



被子植物小热激蛋白家族的比较基因组学分析

范凯, 叶方婷, 毛志君, 潘鑫峰, 李兆伟, 林文雄*

福建农林大学农学院, 作物遗传育种与综合利用教育部重点实验室, 福州 350002

摘要 小热激蛋白(sHSP)是一类重要的响应外界环境变化以及调控植物生长发育的蛋白家族。基于在睡莲(*Nymphaea colorata*)、水稻(*Oryza sativa*)、拟南芥(*Arabidopsis thaliana*)和葡萄(*Vitis vinifera*)中分别鉴定到的33个NcsHSPs、24个OssHSPs、17个AtsHSPs和47个VvsHSPs, 表明sHSP家族可分为12个亚家族, 不同亚家族包含不同的sHSP成员数目、保守基序、基因结构以及复制基因数目。在4种模式被子植物的sHSP成员中共鉴定到12个基因复制事件, 片段复制事件和串联复制事件均与sHSP成员的扩增有关, 且片段复制事件发生的时间早于串联复制事件。在所有sHSP成员中, 拟南芥和葡萄的sHSP成员的同源性最高, 其次为睡莲和葡萄的sHSP成员。sHSP家族在被子植物中可能向更短的氨基酸长度、更小的分子量、更简单的基因结构以及更集中的染色体分布进化。此外, 在睡莲、水稻、拟南芥和葡萄中鉴定了一些可能与调控植物生长发育相关的候选基因。研究结果为4种模式被子植物sHSP家族的比较基因组学研究奠定了重要基础, 并为其它被子植物sHSP家族的研究提供重要参考。

关键词 sHSP家族, 比较基因组学, 扩增, 基因表达

范凯, 叶方婷, 毛志君, 潘鑫峰, 李兆伟, 林文雄 (2021). 被子植物小热激蛋白家族的比较基因组学分析. 植物学报 56, 245–261.

被子植物在日常生活中随处可见, 可为人们提供粮食、纤维、燃料和药品等。被子植物的多样性与外界环境变化息息相关。在众多被子植物中, 基部被子植物睡莲(*Nymphaea colorata*) (Zhang et al., 2020a)、单子叶模式植物水稻(*Oryza sativa*) (Ouyang et al., 2007)、双子叶模式植物拟南芥(*Arabidopsis thaliana*) (Lamesch et al., 2012)以及类似古代双子叶植物基因组的葡萄(*Vitis vinifera*) (The French-Italian Public Consortium for Grapevine Genome Characterization, 2007)在被子植物分类学中具有重要意义。目前, 睡莲、水稻、拟南芥和葡萄的基因组测序已经完成, 从与外界环境变化相关的基因家族层面上对这4类被子植物进化历程进行研究, 不仅有助于解析被子植物进化的方向, 而且可为后续分子育种提供重要参考。

在外界环境发生变化时, 植物在生理和分子水平上发生一系列改变以适应环境。热激蛋白(heat shock

protein, HSP)是一类重要的响应外界环境变化的蛋白家族, 小热激蛋白(small heat shock protein, sHSP)是其中一类分子量较小的热激蛋白(刘德立, 1996; 黄祥富等, 1999; Wang et al., 2004; 栗振义等, 2016)。目前已在古生菌、细菌和真核生物中发现sHSP成员(Waters and Vierling, 2020)。sHSP成员具有保守的 α -晶体结构域, 该结构域由80–100个氨基酸残基构成的结构紧凑的 β 链组成(van Montfort et al., 2001)。在植物中, 首先在烟草(*Nicotiana tabacum*)和大豆(*Glycine max*)中鉴定出sHSP成员(Key et al., 1981), 随后相继在拟南芥(Scharf et al., 2001)、水稻(Sarkar et al., 2009)、葡萄(Ji et al., 2019)、番茄(*Solanum lycopersicum*) (张宁等, 2017)、马铃薯(*S. tuberosum*) (Zhao et al., 2018)、棉花(*Gossypium hirsutum*) (Ma et al., 2016)、银杏(*Ginkgo biloba*) (何福林和张斌, 2019)、辣椒(*Capsicum annuum*) (Guo et al., 2015)和簸箕柳

收稿日期: 2020-12-25; 接受日期: 2021-03-25

基金项目: 国家自然科学基金(No.31701470)、中国博士后基金(No.2017M610388, No.2018T110637)、福建农林大学杰出青年科研人才计划(No.xjq201917)和福建农林大学科技创新专项基金(No.CXZX2020007A)

* 通讯作者。E-mail: lwx@fafu.edu.cn

(*Salix suchowensis*) (Li et al., 2018)等植物中发现了大量的sHSP成员。sHSP家族主要作为分子伴侣,以独立于ATP的形式保护植物免受外界环境的影响(Waters and Vierling, 2020)。大量研究表明, sHSP成员可以响应外界多种环境胁迫(Sun et al., 2010; Zhu et al., 2011; Li et al., 2016; Zhai et al., 2016; Guo et al., 2020)。此外, sHSP家族也参与果实品质形成和植株发芽等植物生长发育过程(Zhang et al., 2018; Ma et al., 2019)。

目前,虽然已在许多植物中鉴定到sHSP成员,但主要集中在对单个物种进行全基因组分析,缺乏对植物整体进化趋势的认识。睡莲、水稻、拟南芥和葡萄在被子植物进化过程中具有重要意义。目前,睡莲(Zhang et al., 2020a)、水稻(Ouyang et al., 2007)、拟南芥(Lamesch et al., 2012)和葡萄(The French-Italian Public Consortium for Grapevine Genome Characterization, 2007)的全基因组测序、组装以及注释工作已经完成,这为在4种被子植物中进行sHSP家族的比较基因组学分析奠定了基础。本研究利用最新的睡莲、水稻、拟南芥和葡萄的全基因组测序信息,对其sHSP家族进行鉴定,并在此基础上进一步对其系统发育、保守基序、基因结构、染色体定位、基因复制事件、共线性和基因表达等方面进行比较基因组学分析,旨在为从基因家族层面进一步揭示植物的进化方向提供重要参考,同时也为分子育种提供具有潜在育种价值的候选基因。

1 材料与方法

1.1 sHSP成员的鉴定以及系统进化分析

水稻(*Oryza sativa* L.)、拟南芥(*Arabidopsis thaliana* L.)和葡萄(*Vitis vinifera* L.)的基因组序列信息从Phytozome数据库(<https://phytozome.jgi.doe.gov/pz/portal.html>)下载。从Waterlily Pond数据库(<http://waterlily.eplant.org>)下载睡莲(*Nymphaea colorata* L.)的基因组序列信息。sHSP家族的HMM模型(PF00011)从pfam数据库(<http://pfam.xfam.org/>)下载。使用HMMER 3.0程序中的hmmsearch工具,利用PF00011作为探针分别在睡莲、水稻、拟南芥和葡萄基因组中搜索sHSP成员,参数除cut_ga和domtblout外均使用默认参数。随后,收集所有可能的

sHSP成员并进一步利用CDD程序(<https://www.ncbi.nlm.nih.gov/cdd>)使用默认参数进行注释,确保每个成员均含有sHSP家族保守的结构域。睡莲、水稻、拟南芥和葡萄中sHSP成员分别记为NcsHSP、OssHSP、AtsHSP和VvsHSP。

通过MAFFT软件对睡莲、水稻、拟南芥和葡萄中sHSP成员的氨基酸序列进行序列连配分析。MAFFT软件参数设置见本实验室之前的研究(Fan et al., 2020b)。采用IQ-tree软件构建所有sHSP成员的系统进化树,其中最适模型主要利用ModelFinder程序进行选择,bootstrap值设定为1 000。

1.2 sHSP成员的序列以及结构分析

睡莲、水稻、拟南芥和葡萄中sHSP成员的蛋白质结构特征在ProtParam网站(<http://web.expasy.org/protparam/>)进行计算。NcsHSP、OssHSP、AtsHSP和VvsHSP的保守基序主要通过MEME程序(<http://meme-suite.org/tools/meme>)进行识别,相关参数设置见本实验室之前的研究(Fan et al., 2020a)。对于通过MEME程序识别的保守基序,可进一步利用CDD程序(<https://www.ncbi.nlm.nih.gov/cdd>)使用默认参数进行注释。从基因组序列信息中提取NcsHSP、OssHSP、AtsHSP和VvsHSP的染色体定位信息以及基因结构信息,利用TBtools软件进行可视化展示(Chen et al., 2020)。

1.3 sHSP成员的基因复制事件分析以及同源基因的识别

采用MCScan软件使用默认参数分别对睡莲、水稻、拟南芥和葡萄中sHSP成员的基因复制事件进行识别。在睡莲、水稻、拟南芥和葡萄中,同源基因的识别主要基于MCScan软件获得的共线性结果进行确定,同源基因之间的共线性关系主要采用MCScan软件进行可视化分析。此外,复制基因之间以及同源基因之间的非同义替换率(K_a)、同义替换率(K_s)以及它们的比值(K_a/K_s)使用TBtools软件计算获得(Chen et al., 2020)。

1.4 sHSP成员的表达分析

拟南芥AtsHSP和水稻OssHSP的表达谱数据主要从Genevestigator数据库下载获得。葡萄中VvsHSP成

员的表达谱数据主要从Grape RNA数据库(<http://www.grapeworld.cn/gt/index.html>)下载获得。睡莲中NcsHSP成员的表达谱数据主要从NCBI-SRA数据库(<https://www.ncbi.nlm.nih.gov/sra>)下载获得(SRA Study: SRP222853)。下载的睡莲RNA-Seq数据主要通过HISAT2、StringTie和Ballgown流程进行转录组分析(Pertea et al., 2016)。拟南芥AtsHSP和水稻OssHSP的表达谱数据通过基因表达潜力的百分比进行计算;葡萄VvsHSP和睡莲NcsHSP的表达谱数据通过将FPKM值进行log₂(FPKM+1)转化获得。所有NcsHSP、OssHSP、AtsHSP和VvsHSP表达谱数据均使用Rstudio中的ComplexHeatmap包进行聚类分析并进行可视化展示。

