

RESEARCH ARTICLE

Testing the Efficacy of DNA Barcodes for Identifying the Vascular Plants of Canada

Thomas W. A. Braukmann*, Maria L. Kuzmina, Jesse Sills, Evgeny V. Zakharov, Paul D. N. Hebert

Centre for Biodiversity Genomics, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada

* tbraukma@uoguelph.ca



Abstract

Their relatively slow rates of molecular evolution, as well as frequent exposure to hybridization and introgression, often make it difficult to discriminate species of vascular plants with the standard barcode markers (*rbcL*, *matK*, ITS2). Previous studies have examined these constraints in narrow geographic or taxonomic contexts, but the present investigation expands analysis to consider the performance of these gene regions in discriminating the species in local floras at sites across Canada. To test identification success, we employed a DNA barcode reference library with sequence records for 96% of the 5108 vascular plant species known from Canada, but coverage varied from 94% for *rbcL* to 60% for ITS2 and 39% for *matK*. Using plant lists from 27 national parks and one scientific reserve, we tested the efficacy of DNA barcodes in identifying the plants in simulated species assemblages from six biogeographic regions of Canada using BLAST and *mothur*. Mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) were strong predictors of barcode performance for different plant families and genera, and both metrics supported ITS2 as possessing the highest genetic diversity. All three genes performed strongly in assigning the taxa present in local floras to the correct genus with values ranging from 91% for *rbcL* to 97% for ITS2 and 98% for *matK*. However, *matK* delivered the highest species discrimination (~81%) followed by ITS2 (~72%) and *rbcL* (~44%). Despite the low number of plant taxa in the Canadian Arctic, DNA barcodes had the least success in discriminating species from this biogeographic region with resolution ranging from 36% with *rbcL* to 69% with *matK*. Species resolution was higher in the other settings, peaking in the Woodland region at 52% for *rbcL* and 87% for *matK*. Our results indicate that DNA barcoding is very effective in identifying Canadian plants to a genus, and that it performs well in discriminating species in regions where floristic diversity is highest.

OPEN ACCESS

Citation: Braukmann TWA, Kuzmina ML, Sills J, Zakharov EV, Hebert PDN (2017) Testing the Efficacy of DNA Barcodes for Identifying the Vascular Plants of Canada. PLoS ONE 12(1): e0169515. doi:10.1371/journal.pone.0169515

Editor: Shilin Chen, Chinese Academy of Medical Sciences and Peking Union Medical College, CHINA

Received: August 29, 2016

Accepted: December 18, 2016

Published: January 10, 2017

Copyright: © 2017 Braukmann et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Complete taxonomic information, collection records, voucher images and sequences for 17,995 specimens are publically available through BOLD (Ratnasingham and Hebert 2007) in the plants of Canada project (Available as of January 4, 2016; doi: dx.doi.org/10.5883/DS-VASCAN).

Funding: This research was funded through the International Barcode of Life project supported by the Genome Canada through Ontario Genomics, Canada Foundation for Innovation, and by the

Introduction

DNA barcoding employs sequence variation in short, standardized gene regions as a tool to discriminate species [1]. The ideal DNA barcode region is reliably amplified and sequenced across large assemblages of taxa and provides a high level of species discrimination [2]. The

Ontario Ministry for Research and Innovation. It is also a contribution to the Food From Thought research program supported by the Canada First Research Excellence Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

success of the 5' region of the mitochondrial cytochrome *c* oxidase I (COI) gene in discriminating animal species motivated efforts to identify gene regions that might deliver similar resolution for plants. Due to the extremely low rates of nucleotide substitution in mitochondrial genes in most plant lineages [3], COI was not a candidate. However, building on their intense use for phylogenetics and molecular systematics, two plastid gene regions were considered as DNA barcodes for vascular plants and the large subunit of RuBisCo (*rbcL*) in combination with an intron maturase (*matK*) were adopted as standards [4; 5]. Because these regions often fail to resolve congeners [6; 7; 8; 9; 10; 11], there has been a subsequent trend, building on earlier suggestions [12; 13; 14], to couple them with the nuclear-encoded ribosomal internal transcribed spacer, ITS2 [2; 15; 16].

A considerable number of studies have now examined the performance of different markers with respect to both their ease of amplification and their capacity to resolve plant species [9; 10; 15; 17; 18; 19; 7; 20; 21; 22; 23; 24; 25; 26]. This work has indicated that *rbcL* has the highest level of sequence recovery (90–100%), followed by ITS2 (~90%), while *matK* is more difficult (56–90%). The efficacy of these gene regions in discriminating species has been determined by tree-based (phylogenetic) or basic local alignment (BLAST) algorithms. ITS2 has been reported to deliver the highest species resolution (79–93%) followed by *matK* (45–80%), and *rbcL* (17%–92%). It was suggested that the efficacy of DNA barcodes in delivering species-level identifications could be improved by developing local libraries [7; 27], and it was later demonstrated that this approach did indeed improve resolution [9; 23]. The effectiveness of such libraries depends upon complete sampling of local floras, accurate identification of the specimens that are analyzed, and quality of the resultant sequences [28].

Comparisons among past studies are difficult due to high variance in taxonomic scope (30–4800 species), biogeographic focus (e.g. Arctic and temperate floras, tropical trees), the number of DNA barcode markers employed (2–8 chloroplast and nuclear), and the methodologies used for making taxonomic assignments. In fact, no prior study has involved a large-scale comparative analysis of the capacity of the standard barcode markers (*rbcL*, *matK*, ITS2) to deliver a species-level identification for different biogeographic communities using a standard barcode library with the same methods. This study addresses this gap by employing a DNA barcode library for the vascular plants of Canada to determine the method that yields the best species resolution and the marker (*rbcL*, *matK*, ITS2) with the highest performance. As well, this study examines the efficacy of custom DNA barcode libraries for identification success, and compares phylogenetic diversity measures between sites and among species-rich families to determine factors affecting species resolution.

Materials and Methods

Taxonomic sampling

Sequences for three DNA barcode regions (*rbcL*, *matK* and ITS2) were generated for the vascular plants of Canada at the Canadian Center for DNA Barcoding [29]. Complete taxonomic information, collection records, voucher images and sequences for 17,995 specimens are publicly available through BOLD [30] in the plants of Canada project (Available as of January 4, 2016; doi: [dx.doi.org/10.5883/DS-VASCAN](https://doi.org/10.5883/DS-VASCAN)). This sequence library includes records for 4923 of the 5108 species of non-hybrid origin (~96%) with coverage for all 1153 genera and 171 families in the Database of Vascular Plants of Canada (VASCAN; [31]). Coverage varies among the three gene regions; the *rbcL* dataset is most complete with 16,008 sequences spanning 4790 species (~93.8%) in 168 families (Table 1). The ITS2 library includes 6630 sequences representing 3044 species (~59.6%) in 125 families while the *matK* dataset includes 6599 sequences covering 2000 species (39%) across 118 families. Overall, 78% of the species (3839)

Table 1. List of localities, corresponding terrestrial ecozones and biogeographic regions used to test the taxonomic resolution of *rbcl*, *matK*, and ITS2 libraries for the vascular plants of Canada. The number of species at each locale is in parentheses.

Park (species)	Terrestrial Ecozones	Region
Ellesmere Island NP (135)	Northern Arctic	Arctic
Ivvavik NP (402)	Southern Arctic	Arctic
Nahannii NP (652)	Taiga Plains	Arctic
Tornгат Mountains NP (368)	Arctic Cordillera	Arctic
Ukkusiksalik NP (150)	Northern Arctic	Arctic
Wapusk NP (272)	Hudson Plains	Arctic
Forillon NP (622)	Atlantic Maritime	Atlantic
Fundy NP (732)	Atlantic Maritime	Atlantic
Kejimikujik NP (578)	Atlantic Maritime	Atlantic
Prince Edward Island NP (648)	Atlantic Maritime	Atlantic
La Mauricie NP (461)	Boreal Shield	Boreal
Mingan Archipelago NP (441)	Boreal Shield	Boreal
Pukaskwa NP (545)	Boreal Shield	Boreal
Terra Nova NP (506)	Boreal Shield	Boreal
Banff NP (910)	Montane Cordillera	Pacific
Glacier NP (634)	Montane Cordillera	Pacific
Mount Revelstoke NP (435)	Montane Cordillera	Pacific
Pacific Rim NP (436)	Pacific Maritime	Pacific
Yoho NP (658)	Montane Cordillera	Pacific
Elk Island NP (482)	Boreal Plains	Prairies
Grasslands NP (427)	Prairies	Prairies
Prince Albert NP (643)	Boreal Plains	Prairies
Riding Mountain NP (714)	Prairies	Prairies
Waterton Lakes NP (976)	Prairies	Prairies
1000 Islands NP (1631)	Mixed wood Plains	Woodland
Bruce Peninsula NP (877)	Mixed wood Plains	Woodland
Koffler Scientific Reserve (621)	Mixed wood Plains	Woodland
Point Pelee NP (858)	Mixed wood Plains	Woodland

doi:10.1371/journal.pone.0169515.t001

possess records for some combination of two markers, but only 1074 species (22%) have data for all three.

To test the taxonomic resolution of the DNA library we created ‘synthetic’ floras based on the checklist of vascular plants for each of 27 Canadian National Parks and the Koffler Scientific Reserve (KSR). Initial checklists were generated using the Parks Canada Biotics Web explorer at http://www.pc.gc.ca/apps/bos/bosfieldselection_e.asp, with more recent updates for Ellesmere, Ivvavik, Nahannii, Point Pelee, Tornгат, Ukkusiksalik, and Wapusk National Parks (Bruce Bennett and Sergei Ponomarenko, personal communication). The species list for KSR was obtained from <http://ksr.utoronto.ca/research/species-list/ksr-plant-list/>. Plant species on the checklists were best represented by *rbcl* (> 95% coverage), followed by ITS2 and *matK* with comparable coverage (54–83% depending on the community; see Fig 1 for details). For the purpose of further analyses, the 28 checklists were clustered into six biogeographic regions: Arctic, Atlantic, Boreal, Pacific, Prairies, and Woodland (Table 1) representing 12 of the 15 terrestrial Canadian ecozones [32]. To ensure standardization of naming, all specimens and checklists used in this study followed the nomenclature accepted by VASCAN [31].

