### Old Dominion University [ODU Digital Commons](https://digitalcommons.odu.edu/)

[Biological Sciences Faculty Publications](https://digitalcommons.odu.edu/biology_fac_pubs) **Biological Sciences** Biological Sciences

2018

# Low- Copy Nuclear Markers in Isoëtes (Isoëtaceae) Identified With **Transcriptomes**

Peter W. Schafran Old Dominion University

Gabriel Johnson

W. Carl Taylor

Elizabeth A. Zimmer

Lytton J. Musselman Old Dominion University, lmusselm@odu.edu

Follow this and additional works at: [https://digitalcommons.odu.edu/biology\\_fac\\_pubs](https://digitalcommons.odu.edu/biology_fac_pubs?utm_source=digitalcommons.odu.edu%2Fbiology_fac_pubs%2F271&utm_medium=PDF&utm_campaign=PDFCoverPages)

 $\bullet$  Part of the [Botany Commons,](https://network.bepress.com/hgg/discipline/104?utm_source=digitalcommons.odu.edu%2Fbiology_fac_pubs%2F271&utm_medium=PDF&utm_campaign=PDFCoverPages) [Plant Biology Commons](https://network.bepress.com/hgg/discipline/106?utm_source=digitalcommons.odu.edu%2Fbiology_fac_pubs%2F271&utm_medium=PDF&utm_campaign=PDFCoverPages), and the [Plant Breeding and Genetics](https://network.bepress.com/hgg/discipline/108?utm_source=digitalcommons.odu.edu%2Fbiology_fac_pubs%2F271&utm_medium=PDF&utm_campaign=PDFCoverPages) [Commons](https://network.bepress.com/hgg/discipline/108?utm_source=digitalcommons.odu.edu%2Fbiology_fac_pubs%2F271&utm_medium=PDF&utm_campaign=PDFCoverPages)

### Original Publication Citation

Schafran, P. W., Johnson, G., Taylor, W. C., Zimmer, E. A., & Musselman, L. J. (2018). Low‐copy nuclear markers in *Isoëtes* (Isoëtaceae) identified with transcriptomes. Applications in Plant Sciences, 6(4), e1142. doi:10.1002/aps3.1142

This Article is brought to you for free and open access by the Biological Sciences at ODU Digital Commons. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu.](mailto:digitalcommons@odu.edu)





# **Low-copy nuclear markers in** *Isoëtes* **(Isoëtaceae) identified with transcriptomes**

Peter W. Schafran<sup>1,2,3</sup> (Cabriel Johnson<sup>2</sup>, W. Carl Taylor<sup>2</sup>, Elizabeth A. Zimmer<sup>2</sup>, and Lytton J. Musselman<sup>1</sup>

Manuscript received 12 December 2017; revision accepted 16 January 2018.

<sup>1</sup> Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA

2 Department of Botany, National Museum of Natural

History, Smithsonian Institution, Washington, D.C. 20560, USA

3 Author for correspondence: [pscha005@odu.edu,](mailto:pscha005@odu.edu) [SchafranP@si.edu](mailto:SchafranP@si.edu)

**Citation**: Schafran, P. W., G. Johnson, W. C. Taylor, E. A. Zimmer, and L. J. Musselman. 2018. Low-copy nuclear markers in *Isoëtes* (Isoëtaceae) identified with transcriptomes. *Applications in Plant Sciences* 6(4): e1142.

**doi:10.1002/aps3.1142**

**PREMISE OF THE STUDY**: Few genetic markers provide phylogenetic information in closely related species of *Isoëtes* (Isoëtaceae). We describe the development of primers for several putative low-copy nuclear markers to resolve the phylogeny of *Isoëtes*, particularly in the southeastern United States.

**METHODS AND RESULTS**: We identified regions of interest in *Isoëtes* transcriptomes based on low-copy genes in other plants. Primers were designed for these regions and tested with 16 taxa of *Isoëtes* and one species of *Lycopodium*. Parts of the *pgiC*, *gapC*, and *IBR*3 gene regions show phylogenetic signal within the North American and Mediterranean clades of *Isoëtes*.

