Climatic niche estimation, trait evolution and species richness in North American *Carex* (Cyperaceae)

by

Jocelyn E. Pender, B.Sc. (Hons.)

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

University of Ottawa

In partial fulfilment of the requirements for the

Degree of Master of Science in the

Ottawa Carleton Institute of Biology

© Jocelyn E. Pender, Ottawa, Canada, 2016

Supervisor:

Dr. Julian Starr

Committee Members:

Dr. Tyler Smith

Dr. Risa Sargent (Fall 2013 - Winter 2015)

Dr. Stéphane Aris-Brosou (Fall 2015 - present)

Abstract

With close to 2100 species, the flowering plant genus *Carex* (Cyperaceae; sedges) is an example of an evolutionary radiation. Despite its potential for use as a model taxon in evolutionary studies, the diversification of sedges remains largely unexplored. This thesis realizes the potential of *Carex* as an evolutionary model group by using it to ask questions about species richness patterns. More specifically, it seeks to determine the relationship, if any, between rates of trait evolution and species richness. This tests the hypothesis that organisms with increased abilities to evolve new traits, speciate more rapidly. Morphological and ecological (habitat and climatic niche) traits are modelled on a nearly complete regional (North America north of Mexico) phylogeny and rates of trait evolution are compared among non-nested sister groups. However, before trait evolution is modelled, this work evaluates the sensitivity of climatic niche estimates to underlying distribution datasets. It tests the agreement of niche estimates derived from the commonly used online repository GBIF (the Global Biodiversity Information Facility) and county-level distributions via BONAP (the Biota of North America Program). Results showed that in the context of phylogenetic comparative analyses, it is not vital to obtain highly accurate climatic niche estimates. The second study found significant positive correlations between the rates of climatic niche, habitat and reproductive morphological evolution and species richness. This result supports the role of high trait lability in generating species richness and more generally, the idea that high trait disparity through evolutionary time leads to species success.

Statement of Contributions

The studies contained in this thesis were conceived and designed by Andrew L. Hipp, Julian R. Starr and Jocelyn E. Pender. Molecular work was performed by J.E. Pender. Additional molecular data were provided by Brianna N. Chouinard, Marlene Hahn and A.L. Hipp. County distribution data were collected, assembled and provided by John T. Kartesz and Misako Nishino. Morphological, specimen-based distribution and climatic niche data were gathered by J.E. Pender. Habitat scoring was by J.R. Starr and A.L. Hipp. All data were analysed, presented and discussed by J.E. Pender.

Acknowledgements

Firstly, thank you to my supervisor Julian Starr. I thoroughly appreciate all of the guidance provided. You have revised many drafts, made many wise suggestions and provided much moral and professional support. Thank you, perhaps most importantly, for providing me with a project that has challenged me intellectually and technically (I am now a proud 'R'tist). I am excited to see where my new skills take me.

Thank you Andrew Hipp, for your enthusiasm, creativity and knowledge. You have been integral to the development of this work. Also, thank you for facilitating my visits to the Arboretum, which were invaluable research getaways. I am further grateful to Marlene Hahn, who provided data download protocols, assisted with molecular data organization, and with other tasks. Thanks to the Morton Arboretum team for providing ITS and ETS sequences.

Valuable feedback and advice was provided by my committee members, Tyler Smith and Risa Sargent. Thank you for your in-depth emails and guidance.

The first Chapter was achieved in collaboration with John Kartesz and Misako Nishino. They graciously supplied the Biota of North America Program (BONAP) data for all North American *Carex* species, undoubtedly requiring much hard work to collect and assemble. Thank you for your cooperation.

Thank you Sarah Chisholm and Elyse Latreille for assistance with portions of the research. The hard work of Brianna Chouinard provided all genomic DNA and *matK* sequences used for the development of the phylogeny. Discussions with Étienne Léveillé-Bourret aided in the development of this work. Also, thank you Étienne for allowing me to use your sequences for outgroups in my phylogeny.

Roger Bull and Paul Sokoloff provided vital training and guidance in DNA extraction, amplification and sequencing protocols. They have been important to the development of this work, thank you. Lastly, thank you Claire Gilmour, Wayne Sawtell, Anna Ginter, Tamara Villaverde, and Warren Cardinal-McTeague for providing advice on the graduate student process.

