



• 研究报告 •

转录组测序揭示翼盖蕨(*Didymochlaena trancatula*)的全基因组复制历史

汪浩^{1,2#} 张锐^{1#} 张娇¹ 沈慧¹ 戴锡玲² 严岳鸿^{1*}

1 (上海辰山植物园, 中国科学院上海辰山植物科学研究中心, 上海 201602)

2 (上海师范大学生命科学学院, 上海 200234)

摘要: 全基因组复制在动植物中普遍存在, 被认为是促进物种进化的重要动力之一。作为蕨类植物的单种科物种, 翼盖蕨(*Didymochlaena trancatula*)是真水龙骨科I的基部类群, 在蕨类中具有独特的演化地位。本研究基于高通量测序, 通过同义替换率(Ks)分析、相对定年分析揭示翼盖蕨的全基因组复制发生情况。Ks分析表明, 翼盖蕨至少经历了两次全基因组复制事件, 其中一次发生于59–62 million years ago (Mya), 另一次发生于90–94 Mya, 这两次全基因组复制事件分别和白垩纪第三纪的Cretaceous-Tertiary (C-T)大灭绝事件以及翼盖蕨的物种分化时间相吻合。进一步对两次全基因组复制保留的基因进行功能注释和富集分析, 结果显示与转录及代谢调控相关的基因优势被保留。翼盖蕨的全基因组复制事件可能促进了该物种的分化及其对极端环境的适应性。

关键词: 全基因组复制; 蕨类植物; 白垩纪大灭绝事件; 进化生物学; 同义替换率

De novo transcriptome assembly reveals the whole genome duplication events of *Didymochlaena trancatula*

Hao Wang^{1,2#}, Rui Zhang^{1#}, Jiao Zhang¹, Hui Shen¹, Xiling Dai², Yuehong Yan^{1*}

1 Shanghai Chenshan Botanical Garden, Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai 201602

2 College of Life Sciences, Shanghai Normal University, Shanghai 200234

Abstract: Whole genome duplication is an important driving force to speciation and evolution. Moreover, most existing plants and animals have experienced whole genome duplication in their evolutionary history. As the basal group of the Eupolypods I, *Didymochlaena trancatula* is the single fern species of Didymochlaenaceae. We performed transcriptome sequencing to detect whole genome duplication (WGD) events by analyzing age distributions built from synonymous substitution rates (Ks). We found that *D. trancatula* has experienced at least two WGDs during its evolutionary history. We dated the two WGDs at 59–62 million years ago (Mya) and 90–94 Mya, corresponding to Cretaceous-Tertiary (C-T) extinction event and the divergence time of *D. trancatula*, respectively. Annotation and functional enrichment analysis showed most duplicated genes that were retained are related to environmental regulation, further emphasizing the role that WGDs may play in the adaptive evolution of *D. trancatula*.

Key words: whole genome duplication; fern; the Cretaceous-Tertiary extinction event; evolutionary biology; Ks

全基因组复制(whole genome duplication, WGD)或者多倍化(polyploid)是促进维管植物物种形成和适应性演化的重要动力之一(Van de Peer et al, 2009)。植物界多数物种都经历了全基因组复制事件(Wolfe

& Shields, 1997; Soltis et al, 2009; Huang et al, 2016; Xiang et al, 2017)。研究显示, 25%的维管植物是多倍体(Barker et al, 2015), 其中15%的被子植物以及31%的蕨类植物以多倍体的形式存在(Wood et al,

收稿日期: 2019-07-25; 接受日期: 2019-12-17

基金项目: 中国科学院战略性先导科技专项(A类)(XDA19050404)和上海市绿化和市容管理局科学技术项目(G152420)

共同第一作者 Co-first authors

* 通讯作者 Author for correspondence. E-mail: yan.yh@126.com

2009; Jiao et al, 2011)。现存种子植物和被子植物分化之前各经历了一次全基因组复制事件, 分别发生在3.19亿年前和1.92亿年前(Jiao et al, 2011), 即使基因组很小的拟南芥(*Arabidopsis thaliana*)也经历了多次全基因组复制事件(Athanasios et al, 2000; Bowers et al, 2003)。全基因组复制使物种有了额外的基因组拷贝, 可以为物种进化提供更多的遗传原材料。全基因组复制发生以后, 复制基因将面临3个选择: 新功能化、亚功能化以及去功能化, 从而促进物种的多样性及其对极端环境的适应性(De Bodt et al, 2005; Semon & Wolfe, 2007; Hegarty & Hiscock, 2008; Freeling, 2009)。

