Molecular Phylogenetics and Evolution 81 (2014) 159–173

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Origin of a novel regulatory module by duplication and degeneration of an ancient plant transcription factor

Sandra K. Floyd [a], Joseph G. Ryan [a], Stephanie J. Conway [b], Eric Brenner [c], Kellie P. Burris [d], Jason N. Burris [d], Tao Chen [e], Patrick P. Edger [f], Sean W. Graham [g], James H. Leebens-Mack [h], J. Chris Pires [f], Carl J. Rothfels [i], Erin M. Sigel [j], Dennis W. Stevenson [k], C. Neal Stewart Jr. [d], Gane Ka-Shu Wong [l,m,n], John L. Bowman [a,o,*]

[a] School of Biological Sciences, Monash University, Melbourne, Victoria 3800, Australia
[b] School of Botany, University of Melbourne, Melbourne, Victoria 3010, Australia
[c] Department of Biology, New York University, 1009 Silver Center, 100 Washington Square East, New York, NY 10003-6688, USA
[d] Department of Plant Science, University of Tennessee, Knoxville, TN 37919, USA
[e] Fairy Lake Botanical Garden, Shenzhen & CAS, Guangdong 518004, China
[f] Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA
[g] Department of Botany, University of British Columbia, UBC Bot Garden, Vancouver, BC V6T 1Z4, Canada
[h] Univ Georgia, Dept Plant Biol, Miller Plant Sci, Athens, GA 30602, USA
[i] Department of Zoology, University of British Columbia, #4200-6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
[j] Department of Biology, Duke University, Box 90338, Durham, NC 27707, USA
[k] New York Bot Garden, Bronx, NY 10458, USA
[l] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada
[m] Department of Medicine, University of Alberta, Edmonton, AB T6G 2E1, Canada
[n] BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China
[o] University of California, Davis, Section of Plant Biology, One Shields Ave., Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 4 April 2014
Revised 2 June 2014
Accepted 2 June 2014
Available online 27 September 2014

Keywords:
Gene family evolution
Gene duplication
Homeodomain Leucine Zipper
LITTLE ZIPPER
Evolution of development

ABSTRACT

It is commonly believed that gene duplications provide the raw material for morphological evolution. Both the number of genes and size of gene families have increased during the diversification of land plants. Several small proteins that regulate transcription factors have recently been identified in plants, including the LITTLE ZIPPER (ZPR) proteins. ZPRs are post-translational negative regulators, via heterodimerization, of class III Homeodomain Leucine Zipper (C3HDZ) proteins that play a key role in directing plant form and growth. We show that ZPR genes originated as a duplication of a C3HDZ transcription factor paralog in the common ancestor of euphyllophytes (ferns and seed plants). The ZPRs evolved by degenerative mutations resulting in loss all of the C3HDZ functional domains, except the leucine zipper that modulates dimerization. ZPRs represent a novel regulatory module of the C3HDZ network unique to the euphyllophyte lineage, and their origin correlates to a period of rapid morphological changes and increased complexity in land plants. The origin of the ZPRs illustrates the significance of gene duplications in creating developmental complexity during land plant evolution that likely led to morphological evolution.

1. Introduction

The increasing abundance of genomic and transcriptomic resources has revealed that a basic genetic toolkit was in place prior to the evolutionary radiation of the land plants (embryophytes) (Banks et al., 2011; Delaux et al., 2012; Floyd and Bowman, 2006, 2007; Rensing et al., 2008; Tanabe et al., 2005; Zalewski et al., 2013). Comparative genomics and phylogenetic analyses of gene families has also shown that there were net gains in the number of gene families at key nodes of the land plant phylogeny, including the vascular plant ancestor and later in the euphyllophyte lineage (monilophytes and seed plants), as well as evidence of gene or genome duplications in numerous lineages, including mosses, lycophytes, euphyllophytes, seed plants, and

Abbreviations: C3HDZ, class III Homeodomain Leucine Zipper; HD, homeodomain; LZ, leucine zipper; SAM, shoot apical meristem; SRP, small regulatory protein; TF, transcription factor; ZPR, LITTLE ZIPPER.