Table of Contents

Abstract ..... **iii**
Statement of Contributions ..... **iv**
Table of Contents ..... **vii**
List of Figures ..... **ix**
List of Tables ..... **xii**
List of Appendices ..... **xiii**
Chapter 1: Introduction **1**
1.1 FIGURES 5
Chapter 2: How sensitive are climatic niche inferences to distribution data sampling? **6**
2.1 INTRODUCTION 6
2.2 METHODS 9
2.2.1 Study taxon 9
2.2.2 Occurrence data 9
2.2.2.1 Global Biodiversity Information Facility (GBIF) dataset 9
2.2.2.2 Biota of North America Program (BONAP) dataset 10
2.2.3 Geographic extent 11
2.2.4 Climate data 11
2.2.5 Principal component analysis 12
2.2.6 Correlates of error 12
2.2.7 Phylogeny 14
2.2.8 Ancestral state reconstruction 14
2.2.9 Additional analyses 15
2.3 RESULTS 16
2.3.1 Principal component analysis 16
2.3.2 Sampling effort influencing error 17
2.3.3 Elevation influencing error 18
2.3.4 Ancestral state reconstruction 18
2.3.5 Additional analyses 19
2.4 DISCUSSION 20
2.4.1 Recommendations 27
2.5 TABLES AND FIGURES 28

Chapter 3: Regional species richness and trait evolution in the genus *Carex* (Cyperaceae) 37
3.1 INTRODUCTION 37
3.2 METHODS 45
3.2.1 Trait definitions 45
3.2.2 Flora of North America region 45
3.2.3 Phylogeny 46
3.2.3.1 Taxon sampling 46
3.2.3.2 Molecular markers 46
3.2.3.3 DNA amplification and sequencing 47
3.2.3.4 Sequence selection 48
3.2.3.5 Outgroup species 50
3.2.3.6 Sequence alignment 50
3.2.3.7 Phylogenetic inference 51
3.2.4 Trait data 52
3.2.5 Trait evolution 54
3.2.6 Sister clade comparison 55
3.3 RESULTS 58
3.3.1 Phylogeny 58
3.3.2 Trait data 59
3.3.3 Trait evolution 59
3.4 DISCUSSION 61
3.5 TABLES AND FIGURES 69
Chapter 4: Conclusions 78
References 80
Appendix 1. Supplementary Figures for Chapter 2 92
Appendix 2. DNA sample data 93
Appendix 3. Supplementary Figures for Chapter 3 102
Appendix 4: Maximum Likelihood gene trees 105

List of Figures

Fig. 1.1 Reproductive morphology of the flowering plant genus *Carex*. (A) A female spikelet showing the perigynium in blue and achene in green. The achene is a one-seeded fruit, surrounded by a sac-like structure called the perigynium. (B) An example of how spikelets are arranged to form an inflorescence. (A) and (B) are adapted from Starr & Ford (2009). (C) Spikelets arranged on a culm or a flower-bearing stalk. (D) Achenes removed from surrounding perigynia. (C) and (D) are from www.phytoimages.siu.edu. 5

Fig. 2.1 Principal component (PC) climatic niche space. PCA was used to condense 19 WorldClim BIOCLIM variables into two PC axes (PC1=40% and PC2=34%) explaining the majority (74%) of bioclimatic variance. Species' GBIF estimates in niche space are represented by circles and BONAP points are represented by triangles. Grey vectors connect species' GBIF estimates to BONAP estimates, from which magnitude and direction are calculated. A sample calcuation in provided on the right. N=378. 29

Fig. 2.2 The relationship between sampling effort and magnitude of differences between GBIF and BONAP climatic niche estimates. Regression lines are presented if a significant correlation was found. Correlation coefficients (ρ), p-values and r^2 values are indicated in the top right-hand corner of plots with regression lines. (A) Sample calculation of GBIF county coverage for *Carex lurida*. Counties included in *C. lurida*'s GBIF data distribution are in blue and *C. lurida* BONAP county data are in red. (B) The relationship between GBIF county coverage and differences in climatic niche space. Each circle is one *Carex* species. N=363. (C) Sample calculation of relative geographic spread for *C. lurida*. GBIF points are in blue and BONAP points are in red. Minimum convex polygons calculated from each dataset are visualized. (D) The relationship between relative geographic spread and differences in climatic niche space. Each circle is one *Carex* species. N=321. 30

