The Utility of Nuclear gapCp in Resolving Polyploid Fern Origins

Eric Schuettpelz, Amanda L. Grusz, Michael D. Windham, and Kathleen M. Pryer
Department of Biology, Duke University, Durham, North Carolina 27708 U.S.A.

Abstract—Although polyploidy is rampant in ferns and plays a major role in shaping their diversity, the evolutionary history of many polyploid species remains poorly understood. Nuclear DNA sequences can provide valuable information for identifying polyploid origins; however, remarkably few nuclear markers have been developed specifically for ferns, and previously published primer sets do not work well in many fern lineages. In this study, we present new primer sequences for the amplification of a portion of the nuclear gapCp gene (encoding a glyceraldehyde-3-phosphate dehydrogenase). Through a broad survey across ferns, we demonstrate that these primers are nearly universal for this clade. With a case study in chelblastins, we show that this rapidly evolving marker is a powerful tool for discriminating between autopolyploids and allopolyploids. Our results indicate that gapCp holds considerable potential for addressing species-level questions across the fern tree of life.

Keywords—allopolyploidy, autopolyploidy, ferns, gapCp, GAPDH, molecular systematics, nuclear marker.

Polyploidy—the multiplication of entire chromosome sets—has been documented in nearly all major eukaryotic lineages (Otto and Whitten 2000; Gregory and Mable 2005; Tate et al. 2005). The process is especially prevalent in ferns, which exhibit both the highest known gametic chromosome numbers (Abraham and Niran 1954) and some of the highest incidences of polyploidy (Walker 1966; Manton and Vida 1968; Löve et al. 1977; Walker 1984). Nearly 50% of fern species that have been studied cytogenetically are polyploids of recent origin (neopolyploids, following Ramsey and Schem-ske 2002), exhibiting chromosome numbers that are multiples of those documented in closely related species (Vida 1976; Walker 1984). In addition, it is estimated that at least 95% of fern species have undergone polyploidization at some point in their evolutionary history (Grant 1981; Hauffier 1987). Polyploidy has left an indelible mark on fern evolution and continues to serve as a dynamic source of genetic variability, ecological innovation, and species diversity (e.g., Klekowski and Baker 1966; Klekowski 1972; Walker 1984; Werth et al. 1985; Soltis and Soltis 1987; Werth and Windham 1991; Hauffier et al. 1995).

Evolutionary biologists typically recognize two major categories of neopolyploids (see Soltis et al. 2007 and references therein). Organisms containing multiple genomes from a single diploid species are generally called autopolyploids; those that incorporate genomes derived from two or more diploid species are called allopolyploids. Although these categories are widely used, the distinction can be difficult to operationalize because it requires detailed knowledge of diploid progenitor populations (which may be unknown or extinct) and is dependent on the species concept applied (Soltis et al. 2007). Nonetheless, it is important to be able to assign individuals or taxa to these categories, because autopolyploids and allopolyploids exhibit fundamental differences in their genetics, ecology, and evolutionary potential (Levin 1983; Thompson and Lumaret 1992; Soltis and Soltis 2000; Wendel 2000; Ramsey and Schemske 2002; Osborn et al. 2003).

Through the years, evolutionary biologists have used a variety of techniques to discriminate between autopolyploids and allopolyploids. Morphology, chromosome pairing behavior, and patterns of genetic segregation (based primarily on allozyme data) all provide important clues regarding polyploid origins (Grant 1981; Jackson 1982; Soltis and Rieseberg 1986). However, these approaches are inherently phenetic, relying exclusively on the genetic similarity of the genomes involved. As we move toward more integrated species concepts, it becomes increasingly important to place the genomes found in polyploid organisms in a phylogenetic context. Thus, DNA sequencing—specifically of nuclear and organellar markers in combination—is emerging as a powerful tool for revealing polyploid origins (Ge et al. 1999; Sang and Zhang 1999; Hoot and Taylor 2001; Popp and OKelso 2001; Popp and Oelixen 2007). Biparentally-inherited nuclear markers provide sequences unique to individual diploid species, which can be isolated from polyploids through cloning and analyzed in a phylogenetic context. This allows for a more objective assessment of whether constituent genomes came from a single diploid species (autopolyploid) or more than one diploid species (allopolyploid). Maternally-inherited organellar markers, in turn, can distinguish the maternal from paternal species.