**CONCLUSIONS**: Transcriptome data prove useful for identification and primer design of lowcopy genes. Three new markers show potential for inferring phylogenies in regional clades of *Isoëtes*, and possibly across the entire genus.

 **KEY WORDS** *gapC*; *IBR*3; *Isoëtes*; *pgiC*; primer design; Sanger sequencing.

*Isoëtes* L. (Isoëtaceae, Lycopodiophyta) is a cosmopolitan genus of ca. 250 recognized species. These heterosporous lycophytes consist of a 2–3-lobed rootstock that bears linear, quill-like, microphyllous leaves or sporophylls. All microphylls have the potential to develop into sporophylls (Foster and Gifford, 1974). Mega- and microsporangia are produced at the base of sporophylls, in some species covered by a layer of tissue called a velum. Traditionally, spore ornamentation and velum coverage have been considered taxonomically important. Although species inhabit a variety of ecological niches, from obligate aquatic to ephemeral terrestrial habitats, their morphology is extremely conserved. Phylogenetic studies in closely related clades of *Isoëtes* have been limited by a dearth of morphological features and molecular markers. Hoot and Taylor (2001) identified the nuclear ribosomal gene internal transcribed spacer (ITS), a *LEAFY* homolog nuclear gene intron (*LFY*), and the plastid *atpB-rbcL* spacer region as informative markers in *Isoëtes*. However, although these markers and the plastid *rbcL* gene show utility in large-scale, global phylogenies, they generally lose resolution at the regional level (Rydin and Wikström, 2002; Hoot et al., 2006; Larsén and Rydin, 2016). *LFY* is more variable than the other three markers and is fairly informative in recently diverged species groups (Taylor et al., 2004; Hoot et al., 2004). With only a single informative nuclear marker within groups such as the eastern North American clade, it is difficult to fully test phylogenetic hypotheses of reticulate evolution and incomplete lineage sorting.

Transcriptomes provide a valuable tool for marker selection and PCR primer design in the absence of a sequenced genome, as is the case in *Isoëtes*. Databases such as the 1000 Plants project [\(http://](http://www.onekp.com) [www.onekp.com;](http://www.onekp.com) Matasci et al., 2014) contain transcriptomes across all major lineages of land plants, allowing identification of unique marker regions for a group of interest. Here we describe use of transcriptome data to develop PCR primers for phylogenetically informative low-copy nuclear markers in *Isoëtes.*

#### **METHODS AND RESULTS**

Markers of interest were selected based on a literature search of reportedly low-copy nuclear markers in ferns and mosses (Table 1; Szövényi et al., 2006; Schuettpelz et al., 2008; Rothfels et al., 2013). Nucleotide sequences for these markers were obtained from the National Center for Biotechnology Information's (NCBI) GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>; Clark et al., 2016) or TreeBASE [\(http://www.treebase.org](http://www.treebase.org); Sanderson et al., 1994) databases. Transcriptomes for three *Isoëtes* taxa were provided by other sources (*I. echinospora* Durieu from S. Hetherington, University of Oxford, Oxford, United Kingdom; and *I. tegetiformans* Rury and an unnamed *Isoëtes* species from the 1000 Plants project [[http://](http://www.oneKP.com) [www.oneKP.com](http://www.oneKP.com)]). Using the BLAST+ 2.4 software package (Camacho et al., 2009), local BLAST databases were constructed from each *Isoëtes* transcriptome. The sequences of selected fern