Fig. 2.3 The relationship between elevation niche and differences in climatic niche space. Regression lines are presented if a significant correlation was found. Correlation coefficients (ρ), p-values and r^2 values are indicated in the top right-hand corner of plots with regression lines. Each circle is one *Carex* species. N=363. (A) Elevation plotted against magnitude of differences in PC climatic niche space. (B) Elevation plotted against direction of differences in PC climatic niche space. Direction is in degrees. 31

Fig. 2.4 Ancestral climatic niche PC1 reconstruction using GBIF estimates (left) and BONAP estimates (right). Colours represent trait values. The reconstruction was visualized on a *Carex* phylogeny using the *contMap* function in R. A plot of BONAP ancestral states versus GBIF ancestral states is presented in the bottom right hand corner, showing a slight offset from state equivalence. Further statistics evaluating ancestral state differences are shown (see Methods for details). 33

Fig. 2.5 The relationship between elevation predicted from GBIF occurrence data and elevation predicted from BONAP county data. A regression line is presented in black and a 1:1 elevation equivalence line is in grey. Elevation data were obtained from WorldClim at a

5km resolution and were combined with locality data to calculate elevation estimates. Each circle is one *Carex* species. N=378. 34

Fig. 2.6 Uncertainty in ancestral state reconstruction from GBIF climatic niche PC1 estimates. Evolutionary time is on the x-axis and trait values on the y-axis. Blue transparent lines extend outwards showing 95% confidence limits. Plot was generated using the *fancyTree* function in the *phytools* package for R. Uncertainty from reconstruction of BONAP climatic niche estimates appeared qualitatively identical and is therefore not presented. N=363. 35

Fig. 2.7 Analysis of the influence of geography on differences in climatic niche estimates. Counties in the eastern US are on average smaller than in the west. Species were separated into strictly eastern and strictly western species. Widespread species were not included. N=179. 36

Fig. 3.1 An example of one set of non-nested sister group comparisons. Each sister contrast is shown in a different colour. Parental nodes are circled in red and sister nodes are circled in black. For each sister comparison, the ratio of species richness and ratio of trait evolution is calculated. The data then are log transformed and plotted. This procedure is performed for ten different sets of sister groups. 72

Fig. 3.2 Maximum clade credibility phylogeny summarizing two BEAST runs of 100 million generations each. Two ougroup species and 477 North American *Carex* species are included. Thickness of branches correspond to Bayesian posterior probabilities. The phylogeny was constructed from the four-gene supermatrix (*matK, rps16*, ITS, ETS). Major clades (from Starr *et al.*, 2015) are indicated and are accompanied by a photo of a member species. Time scale is in millions of years. Labels contain sample information (Morton Arboretum identifier, MOR and Starr Lab identifier, B code), two-letter geography information and section information. 73

Fig. 3.3 Histogram of mean r^2 values and slopes from regressions of species richness against rates of trait evolution (log transformed) using 100 continuous character evolution simulations. See methods for details on the procedure used to obtain regressions. The red line represents the 0.05 significance level. 76

Fig. 3.4 Non-nested sister clade analysis. The ratio of species richness (i.e., the number of descendant taxa) and the ratio of rates of evolution per sister clade contrast are plotted for contrasts satisfying the criterion (= minimum of three descendants per clade). Sets of non-nested sister clades were obtained by maximizing the number of contrasts. Different colours represent different sets of non-nested sister clades, each with the maximum number of contrasts. Mean r^2 and slope values are averaged regression statistics among sets.

Asterisks represent significance according to the null distribution of 100 continuous character simulations on the tree. 77

Fig. A1.1 Supplementary analyses searching for correlates of differences in niche estimates. No significant correlations were recovered. Each circle is one *Carex* species. N=363. 92

Fig. A3.1 Climate data plotted on the *Carex* spp. ultrametric phylogeny. Climate variables were taken from the BIOCLIM WorldClim database. Principal component analysis (PCA) was used to reduce the means of 19 BIOCLIM variables (prcomp in R) into climate PC1 and PC2. The map was generated using the *contMap* function in the *phytools* package for R. N=445. 102

