

## BARCODING PLANTS

# Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae)

JULIAN R. STARR,\*† ROBERT F. C. NACZI‡ and BRIANNA N. CHOUINARD\*†

\*Canadian Museum of Nature, PO Box 3443, Station D, Ottawa, ON, Canada K1P 6P4, †Department of Biology, Gendron Hall, Room 160, 30 Marie Curie, University of Ottawa, Ottawa, ON, Canada K1N 6N5, ‡The New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10458-5126, USA

## Abstract

We investigate the species discriminatory power of a subset of the proposed plant barcoding loci (*matK*, *rbcl*, *rpoC1*, *rpoB*, *trnH-psbA*) in *Carex*, a cosmopolitan genus that represents one of the three largest plant genera on earth (c. 2000 species). To assess the ability of barcoding loci to resolve *Carex* species, we focused our sampling on three of the taxonomically best-known groups in the genus, sections *Deweyanae* (6/8 species sampled), *Griseae* (18/21 species sampled), and *Phyllostachyae* (10/10 species sampled). Each group represents one of three major phylogenetic lineages previously identified in *Carex* and its tribe Cariceae, thus permitting us to evaluate the potential of DNA barcodes to broadly identify species across the tribe and to differentiate closely related sister species. Unlike some previous studies that have suggested that plant barcoding could achieve species identification rates around 90%, our results suggest that no single locus or multilocus barcode examined will resolve much greater than 60% of *Carex* species. In fact, no multilocus combination can significantly increase the resolution and statistical support (i.e.,  $\geq 70\%$  bootstrap) for species than *matK* alone, even combinations involving the second most variable region, *trnH-psbA*. Results suggest that a *matK* barcode could help with species discovery as 47% of *Carex* taxa recently named or resolved within cryptic complexes in the past 25 years also formed unique species clusters in UPGMA trees. Comparisons between the nrDNA internal transcribed spacer region (ITS) and *matK* in sect. *Phyllostachyae* suggest that *matK* not only discriminates more species (50–60% vs. 25%), but it provides more resolved phylogenies than ITS. Given the low levels of species resolution in *rpoC1* and *rpoB* (0–13%), and difficulties with polymerase chain reaction amplification and DNA sequencing in *rbcl* and *trnH-psbA* (alignment included), we strongly advocate that *matK* should be part of a universal plant barcoding system. Although identification rates in this study are low, they can be significantly improved by a regional approach to barcoding.

**Keywords:** *Carex*, Cariceae, DNA barcoding, phylogeny, plants, regional barcoding, species discovery

Received 17 November 2008; revision received 17 January 2009; accepted 30 January 2009

## Introduction

*Carex* L. (Cariceae Pax; Cyperaceae Juss.) is an enormous genus (c. 2000 species) equalled in species diversity only by *Euphorbia* L. and *Piper* L. (Frodin 2004). It is also of global importance as one of the few truly cosmopolitan plant genera (Good 1974) with centres of diversity in the temperate regions of Asia, Europe and the Americas. Throughout its range, *Carex* is found in a multitude of habitats ranging from

deserts to rain forests, and in some regions, such the Arctic, it is one of the dominant components in terms of species diversity and biomass (Scott 1995; Aiken *et al.* 2007). Moreover, its species often demonstrate a high degree of habitat specificity, making them some of the best indicator plants for characterizing habitat types (e.g. Magee & Rorer 1981; Klinka *et al.* 1989; Anderson *et al.* 1996; Ringius *et al.* 1997; Vellend *et al.* 2000; Karlsen & Elvebakk 2003; Gignac *et al.* 2004; Dabros & Waterway 2008). Although *Carex* lacks crop species, it is of indirect economic importance for its weeds. Approximately 18% of the estimated 449 Cyperaceae species explicitly cited as weedy (Bryson & Carter 2008) are found

Correspondence: Julian R. Starr, Fax: 613-364-4027; E-mail: jstarr@mus-nature.ca

within *Carex*, some of which have spread at an astonishing rate (Reznicek & Catling 1987).

Clearly, the biological diversity, ecological significance and economic impact of *Carex* is global in nature, and yet its species are often ignored by scientists and the general public alike. This is largely due to the genus' complex taxonomy and reduced morphology, which can make species identification difficult, particularly in temperate areas where local *Carex* floras can approach or exceed 100 taxa (e.g. City of Ottawa, Ontario, Canada, Brunton 2005; Tompkins Co., New York, USA, Weldy & Werier 2008). In *Carex* species, reliable morphological identification is complicated by the fact that closely related taxa often differ by only a single qualitative character or by the accumulation of many small quantitative differences (Naczi 1992; Standley *et al.* 2002). In addition, identification typically requires reproductively mature material that may be available for only a short period of the year (e.g. Naczi & Bryson 2002) as well as a good understanding of the genus' morphology and terminology in order to navigate through lengthy taxonomic keys (e.g. the key to Canadian *Carex* is 39 pages in Scoggan 1978). This difficulty with identification has many unfortunate consequences: some studies purposely ignore sedge diversity (e.g. Dornbush 2004); invasive species go unnoticed or are mistaken as natives (e.g. Catling & Kostiuik 2003; Janeway 2005); floras are often incomplete or erroneous (e.g. Zika & Kuykendall 2001); and taxonomic confusion related to misidentification is common (e.g. Wheeler 2007; Molina *et al.* 2008). Unfortunately, this problem is only getting worse as numerous, often cryptic species continue to be discovered. Over the past 20 years in North America alone, more than two new *Carex* species per year have been discovered on average with the majority of them in one of the best studied floras in the world, eastern North America (Hartman & Nelson 1998; Ertter 2000; Ford *et al.* 2008). Any methodology that can facilitate the correct identification of *Carex* species and aid in new discoveries, particularly as threats from habitat loss, invasive species and climate change increase globally, would constitute a significant contribution to science and society. Owing to its great diversity, difficult morphology and high potential for still undiscovered species, *Carex* is a prime candidate to develop DNA barcodes.

The promise of DNA barcoding is that it will provide a quick, simple and economic tool for identifying and discovering biological diversity. Ideally, a DNA barcode would require only small amounts of possibly poor quality tissue to be easily amplified and sequenced using universal primers. In addition, its sequences would provide a high level of confidence that the name provided by the barcode database could be treated as an actual 'species identification', and if sufficiently different from any other member in the database, it would indicate that a new species had been discovered (Savolainen *et al.* 2005; Kress & Erickson 2008).

In this study, we evaluate a subset (*matK*, *rbcL*, *rpoC1*, *rpoB*, *trnH-psbA*) of the seven proposed plant barcodes (Pennisi 2007) for their technical practicality (i.e. ease of amplification and sequencing) and potential to identify and discover *Carex* and Cariceae species. We focused our efforts on three of the taxonomically best-known groups in the genus and tribe, *Carex* sections *Deweyanae* (Tuckerm. ex Mack.) Mack., *Griseae* (L. H. Bailey) Kük., and *Phyllostachyae* Tuckerm. ex Kük. Species in these sections are clearly circumscribed by multiple lines of morphological (e.g. Saarela & Ford 2001; Crins *et al.* 2002; Naczi & Bryson 2002; Naczi 2002), anatomical (Starr & Ford 2001), cytological (Naczi 1999), and molecular evidence in the case of sect. *Phyllostachyae* (Ford *et al.* 1998a, 1998b, 1998c; Starr *et al.* 1999; Ford & Naczi 2001). Each section also represents one of three major lineages identified in tribal phylogenies (Yen & Olmstead 2000; Roalson *et al.* 2001; Starr *et al.* 2004, 2008; Waterway & Starr 2007; Starr & Ford 2009), permitting us to evaluate the potential of barcodes to broadly identify species across tribe Cariceae and to differentiate sister species, the most difficult challenge for any barcoding system. In addition, these groups contain species discovered (e.g. Naczi 1993; Ford & Naczi 2001; Saarela & Ford 2001; Naczi *et al.* 2002) and/or species complexes resolved (e.g. Naczi *et al.* 1998; Ford *et al.* 1998a) within the last 25 years, which allowed us to evaluate how successful barcodes might be for discovering new species and for helping to resolve difficult taxonomic problems within the tribe. Since the barcoding regions examined have not previously been used in *Carex* or Cariceae phylogenetics, we also assessed their potential for clarifying relationships within these groups.

