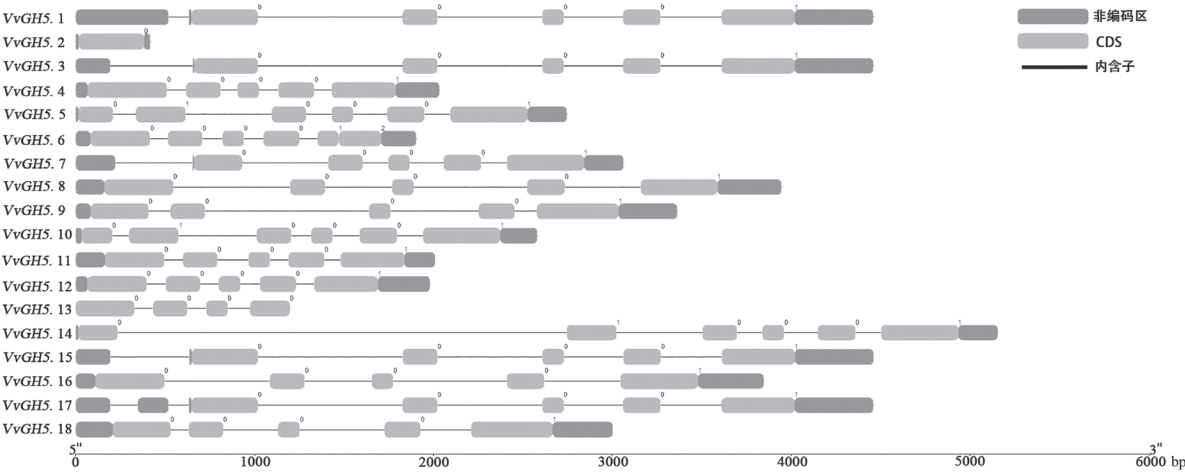


图5 葡萄*VvGH5*基因家族基因结构分析
Fig. 5 Gene structure analysis of grape *VvGH5* gene family

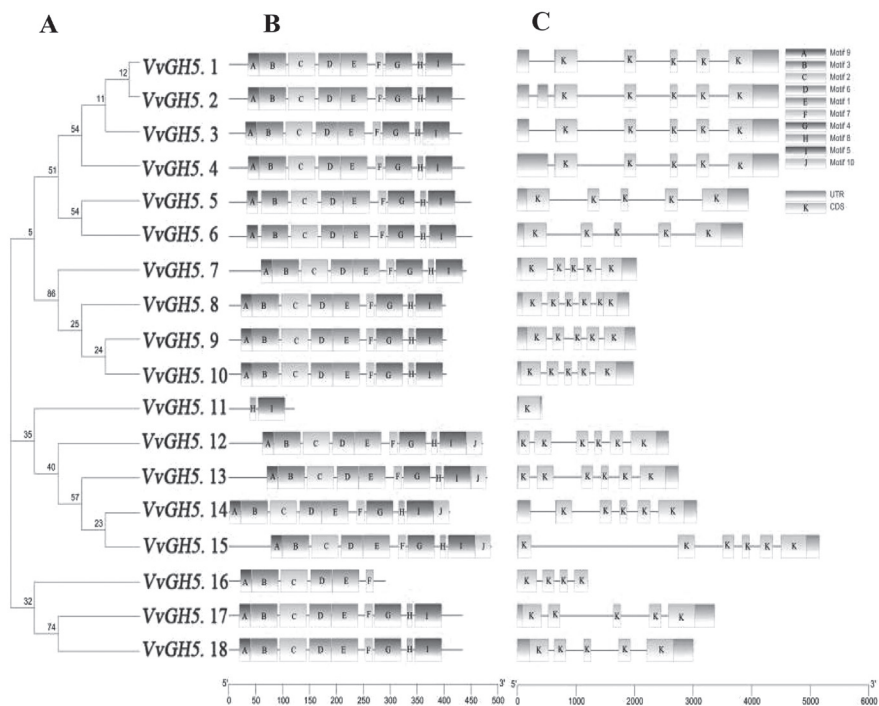


图6 葡萄VvGH5基因家族系统发育树、保守基序及基因结构联合分析
Fig. 6 Combined analysis of phylogenetic tree, conserved motifs and gene structure of VvGH5 gene family in grapevine

A: 葡萄*VvGH5*基因家族系统发育树; B: 葡萄*VvGH5*基因家族保守基序; C: 葡萄*VvGH5*基因家族基因结构。

GH5蛋白质都是由α-螺旋(32.65%~53.72%)、β-折叠(5.46%~8.80%)、延伸链(10.74%~19.68%)和无规卷曲(28.93%~44.15%) 4种二级结构构成。其中, 在VvGH5蛋白二级结构中, 含β-折叠很少; 在所有蛋白的二级结构中, α-螺旋(32.65%~53.72%)和无规则卷曲(28.93%~44.15%)占比最大, 占据主体部分。仅在VvGH5.2、VvGH5.6、VvGH5.9、VvGH5.16、VvGH5.18中, α-螺旋占比>无规则卷曲占比; 其余蛋白二级结构中无规则卷曲占比>α-螺旋占比, 因此可以判断, 无规卷曲(28.93%~44.15%)和α-螺旋(32.65%~53.72%)是葡萄VvGH5蛋白的主要组成成分; β-折叠(5.46%~8.80%)和延伸链(10.74%~19.68%)在蛋白序列中均有分布, 起辅助修饰作用(Qian等2017)。葡萄*VvGH5*基因家族的三维构象同源建模结果表明, VvGH5.1、VvGH5.3、VvGH5.6、VvGH5.11、VvGH5.12、VvGH5.18蛋白的三级结构相似; VvGH5.14和VvGH5.15蛋白的三级结构相似; VvGH5.4、VvGH5.5、VvGH5.7蛋白的三级结