2 结果与分析

2.1 4种植物中sHSP成员的识别以及结构特征

采用HMMER 3.0程序从睡莲、水稻、拟南芥和葡萄的基因组数据库中共获得121个sHSP成员,并且每

个sHSP成员均包含保守的PF00011结构域(表1)。其中,在睡莲中获得33个NcsHSP成员,依据其位点名称的先后顺序,将NcsHSP命名为NcsHSP01–NcsHSP33。在水稻中获得24个OssHSP成员,分别命名为OssHSP01–OssHSP24。在拟南芥中获得17个AtsHSP成员,分别命名为AtsHSP01–AtsHSP17。在葡萄中获得47个VvsHSP成员,分别命名为VvsHSP01–VvsHSP47。随后,对121个sHSP成员的蛋白质结构特征进行分析,发现sHSP成员(除VvsHSP10、NcsHSP04、NcsHSP07、NcsHSP18和NcsHSP26)的氨基酸长度介于60–378个,分子量介于6–43 kDa,理论等电点介于4.82–11.34,蛋白不稳定指数介于29.89–93.80,脂溶指数介于59.53–109.16,总平均疏水性介于–1.343–0.075。

2.2 4种植物中sHSP成员的系统进化关系

利用IQ-tree软件对从睡莲、水稻、拟南芥和葡萄中鉴定到的121个sHSP成员进行系统进化分析(图1)。通过ModelFinder模式我们发现JTT+R4为本研究最合

表1 睡莲、水稻、拟南芥和葡萄中sHSP成员的鉴定及结构分析

Table 1 The identification and structural analysis of the sHSP members in waterlily, rice, *Arabidopsis* and grape

Protein name	Locus name	Subfamily	Protein length (aa)	Molecular weight (kDa)	Theoretical pI	Instability index	Aliphatic index	GRAVY
NcsHSP01	GWHPAAYW001658	CI	60	6.64	9.16	41.10	99.00	–0.218
NcsHSP02	GWHPAAYW001878	CII	308	34.77	9.03	44.31	74.64	–0.624
NcsHSP03	GWHPAAYW002745	CV	276	30.85	4.88	44.78	77.36	–0.291
NcsHSP04	GWHPAAYW004666	PX	1189	132.46	5.86	44.51	85.53	–0.300
NcsHSP05	GWHPAAYW006130	ER	218	24.32	5.46	64.70	95.14	–0.411
NcsHSP06	GWHPAAYW006158	CII	378	43.02	8.36	50.40	74.76	–0.780
NcsHSP07	GWHPAAYW006159	CII	662	75.02	9.13	55.90	79.64	–0.665
NcsHSP08	GWHPAAYW006161	CII	314	35.79	8.86	47.40	79.39	–0.709
NcsHSP09	GWHPAAYW006746	CIII	164	18.56	5.75	53.09	81.28	–0.489
NcsHSP10	GWHPAAYW007213	CIV	132	14.36	6.83	58.83	82.80	–0.117
NcsHSP11	GWHPAAYW009125	CII	165	19.09	5.48	54.83	69.03	–0.921
NcsHSP12	GWHPAAYW010630	MTII	235	26.36	6.79	56.35	67.57	–0.687
NcsHSP13	GWHPAAYW010715	CI	129	15.42	9.12	60.26	59.53	–1.343
NcsHSP14	GWHPAAYW013759	CVI	154	17.47	5.57	46.32	73.96	–0.573
NcsHSP15	GWHPAAYW014339	CI	156	17.60	5.24	55.71	73.72	–0.552
NcsHSP16	GWHPAAYW014344	CI	158	18.29	8.92	54.19	74.05	–0.682
NcsHSP17	GWHPAAYW015449	MTI/CP	208	23.47	6.01	36.92	78.75	–0.667
NcsHSP18	GWHPAAYW018459	CI	833	94.56	6.14	53.05	69.45	–0.668
NcsHSP19	GWHPAAYW019072	CI	163	18.55	6.19	62.44	71.04	–0.601
NcsHSP20	GWHPAAYW019077	CI	337	37.62	7.83	57.71	78.66	–0.457

表1 (续)

Table 1 (continued)

Protein name	Locus name	Subfamily	Protein length (aa)	Molecular weight (kDa)	Theoretical pI	Instability index	Aliphatic index	GRAVY
NcsHSP21	<i>GWHPAAYW019078</i>	CI	156	17.38	5.69	54.61	75.00	-0.449
NcsHSP22	<i>GWHPAAYW019481</i>	CP	222	25.65	9.17	54.03	60.63	-0.880
NcsHSP23	<i>GWHPAAYW022038</i>	CI	159	18.24	5.60	65.94	69.18	-0.648
NcsHSP24	<i>GWHPAAYW025914</i>	CVI	160	18.57	5.82	41.67	70.00	-0.914
NcsHSP25	<i>GWHPAAYW027820</i>	ER	192	21.54	6.54	31.35	102.08	-0.202
NcsHSP26	<i>GWHPAAYW027821</i>	ER	676	74.58	5.31	41.90	85.41	-0.292
NcsHSP27	<i>GWHPAAYW027823</i>	ER	180	20.20	9.21	61.79	73.67	-0.636
NcsHSP28	<i>GWHPAAYW027825</i>	ER	191	21.24	7.93	29.89	91.41	-0.376
NcsHSP29	<i>GWHPAAYW027826</i>	ER	191	21.48	6.00	34.01	98.53	-0.246
NcsHSP30	<i>GWHPAAYW027987</i>	CP	226	25.18	6.65	50.88	75.88	-0.522
NcsHSP31	<i>GWHPAAYW030961</i>	ER	133	15.78	10.21	41.23	77.59	-0.929
NcsHSP32	<i>GWHPAAYW030963</i>	ER	126	14.05	5.63	37.53	109.13	-0.024
NcsHSP33	<i>GWHPAAYW030964</i>	ER	143	15.67	6.08	33.52	109.16	0.075
OssHSP01	<i>LOC_Os01g04340.1</i>	CI	150	16.65	6.31	34.230	77.200	-0.536
OssHSP02	<i>LOC_Os01g04350.1</i>	CI	166	17.89	4.96	55.210	87.590	0.051
OssHSP03	<i>LOC_Os01g04360.1</i>	CI	149	16.90	6.76	47.050	69.870	-0.648
OssHSP04	<i>LOC_Os01g04370.1</i>	CI	150	16.94	6.18	52.240	76.530	-0.605
OssHSP05	<i>LOC_Os01g04380.1</i>	CI	150	16.96	6.18	52.240	73.930	-0.617
OssHSP06	<i>LOC_Os01g08860.1</i>	CII	166	18.03	5.61	45.580	78.130	-0.386
OssHSP07	<i>LOC_Os02g03570.1</i>	CI	177	18.87	6.93	52.450	71.750	-0.406
OssHSP08	<i>LOC_Os02g10710.1</i>	MTII	219	23.59	7.74	47.270	86.030	-0.320
OssHSP09	<i>LOC_Os02g12610.1</i>	CII	175	19.02	5.73	39.010	84.060	-0.486
OssHSP10	<i>LOC_Os02g48140.1</i>	CVII	164	17.77	6.97	46.260	67.320	-0.534
OssHSP11	<i>LOC_Os02g52150.1</i>	MTI/CP	221	24.15	7.96	48.750	77.330	-0.519
OssHSP12	<i>LOC_Os02g54140.1</i>	CIII	172	18.60	7.85	40.550	77.150	-0.538
OssHSP13	<i>LOC_Os03g14180.1</i>	CP	240	26.66	6.78	53.350	75.120	-0.541
OssHSP14	<i>LOC_Os03g15960.1</i>	CI	161	17.91	5.79	57.860	64.780	-0.680
OssHSP15	<i>LOC_Os03g16020.1</i>	CI	154	17.37	6.18	49.280	66.430	-0.706
OssHSP16	<i>LOC_Os03g16030.1</i>	CI	161	18.08	6.77	58.690	61.680	-0.746
OssHSP17	<i>LOC_Os03g16040.1</i>	CI	159	17.66	6.18	55.510	66.790	-0.719
OssHSP18	<i>LOC_Os04g36750.1</i>	ER	215	23.23	5.34	35.060	82.650	-0.365
OssHSP19	<i>LOC_Os05g23140.1</i>	CP	251	27.64	9.42	39.120	69.240	-0.488
OssHSP20	<i>LOC_Os05g42120.1</i>	CV	203	22.28	4.82	54.030	69.800	-0.330
OssHSP21	<i>LOC_Os06g11610.1</i>	MTI/CP	248	26.23	5.27	43.470	79.270	-0.323
OssHSP22	<i>LOC_Os06g14240.1</i>	PX	146	16.02	8.13	36.730	84.790	-0.292
OssHSP23	<i>LOC_Os07g33350.1</i>	CIV	219	23.76	11.74	93.800	71.320	-0.642
OssHSP24	<i>LOC_Os11g13980.1</i>	ER	206	21.87	6.01	48.890	81.500	-0.126
AtsHSP01	<i>AT1G07400.1</i>	CI	157	17.83	5.99	43.380	70.060	-0.680
AtsHSP02	<i>AT1G52560.1</i>	MTII	232	26.54	6.86	45.660	68.880	-0.869
AtsHSP03	<i>AT1G53540.1</i>	CI	157	17.60	5.36	53.800	71.970	-0.578
AtsHSP04	<i>AT1G54050.1</i>	CIII	155	17.36	7.88	58.630	93.680	-0.504
AtsHSP05	<i>AT1G59860.1</i>	CI	155	17.62	6.85	50.240	70.970	-0.678
AtsHSP06	<i>AT2G29500.1</i>	CI	153	17.56	6.33	47.820	71.900	-0.720
AtsHSP07	<i>AT3G46230.1</i>	CI	156	17.44	5.20	46.980	68.590	-0.569

