RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*[1]

Rob Massatti[2], Anton A. Reznicek, and L. Lacey Knowles

PREMISE OF THE STUDY: Relationships among closely related and recently diverged taxa can be especially difficult to resolve. Here we use both Sanger sequencing and next-generation RADseq data sets to estimate phylogenetic relationships among species of *Carex* section *Racemosae* (Cyperaceae), a clade largely restricted to high latitudes and elevations. Interest in relationships among these taxa derives from questions about the species' biogeographic histories and possible links between diversification and Pleistocene glaciations.

METHODS: A combination of approaches and molecular markers were used to estimate relationships among *Carex* species within sect. *Racemosae* and taxa from closely related sections. Nuclear and chloroplast loci generated by Sanger sequencing were analyzed with *BEAST, and SNP data from RADseq loci were analyzed as a concatenated data set using maximum likelihood and as independent loci using SVDquartets.

KEY RESULTS: Sanger sequencing data sets resolved relationships among taxa at intermediate phylogenetic depths (albeit with low levels of support). Only the RADseq data resolved relationships with strong support at all phylogenetic depths. Moreover, different methods and data partitions of the RADseq data resulted in nearly identical topologies. *Carex* sect. *Racemosae* is a strongly supported clade, although a handful of species were found to group with closely related sections. Herbarium specimens up to 35 yr old successfully produced informative RADseq data.

CONCLUSIONS: Despite the short read lengths of RADseq data, they nevertheless resolved relationships that Sanger sequencing data did not. Resolution of the phylogenetic relationships among recently and rapidly diversifying taxa within sect. *Racemosae* clades suggest a role for the Pleistocene glaciations in clade diversification.

KEY WORDS *Carex*; Cyperaceae; glaciation; next-generation sequencing; Pleistocene; section *Racemosae*; SVDquartets

Phylogenetic hypotheses provide a key framework for testing mechanisms underlying macroevolutionary patterns, including processes important in structuring communities (Cavender-Bares et al., 2004; Donoghue, 2008; Wiens et al., 2010), the distribution of morphological traits within and among communities (Kraft et al., 2007), and the dispersal and diversification of clades across geographic barriers (Tkach et al., 2008). Today, phylogenies commonly incorporate nuclear and chloroplast (and/or mitochondrial) loci. Although the inclusion of such sequence data has generally led to a vast improvement over exclusive use of morphological data (e.g., see classifications by Cronquist [1981] or Takhtajan [1987] vs. APG III [2009]), phylogenetic inference with molecular data sets is not free of challenges. For example, gene incongruence due to horizontal gene transfer, gene duplication and loss, hybridization, and/or incomplete lineage sorting lead to difficulties reconstructing phylogenetic relationships (Maddison, 1997). In addition, phylogenetic analyses that rely on limited numbers of loci, even with methods that accommodate processes such as gene-lineage coalescence (Knowles, 2009), may leave a proportion of relationships unresolved. Obtaining resolved phylogenetic relationships becomes especially challenging when the times between speciation events are relatively short, such that a high mutational variance among loci may contribute to insufficient data for estimating relationships (Huang and Knowles, 2014).

Now that genomic data are fairly easy to generate, we have the opportunity to create well-resolved phylogenetic hypotheses while potentially alleviating some of the problems involved with Sanger

[1] Manuscript received 5 July 2015; revision accepted 29 December 2015.
Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor, Michigan, 41809-1079 USA
[2] Author for correspondence (e-mail: massatti@umich.edu)
doi:10.3732/ajb.1500315

sequencing data sets (e.g., lack of power to resolve nodes at various ages). In particular, restriction-associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008; Rowe et al., 2011) data may be useful because it produces abundant, anonymous data from throughout the genome that can be used for phylogenetic inference (Eaton and Ree, 2013; Hipp et al., 2014). Because of the importance of identifying a sufficient number of orthologous restriction sites among species (e.g., Rubin et al., 2012), RADseq data has typically been used in phylogenetic analysis of recently derived clades (Heliconius Genome Consortium, 2013; Eaton and Ree, 2013, Wagner et al., 2013; Takahashi et al., 2014). However, temporal boundaries are being continually expanded. For example, using a RADseq data set, Hipp et al. (2014) inferred relationships within a clade of oaks estimated to be 23–33 million years old. The limitations of RADseq data at greater phylogenetic depths, where homoplasy might erode the phylogenetic signal contained across the single nucleotide polymorphisms (SNPs) of independent loci, is not yet known; at some threshold, older evolutionary relationships will be better assessed using alternative data sets, such as those composed of transcriptome data (e.g., Wickett et al., 2014).

Carex L. is a large and diverse genus for which Sanger sequencing data have often failed to resolve well-supported relationships across multiple phylogenetic levels (Roalson et al., 2001; Hipp et al., 2006; Starr and Ford, 2009). The taxa are ecologically important and distributed throughout many habitat types worldwide (Reznicek, 1990). *Carex* section *Racemosae* G.Don, one of the largest sections within the genus, contains roughly 70 species mostly distributed at higher latitudes and elevations in the northern hemisphere (Egorova, 1999; Murray, 2002; Lunkai et al., 2010; see Gebauer et al., 2015 for a history of the section's circumscription). Section *Racemosae* species are predominantly restricted to cold, harsh habitats that initially became widespread and interconnected during the Pliocene (Matthews, 1979; Matthews and Ovenden, 1990), suggesting that the crown group diversified relatively recently. In addition, montane clades impacted by glaciations may represent more recent, Pleistocene diversifications (e.g., Massatti and Knowles, 2014). As such, sect. *Racemosae* offers an opportunity to test the utility of RADseq data for resolving relationships over recent to intermediate phylogenetic depths. These taxa are united by shared morphological traits including trigonous achenes and multiple spikes, where the terminal spike is gynecandrous or unisexual. Despite this unique combination of traits, morphological similarities do not necessarily translate into close evolutionary relationships (Gebauer et al., 2015). Furthermore, several species, despite morphological similarities, do not belong within the section (Hendrichs et al., 2004; Waterway et al., 2009). While previous studies that included sect. *Racemosae* taxa suggested interesting and complicated morphological dynamics among species, they generally suffered from a paucity of molecular data that leaves many relationships unsupported and from a dependency on high copy nuclear loci (e.g., ITS) that may mislead phylogenetic inference due to intraindividual polymorphisms and incomplete concerted evolution (see King and Roalson, 2008).

By generating phylogenetic hypotheses from both Sanger sequencing and RADseq data sets, here we investigated the utility of different data types to resolve relationships within sect. *Racemosae* across multiple phylogenetic depths (i.e., among and within species groups). We also analyzed the RADseq data set with multiple methodologies to determine the robustness of the inferred topologies. On the basis of these analyses, we evaluated the hypothesis that sect. *Racemosae* sensu lato is a monophyletic clade and assessed its purported distinctiveness from other species and sections, including some postulated close relatives based on morphological similarities. Our analyses provide a highly refined estimate of phylogenetic relationships among *Carex* species in one of the largest sections in the genus, and they also provide general insights about the utility of RADseq data and inference methods that are applicable to nonmodel taxa.

MATERIALS AND METHODS

Taxon sampling—We sampled 111 plant specimens representing 48 species within sect. *Racemosae*, as well as 21 species from eight closely related sections (Appendix S1, see Supplemental Data with the online version of this article). Sampling represents nearly all of the sect. *Racemosae* taxa from North America (excluding a couple of recent segregates, e.g., *Carex orestera* Zika) and roughly 75% of taxa found worldwide. Unsampled taxa occur mainly as narrow endemics in Asian mountain ranges and are not well represented within accessible herbarium collections. Multiple individuals were sampled where possible to facilitate species tree reconstructions (see below). About 0.30 g of leaf tissue was sampled from herbarium specimens housed at A, COLO, DAO, GH, MICH, MO, NAVA, RM, and SI or from live plants; the leaf material was stored in silica gel until DNA was extracted using DNeasy Plant Mini Kits (Qiagen, Valencia, California, USA) following the manufacturer's protocol.