Fig. A3.2 Habitat data plotted on the *Carex* spp. ultrametric phylogeny. Light habitat was scored by experts. Wetness habitat was taken from the National Wetland Plant List (NWPL; Lichvar, 2012). Species missing from the NWPL were scored by experts for wetness habitat. The map was generated using the *contMap* function in the *phytools* package for R. Light habitat N=463, wetness habitat N=460. 103

Fig. A3.3 Morphological data for those individuals with complete data sampling (data for four characters available) plotted on the *Carex* spp. ultrametric phylogeny. Morphological data was extracted from the Flora of North America (2002). Principal component analysis was used to reduce the number of morphological variables. Only PC1 and PC2 were analysed. The map was generated using the *contMap* function in the *phytools* package for R. N=371. 104

List of Tables

Table 2.1. Loadings from principal component analysis of 19 WorldClim BIOCLIM variables estimated from GBIF and BONAP species localities. Only those principal components explaining a larger than expected amount of variation (=5.3%) are included (PC1=40%, PC2=34%, PC3=18%). Loadings are bolded if they are important contributors. This was assessed by comparing against the expected loading value if all variables contributed equally (square root of 1/19=0.229415). BIOCLIM variables are sorted according to contributions to PC1, PC2 and lastly PC3. 28

Table 3.1. Primer sequences used in DNA amplification and sequencing. 69

Table 3.2. Loadings from principal component analysis of 19 WorldClim BIOCLIM variables estimated from GBIF occurrence data. Only those principal components explaining a larger than expected amount of variation are included (>5.3%: PC1=42%, PC2=28%, PC3=16%, PC4=7%). Loadings are bolded if they are important contributors (>square root of 1/19=0.229415). BIOCLIM variables are sorted according to their contributions to PC1, PC2, PC3 and PC4. 70

Table 3.3. Loadings from principal component analysis of six morphological characters (perigynia length, perigynia width, achene length, achene width, culm height, leaf width). Important contributors to principal components explaining a larger than expected amount of variation are bolded (PC1=49%, PC2=19%). 71

Table A2.1 *Carex* individuals selected for phylogenetic analysis. Voucher details can be found in Appendix B of Chouinard (2010). Use B numbers for correspondence. Locations where herbarium specimens were collected (column 3) are abbreviated into states or provinces, when possible. 93

Table A2.2 Voucher information for outgroups used in this study. 101

List of Appendices

Appendix 1. Supplementary Figures for Chapter 2 92
Appendix 2. DNA sample data 93
Appendix 3. Supplementary Figures for Chapter 3 102
Appendix 4: Maximum Likelihood gene trees 105

Chapter 1: Introduction

The study of extant evolutionary radiations has been central to the formulation of diversification theory (Losos & Mahler, 2010). It has allowed biologists to infer, for example, processes leading to speciation, trait variation, and the establishment of geographical ranges and niches (Fritz *et al.,* 2013). Evolutionary radiation is defined as the differentiation of a single ancestor into a diverse set of morphological and ecological forms (Schluter, 2000; Streelman & Danley, 2003; Losos & Mahler, 2010). More concretely, a group should possess species richness of at least one order of magnitude greater than its similarly aged sister group to be considered a radiation (Sanderson & Donoghue, 1994; Bouchenak-Khelladi *et al.,* 2015). Although significant progress has been made in understanding biodiversity processes through the study of radiations, central questions remain. Most significantly, it remains unclear why species richness imbalance persists across phylogenies and how these diversity differences are related to morphological and ecological diversity (termed "disparity" by evolutionary biologists to avoid confusion with species diversity; Losos & Mahler, 2010).

The flowering plant genus *Carex* L. (ca. 2100 spp.; family Cyperaceae or "sedges"; Govaerts *et al.,* 2014; Global *Carex* Group, 2015) is one of the largest angiosperm genera worldwide (4th largest; Frodin, 2004; Global *Carex* Group, 2015) and the most diversified genus in the northern temperate zone (Reznicek, 1990). It possesses 20-fold more species than its sister clade Scirpeae (98 species; Léveillé-Bourret *et al.,* 2014), and therefore constitutes an evolutionary radiation. *Carex* species can be found in forests, grasslands, montane rocky habitats, arctic tundra and moist to wet habitats worldwide, where they are often dominant or co-dominant (Ball & Reznicek, 2002). The extensive habitat diversity of the genus is paralleled by

its global distribution, as *Carex* inhabits every continent except Antarctica. This distribution has recently been shown to be the result of infrequent long-distance dispersal events (Spalink *et al.*, 2015), revealing that *Carex* is geographically coherent at the continental scale. *Carex* is further exceptional in that its species frequently co-occur.