**TABLE 1.** Primers designed for low-copy markers identified in *Isoëtes* transcriptomes.

<b>Marker ID</b>	<b>Primer names</b>	Primer sequences (5'-3')	$T_{\alpha}$ (°C)
pgiC	pgiC_1156F	F: GGTCTCCTAAGTGTCTGGAATGT	55
	pgiC_1900R	R: GTTCTCCAAAATCAATTTCTCC	
<b>IBR3 1</b>	IBR3 2F	E: CTCAAATCAGCTCATGCAATTG	60
	IBR3 6R	R: AGCTCCCAATCCAACACAGC	
<b>IBR3 2</b>	<b>IBR3 13F</b>	F: CAATGACTGAACCGCAAGTTG	60
	<b>IBR3 16R</b>	R: GACCCAACGAGTCTCATGCAG	
Transducin 1	Transducin_1F	F: GATGTGGTTGGTGAGTCTGG	55
	Transducin 1R	R: CACTTCATTGAACCTCAG	
Transducin <sub>2</sub>	Transducin 2F	F: GGAACAAAAGCAGGGACATTAG	55
	Transducin 2R	R: CATCAGAAGAGATGTCCATAC	
qapC_short	gapC_5F	F: GAATCTACTGGTGTCTTCAC	55
	gapC_7R	R: TTCTGGTTTATATTCATGCTCG	
gapC_long	qapC_5F	F: GAATCTACTGGTGTCTTCAC	55
	gapC 9R	R: ATGGTCCATCAACAGTYTTCTG	

*Note:*  $F =$  forward;  $R =$  reverse;  $T_a =$  annealing temperature.

(Rothfels et al., 2013) and moss (Szövényi et al., 2006) low-copy nuclear markers were BLASTed against the transcriptome databases to identify those markers present as single-copy in *Isoëtes*. These single-copy marker regions were extracted from their respective transcriptome and aligned with marker sequences from the literature using Geneious version 7 (Kearse et al., 2012). Primer sequences from the literature were modified to match the *Isoëtes* transcriptome sequences.

Plants were collected from the field, and leaf tissue was desiccated with silica gel. Voucher specimens have been stored at the Old Dominion University herbarium (ODU) and/or the U.S. National Herbarium (US). DNA was extracted from approximately 200 mg of dried tissue with the QIAGEN DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, California, USA) or AutogenPrep 965 (Autogen Inc., Holliston, Mississippi, USA) using standard protocols. Sixteen diploid taxa of *Isoëtes* and one species of *Lycopodium* L. (one individual per taxon) were selected from available DNAs to represent various levels of divergence (Appendix 1).

Markers were amplified by PCR on an ABI 2720 thermocycler (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA), with a reaction mixture of 12.5  $\mu$ L of 2× GoTaq PCR master mix (Promega Corporation, Madison, Wisconsin, USA), 0.5 μL of 0.1 mg/mL bovine serum albumin, 1.0 μL each of 10 μM forward and reverse primer, 7.5 μL of sterile distilled water, and 2.5 μL of DNA template (10–60 ng). PCR reactions were carried out with an initial melting period at 94°C (5 min), followed by 35 cycles of 94°C (30 s), annealing at 55–60°C (30 s), and extension at 72°C (1 min), with a final extension at 72°C (7 min). Amplification success was confirmed by electrophoresis using a 1.5% sodium boric acid– based agarose gel.

**TABLE 2.** Alignment statistics for all sequences with quality scores >85%.



*Note:* PIS = parsimony informative sites.

Markers were selected for Sanger sequencing based on their producing a single band across all samples and for a maximum size of ~1000 bp. PCR products were treated with ExoSAP-IT PCR cleanup enzyme mix (Affymetrix Inc., Santa Clara, California, USA) before cycle sequencing with BigDye Terminator v3.1 (Thermo Fisher Scientific Inc.). The labeled sequencing fragments were read on an ABI 3130xl Genetic Analyzer (Thermo Fisher Scientific Inc.), and the resulting chromatograms were edited and analyzed using Geneious (Kearse et al., 2012).

Initial screening of primers showed that all amplify in at least some of the eastern North American taxa. Gel electrophoresis revealed that IBR3\_1 and Transducin\_2 are too long (~2000 bp) and Transducin\_1 has both short and long copies in some individuals (~500 bp and ~1000 bp), making these poor candidates for a Sanger sequencing approach without needing molecular cloning or gel extraction. Although gapC\_short readily amplified, it is contained within gapC\_long, making sequencing of the shorter fragment redundant. pgiC, IBR3\_2 (hereafter IBR3), and gapC\_long (hereafter gapC) were selected for PCR and sequencing of the full taxa list (Appendices 2, 3).

