Evidence for reciprocal origins in *Polypodium hesperium* (Polypodiaceae): A fern model system for investigating how multiple origins shape allopolyploid genomes

**ERIN M. SIGEL**, **MICHAEL D. WINDHAM**, and **KATHLEEN M. PRYER**

Department of Biology, Duke University, Durham, North Carolina 27708 USA

- **Premise of the study:** Many polyploid species are composed of distinct lineages originating from multiple, independent polyploidization events. In the case of allopolyploids, reciprocal crosses between the same progenitor species can yield lineages with different uniparentally inherited plastid genomes. While likely common, there are few well-documented examples of such reciprocal origins. Here we examine a case of reciprocal allopolyploid origins in the fern *Polypodium hesperium* and present it as a natural model system for investigating the evolutionary potential of duplicated genomes.

- **Methods:** Using a combination of uniparentally inherited plastid and biparentally inherited nuclear sequence data, we investigated the distributions and relative ages of reciprocally formed lineages in *Polypodium hesperium*, an allotetraploid fern that is broadly distributed in western North America.

- **Key results:** The reciprocally derived plastid haplotypes of *Polypodium hesperium* are allopatric, with populations north and south of 42°N latitude having different plastid genomes. Incorporating biogeographic information and previously estimated ages for the diversification of its diploid progenitors, we estimate middle to late Pleistocene origins of *P. hesperium*.

- **Conclusions:** Several features of *Polypodium hesperium* make it a particularly promising system for investigating the evolutionary consequences of allopolyploidy. These include reciprocally derived lineages with disjunct geographic distributions, recent time of origin, and extant diploid progenitors.

**Key words:** allopolyploidy; biogeography; gapCp; multiple origins; plastid haplotype; Polypodiaceae; *Polypodium vulgare* complex.

Polypodioid, or whole genome duplication, played a critical role in the evolution of entire lineages, from fungi to insects and even vertebrates (Otto and Whitton, 2000; Mable, 2004; Albertin and Marullo, 2012). This is particularly true among plants, where it is thought that increases in ploidy level have been involved in ca. 15% and 31% of angiosperm and fern speciation events, respectively (Wood et al., 2009), and that all vascular plants have experienced one or more ancient polyploidization events (Cui et al., 2006; Otto, 2007; Jiao et al., 2011). One particularly intriguing aspect of polyploid species is that many form recurrently, resulting in species composed of multiple, independently derived lineages (henceforth IDLs; Soltis and Soltis, 1993, 1999). In the case of allopolyploid species—polyploids resulting from hybridization between two or more distinct species—IDLs represent the repeated union of divergent genomes (Werth et al., 1985). It has been hypothesized that the increased genetic variation imparted by multiple origins increases the ecological amplitude, geographic range, and evolutionary success of allopolyploid lineages (Meimberg et al., 2009; Madlung, 2013).

Recurrent origins of allopolyploids can result from reciprocal crosses between the same progenitor species, resulting in IDLs with different maternal parents. Although multiple origins are common among polyploid species, there are relatively few, well-documented cases of reciprocal origins (Soltis and Soltis, 1993). The list is growing, however, with examples now known in *Tragopogon* L. (Soltis and Soltis, 1989; Soltis et al., 2004), *Polypodium* L. (Hauffer et al., 1995b), *Platanthera* Rich. (Wallace, 2003), *Senecio* L. (Kadereit et al., 2006), *Aegilops* L. (Meimberg et al., 2009), *Androsace* L. (Dixon et al., 2009), *Asplenium* L. (Perrie et al., 2010; Chang et al., 2013), *Astrolepis* D.M. Benham & Windham (Beck et al., 2010), *Pteris* L. (Chao et al., 2012), and *Nephrolepis* Schott (Kao et al., 2014). Among the cases documented so far is the western North American sexual allotetraploid *Polypodium hesperium* Maxon, a member of the well-studied and morphologically cryptic *P. vulgare* reticulate complex (Fig. 1; summarized by Hauffer et al., 1995b; Sigel et al., 2014b). This species has a broad, but disjunct montane distribution extending from southern British Columbia to the northern Baja Peninsula, east to Montana, the...
In this study, we use nuclear and plastid sequence data to revisit the IDLs of *Polypodium hesperium*. Our goals are to (1) document the geographic distribution of plastid haplotypes within *P. hesperium*; (2) assess the number and relative ages of IDLs within *P. hesperium*; and (3) introduce *P. hesperium* as a natural model system for investigating the evolutionary consequences of allopolyploidy in ferns.

**MATERIALS AND METHODS**

**Taxon sampling**—We sampled 100 specimens of *Polypodium* representing 51 individuals of *P. hesperium*, 10 diploid taxa of the *P. vulgar* complex, and five species belonging to the more distantly related *P. pleiosorum* group (Table 1; Appendix 1). The latter is sister to the *P. vulgar* complex (Sigel et al., 2014b), and its representatives were selected without regard to ploidy because of a lack of cytogenetic data for most members of this lineage. *Pleurosoriopsis makinoi* (Maxim.) Fomin, an uncommon east Asian taxon reported as both triploid and tetraploid (Kuriya and Hasebe, 1977; Iwatsuki et al., 1995), was used as the outgroup based on its position as the closest relative to a clade including the *Polypodium vulgar* complex + *P. pleiosorum* group in recent molecular studies (Schneider et al., 2004; Otto et al., 2009; Sigel et al., 2014b).