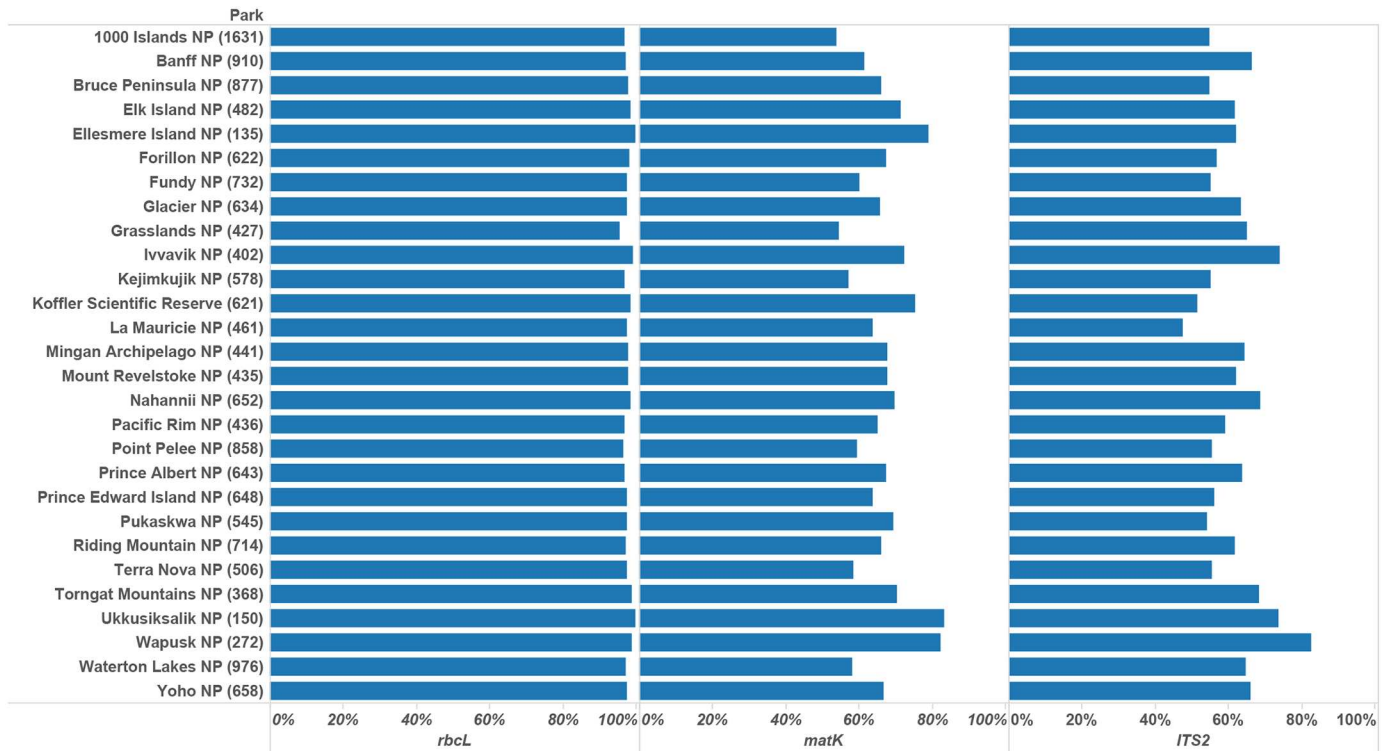


Fig 1. Coverage by barcode locus for the plant communities at 28 Canadian localities. The number of plant species present at each site is indicated in parentheses.

doi:10.1371/journal.pone.0169515.g001

Sequencing and analysis of libraries

Data validation. To reduce redundancy, identical sequences were clustered in UCLUST [33] and each cluster was parsed to its respective species (one species could be represented by more than one cluster). Sequences were then aligned using transAlign [34] for *rbcL* and *matK* (universal codon table), and MAFFT ver 7.221 for ITS2 under default parameters (FFT-NS-2 strategy) [35]. Maximum likelihood phylogenies were inferred for each alignment using RAxML Black box [36] on XCEDE via the CIPRES portal [37]. A dataset of 1074 species with records for all three gene regions was used to evaluate variation in taxonomic resolution (via BLAST and mothur) and phylogenetic metrics (MPD and MNTD). To estimate the number of unique sequences as a proxy for sequence variation, we clustered each marker at 100% using UCLUST [33].

Phylogenetic matrices. We calculated two metrics for each barcode region, mean phylogenetic distance (MPD) and mean nearest taxon distance (MNTD) [38] to examine their potential as predictors of the capacity of each region to resolve species. MPD is the average of the branch lengths (or distances) across all pairs of taxa in a phylogeny. It summarizes the overall phylogenetic diversity of a community and is influenced by the number of taxa in a tree [39]. By comparison, MNTD is an average of the distance between nearest neighbours so it describes the terminal phylogenetic structure. MNTD is the more appropriate measure of species resolution because it excludes internal nodes and instead calculates the mean distance between closely related species. Because both measures are influenced by polytomies in a phylogeny [40], we only included one representative per species to avoid bias created by an unequal number of sequences per species.

MPD and MNTD were estimated using the *picante* package [41] in R ver 3.2.0 [42]. The phylogenetic matrices for each barcode were calculated using the maximum likelihood tree, and were partitioned by family and genus. Regression analysis was used to determine if there was a correlation between each phylogenetic diversity metric and the number of sequences for a family. We also compared MNTD values for the three markers using a common set of genera to determine the strength of the correlation in their divergence values. To determine if significant differences existed between markers, Kruskal-Wallis (KW) tests followed by a Dunn's posthoc were carried out in R ver 3.2.0 [42].

A similar analysis was conducted for each park community using RAxML-based trees to calculate MPD and MNTD partitioned by family. The percentage of congeners in the six large families with low MNTD (Asteraceae, Brassicaceae, Cyperaceae, Poaceae, Rosaceae, and Salicaceae) was evaluated for the datasets representing the three barcodes for the six biogeographic regions.

Taxonomic resolution. Our custom sequence library for Canadian plants was parsed based on the species present at each locality and the taxonomic resolution provided by each barcode was then evaluated using BLAST searches and by *mothur* in Qiime [43]. For both methods, the species known for each park were compared with the parent library to ascertain if barcode records allowed their identification to a family, genus, or species level. The resolution for species with multiple sequences was recorded as that where the taxonomic assignment for all individuals was consistent (e.g. if there were four sequences for species A and three were unambiguously identified to a species and one was to a genus, the recorded level of resolution would be to a genus). This approach generates a 'worst case' outcome for the capacity to identify a particular species. *Mothur* employs a distance matrix to assign a sequence (or cluster) to a species based on a parent library. For its use, identifications were predicted using a posterior probability cut-off of 0.95. We also report the true level of success of *mothur* by comparing the taxonomic identification assigned to a given sequence by *mothur* with its correct assignment. The data for each park was then used to generate a mean level of taxonomic resolution for each family, genus, and species. Data was checked for normality prior to conducting a Kruskal-Wallis (KW) test or one-way ANOVA to test for significant differences in species resolution among the three markers. Any significant test was followed up with the appropriate posthoc tests (Tukey's HSD for ANOVA or Dunn for KW). The parks were then subdivided into six biogeographic regions (Arctic, Atlantic, Boreal, Pacific, Prairies, Woodland) and the data was pooled for each region to estimate the mean level of taxonomic resolution for the floras that were examined. After checking for normality, KW or one-way ANOVA was used to test for a significant difference in species resolution among the regions for a particular barcode marker. We also evaluated taxonomic resolution for the 1074 species with data for all three barcode genes to compare the mean of the parks and the performance of different markers using an identical set of taxa. The performance of the barcodes for 25 families with the most species was then compared based on the BLAST results to identify groups where barcodes delivered low taxonomic resolution. All statistical tests were performed in R ver 3.2.0 [42] with Bonferroni error corrections for multiple tests (adjusted $p = 0.005$).

Results

Clustering and phylogenetic matrices

After the removal of identical sequences within any one species, the read library was reduced to 5919 sequences for *rbcL*, 2891 sequences for *matK*, and 4423 sequences for ITS2. The plastid markers were much less variable than ITS2 as evidenced when the read libraries were clustered at 100% identity which collapsed the sequence count when different species shared a particular