关于全基因组复制的研究目前主要集中在被子植物中, 而在陆生维管植物的重要代表——蕨类植物中鲜见报道。Dyer等(2013)基于细胞学观察间接推测了铁角蕨复合群(*Asplenium monanthes complex*)的多倍化情况。近年来, 随着分子生物学的快速发展, 基于Ks (synonymous substitution rates)或系统发育基因组学检测蕨类植物全基因组复制的报道在逐年增加, 如满江红(*Azolla pinnata*)(Li et al, 2018)、水蕨属(*Ceratopteris* spp.)(Barker, 2009; Zhang et al, 2019)、木贼属(*Equisetum* spp.)(Vanneste et al, 2015; Clark et al, 2019)以及基于大尺度系统发育框架的蕨类植物(Huang et al, 2019; One Thousand Plant Transcriptomes Initiative, 2019)的全基因组复制研究。

翼盖蕨科广布于泛热带地区, 目前被认为是单种科, 它是真水龙骨科I的基部类群(Zhang & Zhang, 2015; PPG I, 2016)。鉴于翼盖蕨(*Didymochlaena trancatula*)特殊的系统位置, 本文通过对其进行转录组测序和Ks分析, 拟揭示: (1)翼盖蕨是否发生了全基因组复制事件? (2)什么时候发生的全基因组复制? (3)全基因组复制对翼盖蕨的物种形成及环境适应性的影响如何?

1 材料与方法

1.1 转录组测序

采集上海辰山植物园温室中种植的翼盖蕨的孢子叶和营养叶, 迅速置于液氮中冷冻。采用TRIzol[®] Reagent试剂(Invitrogen, 上海)提取总RNA, 然后用Plant RNA Purification Reagent试剂(Invitro-

gen, 上海)对提取的总RNA进行纯化, 之后用Agilent 2100检测RNA提取质量。样品检测合格后, 用带有Oligo (dT)的磁珠富集mRNA。随后加入fragmentation buffer将mRNA打断成短片段, 以mRNA为模板, 用六碱基随机引物(random hexamers)合成cDNA(complementary DNA), 然后加入缓冲液、dNTPs、DNA polymerase I和RNase H合成二链cDNA, 再用AMPure XP beads纯化双链cDNA。纯化后的双链cDNA先进行末端修复、加A尾并连接测序接头, 再用AMPure XP beads进行片段大小选择。之后进行PCR扩增, 并用AMPure XP beads纯化PCR产物, 用Illumina HiSeq 2500进行测序。

1.2 序列组装

对原始数据(raw data)进行过滤得到高质量的有效数据(clean data), 利用Trinity进行de novo拼接。采用Cd-hit软件去除冗余序列, 最终得到unigene序列。进一步利用BUSCO评估翼盖蕨序列拼接的完整度。

1.3 编码蛋白序列预测

编码蛋白序列(coding sequence, CDS)预测分为两步: (1)将unigene按照NR蛋白库和Swissprot蛋白库的优先级顺序进行比对, 若比对上, 则从比对结果中提取出转录本的ORF编码框信息, 并按照标准密码子表将编码区序列翻译成氨基酸序列(5'-3'); (2)对于没有比对上NR蛋白库和Swissprot蛋白库的序列, 或者比对上但未预测出结果的序列, 则采用estscan (3.0.3)软件预测其ORF, 从而得到这部分基因编码的核酸序列和氨基酸序列。

