Molecular Phylogenetics and Evolution 112 (2017) 36–46

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Insights into evolution in Andean *Polystichum* (Dryopteridaceae) from expanded understanding of the cytosolic phosphoglucose isomerase gene

Brendan M. Lyons, Monique A. McHenry, David S. Barrington *

Pringle Herbarium, Plant Biology Department, University of Vermont, Torrey Hall, 27 Colchester Ave, Burlington, VT 05405, USA

ARTICLE INFO

Article history:
Received 5 November 2016
Revised 8 March 2017
Accepted 10 April 2017
Available online 12 April 2017

Keywords:
Andes, ferns
Dryopteridaceae
Phosphoglucose isomerase
Polystichum
Allopolyploidy

ABSTRACT

Cytosolic phosphoglucose isomerase (*pgiC*) is an enzyme essential to glycolysis found universally in eukaryotes, but broad understanding of variation in the gene coding for pgiC is lacking for ferns. We used a substantially expanded representation of the gene for Andean species of the fern genus *Polystichum* to characterize *pgiC* in ferns relative to angiosperms, insects, and an amoebozoan; assess the impact of selection versus neutral evolutionary processes on *pgiC*; and explore evolutionary relationships of selected Andean species. The dataset of complete sequences comprised nine accessions representing seven species and one hybrid from the Andes and Serra do Mar. The aligned sequences of the full data set comprised 3376 base pairs (70% of the entire gene) including 17 exons and 15 introns from two central areas of the gene. The exons are highly conserved relative to angiosperms and retain substantial homology to insect *pgiC*, but intron length and structure are unique to the ferns. Average intron size is similar to angiosperms; intron number and location in insects are unlike those of the plants we considered. The introns included an array of indels and, in intron 7, an extensive microsatellite array with potential utility in analyzing population-level histories. Bayesian and maximum-parsimony analysis of 129 variable nucleotides in the Andean polystichums revealed that 59 (1.7% of the 3376 total) were phylogenetically informative; most of these united sister accessions. The phylogenetic trees for the Andean polystichums were incongruent with previously published cpDNA trees for the same taxa, likely the result of rapid evolutionary change in the introns and contrasting stability in the exons. The exons code a total of seven amino-acid substitutions. Comparison of non-synonymous to synonymous substitutions did not suggest that the *pgiC* gene is under selection in the Andes. Variation in *pgiC* including two additional accessions represented by incomplete sequences provided new insights into reticulate relationships among Andean taxa.

1. Introduction

Cytosolic phosphoglucose isomerase (*pgiC*, E.C. 5.3.1.9) is an enzyme essential to glycolysis, the cellular-respiration pathway known in almost all species of living things. Across kingdoms, a coding sequence of approximately 1700 base pairs (bp) codes for the polypeptide that constitutes the pgiC monomer (Thomas et al., 1992). Two monomers assemble to yield the functional pgiC dimer, assembling the catalytically critical region in the process. Although the size of the coding region is fairly similar across kingdoms, the number of exons and the size of the introns can vary dramatically across species—e.g. *Sus pgiC* 32 kb long including 18 exons (Claes et al., 1994), *Colias pgiC* over 10 kb long including 12 exons (Wheat et al., 2006), *Clarkia pgiC* 5.5 kb long including 23 exons (Thomas et al., 1992), and *Arabidopsis* 4.9 kb long including 22 exons (Thomas et al., 1993).*pgiC* polymorphisms are correlated with variation in temperature and related variables such as risk of desiccation across kingdoms (Riddoch, 1993). Riddoch assembled evidence to suggest that allozymes with charge differences represent alternate balances between enzyme stability and functional efficiency. The allozyme variants have in some cases been shown to be under natural selection (Gillespie, 1991; Riddoch, 1993), a common feature of metabolic enzymes in general (Marden, 2013). Both selection for temperature-based kinetic performance and structural stability at different temperatures have been documented. For instance, the sierra willow leaf beetle, *Chrysomela aeneicollis* (Schaeffer), has multiple *pgiC* alleles that vary along a cline that follows latitude and elevation (Dahlhoff and Rank, 2000; Rank and Dahlhoff, 2002). Directional selection of *pgiC* is also

* Corresponding author.
E-mail address: dbarring@uvm.edu (D.S. Barrington).

http://dx.doi.org/10.1016/j.ympev.2017.04.010
1055-7903/

observed along latitudinal clines in *C. aeneicollis*. Temperature-based selection of *pgiC* can favor heterozygous genotypes of *pgiC*, as documented for the sulfur butterflies (*Colias*, e.g. Watt, 1977). Among plants, wild populations of *Arabidopsis thaliana* harbor low levels of population-level DNA polymorphisms and a high level of non-synonymous nucleotide substitution, suggesting that the genetic variants are under natural selection (Kawabe et al., 2000).

Nucleotide variation of *pgiC* has also been studied in ferns, using exons 14–16 and their two included introns, with a focus on intraspecific variation and inferring origins of allopolyploids and hybrids. Recent work includes that on *Acrorumohra* (Chang et al., 2009), closely related species of Asian *Adiantum* (Wang et al., 2016), *Asplenium* (Schneider et al., 2013), *Dryopteris caudipinna* Nakai (Ishikawa et al., 2002), the apomictic complex centered on *Dryopteris varia* (L.) O. Kuntze (Hori et al., 2014), and variation in the fiddlehead fern *Matteuccia struthiopteris* (L.) Todaro (Koenemann et al., 2011). Earlier, isozyme studies often included pgiC as a key marker (e.g. Werth et al., 1985; Gastony, 1986, 1991). Isozyme analysis of *pgiC* in Central American *Polystichum* species revealed that they include at least five different *pgiC* alleles (including two intraspecific variants), implying different amino acid sequences (Barrington, 1990).

Here, we report the results of an inquiry into genetic variation in *pgiC* for selected Andean species of *Polystichum* based on much-expanded coverage of the gene derived from the development of new primers. The Andes host a recently evolved *Polystichum* lineage (McHenry and Barrington, 2014) that comprises an array of species found at diverse elevations and hence temperature regimes; their evolutionary history includes at least three major elevational transitions, both up and down the Andes. Our inquiry resulted in three basic outcomes. First, we provide a profile of variation across a large part of the gene in the context of variation in other lineages. Second, we explore both the utility of *pgiC* to evolutionary studies of the Andean species through consideration of both overall variation and analysis of non-synonymous and synonymous amino-acid substitutions. Finally, as there are a set of hybrids and polyploids in the study set, we use the expanded coverage of *pgiC* to develop new insights into their origin and potential selective advantage.

2. Materials and methods

2.1. Taxonomic sampling

A total of 11 *Polystichum* accessions were included in the study set, comprising eight taxa. Six species and a hybrid were chosen to include closely related taxa from a range of elevations in the central and northern Andes. A single Brazilian species (*P. pallidum* Gardner)—still from the same clade (McHenry and Barrington, 2014)—was included to extend the elevational range downward. The lowest accession sampled (*P. pallidum)* was collected at 480 m, the highest (*P. orbiculatum*) at 4650 m. The collections for this analysis were largely made during the investigation reported in McHenry and Barrington (2014). A complete list of sampled taxa, including accession numbers and voucher information, is provided in Appendix A.

2.2. DNA isolation, primer design, amplification, and sequencing

Leaf material was stored in silica-gel desiccant at −80 °C prior to extraction. Total genomic DNA was extracted from the silica-dried leaf material using a modified CTAB extraction protocol (Doyle and Doyle, 1987). Six primer pairs (Table 1) provided the basis for retrieving sequences from a total of six markers of the *pgiC* gene from Andean *Polystichums*. The first four markers, *pgiC* exons 3–7, 6–8, 7–10, and 9–13, were from the approximately 2000 bp area spanning *pgiC* exons 3–13. The remaining two markers, *pgiC* exons 14–17 and 16–19, were from the 1200 bp area from *pgiC* exons 14–19. The parts of the gene between exons 13 and 14, as well as inferred exons 1–3 and 19–22, failed to amplify.