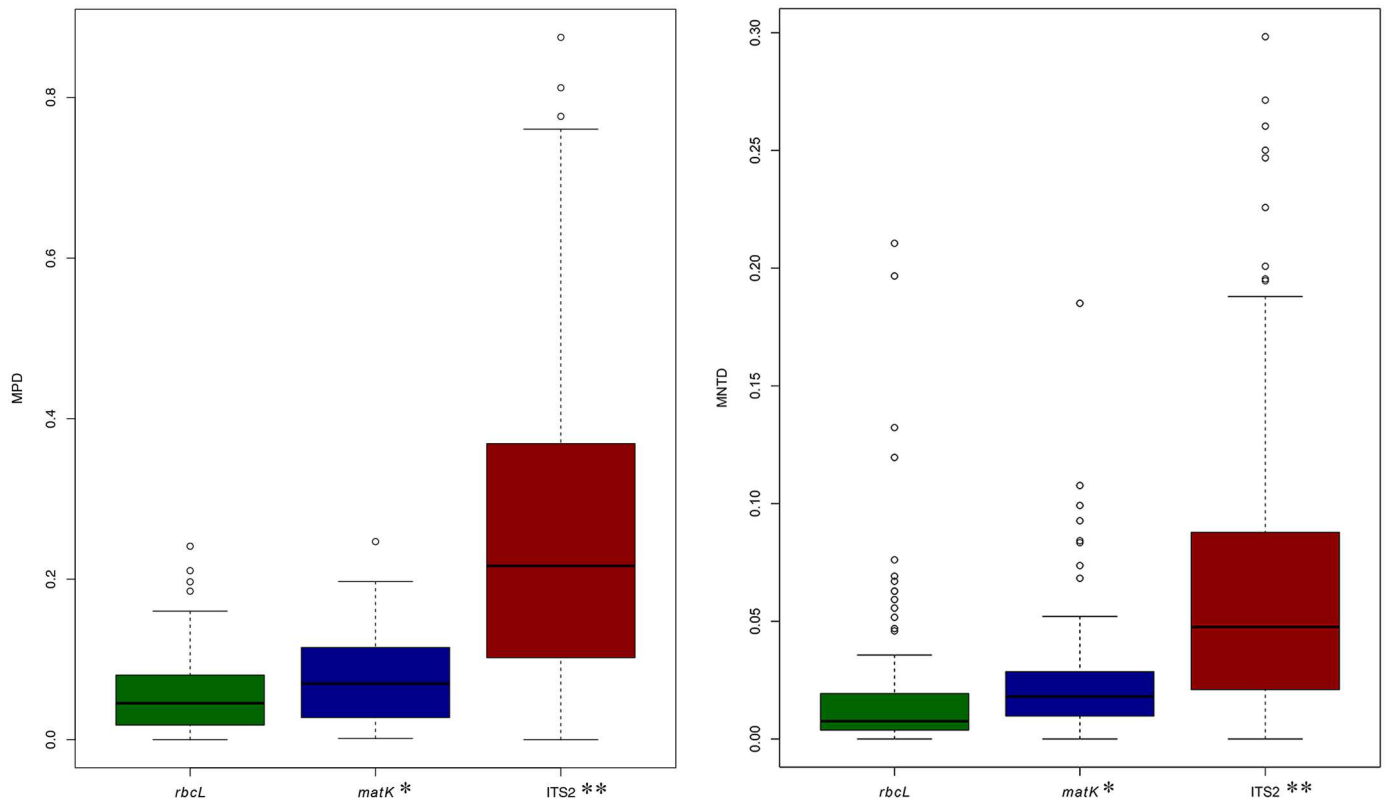


Fig 2. Boxplots of MPD and MNTD for *rbcL*, *matK*, and ITS2. Boxplots comparing MPD and MNTD for the vascular plant families of Canada for *rbcL*, *matK*, and ITS2. Significance (p -adjusted < 0.005) is indicated with an asterisk(s).

doi:10.1371/journal.pone.0169515.g002

sequence. This analysis showed that *rbcL* had considerably less sequence variation (2895 clusters; 5919 sequences) than *matK* (2145 clusters; 2891 sequences) while ITS2 was most diverse (4418 clusters; 4423 sequences). This pattern was reinforced by the global estimates for MPD and MNTD that rated ITS2 as the most variable marker followed by *matK* and *rbcL* (Fig 2, Table A and Table B in S1 File). For each measure, markers were significantly different from one another (KW and Dunn’s posthoc $p < 0.0005$).

The Asteraceae had low values for both metrics across all three barcodes (Table 2), but those for the family Salicaceae were exceptionally so for MPD (*rbcL* = 0.017; *matK* = 0.009; ITS2 = 0.036) and MNTD (*rbcL* = 0.0005; *matK* = 0.0007; ITS2 = 0.013). The latter result reflected the low MNTD values within *Salix* (41–90 species per region; *matK* and *rbcL* = 0.0005; ITS2 = 0.009; Table B in S1 File). The Asteraceae also had low MNTD (*rbcL* = 0.002; *matK* = 0.006; ITS2 = 0.021), strongly influenced by four genera with MNTD $< 2 \times 10^{-6}$. Interestingly, MPD did not predict low species resolution for Asteraceae (*rbcL* = 0.073; *matK* = 0.07; ITS2 = 0.373), because some long internal branches raised this measure (Table B in S1 File).

High phylogenetic diversity for families lacking genera with a low MNTD or MPD is a strong predictor of strong species resolution. For example, the Caryophyllaceae and Fabaceae have high MPD and MNTD for *rbcL* (MPD = 0.081, 0.160; MNTD = 0.008, 0.009 respectively; Table 2), but several of their genera have near zero values for both metrics (< 0.001 ; Table B in S1 File) suggesting that these lineages will have much lower species resolution than highly variable genera. By contrast, nearly all genera of the Orchidaceae and Primulaceae have high MPD

Table 2. The mean MPD and MNTD for 25 species-rich families with the number of sampled species.

Family	<i>rbcL</i>			<i>matK</i>			ITS2		
	No. of species	MPD	MNTD	No. of species	MPD	MNTD	No. of species	MPD	MNTD
Amaranthaceae	85	0.0889	0.0060	36	0.1395	0.0227	75	0.2660	0.0201
Apiaceae	98	0.0315	0.0037	59	0.0621	0.0106	66	0.3154	0.0469
Asteraceae	569	0.0729	0.0022	239	0.0699	0.0059	481	0.3732	0.0209
Boraginaceae	74	0.0579	0.0047	21	0.1065	0.0115	42	0.5789	0.0538
Brassicaceae	233	0.0320	0.0025	82	0.0700	0.0153	208	0.1887	0.0184
Caprifoliaceae	48	0.0756	0.0030	23	0.1138	0.0118	7	0.8123	0.3753
Caryophyllaceae	134	0.0805	0.0075	41	0.1841	0.0287	121	0.3109	0.0325
Cyperaceae	397	0.0329	0.0040	119	0.0552	0.0063	112	0.2724	0.0574
Ericaceae	71	0.0992	0.0120	54	0.1071	0.0135	61	0.1939	0.0236
Fabaceae	211	0.1602	0.0092	104	0.1291	0.0114	156	0.3646	0.0377
Juncaceae	69	0.0952	0.0040	39	0.1742	0.0092	19	0.2826	0.0081
Lamiaceae	92	0.0616	0.0036	22	0.0844	0.0203	36	0.5781	0.1954
Onagraceae	59	0.0558	0.0036	22	0.0543	0.0150	54	0.1015	0.0078
Orchidaceae	61	0.0690	0.0125	24	0.1148	0.0234	50	0.6388	0.0612
Orobanchaceae	79	0.0700	0.0077	24	0.0927	0.0274	49	0.3724	0.0516
Plantaginaceae	95	0.0942	0.0099	37	0.1654	0.0326	77	0.4472	0.0388
Poaceae	438	0.1185	0.0043	242	0.0963	0.0083	235	0.3941	0.0391
Polemoniaceae	43	0.0341	0.0044	7	0.0885	0.0288	39	0.1060	0.0165
Polygonaceae	85	0.0638	0.0053	25	0.1537	0.0521	41	0.5349	0.1005
Primulaceae	40	0.0906	0.0116	16	0.1063	0.0196	37	0.2703	0.0496
Ranunculaceae	120	0.0943	0.0057	39	0.1539	0.0181	93	0.3356	0.0341
Rosaceae	276	0.1021	0.0030	128	0.1183	0.0059	200	0.3677	0.0340
Salicaceae	100	0.0172	0.0005	82	0.0093	0.0007	46	0.0362	0.0128
Saxifragaceae	72	0.0635	0.0057	42	0.1318	0.0158	63	0.3226	0.0355
Violaceae	35	0.0127	0.0026	12	0.0310	0.0138	29	0.1021	0.0262
Top 25 Mean	143.36	0.0710	0.0056	61.56	0.1045	0.0171	95.88	0.3427	0.0559
All Families Mean	28.51	0.0552	0.0180	16.95	0.0770	0.0249	24.54	0.2823	0.0987

doi:10.1371/journal.pone.0169515.t002

and MNTD, ensuring high species resolution (see Table B in [S1 File](#)). Species resolution is also strong for the Lamiaceae (ITS2), Onagraceae (*matK*), and Polygonaceae (ITS2 and *matK*) ([Table 2](#)) due to their high genetic diversity. There was no correlation between the number of species in a family or genus and either MPD or MNTD ($r^2 < 0.05$ for all comparisons). There was also no correlation between markers for MNTD values ($r^2 < 0.007$ for all comparisons; [S1 Fig](#)).

MPD and MNTD were used to predict those parks and biogeographic regions where DNA barcodes would deliver poor taxonomic resolution. Both values were generally lower in the Arctic than in the other biogeographic regions for all three markers (see Table C in [S1 File](#) for details), suggesting that species resolution should be most challenging in the north ([Table C](#) in [S1 File](#)). These estimates of genetic diversity further predict that ITS2 will deliver the best taxonomic resolution followed by *matK* and *rbcL*.

Taxonomic resolution

Overall. Performance comparison of BLAST and mothur in identifying plants from the 28 localities ([Table 3](#); [Fig 1](#)) indicated that BLAST delivered higher species resolution for all three barcodes ([Fig 3a–3c](#)). When employing a posterior probability cut-off of 0.95, mothur

Table 3. Level of species resolution (%) for each barcode for BLAST and mothur. For mothur, species resolution is reported for both a posterior probability cut-off (0.95) and the true level of resolution.