#### **pgiC**

This primer pair is rooted in exons 14 and 16, and amplifies across introns 14, 15, and exon 15 of this locus (Rothfels et al., 2013). The region amplified easily across all taxa of *Isoëtes* and *Lycopodium clavatum* L., and generated consistently high-quality sequence data. All sequences aligned well, with a total alignment length of 466 bp and pairwise identity of 83%. Excluding *L. clavatum*, alignment length decreases to 357 bp and pairwise identity increases to 89%. Sequence length between these species of *Isoëtes* ranges from 310 to 347 bp, with a mean of 324 bp (Table 2). This is approximately half the length of the same region in ferns tested by Rothfels et al. (2013).

#### **gapC**

The *gapC* gene encodes cytosolic glyceraldehyde-3-phosphate and is part of the GAPDH gene family (Strand et al., 1997; Wall, 2002; Szövényi et al., 2006). Primers designed by Szövényi et al. (2006) are rooted in exons 5 and 9 and amplify all exons and introns in between. However, given concern that the resulting marker in *Isoëtes* may be too long for Sanger sequencing, the primers designed for this study were rooted in exons 5 and 8, amplifying introns 5, 6, 7, and exons 6 and 7.

This marker showed the least ability to routinely generate high-quality sequence data. Although not detected in any of the transcriptomes available, it is possible this results from off-target amplification of other members of the GADPH gene family (i.e., *gapCp* or an unnamed *gapC*/*gapCp* relative) (Schuettpelz et al.,

2008; Rothfels et al., 2013). The *Isoëtes*-only alignment is 561 bp and has a pairwise identity of 85% (Table 2).

#### **IBR3**

Unlike *pgiC* and *gapC*, this gene does not have an extensive history of use as a phylogenetic marker. The *IBR3* gene is thought to encode an indole-3-butyric acid–specific peroxisomal enzyme related to acyl-CoA dehydrogenases (Zolman et al., 2007). Rothfels et al. (2013) showed it to be single-copy throughout selected fern lineages, and this also appears to be the case in *Isoëtes*. Primers for the IBR3 marker amplify most species of *Isoëtes* easily, with the exception of two members of the Mediterranean clade (*I. histrix* Bory & Durieu and *I. nuttallii* A. Braun ex Engelm.). Alignment of *Isoëtes* sequences is 700 bp long with 87% pairwise identity (Table 2).

#### **CONCLUSIONS**

Transcriptome mining is shown to be a useful tool for identification of putative low-copy markers for primer design. Despite having access to transcriptomes of just three species of *Isoëtes* in the North American clade, primers could be designed for regions that show phylogenetic signal across widely divergent clades in the genus, and potentially across all Lycopodiophyta. Although techniques such as target enrichment allow for generation of data sets orders of magnitude larger (Mandel et al., 2014), design of primers for Sanger sequencing is still more time- and cost-efficient in taxonomic groups for which just a few markers may be needed to infer well-resolved phylogenies.

#### **ACKNOWLEDGMENTS**

The authors wish to thank Sandy Hetherington of the University of Oxford (Oxford, United Kingdom) and the 1000 Plants project [\(http://www.oneKP.com](http://www.oneKP.com)) for providing *Isoëtes* transcriptomes used for primer design. Funding was provided by the Mary Payne Hogan endowment at Old Dominion University and the Smithsonian Institution Fellowship Program. All or portions of the laboratory and/or computer work were conducted in and with the support of the Laboratories of Analytical Biology of the National Museum of Natural History or its partner labs.

#### **LITERATURE CITED**

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 421.