* Corresponding author at: School of Biological Sciences, Monash University, Melbourne, Victoria 3800, Australia.
E-mail address: john.bowman@monash.edu (J.L. Bowman).

http://dx.doi.org/10.1016/j.ympev.2014.06.017
1055-7903/

angiosperms (Banks et al., 2011; Floyd et al., 2006; Jiao et al., 2011; Rensing et al., 2008; Zalewski et al., 2013). It has long been recognized that gene duplications provide the raw material for biological evolution (Freeling and Thomas, 2006; Jiao et al., 2011; Ohno, 1970; Zhang, 2003). It is likely that gene and genome duplications were the ultimate source for novel gene families as well as increasing gene family size in land plant genomes. Therefore, in order to understand plant morphological evolution, it is important to be able to reconstruct the evolution of complexity in gene families and make associations between gene duplications, the appearance of novel gene function, and increasing developmental and structural complexity (Aburomia et al., 2003; Floyd and Bowman, 2007).

Although the fate of most duplicate genes may be degeneration and loss, the fact remains that numerous paralogs have accumulated and been retained for hundreds of millions of years following gene and genome duplications during the diversification of land plants (Prince and Pickett, 2002). This is particularly true of plant transcription factor genes (Banks et al., 2011; Blanc and Wolfe, 2004; Edger and Pires, 2009; Floyd and Bowman, 2007; Freeling and Thomas, 2006; Seoighe and Gehring, 2004).

Transcription factors (TFs) are proteins that must interact with other molecules to function (Amoutzias et al., 2008; Riechmann et al., 2000. Many transcription factors regulate gene expression during developmental processes. TFs typically interact with the DNA of their target genes, but frequently have multiple functional domains for interaction with other proteins (Amoutzias et al., 2008). Bridgham et al. (2008) demonstrated that for developmental genes that encode proteins with modular subfunctions (eg. distinct DNA binding and dimerization domains), a simple nucleotide substitution can alter the protein of one paralog so that dimerization may occur, but DNA or other ligand binding is prevented due to loss of critical domains. The result of such mutations is that the mutated paralog becomes an antagonistic, negative, post-translational regulator of any non-altered paralogs.

Many plant transcription factors involved in developmental regulation (bHLH, b-ZIP, HD-ZIP, LEAFY, MADS-Box, WRKY) function as dimers, with dimerization critical for DNA binding and mediated by an independently functioning domain (Amoutzias et al., 2008; Chi et al., 2013; Siriwardana and Lamb, 2012). Since dimerization is essential for their function, duplicate paralogs of most plant transcription factor genes have the potential to become (by degenerate mutation) novel regulatory modules for existing developmental pathways, with consequences for fine-tuning developmental control both spatially and temporally. The extent to which this has happened and the possible implications for land plant evolution remain an intriguing area for investigation.

Recently, a variety of small regulatory proteins (SRPs) have been identified in plants. These have been referred to as "small interfering peptides (siPEPS)" (Seo et al., 2011) or "microProteins" (Staudt and Wenkel, 2011). SRPs competitively inhibit transcription factors, either competing for binding sites or creating non-functional dimers and thus post-translationally dampen the effects of the expressed target TF proteins (Hong et al., 2011; Seo et al., 2011; Staudt and Wenkel, 2011). The importance of plant SRPs in flowering plant development is only just beginning to be appreciated (Seo et al., 2011; Staudt and Wenkel, 2011), and the origin and history of SRPs remains largely unexplored.