构也相似; VvGH5.8、VvGH5.9、VvGH5.10蛋白的三级结构也相似; VvGH5.16和VvGH5.17蛋白的三级结构也相似。而VvGH5.2和VvGH5.13结构与其他蛋白的三级结构差异很大, 说明葡萄*VvGH5*基因家族蛋白之间既存在保守性又存在差异性(图9)。葡萄*VvGH5*跨膜螺旋预测结果表明, 葡萄*VvGH5*基因家族蛋白位于膜细胞质侧的总概率最大的是VvGH5.1、VvGH5.15和VvGH5.17, 最小的是VvGH5.7, 范围在0.50%~99.88%之间(Krogh等2001)。共有跨膜螺旋12个, 除了VvGH5.4蛋白有2个跨膜螺旋外, VvGH5.1、VvGH5.3、VvGH5.6、VvGH5.8、VvGH5.10、VvGH5.11、VvGH5.13、VvGH5.15、VvGH5.16和VvGH5.17的跨膜螺旋数量均为1(图10)。

2.6 葡萄VvGH5基因家族顺式作用元件分析

顺式作用元件预测表明, 葡萄*VvGH5*基因家族成员起始密码子上游存在着大量植物激素类响应元件, 包括与分生组织表达相关、防御和应激反

表1 葡萄*VvGH5*蛋白质亚细胞定位

Table 1 Subcellular localization of grape *VvGH5* protein

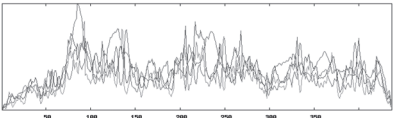
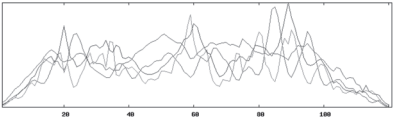
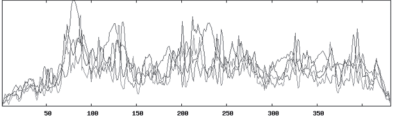
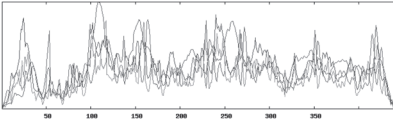
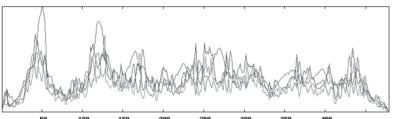
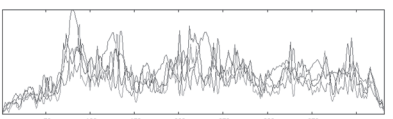
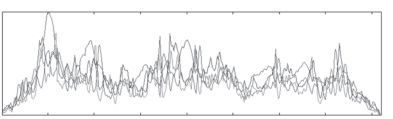
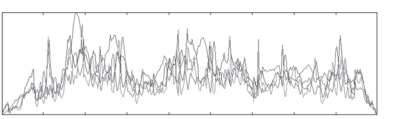
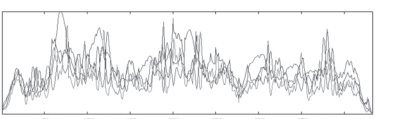
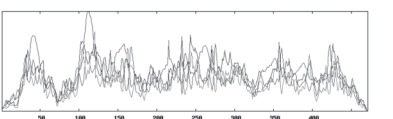
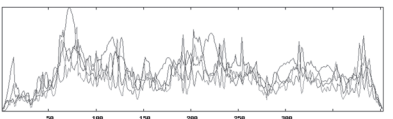
名称	液泡	高尔基体	叶绿体	细胞核	细胞质	质膜	细胞外 基质	内质网	过氧物 酶体	线粒体	细胞骨架
<i>VvGH5.1</i>	4	3	1	1	1	1	1	1	1	0	0
<i>VvGH5.2</i>	0	0	1	19.5	7	0	0	0	0	0	0
<i>VvGH5.3</i>	0	2	1	1	1	1	3	4	0	1	0
<i>VvGH5.4</i>	3	2	8	0	0	0	1	0	0	0	0
<i>VvGH5.5</i>	2	0	1	1	2	7.5	0	7.5	0	1	0
<i>VvGH5.6</i>	6	2	4	1	0	0	1	0	0	0	0
<i>VvGH5.7</i>	1	0	1	9	1	0	0	0	0	2	0
<i>VvGH5.8</i>	2	0	3	6	1	0	0	0	1	1	0
<i>VvGH5.9</i>	1	0	4	4	0	0	3	2	0	0	0
<i>VvGH5.10</i>	1	0	7	3	1	0	0	2	0	0	0
<i>VvGH5.11</i>	7	2	2	1	0	0	2	0	0	0	0
<i>VvGH5.12</i>	6	2	3	1	0	0	2	0	0	0	0
<i>VvGH5.13</i>	6	1	2	0	1	0	3	0	0	1	0
<i>VvGH5.14</i>	1	0	1	5	5	3	0	0	0	0	1.5
<i>VvGH5.15</i>	4	3	1	1	1	1	1	1	1	0	0
<i>VvGH5.16</i>	2	0	2	7	1	0	0	0	1	1	0
<i>VvGH5.17</i>	4	3	1	1	1	1	1	1	1	0	0
<i>VvGH5.18</i>	5	0	3	0	0	0	3	2	0	1	0
合计	55	20	46	61.5	23	14.5	21	20.5	5	8	1.5