**DNA extraction, PCR amplification, cloning, and sequencing**—For each individual sampled, total genomic DNA was isolated from herbarium, silica-dried, or fresh material using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) following the manufacturer’s protocol. The *trnG-trnR* intergenic spacer (hereafter *trnG-R*) was either amplified from the plastid genome and sequenced for each specimen included in this study (Table 1; Appendix 1). Universal *trnG-F* fern primers (TRNG1F, TRNG43F, TRNG63R, and TRNR22R) and protocols used for amplification and sequencing followed Nagalingum et al. (2007). A *trnG-R* sequence was assembled from chromatograms for each specimen using the program Sequencher 4.8 (Gene Codes Corp., Ann Arbor, Michigan, USA).

A portion of the low-copy nuclear locus *gapCp* was amplified, cloned, and sequenced for a subset of 36 specimens included in this study (Table 1; Appendix 1). Sequences for the outgroup taxon, *Pleurosoriopsis makinoi*, were amplified using previously published universal *gapCp* primers, ESGACP8F1 and ESGACP11R1, and protocols (Schuettpelz et al., 2008). Two copies of *gapCp* were recovered in Polypodioideae—one ~900-bp copy and one ~600-bp copy (Rothels et al., 2013); only the ~600-bp copy was sequenced. The need to efficiently separate the two copies of *gapCp* prompted the development of a *Polypodium specific* *gapCp* “short” forward primer, EMgapF1 (5′-GGTGGTCGCTAAGGTTGACAAC-3′). When used with ESGACP11R1, this new forward primer amplified only *gapCp* “short” (hereafter *gapCp*). Amplification, cloning, and sequencing of *gapCp* followed the protocols in Schuettpelz et al. (2008) with the following exceptions: cloning was performed using pGEM-T cloning kits (Promega, Madison, Wisconsin, USA), and two to three separate amplifications per individual were pooled prior to ligation. This pooling was done to mitigate PCR bias occurring in any single amplification (Polz and Cavanaugh, 1998; Beck et al., 2010). A minimum of 16 colonies was amplified for each individual, and sequences were obtained for 8–14 colonies for each specimen sampled.

To minimize variation due to PCR error in the *gapCp* sequences, we filtered the raw data for genetic variants (putative alleles) using the approach described by Grusz et al. (2009). First, a contig sequence was assembled for each clone from chromatograms using the program Sequencher 4.8 (Gene Codes Corp., Ann Arbor, Michigan, USA). Then, all clone sequences obtained from a given individual were combined into a single alignment using Sequencher 4.8. Each alignment was visually inspected and obvious chimeric sequences were removed. A maximum parsimony (MP) tree of each alignment was constructed using PAUP* v. 4.0 a123 (Swoford, 2002). The resulting tree was used to determine the alleles present in an individual by identifying groups of clones united by two or more polymorphisms. A consensus sequence of each group of clones was constructed in Sequencher 4.8, with the resulting consensus sequences representing alleles belonging to a given individual.

**Sequence alignment and data sets**—Separate alignments for *trnG-R* and *gapCp* were manually generated using the program MacClade 4.05 (Maddison and Maddison, 2005). Unsequenced portions of each locus were coded as missing data, and portions of the 5′ and 3′ regions with large amounts of missing data were excluded. Indels due to insertion or deletion events (present in both

---

![Image](58x545 to 302x718)

**Fig. 1.** Summary of relationships among the diploid and allopolyploid taxa of the *Polypodium vulgar* complex as reconstructed by a phylogenetic analysis of a four loci plastid sequence data set (Sigel et al., 2014b) and analysis of isozyme banding patterns (Haufler et al., 1995a). The monophyletic *P. vulgar* complex comprises four clades of diploid species: the *P. appalachianum* clade (A), the *P. glycyrrhiza* clade (G), the *P. cambricum* clade (C), and the *P. scouleri* clade (S). Thicker branches and an asterisk (*) indicate Bayesian inference posterior probability (BIPP) and maximum likelihood bootstrap support (MLBS) values (see inset legend). Dashed lines connect allopolyploid taxa to their progenitor diploid species. Bolded taxon names and dashed lines highlight *P. amorphum* (clade A) and *P. glycyrrhiza* (clade G), the diploid progenitors of the allotetraploid species *P. hesperium*. Numbers in parentheses indicate the ploidy of the polyploid species.

Four Corners states, and Chihuahua, Mexico (Maxon, 1900; Hauffer et al., 1993; Mickel and Smith, 2004; Yatskievych and Windham, 2009).