- Clark, K., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2016. GenBank. *Nucleic Acids Research* 44(D1): D67–D72.
- Foster, A. S., and E. M. Gifford Jr. 1974. Comparative morphology of vascular plants. W. H. Freeman and Company, San Francisco, California, USA.
- Hoot, S. B., and W. C. Taylor. 2001. The utility of nuclear ITS, a *LEAFY* homolog intron, and chloroplast *atpB*-*rbcL* spacer region data in phylogenetic analyses and species delimitation in *Isoëtes*. *American Fern Journal* 91(3): 166–177.
- Hoot, S. B., N. S. Napier, and W. C. Taylor. 2004. Revealing unknown or extinct lineages within *Isoëtes* (Isoëtaceae) using DNA sequences from hybrids. *American Journal of Botany* 91(6): 899–904.
- Hoot, S. B., W. C. Taylor, and N. S. Napier. 2006. Phylogeny and biogeography of *Isoëtes* (Isoëtaceae) based on nuclear and chloroplast DNA sequence data. *Systematic Botany* 31(3): 449–460.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647–1649.
- Larsén, E., and C. Rydin. 2016. Disentangling the phylogeny of *Isoetes* (Isoetales), using nuclear and plastid data. *International Journal of Plant Sciences* 177(2): 157–174.
- Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2(2): 1300085.
- Matasci, N., L. H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17. [https://doi.org/10.1186/2047-217X-3-17.](https://doi.org/10.1186/2047-217X-3-17)
- Rothfels, C. J., A. Larsson, F. W. Li, E. M. Sigel, L. Huiet, D. O. Burge, M. Ruhsam, et al. 2013. Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS One* 8(10): e76957.
- Rydin, C., and N. Wikström. 2002. Phylogeny of *Isoëtes* (Lycopsida): Resolving basal relationships using *rbcL* sequences. *Taxon* 51(1): 83–89.
- Sanderson, M. J., M. J. Donoghue, W. H. Piel, and T. Eriksson. 1994. TreeBASE: A prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. *American Journal of Botany* 81(6): 183.
- Schuettpelz, E., A. L. Grusz, M. D. Windham, and K. M. Pryer. 2008. The utility of nuclear *gapCp* in resolving polyploid fern origins. *Systematic Botany* 33(4): 621–629.
- Strand, A. E., J. Leebans-Mack, and B. G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113–118.
- Szövényi, P., Z. Hock, E. Urmi, and J. Schneller. 2006. New primers for amplifying the GapC gene in bryophytes and its utility in infraspecific phylogenies in the genus *Sphagnum*. *Lindbergia* 31: 78–84.
- Taylor, W. C., A. R. Lekschas, Q. F. Wang, X. Liu, N. S. Napier, and S. B. Hoot. 2004. Phylogenetic relationships of *Isoëtes* (Isoëtaceae) in China as revealed by nucleotide sequences of the nuclear ribosomal ITS region and the second intron of a *LEAFY* homolog. *American Fern Journal* 94(4): 196–205.
- Wall, D. P. 2002. Use of the nuclear gene glyceraldehyde 3-phosphate dehydrogenase for phylogeny reconstruction of recently diverged lineages in *Mitthyridium* (Musci: Calymperaceae). *Molecular Phylogenetics and Evolution* 25(1): 10–26.
- Zolman, B. K., M. Nyberg, and B. Bartel. 2007. IBR3, a novel peroxisomal acyl-CoA dehydrogenase-like protein required for indole-3-butyric acid response. *Plant Molecular Biology* 64: 59–72.

#### **APPENDIX 1**. Collection locations, vouchers, and GenBank accessions for taxa included in this study.



a One individual was sampled per taxon.

b Per Larsén and Rydin (2016).

c Herbaria are abbreviated according to Index Herbariorum [\(http://sweetgum.nybg.org/science/ih/\)](http://sweetgum.nybg.org/science/ih/).

d Tissue samples provided by A. Troia (Università degli Studi di Palermo, Palermo, Italy); not deposited in a recognized herbarium.

e Voucher deposited in P. Schafran's personal collection.



#### **APPENDIX 2**. Amplification and sequence quality of markers across taxa.

*Note:* + = successful amplification or sequence quality >85%; — = no amplification or sequence quality <85%; NA = sequencing not attempted.