Table 1
Primers developed during this study for the amplification of *pgiC* exons 3–13 and 14–19.

Primer	Primer sequence (F are 5′–3′, R are 3′–5′)	Annealing temp (°C)
3F	AGTGACAGCAGAAATCATGGAA	55
7R	AAGCCGACGTCCTTTACAAG	55
6F	AACTGGGAAACCATTGACTGATG	58
8R	TCAATAGGATCCACATTTGCCA	58
7F	TTGTAAAGGACGTCGGCTTC	57
10R	ACGGCATCTGGACTGCAAAT	57
9F	TGGATCACATCAGTCCTTGG	55
13R	GGAGCTGTTTTGAAATGGTTGT	55
14F	GGGTCTTTTGAGTGTTTGG	55
17R	CCAATGAAATCACATGGAATAACACG	55
16F	CTTAGTATGGAAAGCAATGGAAAGG	55
19R	TTATGGGGCACCAAATGTTCC	55

The first *pgiC* primers (to amplify exons 14–16) for *Polystichum* were adapted from those used by Ishikawa et al. (2002) to amplify *Dryopteris caudipinna* Nakai (Dryopteridaceae). Primer sets for *pgiC* exons 16–19 were designed by Barrington lab members from an alignment of an array of *pgiC* sequences available on Genbank. Carl Rothfels aided in design of the remaining primers by sharing the *pgiC* cDNA sequence for *Polystichum acrostichoides* (Michx.) Schott (see Rothfels et al., 2013). This same cDNA was also used to modify previously designed primers in order to improve amplification.

Amplification by the polymerase chain reaction was performed in a TC-3000, TC-3000X or a 3Prime thermal cycler (Techne, Burlington, New Jersey, USA) in 25 μL aliquots with the following components: 150 ng of genomic template; 0.1 μM of each primer; 1x Denville PCR Buffer (Denville Scientific Inc., South Plainfield, New Jersey, USA); 200 μM L^{-1} of each dNTP; and 0.625 U Choice-Taq DNA Polymerase (Denville Scientific Inc.). All sequences were amplified as follows: (1) initial denaturation at 94 °C for 2 min; (2) followed by 35 cycles (94 °C for 30 s, 55–58 °C (annealing temperatures were adjusted for each marker set, see Table 1) for 30 s, 72 °C for 1 min; and (3) a final extension at 72 °C for 10 min. PCR products were visualized on an agarose gel; successfully amplified samples were purified using ExoSAP-IT (USB Corporation, Cleveland, Ohio, USA). Purified PCR products were further cleaned and labeled using the ABI Prism BigDye Terminator cycle sequence ready reaction kit (Applied Biosystems/Life Technologies, Foster City, California, USA).

Two bands were visualized in *pgiC* exons 14–17 for three accessions of *P. orbiculatum* (Desv.) Remy & Fée in Gay (*Barrington 2143*, *Barrington 2144*, and *McHenry 10-06*). Each band was excised for separate amplification and sequencing using a Prep-Ease gel extraction kit (Affymetrix, Santa Clara, California, USA).

Sequences were resolved on an ABI Prism 3730xl DNA analyzer (High Throughput Genomics Center, Seattle, WA, USA). Consensus sequences were assembled from the forward and the reverse sequences for *pgiC* using Geneious Pro version 6.1.5 (Drummond et al., 2007).

2.3. Phylogenetic analysis

Maximum Parsimony (MP) analysis was carried out using TNT v1.1 (Goloboff et al., 2008). All characters were equally weighted and coded as unordered. A heuristic search was performed with 1000 parsimony-ratchet replicates (Nixon, 1999: 200 ratchet iterations, the up and down weights set to 5% each) holding 20 trees

per ratchet, followed by tree-bisection-reconnection (TBR) branch swapping. Bayesian Inference (BI) was carried out on the CIPRES Science Gateway (Miller et al., 2010) using Mr. Bayes v. 3.2 (Ronquist et al., 2012). A Markov Chain Monte Carlo analysis was performed for the concatenated *pgiC* sequences with four independent Markov chains run for 10 million generations and trees sampled every 1000 generations. Stationarity was determined using the log-likelihood scores for each run plotted against generation in the program Tracer version 1.5 (Rambaut and Drummond, 2007). 10% of the trees were discarded as a burn-in phase, and a 50% majority-rule consensus tree was calculated for the remaining trees. Our *Polystichum stuebelii* Hieron. accession was used as the outgroup for the phylogenetic analyses, because it was resolved as sister to the remaining species in McHenry and Barrington (2014).

2.4. Cross-kingdom alignment

With the goal of placing the observed variation in non-synonymous sites in the context of similar variation in other lineages, we developed an alignment of the exons from the most complete *Polystichum sequence* in our dataset (*P. orbiculatum*, *McHenry 10-76*) with pgi coding sequences for an array of taxa for which information is available about variation within and between species. The array comprised a fern—*Dryopteris*; three angiosperms, *Arabidopsis, Clarkia*, and *Dioscorea*; and a butterfly, *Colias*. Also included for further context was the amoebozoan *Acanthamoeba*. We started with complete or largely complete sequences for the gene (details and sources, Appendix B). For comparisons between kingdoms, we aligned the amino-acid translations for these *pgiC* sequences as both intron number and length differ dramatically between kingdoms. Alignment was with the program ProbCons (Do et al., 2005). We then used the coding regions from the original DNA sequences to construct an alignment of the coding DNA. Finally, we located the non-synonymous substitutions in the Andean dataset on the cross-kingdom alignment.

2.5. Tests for selection

We used the CodeML module in PAML (Yang, 2007; Xu and Yang, 2013) to test for selection. CodeML is most appropriate for interspecific comparisons where divergence is sufficient to provide a basis for analysis, as in our Andean dataset. To develop the data for the analysis we built an aligned exon file by removing the non-coding regions from the nine full *Polystichum* PGI sequences. A new phylogenetic analysis was performed using the same parameters as in the full phylogenetic analysis, with *P. stuebelii* as the outgroup, to yield the tree for the PAML analysis. The tree was congruent with the full-data tree.

To test the premise that adaptive evolution is occurring at a few time points and affects a few amino acids, we chose the branch-site model (Model B [Yang and Nielsen, 2002]), which allows the ω ratio (dN/dS) to vary both among sites and among lineages. When the estimate of ω_2 is greater than 1, some sites are under positive selection along the branches of interest. We ran these tests for all the branches in the retrieved exon-only phylogeny. We chose CodonFreq to be 3X4 to implement distinction between transitions and transversions. The remaining parameters were set at default values for the program. To identify amino acid sites potentially under positive selection, the parameter estimates from the branch-site model analysis were used to calculate the posterior probabilities that an amino acid belongs to a class with $\omega > 1$, using the Bayes empirical Bayes (BEB) approaches implemented in PAML (Yang et al., 2005).

3. Results

3.1. A profile for pgiC in Polystichum

The amplification of *pgiC* in Andean *Polystichum* resulted in complete sequences of two central areas of the gene for nine accessions representing seven species and a hybrid (for genbank accession numbers, see Appendix A). We characterized 17 exons and 15 introns in the portion of *Polystichum pgiC* retrieved here. We numbered the exons in our sequences (Fig. 1) by comparison to sequences for *Dryopteris* species (Ishikawa et al., 2002) and *Arabidopsis* species (Thomas et al., 1993). One of our areas comprises between 2076 and 2107 bp and ranges from the end of exon 3 to the beginning of exon 13. The second area comprises between 993 and 1216 bp, starting at the end of exon 14 and ending halfway through exon 19.