On the basis of morphological and cytological data, Lang (1971) hypothesized that *P. hesperium* is an allotetraploid derived from hybridization between the sexual diploids *P. amorphum* Suksd. and *P. glycyrrhiza* D.C. Eaton. The two diploid species belong to distinct lineages within the *P. vulgar* complex (Fig. 1, clades A and G, respectively) that diverged approximately 12 million years ago (Ma; Sigel et al., 2014b), and their present ranges overlap with that of *P. hesperium* only in a limited area of the Pacific Northwest (Hauffer et al., 1993). Lang’s (1971) original hypothesis was supported by an analysis of biparentally inherited nuclear isozyme markers (Hauffer et al., 1995b), and a complementary study utilizing restriction-site mapping of the uniparentally inherited chloroplast genome demonstrated that two specimens of *P. hesperium*, one from eastern Washington and one from Utah, had reciprocal origins (Hauffer et al., 1995a). Presuming maternal inheritance of plastid genomes in ferns (initially demonstrated by Gastony and Yatskievych [1992] and subsequently shown by Vogel et al. [1998] and Guillon and Raquin [2000]), Hauffer et al. (1995a) hypothesized that the maternal parent of the eastern Washington specimen was *P. amorphum*, whereas the maternal parent of the Utah specimen was *P. glycyrrhiza*. Variations in isozyme banding patterns and cryptic morphological characters were proposed as evidence for additional IDLs within *P. hesperium* (Hauffer et al., 1995b).

Recent divergence date estimates for the *P. vulgar* complex (Sigel et al., 2014b) suggest that *P. amorphum* and *P. glycyrrhiza* both arose during the middle to late Pleistocene and, as such, all IDLs of *P. hesperium* likely formed during the last 1.2 Myr.
Phylogenetic analyses—Separate phylogenetic analyses were conducted for the trnG-R and gapCp data sets using maximum likelihood (ML) as implemented in the program Garli 2.0 (Zwickl, 2006) on the Duke Shared Computing Resources Cluster (DSCR; https://wiki.duke.edu/display/SCSC/DSCR). The best model for each data set was determined by the Akaike information criterion (AIC; Table 2) using the program PartitionFinder (Lanfear et al., 2012). For both the trnG-R and gapCp data sets, the binary-coded indel data were assigned the Mkv model (Table 2; Lewis, 2001). Each analysis was run for four independent searches with four replicates each, resulting in 16 optimal maximum likelihood trees. The single most likely ML tree was identified as that having the largest −ln score (tree statistics summarized in Table 2). For each data set, clade support was assessed using 1000 bootstrap pseudoreplicate data sets. The bootstrap majority-rule consensus tree for each data set was compiled using PAUP* v. 4.0a123 (Swofford, 2002).

The trnG-R and gapCp data sets also were analyzed separately using Bayesian inference as implemented in the program MrBayes v 3.1.2 (Ronquist and Huelsenbeck, 2003) on the DSCR computing cluster. To accommodate the range of models accepted by MrBayes, the best-fitting model for trnG-R was implemented with the nst = (0 1 2 2 1 0), statefreqpr = estimate, and rates = gamma settings. The best-fitting model for gapCp was implemented with the nst = 2, statefreqpr = equal, and rates = gamma settings. The Mkv model for both trnG-R and gapCp binary-coded indel data was implemented with the nst = 1, rates = equal, and coding = variable; the average rates for the sequence data partition and binary-coded indel partition were allowed to differ (ratepr = variable). All other priors were left at their default values. Four analyses were run with four chains (one cold, three heated) for 10 million generations with a sample taken every 1000 generations. The sample parameter traces were visualized in the program Tracer v. 1.5 (Rambaut and Drummond, 2007). For each analysis, the four runs converged around 1 million generations. To be conservative, we excluded the first 2 million generations of each run as burn-in, resulting in a final pool of 32 000 samples. A majority-rule consensus tree was generated using PAUP* v. 4.0a123 (Swofford, 2002). Both the trnG-R and gapCp alignments and consensus trees were deposited in TreeBase (treebase.org; study: 15737).

RESULTS

Plastid phylogeny (Fig. 2A)—Of the 101 trnG-R sequences used in this study, 75 were newly generated and deposited in