1.4 全基因组复制事件分析

目前全基因组复制事件检测常用的方法是同义替换率(Ks)法。它不影响氨基酸的正常编码, 为中性突变(Kimura, 1977), 假定其以近似恒定的速率累积, Ks值在某种程度上可指示同源基因的发生时间。正常情况下, 物种的Ks分布图呈现L分布, 分布图前段的峰值反映的是近期的基因复制。多倍化发生后会产生大量同源基因, 映射到Ks分布图上便会有大量Ks值接近的同源基因对产生。如果在分布图中有明显其他峰值存在, 则表明在这个Ks值的时间段内有大量的同源基因产生, 即全基因组复制事件(Lynch & Conery, 2000; Vanneste et al, 2013)。另外, 假定同义替换以恒定速率堆积, 可以根据峰值对应

的Ks值估算全基因组复制事件发生的时间(Badouin et al, 2017)。根据公式 $T = Ks/2r$ 可以推断物种的分化时间, 本研究r选用水蕨(*Ceratopteris thalictroides*)的 1.104×10^{-8} 计算复制发生时间(Zhang et al, 2019)。

对物种内CDS序列进行all-against-all blast, 以 e^{-5} 为阈值筛选基因家族中的旁系同源基因对, 使用mclblastline pipeline构建基因家族(Enright et al, 2002), 然后利用MUSCLE对每个基因家族进行比对(Edgar, 2004)。运用 PAML包中的CODEML软件计算得到Ks值(Goldman & Yang, 1994; Yang, 2007)。

本研究将Ks值小于0.1以及大于5的结果去除, 以排除随机误差以及同义替换饱和效应的影响(Blanc & Wolfe, 2004; Schlueter et al, 2004; Cui et al, 2006)。为了避免峰的假阳性, 运用mclust中高斯混合模型对数据进行正态分布拟合(Fraley & Raftery, 2003)。

1.5 基因功能富集

基于Ks分布结果, 调取峰值范围95%的基因, 对其进行功能注释(Ashburner et al, 2000)。采用agriGO v2.0中的Fisher检验对保留下来的复制基因进行GO功能富集(中国农业大学, 北京) ($P < 0.05$) (Tian et al, 2017)。

2 结果

2.1 转录组数据组装

对翼盖蕨进行转录组测序, 共获得47,555,230个原始读长, 过滤后得到45,564,542个有效读长, 拼接后最终获得77,709个转录本和58,871个unigene, 平均长度分别为986 bp和836 bp (表1)。利用 BUSCO进行序列完整性评估, 结果可以识别96.1%的同源基因, 说明该数据序列完整性较高, 详细结果见

附录1。

2.2 全基因组复制检测

对预测的CDS序列进行Ks计算, 获得物种Ks频数分布图(图1), 进一步运用混合模型对其进行数据拟合(表2)。结果显示, 翼盖蕨中共拟合到7个Ks峰, 其中2个Ks峰比较显著, 峰值分别为1.3359和2.0329。推测翼盖蕨至少经历了两次全基因组复制事件。

2.3 全基因组复制时间估计

根据上文的Ks峰值1.3359和2.0329, 推测翼盖蕨两次全基因组复制的发生时间分别为59–62 Mya和90–94 Mya (95%置信区间)。

表1 翼盖蕨转录组de novo组装结果统计
Table 1 Summary of de novo assembly for *Didymochlaena trancatula* transcriptome

	转录本 Transcript	单基因簇 Unigene
序列数目 Sequence number	77,709	58,871
平均长度 Mean length (bp)	986	836
最大长度 Max. length (bp)	17,374	17,374
最小长度 Min. length (bp)	201	201
N50	1,734	1,546

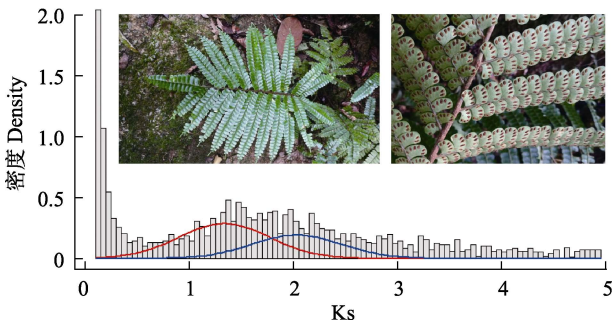


图1 翼盖蕨的形态及其Ks分布图
Fig. 1 Morphology and Ks frequency distribution of *Didymochlaena trancatula*