## Methods and materials

### *Taxonomic sampling*

To determine which of a subset of the proposed barcoding loci are necessary to resolve *Carex* species, we sampled one to four individuals per taxon for sections *Deweyanae* (six taxa analysed from a total of nine in the section), *Griseae* (18 taxa analysed from a total of 21 in the section), and *Phyllostachyae* (10 taxa analysed from a total of 10 in the section) (Appendix S1, Supporting information). Most taxa (31 of 34) were represented by multiple individuals and the majority of these (20 of 31) consisted of three or more samples. Although morphological and molecular evidence suggest that *Carex laeoviculmis* is not a part of sect. *Deweyanae*, it was treated as such for the purposes of this study (see Naczi 2002; Ford *et al.* 2006; Naczi 2009).

### *DNA isolation and amplification*

All regions were polymerase chain reaction (PCR) amplified from total genomic DNA isolated from herbarium specimens

using a modified silica-column-based method (Alexander *et al.* 2007) where the guanidine hydrochloride in the binding buffer was increased from 2 M to 7 M and the end percentage of ETOH was 16.7%. Primer sequences for the coding regions *matK* (*matK*-2.1f, *matK*-5 r), *rpoB* (*rpoB*-2f, *rpoB*-4 r), and *rpoC1* (*rpoC1*-1f, *rpoC1*-4 r) were obtained from the phase 2 protocols available on the Royal Botanic Gardens' (Kew) barcoding website ([www.kew.org/barcoding/protocols.html](http://www.kew.org/barcoding/protocols.html)). A portion of the chloroplast gene *rbcL* was initially amplified using primers Z1 (Arnold *et al.* 1991) and Z-1240R (Herre *et al.* 1996). After problems with sequencing were encountered with these primers, the *rbcL* primer set used by Kress & Erickson (2007) (*rbcL*-a\_f, Tsukaya *et al.* 1997; *rbcL*-a\_r) was also tried. The noncoding *trnH-psbA* region was amplified as in Kress *et al.* (2005) using primers *psbA3'f* (Sang *et al.* 1997) and *trnHf* (Tate & Simpson 2003). Primers were not re-designed when problems with PCR or sequencing were encountered since the purpose of this study was to evaluate both the ability of a barcode to resolve species as well as how easily it can be amplified and sequenced. Each PCR amplification contained the following reactants dissolved in an end volume of 15  $\mu$ L: 1 $\times$  PCR buffer (Sigma, P2317), 0.2 mM of each dNTP, 0.25  $\mu$ M of each primer, 2.5 mM MgCl<sub>2</sub>, 1.33 mg BSA, 10–50 ng of template DNA, and 2.5 U of JumpStart *Taq* DNA Polymerase (Sigma, D4184). PCR products were amplified on a Bio-Rad DNA engine (PTC-200) gradient cyler via 30 cycles of DNA denaturation at 95 °C for 45 s, primer annealing at 45 °C for 45 s, and DNA strand extension at 72 °C for 90 s, with a pre-treatment at 95 °C for 60 s before cycling. The PCR was terminated by a final extension step of 72 °C for 3 min. Minor adjustments to the above PCR protocols were sometimes necessary depending on DNA quality and the primer pair used for amplification. A sample of each reaction was run on 2.0% agarose gels stained with ethidium bromide and successful products were sent to the Canadian Centre for DNA Barcoding at the University of Guelph where they were sequenced in both directions using PCR primers according to the Centre's protocols ([www.dnabarcoding.ca/pa/ge/research/protocols/sequencing](http://www.dnabarcoding.ca/pa/ge/research/protocols/sequencing)).

### Analysis

Bidirectional sequences for each locus were assembled and edited in Sequencher 4.7 (Genecodes). Manual alignment was used for coding loci due to a lack of size differences among taxa; however, alignment for the noncoding *trnH-psbA* region required the use of Clustal\_X (default settings; Thompson *et al.* 1997). To assess the ability of each locus and multilocus combination to resolve species, UPGMA (unweighted pair group method with arithmetic mean) dendrograms based on Kimura 2-parameter distances (K2P, recommended by CBOL, [www.barcoding.si.edu](http://www.barcoding.si.edu)) were constructed in PAUP\* 4.10b (default settings; Swofford 2003). Only UPGMA

analyses were performed as they have consistently resolved the greatest number of species in studies that have used multiple tree building methods (e.g. Lahaye *et al.* 2008a, b). A positive identification by a barcoding locus or combination of loci was counted only when all individuals sampled for a species (i.e.  $\geq 2$ ; includes subspecific taxa) formed a single cluster in trees. Following Fazekas *et al.* (2008), the number of species resolved with a bootstrap (BS) value of  $\geq 70\%$  (heuristic searches, 10 000 random additions with the MULTREES option off; DeBry & Olmstead 2000) was counted and reported as a percentage of the total number of species resolvable to assess support for successful resolution of a species cluster. As numerous difficulties with the amplification and sequencing of the *trnH-psbA* region were encountered for taxa within sections *Deweyanae* and *Phyllostachyae*, the utility of this region for barcoding purposes was only assessed within sect. *Griseae*. The barcoding potential of the *rbcL* gene was not evaluated as both primer sets assayed did not provide a sufficient number of readable sequences.

The potential for barcoding loci to help discover new species was assessed by counting the number of species resolved by the best barcoding locus or multilocus combination for new taxa named in the past 25 years. This represents nine taxa in sect. *Griseae* (*Carex acidicola*, *C. brysonii*, *C. calcifugens*, *C. godfreyi*, *C. ouachitana*, *C. paeninsulae*, *C. pigra*, *C. planispicata*, *C. thornei*; Kral *et al.* 1987; Naczi 1993, 1997, 1999; Naczi *et al.* 2002), three in sect. *Phyllostachyae* (*C. cordillerana*, *C. juniperorum*, *C. timida*; Catling *et al.* 1993; Ford & Naczi 2001; Saarela & Ford 2001), and two in sect. *Deweyanae* (*C. bromoides* ssp. *montana*, *C. infirminevia*; Naczi 1990; Naczi *et al.* 2002). Note that all of these taxa, except *C. brysonii* and *C. bromoides* ssp. *montana*, were represented by two or more individuals. In addition, the ability of the best barcoding locus or multilocus combination for helping to resolve former taxonomic problems was assessed by the level of resolution achieved in the *Carex willdenowii* complex (sect. *Phyllostachyae*), a cryptic group of three species (*C. basiantha*, *C. superata*, *C. willdenowii*) that are separable on the basis of morphological, anatomical, micromorphological and molecular data (Ford *et al.* 1998a; Naczi *et al.* 1998; Starr *et al.* 1999; Starr & Ford 2001).

The utility of coding regions for phylogenetic purposes was also explored in PAUP\* by comparing basic character statistics among loci and the range of sequence divergence across taxa (K2P). The same was not carried out for the noncoding *trnH-psbA* region as it could not be widely amplified, sequenced, or aligned for the sections examined. The utility of *matK* for reconstructing lower-level relationships in *Carex* was assessed by comparing levels of clade support and tree resolution in phylogenies constructed using *matK* vs. the internal transcribed spacer (ITS) region of nuclear ribosomal DNA (nrDNA) in sect. *Phyllostachyae* (Starr *et al.* 1999). Strict consensus trees were assembled from heuristic parsimony searches using a random addition sequence of taxa (10 000 repetitions) with the MULTREES option on.

**Table 1** PCR and sequencing success for the samples used in this study

	<i>matK</i>	<i>rpoB</i>	<i>rpoC1</i>	<i>trnH-psbA</i>	<i>rbcL</i>
PCR – samples tried	93	93	93	93	93
PCR – strong amplification	72 (77%)	60 (65%)	84 (90%)	46 (49%)	60 (65%)
PCR – weak amplification	19 (21%)	33 (35%)	9 (10%)	45 (49%)	31 (33%)
PCR – double banding	0 (0%)	0 (0%)	0 (0%)	16 (17%)	5 (5%)
PCR – failure	2 (2%)	0 (0%)	0 (0%)	2 (2%)	2 (2%)
Sequencing – samples tried	91	93	93	75	86
Sequencing – bidirectional (complete)	86 (95%)	93 (100%)	48 (52%)	54 (72%)	0 (0%)
Sequencing – bidirectional (incomplete)	3 (3%)	0 (0%)	45 (48%)	15 (20%)	15 (18%)
Sequencing – single direction	0 (0%)	0 (0%)	28 (30%)	7 (9%)	0 (0%)
Sequencing – reaction failure	2 (2%)	0 (0%)	0 (0%)	7 (9%)	71 (82%)

**Table 2** Barcoding utility of coding sequences in *Carex*. Only individuals with data for the single locus or multiple loci examined were included in analyses

	Species resolvable ( $\geq 2$ individuals)	Species resolved as single groups	Species resolved with $\geq 70\%$ BS
<i>matK</i>	30	17 (57%)	10 (33%)
<i>rpoB</i>	32	3 (9%)	1 (3%)
<i>rpoC1</i>	33	3 (9%)	1 (3%)
<i>matK + rpoB</i>	29	17 (59%)	11 (38%)
<i>matK + rpoC1</i>	30	18 (60%)	11 (37%)
<i>rpoB + rpoC1</i>	30	5 (17%)	3 (10%)
<i>matK + rpoB</i> + <i>rpoC1</i>	29	17 (59%)	11 (38%)

UPGMA trees were also constructed to see whether more species clusters would be resolved for each data set by this method.

