BMC
Ecology

METHODOLOGY ARTICLE

Open Access

Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library

Maria L Kuzmina[1*], Karen L Johnson[2], Hannah R Barron[3] and Paul DN Hebert[1]

Abstract

Background : Because arctic plant communities are highly vulnerable to climate change, shifts in their composition require rapid, accurate identifications, often for specimens that lack diagnostic floral characters. The present study examines the role that DNA barcoding can play in aiding floristic evaluations in the arctic by testing the effectiveness of the core plant barcode regions (*rbc*L, *mat*K) and a supplemental ribosomal DNA (ITS2) marker for a well-studied flora near Churchill, Manitoba.

Results: This investigation examined 900 specimens representing 312 of the 354 species of vascular plants known from Churchill. Sequencing success was high for *rbc*L: 95% for fresh specimens and 85% for herbarium samples (mean age 20 years). ITS2 worked equally well for the fresh and herbarium material (89% and 88%). However, sequencing success was lower for *mat*K, despite two rounds of PCR amplification, which reflected less effective primer binding and sensitivity to the DNA degradation (76% of fresh, 45% of herbaria samples). A species was considered as taxonomically resolved if its members showed at least one diagnostic difference from any other taxon in the study and formed a monophyletic clade. The highest species resolution (69%) was obtained by combining information from all three genes. The joint sequence information for *rbc*L and *mat*K distinguished 54% of 286 species, while *rbc*L and ITS2 distinguished 63% of 285 species. Discrimination of species within *Salix*, which constituted 8% of the flora, was particularly problematic. Despite incomplete resolution, the barcode results revealed 22 misidentified herbarium specimens, and enabled the identification of field specimens which were otherwise too immature to identify. Although seven cases of ITS2 paralogy were noted in the families Cyperaceae, Juncaceae and Juncaginaceae, this intergenic spacer played an important role in resolving congeneric plant species at Churchill.

Conclusions: Our results provided fast and cost-effective solution to create a comprehensive, effective DNA barcode reference library for a local flora.

Keywords: Arctic, DNA barcoding, *rbc*L, *mat*K, ITS2, Species resolution, Climate change, Biomonitoring

Background

Climate change has already led to substantial modification in the composition of Arctic plant communities [1] as reflected by shifting ranges and genetic differentiation [2]. Many arctic plant species are likely to lose genetic diversity due to their limited dispersal capacity, and consequent range reduction [3], making them particularly vulnerable to climate change. Identifying the impacts of climate change on the composition of plant communities is currently the focus of many studies in the Arctic which employ two main approaches. The first examines the impact of manipulations in light, temperature and nutrient regimes on species composition and richness [4-7]. The second approach involves direct examination of plant community composition to identify species that are particularly effective predictors of shifts in vegetation in response to climate change [8]. Both approaches require rapid and accurate identification of plants, many of which lack diagnostic floral or fruit characters at the time of their collection.

DNA barcoding employs sequence diversity in short, standardized gene region(s) to facilitate species identification [9]. Two gene regions from the chloroplast genome, *rbc*L and *mat*K, have been adopted as the standard barcodes for land plants [10]. Both of these genes have

---
* Correspondence: mkuzmina@uoguelph.ca
[1]Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada
Full list of author information is available at the end of the article

played a very important role in phylogenetic reconstructions for land plants due to their strong phylogenetic signal [11-14]. Their capacity to resolve species in local floras has now been tested in many settings, particularly in species-rich tropical communities. As well, numerous studies have tested the additional discrimination provided by supplemental chloroplast (*trn*H-*psb*A, *atp*B-*atp*H, *rpo*C1) and nuclear (ITS) markers [15-19]. All prior studies have reported 100% success in generic-level assignments, while success in species-level assignment has ranged from 50 – 92% for the two-locus barcode (*rbc*L & *mat*K), and from 70 – 98% with one or more supplementary markers. However, in all analyzed cases DNA barcoding has proven an efficient approach for the evaluation of hyper-diverse floras.

The nuclear ribosomal DNA region ITS and its two components, ITS1 and ITS2, have been extensively utilized for studies on the molecular systematics of plants because of their high rate of nucleotide substitution and relative ease of amplification, sequencing and alignment [20]. Among the varied supplemental barcode markers, ITS2 shows particular promise because its short length (160-320 bp) and the availability of universal primers make it easy to recover. Although it has been suggested that ITS2 exhibits too much paralogy, and is too susceptible to fungal contamination to be adopted as a DNA barcode marker [21,22], it delivered 92.7% discrimination in a recent study on 4800 species of medicinal plants [17]. Given this high performance, ITS2 merits serious consideration as a standard marker for plant barcoding.

Although plant communities in temperate and arctic regions are much less diverse than those in the tropics [23], they may not be easier targets for DNA barcode analysis because rates of molecular evolution in both plastid and nuclear genomes appear lower in groups of flowering plants with low diversity [24] and in plant species from high latitudes [25]. However, there is some evidence that arctic plant communities have experienced more rapid speciation, due to intense processes of hybridization, refugial isolation and range shifts [26]. The question of how this affects the performance of DNA barcoding for the identification of plant species has seen little investigation. However, a recent study of the flora at a temperate site in Canada revealed 93% success in species identification with *rbc*L & *mat*K, while the addition of the *trn*H-*psb*A intergenic spacer raised resolution to 95% [27].

The present study tests the effectiveness of DNA barcoding for the identification of species in the flora at Churchill, Manitoba, Canada. Our decision to work at this locality reflects an ongoing effort to assemble a comprehensive DNA barcode library for all animal and plant species at Churchill. Sequence information was collected for three gene regions (*rbc*L, *mat*K, ITS2) from 312 of the 354 species of vascular plants known from this locale [28,29]. Since herbarium collections can aid the rapid creation of comprehensive DNA barcode libraries [30], we compared the success of barcode recovery from herbarium and freshly collected specimens preserved in silica gel. We also investigated factors affecting sequence recovery for these three gene regions in a high-throughput barcoding setting, and adjusted protocols to enhance success. Finally, we compared the success of species identification in this arctic flora with those reported for temperate and tropical floras.

Methods
Study area
The Churchill area lies within the Hudson Bay Lowlands in a region where quartzite and dolomite bedrock has created a wide range of microhabitats. Poor drainage has led to the formation of extensive peat bogs, while broad tidal flats lie along the margin of Hudson Bay. Churchill is positioned in southern hypoarctic tundra with elements of high boreal subzone following along the Churchill River [31], with stable coexistence of oceanic and continental floristic elements [32]. The community of vascular plants around Churchill has been re-established in the 8000 years since deglaciation [29,33,34].