Unlike famous examples of species radiations (e.g., passerine birds, *Anolis* lizards), investigations into the evolutionary history of sedges have been few. This is a notable loss for a number of reasons. Firstly, the genus' habitat and climate diversity provides a unique opportunity for study of niche disparity and species richness. Furthermore, investigation of niche partitioning and coexistence of related species within communities is possible in *Carex*. Lastly, the exceptional sedge diversity in the Northern Hemisphere would allow for investigation of temperate processes. The potential of *Carex* as a study taxon for evolutionary and ecological studies has been emphasized by sedge systematists (Waterway *et al.,* 2009; Starr & Ford, 2009) but has yet to be fully realized.

Sedge studies have been missing from the evolutionary literature due to their historically difficult taxonomy (Starr & Ford, 2009) and the lack of accurate phylogenetic hypotheses. However, new developments in molecular methods have increased the availability of sequence data and have aided in new resolutions of *Carex* relationships (Starr & Ford, 2009; Léveillé-Bourret *et al.,* 2014; Starr *et al.,* 2015; Global *Carex* Group, 2015). As a result, research has begun to consider character evolution and diversification processes in *Carex* (Gehrke & Linder, 2011; Escudero *et al.,* 2012a; Escudero *et al.,* 2012b; Escudero & Hipp, 2013; Maguilla *et al.,* 2015). Unfortunately, studies continue to be limited in taxonomic and geographic scope.

This thesis realizes the potential of *Carex* to investigate fundamental evolutionary and ecological questions (Chapter 3). It is the first to do so on a broad scale, considering almost all

Carex species existing in North America, north of Mexico (the Flora of North America region; FNA). However, before it attempts to understand *Carex* diversity, it evaluates a commonly employed method of climatic niche estimation (Chapter 2).

Recently, researchers have begun to investigate the phylogenetic conservation, divergence and tempo of change of species' niche (Warren *et al.,* 2008). This has in part been motivated by the increasing abundance of species distribution data and availability of high quality climatic data (Araújo & Peterson, 2012), which are used in combination to estimate a climatic niche. The largest provider of distribution data is the Global Biodiversity Information Facility (GBIF), hosting occurrence data submitted by hundreds of natural history collections across the world. However, despite the acknowledgement of data quality issues in GBIF occurrence datasets (Graham, 2004; Yesson *et al.,* 2007; Newbold, 2010), little information exists on the sensitivity of niche estimates to data inaccuracies and uncertainties. Furthermore, the effect of erroneous GBIF data on phylogenetic comparative analyses, such as the trait evolution analysis presented afterward in this thesis, is unknown. The appropriateness of this method is tested in Chapter 2 by evaluating how GBIF data perform against known county-level species distributions from the Biota of North America Program (BONAP). This study searches for correlates of climatic niche estimation disagreements, by testing GBIF sampling effort and elevation niche. The results are presented and discussed in the context of providing recommendations for evolutionary biologists seeking to use GBIF-enabled climatic niche estimates in their research. Chapter 2 provides the first broad-scale comparison of North American GBIF distribution data to a secondary source of distribution data (BONAP) and provides justification for climatic niche estimation methods used in the following chapter.

Chapter 3 asks the question, why do diversity differences exist across the tree of life? More specifically, why are there so many *Carex* species in North America? It relates this question to a recently proposed hypothesis that higher rates of trait evolution result in higher abilities to evolve new morphological or ecological traits (Lovette *et al.,* 2002; Adamowicz *et al.,* 2008; Pigliucci 2008, Rabosky *et al.,* 2013). Specifically, Chapter 3 estimates rates of evolution of *Carex* morphology (measurements of culms, leaves, perigynia and achenes; see Fig. 1.1 for terminology), habitat niche and climatic niche on a reconstructed phylogeny. Sister clade comparisons are used to assess statistically whether rates of trait evolution are associated with increased species richness within the genus.

Further background information and precise hypotheses are presented at the beginning of each study. Figures and Tables can be found at the end of each Chapter (Chapters 1-3). Lastly, in Chapter 4, conclusions from Chapters 2 and 3 are summarized and discussed.