表1 (续)

Table 1 (continued)

Protein name	Locus name	Subfamily	Protein length (aa)	Molecular weight (kDa)	Theoretical pI	Instability index	Aliphatic index	GRAVY
AtsHSP08	<i>AT4G10250.1</i>	ER	195	22.00	5.58	36.450	93.950	-0.469
AtsHSP09	<i>AT4G21870.1</i>	CIV	134	15.39	5.17	73.930	93.810	-0.325
AtsHSP10	<i>AT4G25200.1</i>	MTI/CP	210	23.61	6.46	58.790	80.760	-0.649
AtsHSP11	<i>AT4G27670.1</i>	CP	227	25.34	8.49	46.860	72.070	-0.641
AtsHSP12	<i>AT5G12020.1</i>	CII	155	17.62	6.32	38.840	80.390	-0.617
AtsHSP13	<i>AT5G12030.1</i>	CII	156	17.69	5.59	44.370	75.000	-0.596
AtsHSP14	<i>AT5G37670.1</i>	PX	137	15.70	7.94	50.680	86.640	-0.512
AtsHSP15	<i>AT5G51440.1</i>	MTI/CP	210	23.47	8.96	52.510	84.000	-0.558
AtsHSP16	<i>AT5G54660.1</i>	CV	192	21.66	5.47	38.030	71.560	-0.500
AtsHSP17	<i>AT5G59720.1</i>	CI	161	18.13	6.77	48.090	67.200	-0.657
VvsHSP01	<i>VIT_200s0707g00010.1</i>	CIV	136	15.69	4.89	53.490	82.430	-0.315
VvsHSP02	<i>VIT_200s0992g00020.1</i>	CIV	136	15.70	5.01	54.190	82.430	-0.312
VvsHSP03	<i>VIT_201s0010g02290.1</i>	CP	226	25.56	6.77	48.710	62.170	-0.706
VvsHSP04	<i>VIT_202s0154g00480.1</i>	MTI/CP	201	22.45	9.24	45.150	77.060	-0.531
VvsHSP05	<i>VIT_202s0154g00490.1</i>	MTI/CP	201	22.55	9.11	51.710	75.670	-0.561
VvsHSP06	<i>VIT_204s0008g01490.1</i>	CII	156	17.34	5.94	41.520	82.500	-0.444
VvsHSP07	<i>VIT_204s0008g01500.1</i>	CII	152	16.69	6.84	42.090	79.540	-0.390
VvsHSP08	<i>VIT_204s0008g01510.1</i>	CII	156	17.40	5.77	44.190	76.220	-0.462
VvsHSP09	<i>VIT_204s0008g01520.1</i>	CII	156	17.58	5.58	40.180	78.080	-0.513
VvsHSP10	<i>VIT_204s0008g01530.1</i>	CII	480	53.16	8.89	39.780	68.600	-0.713
VvsHSP11	<i>VIT_204s0008g01550.1</i>	CII	156	17.41	5.94	39.580	71.790	-0.533
VvsHSP12	<i>VIT_204s0008g01560.1</i>	CII	113	12.66	8.66	38.840	83.540	-0.473
VvsHSP13	<i>VIT_204s0008g01570.1</i>	CII	166	18.60	5.95	40.640	68.670	-0.542
VvsHSP14	<i>VIT_204s0008g01580.1</i>	CII	156	17.42	6.62	42.220	76.790	-0.510
VvsHSP15	<i>VIT_204s0008g01590.1</i>	CII	155	17.29	5.94	37.610	81.030	-0.466
VvsHSP16	<i>VIT_204s0008g01610.1</i>	CII	158	18.14	6.33	40.180	75.820	-0.617
VvsHSP17	<i>VIT_204s0008g01620.1</i>	CII	159	18.42	8.46	42.080	83.900	-0.459
VvsHSP18	<i>VIT_206s0004g05770.1</i>	CI	144	16.31	6.93	33.000	75.070	-0.642
VvsHSP19	<i>VIT_208s0058g00210.1</i>	CI	148	16.88	5.81	64.560	66.490	-0.629
VvsHSP20	<i>VIT_209s0002g00640.1</i>	CIII	160	17.89	6.30	49.690	81.000	-0.516
VvsHSP21	<i>VIT_209s0002g06790.1</i>	MTII	233	26.31	7.78	53.870	82.060	-0.750
VvsHSP22	<i>VIT_212s0035g01910.1</i>	ER	250	28.39	7.94	49.940	89.400	-0.388
VvsHSP23	<i>VIT_213s0019g00860.1</i>	PX	142	15.81	6.75	51.850	85.000	-0.311
VvsHSP24	<i>VIT_213s0019g02740.1</i>	CI	151	17.17	5.81	46.350	74.110	-0.553
VvsHSP25	<i>VIT_213s0019g02760.1</i>	CI	140	15.80	6.77	42.530	72.290	-0.589
VvsHSP26	<i>VIT_213s0019g02770.1</i>	CI	151	17.10	5.81	48.460	72.850	-0.554
VvsHSP27	<i>VIT_213s0019g02780.1</i>	CI	151	17.02	5.80	45.690	70.260	-0.587
VvsHSP28	<i>VIT_213s0019g02820.1</i>	CI	151	17.12	5.81	47.080	70.260	-0.591
VvsHSP29	<i>VIT_213s0019g02840.1</i>	CI	151	17.09	5.54	50.220	72.850	-0.551
VvsHSP30	<i>VIT_213s0019g02850.1</i>	CI	151	17.05	5.80	45.440	70.260	-0.590
VvsHSP31	<i>VIT_213s0019g02900.1</i>	CI	108	12.64	6.18	52.990	80.190	-0.798
VvsHSP32	<i>VIT_213s0019g02920.1</i>	CI	136	15.27	5.70	46.020	79.490	-0.552
VvsHSP33	<i>VIT_213s0019g02930.1</i>	CI	160	18.17	6.78	53.430	76.690	-0.644
VvsHSP34	<i>VIT_213s0019g03000.1</i>	CI	160	18.15	7.93	55.700	76.690	-0.644

表1 (续)

Table 1 (continued)

Protein name	Locus name	Subfamily	Protein length (aa)	Molecular weight (kDa)	Theoretical pI	Instability index	Aliphatic index	GRAVY
VvsHSP35	VIT_213s0019g03010.1	CI	144	16.37	9.21	49.720	69.030	-0.785
VvsHSP36	VIT_213s0019g03030.1	CI	108	12.67	6.76	37.260	77.500	-0.817
VvsHSP37	VIT_213s0019g03050.1	CI	165	19.23	6.46	61.350	79.090	-0.721
VvsHSP38	VIT_213s0019g03090.1	CI	160	18.17	5.43	62.670	73.690	-0.673
VvsHSP39	VIT_213s0019g03160.1	CI	160	18.02	7.94	64.610	73.120	-0.631
VvsHSP40	VIT_213s0019g03170.1	CI	159	18.19	6.17	51.140	71.010	-0.655
VvsHSP41	VIT_216s0022g00510.1	MTI/CP	208	23.74	5.61	60.250	75.870	-0.721
VvsHSP42	VIT_216s0098g01060.1	CP	227	25.03	6.35	44.280	72.600	-0.505
VvsHSP43	VIT_218s0001g01570.1	CVI	163	18.28	6.33	48.000	71.720	-0.790
VvsHSP44	VIT_218s0001g01610.1	CVI	159	18.00	5.74	44.410	72.890	-0.731
VvsHSP45	VIT_218s0089g01270.1	ER	186	21.13	5.89	35.370	90.590	-0.567
VvsHSP46	VIT_219s0014g05050.1	CV	192	22.39	5.35	48.680	70.470	-0.754
VvsHSP47	VIT_219s0085g01050.1	CVII	146	16.45	5.90	50.360	72.740	-0.530

GRAVY: 总平均疏水指数 GRAVY: Grand average of hydropathicity

适的氨基酸替换模型。根据前人的分类标准(Waters and Vierling, 2020), 将sHSP家族进一步分为12个亚家族, 即CI-CVII、CP、ER、MTI/CP、MTII和PX亚家族。在睡莲、水稻、拟南芥和葡萄中, CI亚家族包含的sHSP成员均最多, 其中含有9个NcsHSP成员、10个OssHSP成员、6个AtsHSP成员和19个VvsHSP成员。在水稻和拟南芥中没有发现属于CVI亚家族的成员, 在睡莲和拟南芥中没有发现属于CVII亚家族的成员。此外, 在睡莲中, ER亚家族包含9个NcsHSP成员, CII亚家族包含5个NcsHSP成员, 其余亚家族包含1-2个NcsHSP成员。在葡萄中, CII亚家族包含12个VvsHSP成员, MTI/CP亚家族包含3个VvsHSP成员, 其余亚家族包含1-2个VvsHSP成员。在水稻和拟南芥中, 其余亚家族均包含1-2个sHSP成员。