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Locality: accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polypodium vulgare complex</td>
<td></td>
</tr>
<tr>
<td>Polypodium amorphum Suksd.</td>
<td>Canada, British Columbia, Greater Vancouver District: 6951</td>
</tr>
<tr>
<td>Polypodium appalachianum Hauff &amp; Windham</td>
<td>USA, Oregon, Multnomah Co.: 7556, 7772, 7773, 7778</td>
</tr>
<tr>
<td>Polypodium californicum Kauff.</td>
<td>USA, Washington, King Co.: 7771, 7779; Kittitas Co.: 6952; Mason Co.: 6953; Snohomish Co.: 7770, 7777</td>
</tr>
<tr>
<td>Polypodium cambricum L.</td>
<td>USA, California, Orange Co.: 3829; Riverside Co.: 5909; San Mateo Co.: 7249</td>
</tr>
<tr>
<td>Polypodium fauriei Christ</td>
<td>England: 8786</td>
</tr>
<tr>
<td>Polypodium glycyrrhiza D.C. Eaton</td>
<td>Spain, Valencia: 8787</td>
</tr>
<tr>
<td>Polypodium hesperium Maxon</td>
<td>USA, Arizona, Coconino Co.: 3127, 8171, 8174, 8175, 8176; Gila Co.: 8172, 8177, 8178; Graham Co.: 7217, 8179, 8180, 8181, 8182; Pima Co.: 7793</td>
</tr>
<tr>
<td>Polypodium macaronesium A.E. Bobrov</td>
<td>USA, Oregon, Lincoln Co.: 7559, Multnomah Co.: 7545, 7546, 7768, 7769</td>
</tr>
<tr>
<td>Polypodium pellucidum Kauff.</td>
<td>USA, Washington, King Co.: 7218, 7781; Lewis Co.: 7780, 7783; Snohomish Co.: 7766, 7782</td>
</tr>
<tr>
<td>Polypodium sibiricum Kunze Mexico, Baja California, Ensenada Municipality: 7539</td>
<td></td>
</tr>
<tr>
<td>Polypodium scouleri Hook. &amp; Grev.</td>
<td>USA, Idaho, Clearwater Co.: 7785, 7786; Shoshone Co.: 8288</td>
</tr>
<tr>
<td>Polypodium plesiosorum group</td>
<td>USA, Montana, Flathead Co.: 8277; Glacier Co.: 8278; Lake Co.: 8279, 8280, 8281; Lincoln Co.: 7791, 8268, 8269, 8270, 8271, 8272, 8273, 8274, 8276; Sanders Co.: 8290</td>
</tr>
<tr>
<td>Polypodium colpodes Kunze</td>
<td>USA, New Mexico, Rio Arriba Co.: 8183</td>
</tr>
<tr>
<td>Polypodium martensii Mett.</td>
<td>USA, Oregon, Multnomah Co.: 7763</td>
</tr>
<tr>
<td>Polypodium pellucidum Kauff.</td>
<td>USA, Utah, Beaver Co.: 8160; Salt Lake Co.: 7792, 8159, 8164, 8165; Washington Co.: 8167, 8168, 8169</td>
</tr>
<tr>
<td>Polypodium cambricum L.</td>
<td>USA, Washington, Chelan Co.: 7571; Snohomish Co.: 8206</td>
</tr>
<tr>
<td>Polypodium sibiricum Kunze</td>
<td>USA, Utah, Beaver Co.: 8160; Salt Lake Co.: 7792, 8159, 8164, 8165; Washington Co.: 8167, 8168, 8169</td>
</tr>
<tr>
<td>Polypodium plesiosorum group</td>
<td>USA, Washington, Chelan Co.: 7571; Snohomish Co.: 8206</td>
</tr>
<tr>
<td>Polypodium colpodes Kunze</td>
<td>USA, Utah, Beaver Co.: 8160; Salt Lake Co.: 7792, 8159, 8164, 8165; Washington Co.: 8167, 8168, 8169</td>
</tr>
<tr>
<td>Polypodium sibiricum Kunze</td>
<td>USA, Utah, Beaver Co.: 8160; Salt Lake Co.: 7792, 8159, 8164, 8165; Washington Co.: 8167, 8168, 8169</td>
</tr>
<tr>
<td>Polypodium plesiosorum group</td>
<td>USA, Washington, Chelan Co.: 7571; Snohomish Co.: 8206</td>
</tr>
<tr>
<td>Polypodium colpodes Kunze</td>
<td>USA, Utah, Beaver Co.: 8160; Salt Lake Co.: 7792, 8159, 8164, 8165; Washington Co.: 8167, 8168, 8169</td>
</tr>
<tr>
<td>Polypodium sibiricum Kunze</td>
<td>USA, Washington, Chelan Co.: 7571; Snohomish Co.: 8206</td>
</tr>
</tbody>
</table>

Polypodium ambiguus clade; †P. glycyrrhiza clade. Unique identifiers in italics indicate specimens included in the nuclear gapCp data set. The method of Simmons and Ochoterena (2000), implemented in gapcode.py v. 2.1 (Ree, 2008). The resulting binary characters were appended to the alignment. The corresponding indels were excluded. The resulting binary characters were appended to the alignment. The best model for each data set was determined by the Akaike information criterion (AIC; Table 2) using the program PartitionFinder (Lanfear et al., 2012). For both the trnG-R and gapCp data sets, the binary-coded indel data were assigned the Mkv model (Table 2; Lewis, 2001). Each analysis was run for four independent searches with four replicates each, resulting in 16 optimal maximum likelihood trees. The single most likely ML tree was identified as that having the largest −ln score (tree statistics summarized in Table 2). For each data set, clade support was assessed using 1000 bootstrap pseudoreplicate data sets. The bootstrap majority-rule consensus tree for each data set was compiled using PAUP* v. 4.0a123 (Swofford, 2002).
GenBank (Appendix 1). The best ML tree and Bayesian inference consensus tree yielded similar topologies and comparable support values (see Table 2 for tree statistics). Figure 2A shows the best plastid ML tree as an unrooted phylogram in which well-supported nodes are identified by Bayesian inference posterior probabilities (BIPP) ≥0.95 and maximum likelihood bootstrap support (MLBS) ≥70%.

Our trnG-R phylogeny (Fig. 2A) shows strong support for the reciprocal monophyly of the *P. vulgar* complex (BIPP = 1.0, MLBS = 71%) and the *P. plesiosorum* group (BIPP = 1.0, MLBS = 98%). In agreement with Sigel et al. (in press; findings summarized in Fig. 1), the *P. vulgar* complex consists of four well-supported clades—the *P. cambricum* clade (clade C; BIPP = 1.0, MLBS = 100%), the *P. scouleri* clade (clade S; BIPP = 1.0, MLBS = 99%), the *P. appalachianum* clade (clade A; BIPP = 1.0, MLBS = 92%), and the *P. glycyrrhiza* clade (clade G; BIPP = 1.0, MLBS = 92%). However, relationships among the four clades comprising the *P. vulgar* complex are largely unresolved by the plastid data (Fig. 2A).