The aligned dataset for the nine complete sequences totaled 3376 bp, of which 148 (4.38%) were variable. 2242 bp (66.41%) were intronic (Table 2); 123 of these were variable. The number of introns is the same as in the equivalent portions of *pgiC* in *Arabidopsis*. Introns range from 67 bp (intron 9) to 533 bp (intron 7); the average length was 138 bp. Although none of the known introns in *Polystichum pgiC* is the same size as their corresponding introns in *A. thaliana*, the average size of introns in *Polystichum* (138 bp) and *A. thaliana* (136 bp) is similar. The introns also included 32 indels varying in size from one to 188 bp.

The remaining 1134 bp (33.59%) of the total were exonic (exon characteristics, Table 3); exons range from 43 bp (exon 7) to 156 bp (exon 5) with an average length of 79 bp. All of the 13 complete exons we retrieved from *Polystichum pgiC* match the exact size of the corresponding exons in *A. thaliana*. Only 26 of the 1134 vary: 19 code for synonymous and seven for non-synonymous substitutions (Fig. 2). Three of the amino-acid transformations involve alanine (two, full-dataset characters [see Fig. 2] 95 and 2756, a gain; one, character 416, a loss). A fourth yields a change from negative to positive charge (character 118, glutamic acid to lysine). The remaining three are transformations between two non-polar amino acids (character 541, valine to isoleucine; character 2297, phenylalanine to leucine; and character 3339, proline to glutamine).

3.2. Phylogeny

The Bayesian-inference analysis yielded two clades (Fig. 2). The larger clade, clade one (retrieved with weak support, PP = 0.86), consists of the two *P. orbiculatum* accessions, *P. gelidum* (Kunze ex Klotzsch) Fée, *P. pallidum* Gardner, and *P. polyphyllum* (C.Presl) C.Presl. Within this clade, a subclade comprising the accessions of *P. gelidum* and *P. orbiculatum* was retrieved with strong support (PP = 1). Clade 2 comprises *P. platyphyllum* (Willd.) C.Presl and the hybrid *P. platyphyllum* × *P. bulbiferum* Barrington, retrieved with strong support (PP = 1). *Polystichum dubium* (H.Karst.) Diels lay outside these two clades. Notably, the distribution of character-state changes on the tree is unbalanced. A large number of changes are autapomorphic (73), and among the remainder many (41) unite accession pairs. The early divergences in the tree are supported by single character-state changes, and the early events are unresolved. Results from the parsimony analysis were congruent with those from the Bayesian analysis.

114 transformations (84% of the whole) are intronic; of these 51 are synapomorphic and 63 autapomorphic. Thirty are shared by *P. platyphyllum* and *P. platyphyllum* × *P. bulbiferum*. The two accessions of *P. orbiculatum* share 10 synapomorphies, and the remaining 11 unite the two *P. orbiculatum accessions* and *P. gelidum*. Three introns (7, 15, and 18) contain much more variability than any of

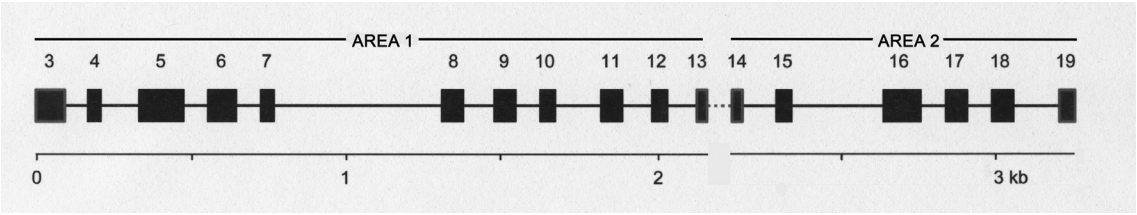

Fig. 1. Proposed gene structure for the two areas of *pgiC* amplified for Andean *Polystichum*. Exons numbered following those for *Arabidopsis* (Thomas et al., 1993). Presumptive intron 13, not retrieved here, represented by a dotted line.

Table 2
Length and nucleotide variation of introns retrieved from *pgiC* in Andean *Polystichum*.

Intron	Length (bp)	Variable characters	Informative characters
3	77	10	7
4	113–124	10	1
5	86	7	3
6	84–85	6	4
7	503–533	28	12
8	92–96	7	3
9	67–68	5	2
10	134–146	3	3
11	88	4	3
12	91–92	5	4
14	92	3	3
15	207–391	14	7
16	80	5	3
17	78	3	2
18	138–183	13	8
Total	1930–2219	123	65

the other sampled introns. In Andean *Polystichum*, *pgiC* intron 15 ranges from 207 to 391 bp long and contains 14 variable characters, five of which are informative. Intron 15 also contains twelve indels, four informative; one of the indels is 188 bp long—the major source of variation in the size of this intron. Despite intron 15's documented variation in the literature, the most variable region we retrieved was intron 7, which contains 28 variable characters, 12 informative. Intron 7 is the longest intron sequenced for Andean *Polystichum pgiC*; it ranges from 503 to 533 bp in length. We also revealed four microsatellite loci from intron 7 in our sample, three of which were variable in repeat number between species (Fig. 3). The third unusually variable region is intron 18, which has 13 variable characters (eight informative). A single autapomorphic indel spanning 58 bp in intron 18 accounts for most of the variation in length, several smaller indels also add to the variation in length; the intron ranges from 137 to 183 bp. As a whole these three introns contain 55 (42.64%) of the 129 variable characters and 25 (49.02%) of the 51 informative characters found in the introns.

Of the 26 substitutions in exons, 12 are informative (Fig. 2). Four of the informative synonymous substitutions are ancestral to clade 2; one unites clade 1, and the other two provide resolution within clade 1. The 10 autapomorphic synonymous substitutions are broadly distributed across accessions. *Polystichum stuebelii* (the outgroup for this study) has the most (four); *P. gelidum* and *P. platyphyllum* × *P. bulbiferum* have none.

Now considering just the seven non-synonymous substitutions, five were informative. In clade one, four united all the accessions but *P. polyphyllum* and three were ancestral to the subclade comprising *P. gelidum* and the two accessions of *P. orbiculatum* (Fig. 2). The remaining informative non-synonymous substitution united the two accessions in clade two. The two autapomorphic non-synonymous substitutions were unique to *P. dubium*.

3.3. Tests for selection

Bayes Empirical Bayes (BEB) analysis (Yang et al., 2005) did not reveal any sites at which there was evidence of positive selection.

3.4. Cross-Kingdom alignment

Alignment of the *pgiC* coding sequences retrieved from *Polystichum* with three angiosperms (*Arabidopsis*, *Clarkia*, and *Dioscorea*), an amoebozoan (*Acanthamoeba*), and the sulfur butterfly (*Colias*) (summary, Table 4; complete alignment, Supplemental Data) provided a context for interpreting the seven non-synonymous substitutions in Andean *Polystichum*. Among these taxa, non-synonymous substitutions are strongly clustered in three exons

Table 3
Size and nucleotide variation of exons retrieved from the *pgiC* gene in Andean *Polystichum*. Partial sequences in italic.