Region	Park	Blast			Mothur PP			Mothur Actual		
		<i>rbcL</i>	<i>matK</i>	<i>ITS2</i>	<i>rbcL</i>	<i>matK</i>	<i>ITS2</i>	<i>rbcL</i>	<i>matK</i>	<i>ITS2</i>
Arctic	Ellesmere Island NP (135, 107, 84)	31.85	59.81	66.67	42.96	48.60	64.29	17.04	45.79	67.86
Arctic	Ivvavik NP (399, 292, 298)	34.59	66.78	69.46	36.09	55.14	63.09	24.81	51.71	68.46
Arctic	Nahannii NP (643, 456, 450)	36.86	72.59	69.11	36.39	54.82	60.22	25.51	55.92	65.56
Arctic	Tornat Mountains NP (364, 260, 252)	39.84	72.31	66.67	38.19	53.46	61.11	25.27	55.38	63.89
Arctic	Ukkusiksalik NP (149, 125, 111)	38.00	67.20	69.37	36.00	56.00	69.37	24.00	53.60	71.17
Arctic	Wapusk NP (269, 224, 225)	35.69	74.11	68.89	37.55	57.59	62.67	27.14	58.48	70.22
Atlantic	Forillon NP (611, 420, 355)	47.46	79.05	70.70	37.97	58.81	62.82	32.90	64.76	67.04
Atlantic	Fundy NP (714, 442, 406)	49.02	83.71	74.63	38.80	59.05	66.50	35.15	65.16	67.98
Atlantic	Kejimikujik NP (561, 332, 319)	52.76	86.75	75.55	41.35	62.35	64.58	40.64	68.98	65.83
Atlantic	Prince Edward Island NP (633, 414, 365)	51.82	86.47	76.16	39.97	61.84	63.84	39.18	67.87	67.40
Boreal	La Mauricie NP (450, 295, 220)	51.33	83.39	70.91	42.00	56.95	65.91	38.00	63.39	68.18
Boreal	Mingan Archipelago NP (432, 299, 285)	46.99	80.94	73.33	37.50	61.87	65.96	31.94	65.55	70.88
Boreal	Pukaskwa NP (532, 379, 296)	46.05	81.53	70.95	37.59	56.46	62.50	31.58	59.89	68.58
Boreal	Terra Nova NP (495, 297, 281)	49.29	85.52	71.89	37.98	64.98	66.19	33.94	68.01	68.33
Pacific	Banff NP (886, 561, 606)	38.37	79.14	70.30	33.97	55.44	60.23	25.96	58.29	65.18
Pacific	Glacier NP (618, 418, 404)	41.68	78.47	70.30	35.54	54.78	63.86	26.82	61.72	66.58
Pacific	Mount Revelstoke NP (427, 296, 271)	40.98	79.39	70.85	35.83	55.41	63.47	29.98	62.50	67.16
Pacific	Pacific Rim NP (423, 285, 258)	54.61	85.96	82.56	35.93	60.35	74.42	35.46	69.12	75.19
Pacific	Yoho NP (643, 440, 436)	39.50	81.63	72.02	33.59	56.92	64.91	27.84	59.86	66.74
Prairies	Elk Island NP (476, 345, 299)	41.81	79.71	73.58	32.77	53.62	59.87	29.41	57.68	69.57
Prairies	Grasslands NP (408, 233, 278)	41.91	86.70	70.14	31.37	52.79	62.23	29.66	58.80	67.27
Prairies	Prince Albert NP (623, 434, 410)	43.02	82.49	72.44	33.07	55.76	61.22	30.02	60.14	70.24
Prairies	Riding Mountain NP (696, 473, 443)	45.55	83.09	75.62	33.62	57.72	62.75	30.32	60.47	69.53
Prairies	Waterton Lakes NP (949, 568, 632)	40.67	81.34	72.94	34.25	54.23	61.55	26.98	59.86	67.41
Woodland	1000 Islands NP (1583, 880, 896)	52.18	87.39	79.02	38.66	60.68	65.63	36.39	67.05	73.10
Woodland	Bruce Peninsula NP (859, 580, 481)	49.71	86.38	75.88	39.12	60.17	66.11	36.55	67.76	70.69
Woodland	Koffler Scientific Reserve (612, 469, 320)	53.92	86.35	73.13	39.05	62.26	64.69	40.85	67.59	68.13
Woodland	Point Pelee NP (832, 513, 477)	53.13	87.52	76.52	39.90	63.16	65.83	37.38	69.01	71.07
Parks mean		44.59	80.20	72.48	37.04	57.54	64.14	31.10	61.58	68.54

doi:10.1371/journal.pone.0169515.t003

underestimated the capacity to make species-level identifications, but overestimated it at a genus level Table D in [S1 File](#)). Both BLAST and mothur indicated that *rbcL* has the lowest species (45% and 31% with BLAST and mothur respectively) and generic (91% and 84% with BLAST and mothur respectively) resolution (Table D in [S1 File](#)), but diverged on which marker provides the highest species resolution. BLAST generates the highest species resolution with *matK* (80%) followed by *ITS2* (73%). By comparison, mothur ranks *ITS2* as the best barcode when resolving taxa with both posterior probability (*ITS2* mean = 64% vs. *matK* mean = 58%) and true species resolution (69% versus 62%). Generic resolution was high for both *matK* (~96–98%) and *ITS2* (96–99%) using either approach (Table 3). The difference in species resolution was significant between markers for both algorithms ($p < 0.005$; Fig 3a–3c). Analysis of the dataset consisting of 1074 species represented by all three barcodes generated similar results to the park data (Table D in [S1 File](#)). Since BLAST yielded the highest species resolution for each marker, these results were employed for the further analyses.

Species resolution by family. For most families, *matK* delivered the highest resolution followed by *ITS2*, but the two gene regions were complementary, jointly delivering 85% species

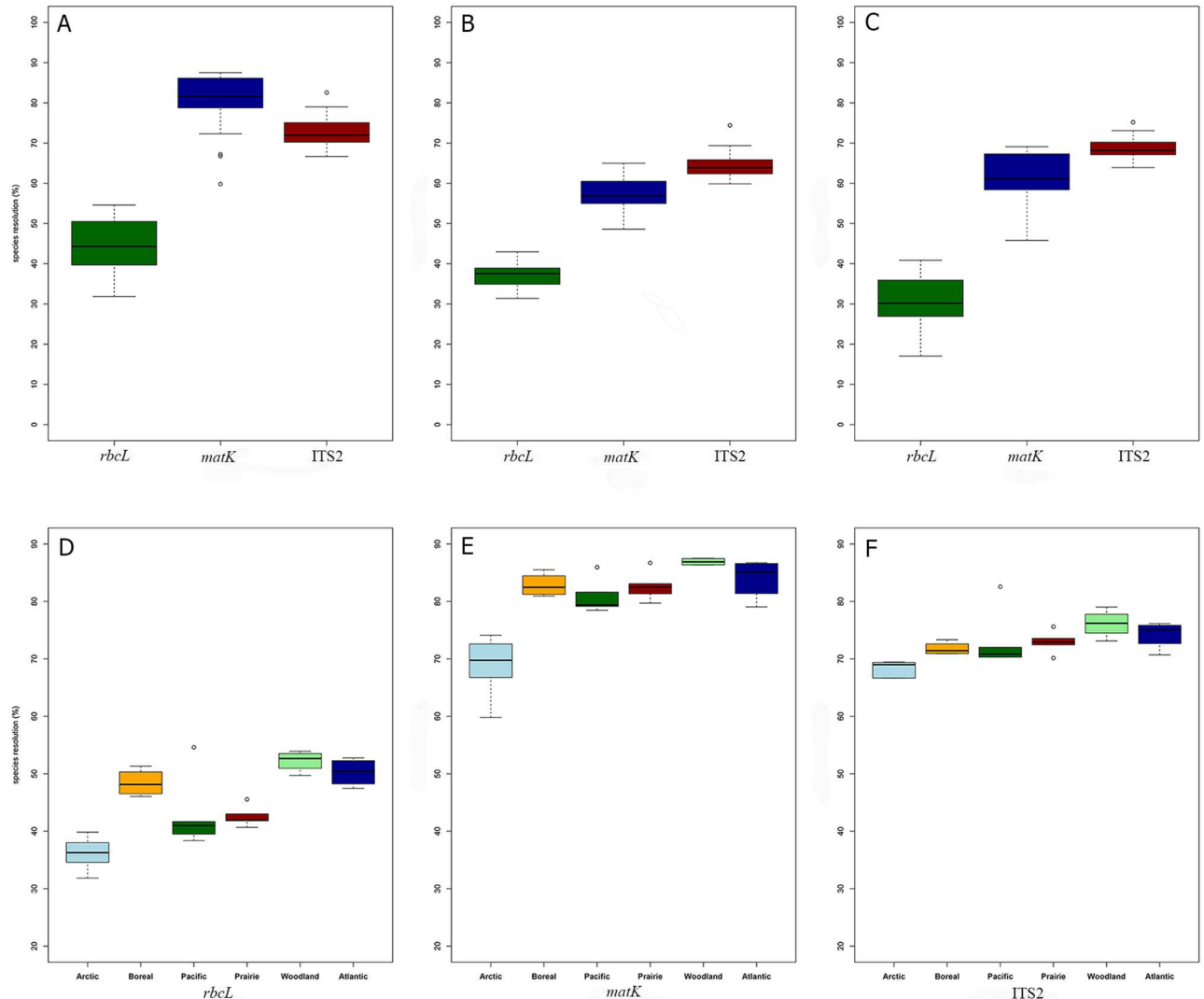


Fig 3. Species resolution for the three DNA barcodes (*rbcL*, *matK*, and ITS2). Species resolution for the three DNA barcodes (*rbcL*, *matK*, and ITS2) based on A) BLAST, B) mother with a posterior probability cut-off 0.95 or C) the actual species resolution of mother. Species resolution in the six biogeographic regions obtained with D) *rbcL*, E) *matK*, F) ITS2.

doi:10.1371/journal.pone.0169515.g003

resolution if two families were excluded (Salicaceae, Asteraceae) (Table 4; Fig 4). In fact, *matK* delivered perfect resolution for four families (Onagraceae, Polemoniaceae, Boraginaceae, Caprifoliaceae), while ITS2 did well for Lamiaceae (98%) and Orchidaceae (92%). By comparison, *rbcL* had low species resolution (<60%) for all families except Orchidaceae (78%), Ericaceae (65%), Plantaginaceae (64%), Primulaceae (71%), and Saxifragaceae (69%). Generic resolution was high for *matK* (98%) and ITS2 (97%) but slightly lower for *rbcL* (91%). Families with compromised generic resolution included the Asteraceae (*rbcL* = 78%; *matK* = 97%; ITS2 = 92%), Fabaceae (*rbcL* = 81%; *matK* = 93%; ITS2 = 89%) and Poaceae (*rbcL* = 82%; *matK* = 95%; ITS2 = 95%) (Table 4; Fig 4).

Consistent with their low values for MNTD and MPD, species resolution was poor for the Salicaceae (<31%) and Asteraceae (<68%). Their low MPD and MNTD values also predicted that certain genes would fail to distinguish species of Violaceae (*rbcL*), Rosaceae (*matK*), and Onagraceae (ITS2). When accounting for genera with low MPD and MNTD within families,

Table 4. Percentage of taxonomic resolution by BLAST to family, genus and species. Taxonomic resolution for *rbcl*, *matK*, and ITS2 for 25 species-rich families.