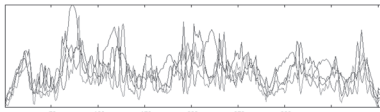
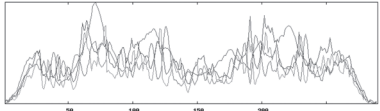
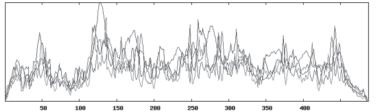
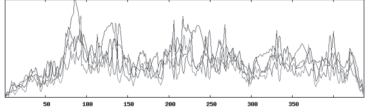
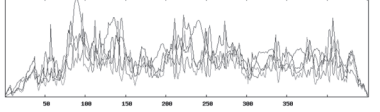
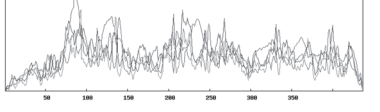
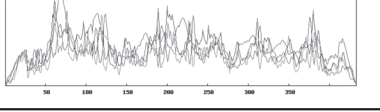
表2 葡萄*VvGH5*基因家族蛋白质基本特性分析

Table 2 Analysis of the basic characteristics of *VvGH5* gene family proteins in grape

基因编号	分子式	氨基酸数目	等电点	分子量	原子数	不稳定系数	脂肪系数	总平均亲水性
<i>VvGH5.1</i>	C ₂₂₇₂ H ₃₄₃₉ N ₆₀₉ O ₆₄₀ S ₁₉	437	8.63	50 134.14	6 979	35.73	72.79	-0.344
<i>VvGH5.2</i>	C ₅₇₉ H ₉₁₂ N ₁₆₄ O ₁₇₅ S ₉	121	9.64	13 259.14	1 839	38.21	65.29	-0.357
<i>VvGH5.3</i>	C ₂₂₄₅ H ₃₃₉₁ N ₅₉₇ O ₆₃₂ S ₁₈	432	8.44	49 433.32	6 883	36.20	73.63	-0.316
<i>VvGH5.4</i>	C ₂₂₁₀ H ₃₃₅₁ N ₅₉₅ O ₆₅₄ S ₂₀	440	7.63	49 360.71	6 830	34.76	73.34	-0.250
<i>VvGH5.5</i>	C ₂₄₅₃ H ₃₆₈₀ N ₆₂₄ O ₇₃₆ S ₂₀	479	5.15	54 329.14	7 513	32.76	73.61	-0.294
<i>VvGH5.6</i>	C ₂₀₁₆ H ₃₀₇₁ N ₅₃₇ O ₆₀₁ S ₁₆	401	5.91	44 959.76	6 241	38.45	78.28	-0.206
<i>VvGH5.7</i>	C ₂₁₁₄ H ₃₁₆₆ N ₅₄₄ O ₆₂₂ S ₁₇	410	5.75	46 698.69	6 463	34.77	70.80	-0.350
<i>VvGH5.8</i>	C ₂₃₁₁ H ₃₄₉₆ N ₆₀₂ O ₆₇₆ S ₁₃	449	5.56	50 945.59	7 098	40.56	82.94	-0.255
<i>VvGH5.9</i>	C ₂₂₀₄ H ₃₃₂₉ N ₆₀₁ O ₆₄₆ S ₁₂	433	9.19	48 966.04	6 792	35.25	70.09	-0.455
<i>VvGH5.10</i>	C ₂₃₈₉ H ₃₆₁₃ N ₆₁₅ O ₇₁₉ S ₁₈	471	5.27	53 030.74	7 354	36.77	76.31	-0.265
<i>VvGH5.11</i>	C ₂₀₄₆ H ₃₁₂₆ N ₅₄₄ O ₆₀₄ S ₁₇	403	7.64	45 553.63	6 337	36.83	77.15	-0.289
<i>VvGH5.12</i>	C ₂₀₁₉ H ₃₀₈₁ N ₅₃₉ O ₆₀₇ S ₁₈	403	5.99	45 194.00	6 264	36.90	77.15	-0.219
<i>VvGH5.13</i>	C ₁₄₈₉ H ₂₂₅₂ N ₄₀₂ O ₄₃₅ S ₁₂	290	6.39	33 129.41	4 590	31.08	73.21	-0.338
<i>VvGH5.14</i>	C ₂₅₁₃ H ₃₇₈₆ N ₆₄₈ O ₇₂₈ S ₂₂	487	6.06	55 428.93	7 697	32.12	32.12	-0.255
<i>VvGH5.15</i>	C ₂₂₇₂ H ₃₄₃₉ N ₆₀₉ O ₆₄₀ S ₁₉	437	8.63	50 134.14	6 979	35.73	72.79	-0.344
<i>VvGH5.16</i>	C ₂₃₁₈ H ₃₅₀₀ N ₆₀₄ O ₆₇₇ S ₁₃	451	5.56	51 077.71	7 112	40.38	81.71	-0.255
<i>VvGH5.17</i>	C ₂₂₇₂ H ₃₄₃₉ N ₆₀₉ O ₆₄₀ S ₁₉	437	8.63	50 134.14	6 979	35.73	72.79	-0.344
<i>VvGH5.18</i>	C ₂₂₃₃ H ₃₃₉₈ N ₆₃₆ O ₆₃₁ S ₁₃	433	9.83	49 666.21	6 911	35.44	72.56	-0.460

