Expression Level Dominance and Homeolog Expression Bias in Recurrent Origins of the Allopolyploid Fern Polypodium hesperium

Authors: Erin M. Sigel, Joshua P. Der, Michael D. Windham, and Kathleen M. Pryer

Source: American Fern Journal, 109(3) : 224-247

Published By: The American Fern Society

URL: https://doi.org/10.1640/0002-8444-109.3.224

American Fern Journal 109(3):224–247 (2019)
Published on 16 September 2019

Expression Level Dominance and Homeolog Expression Bias in Recurrent Origins of the Allopolyploid Fern *Polypodium hesperium*

Erin M. Sigel*
University of Louisiana at Lafayette, Department of Biology, Billeaud Hall, 410 East St Mary Blvd, Lafayette, Louisiana, 70503, USA, email: erinsigel@louisiana.edu

Joshua P. Der
California State University, Fullerton, Department of Biological Science, P.O. Box 6850, Fullerton, CA 92834, USA, email: jder@fullerton.edu

Michael D. Windham
Duke University, Department of Biology, Box 90338, Durham, NC, 27708, USA, email: michael.windham@duke.edu

Kathleen M. Pryer
Duke University, Department of Biology, Box 90338, Durham, NC, 27708, USA, email: pryer@duke.edu

Abstract.—Allopolyploidization is a common mode of speciation in ferns with many taxa having formed recurrently from distinct hybridization events between the same parent species. Each hybridization event marks the union of divergent parental gene copies, or homeologs, and the formation of an independently derived lineage. Little is known about the effects of recurrent origins on the genomic composition and phenotypic variation of allopolyploid fern taxa. To begin to address this knowledge gap, we investigated gene expression patterns in two naturally formed, independently derived lineages of the allotetraploid fern *Polypodium hesperium* relative to its diploid progenitor species, *Polypodium amorphum* and *Polypodium glycyrrhiza*. Using RNA-sequencing to survey total gene expression levels for 19194 genes and homeolog-specific expression for 1073 genes, we found that, in general, gene expression in both lineages of *P. hesperium* was biased toward *P. amorphum*—both by mirroring expression levels of *P. amorphum* and preferentially expressing homeologs derived from *P. amorphum*. However, we recovered substantial expression variation between the two lineages at the level of individual genes and among individual specimens. Our results align with similar transcriptome profile studies of angiosperms, suggesting that expression in many allopolyploid plants reflects the dominance of a specific parental subgenome, but that recurrent origins impart substantial expression, or phenotypic, variation to allopolyploid taxa.

Key Words.—differential expression; independently derived lineages; polyploid; reticulate evolution; RNA-Seq

Once considered an evolutionary dead end, polyploidy is now widely accepted as an evolutionary force with the potential to influence the fate of entire lineages, from plants to fungi to animals, and even protozoans (Stebbins, 1950; Wolfe and Shields, 1997; Gallardo *et al.*, 1999; Sémon and Wolfe, 2007). This is particularly true among vascular plants, where most lineages have

* corresponding author

experienced one or more ancient polyploidization events (Clark and Donoghue, 2018). An increase in ploidy has co-occurred with an estimated 15% and 31% of speciation events in angiosperms and ferns, respectively (Wood *et al.*, 2009). Such a high prevalence of polyploids suggests that genome duplication may be a primary facilitator of evolutionary change, allowing these organisms to adapt to, and persist in, varying environmental conditions (Soltis and Soltis, 2000; Comai, 2005; Otto, 2007; Meimberg *et al.*, 2009; Ramsey, 2011; Madlung, 2013).

Allopolyploids are especially intriguing because they result from interspecific hybridization accompanied by chromosome doubling, and they begin their evolutionary journey with full sets of chromosomes from each progenitor. The union of divergent, or homeologous, genomes is often associated with major genetic and regulatory changes, such as epigenetic modifications (Salmon, Ainouche, and Wendel, 2005; Lukens *et al.*, 2006), activation of transposable elements (Kashkush, Feldman, and Levy, 2002; Senerchia, Felber, and Parisod, 2014), chromosomal rearrangements (Levy and Feldman, 2004; Chester *et al.*, 2012), homeologous recombination (Gaeta and Pires, 2009; Salmon *et al.*, 2010), and gene loss (Thomas, Pederson, and Freeling, 2006; Buggs *et al.*, 2009; Moghe *et al.*, 2014). These genomic changes, in turn, can result in altered gene expression patterns and phenotypes (Comai, 2005; Coate *et al.*, 2012; Madlung, 2013; Wendel *et al.,* 2018).

While patterns of gene expression in allopolyploids can vary significantly among genes and tissues, as well as among individuals and species, two transcriptional phenomena have been widely observed in microarray and RNA-sequencing experiments. The first phenomenon, referred to as expression level dominance, occurs when the gene expression of an allopolyploid for a given gene is statistically identical to the expression level of only one of its parent species (e.g., Rapp, Udall, and Wendel, 2009; Flagel and Wendel, 2010; Bardil *et al.,* 2011; Yoo, Szadkowski, and Wendel, 2013; Cox *et al.*, 2014). The second phenomenon, referred to as homeolog expression bias, is the preferential expression of homeologs, or gene copies, from one parent for a given gene in an allopolyploid (e.g., Chaudhary *et al.*, 2009; Koh, Soltis, and Soltis, 2010; Grover *et al.,* 2012; Combes *et al.*, 2013; Yoo, Szadkowski, and Wendel, 2013; Akama *et al.*, 2014; Cox *et al.*, 2014; Edger *et al.,* 2017). Patterns of expression level dominance and homeolog expression bias across the transcriptome can mirror or contradict the expression of individual genes, often with both phenomena co-occurring. For example, Yoo, Szadkowski, and Wendel (2013) found that 25% of genes surveyed in the natural allotetraploid *Gossypium hirsutum* exhibited expression level dominance in the direction of its *G. arboretum* (A-genome) parent. Most of the genes that exhibited A-genome dominance, however, exhibited near equal expression of homeologs derived from both the *G. arboretum* (A-genome) parent and the second parental species, *G. raimondii* (D-genome). In general, the interplay of expression level dominance and homeolog expression bias in an allopolyploid is thought to reflect a divergence in the number and distribution of transposable elements (TEs) and cis/trans regulatory elements between

parental subgenomes, ultimately mediating the process of genome fractionation, or loss of duplicated gene copies (Steige and Slotte, 2016; Bottani *et al.,* 2018; Edger *et al.,* 2018; Hu and Wendel, 2018; Wendel *et al.,* 2018).

Increasingly, it is recognized that many, if not most, allopolyploid species form recurrently from distinct hybridization events between the same progenitor species (Soltis and Soltis, 1999) and comprise multiple, independently derived lineages (IDLs). In some cases, allopolyploid species form reciprocally, with IDLs having different maternal progenitor species and, consequently, different maternally-inherited plastid genomes (most plants have maternal inheritance of plastids; Reboud and Zeyl, 1994; Birky, 1995). At the time of formation, IDLs contribute to the genetic diversity of an allopolyploid species by incorporating different parental genotypes (Soltis and Soltis, 1999), but the influence of recurrent origins on genetic and phenotypic variation in subsequent generations is less apparent. Studies on a handful of natural, recurrently formed allopolyploid angiosperms, for example, *Tragopogon miscellus* and *T. mirus*, suggest that IDLs have undergone similar patterns of homeolog evolution, resulting in similar patterns of gene expression (Tate *et al.*, 2006, 2009; Koh, Soltis, and Soltis, 2010; Buggs *et al.*, 2009, 2010). For example, plants belonging to IDLs of *T. miscellus*, exhibit strong homeolog-expression bias, with preferential loss of expression of homeologs derived from their *T. dubius* parent (in many cases due to convergent homeolog loss; Tate *et al.,* 2006, 2009; Buggs *et al.*, 2010). Additional studies of recurrently formed allopolyploids, preferably with broad taxonomic representation, are needed to understand the predictability of homeolog evolution and expression.

As the second largest lineage of vascular plants, ferns have a complex history of polyploidization events and are replete with reticulate species complexes (Haufler and Soltis, 1986; Barrington, Haufler, and Werth, 1989; Barker and Wolf, 2010; Sigel, 2016). *Polypodium hesperium* Maxon is a recurrently formed allotetraploid taxon and member of the well-studied *Polypodium vulgare* complex (Sigel, Windham, and Pryer, 2014). Native to western North America, *P. hesperium* is known to have formed reciprocally between two diploid species, *P. amorphum* Suksd. and *P. glycyrrhiza* D.C. Eaton, that belong to genetically and morphologically distinct clades (diverged approximately 12 mya, Sigel *et al.*, 2014; average genetic divergence of 2%; Table S1). *Polypodium hesperium* comprises at least two IDLs with different maternally inherited plastids—populations north of 42°N have plastids inherited from *P. amorphum* (henceforth *P. hesperium* H^a), whereas those south of 42°N have plastids inherited from *P. glycyrrhiza* (henceforth *P. hesperium* H^g; Sigel, Windham, and Pryer, 2014). Geographic data and estimates of divergence times between *P. amorphum*, *P. glycyrrhiza*, and their respective sister taxa suggest that the IDLs of *P. hesperium* likely originated within the last 1 million years (Sigel, Windham, and Pryer, 2014).