Blast Family	<i>rbcl</i>			<i>matK</i>			ITS2		
	Family	Genus	species	Family	Genus	species	Family	Genus	species
Amaranthaceae	100.00%	96.48%	50.70%	100.00%	100.00%	91.09%	100.00%	95.65%	54.35%
Apiaceae	100.00%	87.34%	56.54%	100.00%	97.86%	94.12%	100.00%	97.62%	86.19%
Asteraceae	100.00%	77.97%	26.39%	99.76%	96.76%	67.17%	100.00%	91.50%	58.37%
Boraginaceae	100.00%	88.89%	52.59%	100.00%	100.00%	100.00%	100.00%	100.00%	83.56%
Brassicaceae	100.00%	90.29%	33.44%	100.00%	100.00%	86.60%	100.00%	91.81%	49.67%
Caprifoliaceae	100.00%	85.23%	53.69%	100.00%	100.00%	100.00%	100.00%	100.00%	89.66%
Caryophyllaceae	100.00%	81.49%	33.63%	100.00%	100.00%	92.83%	100.00%	95.61%	67.80%
Cyperaceae	99.94%	95.15%	30.08%	100.00%	97.29%	85.67%	100.00%	100.00%	89.19%
Ericaceae	100.00%	99.62%	64.82%	100.00%	100.00%	93.22%	100.00%	94.68%	72.22%
Fabaceae	95.66%	81.07%	41.42%	100.00%	92.48%	72.09%	100.00%	89.39%	70.88%
Juncaceae	100.00%	99.08%	54.29%	100.00%	100.00%	86.05%	100.00%	100.00%	88.81%
Lamiaceae	100.00%	88.93%	56.15%	100.00%	100.00%	88.04%	100.00%	99.04%	98.08%
Onagraceae	100.00%	99.55%	49.77%	100.00%	100.00%	100.00%	100.00%	100.00%	61.75%
Orchidaceae	100.00%	89.08%	77.59%	100.00%	100.00%	83.70%	100.00%	96.12%	92.24%
Orobanchaceae	100.00%	99.44%	52.54%	100.00%	100.00%	76.40%	100.00%	100.00%	79.35%
Plantaginaceae	100.00%	98.64%	64.07%	100.00%	100.00%	91.88%	100.00%	100.00%	83.54%
Poaceae	100.00%	81.48%	39.89%	100.00%	95.00%	75.25%	100.00%	95.23%	66.84%
Polemoniaceae	100.00%	96.77%	29.03%	100.00%	100.00%	100.00%	100.00%	100.00%	75.86%
Polygonaceae	92.61%	89.79%	48.59%	100.00%	100.00%	94.71%	100.00%	100.00%	88.67%
Primulaceae	100.00%	99.17%	71.07%	100.00%	100.00%	88.16%	100.00%	100.00%	63.11%
Ranunculaceae	100.00%	97.97%	54.07%	100.00%	100.00%	85.37%	100.00%	100.00%	81.75%
Rosaceae	100.00%	88.78%	39.86%	100.00%	100.00%	80.37%	100.00%	99.84%	85.02%
Salicaceae	100.00%	99.76%	12.24%	100.00%	100.00%	26.47%	100.00%	100.00%	30.96%
Saxifragaceae	100.00%	95.06%	68.72%	100.00%	97.69%	91.33%	100.00%	100.00%	89.24%
Violaceae	100.00%	100.00%	30.41%	100.00%	100.00%	90.00%	100.00%	100.00%	69.78%

doi:10.1371/journal.pone.0169515.t004

low species resolution was apparent for Fabaceae (*rbcl* and *matK*) and Caryophyllaceae (*rbcl*), while resolution was high for Orchidaceae (*rbcl*) and Primulaceae (*rbcl*). The lack of low resolution genera in the Polygonaceae (*matK* and ITS2), Caprifoliaceae (ITS2 and *matK*), Polemoniaceae (*matK*), and Lamiaceae (ITS2) accounts for the relatively high success of barcoding in these taxa (Table B in S1 File; Fig 4).

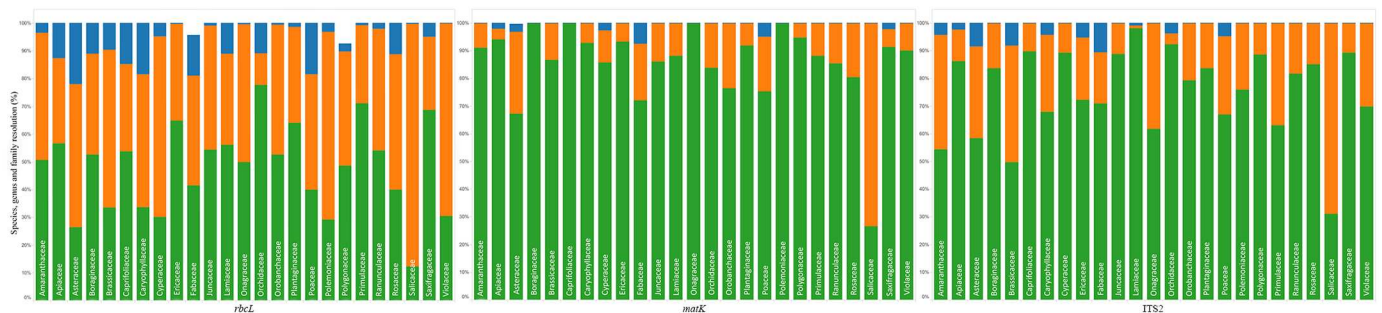


Fig 4. Level of taxonomic resolution provided by *rbcl*, *matK* or ITS2 for 25 families. Level of taxonomic resolution provided by *rbcl*, *matK* or ITS2 for 25 families of vascular plant that are species-rich in Canada. The three colours show the proportion of species identified to a family (blue), genus (orange) or species (green) level.

doi:10.1371/journal.pone.0169515.g004

Species resolution by region. When the 28 localities were organized into six biogeographic regions (Table 1), Arctic sites had significantly lower levels of species resolution than those in the other five regions ($p < 0.05$ for all markers; Fig 3d–3f) although results varied by marker. For *rbcl*, the Atlantic and Woodland regions have significantly higher taxonomic resolution than all others ($p < 0.005$) while Boreal and Pacific communities have significantly higher taxonomic resolution than Arctic and Prairie assemblages ($p < 0.005$). For *matK*, only Arctic communities have significantly lower species resolution ($p < 0.005$) than the other localities. Atlantic and Woodland have significantly higher resolution with ITS2 than the other regions ($p < 0.005$; Fig 3d–3f), as predicted by MPD and MNTD.

Discussion

This study examined the effectiveness of DNA barcoding in the identification of plants from six biogeographic regions of Canada using both local alignment (BLAST) and phylogeny-based (MNTD and MPD) approaches. MPD and MNTD were first proposed as measures of phylogenetic diversity within a community [38], and have commonly been used to study community assembly [44; 45; 46]. MPD was previously used to compare substitution rates among plant families for three barcode regions (*rbcl*, *matK*, ITS2), and a positive correlation was reported between these rates and their capacity to discriminate species [26]. The present study extended this work by examining the utility of MPD and MNTD as predictors of species resolution for the same three gene regions.

Both MPD and MNTD indicated that ITS2 should deliver the best species resolution, an expected result given the higher rates of nucleotide substitution in nuclear than organellar genomes of plants [47; 48]. The prediction was supported when *mothur* was used to generate taxonomic assignments, but *matK* delivered the best species resolution with BLAST. Interestingly, BLAST yielded higher species resolution than *mothur* for all three markers, a result which was maintained even when analysis was restricted to the 1074 species with sequence data for all three regions. BLAST's higher resolution is explained by its greater sensitivity to sequence length [49], as well its inclusion of indel variation, which phylogenetic approaches typically overlook. Although *matK* has less sequence variation than ITS2, it contains more indels which helped it to achieve higher species resolution with BLAST. Our results support the need for DNA barcoding to utilize phylogenetic methods that incorporate indels to maximize the resolving power of a given marker.

The genome compartment exposed to the highest intraspecific gene flow is generally the best suited for making species assignments because it reduces the likelihood that introgressed alleles will gain establishment and blur species diagnosis. Gene flow raises effective population size, reducing exposure to genetic drift, diminishing the chance of introgressed alleles gaining fixation in the gene pool, and increasing the probability that a particular gene will track species relationships [50]. Since the nuclear genome tends to experience greater dispersal and gene flow than the plastid genome, nuclear markers are generally more effective in species diagnosis than their plastid counterparts [50; 51]. Hence, the incorporation of a nuclear marker with the core (plastid) barcodes offers the advantage of compensating for situations where plastid markers fail to provide resolution. ITS2 did outperform its plastid counterparts in several species-rich families (i.e. Lamiaceae, Poaceae, Cyperaceae, and Saxifragaceae; Table 4) examined in this study. However, it was less effective in other families, likely reflecting incomplete lineage sorting stemming from its larger effective population sizes or, in rare cases, when plastid dispersal exceeds that of the nucleus [50; 51]. Despite these situations, the incorporation of ITS2 is preferable over additional plastid markers such as *psbA-trnH* because it occurs in all plant species which is not true for any plastid marker (including *rbcl* and *psbA-trnH*) and

existing primers are nearly universal (as opposed to those for *matK*). Moreover, it delivers high resolution despite its short length (~350 bp), making it an ideal marker for studies using high-throughput sequencing platforms which cannot recover full-length sequences for longer barcodes such as *matK* (~800 bp) or those variable in length (*psbA-trnH*; 50–1000 bp).