## Results

### PCR amplification and DNA sequencing

Of the 93 individuals included in this study, strong or weak PCR amplification was achievable for all loci, although a small percentage of individuals (2%) could not be amplified for either the *trnH-psbA*, *rbcL* (both primer sets) or *matK* regions (Table 1). Double banding was observed in amplifications for the *trnH-psbA* (17%) and *rbcL* (5%) regions (Table 1). Success rates for bidirectional sequences were highest for *rpoB* (100%) and *matK* (95%), followed by *trnH-psbA* (72%) and *rpoC1* (52%). Single direction sequence reads were obtained for 30% of *rpoC1* and 9% of *trnH-psbA* samples. Sequences were unreadable for approximately 9% of *trnH-psbA* and 2% of *matK* sequences. All efforts to sequence *rbcL* in both directions failed, with 82% of sequences being entirely unreadable. Readable sequence could not be obtained from

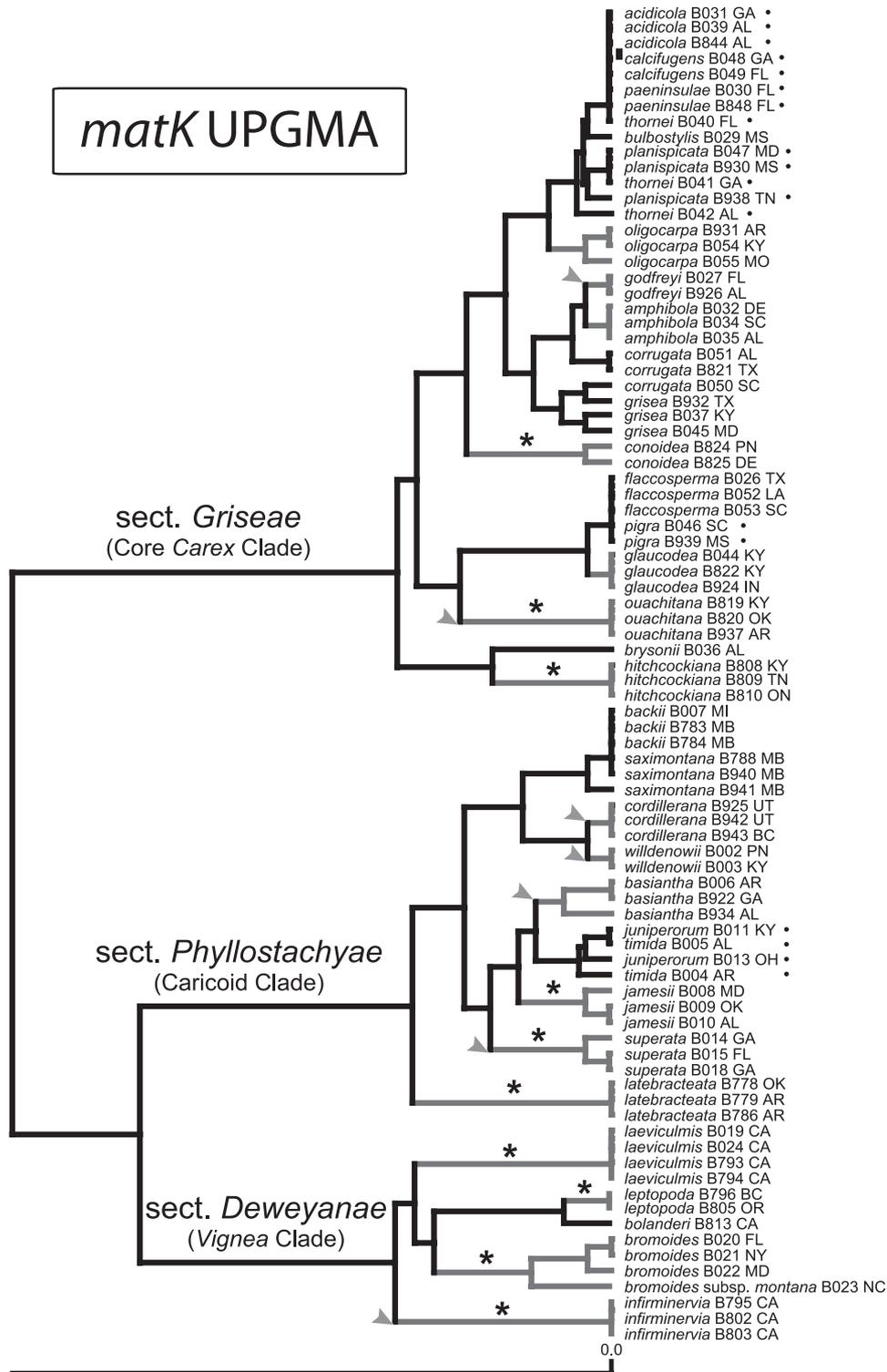
the alternative primer set *rbcL-a\_f* and *rbcL-a\_r*. Although *rpoC1* could not be reliably sequenced in both directions, clear sequence was obtained for all taxa sampled in at least one direction. In this case, the reverse primer (*rpoC1-4 r*) appears to double prime during sequencing.

### Alignment and barcoding

Sequences for coding regions could be aligned easily by eye as no insertion/deletion (indel) events were detected. However, the *trnH-psbA* intergenic spacer showed considerable size variation across the taxa examined (600 to 852 bp). Efforts to align this locus using Clustal\_X could not be easily achieved across all taxa. The *rps19* gene previously detected by Wang *et al.* (2008) in the Cyperaceae genus *Scirpus* L. and across most monocots, is also present in the spacer region between the *trnH* and *psbA* genes of *Carex*. Alignments of *Carex trnH-psbA* sequences with the *rps19* gene of *Scirpus ternatanus* Reinw. ex Miq. (AB331264) revealed the absence of five codons in all *Carex* near the 5' end of the gene (codons 20 to 24) as well as the lack of 27 codons including the stop signal on the 3' end of *rps19* in species of sect. *Griseae* only.

The utility of the coding sequences examined for barcoding, alone and in multigene combinations, is presented in Table 2. Results indicate that *matK* resolves a much higher percentage of species (57%) than either *rpoC1* or *rpoB* (9% each), and it provides more statistical support for taxa  $\geq 70\%$  BS (33%) than either of these loci (3% each). Multigene combinations only marginally resolve a greater percentage of taxa and provide greater support over *matK* alone (Table 2), with the exception of *rpoB + rpoC1* which provides considerably poorer resolution (17%) and support (10%) for the taxa examined. A UPGMA dendrogram using *matK* data is presented in Fig. 1.

In comparisons of the noncoding *trnH-psbA* region vs. all coding sequences and locus combinations in sect. *Griseae*, *matK* provided greater levels of species resolution (53% vs. 44%) and species statistical support (27% vs. 19%) than *trnH-psbA* alone (Table 3). The genes *rpoB* and *rpoC1* provided



**Fig. 1** UPGMA dendrogram of *Carex* taxa using 815 bp of the 3' end of *matK*. Specific epithets are followed by DNA numbers and standard abbreviations for the state or province in which a sample was collected (see Appendix S1). Branches in grey represent species for which all individuals sampled ( $\geq 2$ ) formed a single species cluster. Single species clusters receiving  $\geq 70\%$  BS support are distinguished by asterisks above branches. Arrow heads represent single species clusters ( $\geq 2$ ) named within the past 25 years, including the three species recently resolved within the *Carex willdenowii* complex in sect. *Phyllostachyae*. Individuals from taxa named in the last 25 years that did not group into single species clusters are indicated by black dots to the right of state and province abbreviations. Since the newly discovered *C. brysonii* and *C. bromoides* ssp. *montana* were represented by only a single sample each, they are not distinguished by arrow heads or dots in the figure.