Moving toward the goal of understanding how ferns utilize the genetic diversity imparted by allopolyploidy and recurrent origins, we assessed patterns of gene expression between the IDLs of *P. hesperium*. Because

Polypodium lacks a published genome, we adopted high throughput RNA-sequencing (RNA-Seq) to construct a well-annotated reference transcriptome for *Polypodium* and identify a suite of SNP markers between *P. amorphum* and *P. glycyrrhiza*. By mapping sequencing reads back to the reference transcriptome, we compare gene expression among *P. hesperium* and its diploid progenitors, and assess patterns of homeolog-specific expression in *P. hesperium*. Specifically, we ask the following questions: (1) Do the IDLs of *P. hesperium* exhibit similar patterns total gene expression? (2) Do the IDLs of *P. hesperium* exhibit similar patterns of homeolog-specific expression? (3) Does *P. hesperium* exhibit expression level dominance and homeolog expression bias comparable to other allopolyploid plants?

Material and Methods

Plant material.—For this study we used 12 *Polypodium* individuals—three individuals of *P. amorphum*, three individuals of *P. glycyrrhiza*, three individuals of *P. hesperium* H^a, and three individuals of *P. hesperium* H^g (Table 1). All individuals of *P. hesperium* were included in a previous study investigating its reciprocal origins in which their maternally inherited plastid haplotypes were confirmed by sequencing the *trnG-trnR* intergenic spacer (Sigel, Windham, and Pryer, 2014).

All plants were collected from wild populations between August 2010 and August 2011, and transported live to the Duke University Greenhouses. Each plant was cleaned with water to remove soil from the roots and rhizomes and repotted in Farfard Mix 2 (Sun Gro Horticulture Canada Ltd., USA). Plants were maintained under common glasshouse conditions (photoperiod 18 h: 6 h, light: dark, with luminosity of 200–400 Umol sec^{-1} cm^2; 27–67% humidity; daytime temperature: 18.3–21.1°C; nighttime temperature: 17.8–20.6°C) for a minimum of 18 months prior to sampling for RNA extraction. Material from a single leaf was taken from each individual, flash frozen in liquid nitrogen 21 days after it had fully unfurled but before sporangia had developed, and stored at -80°C until RNA extraction. Material for all individuals was collected between February 28 and March 5, 2013, and always between 10:00 and 11:00 am Eastern Standard Time.

RNA extraction, library construction, and Illumina sequencing.—Total RNA was extracted from 70–100 milligrams of leaf material per *Polypodium* individual using the Spectrum Plant Total RNA kit (Sigma-Aldrich, U.S.A.) without modification. The Aglient DNA 2100 Bioanalyzer (Aglient Technologies, U.S.A.) were used to assess the quality and concentration of each RNA extraction. Twelve mRNA libraries, one for each *Polypodium* individual, were constructed using the TruSeq RNA Sample Prep Kit (Illumina, U.S.A). The standard protocol was modified to produce barcoded, strand-specific, paired-end libraries following Borodina, Adjaye, and Sultan (2011). Equimolar amounts of the 12 libraries were pooled into a single sample and submitted to the Duke University Genome Sequencing and Analysis Core (www.genome.

TABLE 1. *Polypodium* specimens used in this study. DB numbers refer to unique individual identifiers as designated in the Duke Fern Lab Database (https://fernlab.biology.duke.edu). H^a = plastid genome derived from *Polypodium amorphum*; H^g = plastid genome derived from *Polypodium glycyrrhiza*. All vouchers are accessioned at DUKE herbarium.

Taxon (ploidy)	DB number	Voucher information
Polypodium amorphum Suksd. (2*x*)		
P. amorphum	8521	Canada, British Columbia, Squamish-Lillooet Regional District, *Rothfels 4084*
P. amorphum	7773	U.S.A., Oregon, Multnomah County, *Sigel 2010-104*
P. amorphum	7771	U.S.A., Washington, King County, *Sigel 2010-125*
Polypodium glycyrrhiza D.C. Eaton (2*x*)		
P. glycyrrhiza	8493	Canada, British Columbia, Squamish-Lillooet Regional District, *Rothfels 4059*
P. glycyrrhiza	7767	U.S.A., Washington, Snohomish County, *Sigel 2010-81*
P. glycyrrhiza	7559	U.S.A, Oregon, Lincoln County, *Rothfels 3875*
Polypodium hesperium Maxon (4*x*)		
P. hesperium H^a	8179	U.S.A., Idaho, Shoshone County, *Sigel 2011-46*
P. hesperium H^a	8276	U.S.A., Montana, Lincoln County, *Sigel 2011-31C*
P. hesperium H^a	8280	U.S.A., Montana, Lake County, *Sigel 2011-37*
P. hesperium H^g	8175	U.S.A., Arizona, Coconino County, *Sigel 2011-08A*
P. hesperium H^g	8177	U.S.A., Arizona, Pinal County, *Sigel 2011-09A*
P. hesperium H^g	8288	U.S.A., Arizona, Graham County, *Sigel 2011-10*

duke.edu). Sequencing of 100 base paired-end reads was performed on four lanes of the Illumina HiSeq 2000 (Illumina, U.S.A.).

Processing of raw sequencing reads.—Reads from each library were sorted by barcode, and the number and quality of reads for each sample were evaluated with FastQC v. 3-5-2012 (Andrews, 2010). Adapter sequences and low-quality reads were removed using Trimmomatic v. 0.30 (Bolger, Lohse, and Usadel, 2014). All read pairs for which both the forward and reverse read passed quality filtering were used in subsequent analyses.

Generating a de novo Polypodium *reference transcriptome*.—The lack of a published genome for *Polypodium* or closely related genera required us to generate a *de novo* reference transcriptome as a prerequisite for assessing gene expression patterns. To minimize read mapping related technical bias, an initial reference transcriptome was generated using all filtered reads from the three individuals of *P. amorphum* and the three individuals of *P. glycyrrhiza* using Trinity v. r20140413 (Grabherr *et al.*, 2011) on the Duke Shared Cluster Resource. A strand-specific protocol (*RF*) was used with a fixed k-mer size of 25 and a maximum insert size of 800bp. All reads were then mapped back to the initial reference transcriptome using Bowtie v. 1.0.1 (Langmead *et al.*, 2009) and quantified with RSEM v. 1.2.14 (Li and Dewey, 2011) using the wrapper script align_and_estimate_abundance.pl (Grabherr *et al.*, 2011). Transcripts with less than three reads mapped back to them were considered putative artifacts of the transcriptome assembly algorithm and removed from the reference transcriptome (Grabherr *et al.*, 2011; Haas *et al.*, 2013).

To restrict differential gene expression analyses to protein coding genes of known Gene Ontology (GO) classifications (Ashburner *et al.*, 2000), we first translated transcripts using ESTScan, a program that detects coding regions and corrects sequencing errors that cause frameshifts (Iseli, Jongeneel, and Bucher, 1999). We then annotated the reference transcriptome to orthogroup (i.e., narrowly defined gene lineages; Li, Stoeckert, and Roos, 2003) using the global PlantTribes gene family classification (Wall *et al.*, 2008) as described in Sigel *et al.* (2018). All transcripts without hits to the PlantTribes classification were excluded from the *Polypodium* reference transcriptome and not used in subsequent analyses. The completeness of the final *Polypodium* transcriptome assembly was assessed by quantifying the presence of universal single copy orthologs using BUSCO software and the Embryophyta odb09 dataset under the trans setting (Simão *et al.*, 2015).

Read mapping, quantification, and differential expression analysis.—For each of the 12 *Polypodium* individuals, filtered reads were mapped to the *Polypodium* reference transcriptome using Bowtie v. 1.0.1 (Langmead *et al.*, 2009) as implemented in the Trinity package v. r20140413 (Grabherr *et al.*, 2011), using parameters for paired-end, strand-specific reads (as described in Haas *et al.*, 2013). Maximum likelihood read abundances were calculated for each gene using RSEM v. 1.2.13 (Li and Dewey, 2011), and the Trinity package script abundance_estimates_to_matrix.pl (Haas *et al.*, 2013) was used to generate a combined matrix of read abundances of all genes for all samples. Prior to differential expression analyses, the combined read abundance matrix was filtered to remove genes that were not expressed consistently (i.e., expressed in all three biological replicates) by either *P. amorphum* and/or *P. glycyrrhiza*. Additive, or midparent (Grover *et al.,* 2012), expression between *P. amorphum* and *P. glycyrrhiza* was calculated for each gene *in silico* by calculating the mean expression of the six diploid accessions.

Differential expression analyses among individuals were performed using the edgeR release 2.14 in R v. 3.1.0 (Robinson, McCarthy, and Smyth, 2010) with the trimmed mean of M-values method (TMM) to normalize read counts within and across libraries (Robinson and Oshlack, 2010; Dillies *et al.*, 2013). This results in gene expression values that are normalized to transcriptome size, effectively reducing gene expression estimates to per transcriptome concentrations rather than absolute expression. Differential expression analyses were performed in one of two ways: (1) for the three biological replicates of each IDL of *P. hesperium* and (2) for each of the six individuals of *P. hesperium*. For both analyses, differential expression of a gene was assessed by comparing the gene expression among *P. hesperium, P. amorphum, P. glycyrrhiza,* and the *in silico* midparent using Fisher's exact test. The Benjamini and Hochberg (BH) method was used to adjust p-values to account for false discovery (FDR) of differentially expressed genes (Benjamini and Hochberg, 1995; Robinson and Oshlack, 2010). The number of genes exhibiting statistically significant differential expression among *P. hesperium, P. amorphum, P. glycyrrhiza*, and the *in silico* midparent were identified using thresholds of log2 fold change (FC) ≥ 2 and FDR ≤ 0.01. Significant differences

in the number of differentially expressed genes between the IDLs of *P. hesperium* and its diploid progenitors were determined with Fisher's exact test, $\alpha = 0.01$, as implemented in R (R Core Team, 2013).