表3 葡萄VvGH5蛋白质二级结构预测
Table 3 Prediction of the secondary structure of grape VvGH5 protein

基因名	占比/%				二级结构元件分布
	α -螺旋	β -折叠	无规则卷曲	延伸链	
VvGH5.1	35.93	5.72	38.67	19.68	
VvGH5.2	53.72	6.61	28.93	10.74	
VvGH5.3	34.95	8.80	37.96	18.29	
VvGH5.4	34.32	7.50	40.00	18.18	
VvGH5.5	36.95	6.89	39.04	17.12	
VvGH5.6	38.90	7.98	36.16	16.96	
VvGH5.7	33.90	8.54	40.00	17.56	
VvGH5.8	34.74	8.02	39.87	17.37	
VvGH5.9	38.80	7.62	37.18	16.40	
VvGH5.10	33.55	7.64	39.49	19.32	
VvGH5.11	37.22	6.70	37.72	18.36	

基因名	占比/%				二级结构元件分布
	α -螺旋	β -折叠	无规则卷曲	延伸链	
<i>VvGH5.12</i>	35.48	5.46	41.19	17.87	
<i>VvGH5.13</i>	33.45	7.24	40.69	18.62	
<i>VvGH5.14</i>	32.65	6.57	44.15	16.63	
<i>VvGH5.15</i>	35.93	5.72	38.67	19.68	
<i>VvGH5.16</i>	38.58	7.54	37.69	16.19	
<i>VvGH5.17</i>	35.93	5.72	38.67	19.68	
<i>VvGH5.18</i>	40.40	6.93	37.18	15.47	

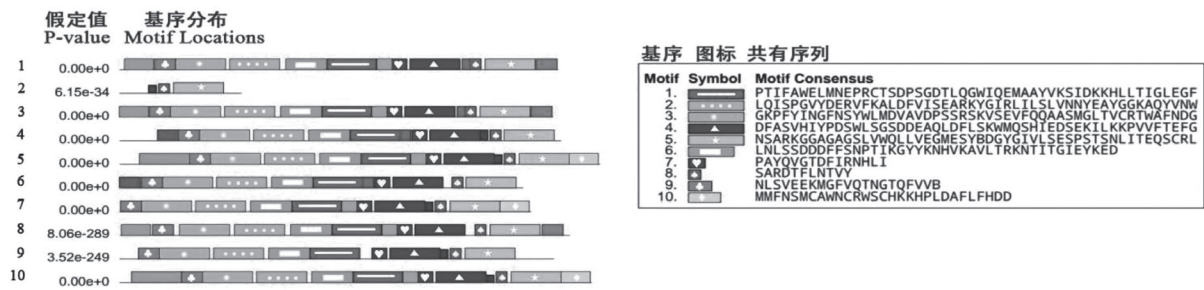


图7 葡萄VvGH5基因家族蛋白保守基序分布

Fig. 7 Distribution of protein conserved motifs of VvGH5 gene family in grapevine

应相关的顺式调控元件; 参与赤霉素应答、生长素应答、脱落酸应答、水杨酸响应、MeJA应答、昼夜节律控制、胚乳表达、低温响应、胚乳特异性负表达、干旱诱导的MYB结合位点的顺式调控元件; 此外, 还有光响应的顺式调节元件和玉米醇溶蛋白代谢调控的顺式调控元件, 在葡萄GH5基因家族成员启动子上也很常见(图11)。上述结果暗示GH5家族基因可能在葡萄生长分化与逆境胁迫