1.1 FIGURES

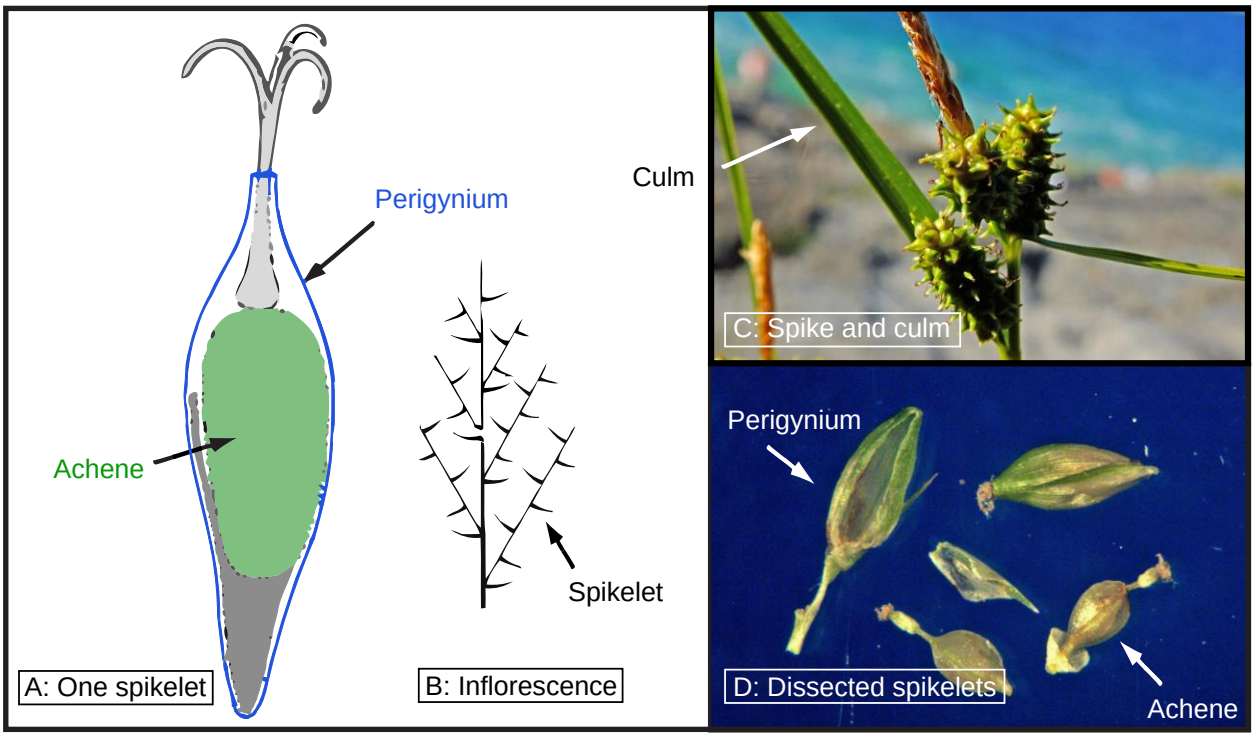


Fig. 1.1 Reproductive morphology of the flowering plant genus *Carex*. (A) A female spikelet showing the perigynium in blue and achene in green. The achene is a one-seeded fruit, surrounded by a sac-like structure called the perigynium. (B) An example of how spikelets are arranged to form an inflorescence. (A) and (B) are adapted from Starr & Ford (2009). (C) Spikelets arranged on a culm or a flower-bearing stalk. (D) Achenes removed from surrounding perigynia. (C) and (D) are from www.phytoimages.siu.edu.

Chapter 2: How sensitive are climatic niche inferences to distribution data sampling?

2.1 INTRODUCTION

The integration of species occurrence and phylogenetic data has allowed researchers to address questions about the role of the ecological niche in speciation and extinction. For example, phylogenetic niche conservatism and niche divergence are two well-known hypotheses about the nature of diversification that have resulted from the study of species' niches (predicted via occurrence data) along phylogenies (Peterson *et al.,* 1999; Graham *et al.,* 2004; Warren *et al.,* 2008; Pyron *et al.,* 2014). In particular, species' realized climatic niche is easily quantified using occurrence datasets and climatic data, and when analyzed in a phylogenetic framework has the possibility to elucidate the role of climate in diversification (Peterson *et al.,* 2010).