Expression level dominance.—Each gene identified as differentially expressed between *P. hesperium* and at least one of its diploid progenitors was assigned to one of 12 possible categories in accordance with Rapp, Udall, and Wendel (2009), which includes multiple categories of additivity (midparent), expression level dominance (statistically identical to only one progenitor), and transgressive regulation (gene expression outside the range of either progenitor). Significant differences in the number of genes belonging to each of the 12 categories within and between the IDLs of *P. hesperium* were determined with Fisher's exact test, $\alpha = 0.01$, as implemented in R (R Core Team, 2013).

Homeolog-specific expression.—One method for assessing subgenomic contributions to gene expression in allopolyploids uses fixed single nucleotide polymorphisms (SNPs) between progenitor species to identify homeologous gene copies in the allopolyploid. SNPs between progenitor species at the time of hybridization are vertically transmitted to the derived allopolyploid. Even after substantial evolutionary time, some SNPs will remain fixed between the diploid species and can be used to identify homeologous gene copies in the allopolyploid (Adams, 2007; Flagel and Wendel, 2009). Encouraged by numerous applications of this method to model and non-model allopolyploids (e.g., Buggs *et al.*, 2010; Combes *et al.*, 2013; Akama *et al.*, 2014; Krasileva *et al.*, 2013; Mithani *et al.*, 2013; Nagy *et al.*, 2013; Page *et al.*, 2013; Cox *et al.*, 2014), we identified putative interspecific SNPs between the diploid species *P. amorphum* and *P. glycyrrhiza* to estimate homeolog-specific contributions to gene expression in the IDLs of *P. hesperium*.

We used GATK Unified Genotyper in combination with previously generated bam alignment files to simultaneously identify SNPs between the *Polypodium* reference transcriptome and each of the 12 *Polypodium* individuals (McKenna *et al.*, 2010; DePristo *et al.*, 2011; Van der Auwera, 2013). The -stand_call_conf and -stand_emit_conf parameters were set to 30 and 10, respectively, to exclude low quality variant calls. The -ploidy parameter was set to 2 for each individual, regardless of whether it was diploid or allopolyploid, to reduce the complexity of variant calls and limit subsequent analyses to intergenomic homeologous SNPs in *P. hesperium* rather than intragenomic variation (i.e., between homologs). GATK Unified Genotyper yielded a single variant call format (VCF) file containing the estimated genotype at each SNP site for each *Polypodium* individual and the read depth of each allele.

Custom scripts were used to identify putative fixed SNPs between the diploid progenitors, and calculate the homeolog expression ratios for each individual of *P. hesperium*. Briefly, SNPs were filtered to exclude sites with low mapping quality (Phred scaled probability score < 30) and genotype calls based on a low number of reads (< 4 reads). Putatively fixed SNPs between *P. amorphum* and *P. glycyrrhiza* were identified as having reciprocal homozygous genotypes in all replicates of each species (i.e., *P. amorphum* genotype =

0/0 and *P. glycyrrhiza* genotype = 1/1, or *P. amorphum* genotype = 1/1 and *P. glycyrrhiza* genotype = 0/0). The homeolog expression ratio for a particular gene for each individual of *P. hesperium* was estimated by dividing the number of *P. amorphum*-derived reads by the total number of reads for all SNPs assigned to that gene. A homeolog expression ratio of 1 indicates that expression for a given gene is solely comprised of homeologs derived from *P. amorphum*, whereas a homeolog expression ratio of 0 indicates that expression for a given gene is comprised solely of homeologs derived from *P. glycyrrhiza*. Intermediate ratios indicate that expression is comprised of homeologs derived from both *P. amorphum* and *P. glycyrrhiza*.

For each gene expressed in all six individuals of *P. hesperium*, we estimated the total number of reads mapping to a *P. amorphum*-derived homeolog and a *P. glycyrrhiza*-derived homeolog by multiplying the total raw read counts as determined by RSEM v. 1.2.13 (Li and Dewey, 2011), by the homeolog expression ratio. EdgeR release 2.14 in R v. 3.1.0 (Robinson, McCarthy, and Smyth, 2010) was used to estimate FPKM values for homeolog-specific read counts (Robinson and Oshlack, 2010; Dillies *et al.*, 2013). We then implemented two likelihood ratio tests (LRTs; Edger *et al.,* 2017; Smith *et al.,* 2019) in MATLAB v. 9.6.0 (MATLAB, 2019) to identify genes with statistically significant differences in homeolog-specific expression for *P. hesperium* H^a and *P. hesperium* H^g. This method determines if there is sufficient evidence to reject a null hypothesis of no homeolog expression bias in favor of an alternative hypothesis of homeolog expression bias for each gene, while accounting for relative read depth within and among *P. hesperium* individuals (Smith *et al.,* 2019). Statistical significance was assessed at α=0.05, applying the BH correct for multiple testing error (Benjamini and Hochberg, 1995). A third LRT test was conducted to identify genes with significant differences in the magnitude of homeolog expression bias between *P. hesperium* H^a and *P. hesperium* H^g (Smith *et al.,* 2019; see File S1 for MATLAB code and data tables).

Results

Illumina sequencing reads.—Approximately 46-68 million 100bp paired-end reads were generated for each *Polypodium* individual. After removing adapter sequences and filtering out low quality reads, 87.9-89.7% of read pairs were retained, as were 4.9-6.2% of forward only reads and 1.9-2.1% of reverse only reads (Table S2). Raw Illumina reads, as well as nucleotide sequences and peptide translations of the final reference transcriptomes are deposited in NCBI's Sequence Read Archive (www.ncbi.nlm.nih.gov/sra; BioProject: PRJNA549695).

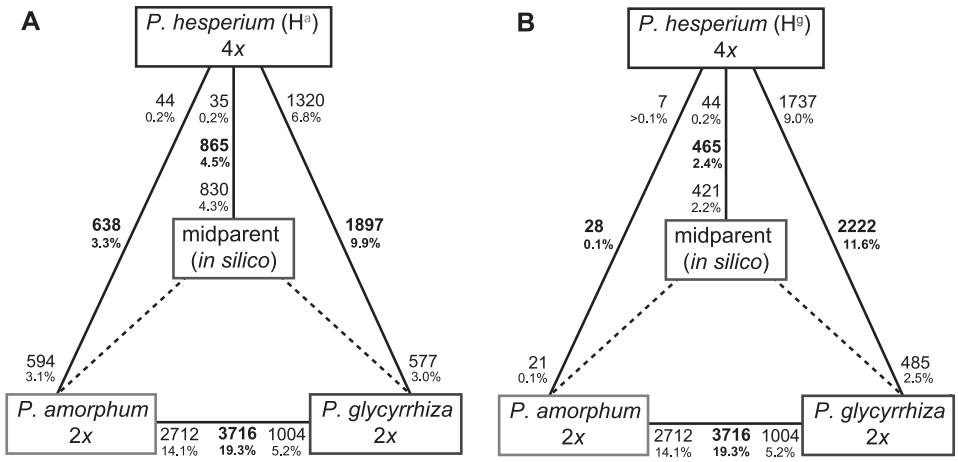
Polypodium *reference transcriptome and annotation*.—The initial *Polypodium* transcriptome comprised 420542 transcripts corresponding to 192418 genes (e.g., Trinity components) with an N50 of 1560 bases and a mean contig length of 814.64 bases. Following filtering to remove potential assembly artifacts and any transcripts without significant similarity (e-value $< 1e^{-8}$) to

one or more orthogroup HMM profiles present in the PlantTribes classification, the final *Polypodium* transcriptome comprised 82893 transcripts corresponding to 25769 genes, with an N50 of 2300 bases and an average transcript length to 1756.44 bases (see Table S3 for assembly statistics and Files S2 and S3 for nucleotide sequences and peptide translations of the final reference transcriptome). A summary of orthogroup numbers and annotation terms associated with each of the 82893 transcripts in the final *Polypodium* reference transcriptome is provided in Table S4. BUSCO assessment determined that 69.1% of universal single-copy Embryophyta orthologs are represented as complete sequences, with another 4.4% represented as partial sequences (Fig. S1). BUSCO representation in our final *Polypodium* reference transcriptome is similar to the previously published *P. amorphum* transcriptome (Sigel *et al.*, 2018)

Differential expression.—Of the 25769 genes in the *Polypodium* reference transcriptome, 19194 were consistently expressed (i.e., expressed in all three biological replicates) by either *P. amorphum* and/or *P. glycyrrhiza* and used for differential expression analyses (Table S5). Comparison of transcriptome-normalized gene expression in *P. amorphum* and *P. glycyrrhiza* using thresholds of FC $\geq$ 2 and FDR $\leq$ 0.01 revealed that 3716 genes (19.4%) were differentially expressed between the diploid species (Fig. 1), with significantly more genes highly expressed in *P. amorphum* (2712 genes, 14.1%) than were highly expressed in *P. glycyrrhiza* (1004 genes, 5.2%; p-value $<$ 0.0001, Fisher's exact test).