2.3 4种植物中sHSP成员的保守基序和基因结构分析

利用MEME程序在睡莲、水稻、拟南芥和葡萄sHSP成员中共鉴定出20个保守基序(图2)。进一步利用CDD程序进行注释, 结果只有3个基序(motif1、motif3和motif5)注释为sHSP家族的保守结构域。利用MEME程序鉴定到20个保守基序的分布, 可以将睡莲、水稻、拟南芥和葡萄中sHSP成员分为类似于系统进化分析中的12个亚家族。Motif2存在于几乎所有

sHSP成员中, 其它motif存在于特定的亚家族中。例如, motif1主要存在于除CP、CV、MTI/CP和MTII亚家族外的其它亚家族中, motif3主要存在于除CP、MTI/CP和MTII亚家族外的其它亚家族中, motif4主要存在于除CIV、CP、CV、MTI/CP和MTII亚家族外的其它亚家族中, motif12和motif15主要存在于CP、MTI/CP和MTII亚家族中, motif17仅存在于CV亚家族中。

对睡莲、水稻、拟南芥和葡萄中sHSP成员基因结构进行分析, 根据外显子和内含子的数量, 可以将121个sHSP成员的基因结构分为3种不同类型(图2)。第1种类型的基因结构包含63个sHSP成员, 它们均只包含1个外显子, 无任何内含子。第2种类型的基因结构包含49个sHSP成员, 它们均包含1个内含子和2个外显子。其余sHSP成员均属于第3种类型的基因结构, 其中NcsHSP04包含的外显子(18个)和内含子(17个)最多。不同sHSP亚家族的基因结构具有一定的保守性。例如, CVII亚家族只包含第1种类型的基因结构, CIII、CVI、MTI/CP和MTII亚家族只包含第2种类型的基因结构。此外, 来源于不同物种的sHSP成员具有不同的基因结构。相比水稻、拟南芥和葡萄的sHSP成员, 睡莲NcsHSP成员的基因结构包含的外显子和内含子更多, 第3种基因结构类型的许多sHSP成员来源于睡莲。另外3个物种的sHSP成员的基因结构无显著区别。

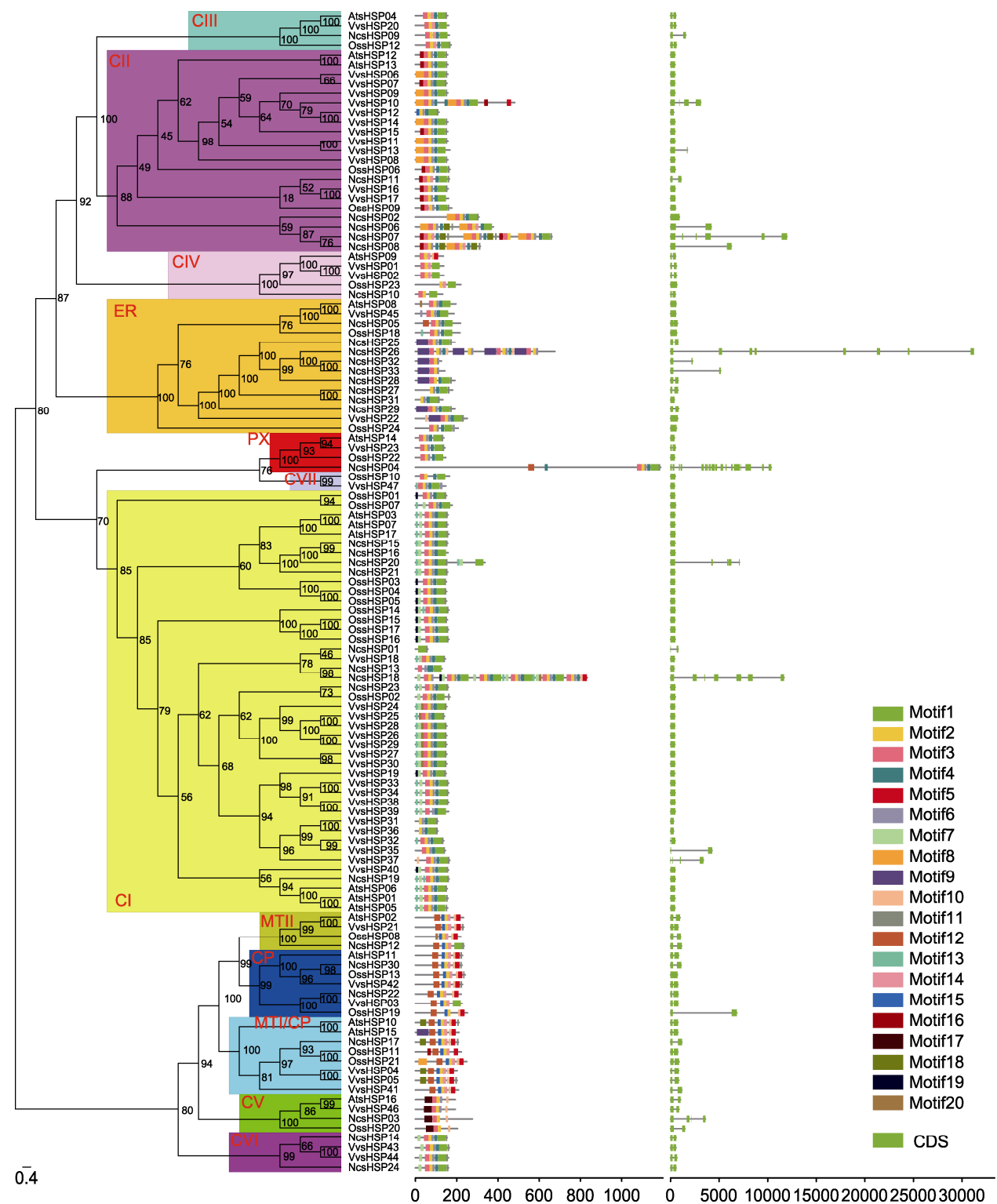


图2 睡莲、水稻、拟南芥和葡萄中sHSP家族的系统进化(左)、保守基序(中)和基因结构(右)分析
通过IQ-tree软件构建系统进化树(左), 分支上的数字代表bootstrap值, sHSP亚家族用红色字母表示。保守基序(中)通过MEME程序识别, 不同颜色代表不同的保守基序, 保守基序的位置使用下方标尺进行估计。根据内含子和外显子的位置构建基因结构示意图(右), 绿色方框表示外显子, 黑色线表示内含子, 内含子和外显子的位置使用下方标尺进行估计。

Figure 2 The phylogenetic analysis (left), conserved motif (middle), and gene structure (right) of the sHSP family in waterlily,

rice, *Arabidopsis* and grape

The phylogenetic tree (left) was constructed by IQ-tree software. The numbers in the clades stood for the bootstrap values, and sHSP subfamilies were indicated by red letters. The conserved motifs (middle) were identified by MEME software. Each conserved motif was marked by a specific color, and the location of motifs can be estimated using the scale at the bottom. The gene structures (right) were visualized according to the location of exons and introns. The exons and introns were represented by green box and black line, respectively. The location of exons and introns can be estimated using the scale at the bottom.

2.4 4种植物中sHSP成员的染色体定位

我们进一步对睡莲、水稻、拟南芥和葡萄中sHSP成员的染色体定位进行分析(图3)。在睡莲中, *NcsHSP*成员定位在12条染色体上, 其中Nc07和Nc13上并不存在*NcsHSP*成员, Nc09和Nc11上包含*NcsHSP*成员最多(6个)(图3A)。在水稻中, *OssHSP*成员定位在8条染色体上, 其中Os08、Os09、Os10和Os12上不存在*OssHSP*成员, Os01和Os02上包含*OssHSP*成员最多(6个)(图3B)。在拟南芥中, *AtsHSP*成员定位于拟南芥所有染色体上, 其中At5包含的*AtsHSP*成员最多(6个)(图3C)。在葡萄中, *VvsHSP*成员定位在11条染色体上, 其中Vv03、Vv05、Vv07、Vv10、Vv11、Vv14、Vv15和Vv17上不存在*VvsHSP*成员, Vv13上包含的*VvsHSP*成员最多(18个)(图3D)。此外, 一些sHSP成员成簇存在, 主要集中在睡莲Nc09的下半部以及Nc11的上半部、水稻Os01和Os03的上半部、拟南芥At1的下半部以及葡萄Vv04和Vv13的上半部。

2.5 4种植物中sHSP成员的基因复制事件分析

在睡莲、水稻、拟南芥和葡萄的sHSP成员中共鉴定到12个基因复制事件, 其中片段复制事件和串联复制事件各6个(表2)。在睡莲中和水稻中片段复制事件和串联复制事件均各有1个; 在拟南芥中片段复制事件有3个, 串联复制事件有1个; 在葡萄中片段复制事件有1个, 串联复制事件有3个(*VvsHSP43*和*VvsHSP44*定位于scaffold上)(图3)。同时, 对这12个基因复制事件进行 K_a 和 K_s 计算, 发现在每个物种中, 片段复制事件的 K_s 值均大于串联复制事件的 K_s 值, 所有基因复制事件的 K_a/K_s 比值均小于0.9, 并且拟南芥中所有3个片段复制事件以及葡萄中仅有的1个片段复制事件的 K_s 值大于1。此外, 在sHSP家族中, 基因复制事件发生在CI、CII、CVI和MTI/CP亚家族中。在CVI亚家族中, 睡莲发生了1个片段复制事件, 葡萄发生了1个串联复制事件。在CII亚家族中, 睡莲、拟南芥和葡萄各发生1个串联复制事件。在MTI/CP亚家

族中, 水稻和拟南芥各发生1个片段复制事件, 葡萄中发生1个串联复制事件。在CI亚家族中, 水稻发生1个串联复制事件, 葡萄发生1个片段复制事件, 拟南芥发生2个片段复制事件。