Tissue collection and identification
We examined 900 specimens including representatives of 312 species, 147 genera, 51 families, and 24 orders. Plant tissue from 540 specimens (60% of total) was collected from 35 localities around Churchill in July 2009, and dried in silica gel at room temperature. The remaining 360 specimens derived from the University of Manitoba Herbarium (WIN) and the Manitoba Museum's Botany Department (MMMN) and had a mean age of 20 years. When available, several individuals (2-5) per species from different populations were analyzed. Freshly collected specimens were identified using standard taxonomic references [35-38], with subsequent confirmation through comparison with specimens in WIN, MMMN, and CAN. The identification of willows (*Salix*) was confirmed by George Argus (CAN). Vouchers for the 540 freshly collected specimens, representing 241 species have been deposited in the BIO Herbarium (OAC) at the University of Guelph with duplicates at the Churchill Northern Studies Centre. The sequences for three barcode markers: *rbc*L, *mat*K, and ITS2 are publicly accessible in the project entitled "Plants of Churchill 2009" on BOLD [39], and are also available on GenBank under the accession numbers shown in Additional file 1.

Botanical nomenclature

We adopted the Checklist of the Panarctic Flora (PAF) Vascular Plants [40] to alleviate problems created by the frequent lack of standardization in name usage for species with Holarctic distributions. Some generic names diverge from those in the most recent taxonomic treatments for the North America flora [35-38,41] where *Oxycoccus* is placed within *Vaccinium* [42], *Cyrtorhyncha* within *Ranunculus* [43], two species of *Chamerion* are assigned to *Epilobium*, and *Comarum palustre* is treated as *Potentilla palustris* [36]. However, some new assignments (*Arctous*, *Orthilia*) have been accepted in both the Flora of North America, and in PAF [40,41]. According to the PAF checklist, 87 of 147 genera in our study were represented by a single species. Family and ordinal assignments follow the Angiosperm Phylogeny Group III [14].

DNA extraction, PCR and sequencing

DNA extraction followed standard protocols at the Canadian Centre for DNA barcoding (CCDB) for plants [44]. In brief, small amounts of dry plant tissue (0.5 cm^2) were placed into racked sterile mini tube strips. A 3.17 mm stainless steel bead was added to each tube before it was closed with a sterile cap strip. The tissue was then ground into fine powder using a Tissue Lyser (Qiagen, USA) with rack adapters at 28 Hz for 30 seconds; the adaptor was then rotated, and one more round of grinding was applied. The powdered tissues were incubated with 2x CTAB buffer at 65°C for 1 hour and DNA was then extracted using semi-automated method employing glass fiber filtration [45,46]. The final concentration of the eluted DNA was 20-40 ng/μL.

Three gene regions (*rbc*L, *mat*K, ITS2) were amplified using the CCDB plant protocol [46,47] with Platinum® Taq DNA polymerase (Invitrogen), and pre-made frozen plates [48]. Different PCR conditions were employed for *mat*K than for the other two regions (Table 1). Strong amplification of *rbc*L and ITS2 was obtained with low concentrations of primers (0.1 μM), dNTPs (0.05 mM), and Taq polymerase (0.024 U/μL). Subsequent 5-10x dilution of the amplicons enabled direct sequencing without PCR purification. One primer set (rbcLa-F [49] and rbcLa-R [16]) was used for all *rbc*L analysis and another for all ITS2 analysis (ITS-S2F [17], and ITS4 [50]). The *mat*K region required higher concentrations of all reagents: primers (0.5 μM), dNTPs (0.2 mM), and Taq polymerase (0.1 U/μL) to optimize amplicon recovery. A 10-fold dilution of DNA (2-4 ng/μL), and a smaller reaction volume (7.5 μL) improved the quality of amplicons for *mat*K, while reducing costs. We employed two primer sets to aid the recovery of *mat*K. The first round was performed with matK-1RKIM-f and matK-3FKIM-r

[pers.com. Ki-Joong Kim]. Failed samples were assembled into a new plate and amplified with matK_390f and matK_1326r [51]. Primer sequences for *rbc*L, *mat*K, and ITS2 are available on the CCDB Protocols website [52] and in Table 1. The cycle sequencing reaction and subsequent clean-up employed standard CCDB protocols [53] with products analyzed on an ABI 3730xl capillary sequencer.

Sequence data analysis

Sequence chromatograms were edited using CodonCode Aligner v.3.7.1 (CodonCode Co, USA). The traces were assembled into bidirectional contigs, primer sequences were removed, and all ambiguous base calls were checked manually. Contigs were compared using the MUSCLE multiple sequence alignment algorithm [54] implemented in CodonCode Aligner. The preliminary alignment with MUSCLE facilitated both the identification and correction of sequencing errors, mostly involving indels in homopolymer regions. BLAST was employed to recognize and exclude any fungal/algal sequences among ITS2 amplicons. The fasta sequences were visualized in BioEdit [55] and MEGA 5 [56], and double-checked for editing errors. A final alignment for the *rbc*L and *mat*K sequences was generated using back-translation in transAlign [57]. This step was particularly important to create the most parsimonious translated alignment for *matK* due to its multiple indels when taxonomically distant groups were compared. TransAlign also helped to reveal ORF shifts caused by editing errors (single missing or extra nucleotides), or pseudogenes. Sequences for ITS2 were clustered into 17 groups composed of species in a single or closely related group of orders. The sequences in each cluster were then aligned with BioEdit using CLUSTAL with a gap penalty of 5 for both pairwise and multiple alignments. An ITS2 sequence was considered as paralogous if it was assigned to a clade (or clades) that conflicted with the species assignment based on chloroplast marker(s) and morphology. The total alignments for *rbc*L, *mat*K and ITS2 alignments were concatenated into a profiled alignment with SequenceMatrix 1.5 alpha9 [58].

Data management and calculation of mean pairwise distances (MPD,%) were performed in BOLD [39]. The standard deviation (σ) was calculated to estimate dispersion of this parameter. The correlation coefficient (r) between three pairs of markers was calculated using MPD values for the families as covariates. Significance test (p-value) was done by Z-score method. All specimens with sequence data for ITS2 and for at least one of the two chloroplast markers (*rbc*L, *mat*K) were included in three datasets (*rbc*L & *mat*K; *rbc*L & ITS2; *rbc*L, *mat*K & ITS2). Analysis was performed using Maximum parsimony (MP) to establish if a particular taxon formed a

Table 1 Primers and PCR protocols

PCR recipe for *rbc*L and ITS2 (total volume of the reaction: 12.5 μL)

Reagents	Final concentration	Volume per reaction (μL)
10% trehalose	5%	6.25
ddH20		2.00
10X buffer	1x	1.25
50 mM MgCl2	2.5 mM	0.625
10 μM primer F	0.1	0.125
10 μM primer R	0.1	0.125
10 mM dNTPs	0.05	0.0625
Polymerase (5 U/μl)	0.024 U/μL	0.06
TOTAL		10.50
DNA template (20-40 ng/μL)		2 .00

PCR recipe for *mat*K (total volume of the reaction: 7.5 μL)