**Table 3** Barcoding utility of the noncoding *trnH-psbA* region versus coding sequences in *Carex* section *Griseae*. Only individuals with data for the single locus or multiple loci examined were included in analyses

	Individuals ( <i>n</i> )	Aligned length	Variable characters	Informative characters	Species resolvable (≥ 2 individuals)	Species resolved as unique clusters	Species resolved with ≥ 70% BS
<i>trnH-psbA</i>	49	600	32	21	16	7 (44%)	3 (19%)
<i>matK</i>	42	815	32	27	15	8 (53%)	4 (27%)
<i>rpoB</i>	42	508	3	2	15	0 (0%)	0 (0%)
<i>rpoC1</i>	42	584	10	7	15	2 (13%)	1 (7%)
<i>trnH-psbA</i> + <i>matK</i>	41	1415	62	39	14	7 (50%)	4 (29%)
<i>trnH-psbA</i> + <i>rpoB</i>	41	1108	33	18	16	5 (31%)	3 (19%)
<i>trnH-psbA</i> + <i>rpoC1</i>	41	1184	39	22	14	5 (36%)	3 (21%)
<i>matK</i> + <i>rpoB</i>	42	1323	35	29	15	7 (47%)	4 (27%)
<i>matK</i> + <i>rpoC1</i>	42	1399	42	34	15	7 (47%)	4 (27%)
<i>rpoB</i> + <i>rpoC1</i>	42	1092	13	9	15	2 (13%)	1 (7%)
<i>trnH-psbA</i> + <i>matK</i> + <i>rpoB</i>	41	1923	65	41	14	7 (50%)	4 (29%)
<i>trnH-psbA</i> + <i>rpoB</i> + <i>rpoC1</i>	41	1692	42	24	14	5 (36%)	3 (21%)
<i>trnH-psbA</i> + <i>matK</i> + <i>rpoC1</i>	41	1999	71	45	14	8 (57%)	5 (36%)
<i>trnH-psbA</i> + <i>matK</i> + <i>rpoB</i> + <i>rpoC1</i>	41	2507	74	47	14	8 (57%)	5 (36%)

very low resolution (0% vs. 13%) and support (0% vs. 7%) for species groups (Table 3). Only *trnH-psbA* + *matK* + *rpoC1* and *trnH-psbA* + *matK* + *rpoB* + *rpoC1* resolved more species and provided better statistical support for taxa than *matK* alone (Table 3), although these increases were small (53% vs. 57% species resolution, 27% vs. 36% of species with ≥ 70% BS).

#### Utility of barcodes for resolving recently named taxa

For recently described taxa (last 25 years), *matK* resolved two (*C. godfreyi*, *C. ouachitana*) of the eight species in sect. *Griseae* for which multiple individuals were available (Fig. 1). The ninth species, *Carex brysonii*, was represented by a single individual isolated on a long branch (Fig. 1). For sect. *Deweyanae*, individuals of *C. infirminevia* were clearly separated from all other species. *Carex bromoides* ssp. *montana*, which was represented by a single sample, clearly separated from the three individuals sequenced for *C. bromoides* ssp. *bromoides*. In sect. *Phyllostachyae*, only one (*C. cordillerana*) of the three species recently named came out as a single cluster (Fig. 1). All members of the *C. willdenowii* complex (*C. basiantha*, *C. superata*, *C. willdenowii*) were resolved as unique clusters (Fig. 1). A pectinate clustering pattern was seen among individuals of *C. juniperorum* and *C. timida* (Fig. 1).

#### Phylogenetic utility of barcoding regions

*MatK* has 3.3 and 2.9 times more informative characters than *rpoB* and *rpoC1* respectively, and it is approximately 3.4 and 2.9 times more variable than *rpoB* and *rpoC1* for the taxa examined (Table 4). For comparisons between phylogenies constructed for sect. *Phyllostachyae* using the ITS and

**Table 4** Phylogenetic utility of barcoding genes

	<i>matK</i>	<i>rpoB</i>	<i>rpoC1</i>
Individuals ( <i>n</i> )	86	91	92
Aligned length	815	508	584
Sequence divergence	0.0–4.2%	0.0–3.0%	0.0–2.7%
Variable characters	82	25	30
Informative characters	72	22	25

**Table 5** Phylogenetic and barcoding utility of *matK* versus ITS at lower taxonomic levels using *Carex* section *Phyllostachyae* as an example

	<i>matK</i>	ITS
Individuals ( <i>n</i> )	27	22
Aligned length (bp)	815	464
Variable characters	23	30
Informative characters	17	25
Species resolvable (≥ 2 individuals)	10	8
Species resolved as single groups (parsimony — UPGMA)	5–6 (50–60%)	2–2 (25–25%)
Species resolved with ≥ 70% BS support	3 (30%)	1 (13%)

*matK* regions, *matK* resolved more taxa and provided more statistical support for clades than ITS regardless of the tree construction method used (Fig. 2; Table 5).

## Discussion

### Barcoding loci: PCR and sequencing success

Beyond simple sequence variability, crucial characteristics for barcoding loci include primer universality and easy

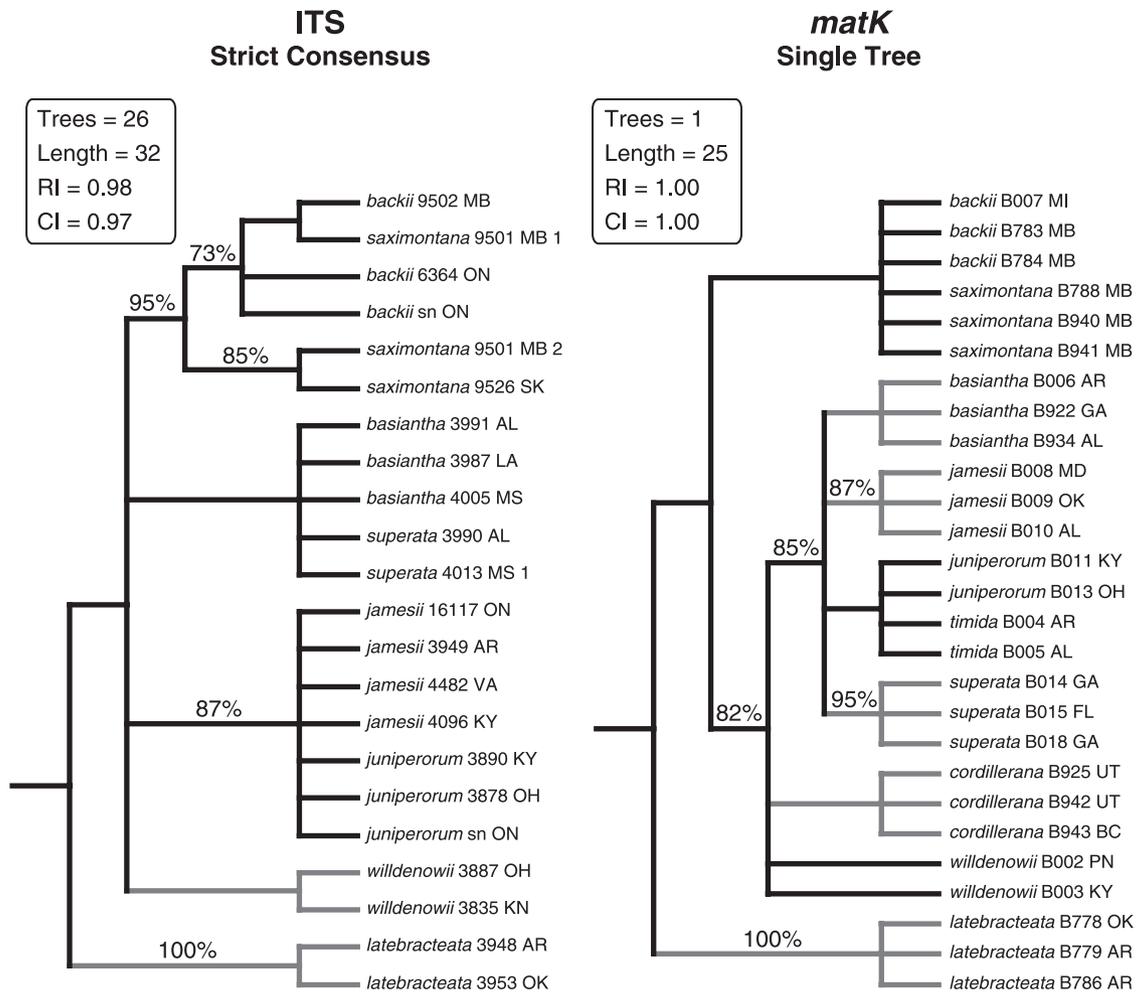


Fig. 2 Trees resulting from the parsimony analysis of nrDNA ITS (strict consensus) and cpDNA *matK* sequences in *Carex* section *Phyllostachyae*. Branches in grey represent monophyletic species. Bootstrap values are given above branches.