2.6 4种植物中sHSP成员之间的共线性分析

对睡莲、水稻、拟南芥和葡萄中sHSP家族之间的共线性关系进行分析, 并鉴定到相应的同源基因(图4)。在睡莲中, 其与水稻以及拟南芥之间分别鉴定到3对同源基因, 而与葡萄之间共鉴定到8对同源基因, 其中CI亚家族的*NcsHSP20*可同时在这3种植物中发现同源基因, CV亚家族的*NcsHSP03*可在拟南芥和葡萄中发现同源基因, MTII亚家族的*NcsHSP12*和CI亚家族的*NcsHSP23*可在水稻和葡萄中发现同源基因, 而CP亚家族的*NcsHSP30*和CIII亚家族的*NcsHSP09*仅在葡萄中发现同源基因(图4A)。在水稻中, 其与拟南芥和葡萄之间分别鉴定到4对同源基因, 共涉及5个*OssHSP*, 其中CI亚家族的*OssHSP14*可在拟南芥和葡萄中同时发现同源基因, CV亚家族的*OssHSP20*和PX亚家族的*OssHSP22*可在拟南芥中发现同源基因, CIII亚家族的*OssHSP12*和CIV亚家族的*OssHSP23*可在葡萄中发现同源基因(图4B)。在拟南芥和葡萄中共发现10对同源基因, 其中有2对同源基因来自CI亚家族(*AtsHSP01/VvsHSP18*和*AtsHSP06/VvsHSP18*), 另外2对同源基因来自MTII/CP亚家族(*AtsHSP10/VvsHSP41*和*AtsHSP15/VvsHSP41*), 而在CII、CIII、CV、CP、ER和MTII亚家族中各存在1对同源基因(图4C)。随后, 对这些同源基因对的 K_a 值进行计算, 发现在不同物种中彼此之间并无显著差异(图4D)。此外, 对这些同源基因对的 K_s 值进行计算, 发现有22对sHSP同源基因的 K_s 值由于序列高度变异无法计算。

2.7 4种植物中sHSP成员在不同组织中的表达模式

本研究对睡莲、水稻、拟南芥和葡萄中sHSP成员在

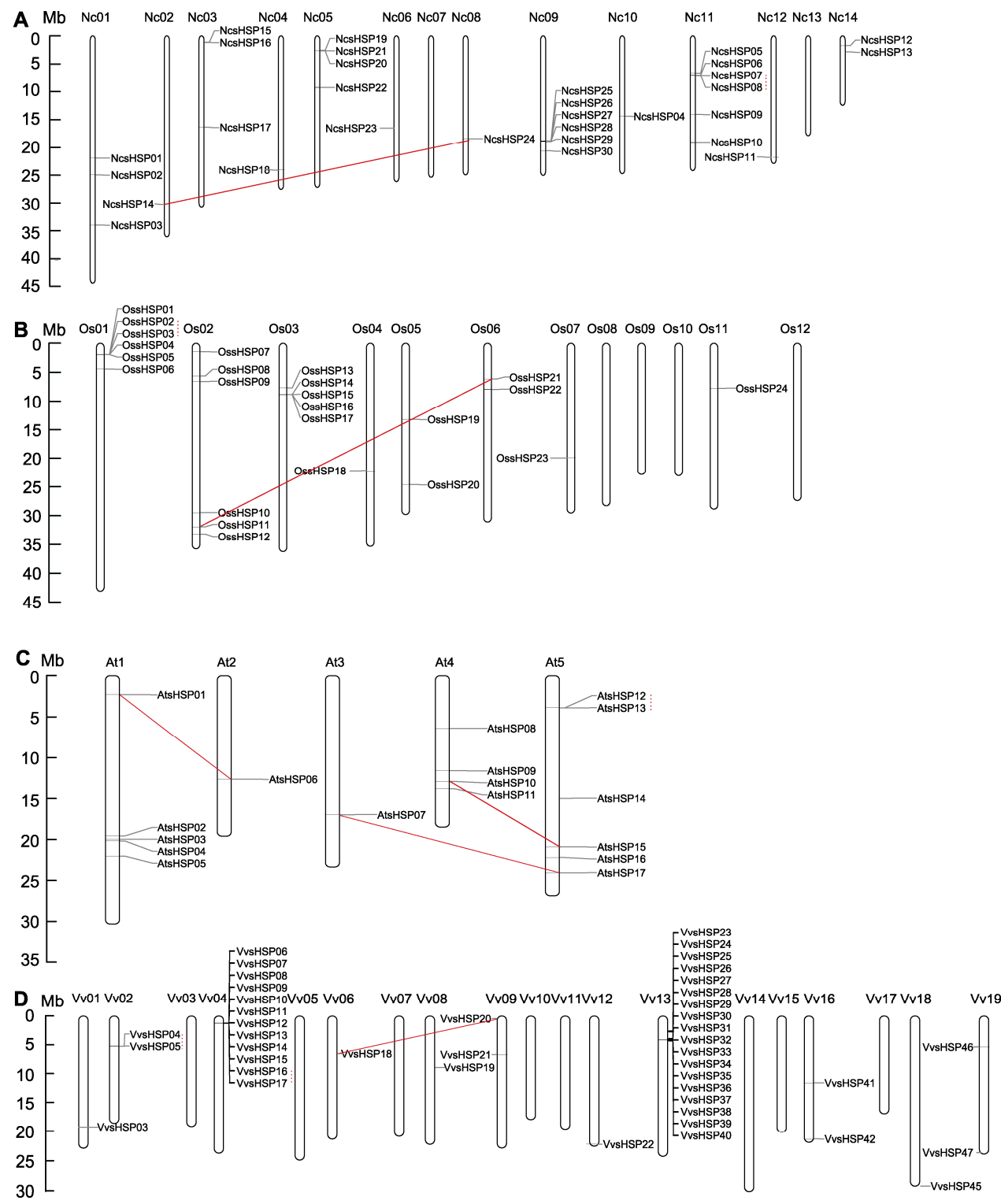


图3 睡莲、水稻、拟南芥和葡萄中*NcsHSP* (A)、*OssHSP* (B)、*AtsHSP* (C)和*VvsHSP*成员(D)的染色体分布
实线连线和虚线连线分别代表片段复制事件和串联复制事件。

Figure 3 Chromosomal locations of *NcsHSP* (A), *OssHSP* (B), *AtsHSP* (C) and *VvsHSP* (D) on the waterlily, rice, *Arabidopsis* and grape chromosomes
The solid lines and dot lines represented the segmental and tandem duplication events, respectively.

表2 在睡莲、水稻、拟南芥和葡萄中复制的sHSP成员的 K_a 和 K_s 分析

Table 2 K_a and K_s analysis for the duplicated sHSP members in waterlily, rice, *Arabidopsis* and grape

Duplicated gene 1	Duplicated gene 2	Subfamily	K_a	K_s	K_a/K_s	Purifying selection	Duplicated type
<i>NcsHSP07</i>	<i>NcsHSP08</i>	CII	0.067786	0.358961	0.188839	Yes	Tandem
<i>NcsHSP14</i>	<i>NcsHSP24</i>	CVI	0.346841	0.789231	0.439467	Yes	Segmental
<i>OssHSP03</i>	<i>OssHSP04</i>	CI	0.02811	0.129535	0.217006	Yes	Tandem
<i>OssHSP11</i>	<i>OssHSP21</i>	MTI/CP	0.303273	0.454245	0.66764	Yes	Segmental
<i>AtsHSP01</i>	<i>AtsHSP06</i>	CI	0.08714	1.101078	0.079141	Yes	Segmental
<i>AtsHSP07</i>	<i>AtsHSP17</i>	CI	0.110454	1.291145	0.085547	Yes	Segmental
<i>AtsHSP10</i>	<i>AtsHSP15</i>	MTI/CP	0.175047	1.17229	0.14932	Yes	Segmental
<i>AtsHSP12</i>	<i>AtsHSP13</i>	CII	0.060844	0.53554	0.113613	Yes	Tandem
<i>VvsHSP04</i>	<i>VvsHSP05</i>	MTI/CP	0.077421	0.10198	0.759181	Yes	Tandem
<i>VvsHSP16</i>	<i>VvsHSP17</i>	CII	0.212704	0.880614	0.241541	Yes	Tandem
<i>VvsHSP18</i>	<i>VvsHSP40</i>	CI	0.206071	1.642523	0.12546	Yes	Segmental
<i>VvsHSP43</i>	<i>VvsHSP44</i>	CVI	0.21995	0.81194	0.270894	Yes	Tandem