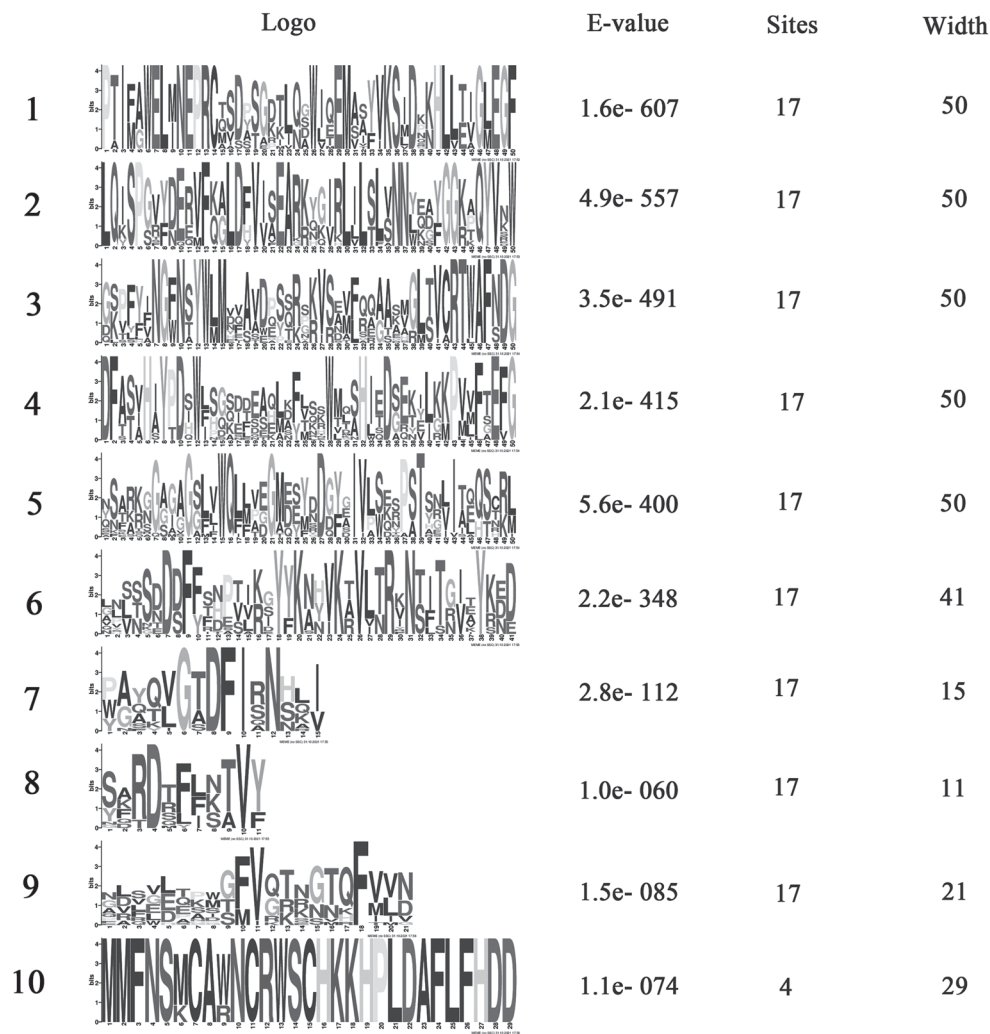


图8 葡萄VvGH5基因家族蛋白保守基序Logo分析
Fig. 8 Logo analysis of protein conserved motifs of grape VvGH5 gene family

过程中发挥作用(Wang等2021; 李濯雪等2015)。

3 讨论

纤维素酶可以将纤维素分解成寡糖或单糖，在细胞壁合成与降解过程中发挥重要作用，高等植物中的CH5基因家族拥有功能最多样化且数量最多的植物纤维素酶基因(李娜等2021; 何芳芳2020)。本研究在全基因组水平上共筛选出在葡萄5条染色体上的18个GH5家族基因。葡萄VvGH5基因结构分析显示，VvCH5家族基因在葡萄第2、3、4号染色体中都存在簇状分布现象，说明葡萄VvGH5

家族基因扩增中可能存在串联复制。对18个葡萄VvGH5基因家族成员进行进化树的构建，可分为三大类，其中I类中的同源性最高为VvGH5.7基因与VvGH5.8基因、VvGH5.9基因、VvGH5.10基因，达到86%，II类中的VvGH5.13基因与VvGH5.14基因、VvGH5.15基因同源性最高为57%，III类中的VvGH5.17基因与VvGH5.18基因同源性最高达74%，验证了葡萄VvGH5基因家族的染色体串联复制几率大(Han等2020; Cao等2016)。基因结构是基因进化的研究依据，从VvGH5基因内含子-外显子结构可以看出，同一进化支的VvGH5基因家族成员虽然

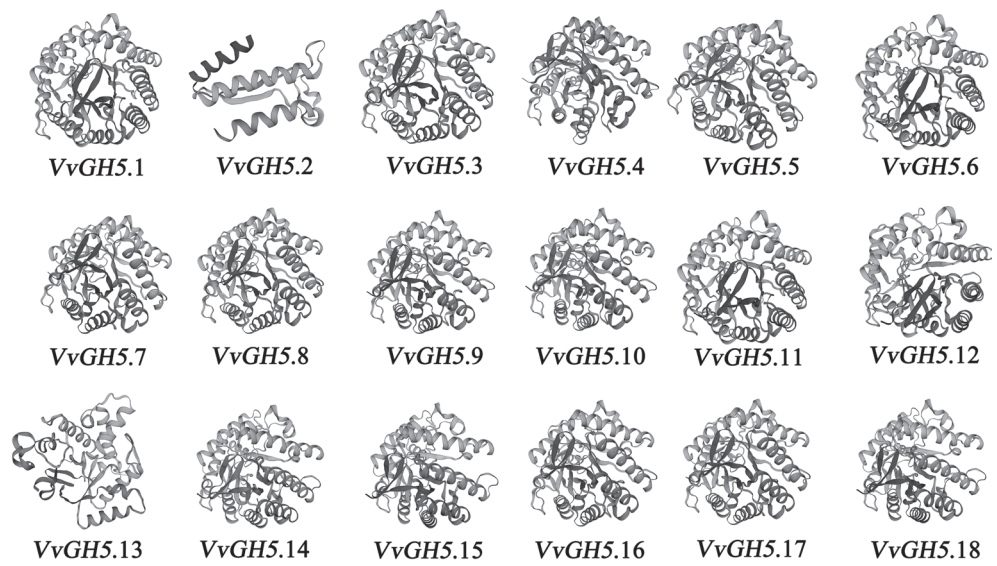


图9 葡萄VvGH5基因家族蛋白三级结构模型
Fig. 9 The tertiary structure model of grape VvGH5 gene family protein