Exon	Length (bp)	Variable characters	Parsimony-informative characters	Non-synonymous substitutions	Synonymous substitutions
3	*16*	*0*	*0*	*0*	*0*
4	77	2	1	2	0
5	156	2	1	1	1
6	97	3	3	1	2
7	43	1	0	0	1
8	68	0	0	0	0
9	85	3	2	0	3
10	49	3	1	0	3
11	73	1	1	0	1
12	61	2	2	0	2
13	*22*	*0*	*0*	*0*	*0*
14	*35*	*0*	*0*	*0*	*0*
15	54	2	1	1	1
16	129	2	1	1	1
17	64	2	1	0	2
18	74	1	1	0	1
19	*31*	*1*	*1*	*1*	*0*
Total	1134	25	16	7	18

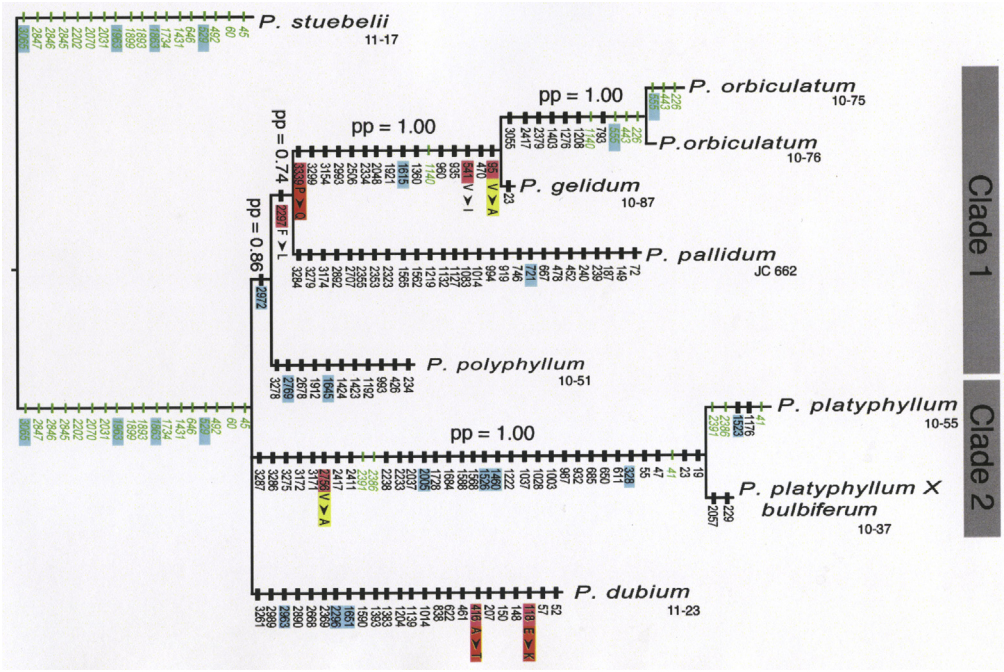

Fig. 2. Phylogeny of Andean *Polystichum* for all amplified regions of *pgiC* (exons 3–19 and introns 3–12 and 14–18) constructed by Bayesian Inference analysis. Each number represents the position of a variable character in the full alignment (including both introns and exons). Characters in green are homoplastic. Synonymous substitutions in blue; non-synonymous substitutions in red (with charge-changing changes highlighted).

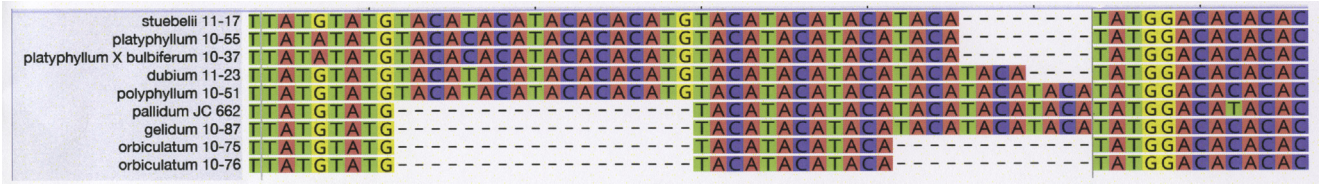

Fig. 3. One of the variable microsatellite loci in intron 7, each based on the TACA quartet motif. The variation in repeats at these loci is consistent with the phylogeny retrieved in the analysis illustrated in Fig. 2.

Table 4
Summary of infra- and interspecific variation for non-synonymous changes in *pgiC* in ferns, angiosperms, and a butterfly. Appendix C provides the data.

Plant Exon Number	Exon length	Total changes	Interspecific changes	Infraspecific changes	Charge changes	Alanine involved
1	53	1		1		
2	101	5	5		2	1
3	82	7	4	3		1
4	48	6	5	1	1	3
5	156	20	12	8	6	6
6	97	5	4	1	1	
7	43					
8	68	3	3			1
9	91	2	1	1		2
10	49	3	2	1	1	
11	73	2	2			1
12	61	1		1		1
13	67	4	1	3	2	1
14	57	1	1		1	
15	54	1		1		
16	132	8	4	4	2	1
17	67	2	2			
18	74					
19	78	10	5	5	4	5
20	69	3	2	1		1
21	98	2	1	1		
22	111	4	4		1	1
23	–					
Totals		19	12	7	21	25

(5, 16, and 19) together having 33 (43%) of the 77 substitutions scored. Considering the seven non-synonymous substitutions retrieved from Andean *Polystichum*, three were from these labile exons (one from exon 5, one from exon 16, and one from exon 19). In contrast, four substitutions lie in exons outside the highly labile group, two in exon 4, one in exon 6, and one in exon 15. The substitution-rich exons are heterogeneous. For instance, in the cross-kingdom alignment, exon 16 is remarkable for two stable regions, one near the beginning and one at the end, that are uniform across all the taxa and contain active sites reported by Troncoso-Ponce et al., 2010 (Fig. 4). Of the 43 amino acids coded by exon 16, the cross-kingdom alignment reveals 39 to be uniform across all plants, including 24 uniform across kingdoms. At the same time, the central region of the exon is quite labile; the catalytically critical region of Wheat et al. (2006) lies here. Exon 5 is similar, with a total of 29 (56%) out of 52 amino acids constant across plants, including 18 constant across kingdoms.

3.5. Two-nucleotide calls

The chromatograms serving as the basis for the Andean *Polystichum* dataset included a total of 16 sites at which we encountered two-nucleotide calls (double peaks, Table 5). All of the two-nucleotide calls included one nucleotide common at that site in the remaining sequences. Ten of the calls were from two complete sequences, those for the putative hybrid *Polystichum platyphyllum* × *P. bulbiferum* and the tetraploid *Polystichum gelidum*. The remaining complete sequences were without double peaks. *Polystichum platyphyllum* × *P. bulbiferum* had five double peaks; for three of these *P. platyphyllum* was the only accession in the dataset with one of the peaks. *Polystichum gelidum* had five double peaks; three of these combine nucleotides common to all the accessions with nucleotides otherwise absent from our sample. One sums a nucleotide unique to the two complete *P. orbiculatum* accessions from Bolivia and a nucleotide unique to *P. polyphyllum*, and one again sums a nucleotide unique to *P. orbiculatum* with a nucleotide common in the dataset.

Three additional accessions of *Polystichum orbiculatum*, two from Costa Rica (*Barrington 2143* and *2144*) and one from northernmost Ecuador (*McHenry 11-06*), yielded only partial sequences. All three had double peaks, and all had a double-banded phenotype for *pgiC* 14–17 in agarose gels not seen in the complete sequences. All told, these partial *P. orbiculatum* sequences yielded seven two-nucleotide calls: five of these combined a nucleotide unique to *P. polyphyllum* (*McHenry 10-51*) with a nucleotide common in the dataset. One summed a nucleotide unique to the two complete *P. orbiculatum* accessions from Bolivia (*McHenry 10-75* and *10-76*) with a common nucleotide. The last combined a common nucleotide with one not found otherwise in the dataset. Sequences of the two bands retrieved for *pgiC* 14–17 isolated from these accessions each yielded one of the nucleotides contributing to the two double peaks in the region. The sequences from the two bands differed by the longest of the indels located in intron 15. The shorter of the two sequences was also recovered from the complete *P. polyphyllum* accession (*McHenry 10-51*); the longer sequence was shared by the remaining accessions.

One of the double peaks segregating in the two sequences isolated from *pgiC* 14–17 from the incomplete Costa Rican accessions, at position 2770 in the complete alignment and 1162 in the exon-only alignment, lies in exon 16; as a result *pgiC* in these plants codes for two different amino acids, glutamic acid and glutamine, at this site (Appendix C).