amplification and sequencing (Chase *et al.* 2007; Fazekas *et al.* 2008). In this study, all regions amplified except for a small percentage of individuals for the *trnH-psbA*, *rbcL* and *matK* regions. Excluding *matK*, these regions were also the most difficult to cleanly amplify and sequence. Although the *rbcL* region typically amplified well, it was prone to double banding, and only a handful of poorly readable sequences were obtained despite the use of two different primer sets. Normally the *rbcL* region is easy to amplify and sequence across a wide range of land plants, but like many other barcoding loci assayed (Fazekas *et al.* 2008), it still requires the use of multiple primer sets to achieve complete taxonomic coverage. If *rbcL* is chosen as a universal plant barcode, as some authors propose (Newmaster *et al.* 2006; Kress & Erickson 2007), primers other than those used here will be necessary to obtain data for *Carex*.

The *trnH-psbA* region was undoubtedly the most difficult to cleanly amplify and sequence. Highly variable in size and difficult to align across *Carex* clades, the *trnH-psbA* region

produced numerous double bands, was difficult to sequence in both directions and yielded a considerable number of sequences that were unreadable. Part of the difficulty associated with its sequencing and alignment can be explained by the presence of numerous homopolymer regions, a fact that also accounts for most of the spacer size differences observed amongst taxa as these and other repeat units varied in number. However, size variation in the *trnH-psbA* region was also due to differences in the length of *rps19*, a gene or pseudogene that has previously been seen in the Cyperaceae genus *Scirpus* as in many other monocots (Wang *et al.* 2008). Alignments of *Scirpus rps19* with *trnH-psbA* sequences from all three *Carex* sections suggest not only the absence of five codons in *Carex* near the 5' end of the gene, but also the lack of 27 possible codons including the stop codon at the 3' end of *rps19* in section *Griseae*, but not in sections *Deweyanae* and *Phyllostachyae*. Even though this region was among the most variable assayed, difficulties with amplification, sequencing and alignment suggest that the *trnH-psbA* region does not

provide a practical barcode for *Carex* or other land plants (Chase *et al.* 2007; Sass *et al.* 2007).

By comparison, the *matK* region was not only easy to amplify, sequence bidirectionally and align (no indels), but it was also more variable than the *trnH-psbA* region. Higher variability for the *matK* region has also been seen in other monocot genera (see Chase *et al.* 2007), a fact that likely accounts for the difficulties encountered with its amplification and sequencing (Fazekas *et al.* 2008). Only *rpoB* was more easily amplified and sequenced for all individuals, but due to low variability and poor species resolving power, we consider it largely impractical for barcoding purposes in *Carex*. The *rpoC1* region amplified well, but it was difficult to sequence due to double priming by the reverse primer *rpoC1-4 r*. This primer could be redesigned, but the region as a whole was among the least variable and poorest species discriminators of all those examined indicating that it is not an effective barcode. Similar results have also been seen in other studies (Sass *et al.* 2007; Lahaye *et al.* 2008a; Newmaster *et al.* 2008; Fazekas *et al.* 2008) suggesting that *rpoC1* may not be an effective barcode for land plants in general.

#### Barcoding loci: species resolution

As a single region barcode, *matK* resolves by far the greatest number of species and provides the highest statistical support for taxa of any marker (Fig. 1; Tables 2 and 3). In fact, no multigene combination greatly improves species resolution or statistical support over the use of *matK* alone (Tables 2 and 3), a result also seen in Lahaye *et al.* (2008a, b). This is even true when the noncoding region *trnH-psbA* is added to the analysis as only marginal gains in taxon resolution (53% vs. 57%) and support (27% vs. 36%) could be achieved over *matK* when *trnH-psbA* was part of a possible three (*trnH-psbA* + *matK* + *rpoC1*) or four locus barcode (*trnH-psbA* + *matK* + *rpoB* + *rpoC1*). Given the low variability of both *rpoB* and *rpoC1* (a trait common for these loci; Chase *et al.* 2007; Fazekas *et al.* 2008; Newmaster *et al.* 2008), and the numerous difficulties with amplification and sequencing in *rbcL* and *trnH-psbA* (alignment included), only *matK* appears to be a viable barcode for *Carex*. Even though we were not able to sequence *rbcL* in this study despite using two different primer sets, it is unlikely that *rbcL* would significantly increase species resolution as it is typically three times less variable than *matK* (Hilu *et al.* 2003). In other words, it would most likely provide the same level of variation as detected in *rpoB* or *rpoC1*, but it would require more sequencing effort since comparisons are based on complete or nearly complete gene sequences (~1300 bp long; Chase *et al.* 2005; Newmaster *et al.* 2006). No matter which single or multilocus comparisons are made, it is clear that *matK* is providing most of the resolving power and statistical support observed in this study.

Although the performance of any one barcoding locus is taxon dependent, most studies have shown that *matK* and

*trnH-psbA* are the most variable chloroplast markers and they consistently provide the best species resolution of proposed loci (e.g. Fazekas *et al.* 2008; Lahaye *et al.* 2008a; Newmaster *et al.* 2008). Fazekas *et al.* (2008) suggest that species resolution is not so dependent on the locus chosen as on number of loci used. However, if a reasonable number of loci (two to four) are used as a universal plant barcode, and the technical and analytical difficulties posed by noncoding regions such as *trnH-psbA* are to be avoided, the exclusion of *matK* from a multilocus barcode will greatly diminish its species discriminatory power in *Carex*, a genus whose species represent 0.5% to 0.9% of seed plant diversity (Govaerts 2001; Scotland & Wortley 2003). Based on our results, an *rpoB* + *rpoC1* or *trnH-psbA* + *rpoB* + *rpoC1* barcode would probably not resolve much more than 13% to 36% of *Carex* species, figures that are unlikely to increase with the addition of *rbcL*. Consequently, as with Lahaye *et al.* (2008a, b), we advocate that *matK* be chosen as part of a universal plant barcode from amongst the seven loci recently proposed (Pennisi 2007) despite the current technical problems associated with its amplification across land plants (Sass *et al.* 2007; Fazekas *et al.* 2008). A single universal primer set for *matK* may not exist, just as there is no single set of primers for *rbcL* or the animal barcode *cox1*; nevertheless, the practicality of a *matK* barcode will be diminished if it cannot be amplified and sequenced with considerably fewer than the ten primer pairs used by Fazekas *et al.* (2008).

Even if *matK* is included in a universal barcode, our results suggest that only 60% of *Carex* species would be resolved by the loci examined, and this would fall to only 38% if the criterion for successful species resolution used by Fazekas *et al.* (2008) is applied (i.e.  $\geq 70\%$  BS; Fig. 1; Tables 2 and 3). Unfortunately, these are probably high-end estimates, as the sections examined here are relatively small and taxonomically well known. For example, the greatest species resolution was achieved in sections *Deweyanae* and *Phyllostachyae* (60% to 100%), groups consisting of eight and 10 species respectively, with species-level boundaries in sect. *Phyllostachyae* having already been tested by multiple lines of morphological and molecular data (evolutionary significant units *sensu* Meyer & Paulay 2005). By comparison, sect. *Griseae* is considerably larger (18 of 21 taxa sampled) and its species have not been tested with genetic tools. In its case, only 57% of taxa clustered into single groups, even when four regions were combined. Bearing in mind that our species sampling represents the greatest for any plant genus thus far and that our sampling represents both broad lineages and numerous close species pairs within sections (the most difficult challenge for barcoding; Hollingsworth 2008; Newmaster *et al.* 2008), we consider it unlikely that the assayed barcodes will achieve a species resolution rate significantly greater than 60% in *Carex* and Cariceae given our current taxonomic understanding of these groups. Although future taxonomic changes may increase this figure,

we expect initial species resolution rates to be worse due to several large and taxonomically difficult groups within the tribe. For example, the inclusion of members of *Carex* section *Ovales* Kunth in our analyses (85 species; Mastrogiuseppe *et al.* 2002 FNA), a large sedge clade that is part of a deeper lineage that appears to have undergone rapid speciation [i.e. *Carex* subg. *Vignea* (P. Beauv. ex Lestib. f.) Perterm.; Ford *et al.* 2006], would have reduced the assignment success of the barcodes examined (B.N. Chouinard *et al.*, unpublished data).