表2 基于混合模型拟合的翼盖蕨的Ks结果
Table 2 Ks results of *Didymochlaena trancatula* based on mixture modeling

加倍数目 No. of duplications	组件数目 No. of components	贝叶斯信息标准 Bayesian information criterion	中值 Median	方差 Variance	比例 Proportion
1,951	7	5,012.931	0.1211	0.0002	0.0953
1,951	7	5,012.931	0.1796	0.0015	0.0897
1,951	7	5,012.931	0.3518	0.0134	0.0584
1,951	7	5,012.931	1.3359	0.1921	0.3296
1,951	7	5,012.931	2.0329	0.1872	0.2107
1,951	7	5,012.931	3.2182	0.5590	0.1958
1,951	7	5,012.931	4.7933	0.0105	0.0205

2.4 复制基因功能注释

对两个Ks峰值95%置信区间内的复制基因分别进行功能富集(图2, 附录2)。结果显示, 两次全基因组复制被保留下来的基因功能基本一致, 其中与囊泡介导转运(GO:0016192)、调控蛋白代谢过程(GO:0051246)、调控胞内蛋白代谢过程(GO:0032268)、调控翻译(GO:0006417)、基因表达转录后调控(GO:0010608)、翻译调控活性(GO:0045182)及蛋白结合(GO:0005515)的GO terms在两次全基因组复制中均富集到; TATA-binding蛋白(GO:0017025)在一次复制事件中特异保留。

3 讨论

3.1 全基因组复制促进翼盖蕨的物种分化

翼盖蕨的系统位置一直存在争议, 过去的研究

认为翼盖蕨属于肿足蕨科成员(Smith & Cranfill, 2002; Schneider et al, 2004; Tsutsumi & Kato, 2005, 2006)。近年来研究发现, 翼盖蕨属于真水龙骨科I (eupolypods I) 的基部类群, 应该单独成立一科(Zhang & Zhang, 2015), 2016年PPG I接受翼盖蕨科的成立(PPG I, 2016)。

本文通过对翼盖蕨进行全基因组复制分析, 发现翼盖蕨经历了两次全基因组复制事件(图3), 其中一次发生在90–94 Mya, 这与翼盖蕨的物种分化时间基本一致(Rothfels et al, 2015); 此外, 翼盖蕨与同为真水龙骨科I的肿足蕨科、鳞毛蕨科等近缘类群的染色体基数一致, 均为 $X = 41$ (Rothfels et al, 2015), 推测这次全基因组复制事件促进了翼盖蕨的物种分化及其物种多样性(Wang et al, 2012; Soltis et al, 2015)。