We now have many primer sets at our disposal for the amplification of organellar (primarily plastid) DNA from ferns (Small et al. 2005), but there are currently few options available for obtaining nuclear sequences in this group of plants. To date, most nuclear sequencing in ferns has focused on ribosomal DNA (rDNA) markers. Although the small ribosomal subunit gene (18S) is broadly amplifiable using published primers (Bult et al. 1992; Wolf 1995), it is of limited utility even at the deepest phylogenetic levels (Schuettpelz et al. 2006), and therefore certainly not useful within a species complex. The large subunit gene (26S) and internal transcribed spacers (ITS) may provide increased phylogenetic signal, but researchers have had little success in sequencing these regions in ferns (Gastony and Rollo 1998; Van den heede et al. 2003; Reid et al. 2006). In any case, rDNA sequences are not an ideal choice for assessing polyploid origins in ferns (or any other group), as the effects of concerted evolution may obscure evidence of reticulation (Wendel et al. 1995; Álvarez and Wendel 2003; Kovarik et al. 2004, 2005; Lihová et al. 2006; Lim et al. 2007; Volkov et al. 2007). Single- or low-copy nuclear genes are not nearly as susceptible to this process (Wendel 2000; Ma and Gustafson 2005), and therefore offer the greatest promise for revealing polyploid origins (Small et al. 2004).

Two single-copy nuclear markers have been developed for
ferns, with primers published for the amplification of portions of the pgIC (Ishikawa et al. 2002) and gapCp (Ebihara et al. 2005) genes. Unfortunately, our efforts to amplify these markers using the published primers were unsuccessful for cheilanthoid ferns—a large clade within the Pteridaceae (Smith et al. 2006; Schuettelpelz et al. 2007) wherein polyploidy is especially rampant (Windham and Yatskievych 2003). So too were our attempts to amplify other low-copy nuclear markers (e.g. portions of the adh, gapC, leafy, and waxxy genes) from cheilanthoids with primer sets not specifically designed for ferns (Frohlich and Meyerowitz 1997; Mason-Gamer et al. 1998; Small et al. 1998; Wall 2002).

In an effort to obtain a marker for use in identifying polyploid origins in cheilanthoid ferns, we developed new primers for the amplification of sequences from the fern *Marsilea* L. (GenBank accession AJ003783; Meyer-Gauen et al. 1998) to published gapCp gene sequences from the seed plant *Pinus sylvestris* L. (GenBank accession AJ001706; Meyer-Gauen et al. 1998) and several other embryophyte species (Petersen et al. 2003). Suitable (i.e. conserved) priming sites were located within exons 8 and 11 of the (the Pinus gapCp gene), and primers (ESGAPCP8F1 and ESGAPCP11R1) were designed to amplify portions of these exons and the elements of the gapCp gene between them (for a map and primer sequences, see Fig. 1).

**Materials and Methods**

**Primer Design**—To identify conserved priming sites that would potentially allow for amplification of gapCp across ferns, we compared a published gapCp cDNA sequence from the fern *Marsilea* L. (GenBank accession AJ003783; Meyer-Gauen et al. 1998) to published gapCp gene sequences from the seed plant *Pinus sylvestris* L. (GenBank accession AJ001706; Meyer-Gauen et al. 1998) and several other embryophyte species (Petersen et al. 2003). Suitable (i.e. conserved) priming sites were located within exons 8 and 11 of the (the Pinus gapCp gene), and primers (ESGAPCP8F1 and ESGAPCP11R1) were designed to amplify portions of these exons and the elements of the gapCp gene between them (for a map and primer sequences, see Fig. 1).

**Taxonomic Sampling**—For our broad survey, to assess the universality of our primers through a broad survey across ferns. Then, with a case study focused on a cheilanthoid species complex (see Discussion for details on this complex), we demonstrate the utility of this newly-developed marker for discriminating between autoploidy and allopolyploidy species.

**RESULTS**

In an effort to obtain a marker for use in identifying polyploid origins in cheilanthoid ferns, we developed new primers for the amplification of sequences from the fern *Marsilea* L. (GenBank accession AJ003783; Meyer-Gauen et al. 1998) to published gapCp gene sequences from the seed plant *Pinus sylvestris* L. (GenBank accession AJ001706; Meyer-Gauen et al. 1998) and several other embryophyte species (Petersen et al. 2003). Suitable (i.e. conserved) priming sites were located within exons 8 and 11 of the (the Pinus gapCp gene), and primers (ESGAPCP8F1 and ESGAPCP11R1) were designed to amplify portions of these exons and the elements of the gapCp gene between them (for a map and primer sequences, see Fig. 1).