Comparisons of transcriptome-normalized gene expression in *P. hesperium* H^a and *P. hesperium* H^g relative to their diploid progenitors recovered similar patterns of differential expression between the IDLs (Fig. 1a, b; also see Fig. S1 for individual *P. hesperium* specimens). For the vast majority of genes surveyed, no significant differences in per transcriptome expression levels were detected between *P. hesperium* and either of its diploid progenitors. Both IDLs of *P. hesperium* had substantially less differentially expressed genes when compared with *P. amorphum*, than with *P. glycyrrhiza* or the *in silico* midparent. Put differentially, per transcriptome gene expression levels in both *P. hesperium* H^a and *P. hesperium* H^g more closely mirrored that of *P. amorphum*. Of the genes that were differentially expressed between *P. amorphum* and *P. hesperium,* significantly more were highly expressed in *P. amorphum*. In contrast, of those genes that were differentially expressed between *P. glycyrrhiza* and *P. hesperium*, significantly more were highly expressed in *P. hesperium* (Fig. 1).

One striking difference between the IDLs of *P. hesperium* is the number of differently expressed genes relative to *P. amorphum*. *Polypodium hesperium* H^g has significantly fewer genes that are differentially expressed with *P. amorphum* than does *P. hesperium* H^a (H^g :28 vs. H^a: 638; Fig. 1). However, this pattern is not consistently recovered in analyses of individual *P. hesperium* accessions (Fig. S2), indicating that there is substantial variation in the number and identity of genes mirroring the expression levels of *P. amorphum* and *P. glycyrrhiza* within and between the IDLs of *P. hesperium*.



C

	Additivity		*P. amorphum*-expression level dominance		*P. glycyrrhiza*-expression level dominance		Transgressive down-regulation			Transgressive up-regulation			DE A and G only	No change
	I	XII	IV	IX	II	XI	III	VII	X	V	VIII	VI		
	A H G	A H G	A H G	A H G	A H G	A H G	A H G	A H G	A H G	A H G	A H G	A H G		
H^a	3	1	1088	392	0	0	7	0	0	0	0	128	2406	15169
H^g	1	0	1666	0	6	19	0	0	0	1	0	0	2096	15406
Common to H^a & H^g	0	0	223	0	0	0	0	0	0	0	0	0	429	12749

FIG. 1. Patterns of differential gene expression in the independently derived lineages of *P. hesperium*. A-B. Number of genes differentially expressed in each contrast between *P. hesperium* H^a and H^g, respectively, and each of its diploid progenitors at thresholds log2 fold change ≥ 2 and FDR ≤ 0.01. Bolded values indicate the total number and percentage of genes differentially expressed in each contrast. Values that are not bolded indicate the number of genes that are more highly expressed in a particular species for each contrast. For example, of the 19194 genes evaluated, 3716 (19.3%) are differentially expressed between *P. amorphum* and *P. glycyrrhiza*, with 2712 (14.1%) more highly expressed in *P. amorphum* and 1004 (5.2%) more highly expressed in *P. glycyrrhiza*. C. The number of genes belonging to 12 categories of differential expression among *P. hesperium, P. amorphum*, and *P. glycyrrhiza*. Roman numerals correspond to the category designations given in Rapp *et al.*, (2009), and graph insets illustrate the expression level of *P. hesperium* (H) relative to *P. amorphum* (A) and *P. glycyrrhiza* (G). "DE A and G only" refers to genes with statistically significant differences in gene expression between *P. amorphum* and *P. glycyrrhiza* but not *P. hesperium*. "No Change" refers to genes without a statistically significant change in gene expression between all three species.

Expression level dominance.—Following the criteria of Rapp, Udall, and Wendel (2009), differentially expressed genes were binned into 12 categories reflecting the per transcriptome expression level in *P. hesperium* relative to its diploid progenitors (Fig. 1; see Fig. S3 for individual *P. hesperium* specimens).

There were no significant changes in expression among *P. hesperium* (for both H^a and H^g), *P. glycyrrhiza*, and *P. amorphum* for the vast majority of the 19194 genes surveyed. For both IDLs of *P. hesperium*, of the genes with differential expression relative to the diploid progenitors, the vast majority exhibited expression level dominance mirroring *P. amorphum*. Notably, this was true for most genes in which expression is upregulated in *P. amorphum* relative to *P. glycyrrhiza* (category IV). Both IDLS had very few genes exhibiting additive expression (categories I and XII) and transgressive regulation (categories III, V, VI, VII, VIII, X). Similarly, a broad pattern of expression level dominance favoring *P. amorphum* over *P. glycyrrhiza,* is present for each of the six individual accessions of *P. hesperium* regardless of IDL, with substantial variation in the number and identity of genes in each expression category (Fig. S3).

Homeolog-specific expression.—A total of 935,207 SNPs were detected among the 12 *Polypodium* samples (File S4). We were able to identify 5564 putatively fixed SNPs between the diploid species. By combining genotype and read depth information from all SNPs assigned to a single gene, we identified 1073 genes for which both *P. amorphum*-derived homeologs and *P. glycyrrhiza*-derived homeologs were expressed in all six individuals of *P. hesperium*.

For a majority of these 1073 genes both *P. hesperium* H^a and *P. hesperium* H^g exhibited strong, statistically significant homeolog expression bias favoring *P. amorphum*-derived homeologs; 923 (86.0%) and 744 (69.3%) genes in *P. hesperium* H^a and *P. hesperium* H^g, respectively. For *P. hesperium* H^a and *P. hesperium* H^g*, P. amorphum*-derived homeologs were expressed, on average, approximately 8 and 12 times greater (H^a: FC=3.06; H^g: FC= 3.62) than *P. glycyrrhiza*-derived homeologs (Fig. 2a,b).

In contrast, both IDLs of *P. hesperium* exhibited few genes with strong expression bias in favor of *P. glycyrrhiza*-derived homeologs (e.g., H^a: 100 genes, H^g: 39 genes). For these genes, there is a notable difference between *P. hesperium* H^a and *P. hesperium* H^g in the magnitude of their homeolog expression biases. On average, *P. hesperium* H^a exhibited a 3-fold bias (FC = 1.54) in favor of *P. glycyrrhiza*-derived homeologs, whereas *P. hesperium* H^g exhibited a 10-fold bias (FC = 3.31) in favor of *P. glycyrrhiza*-derived homeologs (Fig. 2a,b).

When comparing homeolog expression biases between the IDLs of *P. hesperium* at the level of specific, individual genes, we found that the majority (81%) of 1073 genes surveyed did not exhibit a statistically significant difference in the magnitude or direction of the expression bias (Fig. 2c). Of the 209 genes that differed in homeolog expression bias between the IDLs, 128 exhibited preferential expression of *P. glycyrrhiza*-derived homeologs in *P. hesperium* H^g relative to *P. hesperium* H^a.

Discussion

Allopolyploid species are prevalent in both flowering plants and ferns (Wood *et al.,* 2009), but studies investigating gene expression in allopolyploids