#### *Regional barcoding: a solution to poor species resolution?*

Several previous studies have suggested that barcoding in plants could attain taxon assignment levels approximating 90% or greater (e.g. Kress & Erickson 2007; Lahaye *et al.* 2008a, b; Newmaster *et al.* 2008), although these studies have typically included phylogenetically distant taxa where infrageneric sampling was limited and/or regional in nature (Lahaye *et al.* 2008a, b; Newmaster *et al.* 2008). In contrast, Fazekas *et al.* (2008) sampled two to seven species for 32 genera across land plants and concluded that only 69–71% of taxa could be resolved with statistical confidence using a barcoding system based on a moderate number of plastid markers. However, Fazekas *et al.* (2008) considered these figures as an upper-end estimate of the maximum resolution possible since they considered their sampling of close relatives as sparse. Our results are even less encouraging (i.e. a maximum of 60% of species resolved, 38% with statistical confidence) and are probably a reflection of our greater infrageneric and especially infrasectional sampling (see Fazekas *et al.* 2008). This sampling strategy meant that more sister species were present in our analysis than in previous plant barcoding studies.

Figures suggesting that around 60% of *Carex* species would be resolvable using a barcoding system based on the loci examined are discouraging, although successful taxon assignment rates could be increased if a sample's regional origin was known. For example, a multilocus approach to identifying the *Carex* and *Kobresia* Willd. (Cariceae, Cyperaceae) of the Canadian Arctic Archipelago (J. LeClerc-Blain *et al.*, in press) demonstrates that a two-locus barcode involving *matK* can achieve a 100% success rate. Moreover, preliminary data from the development of a *matK* database for all North American *Carex*, north of Mexico (c. 559 taxa), suggests that *matK* can correctly assign sequences to species > 60% to > 90% of the time, depending on which region of Canada is analysed (i.e. Eastern, Central, Western and Northern regions) (B.N. Chouinard *et al.*, unpublished data). In the case of the taxa examined here, which are all woodland groups and whose highest common diversity occurs in Alabama (19 of the 34 taxa examined), a *matK* barcoding survey of Madison and Winston Counties where their syntopy is highest (R.F.C. Naczi, unpublished data),

would achieve successful identification in 88% (seven of eight) and 86% (six of seven) of cases. Although assignment success is often highest where diversity is lowest (e.g. the Arctic; Taberlet *et al.* 2007; LeClerc-Blain *et al.*, in press) and vice versa (B.N. Chouinard *et al.*, unpublished data), such a system could increase the level of certainty in species assignments to a point where it may be deemed practical in applications involving nonspecialists (i.e. nonsystematists). Even if a regional approach could increase species assignment rates to an acceptable level, it would still adversely affect the utility of a barcoding system to nonspecialists since a specimen whose origins were entirely unknown could not be identified with confidence. This does not mean that an imperfect barcode would not be useful to nonspecialists as it could still significantly narrow the number of possible taxa, but for applications where a precise identification was required, further taxonomic expertise would remain indispensable.

#### *Discovering plant species and evolution by DNA barcodes*

Although correct taxon assignment was poor overall, *matK* did show promise as a means to help with the discovery of biodiversity. Unlike previous studies that have suggested that plant barcodes may have already discovered unrecognized diversity (Pennisi 2007; Lahaye *et al.* 2008a), our approach was to examine whether newly named taxa (< 25 years) could be resolved using barcodes, including taxa from a recently resolved cryptic group, the *Carex willdenowii* complex. Of the 15 species this represents with two or more individuals sampled in our analysis, 47% were resolved in *matK* dendrograms. Moreover, if we assume that future samples of *Carex brysonii* will cluster with the single individual sequenced here, as suggested by its long branch in UPGMA trees, *matK* was able to resolve 50% of these taxa. In the case of *Carex timida* and *C. juniperorum*, failure to resolve these newly discovered species could possibly be due to parphyly, hybridization or incomplete lineage sorting. Interestingly, the branching of individuals of *C. timida* and *C. juniperorum* in our UPGMA tree largely reflects the population relationships detected in the isozyme analyses of Ford & Naczi (2001), a pattern they attributed to parphyly which may be common in plants (Rieseberg & Brouillet 1994). Admittedly, a new species barcode discovery rate of 47% to 50% of potential taxa is not incredibly high and the case of *C. timida* and *C. juniperorum* may be isolated, but they do suggest that once a complete *matK* database is available for the North American carices, as is currently underway (Chouinard *et al.* 2008), such a resource could be used to help test species hypotheses, serendipitously discover taxa, or even point to evolutionary phenomena that need further investigation when conducting taxonomic studies.

### Phylogenetic utility of coding regions — comparisons with ITS

As previous studies have found, *matK* is by far the most variable (0.0–4.2%) coding region investigated providing approximately three times as many variable and informative characters than either *rpoB* or *rpoC1*. Although the low levels of sequence variation in the latter two genes (0.0–3.0%, *rpoB*; 0.0–2.7%, *rpoC1*) are not as low as has been seen in some groups (e.g. Myricaceae; Newmaster *et al.* 2008), the resolution they provide is poor, suggesting they are not only poor barcoding regions for *Carex*, but that large scale phylogeny projects would best place their efforts in sequencing other potential chloroplast regions.

Given that the sections we investigated each represent one of the three major clades previously detected in tribal phylogenetic analyses (Starr & Ford 2009), it is not surprising that all of the genes investigated grouped sections into separate clades. In order to investigate the utility of *matK* at lower taxonomic levels, we compared parsimony analyses of species in *Carex* sect. *Phyllostachyae* with a sectional phylogeny constructed using ITS sequences from a previous study (Starr *et al.* 1999). This region is the most widely used marker for lower-level phylogenies in plants (Alvarez & Wendel 2003; Starr *et al.* 2003) and it is generally the most variable region that can be easily amplified and sequenced for nearly any living group. Because of its high variability levels and ease of amplification, the ITS region has been proposed as a barcode for plants (Kress *et al.* 2005) and it is recommended as a barcode for fungi (Seifert 2008). However, despite the desirable trait of being highly variable, several characteristics of the region make it a problematic barcode such as the presence of many paralogues, and in some cases, multiple functional copies (Chase *et al.* 2007; King & Roalson 2008). Surprisingly, at least for section *Phyllostachyae*, *matK* is not only a better phylogenetic marker than ITS, it is also a better barcode. *MatK* may produce fewer variable (40 vs. 55) and informative (17 vs. 25) characters than ITS, but it provides better resolution and support for clades as well as resolving two and a half to three times more species than ITS, and three times as many with  $\geq 70\%$  BS support. This result probably reflects the fact that the ITS region is not simply more variable than *matK*, its variable sites themselves are possibly evolving at a rate so rapid that it creates homoplasy even at the infrasectional level. The low consistency indices seen in tribal ITS phylogenies support such a conclusion as does the common finding of incongruence in comparisons of nrDNA and cpDNA partitions (Starr & Ford 2009) in Cariceae analyses among other reasons (King & Roalson 2008). Our results thus suggest that *matK* could make a significant contribution to resolving phylogeny at multiple taxonomic levels from generic relationships within Cariceae to infrasectional relationships within *Carex*.

### Acknowledgements

The authors thank Bruce Ford (University of Manitoba) for his ongoing assistance in research on *Carex* and Cariceae systematics, and Roger Bull, Jessica LeClerc-Blain, and Jeffery Saarela from the Canadian Museum of Nature for help with this and ongoing barcoding projects. We would also like to thank Janet Topan (Canadian Centre for DNA Barcoding, University of Guelph) for help with DNA sequencing. We are grateful to the curators of the following herbaria for the use of their specimens: DOV, WIN. This research was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to JRS.