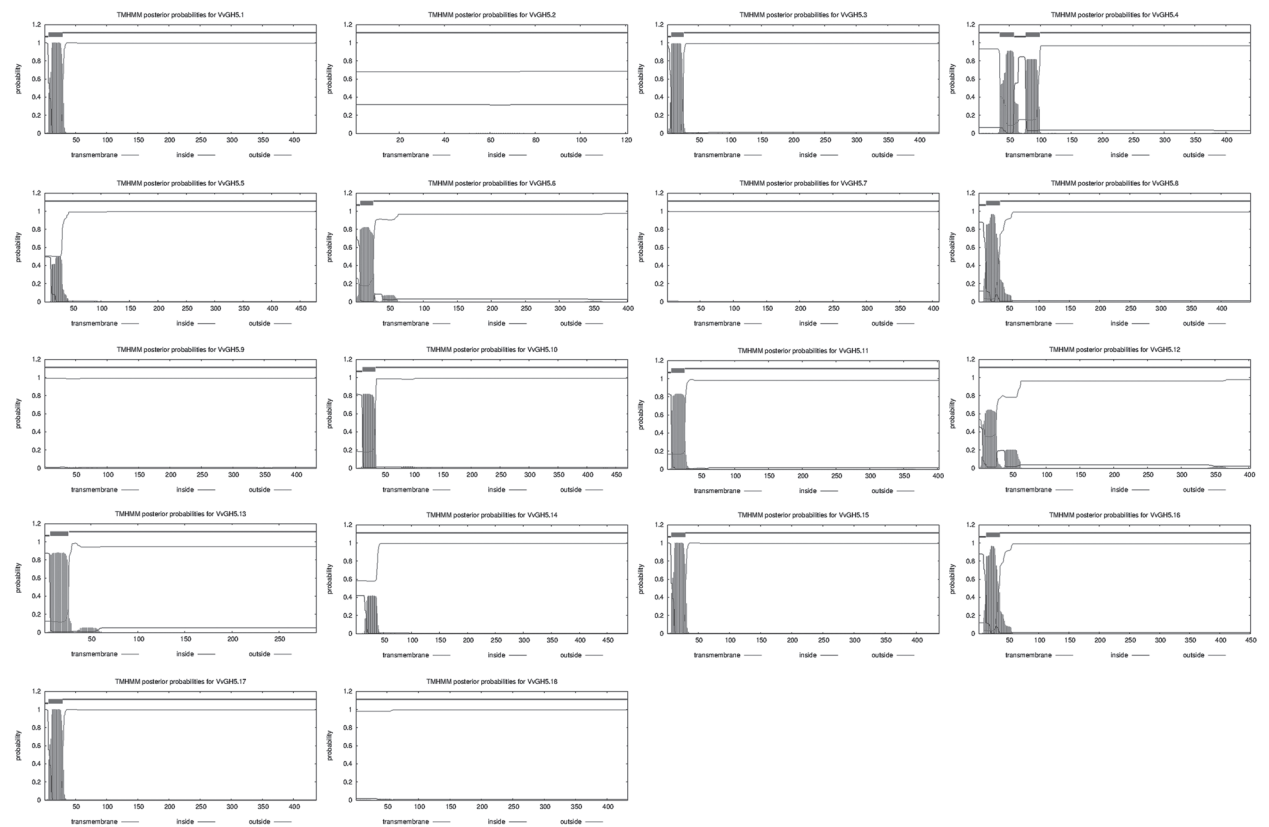


图10 葡萄VvGH5基因家族蛋白跨膜螺旋
Fig. 10 Transmembrane helix of VvGH5 gene family proteins in grapes

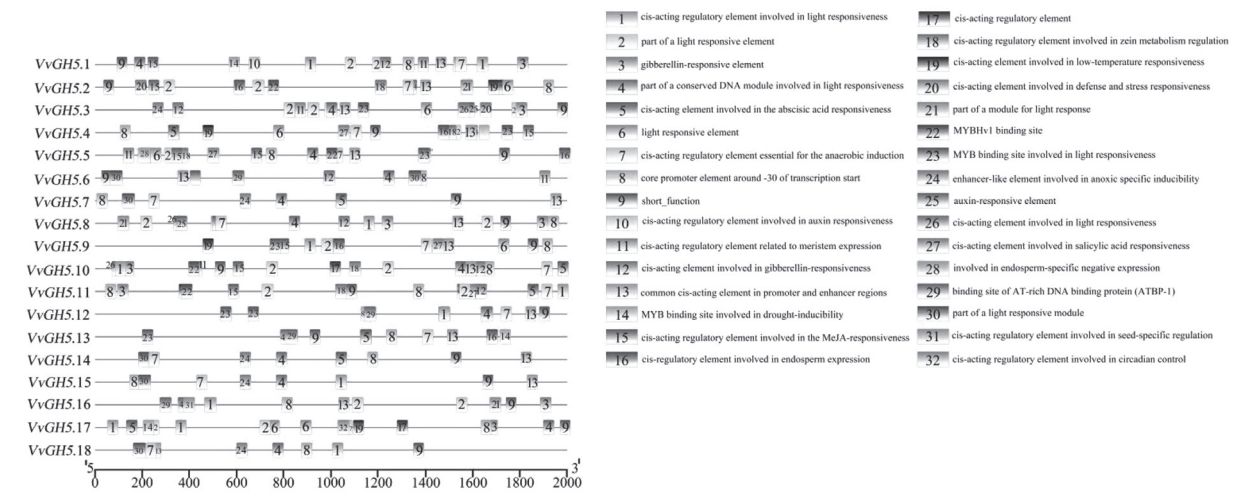


图11 葡萄VvGH5基因家族顺式作用元件分析
Fig. 11 Analysis of cis-acting elements of grape VvGH5 gene family