Reagents	Final concentration	Volume per reaction(μL)
20% trehalose	5%	1.875
ddH20		2.60
10X buffer	1x	0.75
50 mM MgCl2	1.5 mM	0.225
10 μM primer F	0.5	0.375
10 μM primer R	0.5	0.375
10 mM dNTPs	0.2	0.15
Polymerase (5 U/μl)	0.1 U/μL	0.15
TOTAL		6.50
DNA template (2-4 ng/μL)		1.00

Primer sets

	Sequence	Reference
***rbc*L primers**		
rbcLa-F	ATGTCACCACAAACAGAGACTAAAGC	[49] Levin et al. 2003
rbcLa-R	GTAAAATCAAGTCCACCRCG	[16] Kress & Erickson, 2009
***mat*K primers**		
MatK-1RKIM-f	ACCCAGTCCATCTGGAAATCTTGGTTC	Ki-Joong Kim, pers. comm.
MatK-3FKIM-r	CGTACAGTACTTTTGTGTTTACGAG	Ki-Joong Kim, pers. comm.
MatK_390f	CGATCTATTCATTCAATATTTC	[51] Cuenoud et al. 2002
MatK_1326r	TCTAGCACACGAAAGTCGAAGT	[51] Cuenoud et al. 2002
ITS2 primers		
ITS2-S2F	ATGCGATACTTGGTGTGAAT	[17] Chen et al. 2010
ITS4	TCCTCCGCTTATTGATATGC	[50] White et al. 1990

monophyletic clade. Prior to analysis, fasta files were converted into Nexus format using the Nexus Class Library [59]. PAUP was then used to implement Parsimony analysis. Support for branching patterns was assessed using the parsimony ratchet [60]. Consensus tree was generated from a set of the most parsimonious trees, and a single tree for each dataset was visualized in iTOL [61,62]. The species resolution for the supermatrix based on the three markers was compared with those for matrices based on two markers (*rbc*L & *mat*K; *rbc*L & ITS2). Congeneric species were considered as resolved when individuals within one species showed at least one consistent diagnostic difference from other species and produced a monophyletic clade in the MP tree. Since some genera were represented by a single species, their percent of species resolution, which actually reflects generic resolution, was calculated separately from the genera with more than one species. This approach helped to indicate sensitivity of different combinations of markers to species resolution within genera.

Results and discussion

Sequencing success

ITS2 sequences were obtained with equal success from fresh and herbarium material (89%, and 88% respectively), while sequence recovery for *rbc*L was 10% lower for herbarium than fresh specimens (85% versus 95%). The first set of *mat*K primers showed 36% less success from herbarium than fresh specimens (45% versus 76%). The second set of PCR primers delivered *mat*K records for another 7% of the fresh specimens, producing an overall 83% success, while 8% new records were obtained from herbarium samples, raising their overall success to 53% (Figure 1, Additional file 1**). These results indicate a strong correlation between recovery of the markers and their length (circa 800 bp for *mat*K, 552 bp for *rbc*L, and circa 350 bp for ITS2), reflecting DNA degradation in the herbarium specimens. However, the low success in sequence recovery for *mat*K also reflected difficulties in primer binding as evidenced by the 19% lower success in its recovery from fresh specimens in comparison with rbcL. The second round of PCR for *mat*K did amplify some groups that initially failed, but the overall recovery still fell well below those for the other two markers.
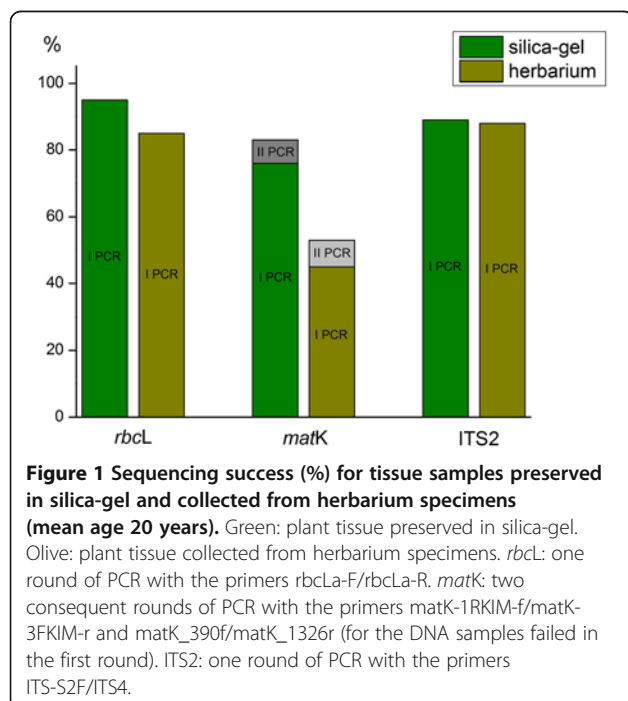
There was no noticeable association between the age of herbarium specimens and sequence recovery (Figure 2). ITS2 showed a high, stable recovery rate across all ages. The *rbc*L barcode also demonstrated good recovery that was not strongly impacted by age of the herbarium samples. Although herbarium specimens had lower recovery for *mat*K, their age did not markedly

Figure 2 Sequencing success (%) versus age of the herbarium samples. Black: *rbc*L. Red: *mat*K. Blue: ITS2.
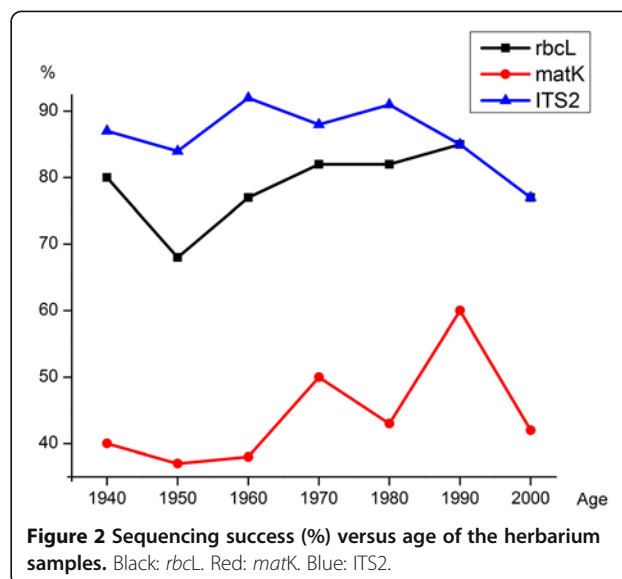
affect success. However, sequencing success was substantially lower for all three genes in genera *(Plantago, Atriplex, Anemone, Amerorchis)* with relatively thick leaves, likely due to slower desiccation of their tissues and consequent DNA degradation. This observation confirms previous evidence that the quality of herbarium preservation plays a more important role for sequence recovery than age of the samples [63]. Most of the herbarium specimens that we examined generated barcode sequences, revealing that herbaria can both accelerate the development of comprehensive barcode reference libraries and allow this work to be completed cost-effectively. Additionally, the analysis of herbarium vouchers and dialogue with curators increased the reliability of the identifications for the specimens that we analyzed. There was a reciprocal benefit for the collections, as barcode screening revealed inconsistencies in identification that provoked reconsideration of initial identifications.

