Molecular Phylogenetics and Evolution 85 (2015) 76–87

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment

Jessica D. Stephens [a,*], Willie L. Rogers [a], Karolina Heyduk [a], Jennifer M. Cruse-Sanders [b], Ron O. Determann [b], Travis C. Glenn [c], Russell L. Malmberg [a]

[a] Department of Plant Biology, University of Georgia, Athens, GA 30602, United States
[b] Atlanta Botanical Garden, 1345 Piedmont Ave NE, Atlanta, GA 30309, United States
[c] Department of Environmental Health Science, University of Georgia, Athens, GA 30602, United States

ARTICLE INFO

Article history:
Received 5 October 2014
Revised 23 January 2015
Accepted 27 January 2015
Available online 14 February 2015

Keywords:
Gene capture
Coalescent
Phylogenomics
Pitcher plants
Species tree estimation

ABSTRACT

The North American carnivorous pitcher plant genus *Sarracenia* (Sarraceniaceae) is a relatively young clade (<3 million years ago) displaying a wide range of morphological diversity in complex trapping structures. This recently radiated group is a promising system to examine the structural evolution and diversification of carnivorous plants; however, little is known regarding evolutionary relationships within the genus. Previous attempts at resolving the phylogeny have been unsuccessful, most likely due to few parsimony-informative sites compounded by incomplete lineage sorting. Here, we applied a target enrichment approach using multiple accessions to assess the relationships of *Sarracenia* species. This resulted in 199 nuclear genes from 75 accessions covering the putative 8–11 species and 8 subspecies/varieties. In addition, we recovered 42 kb of plastome sequence from each accession to estimate a cpDNA-derived phylogeny. Unsurprisingly, the cpDNA had few parsimony-informative sites (0.5%) and provided little information on species relationships. In contrast, use of the targeted nuclear loci in concatenation and coalescent frameworks elucidated many relationships within *Sarracenia* even with high heterogeneity among gene trees. Results were largely consistent for both concatenation and coalescent approaches. The only major disagreement was with the placement of the *purpurea* complex. Moreover, results suggest an Appalachian massif biogeographic origin of the genus. Overall, this study highlights the utility of target enrichment using multiple accessions to resolve relationships in recently radiated taxa.

1. Introduction

The evolution of carnivory in angiosperms has long fascinated evolutionary biologists, with the most notable being Charles Darwin (1875). This interest partially stems from the complex trapping structures used in attraction, retention, and digestion of prey and subsequent absorption of nutrients (Albert et al., 1992; Juniper et al., 1989). These carnivorous adaptations to nutrient poor habitats have independently evolved six times in flowering plants, resulting in approximately 645 species, which often display tremendous morphological diversity at both the infrafamilial and infrageneric level (Albert et al., 1992; Ellison and Gotelli, 2009). Insight into the patterns of structural evolution and diversification across these groups requires an explicit understanding of their evolutionary relationships. Phylogenies currently exist for many carnivorous genera including *Utricularia* (bladderworts, Jobson et al., 2003; Müller and Borsch, 2005), *Drosera* (sundews, Rivadavia et al., 2003), and *Nepenthes* (Old World pitcher plants, Meimberg et al., 2001), yet the evolutionary relationships of one of the more well-studied genera, *Sarracenia* (New World pitcher plants), remain largely ambiguous.

Sarracenia is the most recently diverged group of the three extant genera within the family Sarraceniaceae (Ellison et al., 2012; Neyland and Merchant, 2006). All species within Sarraceniaceae are carnivorous with no geographical overlap among genera. The basal monotypic lineage, *Darlingtonia californica*, is restricted to serpentine seeps in Oregon and California, while the estimated 15 *Heliamphora* species are confined to the Guiana Highlands tepuis in South America (McPherson, 2007). *Sarracenia* is endemic to seepage slopes, wet pine savannas, and fens of North America, predominately the southeastern United States Coastal Plain with one subspecies, *purpurea* ssp. *purpurea*, extending into the

* Corresponding author.
E-mail addresses: jstephens@plantbio.uga.edu (J.D. Stephens), bonjour@uga.edu (W.L. Rogers), kheyduk@plantbio.uga.edu (K. Heyduk), JSanders@atlantabotanicalgarden.org (J.M. Cruse-Sanders), rdetermann@atlantabotanicalgarden.org (R.O. Determann), travisg@uga.edu (T.C. Glenn), russell@plantbio.uga.edu (R.L. Malmberg).

http://dx.doi.org/10.1016/j.ympev.2015.01.015
1055-7903/

northeastern United States and southern Canada. Unfortunately these habitats are being destroyed and estimates suggest less than 3% of historic *Sarracenia* habitat remains (Folkerts, 1982; Folkerts and Folkerts, 1993). This continued habitat loss has resulted in the U.S. Fish and Wildlife and Convention on International Trade in Endangered Species (CITES) listing of three endangered taxa within *Sarracenia* and one taxa considered a candidate for listing (www.cites.org). Complicating protection status of other members of this genus is the disagreement among sources in the number of recognized species, subspecies, and varieties with numbers ranging between 8–11 species and as many as 41 subspecies, varieties, and forms (Ellison et al., 2014).