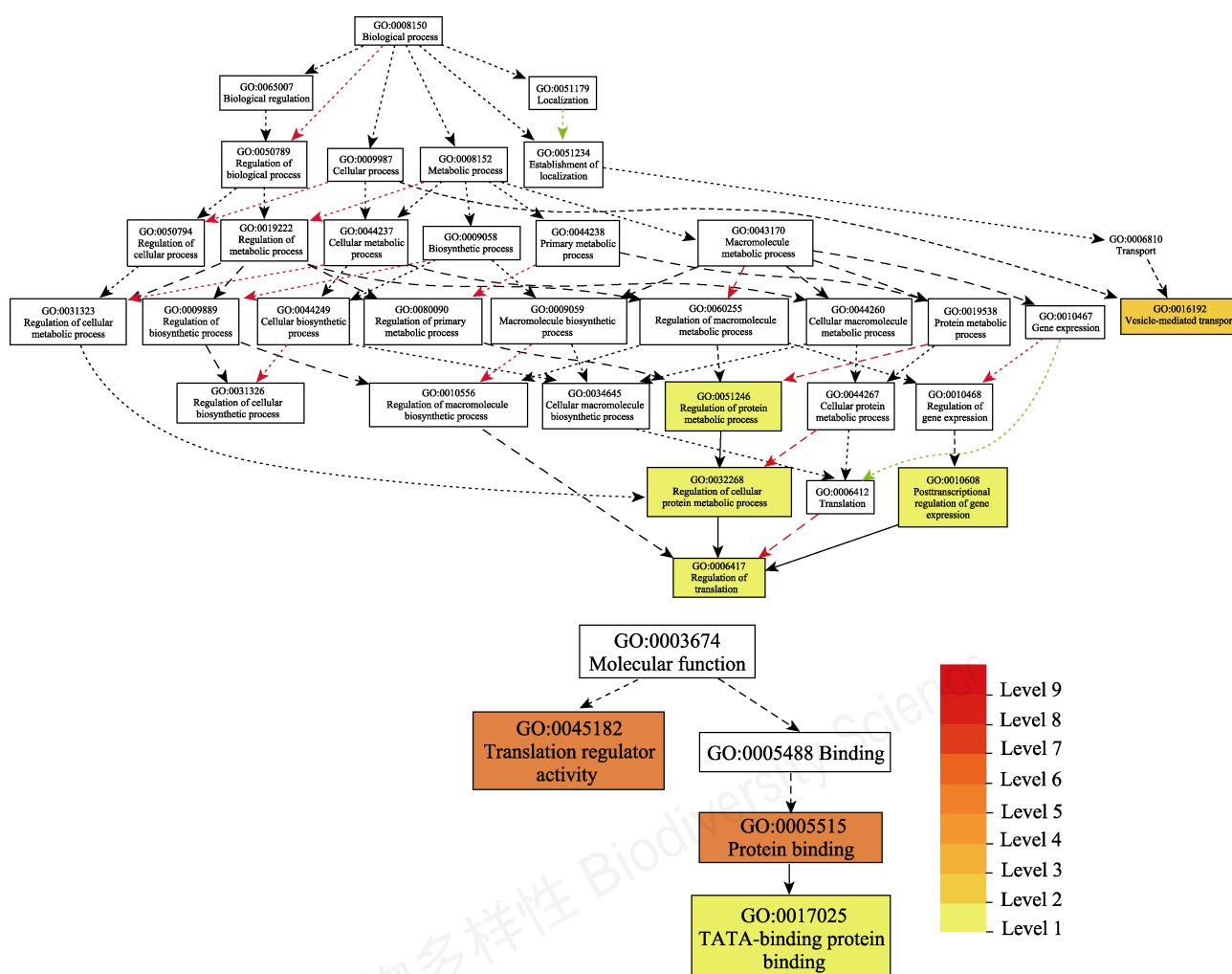


图2 翼盖蕨全基因组复制保留基因的功能富集

Fig. 2 Enriched GO terms of retained duplicates after whole genome duplication events in *Didymochlaena trancatula*

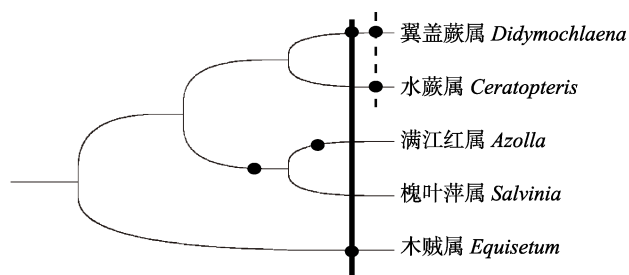


图3 蕨类中已报道的WGDs发生概况。圆圈代表WGD, 实线代表翼盖蕨和巨木贼中(约90 Mya)发生WGD的时间; 虚线代表翼盖蕨和水蕨中(约50 Mya)发生WGD的时间。

Fig. 3 WGDs occurrence in ferns from previous reports. Circle represents WGDs. The solid line denotes the age of WGD at (about 90 Mya) *Didymochlaena trancatula* and *Equisetum giganteum*, the dash line indicates the age of WGD both detected in *Didymochlaena trancatula* and *Ceratopteris thalictroides* (about 50 Mya).