**DNA Isolation, Amplification, Cloning, and Sequencing**—For each sampled individual, genomic DNA was extracted from silica-dried material using the isolation protocol described in Schuettelpelz and Pryer (2007). Amplifications from these extractions were carried out using the polymerase chain reaction (PCR), with 1 × PCR buffer IV containing MgCl2 (Abgene, Epsom, United Kingdom), 200 μM each dNTP, 100 μg/ml BSA, 50 U/ml Taq polymerase, 0.5 μM primer ESGAPCP8F1, 0.5 μM primer ESGAPCP11R1, and 1 μl template DNA eluate in a 25 μl reaction. Thermal cycling programs entailed an initial denaturation step (94°C for 5 min) followed by 35 denaturation, annealing, and elongation cycles (94°C for 1 min, 55°C for 1 min, 72°C for 2 min) and a final elongation step (72°C for 10 min). The results of these reactions were visualized on an agarose gel. PCR products were cloned using the TOPO TA Cloning kit (Invitrogen, Carlsbad, California). Cloning reactions included 0.33 μl salt solution, 0.33 μl water, 0.33 μl TOPO vector, and 0.66 μl PCR product, and were incubated for 30 minutes at room temperature. Transformations included only 0.66 μl of the cloning reaction and 16.7 μl of One Shot TOP10 competent cells, but otherwise followed the manufacturer’s protocol. The transformed cells were spread on selective (ampicillin plus X-gal) LB plates and incubated overnight at 37°C. A minimum of 16 white colonies from each plate were individually picked and resuspended in 25 μl PCR reactions prepared as above, but with the M13 Forward (−20) and M13 Reverse primers supplied by Invitrogen. Thermal cycling was performed as described above, but with elongation cycles of 3 min. The results for these reactions were again visualized on an agarose gel. For the broad survey, multiple reactions from each observed size class were selected for sequencing. For the case study, multiple reactions from the appropriate size class only (see Results section below) were selected for sequencing.

Selected PCR products were purified using Montage PCR Centrifugal Filter Devices (Millipore, Billerica, Massachusetts), following the manufacturer’s protocol. Sequencing of the cleaned PCR products employed the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California). Each 10 μl reaction incorporated 0.375× BigDye Terminator Ready Reaction Mix (Applied Biosystems), 0.625× BigDye Terminator Sequencing Buffer (Applied Biosystems), 1 μl PCR product for inserts up to 700 bp, only the ESGAPCP8F1 and ESGAPCP11R1 primers were used (for longer inserts, the M13 Forward (−20) and M13 Reverse primers were also employed), and 2 μl purified PCR product. Thermal cycling and reaction purification followed the manufacturer’s protocol. Sample electrophoresis and analysis were performed using an ABI 3730xl DNA Analyzer (Applied Biosystems). The multiple sequencing reads obtained as chromatograms from each individual purified PCR product were assembled and edited separately using Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, Michigan). All consensus sequences (219 newly obtained) were subsequently deposited in GenBank (Accessions DQ112921–6051).

**Sequence Alignment and Phylogenetic Analysis**—The consensus sequences obtained for the broad survey and the case study were separately aligned using MacClade 4.08 (Maddison and Maddison 2005). For the broad survey, gapC (GenBank accession L07501) and gapCp (GenBank accession AJ001706) exon sequences from *Pinus* were also included to aid in alignment and analysis. Ambiguously aligned regions (all three introns in the case of the broad survey; a few regions within the introns in the case study) were excluded from subsequent analyses. The included portions of the broad survey and case study data sets had 0.0% and 0.3% missing data, respectively.

Both data sets (TreeBASE study number S2015) were phylogenetically analyzed using GARLI version 0.951 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006), employing the GTR + I + G model of DNA sequence evolution. The analyses comprised eight independent runs, each with a random starting topology and using the weighted priors as recommended by the author. To assess branch support, non-parametric bootstrap analyses (with 100 replicates) were conducted, also in GARLI (with the model and settings as above).

To further assess branch support, both data sets were also analyzed using a Bayesian Markov chain Monte Carlo (MCMC) approach in MrBayes 3.1.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), again employing the GTR + I + G model of DNA sequence evolution. These analyses comprised four independent runs, each with four chains (one cold and three heated), and using the default priors. Chains were run for 10 million generations and trees were sampled from the cold chain every 1000 generations. To identify when analyses had reached stationarity, the standard deviation of the split frequencies among the independent runs (as calculated by MrBayes) was examined, and the output parameter estimates were plotted using Tracer 1.4 (Rambaut and Drummond 2007). Based on these convergence diagn-
nistics, the first 2.5 million generations were (very conservatively) excluded from each analysis before obtaining clade posterior probabilities with the “sumt” command.