The observed differences in taxonomic resolution for the three barcodes are undoubtedly influenced by selection, species demography, hybridization, lineage sorting, and phylogeographic structure (reviewed by [52]). The higher resolution of *matK* compared to *rbcl* reflects the different selective pressures acting on these genes. Because it encodes the large subunit of RuBisCo which has an essential role in photosynthesis, *rbcl* is under strong purifying selection in autotrophic plants [53] reducing its rate of evolution and constraining its utility for distinguishing closely related species. By contrast *matK*, an intron maturase involved in the splicing of group IIa introns, appears to be under relaxed purifying selection as evidenced by nearly equal substitution rates for all three coding positions [53; 54; 55; 56]. The relatively high rates of nucleotide substitution in *matK* compared to other plastid genes is useful for species delimitation, but a lack of conserved priming sites often undermines sequence recovery. Nuclear markers have a larger effective population size than plastid markers and tend to evolve more rapidly [47; 48]. The higher rates of nucleotide substitution and dispersal in plant nuclear genomes support their inclusion for plant DNA barcoding [51]. Additionally, the presence of multiple alleles for nuclear genes makes it possible to identify hybrids. Currently the only plant nuclear locus that meets barcoding criteria is ITS2 (see above) and its inclusion adds depth to barcode reference libraries by tracking a different genomic compartment.

The present analysis shows that MPD and MNTD are strong predictors of barcode resolution, identifying families and genera where taxonomic resolution is low. They were particularly useful in revealing genera with low resolution in families where divergences are high. For example, the Fabaceae has a relatively high MPD for both *rbcl* and *matK* but low species resolution, reflecting its inclusion of several genera (e.g. *Lupinus*, *Oxytropis*) with low MNTD. The latter genera explain the lower than average species resolution in this family for all markers, but this outcome was especially surprising for *Lupinus* because it was previously observed to have high genetic diversity in North America and low genetic diversity in the Andes due to a recent adaptive radiation [57]. Further research is needed to determine if a similar radiation occurred in North America. MNTD is a better predictor of species resolution than MPD because it quantifies the distance between pairs of closely related species and it is also less influenced by polytomies than MPD or PD [40]. As such, it is a better estimator of the efficacy of a given DNA barcode. The low correlation between MNTD values for the three barcode regions in different genera implies they are evolving independently (S1 Fig). As a consequence, the use of multiple barcode markers consistently improves taxonomic resolution because a particular marker can compensate for the deficits in resolution of its counterparts [7; 9; 10; 15; 17; 18; 19; 20; 21; 22; 23; 24; 25; 26]. This complementarity supports the use of specific barcodes that optimize species resolution for different groups [58; 59].

The patterns of variation in the phylogenetic matrices (MPD, MNTD) agree with the earlier conclusion [15; 16] that ITS2 has higher discriminatory power than *matK* or *rbcl* when specimens are analyzed against a local reference library. However, this conclusion may not extend to other situations, such as the present study, where taxonomic resolution is compared against a more comprehensive parent library, an approach which provides a 'real world' outcome of DNA barcoding. For example, Burgess et al. [23] reported 88% species resolution with *matK* and 80% with *rbcl* when identifications were driven by a barcode library comprised solely of plant species known from the site. Analysis of the same community with the barcode library for all Canadian plants lowered species resolution (i.e. 86% for *matK*, 54% for *rbcl*) but with the advantage that newly encountered plants would potentially be identified.

The low levels of sequence variation in several plant families likely reflects the joint impacts of polyploidization, hybridization, phylogeographic effects such as allele surfing during range expansions, and demographic effects including bottlenecks which reduce intra- and inter-specific variation [52; 60; 61]. These effects are more prominent in Arctic communities that might explain the frequent failure of both nuclear and plastid markers in discriminating species in this region [52; 62; 63; 64; 65]. Although these processes (singly or in combination) compromise the effectiveness of DNA barcoding in discriminating plant species, they do provide an opportunity to understand the factors that shape plant populations and genomes. While our dataset lacks the extensive sampling needed to differentiate between these processes, it does reveal taxa that require further study.

Among the 171 plant families in Canada, Salicaceae has the lowest species discrimination, largely due to the very limited genetic diversity among the 90 species of *Salix*. Its lack of variation in seven regions of the plastid genome was linked to frequent hybridization, incomplete lineage sorting, or repeated plastid capture events [66]. However, the same lack of resolution was observed with our ITS2 data and in a more detailed analysis of 22 species [26], indicating that the lack of divergence extends into the nuclear genome. This difficulty in differentiating *Salix* species using molecular markers may reflect hybridization, introgression, recent speciation, allele surfing via range expansion, and low rates of molecular evolution [52; 59; 67; 68; 69]. More extensive phylogenetic studies targeting nuclear markers or whole plastid genomes are necessary to clarify the processes driving the unusually low divergences in *Salix*.

The Asteraceae is another family where DNA barcodes deliver poor species resolution, but the underlying factors differ from those in the Salicaceae. The Asteraceae is a species-rich group that lacks reciprocal monophyly between some closely related species and genera [70], taxa that are difficult to differentiate with molecular data [71]. The Fabaceae also showed poor species resolution with all three loci (41–72%), an expected result given the number of poorly resolved genera in this family [72]. Although species resolution in certain groups may never be resolved by the targeted analysis of a few barcode loci, they do represent interesting models for testing the effectiveness of whole plastid genomes as a tool for species discrimination (reviewed by [58]).

When comparisons were extended across the six biogeographic regions of Canada, DNA barcoding delivered the poorest species discrimination in the Arctic, perhaps reflecting the higher incidence of congeners in Arctic communities (47–53%) than in other regions (39–48%; S2 Fig). As well, the arctic flora is rich in recently radiated congeners that have not achieved reciprocal monophyly [63; 65]. As a consequence, DNA barcoding delivered higher species resolution in the more floristically diverse regions. In fact, the most floristically diverse regions (Atlantic and Woodland) had the highest species resolution for all markers, suggesting that increased species diversity is correlated with genetic diversity. This difference likely reflects both demographic effects and shifts in community composition. For example, species of *Salix* comprise 8% of the flora near Churchill, reducing the overall success of barcoding at this arctic site [26]. We do not observe similar compositional biases in more floristically diverse communities that would influence overall barcoding success.

The present study has established that DNA barcoding delivers approximately 80% species resolution for plant communities in the temperate regions of Canada when either *matK* or ITS2 are employed, meaning that DNA barcoding can provide a standardized, rapid approach for ecological surveys in these settings. The same gene regions deliver near-perfect resolution to a generic level, a level of taxonomic placement useful for characterizing both past [73; 74] and present [11; 75] plant communities, for forensic applications [76; 77; 78], for validating the accuracy of specimen identifications in herbaria [28; 79], and for assessing herbivore diets [11; 80; 81; 82; 83]. The present study also demonstrates the ability of DNA barcoding to

deliver particularly high levels of taxonomic resolution when comprehensive reference libraries are available for *matK* and ITS2, providing motivation for efforts to extend coverage for these genes.

Conclusions

Comprehensive sampling (~ 96% taxonomic coverage) of the Canadian flora provided a unique opportunity to test the efficacy of DNA barcoding across a diverse set of communities. Analyses based on this library indicate that any one of the three barcode regions is very effective (>90%) in delivering a generic assignment while species resolution is often possible with ITS2 (72%) and *matK* (80%). BLAST demonstrated higher performance than mothur in assigning specimens to a species in all datasets, including those at a community level and for 1074 species with data for all three barcode regions. The higher performance of BLAST reflects its consideration of indel variation and absolute length of the marker, leading *matK* to deliver the highest resolution. Although ITS2 showed slightly lower performance, it has two important advantages; its short length makes it suitable for HTS-based applications, and it is readily recovered from diverse taxa, including vascular plants and fungi.

Supporting Information

S1 Fig. MNTD values for the three barcodes for genera. Comparison of MNTD values for the three barcode regions for genera of Canadian vascular plants. A) Three-dimensional scatter plot of 243 genera; B) Three-dimensional scatter plot of a subset of 171 genera with low MNTD values. The r^2 is less than 0.007 for all comparisons.

(PDF)

S2 Fig. The percentage of congeners for the six most species-rich families with low MNTD. The percentage of congeners for the six most species-rich families with low MNTD by barcode and region.

(PDF)

S1 File. A supplemental file containing four tables. Raw MPD and MNTD values for the vascular plant families of Canada (Table A). Raw MPD and MNTD values for the vascular plant genera of Canada (Table B). The biogeographic region, number of families, MPD, and MNTD for the 28 Canadian localities employed as a basis to test barcode resolution (Table C). The mean taxonomic resolution to family, genus, and species for all 28 localities employed as a basis to test barcode resolution and for the subset of 1074 species with sequence data for all three barcode regions (Table D).

(XLSX)

Acknowledgments

We thank two reviewers for helpful comments on an earlier version of this paper. We extend special gratitude to David Erikson and Ina Anreiter for friendly suggestions and comments during the preparation of our manuscript. We also thank Bruce Bennet and Sergei Ponomarenko for providing updated species lists for several national parks. This research was funded through the International Barcode of Life project supported by the Genome Canada through Ontario Genomics, Canada Foundation for Innovation, and by the Ontario Ministry for Research and Innovation. It is also a contribution to the Food From Thought research program supported by the Canada First Research Excellence Fund. The funders

had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: TWAB MLK PDNH EVZ.

Data curation: TWAB MLK JS.

Formal analysis: TWAB.

Funding acquisition: PDNH.

Investigation: TWAB MLK.

Methodology: TWAB MLK.

Supervision: EVZ PDNH.

Validation: TWAB MLK JS.

Visualization: TWAB MLK.

Writing – original draft: TWAB MLK.

Writing – review & editing: TWAB MLK PDNH EVZ.