Within the *P. vulgar* complex, clade C unequivocally unites the diploid taxa *P. cambricum* L. and *P. macaronesicum* A. E. Bobrov (BIPP = 1.0, MLBS = 100%), and the two samples of *P. cambricum* are united with strong support (BIPP = 1.0, MLBS = 74%). Clade S comprises *P. pellucidum* Kauf. and *P. scouleri* Hook. & Grev. (BIPP = 1.0, MLBS = 99%), with the two samples of *P. scouleri* united in a well-supported clade (BIPP = 1.0, MLBS = 83%). Clade A comprises all samples of the diploid taxa *P. amorphum*, *P. appalachianum* Hauffer & Windham, *P. sibiricum* Sipl., as well as 22 samples of *P. hesperium*. The plastid data provide no support for the monophyly of any species in clade A, and relationships among all samples are essentially unresolved (Fig. 2A). Clade G includes all samples of the diploid taxa *P. fauriei* Christ, *P. californicum* Kauf., and *P. glycyrrhiza*, as well as 29 samples of *P. hesperium*. *Polyodium fauriei* is resolved as sister to all other members of clade G with strong support (BIPP = 1.0, MLBS = 99%). Relationships among other samples in clade G are unresolved, with no support for the monophyly of the remaining taxa.

Within the *P. plesiosorum* group (clade P), *P. rhodopleuron* Kunze and *P. colpodes* Kunze are united in a well-supported clade (BIPP = 1.0, MLBS = 96%) that is sister to *P. plesiosorum* (BIPP = 1.0, MLBS = 100%). Relationships among *P. rhodopleuron* + *P. colpodes* + *P. plesiosorum* Kunze, *P. subpetiolatum* Hook., and *P. martensi* Mett. are unresolved (Fig. 2A).

**Nuclear sequence variation and phylogeny** (Fig. 2B)—Cloning and amplification of the nuclear locus gapCp resulted in 380 clones, representing 63 consensus sequences (putative alleles) for 36 samples (Table 1; Appendix 1). All gapCp sequences were newly generated for this study, and consensus allele sequences were deposited in GenBank. Figure 2B depicts the most likely tree (see Table 2 for tree statistics) resulting from ML analysis and is rooted with *Pleurosoripopsi sakmiorii*.

Unlike the trnG-R plastid ML tree (Fig. 2B), the gapCp ML phylogeny does not explicitly support monophyly of the *P. vulgar* complex. In the nuclear phylogeny, the *P. plesiosorum* group (clade P) forms a polytomy with two clades representing the *P. vulgar* complex—one uniting the diploid species of clades C, S, and A with mixed BIPP (≥1.0) and MLBS (≥65%) support and the other comprising the diploid taxa belonging to clade G (BIPP = 1.0, MLBS = 98%). The 28 alleles obtained from the 12 included specimens of *P. hesperium* are divided between these two clades, with each specimen of *P. hesperium* having at least one allele derived from each clade (Fig. 2B).

Our gapCp phylogeny (Fig. 2B) also provides no support for the monophyly of clades C, S, or A. The relationship between *P. cambricum* and *P. macaronesicum* (the two members of clade C) is unresolved. Similarly, the relationship between the two clade S species, *P. scouleri* and *P. pellucidum*, is unclear, though the two specimens of *P. scouleri* are supported (BIPP = 0.99, MLBS = 71%) as sister to one another. Alleles from *P. appalachianum* and *P. sibiricum* (both members of clade A) form a well-supported (BIPP = 1.0, MLBS = 89%) yet unresolved clade. All alleles from *P. amorphum*, as well as 14 alleles obtained from 12 specimens of *P. hesperium*, are united in a moderately supported clade (BIPP = 0.96, MLBS < 50%). Though the relationship among the *P. amorphum* and *P. hesperium* alleles is largely unresolved, five alleles obtained from three specimens of *P. hesperium* (all from the northern portion of its geographic range) form a strongly supported clade (BIPP = 1.0, MLBS = 84%).

Within clade G, *P. fauriei* is weakly supported (BIPP = 1.0, MLBS = 62%) as sister to the other taxa, and *P. californicum* appears sister to a robustly supported (BIPP = 1.0, MLBS = 98%) clade consisting of all *P. glycyrrhiza* samples and 14 alleles obtained from 12 specimens of *P. hesperium*. The relationship among the *P. glycyrrhiza* and *P. hesperium* alleles is largely unresolved, but eight alleles obtained from eight specimens of *P. hesperium* (all from the southern portion of its geographic range) are united in a clade with mixed BIPP (=0.96) and MLBS (=51%) support.