The largest online resource for occurrence data is the Global Biodiversity Information Facility (GBIF), an international infrastructure of millions of biological specimen records published freely by institutions across the world (www.gbif.org). Unfortunately, GBIF suffers from significant data quality issues (Soberón *et al.,* 2002; Graham *et al.,* 2004; Yesson *et al.,* 2007; Newbold, 2010). Most notably, the spatial coverage of GBIF occurrence records does not appropriately represent species ranges (e.g., in invertebrates see Beck *et al.,* 2013). Poor spatial coverage of data may be the result of the biases of collectors, biases of institutions (e.g., predisposition towards collecting in the Western Hemisphere) and which institutions submit to GBIF (Hijmans *et al.,* 2000; Soberón *et al.,* 2000; Reddy & Davalos, 2003; Dennis & Thomas, 2000; Newbold, 2010). These biases pose potential problems for accurate niche estimation and phyloclimatic analyses.

Research on the impact of occurrence data biases on ecological niche estimation has suggested that environmental bias may be avoided if localities are well placed (i.e., covering all main areas of a species' environmental gradient) and if sampling is sufficient (Hirzel & Guisan, 2002; Newbold, 2010). However, this research has been limited in geographic and taxonomic scope, and has produced conflicting results (Kadmon *et al.,* 2004; Hortal *et al.,* 2008; Beck *et al.,* 2014). Furthermore, research has been oriented towards niche modelling applications (e.g., MaxEnt; Phillips *et al.,* 2006), despite the frequent usage of GBIF occurrence data in phylogenetic comparative analyses. An evaluation of the quality of climatic niche inferences from GBIF occurrence data on a broad geographic scale and from a phylogenetic comparative perspective is lacking. More generally, an appropriate strategy to correct for insufficient sampling when environmental bias is present is needed for all users of GBIF data.

This study contributes to the assessment and improvement of climatic niche inferences from GBIF occurrence data across North America. It compares climatic niche estimates obtained from GBIF and from county-level verified distributions (Biota of North America Program; BONAP) for 363 species of the largest angiosperm genus in the temperate zone (genus *Carex*, family Cyperaceae). Specifically, the magnitude and direction of differences between species' GBIF and BONAP-estimated locations in principal component niche space are calculated. Subsequently, this study tests (1) how GBIF and BONAP niche differences vary with GBIF sampling effort, (2) how differences vary with elevation (to determine if county data is biased for high-elevation species, with the aim of supplementing GBIF data with BONAP county data), and tests (3) ancestral state reconstruction with each occurrence dataset. GBIF sampling effort is evaluated by its coverage of BONAP county records and its geographic spread relative to BONAP distributions, capturing coverage and skew of GBIF data respectively. Sampling effort

is expected to be correlated with differences in niche estimates (where poor sampling effort results in greater differences), high elevation species are predicted to be biased because of coarse county-level distribution data resolution, and ancestral state reconstructions are expected to differ between occurrence datasets. The utility of each dataset (GBIF and BONAP) for climatic niche inference is discussed and recommendations for dataset use and the potential for BONAP supplementation are presented.

2.2 METHODS

2.2.1 Study taxon

The genus *Carex* (Cyperaceae; sedges) is one of the largest flowering plant genera in the world (4th largest; Frodin, 2004; Global *Carex* Group, 2015), with centres of diversity in North America (north of Mexcio; ca. 480 species) and eastern Asia (Ball & Reznicek, 2002; Starr & Ford, 2009). Unlike other big plant genera, it inhabits a wide range of habitats (e.g., deciduous forest, freshwater wetland, alpine tundra, deserts; Egorova, 1999; Ball & Reznicek, 2002) and climatic niches (polar, boreal, moderate). Sedges are particularly well known and well collected in North America. Consequently, occurrence data is most abundant and available in North America. The combination of high biodiversity, climatic diversity and occurrence data availability in North American *Carex* result in an ideal opportunity for analysis of the effects of occurrence data quality on climatic niche estimation. The Flora of North America online (www.efloras.org) was used as a checklist for *Carex* species to include in this study. Introduced species were excluded and subspecies and varieties were not recognized (n=363 spp.) for simplicity of analysis.