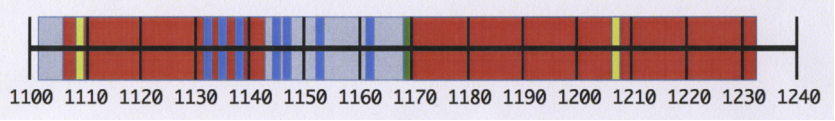

Fig. 4. Profile of variation in Exon 16. Conserved regions in red, sites with non-synonymous variation included in this study in bright blue, location of amino-acid polymorphism in allopolyploid *Polystichum orbiculatum* in green, catalytically active sites reported by Troncoso-Ponce et al. (2010), in yellow. Positions of sites detailed in Appendix C.

Table 5
Variation in *pgiC* nucleotides and indels relevant to interpreting hybrid and polyploid origins in Andean *Polystichum*. Incomplete Costa Rican and Ecuadorian accessions of *P. orbiculatum* shaded. Exon 14–17 region (sites 2137–2984 in the alignment of exons and introns) set off between vertical bars. Queries (?) represent ambiguities imposed by indel.

SITE	41	210	1140	1192	1423	1424	1912	2065	2391	2523	2678	2709	**2769**	**2770**	3055	3077
1076	G	G	A	G	C	A	A	T	T	A	G	A	T	G	A	C
1075			A	G	C	A	A	T	T	A	G	A	T	G	A	C
1087	G	K	R	G	C	A	A	W	T	A	G	R	Y	G	G	C
1106 long									T	?	G	A	C	C		
1106 short									C	?	T	A	C	G		
1106															R	
2144 long									T	?	G	A	C	C		
2143 short									C	?	T	A	C	G		
2143		G	G	R	Y	R	M	T			T	A	C	S	R	
1051	G	G	G	A	T	G	C	T	T		T	A	C	G	G	C
1055	A	G	G	G	C	A	A	T	G	A	G	A	T	G	G	C
1037	R	G	G	G	C	A	A	T	K	W	G	A	T	G	G	Y
1123	G	G	G	G	C	A	A	T	T	A	G	A	T	G	G	C
JC662	G	G	G	G	C	A	A	T	T	A	G	A	T	G	G	C
1117	G	G	G	G	C	A	A	T	T		G	A	T	G	G	C

4. Discussion

4.1. Overview of cytosolic phosphoglucose isomerase in ferns

The basic structure of *pgiC* in *Arabidopsis thaliana* spans about 4700 bp, including 1683 bp of coding sequence or 561 codons, broken up into 22 exons and 21 introns (Thomas et al., 1993). Given the exact match in length of the 13 complete exons we recovered for *Polystichum* to those of *Arabidopsis thaliana*, the similarity in length of the cDNA for the two, as well as their similar average intron size, we project that the complete *pgiC* gene is about 4600 bp long in *Polystichum*. Similarly, a total of 22 exons and 21 introns is likely. The similarity of *pgiC*'s structure in *A. thaliana* and *Polystichum* is striking given that the lineages have been isolated since the Devonian (Magallón et al., 2013), and the *Polystichum acrostichoides* genome is 65 times larger than that of *Arabidopsis* (Bainard et al., 2011; Bennett and Leitch, 2011). At least for *pgiC*, increased intron size does not contribute to increased genome size.

Our comparison of *Polystichum* substitutions to those for angiosperms, an amoebozoan, and a butterfly provide an expanded context for interpreting evolution in the coding regions of *pgiC*. Two exons, 5 and 19, emerge as most labile, along with exon 16—already identified as catalytically critical (Wheat et al., 2006). Most of the variation in exon 16 lies in the region identified by Wheat et al. (2006) as critical to optimizing enzyme function and the target of selection for functional variants (indicated in Appendix C). The two exon-16 *Polystichum* substitutions lie in this immediate region. The upstream *Polystichum* substitution in this region, from valine to alanine, is one of the substitutions reported by Cunningham and Wells (1989) to have significant impact on enzyme function through changes to binding affinity. The downstream substitution from glutamic acid to glutamine, at position 1162 in the exon-only dataset, was only recovered from the incomplete sequence from Costa Rica. By contrast, the pattern of remarkable stability in the upstream and downstream ends of exon 16 that surround the labile region so critical to enzyme function suggests that a tightly controlled structural framework for this region is essential. Similarly, exon 5, identified by Kawabe et al. (2000) as the location of the substitution critical to the key allozyme polymorphism in their sample of *Arabidopsis thaliana* populations and the location of one of our non-synonymous substitutions, is a mosaic of labile and stable amino-acid sequences. However, the single substitution in Andean *Polystichum* most likely to influence enzyme mobility in isozyme studies through change in net charge lies in the less labile exon 4.

4.2. Utility of expanded representation of pgiC for evolutionary studies

What is the value of the expanded coverage of the *pgiC* gene in ferns for phylogenetic inquiries? We retrieved an overwhelming number of autapomorphic changes unique to individual accessions and single species in this inquiry (Fig. 2). This pattern documents a rapid rate of evolutionary change, concentrated in the introns. Most of the characters informing earlier events are in the much less variable exons. Given this pattern, the loss of transformations witnessing deeper evolutionary history of the genus through multiple hits is likely. Alternatively, a rapid radiation leading to the speciation events, followed by post-divergence accumulation of mutations, is possible. In either case, *pgiC* appears to be most valuable for inferring very recent evolutionary history, either at the population level or in closely allied species.

The phylogeny we recovered for *pgiC* for a selected set of Andean *Polystichum* species is incongruent with that of McHenry and Barrington (2014), based on an array of cpDNA markers for a broader sample of Neotropical species. Differences in sampling may be the explanation for the difference, but rapid evolution resulting in loss of witness to earlier history may well account for the incongruence. The problems encountered by Zhang et al. (2016) in their attempt to include pgiC exons 14–16 in phylogenetic work on the Tectariaceae may well be the result of a similar rapid evolutionary rate in that family. However, identification of the highly variable introns 7 and 18 in addition to the well-known intron 15 offers the potential for fine-grain species and population work, for instance in identifying polyploid progenitors and testing hypotheses for multiple origins of allopolyploids—as recently done for *Polystichum braunii* (Spenner) Fée (Jorgensen and Barrington, 2017). The microsatellites encountered in intron 7 are especially suited to this latter agenda.

Our tests for selection did not evidence impact by positive selection on variation in *pgiC*. Though so far there is an absence of evidence for optimization of *pgiC* for different climates in the Andes, an expanded survey of the lineages in the clade from the Andes would be wise before concluding that selection is inoperative.

4.3. Insights into hybridization and polyploidy in Andean Polystichum

In Tropical America, polyploidy has a role in the evolutionary history of *Polystichum*, based on work in the Talamanca Range of Costa Rica and Panamá, where three allopolyploids have been documented among a total of 12 *Polystichum* species (Barrington, 1990, 2003). In the Andes, few polyploids have so far been encountered, but two are documented—both in our sample: *Polystichum orbiculatum* and *P. gelidum* (McHenry, 2012). The work in the Talamanca Range provided isozyme and chromosome-pairing data to argue that *P. orbiculatum* is an allotetraploid there.

When species have different nucleotides at sites in nuclear markers, their hybrids and derived allopolyploids commonly sum these nucleotides (Jorgensen and Barrington, 2017), yielding two-nucleotide calls (double peaks) in samples of nuclear markers. For the testing of hypotheses for hybrid origins, ideally these differences would be fixed across the progenitor species. The accessions in our dataset included a putative hybrid (*Polystichum platyphyllum* × *P. bulbiferum*) and two tetraploid species (*P. gelidum* and *P. orbiculatum*). All of the accessions in our data with two-nucleotide calls (Table 5) were retrieved from these accessions; we take the remaining species to be diploid. Analysis of these accessions yielded a set of insights into hybridization and polyploidy in Andean *Polystichum*.