As with the trnG-R ML tree (Fig. 2B), the gapCp phylogeny (Fig. 2C) provides unequivocal support for the monophyly of the *P. plesiosorum* group (BIPP = 1.0, MLBS = 100%). Within the *P. plesiosorum* group, *P. plesiosorum* is sister to all other taxa (BIPP = 1.0, MLBS = 88%). Two alleles obtained from *P. subpetiolatum* are united with robust support (BIPP = 1.0, MLBS = 100%). Two alleles from *P. rhodopleuron* and two alleles from *P. colpodes* are united with strong support (BIPP = 1.0, MLBS

---

**Table 2.** Tree statistics for the data sets analyzed in this study.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Taxon sampling</th>
<th>Included sites</th>
<th>Variable sites</th>
<th>% Missing data</th>
<th>Best-fitting model</th>
<th>ML score</th>
<th>% Partitions MLBS ≥ 70%</th>
<th>% Partitions BIPP ≥ 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>trnG-R</td>
<td>100</td>
<td>986</td>
<td>123</td>
<td>0.39</td>
<td>K81uf+G</td>
<td>−2881.7743</td>
<td>29.73</td>
<td>32.43</td>
</tr>
<tr>
<td>trnG-R indels</td>
<td>38</td>
<td>38</td>
<td>1.77</td>
<td></td>
<td>K80+G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gapCp</td>
<td>552</td>
<td>552</td>
<td>144</td>
<td>11.14</td>
<td></td>
<td>−2234.9338</td>
<td>51.85</td>
<td>70.37</td>
</tr>
<tr>
<td>gapCp indels</td>
<td>36 (63 consensus alleles)</td>
<td>25</td>
<td>25</td>
<td>3.81</td>
<td>Mkv</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Missing data include both uncertain bases (?, N, R, Y, etc.) and gaps (-); MLBS: maximum likelihood bootstrap support. BIPP, Bayesian inference posterior probability.

---

SIGEL ET AL.––RECI PROCAL ORIGINS IN POLYPODIUM HESPERIUM
= 97%), with the two alleles from *P. colpodes* sister to each other (BIPP = 1.0, MLBS = 97%). The three alleles obtained from the outgroup taxon, *Pleurosoriopsis makinoi*, form an unequivocally supported clade (BIPP = 1.0, MLBS = 100%), with two alleles robustly supported as sister to the third allele (BIPP = 1.0, MLBS = 100%).

**Geographic distribution of plastid haplotypes (Fig. 2C)—**
Our phylogenetic analysis of trnG-R plastid sequence data reveals a striking correspondence between the geographic locality of a *P. hesperium* individual and its plastid haplotype. As shown in Fig. 2C, all specimens of *P. hesperium* sampled from the north—Washington, Oregon, Idaho, and Montana—have plastid genomes contributed by a diploid taxon from the *P. appalachianum* clade (clade A). On the other hand, all specimens from the south—Utah, Colorado, New Mexico, Arizona, and Baja California—have plastid genomes contributed by a diploid taxon from the *P. glycyrrhiza* clade (clade G).

**DISCUSSION**
Using a combination of uniparentally inherited plastid (*trnG-R*) and biparentally inherited nuclear (*gapCp*) sequence data, we were able to corroborate earlier hypotheses based on morphological, cytological, and isozyme data (Lang, 1971; Haufler et al., 1995a) that (1) *P. hesperium* is an allotetraploid derived from hybridization between the diploid species *P. amorphum* and *P. glycyrrhiza*, and (2) that *P. hesperium* encompasses distinct IDLs with reciprocal plastid donors. Chloroplast and mitochondrial plastid genomes are predominantly inherited from a single parent (Bock, 2007), and, while never tested specifically for *Polypodium*, several ferns have been shown to have maternal inheritance of plastids (Pellaea Link., Gastony and Yatskievych, 1992; Asplenium, Vogel et al., 1998; Equisetum L., Guillon and Raquin, 2000). Assuming maternal inheritance of plastids in *Polypodium*, Hauffer et al. (1995a) used chloroplast restriction site data to propose reciprocal hybrid origins in *P. hesperium*. In their study, one specimen of *P. hesperium* from Utah had *P. glycyrrhiza* as the putative maternal progenitor, whereas another specimen of *P. hesperium* from eastern Washington had *P. amorphum* (or a closely related diploid) as the maternal progenitor. By expanding the geographic sampling of *P. hesperium*, we have discovered a broader geographic trend that reinforces the initial findings of Hauffer et al. (1995a). All specimens of *P. hesperium* that we sampled from the north—Washington, Oregon, Idaho, and Montana—have plastid genomes contributed by a diploid taxon from the *P. appalachianum* clade (clade A), whereas all specimens from the south—Utah, Colorado, New Mexico, Arizona, and Baja California—have plastid genomes contributed by a diploid taxon from the *P. glycyrrhiza* clade (clade G; Fig. 2A, C). Our analysis of biparentally inherited *gapCp* sequence data further clarifies the plastid donors, providing support for *P. amorphum* (clade A) and *P. glycyrrhiza* (clade G), specifically, as the diploid progenitors of *P. hesperium* (Fig. 2C). Based on these findings, we propose that *P. hesperium* individuals north of 42°N belong to one or more IDLs that have *P. amorphum* as their maternal progenitor, whereas those south of 42°N belong to one or more IDLs that have *P. glycyrrhiza* as their maternal progenitor (Fig. 2C). Additional sampling of populations in British Columbia, California, and Chihuahua may further corroborate or provide interesting counterexamples to this pattern.

To the best of our knowledge, no other allopolyploid species with reciprocal origins has such a striking association between plastid haplotypes and geography. What might account for the pattern observed in *Polypodium hesperium*? It is possible that it was simply a matter of chance that *P. glycyrrhiza* was the maternal parent during a southern hybridization event, whereas *P. amorphum* was the maternal parent during a northern hybridization event. This scenario seems most probable if there is equal likelihood (e.g., no biological barriers) for reciprocal hybridization, and *P. hesperium* is composed of only two IDLs, one in the northern portion and one in the southern portion of its range. A second possibility is that plastids inherited from different parents confer selective advantages in different portions of the range of *P. hesperium*. While never explicitly analyzed, it is likely that there are significant environmental and ecological differences between *P. hesperium* habitats in the northern vs. southern portions of its range (Windham, 1985). Under this scenario, it is possible that multiple IDLs of *P. hesperium* were formed in both the northern and southern portions of its range.