Our work revealed several sources of interpretational complexity. No cases of ITS2 contamination were detected in the freshly collected specimens, but about 1% of the sequences from herbarium samples were of fungal or algal origin. Several cases of apparent sampling errors were also detected in herbarium specimens of graminoids (Poaceae, Juncaceae, and Cyperaceae) which may reflect mixed samples. Therefore, it is critical to record the exact site on the herbarium sheet where the tissue sample is obtained. Finally, the second round of PCR amplified *mat*K pseudogenes in four species (*Amerorchis rotundifolia, Limnorchis aquilonis, Limnorchis huronensis, Salix myrtillifolia* - Additional file 2). All sequences reflecting contamination or pseudogenes were excluded from further analysis.

Success in sequence recovery ranged from a low of 72% (*mat*K) to a high of 90% (*rbc*L), with 88% success

Figure 1 Sequencing success (%) for tissue samples preserved in silica-gel and collected from herbarium specimens (mean age 20 years). Green: plant tissue preserved in silica-gel. Olive: plant tissue collected from herbarium specimens. *rbc*L: one round of PCR with the primers rbcLa-F/rbcLa-R. *mat*K: two consequent rounds of PCR with the primers matK-1RKIM-f/matK-3FKIM-r and matK_390f/matK_1326r (for the DNA samples failed in the first round). ITS2: one round of PCR with the primers ITS-S2F/ITS4.
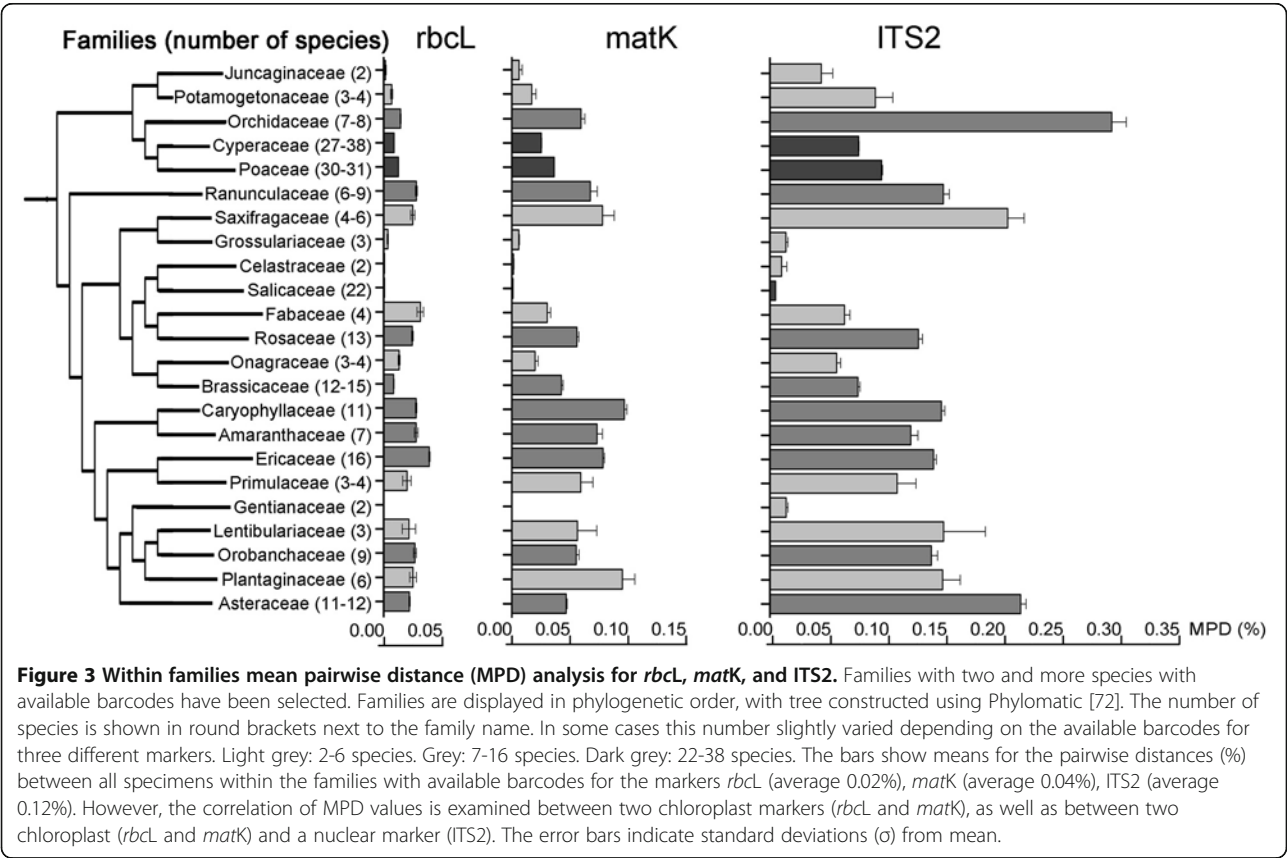
for ITS2. Twenty nine specimens failed to provide a sequence for both plastid markers, creating 871 specimens with sequence records for both ITS2 and one of these genes (*rbc*L & *mat*K from 68%, *rbc*L & ITS2 from 82%). Information for all three genes was available for 63% of the samples. In total 286 of the 312 species (92%) that we examined had sequence data for either *rbc*L or *mat*K allowing their inclusion in our evaluation of species resolution.

Within-family distances

Although mean pairwise distances (MPD) did not show a tight correlation with the resolution capacity of each marker, the comparison of this parameter within families represented by more than one species did reveal differences in the substitution rate among the three markers, and among taxonomic groups (Figure 3). The mean MPD for *rbc*L (0.017%) was noticeably lower than those for *mat*K (0.04%), and ITS2 (0.12%). The MPD for each marker also varied significantly among families, with a strong positive correlation between the two chloroplast markers (*rbc*L/*mat*K: r=0.84; p<0.05) as well as between the chloroplast and nuclear markers (*rbc*L/ITS2: r=0.60; *mat*K/ITS2: r=0.75; p<0.05), confirming prior evidence for a positive relationship between rates of molecular evolution in the chloroplast and nuclear genomes [20].

For example, families (Ranunculaceae, Saxifragaceae, Ericaceae, Caryophyllaceae, Amaranthaceae, Plantaginaceae) with high variation in *rbc*L and *mat*K also showed high sequence variation in ITS2. Conversely, other families (Juncaginaceae, Grossulariaceae, Celastraceae, Salicaceae, Gentianaceae) showed very low genetic variation in all three markers. The Brassicaceae was exceptional as its members showed relatively high sequence variation for *mat*K, and ITS2 despite low divergence for *rbc*L. Six other families (Orchidaceae, Poaceae, Cyperaceae, Rosaceae, Lentibulariaceae, Asteraceae) showed markedly higher variation for ITS2 than for the chloroplast markers, indicating the potential importance of this gene region for discriminating closely related species within these families.