Two-nucleotide calls recovered from the putative hybrid (*Polystichum platyphyllum* × *P. bulbiferum*) corroborated *P. platyphyllum* as one of its progenitors. All of the double peaks comprise a nucleotide found in *P. platyphyllum* with a nucleotide unique in our dataset; in contrast *P. platyphyllum* was without two-nucleotide calls. A similar pattern of double peaks recovered from the tetraploid *P. gelidum* suggests that it is an allopolyploid (hence of hybrid origin). Each of the double peaks includes a nucleotide found elsewhere in the dataset. Three of the five include a second nucleotide unique in the dataset; two combine nucleotides of different accessions among those without double peaks in the dataset. As *P. gelidum* resolves with *P. orbiculatum* and *P. sodiroi* in both the *pgiC* tree (here) and the cpDNA phylogeny (McHenry and Barrington, 2014), its ancestry somehow relates to these species.

Based on the two-nucleotide call data, we argue that our dataset includes among the accessions of *P. orbiculatum* two entities: (1) incomplete sequences of allotetraploids from the northern Andes one of whose progenitors is *P. polyphyllum* and (2)

complete sequences of diploids from the central Andes that were not involved in the origin of the tetraploids. The double peaks and the two mobilities (related to the large indel in intron 15) for *pgiC* we encountered in our incomplete accessions of *P. orbiculatum* from the Northern Andes and Mesoamerica we take as evidence that genomes of two species contributed to an allotetraploid. Supporting evidence is that the tetraploid count for Mesoamerican *P. orbiculatum* (Barrington, 1990) comes from the same population as the Mesoamerican plants sampled for this study. By contrast, the complete sequences retrieved from our accessions of *P. orbiculatum* from Bolivia in the Central Andes (*McHenry 10-75, 10-76*) are probably from a diploid species as they lack double peaks and different PCR-product mobilities in our gels. These two accessions are missing key unique features, both indels and SNPs, found in the presumed tetraploids from Ecuador and Costa Rica. Thus, they do not represent one of the progenitors of tetraploid *P. orbiculatum*.

Further insight comes from consideration of the sequences retrieved from exons 14–17 of the presumed tetraploid *P. orbiculatum* (*Barrington 2143, 2144*, and *McHenry 11-06*). The shorter of the two sequences retrieved from these accessions was otherwise found only in our accession of *P. polyphyllum* from Ecuador (*McHenry 10-51*); we argue that it represents the lineage that contributed one of the genomes to these accessions of *P. orbiculatum* from the northern Neotropics. The five SNPs unique to the tetraploids and *P. polyphyllum* add strength to this conclusion. These observations, considered in the light of the provenance of the accessions, lead to the working hypothesis that *P. orbiculatum—as currently delineated*—comprises at least two entities. The northern Andes host an allotetraploid species; one of its progenitors is *P. polyphyllum*. The central Andes host a diploid species that, though morphologically quite similar to the northern lineage, is not involved in its origin.

Finally, resolution of the two entities in *Polystichum orbiculatum* opens the way to insights into discovering possible evolutionary advantages for the allotetraploid originating in the northern Andes. The incomplete sequences of Northern Andean tetraploids include a two-nucleotide call at our position 2770 (Table 5, position 1162 in the cross-kingdom alignment, Appendix C), which lies close to the catalytically critical site 375 identified for *Colias* (Wheat et al., 2006). This amino-acid polymorphism results in both the negatively charged glutamic acid and the charge-neutral glutamine being coded for at this site, yielding a mix of two monomers that may have different physiological optima. Perhaps the tetraploid, like the heterozygous butterflies (Wheat et al., 2010), profits by balancing selection favoring a mix of these monomers of the pgiC enzyme.

5. Conclusions

Based on our sample of seven *Polystichum* species, the coding sequence for cytosolic phosphoglucose isomerase in ferns is highly conserved, allowing cross-kingdom alignment of the amino-acid sequences. Exon number and size is conserved relative to angiosperms; arthropods and amoebozoans have a different exon design, but share highly conserved sequences across the gene. In contrast, introns are variable and unique, so variable that phylogenetic signal appears to be weak. Three introns (7, 15, and 18) are unusually variable; introns 7 and 18 are new tools for interpreting recent evolutionary history in the ferns. Expanded coverage of *pgiC* allows us to propose that the widespread páramo allotetraploid species *P. orbiculatum* comprises two entities. The northern Andes host an allotetraploid species; one of its progenitors is *P. polyphyllum*. The central Andes host a diploid species that, though morphologically quite similar to the northern lineage, is not involved in its origin. Our future goal will be to recover the full evolutionary history of the Andean páramo reticulate lineage.

Acknowledgments

Funding: This work was supported by CSREES Grant VT-H01405 to DSB and an Undergraduate Research Endeavors Competitive Award from the University of Vermont to BL.

We thank Stacy Jorgensen for her excellent training, the members of BML's Distinguished Undergraduate Defense committee (Douglas Johnson, Jeanne Harris, and Catherine Paris) for helpful suggestions for development of the ideas in this work, and Stephen Keller for his extraordinarily constructive and insightful review of the manuscript and advice on PAML. The members of the Barrington lab at the University of Vermont provided enthusiastic encouragement throughout.

Appendix A

Sources of material for DNA sequencing. Data sequence: Taxon, voucher specimen, country, province, elevation, Herbarium (VT = Pringle Herbarium, University of Vermont), marker or region and Genbank accession number (for each marker, region, or area, separated by semicolons).

A.1. Complete sequences

Polystichum dubium: *M. A. McHenry 11-23*, Ecuador, Pichincha Prov., 2655 m, VT—exons 3 (partial) to 13 (partial), KY610516; exons 14 to 19 (partial) KY610525. **Polystichum gelidum:** *M. A. McHenry 10-87*, Bolivia, La Paz Prov., 3200 m, VT—exons 3 (partial) to 13 (partial), KY610512; intron 14 to exon 19 (partial) KY610521. **Polystichum orbiculatum:** *M. A. McHenry 10-75*, Bolivia, La Paz Prov., 4669 m, VT—exons 6 (partial) to 13 (partial), KY610511; exons 14 to 19 (partial) KY610520. **Polystichum orbiculatum:** *M. A. McHenry 10-76*, Bolivia, La Paz Prov., 4650 m, VT—exons 3 (partial) to 13 (partial), KY569396; exons 14 to 19 (partial) KY610519. **Polystichum pallidum:** *J. Condack 662*, Brazil, Est. Rio de Janeiro., 480 m, VT; exons 3 (partial) to 13 (partial) KY610517; exons 14 to 19 (partial) KY610526. **Polystichum platyphyllum:** *M. A. McHenry 10-55*, Ecuador, Napo Prov., 1729 m, VT—exons 3 (partial) to 13 (partial) KY610514; exons 14 to 19 (partial) KY610523. **Polystichum platyphyllum × bulbiferum:** *M. A. McHenry 10-37*, Ecuador, Zamora-Chinchipe Prov., 1640 m, VT—exons 3 (partial) to 13 (partial) KY610515; exons 14 to 19 (partial) KY610524. **Polystichum polyphyllum:** *M. A. McHenry 10-51*, Ecuador, Pichincha Prov., 3870 m, VT—exons 3 (partial) to 13 (partial) KY610513; exons 14 to 19 (partial) KY610522. **Polystichum stuebelii:** *M. A. McHenry 11-17*, Ecuador, Imbabura Prov., 2417 m, VT—exons 3 (partial) to 13 (partial) KY610518; exons 14 to 19 (partial) KY610527.

A.2. Incomplete sequences

Polystichum orbiculatum: *D. S. Barrington 2143*, Costa Rica, Pr. San José, 3400 m, VT—exon 4 (partial) to intron 19 (partial) KY696278; exons 14 (partial) to 17 (partial) SHORT ALLELE KY696279. *D. S. Barrington 2144*, Costa Rica, Pr. San José, 3400 m, VT—exons 14 (partial) to 16 (partial) LONG ALLELE KY696280; exon 14 (partial) to intron 15 (partial) SHORT ALLELE KY696281. **Polystichum orbiculatum:** *M. A. McHenry 11-06*, Ecuador, Pr. Pichincha, 3495 m, VT—exons 14 (partial) to intron 17 (partial) KY696275; exons 14 (partial) to intron 16 (partial) KY696276; exons 16 (partial) to 19 (partial) KY696277.

