Molecular Phylogenetics and Evolution 127 (2018) 961–977

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families

Xinping Qi[a], Li-Yaung Kuo[b], Chunce Guo[a], Hao Li[a], Zhongyang Li[c], Ji Qi[a], Linbo Wang[a], Yi Hu[d], Jianying Xiang[e], Caifei Zhang[a], Jing Guo[a], Chien-Hsun Huang[a,*], Hong Ma[a,d,*]

[a] *Ministry of Education Key Laboratory of Biodiversity Sciences and Ecological Engineering and State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, Institute of Plant Biology, Institute of Biodiversity Sciences, Center for Evolutionary Biology, School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200433, China*
[b] *Boyce Thompson Institute, Ithaca, NY 14853, USA*
[c] *College of Life and Environmental Sciences, Gannan Normal University, Ganzhou, Jiangxi 341000, China*
[d] *Department of Biology, The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA*
[e] *College of Biodiversity Conservation and Utilization, Southwest Forestry University, 300 Bailong Road, Kunming 650224, China*

ARTICLE INFO

Keywords:
Fern phylogeny
Nuclear genes
Transcriptomes
Divergence time estimation
Transcription factors

ABSTRACT

Ferns account for 80% of nonflowering vascular plant species and are the sister lineage of seed plants. Recent molecular phylogenetics have greatly advanced understanding of fern tree of life, but relationships among some major lineages remain unclear. To better resolve the phylogenetic relationships of ferns, we generated transcriptomes from 125 ferns and two lycophytes, with three additional public datasets, to represent all 11 orders and 85% of families of ferns. Our nuclear phylogeny provides strong supports for the monophyly of all four subclasses and nearly all orders and families, and for relationships among these lineages. The only exception is Gleicheniales, which was highly supported as being paraphyletic with Dipteridaceae sister to a clade with Gleicheniaceae + Hymenophyllales. In addition, new and strongly supported phylogenetic relationships are found for suborders and families in Polypodiales. We provide the first dated fern phylogenomic tree using many nuclear genes from a large majority of families, with an estimate for separation of the ancestors of ferns and seed plants in early Devonian at ~400 Mya and subsequent gradual divergences of fern orders from ~380 to 200 Mya. Moreover, the newly obtained fern phylogeny provides a framework for gene family analyses, which indicate that the vast majority of transcription factor families found in seed plants were already present in the common ancestor of extant vascular plants. In addition, fern transcription factor genes show similar duplication patterns to those in seed plants, with some showing stable copy number and others displaying independent expansions in both ferns and seed plants. This study provides a robust phylogenetic and gene family evolution framework, as well as rich molecular resources for understanding the morphological and functional evolution in ferns.

1. Introduction

Extant ferns (Polypodiopsida) are the second most diverse group of vascular plants, after angiosperms, with an estimated total of ~10,600 species (PPG I, 2016) and are important for the terrestrial biodiversity and ecosystems. Ferns were suggested to have originated in the Devonian (Bateman, 1991; Schneider et al., 2004) and, with lycophytes, became the principal component of the pre-Cretaceous terrestrial ecosystems (Niklas et al., 1983) before most of the extant angiosperm lineages evolved. Among extant ferns, the vast majority (~10,000 species) belong to the subclass Polypodiidae (leptosporangiate ferns) (PPG I, 2016), whose success was thought to be associated with the evolution of epiphytism (Schuettpelz and Pryer, 2009; Watkins and Cardelús, 2012) and possibly with other diversification mechanisms (Lehtonen et al., 2017). The four early-diverging orders Equisetiales (horsetails), Ophioglossales (e.g., adder's-tongues, grape ferns, and their allies), Psilotales (whisk ferns and fork ferns), and Marattiales fall into the other three subclasses, respectively.