The MPD value for a family was not linked to the number of species within it. For example, the most speciose families in this study (Cyperaceae, Poaceae) had moderate or low variation for all three markers. Moreover, although *Salix* was the most diverse genus (22 species) in the flora at Churchill, it showed some of the lowest MPD scores for all three genes. This observation supports prior evidence for differing rates of molecular evolution among plant families [64], and reveals that the most diverse plant families in the arctic do not have the highest rates of molecular evolution, a pattern which

Figure 3 Within families mean pairwise distance (MPD) analysis for *rbc*L, *mat*K, and ITS2. Families with two and more species with available barcodes have been selected. Families are displayed in phylogenetic order, with tree constructed using Phylomatic [72]. The number of species is shown in round brackets next to the family name. In some cases this number slightly varied depending on the available barcodes for three different markers. Light grey: 2-6 species. Grey: 7-16 species. Dark grey: 22-38 species. The bars show means for the pairwise distances (%) between all specimens within the families with available barcodes for the markers *rbc*L (average 0.02%), *mat*K (average 0.04%), ITS2 (average 0.12%). However, the correlation of MPD values is examined between two chloroplast markers (*rbc*L and *mat*K), as well as between two chloroplast (*rbc*L and *mat*K) and a nuclear marker (ITS2). The error bars indicate standard deviations (σ) from mean.

differs from the positive relationship between species number and rates of molecular evolution in flowering plants at the global scale [24].

Intraspecific variation and ITS2 paralogy

Intraspecific variation was not observed for *rbc*L in any of the 205 species with two or more sequence records. However, intraspecific variation was detected in 18 of 156 species with multiple records for *mat*K and in 35 of 189 species with multiple records for ITS2. Twelve of these species demonstrated intraspecific variation for both markers, six only showed intraspecific variation for *mat*K, while twenty three showed intraspecific variation for ITS2 (Additional file 3).

Paralogous sequences were detected in 7 of the 189 species with multiple records for ITS2. Some of these cases involved assignments to closely related species (e.g. *Triglochin palustris, Luzula groenlandica, Carex rotundata, Carex saxatilis*), a pattern which might reflect hybridisation events. However, in other cases (*Carex capillaris, Carex concinna, Kobresia myosuroides*), the ITS variants formed independent, distant clades within the same family, while the corresponding chloroplast markers assigned these individuals to monophyletic clades, excluding DNA contamination as a source of the paralogy.

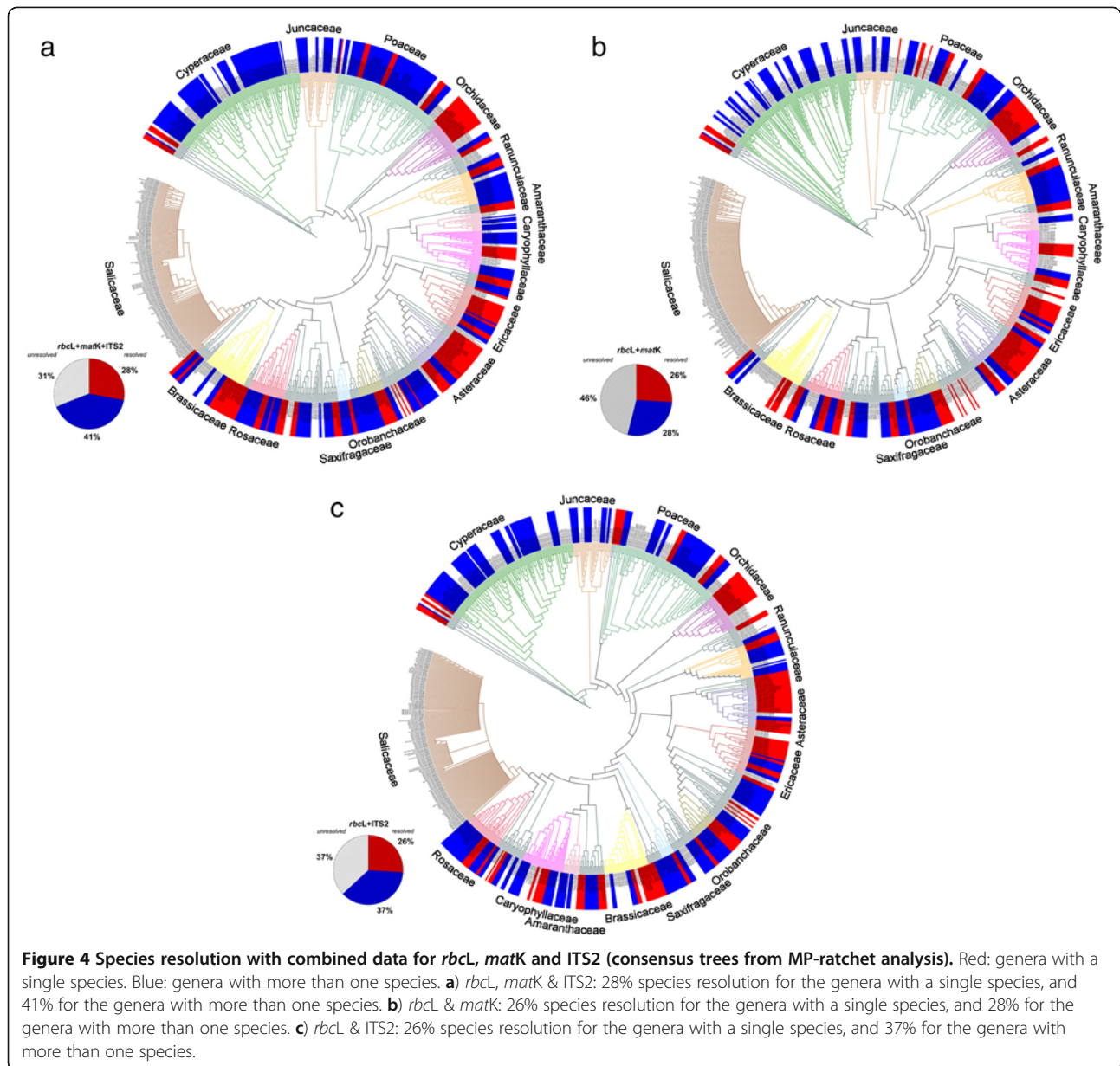### Species resolution with different barcode markers