内含子和外显子数量不同,但是具有相似的排列模式;不同进化枝亚组成员基因结构差异较大,说明*VvGH5*基因家族的不同亚组成员具有不同的进化轨迹(Lu等2020; Marchler-Bauer等2011)。18个葡萄VvGH5蛋白主要分布在液泡、高尔基体、叶绿体、细胞核中,且均为亲水性蛋白,绝大部分蛋白稳定(Cheng等2021; Wang等2012)。含有无规卷曲(28.93%~44.15%)和 α -螺旋(32.65%~53.72%),少量存在的 β -折叠(5.46%~8.80%)和延伸链(10.74%~19.68%)散布于蛋白序列中,起一定辅助修饰的作用,三维构象则更为直观地体现出无规则卷曲和 α -螺旋是葡萄VvGH5蛋白的主要组成部分(Chen等2006)。葡萄*VvGH5*基因家族顺式作用元件分析结果表明,茉莉酸甲酯、生长素、脱落酸、赤霉素等植物激素类响应元件、厌氧诱导元件、光诱导茎叶特异表达作用元件、防卫和胁迫响应元件、低温响应元件、盐胁迫作用元件、昼夜节律调控元件等不同类型元件也均匀分布在*VvGH5*基因家族的启动子区,说明*GH5*家族基因可能在葡萄生长分化与逆境胁迫过程中发挥作用(谷家茂等2021; 孟亚轩等2021; 李濯雪等2015)。然而这些基因具体的功能还需后续研究鉴定。

综上所述,利用生物信息学工具对葡萄*VvGH5*基因家族进行全基因组范围的挖掘鉴定,并对其进行了全面的生物信息学预测分析,结果为进一

步探究葡萄*VvGH5*基因家族功能以及利用该基因家族探索改良葡萄非生物胁迫响应等方面的生物学功能提供了理论依据和数据支持。

参考文献(References)

Alistair MD, Vladimir VLK, Rabie SD, et al (2020). UniRule: a unified rule resource for automatic annotation in the UniProt Knowledgebase. *Bioinformatics*, 36: 4643–4648

Almagro JJ A, Tsigos KD, Sønderby CK, et al (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*, 37: 420–423

Bailey TL, Elkan C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, 2: 28–36

Cao Y, Han Y, Jin Q, et al (2016). Comparative genomic analysis of the *GRF* genes in Chinese pear (*Pyrus bretschneideri* Rehd), poplar (*Populus*), grape (*Vitis vinifera*), *Arabidopsis* and rice (*Oryza sativa*). *Front Plant Sci*, 7: 1750

Chen CJ, Chen H, Zhang Y, et al (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*, 13: 1194–1202

Chen GH, Wang P, Shi L (2021). Review on the dynamic function of cell walls in plant disease resistance. *J Inner Mongolia Agric Univ (Nat Sci Edi)*, 42 (5): 117–120 (in Chinese with English abstract) [陈贵华, 王萍, 石岭(2021). 细胞壁在植物抗病反应中的能动作用研究进展. 内蒙古农业大学学报(自然科学版), 42 (5): 117–120]

Chen R, Chen M, Yang CX, et al (2006). Three-dimensional structure modeling of protein based on SWISS-MODEL. *Chem Life*, 26 (1): 54–56 (in Chinese) [谌蓉, 陈敏, 杨春