**RESULTS**

**Broad Survey**—Initial amplifications with primers ESGAPCP8F1 and ESGAPCP11R1 (Fig. 1) yielded products of one to three distinct lengths for any given taxon, suggesting that either the nuclear gapCp gene underwent duplication at some point in the evolutionary history of ferns, or that some other gene (perhaps coding for another glyceraldehyde-3-phosphate dehydrogenase) was being amplified simultaneously in some instances. Through cloning we were able to isolate and sequence individual fragments, which ranged in size from about 500–1700 bp (Table 1; Appendix 1). All sequences obtained were readily alignable to the gapC and gapCp exon sequences from *Pinus*, without the need for any insertions or deletions in these protein-coding regions, indicating that they encoded glyceraldehyde-3-phosphate dehydrogenases. The presence of three amino acid sequence signatures in the (translated) exon sequences (proline at position 235, asparagine at position 236, and cysteine at position 244; Petersen et al. 2003) further revealed that nearly all were gapCp genes. Introns varied considerably in length (Table 1) and were not alignable across ferns, but all conformed to the GT–AG rule (i.e. all intron sequences began with GT and ended with AG; Breathnach et al. 1978).

Phylogenetic analysis of the exon data (a total of 296 characters) resulted in a tree with mixed levels of branch support (Fig. 2A). Sequences of a given size class, from a particular taxon, were essentially identical (generally differing only by a few substitutions) and thus were always well supported as monophyletic. The relationships of these sequence clusters to one another, however, were often without support, which was expected given the limited number of characters. The gapC/gapCp duplication was well supported, and it is clear that some sequences obtained from *Cheilanthes* and *Adiantum* L. were of gapC rather than gapCp. Furthermore, our data suggest that a gapCp duplication—perhaps yielding “long” and “short” versions of the gene—occurred relatively early in the evolutionary history of ferns, although its precise phylogenetic position cannot yet be determined (Fig. 2B; see Discussion section below).

**Case Study**—As in the broad survey (Table 1), amplifications in the *Cheilanthes* case study yielded products of more than one length. With a desire to focus our efforts on a single paralog, and recognizing that gapC and the “long” version of gapCp in *Cheilanthes* were not readily discernable on a gel (with ranges of 929–930 and 917–967 bp, respectively; Table 1), we targeted only the “short” version of gapCp (594–602 bp; Appendix 2). Both the coding and non-coding regions of these sequences were readily alignable by eye, and it was necessary to exclude only 33 of 609 characters from the phylogenetic analyses due to questionable alignment. This was highly suggestive of homology, as the gapCp “short” introns from *Cheilanthes* are not even alignable to those from its closest sampled relative (*Adiantum*; Fig. 2); homology of these sequences was confirmed through phylogenetic analysis (results not shown).

Phylogenetic analysis of the *Cheilanthes* data set resulted in a tree with four distinct clades (three of which were strongly supported as monophyletic), corresponding to the four sampled diploid species (*C. coovilii*, *C. fendleri*, *C. lindheimeri*, and *C. parryi*; Fig. 3); all sequences obtained from any given diploid fell within a single clade. All sequences obtained from our two accessions of the triploid *C. lindheimeri* (15 plus 19 sequences) fell within the diploid *C. lindheimeri* clade (intermixed with diploid *C. lindheimeri* sequences), supporting an autopolyplloid ancestry for this taxon. By contrast, sequences obtained from two different accessions of the triploid *C. wootonii* (13 plus 15 sequences) grouped with both the *C. fendleri* and *C. lindheimeri* clades (again, intermixed with diploid sequences), indicating allopolyploid ancestry.

**DISCUSSION**

**Glyceraldehyde-3-Phosphate Dehydrogenase Gene Duplications**—Extant land plants possess four distinct glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes (gapA, gapB, gapC, and gapD). The nuclear gapC and gapCp genes are thought to have diverged prior to the origin of land plants, and the gapCp gene was thought to have undergone a duplication event relatively early in the evolutionary history of ferns, although its precise phylogenetic position cannot yet be determined (Fig. 2B; see Discussion section below).