Sanger sequencing data set—We amplified two nuclear and one chloroplast marker using a polymerase chain reaction (PCR) in an Eppendorf Thermal Cycler (Hauppauge, New York, USA); these loci were shown to have phylogenetic utility in previous *Carex* studies (Starr et al., 2003, 2009). All loci were successfully amplified in the following 20 μL reaction mixture: 10× reaction buffer, 10 mM dNTPs, 1 μM of both primers, 1 U TaKaRa Ex Taq (Clontech, Mountain View, California, USA), 25 mM $MgCl_2$, and 1 M betaine. The primers ITS4 and ITS1 (White et al., 1990) were used to amplify the internal transcribed spacer region (ITS) and the included 5.8S rDNA. DNA amplification followed a thermal cycler protocol that included: 95°C initial denaturing for 4 min; followed by 35 cycles of 95°C denaturing for 1 min, 50°C annealing for 1 min, and 72°C extension for 1 min; followed by a final extension at 72°C for 15 min. PCR of the external transcribed fragment (ETS) employed the primers ETS-1f and 18S-R (Starr et al., 2003); amplification followed the thermal cycler protocol of Starr et al. (2003). PCR of the chloroplast *matK* locus was completed using the primers MOmatK480F and trnK2R (Hilu et al., 2003). While most accessions were successfully amplified using the two external primers, some were amplified in two parts using two internal primers: matK-2.1f (5′ CCT ATC CAT CTG GAA ATC TTA G 3′) and matK-5r (5′ GTT CTA GCA CAA GAA AGT CG 3′). Amplification of the chloroplast DNA employed the following thermal cycler program: 94°C initial denaturing for 2 min; followed by 35 cycles of 94°C denaturing for 1 min, 48°C annealing for 30 s, and 72°C extension for 1 min; followed by a final extension at 72°C for 10 min.

The PCR products were electrophoresed in 2% agarose gels in 1× Tris-borate-EDTA (TBE) buffer (pH 8.0) and stained with ethidium bromide. The amplified templates were processed at the University of Michigan Sequencing Core using the same primers used in the PCR reactions. Chromatograms were proofed by visual inspection

and edited using the program Sequencher v.4.2 (Gene Codes Corp., Ann Arbor, Michigan, USA). Sequences were initially aligned in the programs MEGA v.6.06 (Tamura et al., 2013) with MUSCLE (Edgar, 2004) using a 'Gap Open Penalty' of −400 and the UPGMB clustering method. Subsequently, the initial alignment was visually inspected for accuracy. Indels and the poly A tail at the 3′ end of ITS1 were excluded from phylogenetic analyses. Sequences for all specimens are available on GenBank (Appendix S1).

RADseq data set—We selected 58 specimens representing 48 species within sect. *Racemosae* and 7 species from closely related sections for our RADseq data collection efforts (bolded taxa, Appendix S1). For these specimens, extracted genomic DNA was individually barcoded and processed into a reduced complexity library using a restriction-fragment-based procedure (for details, see Peterson et al., 2012). Briefly, DNA was doubly digested with EcoRI and MseI restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique barcodes. Ligation products were pooled across samples and the fragments were amplified by 12 cycles of PCR. A Pippin Prep (Sage Science, Beverly, Massachusetts, USA) was used to size select fragments between 400 and 500 bp. The library was sequenced at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada) in one lane on the Illumina HiSeq2500 platform to generate 150-bp, single-end reads. Sequences were demultiplexed using process_radtags.pl, which is distributed as part of the Stacks pipeline (Catchen et al., 2013); only reads with Phred scores ≥32, no adaptor contamination, and unambiguous barcode and restriction cut sites were retained. Potential chloroplast and mitochondrial sequences were filtered from the processed data set using the program Bowtie 0.12.8 (Langmead et al., 2009). Because of the lack of sequenced chloroplast and mitochondrial genomes in *Carex*, chloroplast and mitochondrial genomes within Poaceae were used. Specifically, the data were compared with genomes downloaded from GenBank, including *Agrostis stolonifera* (NC_008591.1), *Oryza nivara* (AP006728.1), and *Zea mays* (X86563.2) for identifying potential chloroplast genes, and *Zea perennis* (DQ645538.1) and *Triticum aestivum* (EU534409.1) for identifying potential mitochondrial genes. Given the relative slow rates of molecular evolution characterizing chloroplast and plant mitochondrial genomes (Wolfe et al., 1987), a tolerance of 2 mismatches (-v 2) between *Carex* sequences and these genomes was used to identify chloroplast and mitochondrial sequences that were removed from the data set.

Data were analyzed in the pyRAD pipeline, which accounts for indels that may be present among species' homologous loci (for details, see Eaton, 2014). Briefly, sequences of each individual are clustered into highly similar stacks using USEARCH (Edgar, 2010), followed by the estimation of rates of heterozygosity and sequencing error (Lynch, 2008). Heterozygotes are inferred by a binomial probability based on these parameters. Each resulting stack is hereafter referred to as a RADseq locus. Each individual's RADseq loci are independently summarized into consensus sequences, which are subsequently clustered among individuals to generate a data matrix. Because not every individual has a sequence for every RADseq locus, due to both variation in sequencing coverage and mutations in the restriction site defining the RAD loci, the resulting data matrix is expected to be incomplete. Parameter values used for clustering were chosen based on suggestions from the literature (Eaton and Ree, 2013; Eaton, 2014; Takahashi et al., 2014). The percentage similarity required to cluster sequences into a stack and individuals into a locus was 0.88, the minimum stack depth for each individual was 6, and the minimum number of individuals per locus cluster was 4; full parameter details are archived with associated data sets on Dryad (http://doi.org/10.5061/dryad.4b8j7). The number of shared loci among taxa was visualized using the corrplot function in the 'corrplot' package (Wei, 2015) in the program R (R Core Team, 2014).

Phylogenetic inference: Sanger sequencing data set—To allow idiosyncratic patterns of descent from ancestral alleles (e.g., Degnan and Rosenberg, 2009), we estimated a species tree using the *BEAST template in the program BEAST v2.1.3 (Bouckaert et al., 2014). *BEAST accommodates differences in the genealogical histories of loci (in contrast to concatenating the data), while also accommodating uncertainty in the estimates of the gene trees for the independent loci. We used two data sets in species tree reconstructions. The first included 76 accessions representing 47 taxa for which we acquired RADseq data and is hereafter referred to as the "reduced Sanger data set". In contrast, the "full Sanger data set" was composed of 107 accessions representing 70 taxa; this analysis included all Sanger sequencing data that we generated. We linked the ITS and ETS tree models because of their physical linkage, while the *matK* tree model remained unlinked. Substitution (GTR + Γ) and lognormal relaxed clock models were unlinked across loci, and we used a Yule process species tree prior. Preliminary analyses using a range of less complex substitution models resolved relationships within sect. *Racemosae* to a similar extent as the GTR + Γ model; because our data contained sufficient information to estimate the parameters of the GTR + Γ model (as assessed by inspecting marginal densities and ESS scores, see below), we only used this model in our final analyses. Three independent runs per data set (six runs total) were conducted to compare the similarity of the maximum clade credibility trees and nodal posterior probabilities. Each replicate ran for 1e9 Markov chain Monte Carlo (MCMC) iterations with trees sampled every 10,000 generations (for a total of 100,000 sampled trees). Visual inspection of the marginal densities of the estimated parameters and their associated effective sample size (ESS) scores in the program Tracer v1.6 (Rambaut et al., 2014) was used to assess convergence, which occurred within the first 20,000 MCMC generations in the six independent *BEAST runs. Maximum clade credibility trees were generated using TreeAnnotator v.2.1.2 (distributed with BEAST v2.1.3) after discarding the first 20% of sampled trees (resulting in 80,000 trees used per run). Analyses utilizing the chloroplast locus or nuclear loci independently resulted in relationships among taxa that were variable and unsupported; these analyses are not shown.

Phylogenetic inference: SNP data set—A data set consisting of unlinked SNPs created by pyRAD from the RADseq loci was used in multiple phylogenetic analyses. First, a phylogeny was estimated from the concatenated SNP data set using maximum likelihood (ML) in the program RAxML v.8.1.16 (Stamatakis, 2014), with a GTR + Γ model of nucleotide evolution. Likelihood calculations were corrected for ascertainment bias (-m ASC_GTRGAMMA); this option is useful for SNP data sets that contain no invariable sites, which will cause the algorithm to exit with errors. Therefore, before analysis, the data set was filtered with a custom script (http://doi.org/10.5061/dryad.4b8j7) to exclude sites that were potentially invariable due to ambiguous characters. Support was assessed by 500 nonparametric bootstrap replicates, followed by a search for the best-scoring maximum likelihood tree.