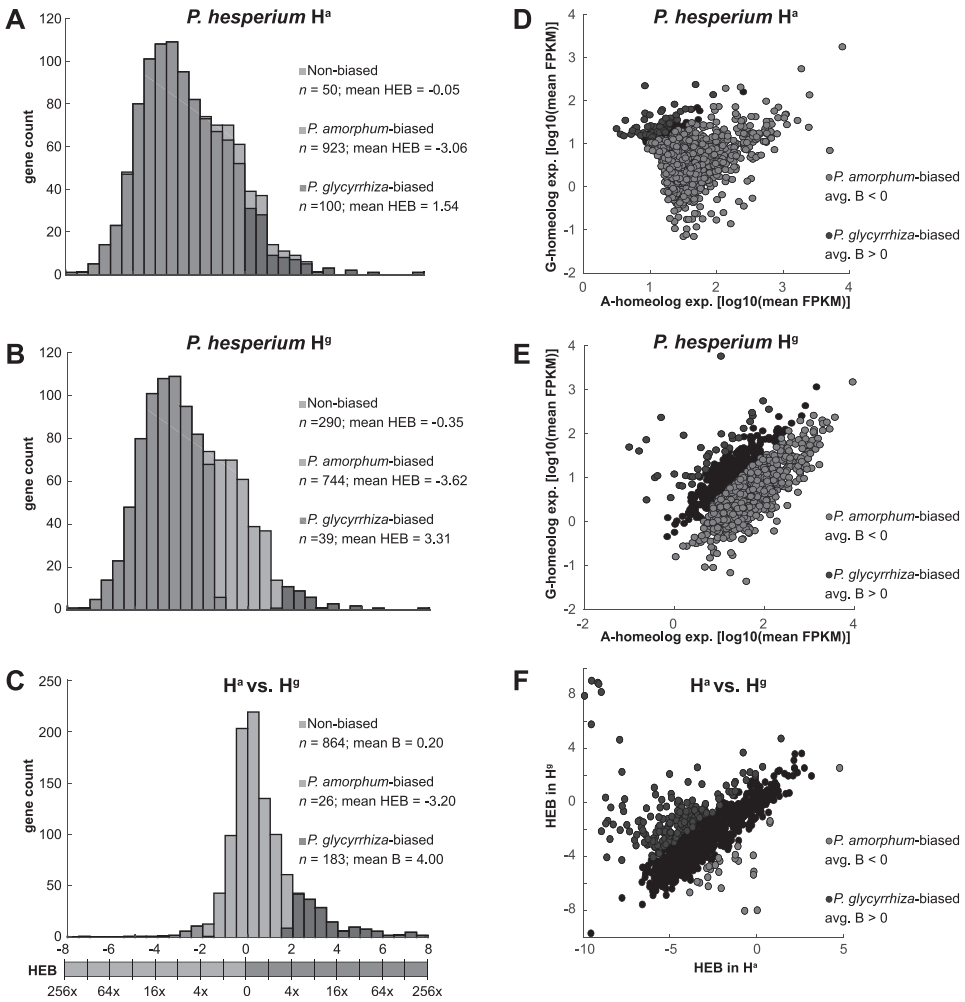


FIG. 2. Results of likelihood ratio tests for homeolog expression bias (HEB) in *P. hesperium* for 1073 genes (BH-adjusted p-value < 0.05). HEB is reported as the log2 fold change in expression between homeologs, with negative values indicating a bias toward *P. amorphum*-derived homeologs and positive values indicating a bias toward *P. glycyrrhiza*-derived homeologs. A-B. Histograms of the distribution of HEB in *P. hesperium* H^a and *P. hesperium* H^g, respectively. Genes significantly biased toward the expression of *P. amorphum*-derived homeologs are shown in red, genes significantly biased toward the expression of *P. glycyrrhiza*-derived homeologs are shown in blue, and non-biased genes are shown in gray. C. Distribution of change in HEB between *P. hesperium* H^a and *P. hesperium* H^g. Genes with significantly greater HEB favoring the *P. amorphum*-derived homeolog in *P. hesperium* H^g relative to *P. hesperium* H^a are shown in red, whereas genes with significantly greater HEB favoring the *P. glycyrrhiza*-derived homeolog in *P. hesperium* H^g relative to *P. hesperium* H^a are shown in blue. Genes without significant change in HEB are shown in gray. D-C. Transcriptome-normalized expression levels of *P. amorphum*-derived homeologs and *P. glycyrrhiza*-derived homeologs in *P. hesperium* H^a and *P. hesperium* H^g, respectively. Color coding is the same as in panels A and B. E. HEB in *P. hesperium* H^a and *P. hesperium* H^g. Color coding is the same as in panel C.

have been largely limited to angiosperms. Here we broaden the evolutionary perspective by using RNA-Seq to assess patterns of gene expression in *Polypodium hesperium*, an allotetraploid fern taxon comprising reciprocally-formed IDLs. Recurrent origins increase the genetic diversity of an allopolyploid species by integrating variation from different progenitor populations (Werth and Lellinger, 1992; Soltis and Soltis, 1999), but little is known about how recurrent origins contribute to variation in gene expression. We find that the IDLs of *P. hesperium* are remarkably consistent in their transcriptome-wide patterns of gene expression and homeolog expression biases. However, we also find substantial variation in the expression of particular genes and among individuals of *P. hesperium*, suggesting that recurrent origins may contribute to phenotypic variation of an allopolyploid taxon. In addition, we report unprecedented levels of unbalanced expression level dominance and unbalanced homeolog expression bias, supporting the notion that these phenomena are pervasive consequences of allopolyploidy in plants.

Additive gene expression in P. hesperium.—It is widely reported that the expression levels of many genes in angiosperm allopolyploids deviate from additive, or midparent, values (e.g., Wang *et al.,* 2006; Chagué, *et al.,* 2010; Yoo, Szadkowski, and Wendel, 2013; Zhao *et al.,* 2013, Jiang *et al.,* 2015). However, for the vast majority of genes surveyed for both *P. hesperium* H^a and *P. hesperium* H^g we failed to recover support for per transcriptome gene expression levels that significantly deviate from *in silico* midparent values (H^a: 1.9%; H^g: 2.4%; Fig. 1a,b). While both ILDs of *P. hesperium* did have many non-additively expressed genes—2.4-4.6% of genes significantly differed from the *in silico* midparent value—our results are lower than reported for other allopolyploid taxa. For example, the percentage of non-additive genes in synthetic *Arabidopsis* (Wang *et al.*, 2006), wheat (Chagué, *et al.,* 2010), and *Brassica* (Zhao *et al.,* 2013) allopolyploids ranged from 4.5-7.8%, with even higher percentages in naturally-formed cotton allotetraploids (16.3-18.3%, Yoo, Szadkowski, and Wendel, 2013) and *Brassica napus* (16%; Zhang *et al.,* 2016).

The low percentage of non-additively expressed genes in *P. hesperium* relative to other allopolyploids largely reflects that the vast majority of genes surveyed in *P. hesperium* H^a and *P. hesperium* H^g have per transcriptome expression levels that do not significantly differ from both *P. amorphum* and *P. glycyrrhiza* (79.0% and 80.2%, respectively; Fig. 1c). This lack of differential expression among all three taxa, or "no change", is a special case of additivity or midparent expression (Rapp, Udall, and Wendel, 2009), and may result from low levels of sequence divergence (2-3% divergence; Table S1) between *P. amorphum* and *P. glycyrrhiza*. However, the percentage of differentially expressed genes between *P. amorphum* and *P. glycyrrhiza* is less than that recovered from comparable studies of natural allopolyploid angiosperm complexes with similar levels of divergence. For example, the diploid progenitors of *Gossypium* and *Coffea* allopolyploids exhibit differential expression in 42-53% and 36-55% of genes, respectively (Rapp, Udall, and Wendel, 2009; Bardil *et al.,* 2011; Yoo, Szadkowski, and Wendel, 2013). This

suggests that in *Polypodium* relative gene expression (i.e., per transcriptome expression) is largely conserved both between divergent diploid taxa, as well as following allopolyploidization, and seemingly at levels much higher than reported for angiosperm allopolyploid complexes.

Unbalanced expression level dominance in P. hesperium.—Expression level dominance, the condition in which the expression of a gene in an allopolyploid mirrors that of one of its progenitor species but not the other, occurs in many taxa including *Brassica* (Zhang *et al.* 2016), *Spartina* (Chelaifa, Monnier, and Ainouche, 2010), cotton (Rapp, Udall, and Wendel, 2009; Flagel and Wendel, 2010), and wheat (Chagué *et al.,* 2010; Li *et al.,* 2014). Here we report expression level dominance for 7.7% and 8.8% of the genes surveyed in *P. hesperium* H^a and *P. hesperium* H^g, respectively (Fig. 1c, categories II, IV, IX, and XI). Despite substantial differences in the identity of specific genes exhibiting expression level dominance between the IDLs (Fig. 1c), both *P. hesperium* H^a and *P. hesperium* H^g exhibit per transcriptome expression levels that almost exclusively mirror those of *P. amorphum*. This is particularly true for genes that are highly expressed by *P. amorphum* relative to *P. glycyrrhiza*. Strikingly, we recovered extremely low numbers of genes exhibiting transgressive expression in *P. hesperium* H^a and *P. hesperium* H^g (Fig. 1c, categories III, V–VIII, X). Hence, genes that are differentially expressed between *P. hesperium* and one progenitor almost exclusively exhibit unbalanced expression-level dominance biased toward *P. amorphum* (85% for H^a and 98% for H^g). Put differently, both *P. hesperium* H^a and *P. hesperium* H^g have per transcriptome gene expression patterns more similar to those of *P. amorphum,* suggesting that *P. amorphum* is the dominant progenitor dictating gene expression in *P. hesperium* (Rapp, Udall, and Wendel, 2009; Grover *et al.,* 2012; Buggs *et al.,* 2014). This result is in line previous studies that report that the dominant progenitor is commonly the more highly expressed (Steige and Slotte, 2016; Bottani *et al.* 2018)—approximately 2.7× more genes are up-regulated in *P. amorphum* than in *P. glycyrrhiza* (Fig. 1a, b).

The extent of expression level dominance reported here is unusual. We can find only one example in which the proportion of genes exhibiting expression level dominance exceeds 25% (Wu *et al.* (2018) reports 74% in resynthesized *Brassica napus*) and no other examples with so few occurrences of transgressive regulation (e.g., Rapp *et al.,* 2009; Chagué *et al.,* 2010; Chelaifa *et al.,* 2010; Flagel & Wendel 2010; Bardil *et al.,* 2011; Yoo, Szadkowski, and Wendel, 2013; Cox *et al.,* 2014; Zhang *et al.,* 2016; Wu *et al.,* 2018). Surprised by these findings and impeded by a general lack of knowledge about genome evolution and regulation in ferns, it is unclear why *P. hesperium* has such high expression level dominance relative to most angiosperm allopolyploids. *Polypodium hesperium,* and perhaps ferns in general, may exhibit extreme manifestations of the factors thought to promote expression level dominance (see discussion below).

While the degree of expression level dominance reported here is intriguing, it is necessary to include an important caveat when interpreting these results. We report expression levels for each individual normalized to its tran-

scriptome size, with expression of individual genes represented as a fraction of total expression (Robinson and Oshlack, 2010). While transcriptome-normalized expression is standard for studies using RNA-Seq to explore gene expression in allopolyploids relative to their diploid progenitors (e.g., Shi *et al.*, 2012; Combes *et al.*, 2013; Page *et al.*, 2013; Yoo, Szadkowski, and Wendel, 2013; Cox *et al.*, 2014; Xu *et al.*, 2014; Wu *et al.*, 2018), it ignores the potential for variation in transcription per cell and implicitly assumes that transcriptome size is constant among taxa. Transcriptome-normalized expression reflects transcript concentration and should not automatically be equated with absolute expression (Coate and Doyle, 2010, 2015; Loven *et al.*, 2012; Buggs *et al.,* 2014; Visger *et al.,* 2019). If the total number of transcripts per cell differs between individuals, as may be the case with a polyploid taxon and its diploid progenitors, a gene that does not exhibit differential expression between taxa at the transcriptome-level may exhibit differential expression at the cell-level, or vice versa. Direct quantification of transcription per cell (e.g., by extracting RNA from known a number of cells in combination with spiked-in exogenous RNA; Visger et al., 2019) is necessary to convert expression per transcriptome to absolute expression per cell.