### Table 1

Summary of sequences obtained in the broad survey using primers ESGAPCP8F1 and ESGAPCP11R1. Assignments to gapC or gapCp are based on the presence of amino acid sequence signatures (see text; Petersen et al. 2003). The “long” and “short” designations for gapCp are based solely on product length; if only one version of gapCp was retrieved, no length designation is given (see text). Exon and intron numbers follow those given in the Appendix.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Product</th>
<th>Clones</th>
<th>Exon 8 (bp)</th>
<th>Intron 8 (bp)</th>
<th>Exon 9 (bp)</th>
<th>Intron 9 (bp)</th>
<th>Exon 10 (bp)</th>
<th>Intron 10 (bp)</th>
<th>Exon 11 (bp)</th>
<th>All exons (bp)</th>
<th>All introns (bp)</th>
<th>Total (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Adiantum</em></td>
<td>gapCp “long”</td>
<td>4</td>
<td>4</td>
<td>78</td>
<td>143</td>
<td>204</td>
<td>84–76</td>
<td>65</td>
<td>296</td>
<td>358</td>
<td>654</td>
<td></td>
</tr>
<tr>
<td><em>Cheilanthes</em></td>
<td>gapCp “long”</td>
<td>2</td>
<td>4</td>
<td>588–593</td>
<td>143</td>
<td>152</td>
<td>84–85–65</td>
<td>65</td>
<td>296</td>
<td>621–671</td>
<td>917–967</td>
<td></td>
</tr>
<tr>
<td><em>Cheilanthes</em></td>
<td>gapCp “short”</td>
<td>16</td>
<td>4</td>
<td>121–122</td>
<td>143</td>
<td>90–93</td>
<td>84–85–65</td>
<td>65</td>
<td>296</td>
<td>300–304</td>
<td>596–600</td>
<td></td>
</tr>
<tr>
<td><em>Diplazierium</em></td>
<td>gapCp</td>
<td>7</td>
<td>4</td>
<td>245–246</td>
<td>143</td>
<td>415</td>
<td>84–100–65</td>
<td>296</td>
<td>761–762</td>
<td>1057–1058</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Drosera</em></td>
<td>gapCp “long”</td>
<td>3</td>
<td>4</td>
<td>90</td>
<td>143</td>
<td>1175–1256</td>
<td>84–85–65</td>
<td>296</td>
<td>1354–1435</td>
<td>1659–1731</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hymenophyllum</em></td>
<td>gapCp</td>
<td>5</td>
<td>4</td>
<td>80</td>
<td>143</td>
<td>107</td>
<td>84–85–65</td>
<td>65</td>
<td>273</td>
<td>569</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hymenophyllum</em></td>
<td>gapCp</td>
<td>5</td>
<td>4</td>
<td>314</td>
<td>143</td>
<td>337</td>
<td>84–91–65</td>
<td>65</td>
<td>742</td>
<td>1038</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lindaea</em></td>
<td>gapCp</td>
<td>5</td>
<td>4</td>
<td>95</td>
<td>143</td>
<td>393</td>
<td>84–85–65</td>
<td>65</td>
<td>572</td>
<td>868</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ligulodendron</em></td>
<td>gapCp</td>
<td>8</td>
<td>4</td>
<td>203–205</td>
<td>143</td>
<td>231</td>
<td>84–82–65</td>
<td>65</td>
<td>516–518</td>
<td>812–814</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pteris</em></td>
<td>gapCp “long”</td>
<td>5</td>
<td>4</td>
<td>239–240</td>
<td>143</td>
<td>275</td>
<td>84–75–65</td>
<td>65</td>
<td>589–590</td>
<td>885–886</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pteris</em></td>
<td>gapCp “short”</td>
<td>5</td>
<td>4</td>
<td>79–90</td>
<td>143</td>
<td>88</td>
<td>84–79–63</td>
<td>65</td>
<td>250–257</td>
<td>546–553</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sticherus</em></td>
<td>gapCp</td>
<td>4</td>
<td>4</td>
<td>299</td>
<td>143</td>
<td>424</td>
<td>84–90–65</td>
<td>65</td>
<td>813</td>
<td>1109</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
gapB, gapC, and gapCp). The gapA gene is believed to be of plastid origin and the gapC gene is thought to be of mitochondrial origin (Martin and Cerff 1986; Brinkmann et al. 1987; Martin and Schnarrenberger 1997), but both were transferred to the host cell nucleus following primary endosymbiosis and both were subsequently duplicated (to yield gapB and gapCp, respectively) well before the emergence of embryophytes (Brinkmann et al. 1989; Meyer-Gauen et al. 1994; Meyer-Gauen et al. 1998; Petersen et al. 2003; Petersen et al. 2006). The early gapC/gapCp duplication can be visualized in the results of our broad survey across ferns (Fig. 2A). Although our primers were generally gapCp-specific, we also amplified gapC from two pteroid genera (Adiantum and Cheilanthes). As expected, these sequences are more closely related to gapC from Pinus than they are to gapCp sequences from ferns (Fig. 2A).