The discriminatory power of the three barcode regions was considered in both pairwise and in the sole three-way combination. The consensus maximum parsimony (MP) tree based on all three markers (Figure 4a) was compared with those for two markers rbcL & matK (Figure 4b), and rbcL & ITS2 (Figure 4c). Because 28% of the species in the analysis belong to genera with a single species at Churchill, their resolution is effectively at a generic level. These taxa are indicated by red labels in the tree diagrams, while species in polytypic genera are represented in blue. The percent of resolved species from monotypic genera was included in overall calculation of the species resolution, but the percent of the species resolution for congeners is provided separately to demonstrate sensitivity of this parameter to the selection of markers. As expected, the combination of three markers (*rbc*L, *mat*K, & ITS2) delivered both the highest overall species resolution (69%) and the highest resolution of congeneric species (41%). By contrast, the tree built with the two standard barcodes (*rbc*L & *mat*K) showed 54% overall resolution, and just 28% resolution of congeners. Interestingly, the combination of *rbc*L & ITS2 delivered 63% overall species resolution and 37% resolution for the species in polytypic genera. Despite the incomplete resolution, the DNA barcode data revealed that 22 of 333

herbarium specimens were misidentified (Additional file 1*). Nearly 8% of the species (21 of 279) were only resolved by ITS2, while 1% of the species (3 of 279) were resolved by both *mat*K and ITS2, while another 1% of the species were only resolved by *mat*K (Additional file 4). Seven congeneric species pairs lacked resolution with all three markers: *Rhododendron tomentosum & R. groenlandicum, Arctous alpina & A. ruber, Cerastium alpinum & C. beeringianum, Limnorchis huronensis & L. aquilonis, Leymus innovatus & L. mollis, Elymus trachycaulis & E. violaceus, Puccinellia nuttaliana & P. lucida*. In addition, two species (*Arctophila fulva & Dupontia fisheri*) in different, but closely allied, genera could not be discriminated.

The 22 species of *Salix* presented a particularly dramatic case of compromised resolution. Despite their morphological distinctiveness, all but one (*Salix serissima*) of these species shared an identical *rbc*L haplotype. Thirteen *mat*K haplotypes were detected, but all were shared by more than one species. ITS2 also formed 11 polyphyletic clades which showed no congruence with the *mat*K clades. These results correspond with previous studies of phylogenetic relationships in willows based on *rbc*L, *mat*K and ITS [65,66], and likely reflect the impact of introgressive hybridisation [67,68]. Since members of the genus *Salix* comprised 8% of all species analyzed from Churchill, they substantially lowered the overall resolution of barcodes for the local flora.

Conclusions

Rates of species-level resolution have varied widely in past tests of the efficacy of plant DNA barcodes. Some of this variation reflects the differing nature of the studies; some have adopted a floristic approach [15-19], while others have targeted a taxonomic assemblage [69,70]. Other variation in resolution success reflects differences in the number and combination of gene markers used. Our study tested the capacity of the standard plant markers, *rbc*L and *mat*K, and one of the most frequent supplemental markers, ITS2, to resolve the species-poor flora at Churchill. We observed 69% species-level resolution with the three-locus data including 41% resolution for congeneric taxa. The two marker combinations were less effective: *rbc*L & *mat*K delivered 54% (28% congeners) resolution, while *rbc*L & ITS2 produced 63% (37% congeners) resolution. As a result, the inclusion of *mat*K raised overall species resolution by 6% and the resolution of congeners by 4% from results based on just *rbc*L and ITS2. Comparison of *rbc*L & ITS2 versus *rbc*L & *mat*K demonstrated an overall rise of 9% in species resolution, all reflecting better resolution of congeners, suggesting that ITS2 will aid resolution of congeners in the more diverse floras.

Figure 4 Species resolution with combined data for *rbc*L, *mat*K and ITS2 (consensus trees from MP-ratchet analysis). Red: genera with a single species. Blue: genera with more than one species. **a)** *rbc*L, *mat*K & ITS2: 28% species resolution for the genera with a single species, and 41% for the genera with more than one species. **b)** *rbc*L & *mat*K: 26% species resolution for the genera with a single species, and 28% for the genera with more than one species. **c)** *rbc*L & ITS2: 26% species resolution for the genera with a single species, and 37% for the genera with more than one species.

ITS2 data helped to resolve closely related species in some families (Brassicaceae, Rosaceae, Poaceae, Cyperaceae) with relatively high MPD. However, it did not aid resolution in the Orchidaceae although this family had the highest MPD (3%), perhaps because it reflected high intraspecific variation in two monotypic orchid genera (*Corallorhiza trifida, Amerorchis rotundifolia*). Several cases of ITS2 paralogy were noted in the Juncaginaceae, Juncaceae, and Cyperaceae (2%), and a few cases (1%) of fungal and algal contamination were detected in herbarium samples, but these complications did not seriously compromise the utility of ITS2 in discriminating congeneric species.

The *mat*K sequence information did increase the stability of the phylogeny as the three gene phylogeny had higher MP-ratchet support scores than either of the two gene analyses. It also provided additional, and in some cases unique, diagnostic traits for species resolution (e.g., *Calamagrostis*). However, the sensitivity of this marker to DNA degradation, the lack of universal primers and the complexity of sequence editing and alignment meant that the recovery of sequence information for this gene region was expensive. At present, the most cost-effective, rapid screening for the Churchill flora lies in the analysis of *rbc*L and ITS2.

The two standard DNA barcodes together with ITS2 delivered lower species-level resolution than those obtained in most prior floristic studies in the tropics and temperate zone. In part, this may reflect the fact that

our study examined a higher percentage of the local flora than many earlier investigations, and may therefore, have included a higher percentage of closely related taxa. Although the number of species examined in previous analyses was similar to that in our study (approx. 300 species), they derived from much more diverse floras, e.g. the tree plots in Amazon and Panama examined just about 3% of the resident species [15,16,71]. The higher success in species resolution for 4600 species of medicinal plants [17] may reflect the inclusion of phylogenetically diverse taxa. The low species resolution at Churchill might additionally reflect lowered rates of molecular evolution in low diversity plant communities [24,25] and in low temperate settings, but disentangling the impacts of these varied factors will require further study.

Additional files

Additional file 1: Specimen information with GenBank accession numbers.

Additional file 2: Pseudogenes amplified with matK_390f/ matK_1326r.

Additional file 3: Cases of intraspecific variation.

Additional file 4: Species resolution with different markers.

Abbreviations

BOLD: Barcode of Life Data System; CAN: National Herbarium of Canada, Canadian Museum of Nature; CCDB: Canadian Centre for DNA Barcoding; CNSC: Churchill Northern Studies Centre; KSR: Koffler Scientific Reserve, Ontario; MMMN: Herbarium at the Manitoba Museum, Winnipeg; MP: Maximum Parsimony analysis; MPD: Mean pairwise distance; OAC: Herbarium at the Biodiversity Institute of Ontario, University of Guelph; PAF: Checklist of the Panarctic Flora Vascular Plants; WIN: University of Manitoba Herbarium.

Authors' contributions

MK designed the experiment, collected and identified plant material, generated and analysed molecular data, performed statistical analysis, and also drafted the manuscript. KJ planned and performed field investigations, and also collected and identified plant material. HB collected and identified plant material, aided in molecular analysis, and also helped draft the manuscript. PH aided study design, revised the manuscript, and also gained support for the work. All authors read and approved the final manuscript.