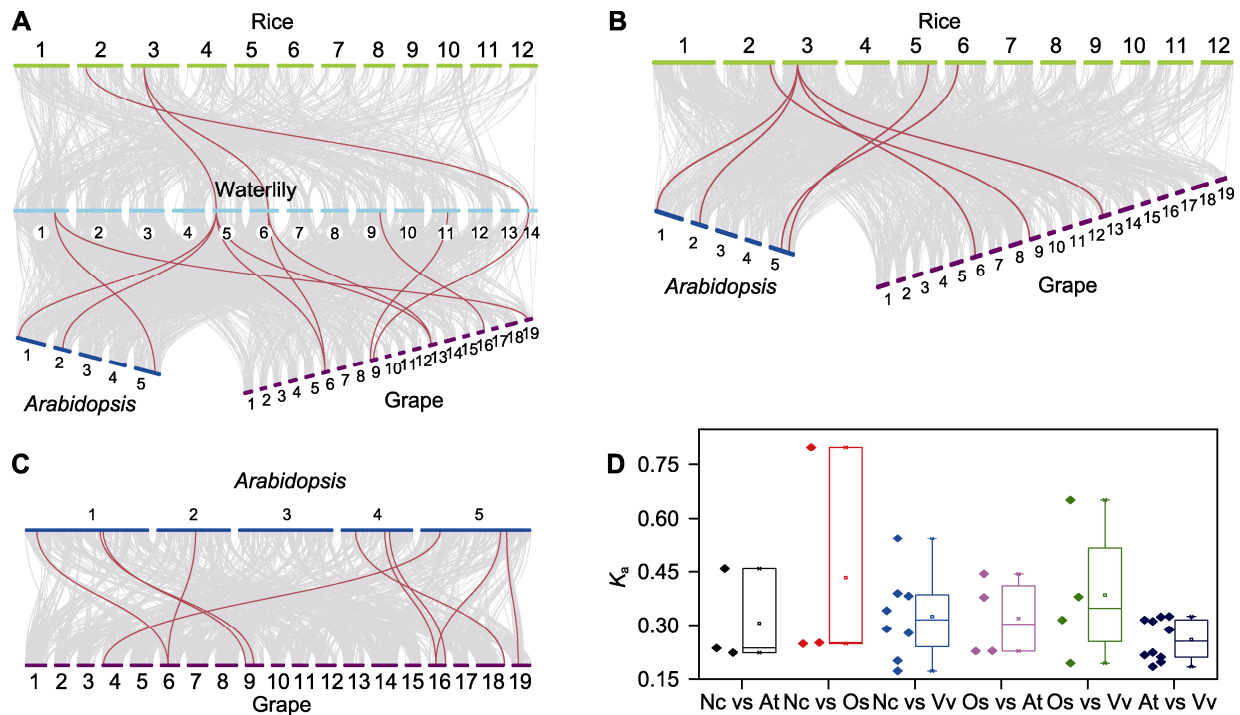


图4 睡莲、水稻、拟南芥和葡萄中sHSP成员的共线性分析

(A) 睡莲与水稻、拟南芥和葡萄中sHSP成员的共线性分析; (B) 水稻与拟南芥和葡萄sHSP成员的共线性分析; (C) 拟南芥与葡萄sHSP成员的共线性分析; (D) 在不同物种中sHSP同源基因对的 K_a 值分析。(A)–(C) 灰线代表所有的同源基因, 红线代表sHSP同源基因。

Figure 4 Synteny analysis of sHSP members in waterlily, rice, *Arabidopsis* and grape

(A) Synteny analysis of sHSP members between waterlily and rice, *Arabidopsis* and grape; (B) Synteny analysis of sHSP members between rice and *Arabidopsis* and grape; (C) Synteny analysis of sHSP members between *Arabidopsis* and grape; (D) Comparison of K_a values of orthologous sHSPs between different species. (A)–(C) Gray lines connected orthologous genes, and orthologous sHSPs were marked with red lines.

不同组织中的表达模式进行了分析(图5)。在睡莲中, *NcsHSP* 成员的表达谱包括3种类型: 第1种类型有26个 *NcsHSPs*, 它们在大多数组织中不表达或者仅在个别组织中表达。例如, *NcsHSP30* 在幼小的叶柄、花和雄蕊中表达, *NcsHSP10* 主要在叶柄中表达。第2种类型为在所有组织中均高度表达的 *NcsHSP19*。第3种类型为在所有组织中中度表达的6个 *NcsHSPs* (*NcsHSP02*、*NcsHSP03*、*NcsHSP04*、*NcsHSP17*、*NcsHSP22* 和 *NcsHSP29*) (图5A)。在拟南芥中, *AtsHSP* 成员的表达谱包括3种类型: 第1种类型为在所有组织中均较低表达的 *AtsHSP02* 和 *AtsHSP11*; 第2种类型为在所有组织中均较高表达的 *AtsHSP04* 和 *AtsHSP09*; 其余 *AtsHSP* 成员均为第3种类型, 主要在花瓣中具有较高的表达水平(图5B)。在水稻中, *OssHSP* 成员的表达谱也包括3种类型: 第1种类型为在所有组织中均较低表达的10个 *OssHSPs*; 第2种类型为主要在花器官相关组织中有较高表达的 *OssHSP03*、*OssHSP04*、*OssHSP06*、*OssHSP12*、*OssHSP16* 和 *OssHSP21*; 第3种类型为在所有组织中均较高表达的 *OssHSP02*、*OssHSP11*、*OssHSP14* 和 *OssHSP23* (图5C)。在葡萄中, *VvsHSP* 成员的表达谱也包括3种类型: 第1种类型为在所有组织中均较高表达的 *VvsHSP40*; 第2种类型为主要在浆果发育过程中表达的18个 *VvsHSPs*; 其余 *VvsHSPs* 在大多数组织中不表达或者仅在个别组织中表达。例如, *VvsHSP20* 主要在芽的发育阶段大量表达(图5D)。

3 讨论

本研究在睡莲、水稻、拟南芥和葡萄中分别鉴定到33个 *NcsHSPs*、24个 *OssHSPs*、17个 *AtsHSPs* 和47个 *VvsHSPs* (表1)。4种植物中, 在葡萄中鉴定到的 *sHSP* 成员最多, 而在拟南芥中鉴定到的 *sHSP* 成员最少, 这主要与4种植物基因组大小有关(The French-Italian Public Consortium for Grapevine Genome Characterization, 2007; Ouyang et al., 2007; Lamesch et al., 2012; Zhang et al., 2020a)。通过系统进化分析, 我们将 *sHSP* 家族进一步分为12个亚家族(图1)。不同的亚家族含有不同数目的 *sHSP* 成员, 其中CI亚家族包含最多的 *sHSP* 成员, 这可能与CI亚家族包含最多的基因复制事件(4个)有关(表2)。相似的分布也出现在番茄(张宁等, 2017)、马铃

薯(Zhao et al., 2018)、棉花(Ma et al., 2016)、银杏(何福林和张斌, 2019)和辣椒(Guo et al., 2015)等植物中。因此, 在植物中 *sHSP* 家族具有保守的亚家族分布。在拟南芥中未发现任何属于CVI和CVII亚家族的成员, 而在葡萄中CI和CII亚家族的成员数目远多于另外3个物种, 这直接导致拟南芥和葡萄中含有不同数目的 *sHSP* 成员。类似现象也出现在棉花扩展蛋白基因家族(张奇艳等, 2019)、玉米(*Zea mays*) NAC家族(Fan et al., 2014)及大豆PP2C家族(Fan et al., 2020a)。*sHSP* 家族的保守基序分布和基因结构分布也与通过系统进化分析得出的亚家族分布一致(图2)。

在睡莲、水稻、拟南芥和葡萄中共鉴定到12个与 *sHSP* 家族有关的基因复制事件(表2)。相比其它2种植物, 在葡萄和拟南芥中发生的基因复制事件最多(各4个)。葡萄中 *sHSP* 成员数量最多可能与葡萄发生的基因复制事件最多有关, 相似的研究结果也出现在葡萄的CDPL家族(Zhang et al., 2015)、卵形家族蛋白(Wang et al., 2018)以及WRKY家族中(Guo et al., 2014)。在拟南芥中, 虽然与 *AtsHSP* 成员相关的基因复制事件最多, 但成员数量较少, 这可能与拟南芥基因复制事件后大量基因丢失有关(Cannon et al., 2004; Maere et al., 2005)。结合染色体定位结果, 可以将12个与 *sHSP* 家族有关的基因复制事件进一步分为6个片段复制事件和6个串联复制事件, 并且每个物种中均含有这2种基因复制事件, 这表明 *sHSP* 家族的扩增不仅与大规模的片段复制事件有关, 还与小规模串联复制事件有关(图3)。相似的 *sHSP* 家族扩增模式也出现在辣椒(Guo et al., 2015)、簸箕柳(Li et al., 2018)和马铃薯(Zhao et al., 2018)等植物中。在每个物种中, 片段复制事件的 K_a 值均大于串联复制事件的 K_s 值, 这表明在 *sHSP* 家族中片段复制事件发生的时间均早于串联复制事件, 相似的现象也出现在梅花(*Prunus mume*) NAC家族(Zhuo et al., 2018)和棉花PP2C家族(Shazadee et al., 2019)中。基于复制事件的 K_a/K_s 比值表明, *sHSP* 成员在发生复制事件以后会发生强烈的纯化选择作用, 类似的 K_a/K_s 比值分布也出现在其它植物的 *sHSP* 家族中(Ma et al., 2016; Li et al., 2018)。与此同时, 与其它基因家族相比(Zhuo et al., 2018; Fan et al., 2020b), *sHSP* 家族发生的基因复制事件较少(表2), 这也直接导致 *sHSP* 家族成员数相对较少, 表明在植物的进化过程中, *sHSP* 成