The relationships among and within clades suggested by the ML analysis of the concatenated SNP data set were further investigated by analyses that account for differences in the genealogical histories of individual loci. Specifically, we used the program SVDquartets (Chifman and Kubatko, 2014), which is a quartet sampling method that accounts for sequence variability due to both mutational and coalescent variance. Because SVDquartets uses the sequence data directly and it does not require long sequence reads to estimate a gene tree, it provides an advantage over summary-statistics methods for estimating species trees such as STEM (Kubatko et al., 2009), STAR (Liu et al., 2009), and MP-EST (Liu et al., 2010). In addition, the method is rapid and results are straightforward to interpret, in contrast to other SNP-based approaches that use MCMC frameworks (e.g., SNAPP, Bryant et al., 2012), which can be slow for large data sets and difficult to assess convergence. Note, however, that a large number of quartets must be sampled to estimate phylogenetic relationships. To maximize the coverage of shared data across taxa (i.e., the overlap in shared loci decreases as a function of the time since common ancestry; Huang and Knowles, 2014), only one outgroup and sect. *Racemosae* s.s. (39 species total; see Results) were included in the analyses. Three independent runs of SVDquartets were conducted in the program PAUP* 4.0 (Swofford, 2003) to assess topological convergence, each of which included 500 bootstrap replicates and exhaustive quartet sampling.

RESULTS

Sequence analysis: Sanger sequencing data set—The ITS and ETS nuclear regions and the *matK* chloroplast region were successfully amplified in 106, 105, and 95 specimens, respectively (Appendix S1). Alignment of the full nuclear data set (ITS + ETS) yielded 1274 characters, of which 474 were variable and 199 (15.6%) were phylogenetically informative. Alignment of all chloroplast sequences (*matK*) yielded 1209 characters, of which 192 were variable and 88 (7.3%) were phylogenetically informative. All phylogenetically informative sites were retained after pruning out taxa to create the reduced Sanger data set.

Sequence analysis: SNP data set—Of the 58 specimens included in the Illumina library, nine were discarded because they had low numbers of raw reads and few loci (online Appendix S2). While four of the excluded taxa were sampled between 1984 and 1997, the other five represent 5 of 6 taxa sampled from vouchers before 1980, suggesting a limit to the usefulness of herbarium collections more than roughly 35 yr old. In addition, the *Carex atrosquama* Mack. accession was found to be misidentified (likely a specimen of *Carex scopulorum* Holm) and was excluded from analyses.

After data processing, taxa had on average 12,598 unlinked SNPs in the final data set, based on an average of 1,674,947 reads per specimen (Table 1; Appendix S2). However, the proportion of shared loci varied among taxa. Specifically, the initial concatenated data matrix contained 41,615 loci, with as few as 1.2% shared between more distantly related taxa (e.g., *Carex adelostoma* V.I.Krecz. and *C. vaginata* Tausch) and as many as 37.2% shared between closely related taxa (e.g., *Carex nova* L.H.Bailey and *Carex nelsonii* Mack.; online Appendix S3). After filtering out invariable sites due to ambiguous characters, 12,689 SNPs (http://doi.org/10.5061/dryad.4b8j7) remained for the maximum likelihood analysis of the concatenated data with RAxML, while the entire data matrix (41,615 unlinked SNPs) was used for the SVDquartets analyses.

Phylogenetic inference: Sanger sequencing data set—The three loci comprising the full and reduced Sanger data sets consistently resolved recently derived clades (numbered circles in Fig. 1 and online Appendix S4), except for the placements of *Carex obscura* Nees and *Carex melanocephala* Turcz., which were nested within Clade 5 in analyses on the full Sanger data set, but independent of defined clades in analyses on the reduced Sanger data set. However, relationships among these clades (i.e., the backbone topology) and those within many clades were unresolved and differed between the two data sets. In addition, relationships within and among recently derived clades were variable across independent *BEAST runs, even though convergence appeared to be reached in every instance (see details in Methods). We only present the maximum clade credibility trees from the first of the three *BEAST runs, and except where noted, we only refer to the tree generated with the reduced Sanger data set (Fig. 1). All posterior probabilities (PPs) were low, except on some branches subtending aforementioned recently derived clades and a couple of branches within these clades (Fig. 1; Appendix S4). Nevertheless, moderate support (82% PP) was found for the monophyly of sect. *Racemosae* s.s., which excludes five taxa typically treated as members of this clade: *Carex augustinowiczii* Meinsh. ex Korsh., *Carex curvicollis* Franch. & Sav., *Carex mertensii* J.D.Prescott ex Bong., *Carex meyeriana* Kunth, and *Carex stylosa* C.A.Mey.

Phylogenetic inference: SNP data set—The ML and SVDquartets analyses of the respective SNP data sets produced very consistent results (Figs. 2 and 3). The monophyly of sect. *Racemosae* s.s. was highly supported, as was the sister relationship between sect.

TABLE 1. Data processing summary statistics detailing next-generation sequencing and the single nucleotide polymorphism (SNP) data set. Raw reads refers to the total reads produced by Illumina sequencing; postprocessing reads are those that remained after filtering for adaptor contamination, quality, ambiguous barcodes, and chloroplast/mitochondrial sequences. The postprocessing reads were used by pyRAD to create clusters of homologous sequences per species (Total clusters, Mean depth of clusters). After the heterozygosity (*H*) and error-rate (*E*) were estimated across clusters, consensus sequences were created for each cluster; those that passed the pyRAD filtering parameters were retained (Loci). Variable and invariable DNA sites were summed across all loci (Total sites), and the percentage of polymorphic sites (Percent poly) is reported. Consensus sequences were clustered across species, and loci that passed filtering parameters were included in the final data matrix (Final loci). Refer to online Appendix S2 for a breakdown by species, as well as for the taxa not included in the final SNP data set.

Summary statistic	Raw reads	Postprocessing reads	Percent retained	Total clusters	Mean depth of clusters	*H*	*E*	Loci	Total sites	Percent poly	Final loci
Average	1,882,068	1,674,947	88.0	23,741	53.4	0.0104	0.0009	20,864	2,815,076	0.18	12,598
Minimum	138,966	107,088	77.1	4214	13.0	0.0068	0.0004	3234	436,159	0.06	1564
Maximum	6,307,952	5,746,439	93.1	49,313	114.8	0.0228	0.0032	42,722	5,766,499	0.47	27,988
SD	1,321,887	1,195,946	3.0	10,853	21.3	0.0036	0.0005	9933	1,340,434	0.12	6693