The most recent fern classification recognizes 48 families in 11 orders (PPG I, 2016). At the levels of family and order, fern systematics have mostly been supported by analyses using plastid sequences, including studies with extensive taxon sampling representing most

---

* Corresponding authors at: School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200433, China (C.H. Huang and H. Ma).
E-mail addresses: huang_ch@fudan.edu.cn (C.-H. Huang), hongma@fudan.edu.cn, hxm16@psu.edu (H. Ma).

https://doi.org/10.1016/j.ympev.2018.06.043
Received 22 March 2018; Received in revised form 27 June 2018; Accepted 27 June 2018

families, albeit with only three to six plastid sequences (Pryer et al., 2001, 2004; Wikström and Pryer, 2005; Qiu et al., 2006; Schuettpelz et al., 2006; Schuettpelz and Pryer, 2007; Kuo et al., 2011; Lehtonen, 2011; Rothfels et al., 2012; Knie et al., 2015; Testo and Sundue, 2016). Fern phylogenies with 20–36 species have also been inferred using a combination of three or four plastid, one mitochondrial and one nuclear genes (Wikström and Pryer, 2005; Qiu et al., 2006), five plastid and four mitochondrial genes (Knie et al., 2015), or three or four plastid genes plus one nuclear gene (Pryer et al., 2001, 2004; Schuettpelz et al., 2006). In addition, plastome phylogenomics were used to infer relationships among fern orders (Kim et al., 2014; Lu et al., 2015; Labiak and Karol, 2017) or within the suborder Aspleniineae (eupolypods II) (Wei et al., 2017). Recent analyses of fern phylogeny have also been conducted with 25 nuclear genes from 73 species in 33 families (Rothfels et al., 2015), using 2024 or 1334 nuclear genes from 69 species (Shen et al., 2018), or employing target sequence capture of nuclear genes (Wolf et al., 2018).

However, conflicts or poorly resolved relationships still remain among several families or orders, particularly among the four fern subclasses. In the two studies using four plastid genes or three plastid genes plus one nuclear gene, Ophioglossidae was resolved as the sister of all other extant ferns, with uncertain relationships among the other three subclasses (Pryer et al., 2001, 2004; Lehtonen, 2011). However, analyses using either a few plastid genes or plastid and mitochondria genes placed the subclasses Equisetidae (horsetails), Ophioglossidae and Marattiidae as successive sisters of the other derived groups (Rai and Graham, 2010; Kuo et al., 2011; Knie et al., 2015; Testo and Sundue, 2016), as did two studies using plastome data (Kim et al., 2014; Labiak and Karol, 2017). In other studies using plastome data, Equisetidae was grouped with Ophioglossidae in a clade sister to all other ferns (Karol et al., 2010; Gao et al., 2013; Grewe et al., 2013; Ruhfel et al., 2014; Zhong et al., 2014; Lu et al., 2015). Also, the study with 25 nuclear genes supported Equisetidae, Ophioglossidae, and Marattiidae as successive sisters to Polypodiidae (Rothfels et al., 2015), whereas a phylotranscriptomic study with all major lineages of land plants, including six fern species, strongly supported Ophioglossidae and Marattiidae as sisters among the four subclasses (Wickett et al., 2014). Even the same study using nuclear genes resulted in conflicting relationships when different methods were used: supermatrix analyses supported Equisetidae, Ophioglossidae, Marattiidae being successive sisters to Polypodiidae, while the coalescent analyses suggested a sister relationship between Ophioglossidae and Marattiidae (Shen et al., 2018).

At the order level, Hymenophyllales and Gleicheniales were found to be either sisters in a clade (Pryer et al., 2004) or successive sister clades to the remaining non-Osmundales leptosporangiates (Qiu et al., 2006; Schuettpelz and Pryer, 2007; Rai and Graham, 2010; Lehtonen, 2011; Testo and Sundue, 2016). In some studies, the order Gleicheniales was not even monophyletic (Kuo et al., 2011; Rothfels et al., 2015; Shen et al., 2018). In addition, relationships among suborders are unresolved within the largest order Polypodiales, which comprises the majority of the extant fern species (more than 80% of fern species in 26/48 families). Particularly, the positions of suborders Saccolomatineae and Lindsaeineae and their relationships with the other suborders in Polypodiales are uncertain from studies using plastid genes (Schuettpelz and Pryer, 2007; Rai and Graham, 2010; Kuo et al., 2011; Lehtonen, 2011; PPG I, 2016; Testo and Sundue, 2016); most importantly, the position of Saccolomatineae has only been investigated using nuclear genes from target capture but with low support and conflicting results (Wolf et al., 2018). In addition, the relationships of the early-diverging families in the suborder Aspleniineae and the positions of the families Nephrolepidaceae and Lomariopsidaceae in Polypodiineae are still unclear (PPG I, 2016).