3.2 全基因组复制促进翼盖蕨对环境的适应性

研究发现, 许多物种的全基因组复制事件聚集在白垩纪的物种灭绝时期(60–70 Mya)(Lynch & Conery, 2000; Schlueter et al, 2004; Cui et al, 2006), 该时期地球环境发生剧烈变化, 物种大规模灭绝, 部分物种能适应这种极端环境并存活至今可能归因于物种的全基因组加倍(Vajda et al, 2001; Fawcett et al, 2009)。本研究发现翼盖蕨至少经历了两次全基因组复制事件(图3), 其中一次发生在90–94 Mya, 这与已报道的巨木贼(*Equisetum giganteum*)中全基因组复制发生时间相一致(Vanneste et al, 2015); 另一次发生于59–62 Mya (Cretaceous-Tertiary C-T大灭绝事件), 这与水蕨中报道的全基因组复制发生时间一致, 推测翼盖蕨的这次全基因组复制事件使其成功度过物种大灭绝时期并存活下来(Vanneste et al, 2014; Zhang et al, 2019)。然而, 全基因组复制如何促进翼盖蕨对环境的适应性?

通过对保留下来的复制基因进行功能富集分析(图2), 发现调控蛋白代谢过程(GO:0051246)、调控胞内蛋白代谢过程(GO:0032268)、调控翻译(GO:0006417)、基因表达转录后调控(GO:0010608)及翻译调控活性(GO:0045182)等GO terms被优势保留, 翼盖蕨可能通过调控转录、转录后调控、影响翻译起始因子活性, 并通过调控蛋白代谢, 参与逆境胁迫调控网络, 从而增强翼盖蕨对极端环境的适应性(Muñoz & Castellano, 2012; Guerra et al, 2015); 此外, 还富集到 TATA-binding protein binding (GO:0017025), 其主要功能是通过控制 Transcrip-

tion factor IIB-related protein, 影响植物的萌发率(Niu et al, 2013)。推测这些复制基因的偏倚性保留对于翼盖蕨的环境适应性起重要作用。

参考文献

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: Tool for the unification of biology. *Nature Genetic*, 25, 25–29.
- Athanasios T, Joseph RE, Ronald WD (2000) Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature*, 408, 816–820.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Bérard A, Bergès H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadiou C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Varès D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Muñoz S, Vincourt P, Rieseberg LH, Langlade NB (2017) The sunflower genome provides insights into oil metabolism, flowering and asterid evolution. *Nature*, 546, 148–152.
- Barker MS (2009) Evolutionary genomic analyses of ferns reveal that high chromosome numbers are a product of high retention and fewer rounds of polyploidy relative to angiosperms. *American Fern Journal*, 99, 136–141.
- Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA (2015) On the relative abundance of auto- and allopolyploids. *New Phytologist*, 210, 391–398.
- Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 16, 1667–1678.
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422, 433–438.
- Clark JW, Puttick MN, Donoghue PCJ (2019) Origin of horetails and the role of whole-genome duplication in plant macroevolution. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20191662.
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Research*, 16, 738–749.