Appendix B. Sources of the complete or broadly representative sequences for the cross-kingdom comparisons

Organism	Genbank Accession Numbers	Coverage	Literature Citation
Acanthamoeba castellanii	XM_004336611	Single sequence	Clarke et al. (2013)
Arabidopsis lyrata	AB044948–67	Intraspecific only	Kawabe et al. (2000)
Clarkia lewisii	X71085, X89385	Intraspecific (*pgiC*2)	Thomas et al. (1992)
	X89384,86,89,90,92,94–97	Interspecific (*pgiC*1)	Gottlieb and Ford (1996)
Colias eurytheme	DQ205063–76	Intraspecific only	Wheat et al. (2006)
Dryopteris caudipinna	AB066272	Single sequence	Ishikawa et al. (2002)
Dioscorea tokoro	D88920–30, AB006088	Intraspecific and Interspecific	Terauchi et al. (1997)
Polystichum orbiculatum	KY569396, KY610519	Interspecific	Lyons et al. (this publication)

Appendix C. Intra- and interspecific variation for non-synonymous changes in *pgiC* in ferns, angiosperms, and a butterfly. Sites in bold were retrieved in this study. Infraspecific changes shaded. The catalytically critical region identified in Fig. 3 of Wheat et al. (2006) (characters 1135–1162) is enclosed in a box

Exon Alignment DNA Character number	Plant Exon number	Position and Taxon or Orientation to Exons	Amino acid substitution	Charge change/ Alanine ±
1	1	PLANT EXON 1 BEGINS		
14	1	Arabidopsis	Asn > Thr	0
55	2	PLANT EXON 2 BEGINS		
55	2	Clarkia	Ala > Ser	A
67	2	Colias 21	Asn > Asp	–1
83	2	Clarkia	Met > Arg	+1
86	2	Clarkia	Thr > Ile	0
139	2	Clarkia	Val > Leu	0
156	3	PLANT EXON 3 BEGINS		
154	2–3	Colias 50	Pro > Leu	0
157	3	Colias 51	Asn > Lys	0
199	3	Clarkia	Ile > Ser	0
202	3	Clarkia	Pro > Ser	0
205	3	Colias 67	Thr > Ala	A
214	3	Clarkia	Ser > Asn	0
227	3	Clarkia	Lys > Arg	0
238	4	PLANT EXON 4 BEGINS		
239	**4**	**Polystichum**	**Val > Ala**	**A**
239	4	Clarkia	Ala > Glu	A
241	4	Arabidopsis	Ser > Ala	A
262	**4**	**Polystichum**	**Glu > Lys**	**+2**
266	4	Clarkia	Arg > His	0
286	5	PLANT EXON 5 BEGINS		
293	5	Clarkia	Thr > Ser	0
307	5	Arabidopsis	Ser > Ala	A
319	5	Arabidopsis	Val > Iso	0
334	5	Clarkia	Pro > Ser	0
334	5	Colias 110	Arg > His	0
340	5	Clarkia	Asn > Asp	–1
343	5	Clarkia	Ser > Leu	0
346	5	Dioscorea 1	Ala > Val	A
352	5	Clarkia	Cys > Ser	0
352	5	Dioscorea 2	Lys > Glu	–2
361	5	Dioscorea 3	Gly > Asp	–1
365	5	Arabidopsis	Met > Lys	+1
388	5	Colias 128	Gly > Ala	0
389	5	Clarkia	Asn > Ser	0
391	5	Dioscorea 4	Val > Ala	A
398	5	Colias 110	Ala > Glu	A –1
412	5	Dioscorea 5	Phe > Val	0
423	5	Clarkia	Ser > Arg	+1
436	**5**	**Polystichum**	**Ala > Thr**	**A**
436	5	Colias 144	Ala > Pro	A
442	6	PLANT EXON 6 BEGINS		
442	6	Clarkia	Ile > Val	0
460	6	Colias 152	Pro > Ser	0
460	6	Clarkia	Glu > Lys	+2
475	**6**	**Polystichum**	**Val > Ile**	**0**
539	7	PLANT EXON 7 BEGINS		
582	8	PLANT EXON 8 BEGINS		
622	8	Dioscorea 6	Thr > Ala	A
632	8	Clarkia	Asn > Ser	0
641	8	Clarkia	Thr > Ser	0
650	9	PLANT EXON 9 BEGINS		

724	9	Dioscorea 7	Ala > Gly	A
727	9	Colias 235	Ala > Ser	A
741	10	PLANT EXON 10 BEGINS		
745	10	Clarkia	Ser > Thr	0
745	10	Dioscorea 8	Gln > Arg	+1
763	10	Clarkia	Leu > Ile	0
790	11	PLANT EXON 11 BEGINS		
818	11	Clarkia	Asn > Ile	0
823	11	Clarkia	Ala > Thr	A
863	12	PLANT EXON 12 BEGINS		
910	12	Clarkia 1	Ala > Ser	A
924	13	PLANT EXON 13 BEGINS		
940	13	Dioscorea 9	Ala > Arg	A +1
943	13	Clarkia 2	Asn > Ser	0
970	13	Clarkia	Pro > Ser	0
980	13	Colias 317	Lys > Met	–1
991	14	PLANT EXON 14 BEGINS		
1033	14	Dioscorea 10	Arg > Gly	–1
1048	15	PLANT EXON 15 BEGINS		
1081	**15**	**Polystichum**	**Leu > Phe**	**0**
1102	16	PLANT EXON 16 BEGINS		
1132	16	Dioscorea 11	Ser > Thr	0
1135	16	Colias 369	Arg > Cys	–1
1135	16	Clarkia	Ile > Val	0
1138	16	Colias 370	Ser > Gly	0
1144	16	Dioscorea 12	Val > Ile	0
1145	**16**	**Polystichum**	**Val > Ala**	**A**
1153	16	Colias 375	Glu > Gln	+1
1162	**16**	**Polystichum**	**Glu > Gln**	**+1**
1189	16	Dioscorea 13	Pro > Ser	0
1234	17	PLANT EXON 17 BEGINS		
1267	17	Dioscorea 14	Val > Ile	0
1267	17	Clarkia	Val > Ile	0
1301	18	PLANT EXON 18 BEGINS		
1375	19	PLANT EXON 19 BEGINS		
1379	**19**	**Polystichum**	**Pro > Gln**	**0**
1381	19	Dioscorea 15	Glu > Asp	0
1381	19	Clarkia 3	Glu > Ala	A+1
1384	19	Clarkia 4	Glu > Gln	+1
1393	19	Colias 450	Ala > Ser, Ala > Val	A
1395	19	Clarkia	Lys > Asn	–1
1417	19	Colias 458	Ala > Val	A
1429	19	Clarkia	Ser > Ala	A
1432	19	Colias 463	Ala > Thr	A
1432–3	19	Dioscorea 16	Asp > Asn	+1
1453	20	PLANT EXON 20 BEGINS		
1474	20	Clarkia 5	Cys > Ser	0
1477	20	Clarkia	Ile> Leu	0
1505	20	Clarkia	Ala > Gly	A
1522	21	PLANT EXON 21 BEGINS		
1522	21	Clarkia	Leu > Val	0
1570	21	Dioscorea 17	Gly > Val	0
1620	22	PLANT EXON 22 BEGINS		
1646	22	Clarkia	Gly > Ala	A
1657	22	Colias 538	Ile > Val	0
1712	22	Clarkia	Arg > Lys	0
1720	22	Clarkia	Gln > Glu	–1
1731	23	PLANT EXON 23 BEGINS		

Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.04.010.