Only two studies of diploid-polyploid changes in gene expression changes have accounted for variation in total transcription per cell: Coate and Doyle (2010) report that the average transcription per cell of the allotetraploid *Glycine dolichocarpa* is 1.4× that of its diploid progenitors, and Visger *et al.* (2019) report that the autotetraploid *Tolmiea menziesii* has average per cell transcription levels 2.1× that of the diploid *T. diplomenziesii*. Hence, doubling of ploidy does not necessarily lead to a doubling of transcription per cell. Transcription per cell may vary widely among taxa, and to the best of our knowledge, no study has ever quantified absolute expression per cell in polyploid and diploid ferns. Interestingly, Visger et al. (2019) reports that 61% of differential expressed genes between *T. menziesii* and *T. diplomenziesii* detected with per cell-normalized expression data were also considered differentially expressed using transcriptome-normalized data. These genes have high magnitudes of per cell differential expression between the diploid and polyploid taxa, suggesting that transcriptome-normalization methods may be adequate for detecting genes with high-levels of absolute differential expression (enough to change the per transcriptome concentrations) but fail to detect genes with low-levels of absolute differential expression.

While we are confident the results of this study are comparable to most studies utilizing transcriptome normalized RNA-Seq data, the results of all these studies must be approached judiciously. In the case of *P. hesperium* and many other RNA-Seq studies of allopolyploid plants using transcriptome-normalized data, the number of genes exhibiting mid-parent expression levels and expression level dominance may be inflated. By utilizing both transcriptome-normalization and per cell-normalization methods in concert, future studies comparing gene expression in diploid and polyploid ferns could provide broader insights into gene dosage responses, the conservation of

expression levels with changes in biomass, and selection on absolute expression levels (Visger *et al.,* 2019; Doyle and Coate, 2019).

Homeolog expression bias in P. hesperium.—To determine how gene copies derived from different progenitors contribute to total gene expression in *P. hesperium*, we examined homeolog-specific expression for 1073 genes in both IDLs of *P. hesperium*. By identifying SNPs that are putatively fixed between *P. amorphum* and *P. glycyrrhiza*, we were able to infer the parental origin of homeologs in *P. hesperium* and calculate their relative expression. Just as we discovered unbalanced expression level dominance with *P. amorphum* as the dominant parent, we recovered unbalanced homeolog expression bias with a disproportionate number of genes in *P. hesperium* preferentially expressing *P. amorphum*-derived homeologs (Fig. 2). It is important to note that there are differences in the degree and direction of homeolog expression bias between the IDLs. For example, *P. hesperium* H^a has more genes with homeolog expression bias than *P. hesperium* H^g (1023 vs. 783), but *P. hesperium* H^g has a larger average magnitude of bias favoring *P. amorphum*-derived homeologs (12-fold vs. 8-fold in H^a, Fig. 2a, b). Furthermore, at the level of specific genes, *P. hesperium* H^g tends to express more *P. glycyrrhiza*-derived homeologs (Fig. 2c). While it appears that all individuals of *P. hesperium*, regardless origin, generally exhibit a strong bias for the expression of *P. amorphum*-derived homeologs across the transcriptome, statistically significant differences in homeolog expression bias between the IDLs of *P. hesperium* for individual genes (Fig. 2f), suggest that homeolog expression in these genes may be dictated by lineage-specific factors.

Preferential expression of homeologs derived from a particular progenitor is commonly observed in other allopolyploids (e.g., *Arabidopsis*: Wang *et al.,* 2006; *Coffea*: Combes *et al.,* 2013; *Glycine*: Coate, Bar, and Doyle, 2014; *Gossypium*: Chaudary *et al.,* 2009; Flagel and Wendel, 2009, 2010; Rapp, Udall, and Wendel, 2009; *Mimulus:* Edger *et al.,* 2017*; Tragopogon*: Tate *et al.,* 2006; Buggs *et al.,* 2010; Koh, Soltis, and Soltis, 2010; *Zea:* Schnable, Springer, and Freeling, 2011), and appears to be a prevalent consequence of allopolyploidization. Once again though, the percentage of genes exhibiting significant homeolog expression bias in *P. hesperium* is greater than in most surveyed allopolyploids. To the best of our knowledge, *Tragopogon miscellus* and *T. mirus* are the only other allopolyploids with IDLs for which homeolog-specific expression has been investigated. As in *P. hesperium*, all IDLs of both *T. miscellus* and *T. mirus* preferentially express homeologs derived from the same progenitor (Tate *et al.,* 2006, 2009; Koh, Soltis, and Soltis, 2010; Buggs *et al.,* 2009, 2010). These findings suggest that recurrent origins of natural allopolyploids consistently yield similar transcriptome-wide patterns of homeolog expression.

Possible mechanisms for expression level dominance and homeolog expression bias in P. hesperium.—Both expression level dominance and homeolog expression bias are broadly observed phenomena, prompting much recent interest in underlying genomic and regulatory mechanisms (e.g., Hollister and Gaut, 2009; Freeling *et al.,* 2012; Steige and Slotte, 2016; Bird *et al.,* 2018; Bottani *et al.,* 2018; Wendel *et al.,* 2018). While little is known

about gene regulation and genome evolution in ferns (Sessa *et al.*, 2014; Sessa and Der, 2016), mechanisms in angiosperm allopolyploids may help explain patterns seen in *P. hesperium*. The prevailing explanation for expression biases in allopolyploids is a difference in the number and distribution of transposable elements (TEs) in their progenitor-derived subgenomes. TEs inserted in or near genes become the targets of RNA-directed DNA methylation and heterochromatic modification, causing the repression of gene expression (Parisod *et al.,* 2010; Vicient and Casacuberta, 2017). Furthermore, it has been demonstrated in *Arabidopsis* allopolyploids that TE silencing leads to *cis*-regulatory variation (Shi *et al.,* 2012). Under this scenerio, the subgenome with the lower density of TEs would be more likely to have higher gene expression and be the dominant progenitor genome dictating total and homeolog-specific gene expression levels in an allopolyploid (Bottani *et al.,* 2018). Ultimately, the homeologs from the repressed progenitor subgenome may become lost, resulting in biased fractionation of the allopolyploid genome (Steige and Slotte, 2016). Empirical evidence for the role of TE mediated gene expression biases has been found in allopolyploids and F1 hybrids of *Brassica* (Woodhouse *et al.*, 2014) and *Zea* (Pophaly and Tellier, 2015). However, little evidence was found for the role of TE gene silencing in *Capsella* (Steige *et al.,* 2016) and *Gossypium* (Renny-Byfield *et al.,* 2015), suggesting that other processes may contribute to expression biases.

Looking ahead, *P. hesperium* offers an ideal opportunity to test for a correlation between the divergence of TE abundance and distribution between subgenomes and expression biases in allopolyploid ferns. Specifically, we hypothesize that its *P. glycyrrhiza*-derived subgenome will harbor a greater TE load than its *P. amorphum*-derived genome, particularly near genes exhibiting preferential expression of *P. amorphum*-derived homeologs. By surveying the genomes of *P. amorphum*, *P. glycyrrhiza,* and both IDLs of *P. hesperium*, we may be able to determine if the abundance and distribution of TEs in the *P. hesperium* subgenomes reflects preexisting differences in its progenitor species or is due to the expansion of TEs following allopolyploidization, as has been observed for some angiosperms (e.g., Vicient and Casacuberta, 2017; Mhiri *et al.,* 2018; Palacios *et al.,* 2019). Furthermore, we will be able to assess the fidelity of the underlying mechanisms resulting in the similar transcriptome-scale patterns of gene expression between *P. hesperium* H^a and *P. hesperium* H^g reported here.

Acknowledgements

The authors thank E. Schuettpelz and A. Grusz for their suggestions on this manuscript. We are particularly grateful to C. dePamphilis and P. Ralph for building the RNA-Seq libraries for this project. This research was funded by NSF DEB-1110775 to K. M. Pryer and E. M. Sigel.

Literature Cited

Adams, K. L. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. Journal of Heredity 98:136–141.

Akama, S., R. Shimizu-Inatsugi, K. K. Shimizu, and J. Sese. 2014. Genome-wide quantification of homoeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. Nucleic Acids Research 42.6:e46.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Ashburner, M., C.A Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, and M. A. Harris. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics, 25:25–29.

Bardil, A., J. Dantas de Almeida, M.-C. Combes, P. Lashermes, and B. Bertrand. 2011. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytologist 192:760–774.

Barker, M. S., and P. G. Wolf. 2010. Unfurling fern biology in the genomics age. Bioscience 60:177–185.

Barrington, D. S., C. H. Haufler, and C. R. Werth. 1989. Hybridization, reticulation, and species concepts in the ferns. American Fern Journal 79:55–64.

Benjamini Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B 57:289–300.

Bird, K.A., R. VanBuren, J. R. Puzey, and P. P. Edger. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytologist 220:87–93.

Birky, C. W. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences, U.S.A. 92:11331–11338.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borodina, T., J. Adjaye, and M. Sultan. 2011. A strand-specific library preparation protocol for RNA sequencing. Pp. 79-98, *in* D. Jameson, M. Verma, and H. V. Westerhoff (eds.), *Methods in Enzymology. Volume 500*. Academic Press, San Diego CA.