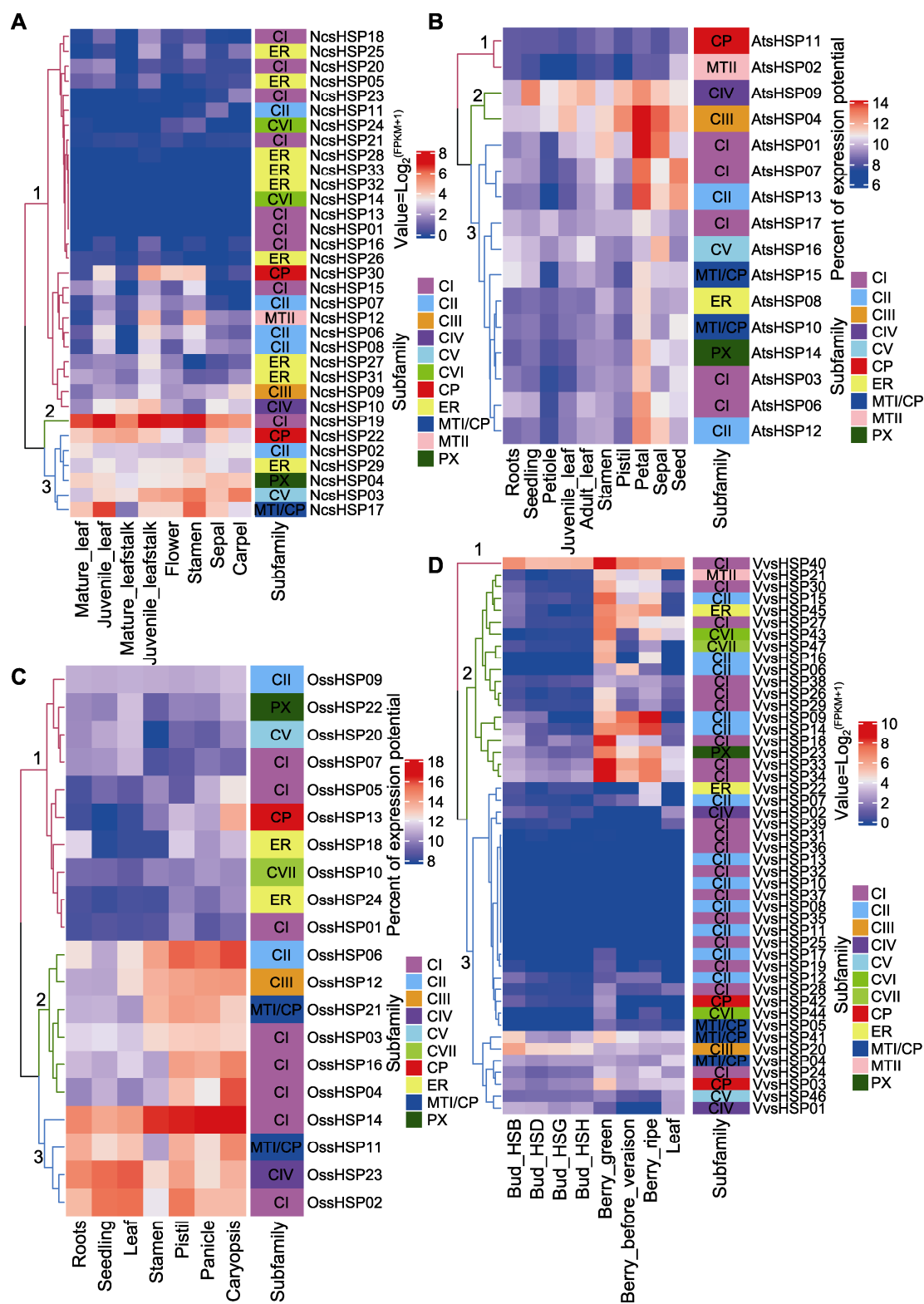


图5 *NcsHSP* (A)、*AtsHSP* (B)、*OssHSP* (C)和 *VvsHSP* (D)在不同组织中的表达分析
两种颜色的标识分别代表表达值和sHSP亚家族。

Figure 5 Expression profiles of *NcsHSP* (A), *AtsHSP* (B), *OssHSP* (C) and *VvsHSP* (D) in different tissues. Two color bars represented expression level and sHSP subfamilies, respectively.

员可能具有相对保守的功能(如响应高温胁迫) (Waters and Vierling, 2020)。此外, 本研究发现4个片段复制事件的 K_s 值大于1 (表2), 其中在拟南芥中有1个片段复制事件(*AtsHSP01/AtsHSP06*), 在另外3种植物中均发现对应的同源基因(图4), 该片段复制事件发生的时间大约为1–1.2亿年前温暖干旱的白垩纪(Zhang et al., 2020b), 这表明*AtsHSP01/AtsHSP06*及其在植物界的同源基因可能与植物适应高温环境有关, 可以作为今后研究高温胁迫的重要候选基因。

之前的研究表明, sHSP家族可以参与许多植物生长与发育过程(Zhang et al., 2018; Ma et al., 2019)。本研究在睡莲、水稻、拟南芥和葡萄中新发现一些与植物生长发育过程相关的sHSP成员(图5)。CI亚家族的*NcsHSP19*、*OssHSP02*、*OssHSP14*和*VvsHSP40*以及CIV亚家族的*AtsHSP09*和*OssHSP23*在几乎所有组织中均高度表达, 其在棉花和马铃薯中的同源基因也具有类似的表达模式(Ma et al., 2016; Zhao et al., 2018), 这表明sHSP家族在植物中的功能可能具有一定的保守性。此外, 本研究还发现一些与水稻花器官发育相关的基因(*OssHSP03*、*OssHSP04*、*OssHSP06*、*OssHSP12*、*OssHSP16*和*OssHSP21*)。有报道显示, 在玫瑰(*Rosa hybrida*)中与CI亚家族*OssHSP03*、*OssHSP04*和*OssHSP16*相同亚家族的*sHSP17.5-CI*可能与开花过程中保护细胞蛋白有关(Dafny-Yelin et al., 2008)。

对睡莲、水稻、拟南芥和葡萄中sHSP家族之间的共线性分析(图4)表明, 拟南芥和葡萄中的同源基因对最多(10对), 这主要与拟南芥和葡萄之间具有较近的进化关系有关(Jiao et al., 2012)。虽然睡莲(基部被子植物)和葡萄(双子叶植物)具有较远的进化距离, 但其sHSP成员却有较多的同源基因对(8对), 这主要与葡萄基因组并未发生最近的基因复制事件从而导致基因组有很多古代植物基因组的特征有关(The French-Italian Public Consortium for Grapevine Genome Characterization, 2007; Zhang et al., 2020a)。此外, 本研究通过 K_s 值计算发现同源基因之间的进化同步, 即在植物中sHSP家族的进化相对保守, 这可能与sHSP家族的功能有一定关系(Waters and Vierling, 2020)。

睡莲是现存的早期被子植物之一(Zhang et al., 2020a), 通过与另外3种模式被子植物的sHSP家族

进行对比, 发现睡莲sHSP家族中氨基酸长度较长以及分子量较大的成员较多(表1), 其次为具有很多古代植物基因组特征的葡萄sHSP成员(The French-Italian Public Consortium for Grapevine Genome Characterization, 2007), 表明sHSP家族在被子植物中可能向更短的氨基酸长度和更小的分子量进化, 这种进化趋势与已报道的rhomboid家族不同(Li et al., 2015), 但与PG家族类似(Park and Kim, 2010)。其次, 水稻、拟南芥和葡萄大部分sHSP成员的内含子不多于1个(图2), 而睡莲含有较多基因结构较为复杂的sHSP成员, 这表明sHSP家族在被子植物中可能向更简单的基因结构进化, 这种趋势与已报道的NAP亚家族不同(Fan et al., 2015)。此外, 睡莲sHSP成员分布在12条染色体上(图3), 葡萄sHSP成员分布在11条染色体上, 而拟南芥和水稻的sHSP成员分别分布在5条和8条染色体上, 表明sHSP成员在被子植物中可能更集中分布于某几条染色体上, 这种集中分布模式可能更有利于sHSP成员协同发挥作用。

综上所述, 本研究在睡莲、水稻、拟南芥和葡萄中分别鉴定到33个*NcsHSPs*、24个*OssHSPs*、17个*AtsHSPs*和47个*VvsHSPs*。上述sHSP成员可以进一步分为12个亚家族, 不同的亚家族含有不同的成员数目、基序分布、基因结构以及复制基因数目。sHSP家族的扩增与片段复制事件和串联复制事件均有关, 同时, sHSP家族在被子植物中可能朝着更短的氨基酸长度、更小的分子量、更简单的基因结构以及更集中的染色体分布进化, 这种进化模式可能与sHSP家族相关的热激转录因子家族进化有关(Zhu et al., 2016)。此外, 在睡莲、水稻、拟南芥和葡萄中还鉴定到一些可能参与调控植物生长发育的候选基因。

参考文献

- 何福林, 张斌 (2019). 银杏(*Ginkgo biloba*) GbHsp20基因家族的鉴定及系统进化分析. 分子植物育种 17, 7368–7376.
- 黄祥富, 黄上志, 傅家瑞 (1999). 植物热激蛋白的功能及其基因表达的调控. 植物学通报 16, 530–536.
- 栗振义, 龙瑞才, 张铁军, 杨青川, 康俊梅 (2016). 植物热激蛋白研究进展. 生物技术通报 32(2), 7–13.
- 刘德立 (1996). 植物热激蛋白及其功能. 植物学通报 13, 14–19.
- 张宁, 姜晶, 史洁玮 (2017). 番茄HSP20基因家族的全基因