- De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution*, 20, 591–597.
- Dyer RJ, Pellicer J, Savolainen V, Leitch IJ, Schneider H (2013) Genome size expansion and the relationship between nuclear DNA content and spore size in the *Asplenium monanthes* fern complex (Aspleniaceae). *BMC Plant Biology*, 13, 219.
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.
- Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30, 1575–1584.
- Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences, USA*, 106, 5737–5742.
- Fraley C, Raftery AE (2003) Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *Journal of Classification*, 20, 263–286.
- Freeling M (2009) Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*, 60, 433–453.
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, 11, 725–736.
- Guerra D, Crosatti C, Khoshro HH, Mastrangelo AM, Mica E, Mazzucotelli E (2015) Post-transcriptional and post-translational regulations of drought and heat response in plants: A spider's web of mechanisms. *Frontier in Plant Science*, 6, 57.
- Hegarty MJ, Hiscock SJ (2008) Genomic clues to the evolutionary success of polyploid plants. *Current Biology*, 18, 435–444.
- Huang CH, Qi X, Chen D, Qi J, Ma H (2019) Recurrent genome duplication events likely contributed to both the ancient and recent rise of ferns. *Journal of Integrative Plant Biology*, doi: 10.1111/jipb.12877.
- Huang CH, Zhang C, Liu M, Hu Y, Gao T, Qi J, Ma H (2016) Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Molecular Biology and Evolution*, 33, 2820–2835.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473, 97–100.
- Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, 267, 275–276.
- Li FW, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux PM, Eily A, Koppers N, Kuo LY, Li Z, Simenc M, Small I, Wafula E, Angarita S, Barker MS, Bräutigam A, dePamphilis C, Gould S, Hosmani PS, Huang YM, Huettel B, Kato Y, Liu X, Maere S, McDowell R, Mueller LA, Nierop KGJ, Rensing SA, Robison T, Rothfels CJ, Sigel EM, Song Y, Timilsena PR, Van de Peer Y, Wang H, Wilhelmsson PKI, Wolf PG, Xu X, Der JP, Schluepmann H, Wong GK, Pryer KM (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, 4, 460–472.
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290, 1151–1155.
- Muñoz A, Castellano MM (2012) Regulation of translation initiation under abiotic stress conditions in plants: It is a conserved or not so conserved process among eukaryotes? *Comparative and Functional Genomics*, 2012, 406357.
- Niu QK, Liang Y, Zhou JJ, Dou XY, Gao SC, Chen LQ, Zhang XQ, Ye D (2013) Pollen-expressed transcription factor 2 encodes a novel plant-specific TFIIB-related protein that is required for pollen germination and embryogenesis in *Arabidopsis*. *Molecular Plant*, 6, 1091–1098.
- One Thousand Plant Transcriptomes Initiative (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, 574, 679–685.
- PPG I (2016) A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution*, 54, 563–603.
- Rothfels CJ, Li FW, Sigel EM, Huie L, Larsson A, Burge DO, Ruhsam M, Deyholos M, Soltis DE, Stewart CN Jr, Shaw SW, Pokorny L, Chen T, dePamphilis C, DeGironimo L, Chen L, Wei X, Sun X, Korall P, Stevenson DW, Graham SW, Wong GK, Pryer KM (2015) The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *American Journal of Botany*, 102, 1089–1107.
- Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome*, 47, 868–876.
- Schneider H, Schuettelpelz E, Pryer KM, Cranfill R, Magallon S, Lupia R (2004) Ferns diversified in the shadow of angiosperms. *Nature*, 428, 553–557.
- Semon M, Wolfe KH (2007) Reciprocal gene loss between tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends in Genetics*, 23, 108–112.
- Smith AR, Cranfill RB (2002) Intrafamilial relationships of the thelypteroid ferns (Thelypteridaceae). *American Fern Journal*, 92, 131–149.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. *American Journal of Botany*, 96, 336–348.
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015)

- Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, 35, 119–125.
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, 45, 122–129.
- Tsutsumi C, Kato M (2005) Molecular phylogenetic study on Davalliaceae. *Fern Gazette*, 17, 147–162.
- Tsutsumi C, Kato M (2006) Evolution of epiphytes in Davalliaceae and related ferns. *Botanical Journal of the Linnean Society*, 151, 495–510.
- Vajda V, Raine JI, Hollis CJ (2001) Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. *Science*, 294, 1700–1702.
- Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10, 725–732.
- Vanneste K, Baele G, Maere S, Van de Peer Y (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Research*, 24, 1334–1347.
- Vanneste K, Sterck L, Myburg AA, Van de Peer Y, Mizrahi E (2015) Horsetails are ancient polyploids: Evidence from *Equisetum giganteum*. *The Plant Cell*, 27, 1567–1578.
- Vanneste K, Van de Peer Y, Maere S (2013) Inference of genome duplications from age distributions revisited. *Molecular Biology and Evolution*, 30, 177–190.
- Wang JT, Li JT, Zhang XF, Sun XW (2012) Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genomics*, 13, 96.
- Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387, 708–713.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA*, 106, 13875–13879.
- Xiang YZ, Huang CH, Yi H, Wen J, Li S, Yi T, Chen H, Xiang J, Hong M (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution*, 34, 262–281.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586–1591.
- Zhang LB, Zhang L (2015) Didymochlaenaceae: A new fern family of eupolypods I (Polypodiales). *Taxon*, 64, 27–38.
- Zhang R, Wang FG, Zhang J, Shang H, Liu L, Wang H, Zhao GH, Shen H, Yan YH (2019) Dating whole genome duplication in *Ceratopteris thalictroides* and potential adaptive values of retained gene duplicates. *International Journal of Molecular Science*, 20, 1926.