Author details

[1]Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada. [2]The Manitoba Museum, Botany Department, Manitoba, Canada. [3]Trent University, Environmental and Life Sciences, Peterborough, ON, Canada.

References

1. Henry GHR, Molau U: **Tundra plants and climate change: the International Tundra Experiment (ITEX).** *Glob Change Biol* 1997, **3**(Suppl 1):1–9.
2. Davis MB, Shaw RG, Etterson JR: **Evolutionary responses to changing climate.** *Ecology* 2005, **86**:1704–1714.
3. Alsos IG, Ehrich D, Thuiller W, Eidesen PB, Tribsch A, Schönswetter P, Claire Lagaye C, Taberlet P, Brochmann C: *Genetic consequences of climate change for northern plants.* Proc Soc Lond B Bio: Proc Soc Lond B Bio; 2012. doi:101098/rsb20112363.
4. Chapin FS III, Shaver GR, Giblin AE, Nadelhoffer KJ, Laundre JA: **Responses of Arctic tundra to experimental and observed changes in climate.** *Ecology* 1995, **76**:694–711.
5. Robinson CH, Wookey PA, Lee JA, Callaghan TV, Press MC: **Plant community responses to simulated environmental change at a high Arctic polar semi-desert.** *Ecology* 1998, **79**:856–866.
6. Walker MD, Wahren CH, Hollister RD, Henry GHR, Ahlquist LE, Alatalo JM, Bret-Harte MS, Calef MP, Callaghan TV, Carroll AB, Epstein HE, Jonsdottir IS, Klein JA, Magnusson B, Molau U, Oberbauer SF, Rewa SP, Robinson CH, Shaver GR, Suding KN, Thompson CC, Tolvanen A, Totland O, Turner PL, Tweedie CE, Webber PJ, Wookey PA: **Plant community responses to experimental warming across the tundra biome.** *Proc Natl Acad Sci U S A* 2006, **103**:1342–1346.
7. Elmendorf SC, Henry GHR, Hollister RD, Björk RG, Bjorkman AD, Callaghan TV, Collier LS, Cooper EJ, Cornelissen JHC, Day TA, Fosaa AM, Gould WA, Grétarsdóttir J, Harte J, Hermanutz L, Hik DS, Hofgaard A, Jarrad F, Jónsdóttir IS, Keuper F, Klanderud K, Klein JA, Koh S, Kudo G, Lang SI, Loewen V, May JL, Mercado J, Michelsen A, Molau U, et al: **Global assessment of experimental climate warming on tundra vegetation: heterogeneity over space and time.** *Ecol Lett* 2012, **15**:164–175.
8. Chapin FS III, Bret-Harte MS, Hobbie SE, Zhong H: **Plant functional types as predictors of transient responses of arctic vegetation to global change.** *J Veg Sci* 1996, **7**:347–358.
9. Hebert PDN, Cywinska A, Ball SL, DeWaard JR: **Biological identification through DNA barcodes.** *Proc Soc Lond B Bio* 2003, **270**:313–321.
10. CBOL Plant Working Group: **A DNA barcode for land plants.** *Proc Natl Acad Sci USA* 2009, **106**:12794–12797.
11. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG QIUYL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, et al: **Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL.** *Ann Mo Bot Gard* 1993, **80**:528–580.
12. Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS CS, Chatrou LW: **Angiosperm phylogeny based on matK sequence information.** *Am J Bot* 2003, **90**:1758–1776.
13. The Angiosperm Phylogeny Group II: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II.** *Bot J Linn Soc* 2003, **141**:399–436.
14. The Angiosperm Phylogeny Group III: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III.** *Bot J Linn Soc* 2009, **161**:105–121.
15. Gonzalez MA, Baraloto C, Engel J, Mori SA, Petronelli P, Riéra B, Roger A, Thébaud C, Chave J: **Identification of Amazonian trees with DNA barcodes.** *PLoS One* 2009, **4**:e7483. doi:101371/journalone0007483.
16. Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E: **Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama.** *Proc Natl Acad Sci USA* 2009, **106**:18621–18626.
17. Chen S, Yao H, Han J, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C: **Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species.** *PLoS One* 2010, **5**:e8613. doi:101371/journalone0008613.
18. Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A: **Plant DNA barcodes can accurately estimate species richness in poorly known floras.** *PLoS One* 2011, **6**:e26841. doi:101371/journalone0026841.

19. Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, Chuyong GB, Cruaud C, Hardy OJ: *How effective are DNA barcodes to identify African rain forest trees?* PLoS ONE, in progress.

20. Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ: **The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny.** *Ann Mo Bot Gard* 1995, **82**:247–277.

21. Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M: **A proposal for a standardised protocol to barcode all land plants.** *Taxon* 2007, **5**:295–299.

22. Hollingsworth PM, Graham SW, Little DP: **Choosing and using a plant DNA barcode.** *PLoS One* 2011, **6**:e19254. doi:101371/journalone0019254.

23. Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, McCain CM, McCune AR, McDade LA, McPeek MA, Near TJ, Price TD, Ricklefs RE, Roy K, Sax DF, Schluter D, Sobel JM, Turelli M: **Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography.** *Ecol Lett* 2007, **10**:315–331.

24. Barraclough TG, Savolainen V: **Evolutionary rates and species diversity in flowering plants.** *Evolution* 2001, **55**:677–683.

25. Gillman LN, Keeling DJ, Gardner RC, Wright SD: **Faster evolution of highly conserved DNA in tropical plants.** *Evol Biol* 2010, **23**:1327–1330.

26. Brochmann C, Brysting AK: **The Arctic – an evolutionary freezer?** *Plant Ecol Divers* 2008, **2008**(1):181–195.

27. Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH: **Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode.** *Method Ecol Evol* 2011, **2**:333–340.

28. Ritchie JC: **The native plants of Churchill.** *Manitoba. Can J Bot* 1956, **34**:269–320.

29. Scoggan HJ: *The native flora of Churchill, Manitoba with notes on the history, geology, and climate of the area.* Ottawa: Ministry of Northern Affairs and National Resources; 1959.

30. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT: **How to Open the treasure chest? Optimising DNA extraction from herbarium specimens.** *PLoS ONE* 2012, **7**:e43808. doi:10.1371/journal.pone.0043808.

31. Yurtsev BA: **The floristic division of the Arctic.** *J Veg Sci* 1994, **5**:765–776.

32. Walker DA, Maier HA, Barbour EM, Raynolds MK: *Toolik-Arctic Geobotanical Atlas.* 2011 http://www.arcticatlas.org/maps/.

33. Riley JL: *Flora of the Hudson Bay Lowland and its Postglacial Origins.* Ottawa: National Research Council of Canada; 1950.

34. Sims RA, Riley JL, Jeglum JK: *Vegetation, flora and vegetational ecology of the Hudson Bay Lowland: a literature review and annotated bibliography, Report 0-X-297.* Great Lakes Forest Research Centre, Canadian Forestry Service, Department of the Environment: Sault Ste. Marie, Ontario; 1979.