Bottani, S., N. R. Zabet, J. F. Wendel, and R. A. Veitia. 2018. Gene expression dominance in allopolyploids: hypotheses and models. Trends in Plant Science 23:393–402.

Buggs, R. J. A., A. N. Doust, J. A. Tate, J. Koh, K. Soltis, F. A. Feltus, A. H. Paterson, P. S. Soltis, and D. E. Soltis. 2009. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103:73–81.

Buggs, R. J. A., S. Chamala, W. Wu, L. Gao, G. D. May, P. S. Schnable, D. E. Soltis, P. S. Soltis, and W. B. Barbazuk. 2010. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Molecular Ecology 19:132–146.

Buggs, R. J., J. F, Wendel, J. J. Doyle, D. E. Soltis, P. S. Soltis, and J. E. Coate. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical Transactions of the Royal Society B 369:20130354.

Chagué, V., J. Just, I. Mestiri, S. Balzergue, A.-M. Tanguy, C. Huneau, V. Huteau, H. Belcram, O. Coriton, J. Jahier, and B. Chaloub. 2010. Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New Phytologist 187:1181–1194.

Chaudhary, B., L. Flagel, R. M. Stupar, J. A. Udall, N. Verma, N. M. Springer, and J. F. Wendel. 2009. Reciprocal silencing, transcriptional bias and functional divergence of homoeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517.

Chelaifa, H., A. Monnier, and M. Ainouche. 2010. Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). New Phytologist 186:161–174.

Chester, M., J. P. Gallagher, V. Vaughan Symonds, A. V. Cruz da Silva, E. V. Mavrodiev, A. R. Leitch, P. S. Soltis, and D. E. Soltis. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proceedings of the National Academy of Sciences, U.S.A. 109:1176–1181.

Clark, J.W. and P. C. Donoghue. 2018. Whole-genome duplication and plant macroevolution. Trends in Plant Science 23:933–945.

Coate, J.E. and J. J. Doyle. 2010. Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. Genome Biology and Evolution 2:534–546.

Coate, J.E. and J. J. Doyle. 2015.Variation in transcriptome size: are we getting the message? Chromosoma 124:27–43.

Coate, J. E., A. K. Luciano, V. Seralathan, K. J. Minchew, T. G. Owens, and J. J. Doyle. 2012. Anatomical, biochemical, and photosynthetic responses to recent allopolyploidy in *Glycine dolichocarpa* (Fabaceae). American Journal of Botany 99:55–67.

Coate, J.E., H. Bar, and J. J. Doyle. 2014. Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). The Plant Cell 26:136–150.

Comai, L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6:836–846.

Combes, M.-C., A. Dereeper, D. Severac, B. Bertrand, and P. Lashmeres. 2013. Contribution of subgenomes to the transcriptome and their intertwined regulation in the *allopolyploid Coffea arabica* grown at contrasted temperatures. New Phytologist 200:251–260.

Cox, M. P., T. Dong, G. Shen, Y. Davi, D. B. Scott, and A. R. D. Ganley. 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLOS Genetics 10:e1004180.

DePristo, M., E. Banks, R. Poplin, K. Garimella, J. Maguire, C. Hartl. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. Fennell, *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491–498.

Dillies, M.-A., A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, *et al.* 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings in Bioinformatics 14:671–683.

Doyle, J. J. and J. E. Coate. 2019. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. International Journal of Plant Sciences 180:1–52.

Edger, P.P., R. D. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin, Y. W. Yuan, A. J. Bewick, L. Ji, A. E. Platts, M. J. Bowman, and K. Childs. 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell 29:2150–2167.

Edger, P.P., M. R. McKain, K. A. Bird, and R. VanBuren. 2018. Subgenome assignment in allopolyploids: challenges and future directions. Current Opinion in Plant Biology 42:76–80.

Flagel, L. E. and J. F. Wendel. 2009. Gene duplication and evolutionary novelty in plants. New Phytologist 183:557–564.

Flagel, L. E. and J. F. Wendel. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytologist 186:184–193.

Freeling, M., M. R. Woodhouse, S. Subramaniam, G. Turco, D. Lisch, and J. C. Schnable. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Current Opinion in Plant Biology 15:131–139.

Gaeta, R. T. and J. C. Pires. 2009. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist 186:18–28.

Gallardo, M. H., G. Kausel, A. Jiménez, C. Bacquet, C. González, J. Figueroa, N. Köhler, and R. Ojeda. 2004. Whole-genome duplications in South American desert rodents (Octodontidae). Biological Journal of the Linnean Society 82:443–451.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Maucell, *et al.* 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29:644–652.

Grover, C. E., J. P. Gallagher, E. P. Szadkowski, M. J. Yoo, L. E. Flagel, and J. F. Wendel. 2012. Homoeolog expresssion bias an expression level dominance in allopolyploids. New Phytologist 196:966–971.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, *et al.* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8:1494–512.

Haufler, C. H., and D. E. Soltis. 1986. Genetic evidence suggests that homosporous ferns with high chromosome numbers are diploid. Proceedings of the National Academy of Sciences, U.S.A. 83:4389–4389.

Hollister, J. D., and B. S. Gaut. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research 19:1419–1428.

Hu, G. and J. F. Wendel. 2018. Cis–trans controls and regulatory novelty accompanying allopolyploidization. New Phytologist. doi:10.1111/nph.15515

Iseli, C., C. V. Jongeneel, and P. Bucher. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proceedings of the International Conference on Intelligent Systems for Molecular Biology 99:138–148.

Jiang, J., Y. Wang, B. Zhu, T. Fang, Y. Fang, and Y. Wang. 2015. Digital gene expression analysis of gene expression differences within *Brassica* diploids and allopolyploids. BMC Plant Biology 15:22.

Kashkush, K., M. Feldman, and A. A. Levy. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659.

Koh, J., P. S. Solitis, and D. E. Soltis. 2010. Homoeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics 11:97–121.

Krasileva, K. V., V. Buffalo, P. Bailey, S. Pearce, S. Ayling, F. Tabbita, M. Soria, S. Wang, IWGS Consortium, E. Akhunov, C. Uauy, and J. Dubcovsky. 2013. Separating homoeologs by phasing in the tetraploid wheat transcriptome. Genome Biology 14:R66.

Langmead B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Levy, A. A. and M. Feldman. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biological Journal of the Linnean Society 82:607–613.

Li, A., D. Liu, J. Wu, X. Zhao, M. Hao, S. Geng, J. Yan, X. Jiang, L. Zhang, J. Wu, and L. Yin. 2015 mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. The Plant Cell 5:1878–900.

Li, B. and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics 12:323–338.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13:2178–2189.

Lovén, J., D. A. Orlando, A. A. Sigova, C. Y. Lin, P. B. Rahl, C. B. Burge, D. L. Levens, T. I. Lee, and R. A. Young. 2012. Revisiting global gene expression analysis. Cell 151:476–482.

Lukens, L., J. Pires, E. Leon, and R. Vogelzang. 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. Plant Physiology 140:336–348.

Madlung, A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110:99–104.

MATLAB. 2019. version 9.6.0 (R2019a). Natick, Massachusetts: The MatWorks Inc.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303.

Meimberg, H., K. J. Rice, N. F. Milan, C. C. Njoku, and J. K. McKay. 2009. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). American Journal of Botany 96:1262–1273.

Mhiri, C., C. Parisod, J. Daniel, M. Petit, K. Y. Lim, F. Dorlhac de Borne, A. Kovarik, A. R. Leitch, and M. A. Grandbastien. 2018. Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids. New Phytologist 221:1619–1633.

Mithani, A., E. J. Belfield, C. Brown, C. Jiang, L. J. Leach, and N. P. Harberd. 2013. HANDS: a tool for genome-wide discovery of subgenome-specific base-identity in polyploids. BMC Genomics 14:653–660.

Moghe, G. D., D. E. Hufnagel, H. Tang, Y. Xiao, I. Dworkin, C. D. Town, J. K. Conner, and S. H. Shiu. 2014. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. The Plant Cell 26:1925–1937.

Nagy I., S. Barth, J. Mehenni-Ciz, M. T. Abberton, and D. Milbourne. 2013. A hybrid next generation transcript sequencing-based approach to identify allelic and homoeolog-specific single nucleotide polymorphisms in allotetraploid white clover. BMC Genomics 14:100–118.

Otto, S. P. 2007. The evolutionary consequences of polyploidy. Cell 131:452–462.

Page, J. T., A. R. Gingle, and J. A. Udall. 2013. PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3 3:517–525.

Palacios, P. M., M.-P. Jacquemot, M. Tapie, A. Rousselet, M. Diop, C. Remoue, M. Falque, A. Lloyd, E. Jenczewski, G. Lassalle, A.-M. Chevre, C. Lelandais, *et al* 2019. Assessing the response of small RNA populations to allopolyploidy using resynthesized *Brassica napus* allotetraploids. Molecular Biology and Evolution doi: 10.1093/molbev/msz007.

Parisod, C., K. Alix, J. Just, M. Petit, V. Sarilar, C. Mhiri, M. Ainouche, B. Chalhoub, and M. A, Grandbastien. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytologist 186:37–45.

Pophaly, S. D. and A. Tellier. 2015. Population level purifying selection and gene expression shape subgenome evolution in maize. Molecular Biology and Evolution 32:3226–3235.

Ramsey, J. 2011. Polyploidy and ecological adaptation in wild yarrow. Proceedings of the National Academy of Sciences, U.S.A. 108:7096–7101.