More recent GAPDH gene duplications have been uncovered in several groups of plants (see Petersen et al. 2003 for a summary), but none have yet been characterized within ferns. Our broad survey, however, indicates that a gapCp duplication occurred relatively early in the evolutionary history of this clade. We recovered both “long” and “short” gapCp sequences from all sampled pteroid genera (Adiantum, Cheilanthes, and Pteris L.), as well as from one of two sampled eupolypod genera (Dryopteris Adans.; Table 1). Because pteroids and eupolypods are understood to be sister clades

---

**FIG. 2.** Genealogy (A) resulting from maximum likelihood analysis of the broad survey gapC/gapCp data set (exons only; 296 bp), and summary cladogram (B) resulting from previous analyses of plastid data (Pryer et al. 2004; Schuettpelz et al. 2006; Schuettpelz and Pryer 2007). Genealogy resulting from Bayesian analysis of the broad survey data set yielded an essentially congruent topology (not shown). In both the genealogy (A) and the summary cladogram (B), heavily thickened branches (most subterminal branches in gapCp genealogy and nearly all branches in summary cladogram) received both a maximum likelihood bootstrap score ≥70% and a Bayesian posterior probability ≥0.95; slightly thickened branches received either a maximum likelihood bootstrap score ≥70% or a Bayesian posterior probability ≥0.95. In the gapC/gapCp genealogy (A), the number of clones in each terminal clade is indicated in parentheses. The gapCp/gapC gene duplication (filled arrow) occurred well before the divergence of seed plants (Pinus) from ferns (all other genera), and represents the root of the presented genealogy. A gapCp duplication (open arrow, B) occurred sometime before the divergence of pteroid ferns (Pt) from eupolypod ferns (Eu), as “long” and “short” versions of the gene were recovered from both of these sister clades; however, its precise phylogenetic position remains unclear (see text).
(Schuettpelz and Pryer 2007), one might hypothesisize that the gapCp duplication (yielding “long” and “short” versions of the gene) occurred immediately prior to their divergence from one another (Fig. 2B). However, the gapC/gapCp genealogy is not entirely consistent with this scenario, because the “long” and “short” versions do not form reciprocally monophyletic sister clades (Fig. 2A). While this result is likely the product of too few data (only 296 bp were analyzed and there is little support across the backbone of the genealogy), it does cast doubt on the working hypothesis. It is entirely possible that the gapCp duplication occurred much earlier in the evolutionary history of ferns, followed by the loss of one copy in most of the sampled lineages. Alternatively, the observed pattern may simply reflect our inability to recover all copies from all lineages or a much more complex evolutionary history (with many underlying duplications and losses).

**Identifying Polyploid Origins in Cheilanthoid Ferns**—Cheilanthoid ferns compose a large, mostly xeric-adapted clade within the Pteridaceae (Smith et al. 2006; Schuettpelz et al. 2007). Both polyploidy and apomixis are rampant, and evolutionary histories can be very complex (Windham and Rabe 1993; A. L. Grusz et al. unpubl. results). Despite considerable morphological, cytogenetic, and allozyme data, the origins of many polyploid species remain unclear. Such was the case for the two apomictic triploids included in our case study: *Cheilanthes lindheimeri* and *C. wootonii*.

The apomictic triploid *Cheilanthes lindheimeri* has a broad range extending from southern Arizona and central Texas to Hidalgo, Mexico. Because its rather extreme morphology...
 Reeves (1979) cannot be readily explained by hybridization among any other species of Cheilanthes. C. lindheimeri was generally assumed to be an autopolyploid whose diploid progenitor was undiscovered (due to its morphological similarity to the widespread triploid). Interestingly, the existence of this diploid “C. lindheimeri” was also implicit in the proposed origin of another apomictic triploid. Based on morphological comparisons, Reeves (1979) hypothesized that C. wootonii (subsequently shown to be an apomictic triploid; Windham and Yatskievych 2003) was produced via hybridization between C. lindheimeri and C. fendleri. However, there is no known mechanism by which the apomictic triploid C. lindheimeri could contribute to a hybrid with the same ploidy level (Gastony and Windham 1989). Thus, if C. lindheimeri was indeed one of the parents of C. wootonii, the former had to include more than one ploidy level. A recent survey of herbarium specimens (A. L. Grusz et al. unpubl. results) resulted in the discovery of three diploid populations of C. lindheimeri. Two individuals from these populations were included here, as hypothetical parents of both focal triploids.