- 组鉴定、系统进化及表达分析. 沈阳农业大学学报 **48**, 137–144.
- 张奇艳, 雷忠萍, 宋银, 海江波, 贺道华 (2019). 陆地棉扩展蛋白基因的鉴定与特征分析. 中国农业科学 **52**, 3713–3732.
- Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* **4**, 10.
- Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* **13**, 1194–1202.
- Dafny-Yelin M, Tzfira T, Vainstein A, Adam Z (2008). Non-redundant functions of sHSP-CIs in acquired thermotolerance and their role in early seed development in *Arabidopsis*. *Plant Mol Biol* **67**, 363–373.
- Fan K, Chen YR, Mao ZJ, Fang Y, Li ZW, Lin WW, Zhang YQ, Liu JP, Huang JW, Lin WX (2020a). Pervasive duplication, biased molecular evolution and comprehensive functional analysis of the PP2C family in *Glycine max*. *BMC Genomics* **21**, 465.
- Fan K, Mao ZJ, Zheng JX, Chen YR, Li ZW, Lin WW, Zhang YQ, Huang JW, Lin WX (2020b). Molecular evolution and expansion of the KUP family in the allopolyploid cotton species *Gossypium hirsutum* and *Gossypium barbadense*. *Front Plant Sci* **11**, 545042.
- Fan K, Shen H, Bibi N, Li F, Yuan SN, Wang M, Wang XD (2015). Molecular evolution and species-specific expansion of the NAP members in plants. *J Integr Plant Biol* **57**, 673–687.
- Fan K, Wang M, Miao Y, Ni M, Bibi N, Yuan SN, Li F, Wang XD (2014). Molecular evolution and expansion analysis of the NAC transcription factor in *Zea mays*. *PLoS One* **9**, e111837.
- Guo CL, Guo RR, Xu XZ, Gao M, Li XQ, Song JY, Zheng Y, Wang XP (2014). Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family. *J Exp Bot* **6**, 1513–1528.
- Guo LM, Li J, He J, Liu H, Zhang HM (2020). A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Sci Rep* **10**, 1383.
- Guo M, Liu JH, Lu JP, Zhai YF, Wang H, Gong ZH, Wang SB, Lu MH (2015). Genome-wide analysis of the *Ca-Hsp20* gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Front Plant Sci* **6**, 806.
- Ji XR, Yu YH, Ni PY, Zhang GH, Guo DL (2019). Genome-wide identification of small heat-shock protein (HSP20) gene family in grape and expression profile during berry development. *BMC Plant Biol* **19**, 433.
- Jiao YN, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain M, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, Wu XL, Zhang Y, Wang J, Zhang YT, Carpenter EJ, Deyholos MK, Kutchan TM, Chanderbali AS, Soltis PS, Stevenson DW, McCombie R, Pires JC, Wong GKS, Soltis DE, DePamphilis DW (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biol* **13**, R3.
- Key JL, Lin CY, Chen YM (1981). Heat shock proteins of higher plants. *Proc Natl Acad Sci USA* **78**, 3526–3530.
- Lamesch P, Berardini TZ, Li DH, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* **40**, D1202–D1210.
- Li JB, Zhang J, Jia HX, Li Y, Xu XD, Wang LJ, Lu MZ (2016). The *Populus trichocarpa* PtHSP17.8 involved in heat and salt stress tolerances. *Plant Cell Rep* **35**, 1587–1599.
- Li JB, Zhang J, Jia HX, Yue ZQ, Lu MZ, Xin XB, Hu JJ (2018). Genome-wide characterization of the *sHsp* gene family in *Salix suchowensis* reveals its functions under different abiotic stresses. *Int J Mol Sci* **19**, 3246.
- Li Q, Zhang N, Zhang LS, Ma H (2015). Differential evolution of members of the *rhomboid* gene family with conservative and divergent patterns. *New Phytol* **206**, 368–380.
- Ma W, Guan XY, Li J, Pan RH, Wang LY, Liu FJ, Ma HY, Zhu SJ, Hu J, Ruan YL, Chen XY, Zhang TZ (2019). Mitochondrial small heat shock protein mediates seed germination via thermal sensing. *Proc Natl Acad Sci USA* **116**, 4716–4721.
- Ma W, Zhao T, Li J, Liu BL, Fang L, Hu Y, Zhang TZ (2016). Identification and characterization of the *GhHsp20* gene family in *Gossypium hirsutum*. *Sci Rep* **6**, 32517.
- Maere S, de Bodt S, Raes J, Casneuf T, van Montagu M, Kuiper M, van de Peer Y (2005). Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* **102**, 5454–5459.
- Ouyang S, Zhu W, Hamilton J, Lin HN, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng

- L, Orvis J, Haas B, Wortman J, Buell CR (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* **35**, D883–D887.
- Park KC, Kim NS (2010). Intron loss mediated structural dynamics and functional differentiation of the polygalacturonase gene family in land plants. *Genes Genom* **32**, 570–577.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**, 1650–1667.
- Sarkar NK, Kim YK, Grover A (2009). Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics* **10**, 393.
- Scharf KD, Siddique M, Vierling E (2001). The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α -crystallin domains (Acid proteins). *Cell Stress Chaperones* **6**, 225–237.
- Shazadee H, Khan N, Wang JJ, Wang CC, Zeng JG, Huang ZY, Wang XY (2019). Identification and expression profiling of protein phosphatases (PP2C) gene family in *Gossypium hirsutum* L. *Int J Mol Sci* **20**, 1395.
- Sun WN, Bernard C, Van De Cotte B, Van Montagu M, Verbruggen N (2010). *At-HSP17.6A*, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression. *Plant J* **27**, 407–415.
- The French-Italian Public Consortium for Grapevine Genome Characterization (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467.
- van Montfort RLM, Basha E, Friedrich KL, Slingsby C, Vierling E (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol* **8**, 1025–1030.
- Wang L, Zhang SL, Zhang XM, Hu XY, Guo CL, Wang XP, Song JY (2018). Evolutionary and expression analysis of *Vitis vinifera* OFP gene family. *Plant Syst Evol* **304**, 995–1008.
- Wang WX, Vinocur B, Shoseyov O, Altman A (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci* **9**, 244–252.
- Waters ER, Vierling E (2020). Plant small heat shock proteins—evolutionary and functional diversity. *New Phytol* **227**, 24–37.
- Zhai MZ, Sun YD, Jia CX, Peng SB, Liu ZX, Yang GY (2016). Over-expression of *JrsHSP17.3* gene from *Juglans regia* confer the tolerance to abnormal temperature and NaCl stresses. *J Plant Biol* **59**, 549–558.
- Zhang K, Han YT, Zhao FL, Hu Y, Gao YR, Ma YF, Zheng Y, Wang YJ, Wen YQ (2015). Genome-wide identification and expression analysis of the CDPK gene family in grape, *Vitis* spp. *BMC Plant Biol* **15**, 164.
- Zhang LS, Chen F, Zhang XT, Li Z, Zhao YY, Lohaus R, Chang XJ, Dong W, Ho SYW, Liu X, Song AX, Chen JH, Guo WL, Wang ZJ, Zhuang YY, Wang HF, Chen XQ, Hu J, Liu YH, Qin Y, Wang K, Dong SS, Liu Y, Zhang SZ, Yu XX, Wu Q, Wang LS, Yan XQ, Jiao YN, Kong HZ, Zhou XF, Yu CW, Chen YC, Li F, Wang JH, Chen W, Chen XL, Jia QD, Zhang C, Jiang YF, Zhang WB, Liu GH, Fu JY, Chen F, Ma H, van de Peer Y, Tang HB (2020a). The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79–84.
- Zhang LS, Wu SD, Chang XJ, Wang XJ, Zhao YP, Xia YP, Trigliano RN, Jiao YN, Chen F (2020b). The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant Cell Environ* **43**, 2847–2856.
- Zhang N, Shi JW, Zhao HY, Jiang J (2018). Activation of small heat shock protein (SIHSP17.7) gene by cell wall invertase inhibitor (SICIF1) gene involved in sugar metabolism in tomato. *Gene* **679**, 90–99.
- Zhao P, Wang DD, Wang RQ, Kong NN, Zhang C, Yang CH, Wu WT, Ma HL, Chen Q (2018). Genome-wide analysis of the potato *Hsp20* gene family: identification, genomic organization and expression profiles in response to heat stress. *BMC Genomics* **19**, 61.
- Zhu W, Lu MH, Gong ZH, Chen RG (2011). Cloning and expression of a small heat shock protein gene *CaHSP24* from pepper under abiotic stress. *Afr J Biotechnol* **10**, 4968–4975.
- Zhu YX, Yan HW, Wang YY, Feng L, Chen Z, Xiang Y (2016). Genome duplication and evolution of heat shock transcription factor (HSF) gene family in four model angiosperms. *J Plant Growth Regul* **35**, 903–920.
- Zhuo XK, Zheng TX, Zhang ZY, Zhang YC, Jiang LB, Ahmad S, Sun LD, Wang J, Cheng TR, Zhang QX (2018). Genome-wide analysis of the NAC transcription factor gene family reveals differential expression patterns and cold-stress responses in the woody plant *Prunus mume*. *Genes (Basel)* **9**, 494.

Comparative Genomics of the Small Heat Shock Protein Family in Angiosperms

Kai Fan, Fangting Ye, Zhijun Mao, Xinfeng Pan, Zhaowei Li, Wenxiong Lin*

Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002

Abstract Small heat shock protein (sHSP) can respond to environmental stresses and regulate plant growth and development. In this study, 33 NcsHSPs, 24 OssHSPs, 17 AtsHSPs and 47 VvsHSPs were found in waterlily, rice, *Arabidopsis*, and grape, respectively. The sHSP family can be divided into 12 subfamilies, and different subfamily had different number of sHSP members, conserved motifs, gene structures and number of duplicated genes. Then, 12 gene duplication events in the sHSP family were identified from 4 model angiosperms of sHSP family. The segmental and tandem duplication events were related to the gene expansion in the sHSP family, and the segmental duplication events occurred earlier than the tandem duplication events. Moreover, sHSP members in *Arabidopsis* and grape had the highest orthologous similarity, followed by sHSP members in waterlily and grape. Furthermore, sHSP family in angiosperms evolved towards sHSP members with smaller length of amino acid, lower molecular weight, simpler gene structure, and concentrated chromosomal localization. Besides, some sHSP members may be related to plant growth and development in waterlily, rice, *Arabidopsis*, and grape. Above all, these results could lay vital foundations on comparative genomics of the sHSP family in four model angiosperms and advance the sHSP family in other angiosperms.

Key words sHSP family, comparative genomics, expansion, gene expression

Fan K, Ye FT, Mao ZJ, Pan XF, Li ZW, Lin WX (2021). Comparative genomics of the small heat shock protein family in angiosperms. *Chin Bull Bot* **56**, 245–261.

* Author for correspondence. E-mail: lwx@fafu.edu.cn

(责任编辑: 朱亚娜)