In recent years, phylogenomic and phylotranscriptomic approaches have been demonstrated to be powerful in resolving deep nodes among land plants or within large families and orders (Wen et al., 2013; Wickett et al., 2014; Zeng et al., 2014, 2017; Yang et al., 2015; Huang et al., 2016a, 2016b; Xiang et al., 2017; Morris et al., 2018; Yang et al., 2018). Importantly, these studies provide large-scale analyses of land plant tree of life using sequences from the nuclear genome, which contains the majority of genetic information among the three genomes in plants. Phylogenetic analyses with nuclear genes have also been applied to fern phylogeny recently and have resolved several relationships that were uncertain in previous studies (Rothfels et al., 2015; Shen et al., 2018). However, even when thousands of genes were used, the relationships among the four subclasses were inconsistent between the coalescent and supermatrix analyses (Shen et al., 2018). These conflicts might be affected by missing data, short sequences and possible paralogous genes among the very large number of gene sequences selected without stringent criteria. Such genes might produce wrong topologies, especially for coalescent analysis, which rely on individual gene trees. Also, the sampling of the largest order Polypodiales lacked the earliest-diverging suborder (Saccolomatineae) in a recent study with nuclear genes (Shen et al., 2018).

It has been noted that different nuclear genes can produce different topologies for organismal phylogenies (Zeng et al., 2014, 2017; Huang et al., 2016a, 2016b). The possible reasons could be: First, many nuclear genes have experienced duplication during evolution, and inclusion of paralogs could give rise to wrong topologies; second, short gene sequences can result in incorrect gene tree estimations due to lack of phylogenetic signals; third, incomplete gene sequences found in transcriptomic datasets could result in a large amount of missing data. Thus, it is important to screen nuclear genes for evidence of paralogy or noise, and exclude such misleading sequences from matrices for phylogenetic analysis. In this study, we have generated transcriptomes from 125 fern species and two lycophytes; these and three public datasets of two ferns and one lycophyte (*Selaginella moellendorffii*) together represent all 11 fern orders and 85% of fern families, as well as all three lycophyte orders. We analyzed over 5000 low-copy nuclear candidate marker genes and identified 1000 genes to reconstruct a robust and consistent fern phylogeny with strong supports from multiple phylogenetic analyses, with Equisetidae, Ophioglossidae, and Marattiidae as the earliest successive diverging subclasses. In addition, the topology here includes well-supported relationships among all orders and almost all the families. Most importantly, we provide the first insights of nuclear phylogenomics in Polypodiales based on a sampling covering all suborders and 23 among the 26 families, and all subfamilies for the sampled families, except the subfamily Polybotryoideae in Dryopteridaceae. We found strong support for sister relationship of the basal suborders Saccolomatineae and Lindsaeineae in Polypodiales. In addition, Hypodematiaceae was strongly supported as the earliest-diverging lineage in the suborder Polypodiineae, and such a relationship is different from those found in most of the previous reports (Smith and Cranfill, 2002; Schneider et al., 2004; Liu et al., 2007, 2013; Schuettpelz and Pryer, 2007; Kuo et al., 2011; Lehtonen, 2011; Rothfels et al., 2012, 2015; Zhang and Zhang, 2015; Testo and Sundue, 2016).

Studies on the evolutionary history of multiple gene families can provide valuable insights into the complexity and diversification of organisms. Transcription factors (TFs) play crucial regulatory roles in multiple developmental and physiological processes, including cell division and differentiation, organ initiation and growth, and responses to environmental stimuli (de Mendoza et al., 2013; Weirauch et al., 2014). Previous studies indicated that the expansion, diversification and recruitment of TFs contributed to the morphological innovations and greater complexity in structural and functional evolution during land plant history (Floyd and Bowman, 2007; Pires and Dolan, 2012). Although comprehensive genome-wide analyses have been performed in green algae, gymnosperms and angiosperms (Wilhelmsson et al., 2017), the understanding of evolution of fern transcription factors is still limited. The extensive transcriptome datasets newly generated here provide a great opportunity to examine the histories of such potentially important regulatory genes. Here, we present a profile of transcription factor (TF) gene families across the fern diversity, including estimates