### Conflict of interest statement

The authors have no conflict of interest to declare and note that the funders of this research had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

- Aiken SG, Dallwitz MJ, Consaul LL *et al.* (2007) *Flora of the Canadian Arctic Archipelago: Descriptions, Illustrations, Identification, and Information Retrieval*. NRC Research Press, Ottawa [CD-ROM].
- Alexander PJ, Rajanikanth G, Bacon CD, Bailey CD (2007) Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes*, **7**, 5–9.
- Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.
- Anderson DS, Davis RB, Rooney SC, Campbell CS (1996) The ecology of sedges (Cyperaceae) in Maine peatlands. *Bulletin of the Torrey Botanical Club*, **123**, 100–110.
- Arnold ML, Buckner CM, Robinson JJ (1991) Pollen-mediated introgression and hybrid speciation in Louisiana irises. *Proceedings of the National Academy of Sciences, USA*, **88**, 1398–1402.
- Brunton DF (2005) *Vascular Plants of the City of Ottawa, with Identification of Significant Species, Appendix A of Muncaster Environmental Planning and Brunton Consulting Services. 2005. Urban Natural Areas Environmental Evaluation Study, Final Report*. Planning and Growth Management Department, City of Ottawa, Ottawa, Canada.
- Bryson CT, Carter R (2008) The significance of Cyperaceae as weeds. In: *Sedges: Uses, Diversity, and Systematics of the Cyperaceae* (eds Naczi RFC, Ford BA), pp. 15–101. Monographs in Systematic Botany from the Missouri Botanical Garden, St. Louis, Missouri.
- Catling PM, Kostjuk B (2003) *Carex acutiformis* — dominance of a cryptic invasive sedge at Ottawa. *Lingua Botanica*, **4**, 12–15.
- Catling PM, Reznicek AA, Crins WJ (1993) *Carex juniperorum* (Cyperaceae), a new species from northeastern North America, with a key to *Carex* sect. *Phyllostachys*. *Systematic Botany*, **18**, 496–501.
- Chase MW, Cowan RS, Hollingsworth PM *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.
- Chase MW, Salamin N, Wilkinson M *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1889–1895.
- Chouinard BN, Starr JR, Naczi RFC (2008) Barcodes and relationships of the North American sedges, north of Mexico (*Carex*,

- Cariceae, Cyperaceae). Abstract ID. 811, Botany 2008, Vancouver, British Columbia, Canada <http://2008.botanyconference.org/engine/search/index.php?func=detail&aid=811>.
- Crins WJ, Naczi RFC, Reznicek AA, Ford BA (2002) *Carex* sect. *Phyllostachyae*. In: *Flora of North America, Volume 23, Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Ball PW, Gandhi K, Kiger RW *et al.*), pp. 558–563. Oxford University Press, Oxford, UK.
- Dabros A, Waterway MJ (2008) Segregation of sedge species (Cyperaceae) along environmental gradients in fens of the Schefferville region, northern Quebec. In: *Sedges: Uses, Diversity, and Systematics of the Cyperaceae* (eds Naczi RFC, Ford BA), pp. 145–161. Monographs in Systematic Botany from the Missouri Botanical Garden, St. Louis, Missouri.
- DeBry RW, Olmstead RG (2000) A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology*, **49**, 171–179.
- Dornbush ME (2004) Plant community change following fifty-years of management at Kalsow Prairie Preserve, Iowa, U.S.A. *American Midland Naturalist*, **151**, 241–250.
- Ertter B (2000) Floristic surprises in North America north of Mexico. *Annals of the Missouri Botanical Garden*, **87**, 81–209.
- Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **3**, e2802. doi: 10.1371/journal.pone.0002802.
- Ford BA, McQueen DAR, Naczi RFC, Reznicek AA (1998a) Allozyme variation and genetic relationships among species in the *Carex willdenowii* complex (Cyperaceae). *American Journal of Botany*, **85**, 546–552.
- Ford BA, McQueen DAR, Starr JR, Naczi RFC (1998b) The impact of species-specific traits and phylogenetic relatedness on allozyme diversity in *Carex* section *Phyllostachys* (Cyperaceae). *Plant Systematics and Evolution*, **212**, 13–29.
- Ford BA, Starr JR, McQueen DAR, Naczi RFC (1998c) Relationships among species in *Carex* section *Phyllostachys* (Cyperaceae) based on allozyme divergence. *Plant Systematics and Evolution*, **212**, 31–51.
- Ford BA, Naczi RFC (2001) Genetic diversity in the *Carex jamesii* complex (Cyperaceae: sect. *Phyllostachyae*) with insights into the evolution and origin of the newly described species *Carex timida*. *Sida*, **19**, 885–897.
- Ford BA, Iranpour M, Naczi RFC, Starr JR, Jerome CA (2006) Phylogeny of *Carex* subg. *Vignea* (Cyperaceae) based on non-coding nrDNA sequence data. *Systematic Botany*, **31**, 70–82.
- Ford BA, Naczi RFC, Starr JR (2008) *Carex* section *Phyllostachyae*: the value of a multidisciplinary approach in conducting systematics studies in sedges. In: *Sedges: Uses, Diversity, and Systematics of the Cyperaceae* (eds Naczi RFC, Ford BA). Monographs in Systematic Botany from the Missouri Botanical Garden, St. Louis, Missouri.
- Frodin DG (2004) History and concepts of big plant genera. *Taxon*, **53**, 753–776.
- Gignac LD, Gauthier R, Rochefort L, Bubier J (2004) Distribution and habitat niches of 37 peatland Cyperaceae species across a broad geographic range in Canada. *Canadian Journal of Botany*, **82**, 1292–1313.
- Good R (1974) *The Geography of the Flowering Plants*. Longman, London.
- Govaerts R (2001) How many species of seed plants are there? *Taxon*, **50**, 1085–1090.
- Hartman RL, Nelson BE (1998) Taxonomic novelties from North America north of Mexico: a 20-year vascular plant diversity baseline. *Monographs in Systematic Botany from the Missouri Botanical Garden*, **67**, 1–59.
- Herre EA, Machado CA, Bermingham E *et al.* (1996) Molecular phylogenies of figs and their pollinator wasps. *Journal of Biogeography*, **23**, 521–530.
- Hilu KW, Borsch T, Müller K *et al.* (2003) Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany*, **90**, 1758–1776.
- Hollingsworth PM (2008) Progress and outstanding questions. *Heredity*, **101**, 1–2.
- Janeway LP (2005) Noteworthy collections: *Carex pendula*. *Madrono*, **52**, 125.
- Karlsen SR, Elvebakk A (2003) A method using indicator plants to map local climatic variation in the Kangerlussuaq/Scoresby Sund area, East Greenland. *Journal of Biogeography*, **30**, 1469–1491.
- King MG, Roalson EH (2008) Exploring evolutionary dynamics of nrDNA in *Carex* subgenus *Vignea* (Cyperaceae). *Systematic Botany*, **33**, 514–524.
- Klinka K, Krajina VJ, Ceska A, Scagel AM (1989) *Indicator Plants of Coastal British Columbia*. UBC Press, Vancouver, Canada.
- Kral R, Manhart J, Bryson CT (1987) A new *Carex* sect. *Oligocarpae* (Cyperaceae) from western Arkansas and eastern Oklahoma. *Annals of the Missouri Botanical Garden*, **74**, 440–442.
- Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE*, **2**, e508. doi: 10.1371/journal.pone.0000508.
- Kress WJ, Erickson DL (2008) DNA barcodes: genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences, USA*, **105**, 2761–2762.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA*, **102**, 8369–8374.
- Lahaye R, van der Bank M, Bogarin D *et al.* (2008a) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA*, **105**, 2923–2928.
- Lahaye R, Savolainen V, Duthoit S, Maurin O, van der Bank M (2008b) A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nature Precedings* Available from URL: <http://hdl.handle.net/10101/npre.2008.1896.1>
- LeClerc-Blain J, Starr JR, Bull RD, Saarela JM (2009) A regional approach to plant DNA barcoding provides high species resolution of sedges (*Carex* and *Kobresia*, Cyperaceae) in the Canadian Arctic Archipelago. *Molecular Ecology Resources*, in press.
- Magee DW, Rorer A (1981) *Freshwater Wetlands: A Guide to Common Indicator Plants of the Northeast*. University of Massachusetts Press, Amherst, Massachusetts.
- Mastrogriuseppe J, Rothrock PE, Dibble AC, Reznicek AA (2002) *Carex* sect. *Ovales*. In: *Flora of North America, Vol. 23, Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Ball PW, Gandhi K, Kiger RW *et al.*), pp. 332–378. Oxford University Press, Oxford, UK.
- Meyer C, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Plos Biology*, **12**, 2229–2238.
- Molina A, Acedo C, Llamas F (2008) Taxonomy and new taxa of the *Carex divulsa* aggregate in Eurasia (section *Phaestoglochin*, Cyperaceae). *Botanical Journal of the Linnean Society*, **156**, 385–409.
- Naczi RFC (1990) The taxonomy of *Carex bromoides* (Cyperaceae). *Contributions from the University Michigan Herbarium*, **17**, 215–222.