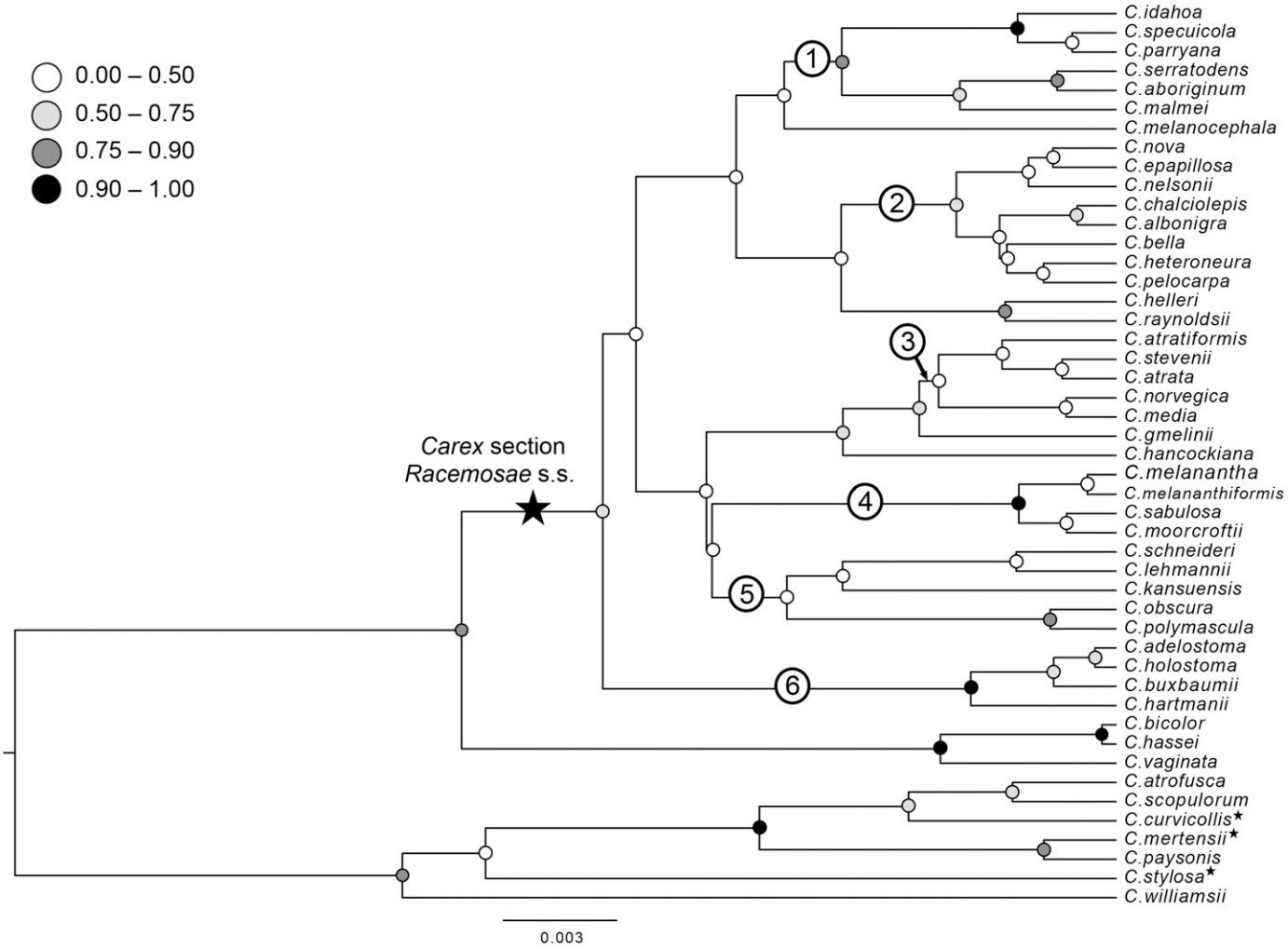

FIGURE 1 Maximum clade credibility tree detailing the relationships of sect. *Racemosae* and closely related taxa estimated using *BEAST and the reduced Sanger data set, which is composed of ITS and ETS nuclear loci and the *matK* chloroplast locus for 76 accessions representing 47 taxa. Shaded circles on nodes represent posterior probabilities estimated from 80,000 post burn-in trees. Numbered circles representing clades discussed in the text are used throughout the figures. The branch subtending sect. *Racemosae* s.s. is starred, as are the sect. *Racemosae* taxa that cluster with taxa representing other sections. See online Appendix S4 for the phylogeny estimated using all taxa for which Sanger sequencing data were collected.

Racemosae and *C. bicolor* Bellardi ex All. (and additionally *C. hassei* L.H.Bailey, *C. vaginata,* and *C. williamsii* Britton in the ML analyses that included more outgroup taxa). These analyses also corroborated phylogenetic estimates of the recently derived clades that were consistently resolved using the Sanger data sets (as noted by the same circled numbers in Figs. 2 and 3 as in Fig. 1), including the placement of *C. obscura* and *C. melanocephala* as independent of defined clades, similar to the reduced Sanger data set (see above). The best-scoring ML tree contained many well-supported nodes, most with 100% bootstrap support, although multiple nodes had lower bootstrap values (Fig. 2). Bootstrap values in the SVDquartets phylogeny largely reiterated those resolved in the ML analysis, with the lowest support values clustered in and around Clade 2 and Clade 3, as well as in the placement of *C. obscura* (Fig. 3). The only difference between the ML and SVDquartets topologies was the placement of the clade containing *Carex helleri* Mack. and *Carex raynoldsii* Dewey (see Figs. 2 and 3); while these taxa were placed as sister to Clade 2 with strong support in the ML phylogeny, the quartet phylogeny placed them sister to the clade containing Clade 2, Clade 3, *C. melanocephala*, *Carex gmelinii* Hook. & Arn., and *Carex hancockiana* Maxim., albeit with weak support.

DISCUSSION

Our study highlights how next-generation sequencing, and in particular RADseq data, can provide phylogenetic information for resolving relationships among taxa over a variety of temporal depths, including for relationships that lack resolution because Sanger sequencing data are uninformative or morphological characters are ambiguous. Moreover, recently derived clades suggested by Sanger sequencing data, and in some instances morphology, are not only corroborated by SNP data, but relationships within and among these clades are consistent across multiple inference methods. Our analyses suggest that RADseq data have high potential for future phylogenetic research, including instances where conflict among

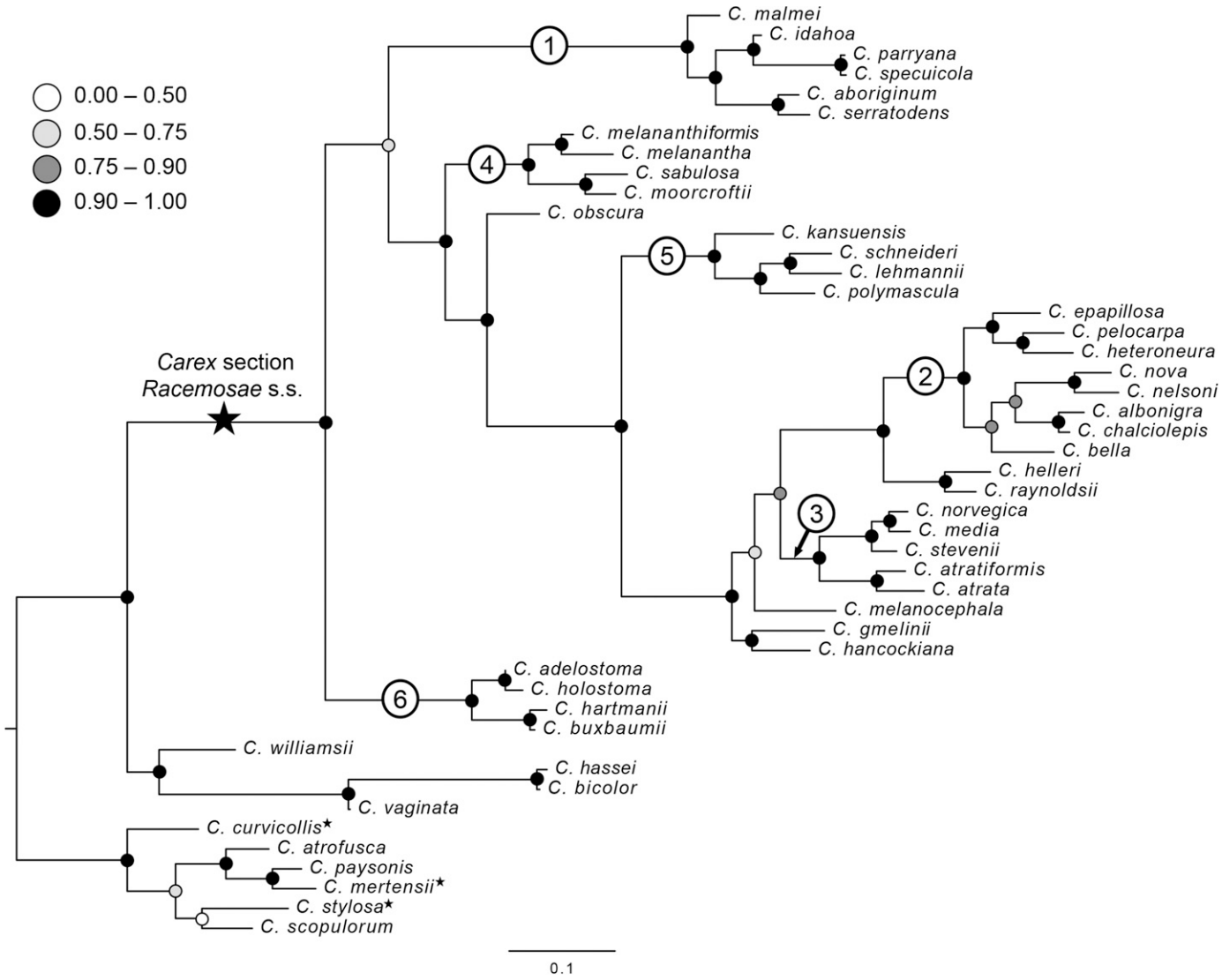

FIGURE 2 The best-scoring maximum likelihood tree estimated by RAxML using 12,689 unlinked SNPs. Shaded circles on the nodes illustrate nonparametric bootstrap values. Numbered circles correspond to clades depicted in other figures and are discussed in the text. The branch subtending sect. *Racemosae* s.s. is starred, as are the sect. *Racemosae* taxa that cluster with taxa representing other sections.