References

Bainard, J.D., Henry, T.A., Bainard, L.D., Newmaster, S.G., 2011. DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. Chromosome Res. 19, 763–775.

Barrington, D.S., 1990. Hybridization and allopolyploidy in Central American *Polystichum*: cytological and isozyme documentation. Ann. Missouri Bot. Gard. 77, 297–305.

Barrington, D.S., 2003. *Polystichum lilianae* sp. nov. (Dryopteridaceae) and its relationships to *P. fournieri* and *P. turrialbae*. Brittonia 55, 317–325.

Bennett, M.D., Leitch, I.J., 2011. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann. Bot. 107, 467–590.

Chang, H.M., Chiou, W.L., Wang, J.C., 2009. Molecular evidence for genetic heterogeneity and the hybrid origin of *Acrorumohra subreflexipinna* from Taiwan. Am. Fern J. 9, 61–77.

Claes, V., Kettmann, R., Burny, A., 1994. Structure of the gene encoding pig phosphoglucose isomerase. Gene 150, 235–241.

Clarke, M., Lohan, A.J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., Bertelli, C., Schilde, C., Kianianmomeni, A., Bürglin, T.R., Frech, C., 2013. Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol. 14, R11. http://dx.doi.org/10.1186/gb-2013-14-2-r11.

Cunningham, B.C., Wells, J.A., 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244, 1081–1085.

Dahlhoff, E.P., Rank, N.E., 2000. Functional and physiological consequences of genetic variation at phosphoglucose isomerase: heat shock protein expression is related to enzyme genotype in a montane beetle. Proc. Natl. Acad. Sci. USA 97, 10056–10061.

Do, C.B., Mahabhashyam, M.S., Brudno, M., Batzoglou, S., 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. 15, 330–340.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., 2007. Geneious version 5.4 Website <http://www.geneious.com/> (accessed 13.03.03).

Gastony, G.J., 1986. Electrophoretic evidence for the origin of fern species by unreduced spores. Am. J. Bot. 73, 1563–1569.

Gastony, G.J., 1991. Gene silencing in a polyploid homosporous fern: paleopolyploidy revisited. Proc. Natl. Acad. Sci. USA 88, 1602–1605.

Gillespie, J.F., 1991. The Causes of Molecular Evolution. Oxford University Press, New York.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT a free program for phylogenetic analysis. Cladistics 24, 774–786.

Gottlieb, L.D., Ford, V.S., 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of PgiC. Syst. Bot. 21, 45–62.

Hori, K., Tono, A., Fujimoto, K., Kato, J., Ebihara, A., Watano, Y., Murakami, N., 2014. Reticulate evolution in the apogamous *Dryopteris varia* complex (Dryopteridaceae, subg. Erythrovariae, sect. Variae) and its related sexual species in Japan. J. Plant Res. 127, 661–684.

Jorgensen, S.A., Barrington, D.S., 2017. Two Beringian origins for the allotetraploid fern *Polystichum braunii* (Dryopteridaceae). Syst. Bot. 42, 6–16.

Ishikawa, H., Watano, Y., Kano, K., Ito, M., Kurita, S., 2002. Development of primer sets for PCR amplification of the *pgiC* gene in ferns. J. Plant Res. 115, 65–70.

Kawabe, A., Yamane, K., Miyashita, N.T., 2000. DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. Genetics 156, 1339–1347.

Koenemann, D.M., Maisonpierre, J.A., Barrington, D.S., 2011. Broad-scale integrity and local divergence in the fiddlehead fern *Matteuccia struthiopteris* (L.) Todaro (Onocleaceae). Am. Fern J. 101, 213–230.

Magallón, S., Hilu, K.W., Quandt, D., 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am. J. Bot. 100, 556–573.

Marden, J.H., 2013. Nature's inordinate fondness for metabolic enzymes: why metabolic enzyme loci are so frequently targets of selection. Mol. Ecol. 22, 5732–5764.

McHenry, M.A., 2012. Systematics and biogeography of Andean ferns in the genus *Polystichum* Roth (Dryopteridaceae) Unpublished Ph.D. Dissertation. Univ Vermont.

McHenry, M.A., Barrington, D.S., 2014. Phylogeny and biogeography of exindusiate Andean *Polystichum* (Dryopteridaceae). Am. J. Bot. 101, 365–375.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, pp. 1–8.

Nixon, K., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.

Rambaut, A., Drummond, A.J., 2007. Tracer. Version 1.5. <http://beast.bio.ed.ac.uk/Tracer>.

Rank, N.E., Dahlhoff, E.P., 2002. Allele frequency shifts in response to climate change and physiological consequences of allozyme variation in a montane insect. Evolution 56, 2278–2289.

Riddoch, B., 1993. The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). Biol. J. Linn. Soc. 50, 1–17.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rothfels, C.J., Larsson, A., Li, F.W., Sigel, E.M., Huiet, L., Burge, D.O., Pryer, K.M., 2013. Transcriptome-mining for single-copy nuclear markers in ferns. PLoS ONE 8, e76957. http://dx.doi.org/10.1371/journal.pone.0076957.

Schneider, H., Navarro-Gomez, A., Russell, S.J., Ansell, S., Grundmann, M., Vogel, J., 2013. Exploring the utility of three nuclear regions to reconstruct reticulate evolution in the fern genus *Asplenium*. J. Syst. Evol. 51, 142–153.

Terauchi, R., Terachi, T., Miyashita, N.T., 1997. DNA polymorphism at the pgi locus of a wild yam, *Dioscorea tokoro*. Genetics 147, 1899–1914.

Thomas, B.R., Laudencia-Chingcuanco, D., Gottlieb, L.D., 1992. Molecular analysis of the plant gene encoding cytosolic phosphoglucose isomerase. Plant Mol. Biol. 19, 745–757.

Thomas, B.R., Ford, V.S., Pichersky, E., Gottlieb, L.D., 1993. Molecular characterization of duplicate cystolic phosphoglucose isomerase genes in *Clarkia* and comparison to the single gene in *Arabidopsis*. Genetics 135, 895–905.

Troncoso-Ponce, M.A., Rivoal, J., Cejudo, F.J., Dorion, S., Garcés, R., Martínez-Force, E., 2010. Cloning, biochemical characterisation, tissue localization and possible post-translational regulatory mechanism of the cytosolic phosphoglucose isomerase from developing sunflower seeds. Planta 232, 845–849.

Wang, F.H., Lu, J.M., Wen, J., Ebihara, A., Li, D.Z., 2016. Applying DNA barcodes to identify closely related species of ferns: a case study of the Chinese *Adiantum* (Pteridaceae). PLoS ONE 11, e0160611.

Watt, W.B., 1977. Adaptation at specific loci I. natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. Genetics 87, 177–194.

Werth, C.R., Guttman, S.I., Eshbaugh, W.H., 1985. Electrophoretic evidence of reticulate evolution in the Appalachian *Asplenium* complex. Syst. Bot. 10, 184–192.

Wheat, C.W., Watt, W.B., Pollock, D.D., Schulte, P.M., 2006. From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). Mol. Biol. Evol. 23, 499–512.

Wheat, C.W., Haag, C.H., Marden, J.H., Hanski, I., Frilander, M.J., 2010. Nucleotide polymorphism at a gene (Pgi) under balancing selection in a butterfly metapopulation. Mol. Biol. Evol. 27, 267–287.

Xu, B., Yang, Z., 2013. PAMLX: a graphical user interface for PAML. Mol. Biol. Evol. 30, 2723–2724.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908–917.

Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.

Zhang, L., Schuettpelz, E., Rothfels, C.J., Zhou, X.M., Gao, X.F., Zhang, L.B., 2016. Circumscription and phylogeny of the fern family Tectariaceae based on plastid and nuclear markers, with the description of two new genera: *Draconopteris* and *Malaifilix* (Tectariaceae). Taxon 65, 723–738.