- Naczi RFC (1992) *Systematics of carex section griseae (Cyperaceae)*. Ph.D. Dissertation. University of Michigan, Ann Arbor.
- Naczi RFC (1993) *Carex brysonii* and *Carex godfreyi*, new species of *Carex* section *Griseae* (Cyperaceae) from the southeastern United States. *Contributions from the University of Michigan Herbarium*, **19**, 195–205.
- Naczi RFC (1997) *Carex pigra*, a new species of *Carex* section *Griseae* (Cyperaceae) from the southeastern United States of America. *Novon*, **7**, 67–71.
- Naczi RFC (1999) *Carex planispicata*, a widespread and frequent new species of *Carex* section *Griseae* (Cyperaceae) from the eastern United States of America. *Journal of the Kentucky Academy of Science*, **60**, 37–44.
- Naczi RFC (2002) *Carex* sect. *Deweyanae*. In: *Flora of North America, volume 23, Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Ball PW, Gandhi K, Kiger RW *et al.*), pp. 321–325. Oxford University Press, Oxford, UK.
- Naczi RFC (2009) Insights on using morphologic data for phylogenetic analysis in sedges (Cyperaceae). *The Botanical Review*, **75**, 67–95.
- Naczi RFC, Bryson CT (2002) *Carex* sect. *Griseae*. In: *Flora of North America, Vol. 23, Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Ball PW, Gandhi K, Kiger RW *et al.*), pp. 448–461. Oxford University Press, Oxford, UK.
- Naczi RFC, Reznicek AA, Ford BA (1998) Morphological, geographical, and ecological differentiation in the *Carex willdenowii* complex (Cyperaceae). *American Journal of Botany*, **85**, 434–447.
- Naczi RFC, Bryson CT, Cochrane TS (2002) Seven new species and one new combination in *Carex* (Cyperaceae) from North America. *Novon*, **12**, 508–532.
- Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: evaluation of *rbcl* in a multigene tiered approach. *Canadian Journal of Botany*, **84**, 335–341.
- Newmaster SG, Fazekas AJ, Steeves AD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*, **8**, 480–490.
- Pennisi E (2007) Wanted: a barcode for plants. *Science*, **318**, 190–191.
- Reznicek AA, Catling PM (1987) *Carex praegracilis* (Cyperaceae) in eastern North America: a remarkable case of rapid invasion. *Rhodora*, **89**, 205–216.
- Rieseberg LH, Brouillet L (1994) Are many plant species paraphyletic? *Taxon*, **43**, 21–32.
- Ringius GS, Sims RA, Meades SJ (1997) *Indicator Plant Species in Canadian Forests*. Canadian Forest Service, Ottawa, Canada.
- Roalson EH, Columbus JT, Friar EA (2001) Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and *trnT-L-F* (cpDNA) region sequences: assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. *Systematic Botany*, **26**, 318–341.
- Saarela JM, Ford BA (2001) Taxonomy of the *Carex backii* complex (section *Phyllostachyae*, Cyperaceae). *Systematic Botany*, **26**, 704–721.
- Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120–1136.
- Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE*, **11**, e1154. doi: 10.1371/journal.pone.0001154.
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1805–1811.
- Scoggan HJ (1978) *The Flora of Canada Part 2 Pteridophyta, Gymnospermae, Monocotyledonae*. National Museum of Natural Sciences Publications in Botany, no. 7, Ottawa, Canada.
- Scotland RW, Wortley AH (2003) How many species of seed plants are there? *Taxon*, **52**, 101–104.
- Scott GAJ (1995) *Canada's Vegetation: a World Perspective*. McGill-Queen's University Press, Montréal, Canada.
- Seifert KA (2008) Recent developments and progress on fungal dna barcoding. In: *Towards a Barcoded Planet, Canadian Barcode of Life Network 2008 Scientific Symposium*, p. 24 Royal Ontario Museum, Toronto, Canada.
- Standley LA, Cayouette J, Bruederle L (2002) *Carex* sect. *Phacocystis*. In: *Flora of North America, Vol. 23, Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Ball PW, Gandhi K, Kiger RW *et al.*), pp. 379–401. Oxford University Press, Oxford, UK.
- Starr JR, Bayer RJ, Ford BA (1999) The phylogenetic position of *Carex* section *Phyllostachys* and its implications for phylogeny and subgeneric circumscription in *Carex* (Cyperaceae). *American Journal of Botany*, **86**, 563–577.
- Starr JR, Ford BA (2001) The taxonomic and phylogenetic utility of vegetative anatomy and fruit epidermal silica bodies in *Carex* section *Phyllostachys* (Cyperaceae). *Canadian Journal of Botany*, **79**, 362–379.
- Starr JR, Ford BA (2009) Phylogeny and evolution in Cariceae (Cyperaceae): current knowledge and future directions. *The Botanical Review*, **75**, 110–137.
- Starr JR, Harris SA, Simpson DA (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 JR, Harris SA, Simpson DA (2004) Phylogeny of the unispicate taxa in Cyperaceae tribe Cariceae I: generic relationships and evolutionary scenarios. *Systematic Botany*, **29**, 528–544.
- Starr JR, Harris SA, Simpson DA (2008) Phylogeny of the unispicate taxa in Cyperaceae tribe Cariceae II: the limits of *Uncinia*. In: *Sedges: Uses, Diversity, and Systematics of the Cyperaceae* (eds Naczi RFC, Ford BA). *Monographs in Systematic Botany from the Missouri Botanical Garden*, **108**, 243–267.
- Swofford DL (2003) *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sinauer & Associates, Sunderland, Massachusetts.
- Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, doi: 10.1093/nar/gkl938. PMID: PMC1807943.
- Tate JA, Simpson BB (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany*, **28**, 723–737.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal\_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.
- Tsukaya H, Yokoyama J, Ikeda H, Kuroiwa H, Kuroiwa T, Iwatsuki K (1997) Morphological, Physiological and Molecular Genetic Characterization of *Arabidopsis himalaica*, with reference to *A. thaliana*. *Journal of Plant Research*, **110**, 15–23.
- Vellend M, Lechowicz MJ, Waterway MJ (2000) Environmental distribution of four *Carex* species (Cyperaceae) in an old-growth forest. *American Journal of Botany*, **87**, 1507–1516.
- Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM (2008) Dynamics and evolution of the inverted repeat-large single

- copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology*, **8**, 36.
- Waterway MJ, Starr JR (2007) Phylogenetic relationships in the tribe Cariceae (Cyperaceae) based on nested analyses of three molecular data sets. *Aliso*, **23**, 165–192.
- Weldy T, Werier D (2008) *New York Flora Atlas*. New York Flora Association, Albany, New York. Available from URL: <http://www.newyork.plantatlas.usf.edu>.
- Wheeler GA (2007) *Carex sagei* (Cyperaceae), the correct name for *C. barrosii*. *Darwiniana*, **45**, 231–235.
- Yen AC, Olmstead RG (2000) Molecular systematics of Cyperaceae tribe Cariceae based on two chloroplast DNA regions: *ndhF* and *trnL* intron-intergenic spacer. *Systematic Botany*, **25**, 479–494.
- Zika P, Kuykendall K (2001) Some sedges (*Carex*) never found in Oregon. *Kalmiopsis*, **7**, 24–30.

## Supporting information

Additional supporting information may be found in the online version of this article:

**Appendix S1** Voucher data for all individuals used in this study. DNA and GenBank numbers are given in the following order in parentheses after collection data and herbarium acronyms: DNA number, *matK*, *trnH-psbA*, *rpoB*, *rpoC1*. All taxa are arranged alphabetically by section.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.