subsets of Sanger sequencing data and/or methods of analyses is commonplace. The utility of RADseq data is especially heartening in groups where particular biological phenomena that directly contribute to difficulties with phylogenetic analysis may be of interest (e.g., rapid and recent species diversification). In contrast, when using only Sanger sequencing data, it is unclear whether low support for relationships among taxa should be attributed to biological phenomena (e.g., short internodes associated with rapid diversification) vs. a paucity of data (e.g., relying on traditional markers that amplify well, but have low mutation rates; Lanier et al., 2014). Finally, our data suggest that herbarium specimens up to 35 yr old may by used within library construction protocols with no modification for specimen age.

Sanger vs. RADseq data for phylogenetic inference—Several patterns stand out when comparing the phylogenetic hypotheses inferred utilizing Sanger sequencing vs. the SNP data sets. First, both types of data and all methods of inference resolved the same recently derived clades (albeit with low support in the Sanger data sets; see numbered clades throughout all figures), supporting the robustness of these relationships and suggesting that both types of data are informative at this phylogenetic depth (see Escudero et al., 2014 for a similar result in *Carex* section *Ovales* Kunth). However, whereas ML and quartet analyses of the SNP data sets resolved largely the same, highly supported topology (Figs. 2 and 3; see below for differences), no consistent relationships among recently derived clades are apparent among the Sanger data sets (either the full data set or the reduced data set that matches the SNP data set; Fig. 1; Appendix S4). Similarly, relationships within recently derived clades are labile among analyses on the Sanger data sets, with the exception of a few highly supported nodes that are consistent across analyses (e.g., *Carex serratodens* W.Boott–*Carex aboriginum* M.E.Jones and *Carex moorcroftii* Falc. ex Boott–*Carex sabulosa* Turcz. ex Kunth; Figs. 1–3).

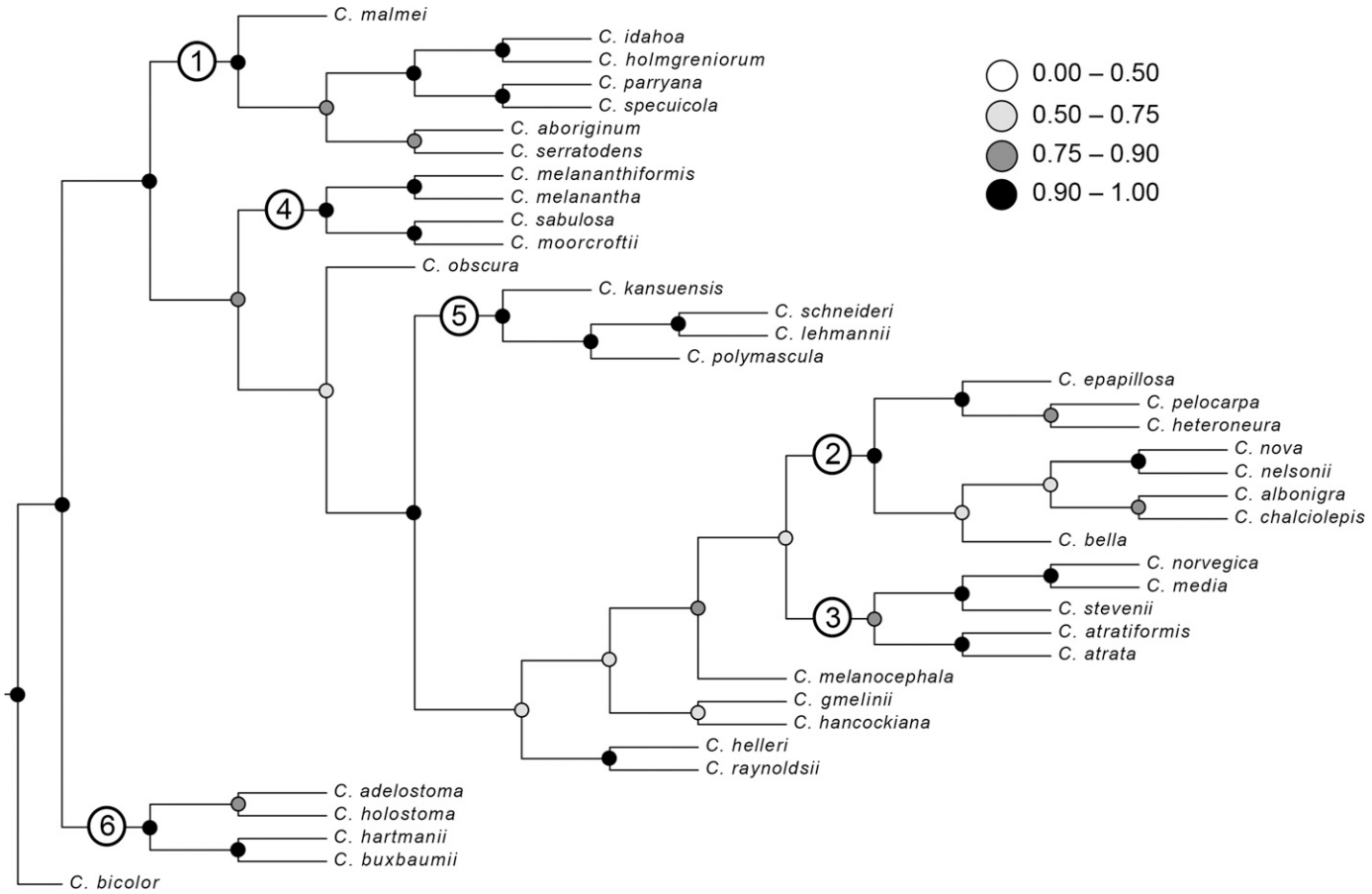

FIGURE 3 Relationships among sect. *Racemosae* species as resolved by exhaustive quartet sampling in a data matrix containing 41,615 SNPs and 39 taxa using SVDquartets. Shaded circles on the nodes illustrate the proportion of bootstrap replicates that support the respective relationship. Numbered circles correspond to clades depicted in other figures and are discussed in the text. All species belong to sect. *Racemosae* s.s. except for one outgroup taxon (*Carex bicolor*). Branch lengths are not meaningful.

While phylogenetic inference of the SNP data set produced a strongly supported and robust hypothesis for sect. *Racemosae,* there are several issues that need special attention before this methodology is broadly applied across clades. Where divergence times between taxa or clades are known to be old, care should be taken to identify and exclude loci that have high mutation rates because they may display extensive homoplasy. Our investigation of mutation rates across SNP loci reinforced the phylogenetic utility of the SNPs with the fastest rates within our data set, and therefore they were not excluded (data not shown). As mentioned previously, because the taxa in sect. *Racemosae* are almost exclusively alpine and arctic, and this flora has only been well established and interconnected since roughly the Pliocene, it is possible that the genetic variation within these taxa has not reached the threshold beyond which homoplasy degrades phylogenetic signal. In groups where species and/or clades are known to be more distantly related, targeted enrichment approaches (e.g., Nicholls et al., 2015) may alleviate the problematic issue of identifying orthologous loci.