- 贤等(2006). 基于SWISS-MODEL的蛋白质三维结构建模. 生命的化学, 26 (1): 54–56]
- Chen Y, Yu P, Luo J, et al (2003). Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. Mamm Genome, 14: 859–865
- Cheng X (2021). Bioinformatics analysis of tomato *OFP* gene family. Jiangsu Agric Sci, 49 (18): 39–48 (in Chinese) [程欣(2021). 番茄*OFP*基因家族生物信息学分析. 江苏农业科学, 49 (18): 39–48]
- Chou KC, Shen HB (2010). Cell-PLoc 2.0: an improved package of web-servers for predicting subcellular localization of proteins in various organisms. Nat Sci, 2: 1090–1103
- Dong X, Chen W, Liang Z, et al (2020). VitisGDB: the multi-functional database for grapevine breeding and genetics. Mol Plant, 13: 1098–1100
- Florian K, Konstantin A, Michael K, et al (2009). The SWISS-MODEL repository and associated resources. Nucleic Acids Res, 37: 387–392
- Gu JM, Liu YF, Wang F, et al (2021). Advances of plant *ABCIK* gene family. Plant Physiol J, 57 (1): 33–40 (in Chinese with English abstract) [谷家茂, 刘玉凤, 王峰等(2021). 植物*ABCIK*基因家族研究进展. 植物生理学报, 57 (1): 33–40]
- Han B, Liu J, Zhao XH, et al (2020). Bioinformatics analysis of *JAZ* gene family in maize. J Shanxi Agric Sci, 48 (10): 1552–1556 (in Chinese with English abstract) [韩斌, 刘江, 赵兴华等(2020). 玉米*JAZ* 基因家族的生物信息学分析. 山西农业科学, 48 (10): 1552–1556]
- Han XY, Liu YH (2020). Protein secondary structure prediction server PSRSM. Bioinformatics, 18 (2): 116–126 (in Chinese with English abstract) [韩心怡, 刘毅慧(2020). 蛋白质二级结构预测服务器PSRSM. 生物信息学, 18 (2): 116–126]
- He FF, Wang HJ, Wang XY (2020). Research progress of cellulase. Paper Sci Technol, 39 (4): 1–8 (in Chinese with English abstract) [何芳芳, 王海军, 王雪莹(2020). 纤维素酶的研究进展. 造纸科学与技术, 39 (4): 1–8]
- Höfte H, Voxeur A (2017). Plant cell walls. Curr Biol, 17: 865–870
- Jiang QQ, Wang YT, Hui ZM (2021). Identification and expression analysis of BZR gene family in grapevine. Plant Physiol J, 57 (6): 1218–1228 (in Chinese with English abstract) [江倩倩, 王雨婷, 惠竹梅(2021). 葡萄BZR基因家族的鉴定及表达分析. 植物生理学报, 57 (6): 1218–1228]
- Khan Y, Khan SM, ul Haq I, et al (2021). Antioxidant potential in the leaves of grape varieties (*Vitis vinifera* L.) grown in different soil compositions. Arab J Chem, 11: 103412
- Krogh A, Larsson B, Heijne GV, et al (2001). Predicting trans-membrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol, 305: 567–580
- Larkin MA, Blackshields G, Brown NP, et al (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947–2948
- Li N, Xia H, Jiang YB (2021). Research progress of fusion glycoside hydrolase in biomass conversion. Sci Sin, 51 (7): 831–843 (in Chinese with English abstract) [李娜, 夏欢, 江燕斌(2021). 融合糖苷水解酶在生物质转化中的研究进展. 中国科学, 51 (7): 831–843]
- Li S, Lin H, J Huang, et al (2021). Preparation of soybean β -conglycinin epitope antibody and its preliminary application in frozen surimi detection. Eur Food Res Technol, 247: 1411–1423
- Li ZX, Chen XB (2015). Research advances on plant inducible promoters and related *cis*-acting elements. Biotech Bull, 31 (10): 8–15 (in Chinese with English abstract) [李濯雪, 陈信波(2015). 植物诱导型启动子及相关顺式作用元件研究进展. 生物技术通报, 31 (10): 8–15]
- Lu SN, Wang JY, Farideh C, et al (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res, 48: 265–268
- Marchler-Bauer A, Anderson JB, Chitaz F, et al (2011). CDD: specific functional annotation with the conserved domain database. Nucleic Acids Res, 9: 225–229
- Meng YX, Sun YQ, Zhao XY, et al (2021). Identification and expression analysis of millet *GH5* gene family. Acta Agric Zhejiangensis, 33 (10): 1797–1807 (in Chinese with English abstract) [孟亚轩, 孙颖琦, 赵心月等(2021). 谷子*GH5*基因家族全基因组鉴定和表达分析. 浙江农业学报, 33 (10): 1797–1807]
- Paul H, Keun-Joon P, Takeshi O, et al (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res, 35: 585–587
- Qian J, Xin J, Lee SJ, et al (2017). Protein secondary structure prediction: A survey of the state of the art. J Mol Graph, 76: 379–402
- Rosa SBA, Csordas, BG, Oliveira SMDVL, et al (2020). Prediction of conserved peptides of *Paracoccidioides* for Interferon- γ release assay: the first step in the development of a Lab-Based Approach for immunological assessment during antifungal therapy. J Fungi, 6: 378–379
- Tamura K, Stecher G, Peterson D, et al (2013). MEGA6: molecular evolutionary genetics analysis Version 6.0. Mol Biol Evol, 30: 2725–2729
- Wang W, Quan C, Zheng S, et al (2021). OsPM1 is a positive regulator of rice tolerance to drought stress but a negative regulator of rice tolerance to salt stress. J Plant Interact,

- 16: 213–221
- Wang Y, Deng D, Shi Y, et al (2012). Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights gained from analyzing maize *ARF* genes. *Mol Biol Rep*, 39: 2401–2415
- Wintersinger JA, Wasmuth JD (2015). Kablammo: an interactive, web-based BLAST results visualizer. *Bioinformatics*, 31: 1305–1306
- Yuan HM, Guo WD, Zhao LJ, et al (2016). Bioinformatics and expression analysis of the cellulose synthase supergene family in flax. *Sci Agric Sin*, 49 (23): 4656–4668 (in Chinese with English abstract) [袁红梅, 郭文栋, 赵丽娟等(2016). 亚麻纤维素合酶超基因家族的生物信息学及表达分析. *中国农业科学*, 49 (23): 4656–4668]
- Zeng JJ, Chen J, Chen Z, et al (2021). Screening of compound microorganism and application effect of straw fermentation. *China Brew*, 40 (3): 124–128 (in Chinese with English abstract) [曾佳佳, 陈静, 陈振等(2021). 复合微生物的筛选及其发酵秸秆的应用效果. *中国酿造*, 40 (3): 124–128]