**Fig. 2.** Summary of the evolutionary relationships and geographic distributions of independently derived lineages (IDLs) of the allotetraploid species *Polypodium hesperium*. (A) Best unrooted phylogram recovered by maximum likelihood (ML) analysis of the maternally inherited plastid *trnG-R* data set. Lettered clades correspond to those in Fig. 1. Black branches show relationships within the *P. vulgare* complex (clades A, G, C, and S); gray branches indicate the *P. plesiosorum* group (clade P) and *Pleurosoriopsis makinoi* (O). Red and blue circles highlight members of the *P. appalachianum* clade (A) and *P. glycyrrhiza* clade (G), respectively, with the number of individuals of each species noted in parentheses. *Polypodium hesperium* individuals contained within the red circle have a diploid member of the *P. appalachianum* clade (A) as a plastid donor, whereas *P. hesperium* individuals contained within the blue circle have a diploid member of the *P. glycyrrhiza* clade (G) as a plastid donor. Thicker branches display Bayesian inference posterior probability (BIPP) and maximum likelihood bootstrap (MLBS) support values (see inset legend). (B) Best phylogram recovered by ML analysis of the biparentally inherited nuclear *gapCp* data set. Lettered clades and branch colors are as specified in Figs. 1 and 2A. Bolded branches display BIPP and MLBS support values (see inset legend). Four-digit numbers following taxon names are unique specimen identifiers, and numbers in parentheses indicate the number of clones compiled to determine the corresponding consensus allele sequence (see Table 1 and Appendix 1). Red font indicates consensus allele sequences derived from *P. hesperium* specimens with a *P. appalachianum* clade (A) plastid donor, whereas blue font indicates consensus allele sequences derived from *P. hesperium* specimens with a *P. glycyrrhiza* clade (G) plastid donor. Curved lines join consensus allele sequences recovered from the same specimen. (C) Geographic distribution of *P. hesperium* and specific collection localities of *P. hesperium* specimens included in this study. Gray patches represent the known geographic range of *P. hesperium* as interpreted from voucher coordinate information from the Global Biodiversity Informatics Facility (http://www.gbif.com; Edwards et al., 2000), the Intermountain Region Herbarium Network (http://intermountainbiota.org), and the Consortium of California Herbaria (http://ucjeps.berkeley.edu/consortium). Squares indicate the collection localities of particular specimens used in this study (see Table 1 and Appendix 1 for locality information and coordinates). Square color corresponds to the plastid (presumably maternal) donor as determined by phylogenetic analysis of the *trnG-R* data set: red indicates *P. hesperium* specimens with plastids inherited from a member of the *P. appalachianum* clade (A); blue indicates *P. hesperium* specimens with plastids inherited from a member of the *P. glycyrrhiza* clade (G).
but only those having the more advantageous plastid haplotype—*P. amorphum* in the north and *P. glycyrrhiza* in the south—remain extant.

**Relative ages of northern and southern IDLs of *Polypodium hesperium***—Because *P. amorphum* and *P. glycyrrhiza* both diverged from their closest relatives during the early to middle Pleistocene (approximately 1.2 Ma and 2.5 Ma, respectively; Sigel et al., 2014b), all IDLs of *P. hesperium* were likely formed within the last 1.2 Myr. The present-day distributions of *P. amorphum* and *P. glycyrrhiza* are characteristic of species whose evolutionary histories have been shaped by cycles of glacial advances and retreats during the Pleistocene (Pielou, 1991; Soltis et al., 1997; Hauffler et al., 2000; Sigel et al., 2014b). *Polypodium amorphum* is primarily found in the Cascade and Olympic Mountains of the Pacific Northwest, whereas *P. glycyrrhiza* grows near the Pacific coast from the Kamchatka Peninsula to central California (Hauffler et al., 1993; Douglas et al., 2000). The geographic ranges of *P. hesperium* and its progenitors currently overlap in a narrow band extending from the Columbia River gorge on the Oregon/Washington border to the Frasier River gorge in southern British Columbia (Lang, 1971). During glacial advances, the ranges of *P. glycyrrhiza* and *P. amorphum* likely extended south and overlapped more extensively to form one (or more) southern IDLs of *P. hesperium*. During glacial retreats, the ranges of *P. glycyrrhiza* and *P. amorphum* likely contracted northward and formed at least one northern IDL of *P. hesperium*. The northernmost distribution of *P. hesperium* was glaciated or under permafrost during the last glacial maximum (approximately 20000–19000 yr ago; Pielou, 1991; Taberlet, 1998; Clark and Mix, 2002; Clark et al., 2009), and sediment cores from Fraser River Canyon reveal the appearance of *Polypodium* spores approximately 11000 yr ago (Matthews and Rouse, 1975). We hypothesize that the northern IDL(s) of *P. hesperium* may have formed during the last 20000 yr and are potentially younger that the southern IDL(s).