All of the morphological and cytogenetic hypotheses outlined above were corroborated by our gapCp analysis. The sequences isolated from diploid and triploid individuals of C. lindheimeri were intermixed with one another and composed a single well-supported clade (Fig. 3), supporting an autoploid origin for the latter. Sequences obtained from accessions of C. wootonii revealed contributions from two phylogenetically divergent diploid species, indicating that C. wootonii is indeed an alloploid and that C. lindheimeri and C. fendleri were involved in its formation (Fig. 3). Additional work is underway to more fully characterize the origins of triploid C. lindheimeri, C. wootonii, and other polyploid species of Cheilanthes (A. L. Grusz et al. unpubl. results).

**The Utility of the Nuclear gapCp Gene**—It is clear from the lack of branch support in our broad survey genealogy (Fig. 2A) that the small portion of the nuclear gapCp gene employed herein will be of little use for the deepest phylogenetic questions (only 296 bp of exon data are available for analysis; introns are unalignable at this level). However, our findings in cheilanthoids (Fig. 3), combined with the apparent universality of our primers (Table 1), indicate that this marker holds considerable potential for addressing species-level problems across the fern tree of life. Whether or not gapCp will be useful for addressing relationships at the generic level (i.e., within families) remains to be explored.

When working with gapCp, as with any nuclear marker, it will be important for researchers to appreciate the possibly confounding nature of multiple copies (Small et al. 2004). We have not yet determined the precise phylogenetic position of the gapCp duplication (Fig. 2); however, it is obvious that many ferns will have both “long” and “short” versions (Table 1), and there is certainly the potential for other more recent duplications. In addition to amplifying multiple copies of gapCp, our primers may even (in certain circumstances) amplify the nuclear gapCp gene. This mixture of PCR products in a single reaction could be viewed as problematic, but may also be seen as beneficial. More markers (i.e. more data) can be obtained from a single amplification. Cloning is required regardless, and can effectively isolate individual gene copies. Most gapCp paralogs are easily separated on an agarose gel (Table 1). If a specific copy of gapCp is desired, it can be sequenced and other products ignored. In any case, the risk of constructing a data set that combines different gapCp paralogs is alleviated by the fact that they are highly divergent, with introns that are not at all alignable. Of course, in some instances (e.g. if the desired product is not preferentially amplified) it would be advisable to develop copy-specific primers.

In our study, more than two unique but highly similar sequences corresponding to a particular paralog, and forming a clade, were occasionally recovered from a single diploid individual. Similarly, more than three sequences were sometimes recovered from a single triploid individual. As only two (or three) allelic sequences should have been found, these results suggest that either errors had been introduced by PCR amplification or that a species-specific duplication had occurred. Because relatively few clones were sequenced from any given individual, we made no attempt here to discriminate among allelic diversity, substitutional artifacts, and recent duplication events. However, future studies will need to consider these possibilities. Furthermore, those choosing to work with gapCp or any other nuclear marker must also be wary of the possible formation of chimeric sequences, especially when working with polyploid species (Bradley and Hillis 1997; Cronn et al. 2002).

**ACKNOWLEDGMENTS.** We are grateful to Donovan Bailey, Ann Kelsey, Jordan Metzgar, Nathalie Nagalingum, and Harald Schneider, as well as the New Mexico State University Herbarium, United States Department of Agriculture—Forest Service, United States Department of Defense—Fort Huachuca, United States Department of the Interior—Bureau of Land Management, and University of Utah Herbarium for their assistance in the acquisition of material newly utilized in this study. James Beck, Gordon Burleigh, Layne Huiet, and Carl Rothfels provided valuable comments and criticism. We thank David Hearn and Channa Pickett for their help in evaluating the utility of previously published primer sets. Funding for this study was provided in part by an NSF award to K.M.P. and M.D.W. (DEB-0717388), an NSF DDIG award to K.M.P. and E.S. (DEB-0408077), and an NSF CAREER award to K.M.P. (DEB-0347840).