References

1. Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proc Biol Sci.* 2003; 270(1512): 313–321 doi: [10.1098/rspb.2002.2218](https://doi.org/10.1098/rspb.2002.2218) PMID: [12614582](https://pubmed.ncbi.nlm.nih.gov/12614582/)
2. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLOS One* 2011; 6(5): e19254 doi: [10.1371/journal.pone.0019254](https://doi.org/10.1371/journal.pone.0019254) PMID: [21637336](https://pubmed.ncbi.nlm.nih.gov/21637336/)
3. Santamaria M, Vicario S, Pappadà G, Scioscia G, Scazzocchio C, Saccone C. Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. *BMC Bioinformatics.* 2009; 10 Suppl 6: S15
4. Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G. A proposal for a standardised protocol to barcode all land plants. *Taxon.* 2007; 56(2): 295–299
5. CBOL Plant Working Group. A DNA barcode for land plants. *Proc Natl Acad Sci USA.* 2009; 106(31): 12794–12797 doi: [10.1073/pnas.0905845106](https://doi.org/10.1073/pnas.0905845106) PMID: [19666622](https://pubmed.ncbi.nlm.nih.gov/19666622/)
6. Seberg O, Petersen G. How many loci does it take to DNA barcode a crocus? *PLOS One.* 2009; 4(2): e4598 doi: [10.1371/journal.pone.0004598](https://doi.org/10.1371/journal.pone.0004598) PMID: [19240801](https://pubmed.ncbi.nlm.nih.gov/19240801/)
7. Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, et al. Identification of Amazonian trees with DNA barcodes. *PLOS One.* 2009; 4(10): e7483 doi: [10.1371/journal.pone.0007483](https://doi.org/10.1371/journal.pone.0007483) PMID: [19834612](https://pubmed.ncbi.nlm.nih.gov/19834612/)
8. Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, et al. Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. *PLOS One.* 2010; 5(10): e13674 doi: [10.1371/journal.pone.0013674](https://doi.org/10.1371/journal.pone.0013674) PMID: [21060687](https://pubmed.ncbi.nlm.nih.gov/21060687/)
9. Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, et al. How effective are DNA barcodes in the identification of African rainforest trees? *PLOS One.* 2013; 8(4): e54921 doi: [10.1371/journal.pone.0054921](https://doi.org/10.1371/journal.pone.0054921) PMID: [23565134](https://pubmed.ncbi.nlm.nih.gov/23565134/)
10. Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL, Bull RD. DNA barcoding the Canadian Arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLOS One.* 2013; 8(10): e77982 doi: [10.1371/journal.pone.0077982](https://doi.org/10.1371/journal.pone.0077982) PMID: [24348895](https://pubmed.ncbi.nlm.nih.gov/24348895/)
11. Bello A, Daru BH, Stirton CH, Chimphango SBM, van der Bank M, Maurin O, Muasya AM. DNA barcodes reveal microevolutionary signals in fire response trait in two legume genera. *AoB Plants.* 2015; 7: 124
12. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, et al. Land plants and DNA barcodes: short-term and long-term goals. *Phil Trans R Soc Lond B Biol Sci.* 2005; 360(1462): 1889–1895

13. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA*. 2005; 102(23): 8369–8374 doi: [10.1073/pnas.0503123102](https://doi.org/10.1073/pnas.0503123102) PMID: [15928076](https://pubmed.ncbi.nlm.nih.gov/15928076/)
14. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: evaluation of *rbcL* in a multi-gene tiered approach. *Can J Bot*. 2006; 84(3): 335–341
15. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLOS One*. 2010; 5(1): e8613 doi: [10.1371/journal.pone.0008613](https://doi.org/10.1371/journal.pone.0008613) PMID: [20062805](https://pubmed.ncbi.nlm.nih.gov/20062805/)
16. Li D–Z, Gao L–M, Li H–T, Wang H, Ge X–J, Liu J–Q, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci USA*. 2011; 108(49): 19641–19646 doi: [10.1073/pnas.1104551108](https://doi.org/10.1073/pnas.1104551108) PMID: [22100737](https://pubmed.ncbi.nlm.nih.gov/22100737/)
17. Kress WJ, Erickson DL. A two–locus global DNA barcode for land plants: The coding *rbcL* gene complements the non–coding *trnH–psbA* spacer region. *PLOS One*. 2007; 2(6): e508 doi: [10.1371/journal.pone.0000508](https://doi.org/10.1371/journal.pone.0000508) PMID: [17551588](https://pubmed.ncbi.nlm.nih.gov/17551588/)
18. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, et al. DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA*. 2008; 105(8): 2923–2928 doi: [10.1073/pnas.0709936105](https://doi.org/10.1073/pnas.0709936105) PMID: [18258745](https://pubmed.ncbi.nlm.nih.gov/18258745/)
19. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLOS One*. 2008; 3(7): e2802 doi: [10.1371/journal.pone.0002802](https://doi.org/10.1371/journal.pone.0002802) PMID: [18665273](https://pubmed.ncbi.nlm.nih.gov/18665273/)
20. Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjurjo O, et al. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc Natl Acad Sci USA*. 2009; 106(44): 18621–18626 doi: [10.1073/pnas.0909820106](https://doi.org/10.1073/pnas.0909820106) PMID: [19841276](https://pubmed.ncbi.nlm.nih.gov/19841276/)
21. Starr JR, Naczi RFC, Chouinard BN. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Mol Ecol Res*. 2009; 9 Suppl S1: 151–163
22. Clerc–Blain JLE, Starr JR, Bull RD, Saarela JM. A regional approach to plant DNA barcoding provides high species resolution of sedges (*Carex* and *Kobresia*, Cyperaceae) in the Canadian Arctic Archipelago. *Mol Ecol Resour*. 2010; 10(1): 69–91 doi: [10.1111/j.1755-0998.2009.02725.x](https://doi.org/10.1111/j.1755-0998.2009.02725.x) PMID: [21564992](https://pubmed.ncbi.nlm.nih.gov/21564992/)
23. 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. *Methods Ecol Evol*. 2011; 2(4): 333–340
24. 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(11): e26841 doi: [10.1371/journal.pone.0026841](https://doi.org/10.1371/journal.pone.0026841) PMID: [22096501](https://pubmed.ncbi.nlm.nih.gov/22096501/)
25. deVere N, Rich TCG, Ford CR, Trinder SA, Long C, Moore CW, et al. DNA barcoding the native flowering plants and conifers of Wales. *PLOS One*. 2012; 7(6): e37945 doi: [10.1371/journal.pone.0037945](https://doi.org/10.1371/journal.pone.0037945) PMID: [22701588](https://pubmed.ncbi.nlm.nih.gov/22701588/)
26. Kuzmina ML, Johnson KL, Barron HR, Hebert PDN. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. *BMC Ecol*. 2012; 12(1): 25
27. Chase MW, Fay MF. Barcoding of plants and fungi. *Science*. 2009; 325: 682–683 doi: [10.1126/science.1176906](https://doi.org/10.1126/science.1176906) PMID: [19644072](https://pubmed.ncbi.nlm.nih.gov/19644072/)
28. Elliott T, Davies J. Challenges to barcoding an entire flora. *Mol Ecol Res*. 2014; 14(5): 883–891
29. Kuzmina ML, Braukmann TWA, Rodrigues A, deWaard S, Graham S, et al. A DNA barcode library for the vascular plants of Canada. 2016; In Prep.
30. Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes*. 2007; 7(3): 355–364
31. Brouillet L, Coursol F, Meades SJ, Favreau M, Anions M, Bélisle P, and Desmet P. VASCAN, the Database of Vascular Plants of Canada. 2010; <http://data.canadensys.net/vascan/>
32. Ecological Stratification Working Group. A National Ecological Framework for Canada. Agriculture and Agri–Food Canada, Research Branch, Centre for Land and Biological Resources Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/ Hull. Report and national map at 1:7,500,000 scale. 1996.
33. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010; 26(19): 2460–2461 doi: [10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461) PMID: [20709691](https://pubmed.ncbi.nlm.nih.gov/20709691/)
34. Bininda–Emonds ORP. transAlign: using amino acids to facilitate the multiple alignment of protein–coding DNA sequences. *BMC Bioinformatics*. 2005; 6(1): 156

35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4): 772–780 doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
36. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol.* 2008; 57(5): 758–771 doi: [10.1080/10635150802429642](https://doi.org/10.1080/10635150802429642) PMID: [18853362](https://pubmed.ncbi.nlm.nih.gov/18853362/)
37. Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway. In Proceedings of the 2011 TeraGrid Conference on Extreme Digital Discovery—TG '11. New York, USA: ACM Press; 2011. pp. 1–8
38. Webb CO. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *Am Nat.* 2000; 156(2): 145–155 doi: [10.1086/303378](https://doi.org/10.1086/303378) PMID: [10856198](https://pubmed.ncbi.nlm.nih.gov/10856198/)
39. Swenson NG. Phylogenetic analyses of ecological communities using barcode data. In: Kress WJ and Erickson D L, editors. DNA barcodes. Methods and protocols. Humana Press; 2012. pp. 409–419.
40. Swenson NG. Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLOS One.* 2009; 4(2): e4390 doi: [10.1371/journal.pone.0004390](https://doi.org/10.1371/journal.pone.0004390) PMID: [19194509](https://pubmed.ncbi.nlm.nih.gov/19194509/)
41. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics.* 2010; 26(11): 1463–1464 doi: [10.1093/bioinformatics/btq166](https://doi.org/10.1093/bioinformatics/btq166) PMID: [20395285](https://pubmed.ncbi.nlm.nih.gov/20395285/)
42. R Development Core Team. R: A language and environment for statistical computing. R Foundation, Vienna, Austria. 2008
43. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010; 7(5): 335–336 doi: [10.1038/nmeth.f.303](https://doi.org/10.1038/nmeth.f.303) PMID: [20383131](https://pubmed.ncbi.nlm.nih.gov/20383131/)
44. Uriarte M, Swenson NG, Chazdon RL, Comita LS, Kress WJ, Erickson D, et al. Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: implications for community assembly. *Ecol Lett.* 2010; 13(12): 1503–1514 doi: [10.1111/j.1461-0248.2010.01541.x](https://doi.org/10.1111/j.1461-0248.2010.01541.x) PMID: [21054732](https://pubmed.ncbi.nlm.nih.gov/21054732/)
45. Liu X, Swenson NG, Zhang J, Ma K. The environment and space, not phylogeny, determine trait dispersion in a subtropical forest. *Funct Ecol.* 2013; 27(1): 264–272
46. Lebrija-Trejos E, Wright SJ, Hernandez A, Reich PB. Does relatedness matter? Phylogenetic density-dependent survival of seedlings in a tropical forest. *Ecology.* 2014; 95(4): 940–951 PMID: [24933813](https://pubmed.ncbi.nlm.nih.gov/24933813/)
47. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA.* 1987; 84(24): 9054–9058 PMID: [3480529](https://pubmed.ncbi.nlm.nih.gov/3480529/)
48. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol.* 2008; 49(3): 827–831 doi: [10.1016/j.ympev.2008.09.009](https://doi.org/10.1016/j.ympev.2008.09.009) PMID: [18838124](https://pubmed.ncbi.nlm.nih.gov/18838124/)
49. Nalbantoglu OU, Way SF, Hinrichs SH, Sayood K. RAphy: phylogenetic classification of metagenomics samples using iterative refinement of relative abundance index profiles. *BMC Bioinformatics.* 2011; 12: 41 doi: [10.1186/1471-2105-12-41](https://doi.org/10.1186/1471-2105-12-41) PMID: [21281493](https://pubmed.ncbi.nlm.nih.gov/21281493/)
50. Petit RJ, Excoffier L. Gene flow and species delimitation. *Trends Ecol Evol.* 2009; 24: 386–393 doi: [10.1016/j.tree.2009.02.011](https://doi.org/10.1016/j.tree.2009.02.011) PMID: [19409650](https://pubmed.ncbi.nlm.nih.gov/19409650/)
51. Naciri Y, Caetano S, Salamin N. Plant DNA barcodes and the influence of gene flow. *Mol Ecol Res.* 2012; 12: 575–580
52. Naciri Y, Linder P. Species identification and delimitation: the dance of the seven veils. *Taxon.* 2015; 64: 3–16
53. Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011; 76: 273–297 doi: [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4) PMID: [21424877](https://pubmed.ncbi.nlm.nih.gov/21424877/)
54. Hilu KW, Liang H. the *matK* gene: sequence variation and application in plant systematics. *Am J Bot.* 1997; 84(6): 830–839 PMID: [21708635](https://pubmed.ncbi.nlm.nih.gov/21708635/)
55. Müller KF, Borsch T, Hilu KW. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. *Mol Phy Evol.* 2006; 41: 99–117
56. Duffy AM, Kelchner SA, Wolf PG. Conservation of selection on *matK* following an ancient loss of its flanking intron. *Gene.* 2009; 438: 17–25 doi: [10.1016/j.gene.2009.02.006](https://doi.org/10.1016/j.gene.2009.02.006) PMID: [19236909](https://pubmed.ncbi.nlm.nih.gov/19236909/)
57. Hughes C, Eastwood R. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proc Natl Acad Sci USA.* 2006; 103: 10334–10339 doi: [10.1073/pnas.0601928103](https://doi.org/10.1073/pnas.0601928103) PMID: [16801546](https://pubmed.ncbi.nlm.nih.gov/16801546/)
58. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: From gene to genome. *Biol Rev.* 2015; 90(1): 157–166 doi: [10.1111/brv.12104](https://doi.org/10.1111/brv.12104) PMID: [24666563](https://pubmed.ncbi.nlm.nih.gov/24666563/)