Rapp, R. A., J. A. Udall, and J. F. Wendel. 2009. Genomic expression dominance in allopolyploids. BMC Biology 7:18.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reboud, X. and C. Zeyl. 1994. Organelle inheritance in plants. Heredity 72:132–140.

Renny-Byfield, S., L. Gong, J. P. Gallagher, and J. F. Wendel. 2015. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. Molecular Biology and Evolution 32:1063–1071.

Robinson, M. D. and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-Seq data. Genome Biology 11:R25.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2:139–140.

Salmon, A., M. L. Ainouche, and J. F. Wendel. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Molecular Ecology 14:1163–1175.

Salmon, A., L. Flagel, B. Ying, J. A. Udall, and J. F. Wendel. 2010. Homoeologous nonreciprocal recombination in polyploid cotton. New Phytologist 186:123–134.

Schnable, J. C., N. M. Springer, and M. Freeling. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences, U.S.A. 108:4069–4074.

Sémon, M. and K. H. Wolfe. 2007. Consequences of genome consequences. Current in Genetics and Development 17:505–512.

Senerchia, N., F. Felber, and C. Parisod. 2014. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. New Phytologist 202:975–985.

Sessa, E. B., J. A. Banks, M. S. Barker, J. P. Der, A. M. Duffy, S. W. Graham, M. Hasebe, J. Langdale, F.-W. Li, D. B. Marchant, K. M. Pryer, C. J. Rothfels, *et al*. 2014. Between two fern genomes. GigaScience 3:15.

Sessa, E. B. and J. P. Der. 2016. Evolutionary genomics of ferns and lycophytes. Pp. 215–254, *in* S. A. Rensing (ed.), *Genomes and Evolution of Charophytes, Bryophytes, and Ferns. Advances in Botanical Research, Volume78*. Academic Press, London.

Shi, X., D. W. Ng, C. Zhang, L. Comai, W. Ye, and Z. J. Chen. 2012. Cis-and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nature Communications 3:950.

SIGEL, E. M., M. D. WINDHAM, and K. M. PRYER. 2014. Evidence for reciprocal origins in *Polypodium hesperium* (Polypodiaceae): a fern model system for investigating how multiple origins shape allopolyploid genomes. American Journal of Botany 101:1476–1485.

SIGEL, E. M., M. D. WINDHAM, C. HAUFLER, and K. M. PRYER. 2014. Phylogeny, divergence time estimates, and phylogeography for the diploid species of the *Polypodium vulgare* complex (Polypodiaceae). Systematic Botany 39:1042–1055.

SIGEL, E. M. 2016. Genetic and genomic aspects of hybridization in ferns. Journal of Systematics and Evolution 54:638–655.

SIGEL, E. M., E. SCHUETTPELZ, K. M. PRYER, and J. P. DER. 2018. Overlapping patterns of gene expression between gametophyte and sporophyte phases in the fern *Polypodium amorphum* (Polypodiales). Frontiers in Plant Sciences 9:1450.

SIMÃO, F.A., R. M. WATERHOUSE, P. IOANNIDIS, E. V. KRIVENTSEVA, and E. M. ZDOBNOV. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31:3210–3212.

SOLTIS, D. E. and P. S. SOLTIS. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology and Evolution 14:348–351.

SOLTIS, P. S. and D. E. SOLTIS. 2000. The roles of genetic and genomic attributes in success of polyploids. Proceedings of the National Academy of Sciences, U.S.A. 97:7051–7057.

SMITH, R. D., T. J. KINSER, G. D. C. SMITH , and J. R. PUZEY. 2019. A likelihood ratio test for changes in homeolog expression bias. BMC Bioinformatics 20:149.

STEBBINS, G. L. 1950. Variation and Evolution in Plants. New York: Columbia University Press.

STEIGE, K.A. and T. SLOTTE. 2016. Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. Current Opinion in Plant Biology 30:88–93.

STEIGE, K. A., J. REIMEGÅRD, C. A. REBERNIG, C. KÖHLER, D. G. SCOFIELD, and T. SLOTTE. 2016. The role of transposable elements for gene expression in *Capsella* hybrids and allopolyploids. bioRxiv doi: 10.1101/044016.

TATE, J. A., Z. NI, A. C. SCHEEN, J. KOH, C.A, GILBERT, D. LEFKOWITZ, Z. J. CHEN, P. S. SOLTIS, and D. E. SOLTIS. 2006. Evolution and expression of homoeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173:1599–1611.

TATE, J. A., P. JOSHI, K. A. SOLTIS, P. S. SOLTIS, and D. E. SOLTIS. 2009. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biology 9:80–89.

THOMAS, B. C., B. PEDERSEN, and M. FREELING. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homoeolog leaving clusters enriched in dose-sensitive genes. Genome Research 16:934–946.

VAN DER AUWERA, G. A., M. CARNEIRO, C. HARTL, R. POPLIN, G. DEL ANGEL, A. LEVY-MOONSHINE, T. JORDAN, K. SHAKIR, D. ROAZEN, J. THIBAULT, E. BANKS, K. GARIMELLA, *et al.* 2013. From fastq data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics 43:11.10.1–11.10.33.

VICIENT, C. M. and J. M. CASACUBERTA. 2017. Impact of transposable elements on polyploid plant genomes. Annals of Botany 120:195–207.

VISGER, C. J., WONG, G. K. S., ZHANG, Y., SOLTIS, P. S. and D. E. SOLTIS. 2019. Divergent gene expression levels between diploid and autotetraploid *Tolmiea* relative to the total transcriptome, the cell, and biomass. American Journal of Botany, 106: 280–291.

WALL, P. K., J. LEEBENS-MACK, K. F. MÜLLER, D. FIELD, N. S. ALTMAN, and C. W. DEPAMPHILIS. 2008. PlantTribes: a gene and gene family resource for comparative genomics in plants. Nucleic acids research 36:D970–D976.

WANG, J., L. TIAN, H.-S. LEE, M. E. WEI, H. JIANG, B. WATSON, A. MADLUNG, T. C. OSBORN, R. W. DOERGE, L. COMAI, and Z. J. CHEN. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517.

WENDEL, J.F., D. LISCH, G. HU, and A. S. MASON. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current Opinion in Genetics and Development 49:1–7.

WERTH, C.R. and D. B. LELLINGER. 1992. Genomically preserved plants and their nomenclature. Taxon 41:513–521.

Wolfe, K. H. and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–712.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences, U.S.A. 106:13875–13879.

Woodhouse, M. R., F. Cheng, J. C. Pires, D. Lisch, M. Freeling, and X. Wang. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proceedings of the National Academy of Sciences, U.S.A. 111:5283–5288.

Wu, J., L. Lin, M. Xu, P. Chen, D. Liu, Q. Sun, L. Ran, and Y. Wang. 2018. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics 19:586.

Yoo, M.-J., E. Szadkowski, and J. F. Wendel. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110:171–180.

Zhang, D., Q. Pan, C. Tan, B. Zhu, X. Ge, Y. Shao, and Z. Li. 2016. Genome-wide gene expressions respond differently to A-subgenome origins in *Brassica napus* synthetic hybrids and natural allotetraploid. Frontiers in Plant Science 7:1508.

Zhao, Q., J. Zou, J. Meng, S. Mei, and J. Wang. 2013. Tracing the transcriptomic changes in synthetic Trigenomic allohexaploids of *Brassic*a using an RNA-Seq approach. PLoS ONE 8:e68883.

Supplementary Materials

All supplementary files, table, and figures are archived on FigShare, https://doi.org/10.6084/m9.figshare.c.4388981.v1

File S1. MATLAB scripts for homeolog expression bias likelihood ratio tests.

File S2. Final *Polypodium* reference transcriptome, nucleotide coding sequences.

File S3. Final *Polypodium* reference transcriptome, peptide translations of coding sequences.

File S4. Variant call file for all 12 *Polypodium* individuals.

Table S1. Estimate of genetic distance between *P. amorphum* and *P. glycyrrhiza* based on coding regions of 24 single copy nuclear loci.

Table S2. Summary of the number of Illumina 100bp paired-end reads for each *Polypodium* individual.

Table S3. Summary statistics for the *Polypodium* reference transcriptome.

Table S4. Details of orthogroup membership for *Polypodium* transcripts corresponding to the PlantTribes gene family classification.

Table S5. Read counts for 19194 genes used for differential expression analyses.

Table S6. Results of likelihood ratio tests for homeolog expression bias for 1073 genes in *Polypodium hesperium* H^a.

Table S7. Results of likelihood ratio tests for homeolog expression bias for 1073 genes in *Polypodium hesperium* H^g.

Table S8. Results of likelihood ratio tests for differences in homeolog expression bias between *Polypodium hesperium* H^a and *P. hesperium* H^g for 1073 genes.

Fig. S1. BUSCO scores for the final *Polypodium* transcriptome assembly.

Fig. S2. The number of genes differentially expressed in each contrast between the allotetraploid *Polypodium hesperium* and its diploid progenitors,

P. amorphum and *P. glycyrrhiza,* at a threshold of log2 fold change $\geq$ 2 and FDR $\leq$ 0.01.

FIG. S3. The number of genes belonging to 12 categories of differential expression among the allotetraploid *P. hesperium* and its diploid progenitors, *P. amorphum* and *P. glycyrrhiza,* at a threshold of log2 fold change $\geq$ 2 and FDR $\leq$ 0.01.