(责任编委: 王艇 责任编辑: 闫文杰)

附录 Supplementary Material

附录1 基于BUSCO评价翼盖蕨转录组数据组装完整度

Appendix 1 Assessment of the transcriptome assembly completeness of *Didymochlaena trancatula* based on BUSCO results
<http://www.biodiversity-science.net/fileup/PDF/2019236-1.pdf>

附录2 全基因组复制后保留基因的GO功能富集结果

Appendix 2 Enriched GO terms for the retained genes following the whole genome duplication event of *Didymochlaena trancatula*
<http://www.biodiversity-science.net/fileup/PDF/2019236-2.pdf>

汪浩, 张锐, 张娇, 沈慧, 戴锡玲, 严岳鸿. 转录组测序揭示翼盖蕨(*Didymochlaena trancatula*)的全基因组复制历史. 生物多样性, 2019, 27 (11): 1221–1227. <http://www.biodiversity-science.net/CN/10.17520/biods.2019236>

附录1 基于BUSCO评价翼盖蕨转录组数据组装完整度
Appendix 1 Assessment of the transcriptome assembly completeness of *Didymochlaena trancatula* based on BUSCO results

	比例 Proportion (%)	数目 Number
完整的BUSCOs Complete BUSCOs	96.1	291
完整和单一拷贝的BUSCOs Complete and single-copy BUSCOs	61.4	186
完整和加倍的BUSCOs Complete and duplicated BUSCOs	34.7	105
片段的BUSCOs Fragmented BUSCOs	3.0	9
丢失的BUSCOs Missing BUSCOs	0.9	3
总的BUSCOs Total BUSCOs	100	303

附录2 全基因组复制后保留基因的GO功能富集结果
Appendix 2 Enriched GO terms for the retained genes following the whole genome duplication event of *Didymochlaena trancatula*

GO条目 GO term	本体 Ontology	描述 Description	数目 Number	<i>p</i>	错误发现率 FDR
GO:0051246	生物学过程 Biological process	蛋白质代谢过程调控	94*	8.70e-06*	0.02*
		Regulation of protein metabolic process	98 [#]	2.6e-06 [#]	0.0057 [#]
GO:0006417	生物学过程 Biological process	翻译调控	90*	1.90e-05*	0.027*
		Regulation of translation	95 [#]	3.2e-06 [#]	0.0057 [#]
GO:0010608	生物学过程 Biological process	基因表达的转录后调控	93*	1.70e-05*	0.027*
		Posttranscriptional regulation of gene expression	98 [#]	3e-06 [#]	0.0057 [#]
GO:0032268	生物学过程 Biological process	细胞蛋白质代谢过程调控	94*	5.80e-06*	0.02*
		Regulation of cellular protein metabolic process	98 [#]	1.7e-06 [#]	0.0057 [#]
GO:0016192	生物学过程 Biological process	囊泡介导转运	144*	2.70e-07*	0.0019*
		Vesicle-mediated transport	140 [#]	8.5e-06 [#]	0.012 [#]
GO:0045182	分子功能 Molecular function	翻译调控活性	71*	1.90e-10*	5.30e-07*
		Translation regulator activity	183 [#]	2.9e-11 [#]	8.4e-08 [#]
GO:0005515	分子功能 Molecular function	蛋白结合	1,126*	1.90e-09*	2.70e-06*
		Protein binding	1,170 [#]	7.6e-11 [#]	1.1e-07 [#]
GO:0017025 [#]	分子功能 Molecular function	TATA-binding蛋白结合 TATA-binding protein binding [#]	17 [#]	5e-05 [#]	0.048 [#]

*第一次全基因组复制特有的功能富集结果; [#]第二次全基因组复制特有的功能富集结果。
* Represents the specific functional enrichment results of first whole genome duplication; [#] represents the specific functional enrichment results of the second whole genome duplication.