Previous attempts at constructing a phylogeny for *Sarracenia* from nuclear (Ellison et al., 2012; Neyland and Merchant, 2006), chloroplast (Bayer et al., 1996; Ellison et al., 2012), and mitochondrial regions (Ellison et al., 2012) have been inconsistent, typically with numerous polytomies within the genus. In addition, the relatively short branch lengths dated at roughly 0.5–3 million years ago (mya) (Ellison et al., 2012) indicate that this group may have undergone a recent, rapid diversification. Further complicating phylogenetic resolution is frequent hybridization among sympatric species (Furches et al., 2013; Mellichamp and Case, 2009). Both short branches and hybridization can have dramatic effects on species tree estimation. In particular, a recent radiation increases the chance that genes retain ancestral polymorphisms, resulting in incomplete lineage sorting (Pamilo and Nei, 1988); additionally, hybridization can lead to reticulation within gene trees (Hennig, 1966). Using multilocus data and modeling differences in gene history with use of the multispecies coalescent model can mitigate these potential sources of gene tree discordance within the species tree (Degnan and Rosenberg, 2009; Knowles, 2009; Liu et al., 2009). Increasing loci is expected to produce more accurate model parameters and therefore increase nodal support values in phylogenetic analyses (Maddison, 1997; Song et al., 2012), and use of multispecies coalescence has repeatedly outperformed concatenation methods under simulated and empirical data (Kubatko and Degnan, 2007; McCormack et al., 2012; Song et al., 2012). Including multiple accessions per species can also decrease the variance around the effective population size parameter within the coalescent framework (Heled and Drummond, 2010).

To further our understanding of evolutionary relationships of *Sarracenia* we conducted target enrichment of nuclear genes from multiple accessions per species sequenced on an Illumina HiSeq platform. Target enrichment involves the use of oligonucleotide probes that retain selected genomic regions for sequencing while reducing non-selected DNA (Mamanova et al., 2010). Target enrichment is highly applicable for phylogenetics as it works well for non-model organisms, is cost-efficient, and allows for an increase in the number of species and individuals for phylogenetic analysis (Faircloth et al., 2012a; Lemmon and Lemmon, 2013). Here, we (1) assessed the utility of this method for a recently radiated, non-model genus, (2) compared the multispecies coalescent approach with a concatenation approach, and (3) determined the evolutionary relationships within *Sarracenia*. The resolved species level phylogeny is then discussed in regard to the current taxonomy, biogeography, and conservation status of this group. Taken together, this multilocus and multiaccessional approach represents the most robust attempt to resolve the *Sarracenia* phylogeny to date and has implications for other recently radiated groups.

2. Material and methods

2.1. Taxon sampling

The majority of leaf tissue was sampled from the Atlanta Botanical Garden, which maintains an extensive living collection of *Sarracenia* species from various localities for conservation and as a reference for the North American Plant Collections Consortium. The remaining samples were collected from plant stocks maintained at the University of Georgia Plant Biology greenhouse and field collections. Current estimates list between 8 to 11 species with many varieties and subspecies being designated to the species level based on differing taxonomic schemes (Ellison et al., 2014). We sampled 71 *Sarracenia* accessions covering putative species, varieties, and subspecies. These include the eleven species recognized by Mellichamp and Case (2009) (*alabamensis*, *alata*, *flava*, *jonesii*, *leucophylla*, *minor*, *oreophila*, *psittacina*, *purpurea*, *rosea*, *rubra*) with 1–8 localities spanning the southeastern range of each species (see Table A.1) and additional samples from Maryland, Nova Scotia, and Wisconsin for *purpurea* ssp. *purpurea*. The 71 accessions also include three subspecies/varieties from the *purpurea* complex (ssp. *purpurea*, ssp. *venosa*, ssp. *venosa* var. *montana*), two subspecies from the *rubra* complex (ssp. *gulfensis*, ssp. *wherryi*), one *minor* variety (var. *okefenokeensis*), and two *flava* varieties (var. *rugelii*, var. *rubricorpora*). These putative subspecies and varieties are based on a combination of taxonomic descriptions between Mellichamp and Case (2009) and McPherson and Schnell (2011). Taxonomic descriptions have frequently designated *alabamensis* and *jonesii* as subspecies within the *rubra* complex and *rosea* as *purpurea* ssp. *venosa* var. *burkii*. Three *Darlingtonia californica* and one *Heliamphora minor* (both within Sarraceniaceae) were used as outgroups for the genus. This coverage of varieties, subspecies, and range distribution of putative species allows for a comprehensive analysis of this genus. Voucher specimens were deposited in either the University of Georgia Herbarium (UGA) or the Texas A&M Herbarium (TAES) (Table A.1).

2.2. Probe design

Targets for enrichment were initially identified by aligning *Sarracenia psittacina* and *S. purpurea* transcriptomes (Srivastava et al., 2011). All repeat-like regions were masked using RepeatMasker (http://www.repeatmasker.org/) prior to probe design. Targets with promising single nucleotide polymorphisms for phylogenetic analyses and at least two independent reads from each species were selected for further processing (~1000 contigs). Because paralogous sequences are not ideal for phylogenetic inference due to their independent evolutionary histories, potential targets were screened for paralogous signals using two methods. First, a within-species BLAST (Altschul et al., 1997) search of possible paralogous sequences was conducted with a stringent e-value cut off of $<3 \times 10^{-20}$. A reciprocal best BLAST(blastn) hit approach was then used on the subsequent targets to determine orthologous sequences between the two species. Targets that did not meet the cut off criterion were discarded from the potential target database; this resulted in 646 genes for target sequencing. Previous work suggests that *Sarracenia* may be a partial polyploid (Srivastava et al., 2011); however, we are confident that our stringent screening of paralogs prior to probe design and additional downstream removal of duplicates adequately addresses this possible source of conflict. Approximately three 120-mer oligonucleotide probes were designed for each gene per the manufacturer's probe design specifications. These probes were commercially synthesized by Mycroarray® into a custom MYbaits kit (http://www.mycroarray.com; Ann Arbor, MI).

2.3. DNA extraction, library preparation, sequencing

All leaves (i.e. pitchers) were cut near the base of the plant, sliced open, and cleaned of any insect residue, algae, soil, and other particulates. Areas of the leaf that were senescing, discolored, or greatly impacted from decomposing insect prey were removed