35. Scoggan HJ: *The Flora of Canada.* Ottawa: National Museum of Natural Sciences: Pteridophyta, Gymnospermae, Monocotyledoneae; 1978-1979.

36. Porsild AE, Cody WJ: *Vascular plants of continental Northwest Territories, Canada.* Ottawa: National Museum of Natural Sciences; 1980.

37. Johnson K: *Wildflowers of Churchill and the Hudson Bay Region.* Winnipeg: University of Manitoba Press; 1998.

38. Aiken SG, Dallwitz MJ, Consaul LL, McJannet CL, Gillespie LJ, Boles RL, Argus GW, Gillett JM, Scott PJ, Elven R, LeBlanc MC, Brysting AK, Solstad H: *Flora of the Canadian Arctic Archipelago.* 2003 http://www.mun.ca/biology/delta/arcticf/.

39. Ratnasingham S, Hebert PDN: **BOLD: The Barcode of Life Data System (www.barcodinglife.org).** *Mol Ecol Notes* 2007, **7**:355–364.

40. Elven R, Murray DF, Razzhivin V, Yurtsev BA: *Checklist of the Panarctic Flora (PAF).* 2011. URL http://gbif.no/paf.

41. e. 1: *Flora of North America Editorial Committee.* New York: Oxford University Press: Flora of North America North of Mexico 16+ vols.

42. Tucker GC: *Ericaceae Jussieu.* New York: Oxford University Press: In Flora of North America North of Mexico. 16+ vols. Volume 8; 2009:370–535.

43. Whittemore AT, Parfitt BD: **Ranunculaceae Jussieu Crowfoot Family.** In *Flora of North America North of Mexico 16+ vols*, Volume 3. New York: Oxford University Press; 1997:85–271.

44. Ivanova N, Kuzmina M, Fazekas A: *CCDB Protocols.* Manual Protocol Employing Centrifugation: Glass Fiber Plate DNA Extraction Protocol For Plants, Fungi, Echinoderms and Mollusks; 2011 http://www.ccdb.ca/CCDB_DOCS/CCDB_DNA_Extraction-Plants.pdf.

45. Ivanova NV, Fazekas AJ, Hebert PDN: **Semi-automated, membrane-based protocol for DNA isolation from plants.** *Plant Mol Biol Rep* 2008, **26**:186–198.

46. Fazekas AJ, Kuzmina ML, Newmaster SG, Hollingsworth PM: *DNA barcoding methods for land plants.* Methods in Molecular Biology: In DNA barcodes: methods and protocols; 2012:858. doi:101007/9781-61779-591-6_11.

47. Kuzmina M, Ivanova N: *PCR Amplification for Plants and Fungi.* 2011 http://www.ccdb.ca/CCDB_DOCS/CCDB_Amplification-Plants.pdf.

48. Ivanova NV, Grainger C: *Pre-made frozen PCR and sequencing plates.* 2006 www.dnabarcoding.ca/CCDB_DOCS/CCDB_Advances_Methods_Release_No4_Dec1st_2006.pdf.

49. Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KJ: **Family-level relationships of Onagraceae based on chloroplast *rbc*L and *ndh*F data.** *Am J Bot* 2003, **90**:107–115.

50. White TJ, Bruns T, Lee S, Taylor J: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.** In *PCR Protocols: a guide to methods and amplifications.* New York: Academic; 1990:315–322.

51. Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW: **Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K DNA sequences.** *Am J Bot* 2002, **89**:132–144.

52. Kuzmina M, Ivanova N: *CCDB Protocols.* 2011 http://www.ccdb.ca/CCDB_DOCS/CCDB_PrimerSets-Plants.pdf.

53. Ivanova NV, DeWaard JR, Hajibabaei M, Hebert PDN: *Protocols for high volume DNA barcoding.* 2005 http://www.dnabarcoding.ca/.

54. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.

55. Hall TA: **BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95–98.

56. Tamura K, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Mol Biol Evol* 2011, **28**:2731–2739.

57. Bininda-Emonds ORP: **Software transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences.** *BMC Bioinforma* 2005, **6**:156.

58. Vaidya G, Lohman D, Meier R: **Sequence Matrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information.** *Cladistics* 2011, **27**:171–180.

59. Lewis PO, Holder MT: *Nexus Class Library (NCL).* 2008 http://sourceforge.net/projects/ncl/.

60. Nixon KC: **The parsimony ratchet, a new method for rapid parsimony analysis.** *Cladistics* 1999, **15**:407–414.

61. Letunic I, Bork P: **Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.** *Bioinformatics* 2006, **23**:127–128.

62. Letunic I, Bork P: **Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.** *Nucleic Acids Res* 2011, :1–4. doi:101093/nar/gkr201.

63. Staats M, Cuenca A, Richardson JE, Vrielink-Van Ginkel R, Petersen G, Seberg O, Freek T, Bakker FT: **DNA damage in plant herbarium tissue.** *PLoS One* 2011, **6**:28448. doi:10.1371/journal.pone.0028448.

64. Clegg MT, Gaut BS, Learn GH, Morton BR: **Rates and patterns of chloroplast DNA evolution.** *Proc Natl Acad Sci USA* 1994, **91**:6795–6801.

65. Azuma T, Kajita T, Yokoyama J, Ohashi H: **Phylogenetic relationships of *Salix* (Salicaceae) based on rbcL sequence data.** *Am J Bot* 2000, **87**:67–75.

66. Hardig TM, Anttila CK, Brunsfeld SJ: **A phylogenetic analysis of *Salix* (Salicaceae) based on matK and ribosomal DNA sequence data.** *J Bot* 2010, doi:101155/2010/197696.

67. Hardig TM, Brunsfeld SJ, Fritz RS, Morgan M, Orians CM: **Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone.** *Mol Ecol* 2000, **9**:9–24.

68. Mosseler A: **Hybrid performance and species crossability relationships in willows (*Salix*).** *Can J Bot* 1990, **68**:2329–2338.

69. Sass C, Little DP, Stevenson DW, Specht CD: **DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of cycads.** *PLoS One* 2007, **11**:e1154. doi:10.1371/journal.pone.0001154.

70. Yao H, Song JY, Ma XY, Liu C, Li Y, Xu HX, Han JP, Duan LS, Chen SL: **Identification of *Dendrobium* species by a candidate DNA barcode**

sequence: the chloroplast psbA-trnH intergenic region. *Planta Med* 2009,
75:667–669.

71. Hubbell SP, He F, Condit R, Borda de A' Gua L, Kellner J, Ter Steege H: **How
    many tree species are there in the Amazon and how many of them will
    go extinct?** *PNAS* 2008, **105**:11498–11504.

72. Webb CO, Donoghue MJ: **Phylomatic: tree assembly for applied
    phylogenetics.** *Mol Ecol Notes* 2005, **5**:181–183.