**LITERATURE CITED**


clone 7, 600 bp, EU551287, clone 8, 600 bp, EU551288; *Cheilanthes fendleri* Hook., diploid, DB3177, *Schuettpelz 470* (DUKE), clone 7, 602 bp, EU551289, clone 8, 602 bp, EU551290, clone 13, 602 bp, EU551291, clone 14, 602 bp, EU551292, clone 15, 602 bp, EU551293, clone 16, 602 bp, EU551294, clone 17, 602 bp, EU551295, clone 18, 602 bp, EU551296, DB3690, *Windham 3468* (UT), clone 1, 602 bp, EU551297, clone 2, 602 bp, EU551298, clone 3, 602 bp, EU551299, clone 4, 602 bp, EU551300, clone 5, 602 bp, EU551301, clone 6, 602 bp, EU551302; *Cheilanthes lindheimeri* Hook., diploid, DB3157, Schuettpelz 450 (DUKE), clone 1, 601 bp, EU551318, clone 2, 601 bp, EU551319, clone 3, 601 bp, EU551320, clone 4, 601 bp, EU551321, clone 5, 601 bp, EU551322, clone 6, 601 bp, EU551323, clone 7, 601 bp, EU551324, clone 8, 601 bp, EU551325, DB3692, *Spellenberg 5065* (NMC), clone 4, 600 bp, EU551345, clone 5, 600 bp, EU551346, clone 6, 601 bp, EU551347, clone 7, 600 bp, EU551348, clone 8, 601 bp, EU551349, triploid, DB3147, *Schuettpelz 440* (DUKE), clone 2, 601 bp, EU551303, clone 3, 601 bp, EU551304, clone 5, 601 bp, EU551305, clone 6, 601 bp, EU551306, clone 7, 601 bp, EU551307, clone 8, 601 bp, EU551308, clone 9, 601 bp, EU551309, clone 10, 601 bp, EU551310, clone 11, 601 bp, EU551311, clone 12, 601 bp, EU551312, clone 13, 601 bp, EU551313, clone 14, 600 bp, EU551314, clone 15, 601 bp, EU551315, clone 16, 601 bp, EU551316, clone 17, 601 bp, EU551317, DB3196, *Schuettpelz 489* (DUKE), clone 1, 601 bp, EU551326, clone 2, 600 bp, EU551327, clone 3, 601 bp, EU551328, clone 4, 601 bp, EU551329, clone 5, 600 bp, EU551330, clone 6, 601 bp, EU551331, clone 7, 601 bp, EU551332, clone 9, 601 bp, EU551333, clone 10, 601 bp, EU551334, clone 11, 601 bp, EU551335, clone 12, 601 bp, EU551336, clone 13, 601 bp, EU551337, clone 14, 601 bp, EU551338, clone 15, 601 bp, EU551339, clone 16, 601 bp, EU551340, clone 17, 601 bp, EU551341, clone 18, 601 bp, EU551342, clone 20, 601 bp, EU551343, clone 21, 601 bp, EU551344; *Cheilanthes parryi* (D. C. Eaton) Domin, diploid, DB3802, Metzgar 149 (DUKE), clone 1, 596 bp, EU551350, clone 2, 596 bp, EU551351, clone 3, 596 bp, EU551352, clone 4, 595 bp, EU551353, DB3846, *Windham 3440* (UT), clone 1, 595 bp, EU551354, clone 2, 595 bp, EU551355, clone 3, 595 bp, EU551356, clone 4, 595 bp, EU551357, clone 5, 595 bp, EU551358, clone 6, 595 bp, EU551359, clone 7, 594 bp, EU551360, clone 8, 595 bp, EU551361, clone 9, 595 bp, EU551362, clone 10, 595 bp, EU551363; *Cheilanthes wootonii* Maxon, triploid, DB3693, *Windham 3409* (DUKE), clone 9, 600 bp, EU551379, clone 10, 602 bp, EU551380, clone 14, 600 bp, EU551381, clone 15, 600 bp, EU551382, clone 16, 602 bp, EU551383, clone 17, 599 bp, EU551384, clone 18, 602 bp, EU551385, clone 19, 602 bp, EU551386, clone 20, 600 bp, EU551387, clone 21, 600 bp, EU551388, clone 22, 602 bp, EU551389, clone 23, 602 bp, EU551390, clone 24, 602 bp, EU551391, DB3694, *Spellenberg 10467* (NMC), clone 1, 602 bp, EU551364, clone 3, 600 bp, EU551365, clone 5, 602 bp, EU551366, clone 6, 600 bp, EU551367, clone 7, 602 bp, EU551368, clone 8, 602 bp, EU551369, clone 9, 602 bp, EU551370, clone 10, 601 bp, EU551371, clone 11, 600 bp, EU551372, clone 12, 602 bp, EU551373, clone 13, 600 bp, EU551374, clone 14, 602 bp, EU551375, clone 16, 602 bp, EU551376, clone 17, 602 bp, EU551377, clone 18, 602 bp, EU551378.