Library construction is another important issue to consider when using RADseq data for phylogenetic inference. The number of samples to include in highly multiplexed libraries must be carefully calibrated to ensure sufficient depth is achieved across all taxa. Even with the most careful preparation, significant variability in coverage among taxa may be present and may necessitate excluding individuals from downstream analyses. In our analyses, nine individuals were excluded; five of these may have suffered from DNA degradation that inhibited sequencing due to their ages (i.e., the specimens were older than 35 yr). While our data suggest that herbarium specimens yield high-quality sequences, care must be taken to select specimens that have been appropriately collected and stored (Staats et al., 2011). Herbarium specimens older than 35 yr may be amenable to RADseq data generation, although more work may be required during specimen selection, DNA extraction, and quality control before library construction. Specifically, our experience suggests that older specimens may warrant qPCR assays or quantifications of the size fractionations with a Bioanalyzer prior to their inclusion within the library (see Staats et al., 2011 for additional information). While the variability of coverage among taxa and missing data may affect the quality of the phylogenetic inference, it may not be a problem in many cases (see Hovmöller et al., 2013) and excluding data can create a host of potential problems (Huang and Knowles, 2014). Within our SNP data sets, there is not a correlation between the number of loci shared among taxa and the level of support between pairs of taxa or clades, despite high total levels of missing data. For example, the position within the phylogeny of *C. melanocephala* is not well supported (Fig. 2), even though it shares relatively many loci with other

species (Appendix S3). In contrast, the relationships between *C. adelostoma* and *Carex stevenii* (Holm) Kalela and their respective sister taxa are strongly supported, even though they share relatively few loci (Fig. 2; Appendix S3). In addition, some relationships within Clade 2 are not well supported, even though these taxa are among the best represented within the RADseq data set (Appendix S3). That said, there are examples of taxa that share relatively few loci and that also have low support values in their respective clades, for example *C. scopulorum* and *C. stylosa*, which are outgroup species (Fig. 2). Overall, better coverage and a more complete data set should have benefits for creating a well-supported topology, reinforcing the necessity of careful library planning and preparation to ensure that data are generated evenly and in sufficient quantities across individuals included within the library.

Finally, another complication with using RADseq data for phylogenetic estimation regards the variability of topologies inferred from different subsets of a RADseq data set. For example, Takahashi et al. (2014) used a range of criteria to filter a RADseq data set and identify orthologous loci, which resulted in different topologies among 15 species in a ground beetle species flock (see also Wagner et al., 2013; Cruaud et al., 2014). In contrast, our analyses of different data sets identified the same recently derived clades, some of which are also highly supported by the morphological similarity among species (discussed in detail below). In addition, multiple analyses of the SNP data set largely resolved the same backbone topology and relationships within recently derived clades, suggesting that phylogenetics utilizing SNP data within sect. *Racemosae* may be less prone to these complications. However, while we used parameter values informed by the literature, we acknowledge that a more thorough exploration of how parameter space affects our RADseq data set may benefit our final phylogenetic hypothesis. Overall, concerns about the information contained within a SNP data set highlight the importance of investigating data using multiple methodologies and an appropriate data set and reinforce the potential utility of resolving discordance within a SNP data set as a means for generating biologically interesting hypotheses.

General biological phenomena that impact phylogenetic resolution—Backbone nodes with lower support in the ML and SVDquartets analyses of the RADseq data may reflect evolutionary scenarios that facilitated rapid diversification within sect. *Racemosae*. For example, Clade 3 is highly nested within the pectinate sect. *Racemosae* phylogeny (when considering the remaining numbered clades), and it contains species with a mixture of distributions from widespread at higher latitudes to narrowly endemic. Furthermore, all of the remaining numbered clades, except for Clade 6, are endemic to single continents (or North America + South America for Clade 1). If the ancestor of Clade 3 was distributed across high latitudes similar to extant species, multiple instances of peripatric speciation over a short period of time may have facilitated the generation of the ancestors to the other clades/independent species composing sect. *Racemosae*. This scenario is biologically plausible because these species are all adapted to habitats heavily disturbed by Pleistocene climatic oscillations (e.g., Hewitt, 1996; Weir and Schluter, 2004; Carstens and Knowles, 2007). In turn, this process may have created significant conflict among loci and unresolved gene trees, which would manifest as low nodal support between clades or between a clade and an individual species such as *C. obscura* or *C. melanocephala* (Huang et al., 2010, 2014). A similar process may have been reiterated at a smaller scale within Clade 2, which also contains nodes that have lower support values as determined by the ML and SVDquartets analyses of the RADseq data (Figs. 2 and 3). In fact, preliminary analyses on a data set including more individuals per species exclusively from Clade 2 suggest these taxa diversified rapidly during the Pleistocene in conjunction with glacial cycles and may be subject to incomplete lineage sorting, in addition to present-day introgression (e.g., Massatti and Knowles, 2014; R. Massatti, unpublished data).

Phylogenetic relationships of sect. Racemosae—The data presented herein suggest that sect. *Racemosae* is a strongly supported, monophyletic clade. However, the boundary of sect. *Racemosae* needs to be modified to exclude *C. augustinowiczii*, *C. curvicollis*, *C. mertensii*, *C. meyeriana*, and *C. stylosa*; these species have been suggested by previous studies as not belonging with other members of this section (Hendrichs et al., 2004; Waterway et al., 2009; Gebauer et al., 2015). Elucidating the true sectional affiliations of the excluded taxa will require an analysis of all sections closely related to sect. *Racemosae*. Within sect. *Racemosae* s.s., six clades were consistently resolved throughout our phylogenetic inferences. While relationships among these clades differed across analyses based on the Sanger data sets, a backbone was firmly established and supported in both gene tree and species tree inferences of the SNP data sets. Differences among these analyses reflect only the placement of individuals or pairs of taxa (discussed in detail below).

Clades 1 and 2 are exclusive to the New World. Clade 1 includes multiple species endemic to western North America and one endemic to South America. The North American group (seven species included here, *C. utahensis* Reznicek & D.F.Murray was not sampled) has not been previously included in a subsection within sect. *Racemosae*, though they were thought to be closely related based on morphology (e.g., "*Carex parryana* complex" of Mackenzie, 1935 and Reznicek and Murray, 2013). The South American taxon sampled here was included in subsect. *Atropictae* G.Wheeler, which also includes the only two other South American sect. *Racemosae* species [*Carex atropicta* Steud. and *Carex monodynama* (Griseb.) G.Wheeler]. These three South American species are very similar morphologically. This study is the first to suggest that subsect. *Atropictae* and the North American "*Carex parryana* complex" should be treated as one clade. Clade 2 contains eight species that are endemic to montane habitats in western North America. Several species, including *C. nelsonii*, *C. nova*, and *Carex pelocarpa* F.J.Herm., have previously been included in subsect. *Alpinae* Kalela (Murray, 1969; Egorova, 1999), along with species that, in this study, are either members of Clade 3 or unresolved (e.g., *C. melanocephala*); the remaining Clade 2 species have not been previously included within a subsection. This clade also contains *C. orestera* (not included in the current study), which is morphologically similar to *C. albonigra* Mack. (Zika, 2012), and may include *C. atrosquama*, which groups with Clade 2 in the full *BEAST cladogram (Appendix S4).

Clade 3 contains a mixture of species that are geographically widespread (*C. atrata* L.) to narrowly endemic (*C. stevenii*). The five species included in this study have not previously been treated as a clade, most commonly being separated into subsect. *Alpinae*, subsect. *Aterrimae* T.V.Egorova, or remaining unplaced (e.g., Egorova, 1999). Unlike in Clades 1 and 2, it is unclear which additional taxa belong in this group, but some of the unsampled Asian endemics may cluster with these species (see below for details). The full *BEAST analysis suggests that *C. aterrima* Hoppe [alternatively treated as *C. atrata*

subsp. *aterrima* (Hoppe) S.Yun Liang] and *Carex parviflora* Host may belong in this group, along with *C. gmelinii* and *C. hancockiana* (see below for details on the alternative placement of the two latter taxa). The ML and quartet analyses of the SNP data set were in agreement regarding the relationships within this clade.