Our hypothesis about the relative ages of the northern and southern IDLs of *P. hesperium* challenges an earlier hypothesis by Hauffler et al. (1995a) that populations from western Montana were of older origin than other northern or southern populations. Because specimens from western Montana and eastern Washington exhibit nuclear-encoded hexokinase (HK) isozyme alleles and spore ornamentation patterns resembling those of *P. sibiricum*, a diploid species closely allied to *P. amorphum* (Fig. 1, clade A), Hauffler et al. (1995b) proposed that *P. hesperium* in western Montana and eastern Washington may have resulted from an ancient hybridization event between *P. glycyrrhiza* and the common ancestor of *P. amorphum* and *P. sibiricum*. However, neither our *trnG*-R nor gapCp data sets provide evidence that specimens from western Montana belong to a distinct, ancient IDL; the presence of *P. sibiricum* characters might be explained by introgression instead. *Polypodium saximontanum* Windham, an allotetraploid derived from *P. amorphum* and *P. sibiricum* (Fig. 1), recently has been found in Ravalli County, Montana, where it overlaps with *P. hesperium* (Sigel et al., 2014a). It is feasible that these two allopolyploid species may have hybridized, providing a mechanism for the introgression of *P. sibiricum* genes into the *P. hesperium* population (Rieseberg, 1995; Hegarty and Hiscock, 2005).

**Are there additional IDLs of *Polypodium hesperium***?—While our *trnG*-R plastid phylogeny (Fig. 2B) strongly supports reciprocal origins of *P. hesperium*, it is possible that the northern and southern populations of *P. hesperium* do, in fact, include multiple, “cryptic” IDLs that share a plastid haplotype. This has been shown for several natural allopolyploids, including species of *Platanthera* (Wallace, 2003), *Aegilops* (Meinberg et al., 2009), and *Androsace* (Dixon et al., 2009). The most extensively studied example involves the allopolyploid angiosperm *Tragopogon miscellus* Ownbey (Asteraceae; Soltis and Soltis, 1991; Soltis et al., 1995, 2004). In addition to having well-documented reciprocal IDLs, different populations of *T. miscellus* with identical plastid haplotypes exhibit significant variation in random amplified polymorphic DNA (RAPD) markers (Cook et al., 1998). Because *T. miscellus* is of very recent origin (~80 yr ago), this variation has been attributed to repeated allopolyploid formation among genetically variable diploid populations, not subsequent genetic evolution.

In the case of *P. hesperium*, our nuclear gapCp data support the distinction between northern and southern populations suggested by the plastid data (Fig. 2B). Specimens from northern populations (highlighted in red) have *P. amorphum*–derived alleles that form a well-supported (BIPP = 1.0; MLBS = 89%) clade not represented among the sampled diploids of that species. On the other hand, the *P. amorphum*–derived alleles observed in the southern populations of *P. hesperium* are identical (or very similar) to one of the two alleles recovered from *P. amorphum* 7773 (Fig. 2B). With regard to the *P. glycyrrhiza*–derived alleles found in *P. hesperium*, the northern (red) and southern (blue) populations do not share alleles with each other, and none are identical to alleles recovered from the three samples of diploid *P. glycyrrhiza*. The lack of resolution among alleles within the northern and southern populations of *P. hesperium* seems to argue against additional IDLs of *P. hesperium*. The southern populations in particular show minimal variation in gapCp, and the majority of specimens have identical alleles (Fig. 2C). While the gapCp data do not provide compelling evidence for additional IDLs in *P. hesperium*, the possibility of their existence cannot be dismissed. It is possible, even likely, that gapCp is insufficiently variable to allow detection of cryptic IDLs, and additional nuclear markers (such as the RAPDs used in *Tragopogon miscellus*) will be necessary to address this question.

**Polypodium hesperium: A natural fern model system for investigating how reciprocal origins shape allopolyploids**—With few exceptions, research on polyploid plants has been dominated by studies of angiosperms—particularly crops, synthesized polyploids, and model organisms (Buggs, 2008). It remains unclear whether conclusions drawn from these particular taxa may be extended to broadly divergent evolutionary lineages. Here, we propose *P. hesperium* as a natural fern model system for studying the effects of gene duplication and hybridization on adaptation and evolution at the phenotypic, genetic, and genomic levels. Several features of *Polypodium hesperium* make it a particularly promising study system, including reciprocal IDLs with disjunct geographic distributions, recent (Pleistocene) time of origin, divergent nuclear genomes with homoeologous gene copies that can be phylogenetically distinguished from one another, extant diploid progenitors, and possible evidence for introgression with a closely related species. In addition, *P. hesperium*, as well as *P. glycyrrhiza* and *P. amorphum*, survive well when transplanted from the field into controlled growth conditions, can be asexually propagated, and have spores that readily germinate into gametophytes in standard media (Hoshizaki, 1975; E. M. Sigel, personal observation). At present, no other allopolyploid fern is known that is
better suited to improving our understanding of the role that polyploidy and hybridization have played in plant evolution.

LITERATURE CITED


APPENDIX 1. Specimens of Pleurozium and Polypodium included in this study.

**Taxon name—Fern Laboratory Database accession number** (http://fernlabbiology.duke.edu): voucher information (herbarium); trnG-R: GenBank accession; gapCp: GenBank accession (number of clones in consensus allele sequence) [repeated for multiple gapCp consensus alleles]; latitude and longitude coordinates are not available.