The ability of SRPs to interact or compete with transcription factors through shared domains suggests that the SRPs may be evolutionary related to their targets and may represent degenerate paralogs. The origin of such post-translational competitive inhibitors would represent a mechanism for the origin of novel regulatory modules imposed on the ancestral developmental tool kit, and may have had a significant role in morphological change during embryophyte evolution. Thus far, most SRPs are known only from flowering plants, suggesting a relatively recent origin (Magnani and Hake, 2008; Staudt and Wenkel, 2011; Wenkel et al., 2007). However, this may reflect a discovery bias as most developmental research focuses on angiosperms. Recently, Hu et al. (2008) identified MINI ZINC FINGER (MIF) sequences in gymnosperm taxa, indicating that this class of plant SRPs predates the origin of flowering plants. The Aux/IAA genes that competitively regulate AUXIN RESPONSE FACTOR (ARF) transcription factors involved in auxin response are ancient, as are the ARFs themselves. Both ARFs and their Aux/IAA regulators have been identified in the genomes of the moss, *Physcomitrella* and the lycophyte, *Selaginella*, indicating that they may be present in all land plants (Banks et al., 2011). Extensive searches for other SRP homologs in non-flowering plants have not yet been undertaken and have the potential to reveal a more ancient origin for many SRPs than is currently hypothesized.

To resolve the phylogenetic relationships of SRPs and their targets, and therefore infer the timing of the duplication(s) that gave rise to novel regulatory modules, it is essential to have a well-resolved phylogenetic history of both the SRPs and their targets. Very few plant developmental gene families have been broadly studied from a phylogenetic perspective. One notable exception is the class III Homeodomain Leucine Zipper (C3HDZ) gene family. C3HDZs are essential for patterning and differentiation in the *Arabidopsis* shoot, including establishment of the embryonic meristem, patterning and polarity of vascular tissues, and establishment of leaf polarity (Emery et al., 2003; McConnell et al., 2001; Otsuga et al., 2001; Prigge et al., 2005; Talbert et al., 1995). Expression data for C3HDZs in gymnosperms and the lycophyte *Selaginella* suggest that a role in the SAM and vascular patterning may be quite ancient (Floyd and Bowman, 2006; Floyd et al., 2006).

C3HDZ genes have been identified in all land plant lineages and charophycean algae. Phylogenetic analyses of Floyd et al. (2006) and Prigge and Clark (2006) both suggested that C3HDZ gene duplications occurred in a common ancestor of monilophytes (ferns, whiskferns, and horsetails) and seed plants. Comparison of the two independent analyses also suggests that sampling of gymnosperms and monilophytes has not yet revealed the full complement of C3HDZs in those taxa. Full resolution of C3HDZ phylogeny in euphyllophytes requires the addition and analysis of missing homologs.

C3HDZs in *Arabidopsis* are post-translationally regulated by LITTLE ZIPPER (ZPR) proteins (Kim et al., 2008; Staudt and Wenkel, 2011; Wenkel et al., 2007). ZPR genes were first identified as direct targets of the *Arabidopsis* C3HDZ protein REVOLUTA (REV) (Wenkel et al., 2007). Each of the four ZPR genes in *Arabidopsis* (*ZPR1-4*) encodes a short protein (67–105 amino acids) with a single leucine-zipper (LZ) domain that is most similar to the LZ domains of C3HDZ proteins. All four ZPR genes are known to be upregulated by three of the five *Arabidopsis* C3HDZ proteins, REV, PHB, and PHB, (unknown for ATHB8 and ATHB15/CNA) (Brandt et al., 2013; Kim et al., 2008; Wenkel et al., 2007). All four ZPR proteins dimerize with all five C3HDZ proteins via the LZ domains, forming heterodimers that are unable to bind to DNA (Brandt et al., 2013; Kim et al., 2008; Wenkel et al., 2007). Overexpression of ZPR genes causes phenotypic defects that mimic C3HDZ loss-of-function phenotypes (Wenkel et al., 2007), and can partially rescue the *phb-1d* dominant gain-of-function phenotype (Kim et al., 2008). Analysis of ZPR loss-of-function mutants indicates that ZPR function is required for SAM structure and function. *zpr3-2 zpr4-2* double mutants exhibited a variety of defects including abnormal SAM structure, disruption of normal phyllotaxis, production of extra cotyledons and leaves, as well as ectopic axillary meristems (Kim et al., 2008). Together these data suggest that normal SAM maintenance and function (including phyllotaxis) in *Arabidopsis* requires ZPR proteins to maintain the balance of C3HDZ activity (Brandt et al., 2013; Kim et al., 2008).