Clades 4 and 5 are almost exclusively confined to the Asian continent, parallel to the pattern of Clades 1 and 2 being restricted to the New World. Species composing Clade 4 have been previously considered closely related based on morphology (subsect. *Sabulosae* T.V.Egorova; Kreczetovicz, 1935), and the clade likely does not contain additional species. *Carex sabulosa* is the only Clade 4 or 5 species distributed outside of Asia, with a few occurrences in Alaska and Yukon Territory in North America. Clade 5 species have previously been considered members of multiple subsections, including subsect. *Longibracteatae* T.V.Egorova and subsect. *Aterrimae* (Egorova, 1999). Much like Clade 3, it is unclear which additional taxa may be closely related to these species. The full *BEAST analysis suggests that *Carex infuscata* Nees and *Carex caucasica* Steven may belong in this group, along with *C. obscura* and *C. melanocephala*.

Finally, Clade 6 is well supported in analyses of both the Sanger sequencing and SNP data sets (Figs. 1–3). This clade consistently occupies a basal position within sect. *Racemosae* s.s., except in the *BEAST analysis on the full Sanger sequencing data set (Appendix S4). The close relationship among these species has been previously established (e.g., subsect. *Papilliferae* T.V.Egorova), except for *C. holostoma,* which has been treated as a monospecific subsection (i.e., subsect. *Holostomae* T.V.Egorova). The only additional species that Clade 6 likely includes is *Carex tarumensis* Franch., which is scarcely different from *Carex buxbaumii* Wahlenb. Both the ML and quartet analyses support the same relationships among Clade 6 species.

Differences between the ML and quartet phylogenies are apparent in the placement of the clade containing *C. helleri* and *C. raynoldsii* as either sister to Group 2 or sister to the combination of Group 2, Group 3, *C. melanocephala*, *C. gmelinii* (previously treated as a monospecific subsect. *Longiaristatae* T.V.Egorova by Egorova, 1999), and *C. hancockiana* (previously treated as subsect. *Longibracteatae* by Egorova, 1999; Figs. 2 and 3). Both of these taxa are endemic to montane habitat in western North America, although *C. raynoldsii* has a much broader distribution than *C. helleri*. In addition to morphological similarities shared with species included in Clade 2, their distributions suggest that a close relationship with Clade 2 may be more appropriate than a more basal position (i.e., see Fig. 3). The placements of *C. obscura*, *C. melanocephala*, and the clade containing *C. gmelinii* and *C. hancockiana* are also questionable, based on their lability within the phylogeny at low and intermediate levels of quartet sampling by SVDquartets (data not shown), as well as lower bootstrap support for relevant nodes in the ML and SVDquartets analyses. The most common alternative topology nested *C. melanocephala* and the *C. gmelinii*–*C. hancockiana* clade within Clade 3. *Carex obscura* may have a close relationship with the Asian endemic Clade 5, based on its well-supported sister relationship to *C. polymascula* P.C.Li in the analyses on the Sanger data sets (Fig. 1 and Appendix S4).

Relationships among outgroup taxa—Taxa from two sections [*Bicolores* (Tuck. ex L.H.Bailey) Rouy and *Paniceae* G.Don] were resolved as sister to sect. *Racemosae* in all analyses, including *C. bicolor*, *C. hassei*, and *C. vaginata*; in addition, *C. williamsii* (sect. *Chlorostachyae* Tuck. ex Meinsh.) nested with these taxa only in the ML analysis of the SNP data set. While sect. *Bicolores* is monophyletic within this study, sect. *Paniceae* is not, as *Carex livida* (Wahlenb.) Willd. does not cluster with the immediate sister group to sect. *Racemosae* in the full *BEAST analysis (Appendix S4). This pattern is reiterated among the outgroup taxa, in that no traditionally circumscribed sections appear to be monophyletic, either because the taxa are split among multiple clades, or because of the inclusion of taxa formerly considered to be part of sect. *Racemosae* (current sectional assignments of the outgroup taxa are given in Appendix S1). Furthermore, while many authors have considered taxa from sect. *Scitae* Kük. to be very closely related to sect. *Racemosae* (some authors even treat them together; Kreczetovicz, 1935; Mackenzie, 1935), we note here that taxa from sect. *Scitae* are more distantly related than the immediate sister clade to sect. *Racemosae*. Given the lack of support among outgroup taxa in the phylogenies inferred with the Sanger data sets, it may be wise to use next-generation sequencing to develop robust hypotheses for these taxa.

Implications of findings for morphological evolution and taxonomy—Our data suggest that *Carex* morphology, both within and among sections, is extremely labile over evolutionary time. For example, despite previous suggestions that *C. nova* and *C. melanocephala* might better be treated as the same species based on morphological similarities (see Murray, 1969), no evidence for a close relationship is suggested by molecular data. Alternatively, this study is the first to treat the South American taxa in Clade 1 as sister to the remaining North American "*Carex parryana* complex", because morphological traits linking these clades have not previously been identified. However, some recently derived clades (e.g., the North American species in Clade 1, Clade 4, and the majority of Clade 6) are strongly united by morphology, which may suggest that revisiting newly resolved clades and searching for uniting morphological characters will be fruitful.

Given the difficulty in creating monophyletic groups of species based on morphology (e.g., Clades 1–6 often contain species formerly thought to belong in various subsections, or at least not previously grouped together), we do not recommend classifying all but the most morphologically similar taxa (e.g., *C. orestera* in Clade 2, *C. utahensis* in Clade 1, and *C. tarumensis* in Clade 6) into one of the clades recovered herein by morphological features alone. That said, the remaining sect. *Racemosae* species not hitherto discussed likely belong in Clades 3, 4, or 5, or will group independently, with *C. obscura*, *C. melanocephala*, or outside of sect. *Racemosae*. These species are mostly Asian endemics and include: *Carex alsia* Raymond, *Carex aristulifera* P.C.Li, *Carex bijianensis* S.Yun Liang & S.R.Zhang, *Carex decaulescens* V.I.Krecz., *Carex hongyuanensis* Y.C.Tang & S.Yun Liang, *Carex macrostigmatica* Kük., *Carex minxianensis* S.Yun Liang, *Carex montis-wutaii* T.Koyama, *Carex nigerrima* Nelmes, *Carex obliquitruncata* Y.C.Tang & S.Yun Liang, *Carex oligantha* Steud., *Carex peiktusanii* Kom., *Carex pirinensis* Acht., *Carex praeclara* Nelmes, *Carex pseudobicolor* Boeckeler, *Carex serreana* Hand.-Mazz., and *Carex tatjanae* Malyschev. The last sect. *Racemosae* taxon, *C. urostachys,* will likely not fall within sect. *Racemosae* based on its morphological similarity to *C. mertensii*, which is not included in sect. *Racemosae* s.s.

CONCLUSIONS

Next-generation sequencing, in particular RADseq data, provides resolution of the phylogeny of sect. *Racemosae* across multiple

taxonomic levels; no previous study within *Carex* utilizing morphological traits or Sanger sequencing data has achieved a comparable resolution. In fact, our data suggest that, while morphology may accurately group related species, at least across deeper phylogenetic depths, it can be especially misleading when used to infer more recent relationships. While support values for species' relationships can be low in inferences utilizing the SNP data set, lower values may also suggest biologically interesting phenomena that affected the taxa over evolutionary time. As next-generation sequencing technologies continue to develop, it will become easier and cheaper to generate large data sets of orthologous loci distributed throughout organisms' genomes. Furthermore, refined library construction protocols may alleviate technical problems associated with the variability of coverage and missing data among the specimens included in the library and may allow researchers to easily use older museum specimens. Nevertheless, our study supports the utility of RADseq data for phylogenetic estimation, including for clades where the lack of well-resolved and accurate phylogenies have hampered downstream evolutionary inferences.

ACKNOWLEDGEMENTS

The authors are grateful for the support and helpful advice of their colleagues, especially J.-P. Huang for his suggestions on the analyses. In addition, two anonymous reviewers and the Associate Editor provided valuable comments on an earlier draft of this manuscript. This work was funded by NSF (DEB 1309072 to R.M. and L.L.K. and DEB 1118815 to L.L.K.) and Graduate Student Research Awards from the American Society of Plant Taxonomists and the Botanical Society of America to R.M. The authors declare no conflicts of interest.

LITERATURE CITED

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, et al. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Bryant, D., R. Bouckaert, J. Felsenstein, N.A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* 29: 1917–1932.