59. Caetano Wyler S, Naciri Y. Evolutionary histories determine DNA barcoding success in vascular plants: seven case studies using intraspecific broad sampling of closely related species. *BMC Evol Biol.* 2016; 16: 103 doi: [10.1186/s12862-016-0678-0](https://doi.org/10.1186/s12862-016-0678-0) PMID: [27178437](https://pubmed.ncbi.nlm.nih.gov/27178437/)
60. Klopstein S, Currat M, Excoffier L. The fate of mutations surfing on the wave of range expansion. *Mol Biol Evol.* 2006; 23: 482–490 doi: [10.1093/molbev/msj057](https://doi.org/10.1093/molbev/msj057) PMID: [16280540](https://pubmed.ncbi.nlm.nih.gov/16280540/)
61. Excoffier L, Ray L. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol.* 2008; 23: 347–351 doi: [10.1016/j.tree.2008.04.004](https://doi.org/10.1016/j.tree.2008.04.004) PMID: [18502536](https://pubmed.ncbi.nlm.nih.gov/18502536/)
62. Brochmann C, Brysting AK. The Arctic—an evolutionary freezer? *Plant Ecol Divers.* 2010; 1(2): 181–195
63. Grundt HH, Kjølnner S, Borgen L, Rieseberg LH, Brochmann C. High biological species diversity in the arctic flora. *Proc Natl Acad Sci USA.* 2006; 103(4): 972–975 doi: [10.1073/pnas.0510270103](https://doi.org/10.1073/pnas.0510270103) PMID: [16418291](https://pubmed.ncbi.nlm.nih.gov/16418291/)
64. Skrede I, Borgen L, Brochmann C. Genetic Structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic-alpine distributions. *Heredity.* 2009; 102: 293–302 doi: [10.1038/hdy.2008.120](https://doi.org/10.1038/hdy.2008.120) PMID: [19066622](https://pubmed.ncbi.nlm.nih.gov/19066622/)
65. Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R. Polyploidy in arctic plants. *Biol J Linn Soc.* 2004; 82: 521–536
66. Percy DM, Argus GW, Cronk QC, Fazekas AJ, Kesanakurti PR, Burgess KS, et al. 2014. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? *Mol Ecol.* 2014; 23(19): 4737–4756 doi: [10.1111/mec.12837](https://doi.org/10.1111/mec.12837) PMID: [24944007](https://pubmed.ncbi.nlm.nih.gov/24944007/)
67. Leskinen E, Alstrom-Rapaport C. Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: evidence from 5.8S, ITS1, and ITS2 of the rDNA. *Plant Syst Evol.* 1999; 215: 209–227
68. Fogelqvist J, Verkhovina AV, Katyshev AI, Pucholt P, Dixelius C, Ronnberg-Wastjung AC, et al. Genetic and morphological evidence for introgression between three species of willows. *BMC Evol Biol.* 2015; 15: 193 doi: [10.1186/s12862-015-0461-7](https://doi.org/10.1186/s12862-015-0461-7) PMID: [26376815](https://pubmed.ncbi.nlm.nih.gov/26376815/)
69. Lauron-Moreau A, Pitre FE, Argus GW, Labrecque M, Brouiller L. Phylogenetic relationships of American willows (*Salix* L., Salicaceae). *PLOS One.* 2015; 10(4): e0121965 doi: [10.1371/journal.pone.0121965](https://doi.org/10.1371/journal.pone.0121965) PMID: [25880993](https://pubmed.ncbi.nlm.nih.gov/25880993/)
70. Mandel JR, Dikow RB, Funk VA. Using phylogenomics to resolve mega-families: An example from Compositae. *J Syst Evol.* 2015; 53(5): 391–402
71. Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *J. Ethnopharmacol.* 2010; 130(1): 116–121 doi: [10.1016/j.jep.2010.04.026](https://doi.org/10.1016/j.jep.2010.04.026) PMID: [20435122](https://pubmed.ncbi.nlm.nih.gov/20435122/)
72. Tekpinar AD, Erkul SK, Aytaç Z, Kaya Z. Phylogenetic relationships between *Oxytropis* DC. and *Astragalus* L. species native to an Old World diversity center inferred from nuclear ribosomal ITS and plastid *matK* gene sequences. *Turkish J Biol.* 2016; 40(1): 250
73. Sønstebo JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J, Willerslev E, et al. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol Ecol Res.* 2010; 10(6): 1009–1018
74. Puente-Lelièvre C, Harrington MG, Brown EA, Kuzmina M, Crayn DM. Cenozoic extinction and recolonization in the New Zealand flora: The case of the fleshy-fruited epacrids (Styphelioideae, Styphelioideae, Ericaceae). *Mol Phylogenet Evol.* 2013; 66(1): 203–214 doi: [10.1016/j.ympev.2012.09.027](https://doi.org/10.1016/j.ympev.2012.09.027) PMID: [23044402](https://pubmed.ncbi.nlm.nih.gov/23044402/)
75. Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, et al. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature.* 2007; 445(7129): 757–760 doi: [10.1038/nature05587](https://doi.org/10.1038/nature05587) PMID: [17301791](https://pubmed.ncbi.nlm.nih.gov/17301791/)
76. Kumar S, Kahlon T, Chaudhary S. A rapid screening for adulterants in olive oil using DNA barcodes. *Food Chem.* 2011; 127(3): 1335–41 doi: [10.1016/j.foodchem.2011.01.094](https://doi.org/10.1016/j.foodchem.2011.01.094) PMID: [25214135](https://pubmed.ncbi.nlm.nih.gov/25214135/)
77. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial teas highlight plant DNA barcode identification successes and obstacles. *Sci Rep.* 2011; 1: 42 doi: [10.1038/srep00042](https://doi.org/10.1038/srep00042) PMID: [22355561](https://pubmed.ncbi.nlm.nih.gov/22355561/)
78. Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V. 2016. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J.* 2016; 14(1): 8–21 doi: [10.1111/pbi.12419](https://doi.org/10.1111/pbi.12419) PMID: [26079154](https://pubmed.ncbi.nlm.nih.gov/26079154/)
79. Bruni I, De Mattia F, Martellos S, Galimberti A, Savadori P, Casiraghi M, Nimis PL, Labra M. DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). *PLOS One.* 2012; 7(9): e43256 doi: [10.1371/journal.pone.0043256](https://doi.org/10.1371/journal.pone.0043256) PMID: [22970123](https://pubmed.ncbi.nlm.nih.gov/22970123/)

80. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, et al. Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front Zool.* 2009; 6(1): 16
81. Avanesyan A. Plant DNA detection from grasshopper guts: A step-by-step protocol, from tissue preparation to obtaining plant DNA sequences. *Appl Plant Sci.* 2014; 2(2): apps.1300082
82. McClenaghan B, Gibson JF, Shokralla S, Hajibabaei M. Discrimination of grasshopper (Orthoptera: Acrididae) diet and niche overlap using next-generation sequencing of gut contents. *Ecol Evol.* 2015; 5(15): 3046–3055 doi: [10.1002/ece3.1585](https://doi.org/10.1002/ece3.1585) PMID: [26356479](https://pubmed.ncbi.nlm.nih.gov/26356479/)
83. Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, et al. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc Natl Acad Sci USA.* 2015; 112(26): 8019–8024 doi: [10.1073/pnas.1503283112](https://doi.org/10.1073/pnas.1503283112) PMID: [26034267](https://pubmed.ncbi.nlm.nih.gov/26034267/)