Carstens, B. C., and L. L. Knowles. 2007. Shifting distributions and speciation: Species divergence during rapid climate change. *Molecular Ecology* 16: 619–627.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. *American Naturalist* 163: 823–843.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

Cronquist, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York, New York, USA.

Cruaud, A., M. Gautier, M. Galan, J. Foucaud, J. Saune, G. Genson, E. Dubois, et al. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. *Molecular Biology and Evolution* 31: 1272–1274.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

Donoghue, M. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549–11555.

Eaton, D. A. R. 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30: 1844–1849.

Eaton, D. A. R., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

Edgar, R. C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461.

Egorova, T. V. 1999. The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). St. Petersburg State Chemical-Pharmaceutical Academy, St. Petersburg, Russia; Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Escudero, M., D. A. R. Eaton, M. Hahn, and A. L. Hipp. 2014. Genotype-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* 79: 359–367.

Gebauer, S., M. Roeser, and M. H. Hoffmann. 2015. Molecular phylogeny of the species-rich *Carex* sect. *Racemosae* (Cyperaceae) based on four nuclear and chloroplast markers. *Systematic Botany* 40: 433–447.

Heliconius Genome Consortium. 2013. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.

Hendrichs, M., F. Oberwinkler, D. Begerow, and R. Bauer. 2004. *Carex*, subgenus *Carex* (Cyperaceae) – A phylogenetic approach using ITS sequences. *Plant Systematics and Evolution* 246: 89–107.

Hewitt, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.

Hilu, K., T. Borsch, K. Muller, D. Soltis, P. Soltis, V. Savolainen, M. Chase, et al. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758.

Hipp, A. L., D. A. R. Eaton, J. Cavender-Bares, E. Fitzek, R. Nipper, and P. S. Manos. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS One* 9: e93975.

Hipp, A. L., A. A. Reznicek, P. E. Rothrock, and J. A. Weber. 2006. Phylogeny and classification of *Carex* section *Ovales* (Cyperaceae). *International Journal of Plant Sciences* 167: 1029–1048.

Hovmöller, R., L. L. Knowles, and L. S. Kubatko. 2013. Effects of missing data on species tree estimation under the coalescent. *Molecular Phylogenetics and Evolution* 69: 1057–1062.

Huang, H., Q. He, L. S. Kubatko, and L. L. Knowles. 2010. Sources of error inherent in species-tree estimation: Impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Systematic Biology* 59: 573–583.

Huang, H., and L. L. Knowles. 2014. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology* 0: 1–9.

Huang, H., L. A. P. Tran, and L. L. Knowles. 2014. Do estimated and actual species phylogenies match? Evaluation of East African cichlid radiations. *Molecular Phylogenetics and Evolution* 78: 56–65.

King, M. G., and E. H. Roalson. 2008. Exploring evolutionary dynamics of nrDNA in *Carex* subgenus *Vignea* (Cyperaceae). *Systematic Botany* 33: 514–524.

Knowles, L. L. 2009. Statistical phylogeography. *Annual Review of Ecology Evolution and Systematics* 40: 593–612.

Kraft, N. J. B., W. K. Cornwell, C. O. Webb, and D. D. Ackerly. 2007. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist* 170: 271–283.

Kreczetovicz, V. I. 1935. *Carex* L. *In* V. L. Komarov [ed.], Flora USSR, vol. 3, 114–464. Academy of Science, Leningrad, USSR.

Kubatko, L. S., B. C. Carstens, and L. L. Knowles. 2009. STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25: 971–973.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.

Lanier, H. C., H. Huang, and L. L. Knowles. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. *Molecular Phylogenetics and Evolution* 70: 112–119.

Liu, L., L. Yu, and S. V. Edwards. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 10: 302.

Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58: 468–477.

Lunkai, D., L. Songyun, Z. Shuren, T. Yancheng, T. Koyama, and C. G. Tucker. 2010. Flora of China, vol. 23, Acoraceae through Cyperaceae. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Lynch, M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Molecular Biology and Evolution* 25: 2409–2419.

Mackenzie, K. K. 1935. Cyperaceae. *North American Flora* 18: 1–478 New York Botanical Garden, Bronx, New York, USA.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Massatti, R., and L. L. Knowles. 2014. Microhabitat differences impact phylogeographic concordance of codistributed species: Genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution* 68: 2833–2846.

Matthews, J. V. 1979. Tertiary and Quaternary environments: Historical background for an analysis of the Canadian insect fauna. *In* H. V. Danks [ed.], Canada and its insect fauna, 31–86. Entomological Society of Canada, Ottawa, Canada.

Matthews, J. V., and L. E. Ovenden. 1990. Late Tertiary plant macrofossils from localities in Arctic/Subarctic North America: A review of the data. *Arctic* 43: 364–392.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.

Murray, D. F. 1969. Taxonomy of *Carex* sect. *Atratae* (Cyperaceae) in the southern Rocky Mountains. *Brittonia* 21: 55–76.

Murray, D. F. 2002. *Carex* L. sect. *Racemosae* G. Don. *In* Flora of North America Editorial Committee [eds.], Flora of North America, north of Mexico, vol. 23, 401–414. Oxford University Press, New York, New York, USA.

Nicholls, J. A., R. T. Pennington, E. J. M. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7: e37135.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website http://www.R-project.org/.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Computer program and documentation distributed by the author, website http://beast.bio.ed.ac.uk/Tracer [accessed 1 July 2014].

Reznicek, A. A. 1990. Evolution in sedges (*Carex*, Cyperaceae). *Canadian Journal of Botany* 68: 1409–1432.

Reznicek, A. A., and D. F. Murray. 2013. A re-evaluation of *Carex specuicola* and the *Carex parryana* complex (Cyperaceae). *Journal of the Botanical Research Institute of Texas* 7: 37–51.

Roalson, E., J. Columbus, and E. Friar. 2001. Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and *trnT-LF* (cpDNA) region sequences: Assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis. Systematic Botany* 26: 318–341.

Rowe, H. C., S. Renaut, and A. Guggisberg. 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 3499–3502.

Rubin, B. E. R., R. H. Ree, and C. S. Moreau. 2012. Inferring phylogenies from RAD SEQUENCE DATA. *PLoS One* 7: e33394.

Staats, M., A. Cuenca, J. E. Richardson, R. Vrielink-van Ginkel, G. Petersen, O. Seberg, and F. T. Bakker. 2011. DNA damage in plant herbarium tissue. *PLoS One* 6: e28448.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Starr, J., and B. Ford. 2009. Phylogeny and evolution in Cariceae (Cyperaceae): Current knowledge and future directions. *Botanical Review* 75: 110–137.

Starr, J., S. Harris, and D. Simpson. 2003. Potential of the 5′ and 3′ ends of the intergenic spacer (IGS) of rDNA in the Cyperaceae: New sequences for lower-level phylogenies in sedges with an example from *Uncinia* Pers. *International Journal of Plant Sciences* 164: 213–227.

Starr, J. R., R. F. C. Naczi, and B. N. Chouinard. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources* 9: 151–163.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0a147. Sinauer, Sunderland, Massachusetts, USA.

Takahashi, T., N. Nagata, and T. Sota. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution* 80: 137–144.

Takhtajan, A. 1987. Systema magnoliophytorum. Editoria Nauka, Leningrad (St. Petersburg), Russia.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Tkach, N. V., M. H. Hoffmann, M. Röser, A. A. Korobkov, and K. B. von Hagen. 2008. Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae). *Evolution* 62: 184–198.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.

Waterway, M. J., T. Hoshino, and T. Masaki. 2009. Phylogeny, species richness, and ecological specialization in Cyperaceae tribe Cariceae. *Botanical Review* 75: 138–159.

Wei, T. 2015. Package 'corrplot'. *Statistician* 56: 316–324.

Weir, J. T., and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London, B, Biological Sciences* 271: 1881–1887.

White, T. J., T. D. Bruns, S. B. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White [eds.], PCR protocols: A guide to methods and applications, 315–322. Academic Press, San Diego, California, USA.

Wickett, N.J., S. Mirarib, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* E4859–E4868.

Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13: 1310–1324.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

Zika, P.F. 2012. *Carex orestera* (Cyperaceae), a new sedge from the mountains of California. *Novon* 22: 118–124.