2.2.2 Occurrence data

2.2.2.1 Global Biodiversity Information Facility (GBIF) dataset – *Carex* occurrence data were obtained using the function *gbif* in the package *dismo* for R version 3.0.1 (www.r-project.org). As occurrence data can be removed or uploaded by participating institutions to GBIF at any time, to ensure the capture of the highest quantity of specimen data possible, data were downloaded at three time points; from 16 May to 28 May 2013, on 13 January 2014 and on

3 July 2014. These data were then merged, yielding from one to 216,475 GBIF records per *Carex* species. GBIF occurrence data obtained in R contained fields describing the locality (continent, country, administrative unit level 1, administrative unit level 2, locality, latitude, longitude, altitude), the collection (institution, collection, catalogNumber, basisOfRecord, collector, earliestDateCollected, latestDateCollected) and the download (gbifNotes, downloadDate). All records missing latitude and longitude coordinates were excluded. Additionally, duplicate records and records with unacceptably low precision were removed. Records were deemed to have low precision if either latitude or longitude had one decimal place or less, and these were removed because they were determined to be erroneous.

2.2.2.2 Biota of North America Program (BONAP) dataset –BONAP (www.bonap.org) is a database containing U.S. county-level botanical species distributions. BONAP maps are generated with the aid of floristic experts and are verified by dataset curators. Therefore, these maps represent current knowledge on species distributions and are the highest accuracy distribution data available. In their raw format, BONAP maps are county presence lists. To facilitate niche estimation, the BONAP data were transformed into point data in R by calculation of county polygon centroids (*map_data* and *Polygon* functions in the *ggplot2* package). Unfortunately, the resolution of this dataset depends on the size of the counties in the species' distributions. However, no finer scale accurate data are readily available for all *Carex* species at this time.

2.2.3 Geographic extent

Extent was limited to the continental United States due to the quality of GBIF data and the availability of BONAP data, and to achieve a contiguous study area. The R packages *maps, sp,* and *maptools* were used in combination to extract US border polygons from the *usa* database (including mainland borders and borders of Long Island, Manhattan, Martha's Vineyard, Nantucket Island, Orcas Island, San Juan Island, Staten Island and Whidbey Island). The function *pnt.in.poly* from the *SDMTools* package was used to verify whether occurrence points fell on or within US borders. Points that fell outside of polygons were excluded.

2.2.4 Climate data

Climatic niche estimates were enabled by the WorldClim Global Climate Dataset (www.worldclim.org; Hijmans *et al.,* 2005) using the R package *dismo*. WorldClim data are a set of global climate raster layers at a maximum resolution of one square kilometre. The BIOCLIM dataset (www.worldclim.org/bioclim) is composed of nineteen biologically meaningful variables derived from the WorldClim data set. Data were downloaded in R using the *getData* function and name='worldclim' in the *nlme* package. WorldClim data at a resolution of 2.5 arc-minutes (=~5 km) were used to remain conservative in climate estimates.

All 19 BIOCLIM variables were extracted for each locality point in both the GBIF and BONAP datasets. For each species and dataset, the mean per BIOCLIM variable was calculated. This resulted in 19 BIOCLIM-GBIF means and 19 BIOCLIM-BONAP means for each *Carex* species.

2.2.5 Principal component analysis

Climatic niche was quantified using principal component analysis (PCA). PCA is an ordination method that reduces a number of correlated variables into a smaller number of uncorrelated principal components. By including multiple species in a PCA, direct comparison of species' positions in climatic space is possible (Jongman *et al.,* 1995). The *prcomp* function in the *stats* package for R was used to perform PCA on BIOCLIM-GBIF and BIOCLIM-BONAP means and to rotate, centre and scale all variables. All means were included in one PCA, to allow for comparison of estimates.

The difference between GBIF and BONAP niche positions in principal component climatic space was computed using vectors projected along PC1 and PC2 axes. The magnitude and angle of the vector connecting the estimates were calculated. Vector magnitude should be interpreted in this study as the disagreement or error separating GBIF and BONAP niche positions and the vector angle as the direction of bias.

2.2.6 Correlates of error

Sampling effort of species GBIF data was evaluated in two ways; (1) by the number of BONAP distribution counties corresponding GBIF data cover (i.e., county coverage) and (2) by comparing the geographic spread of GBIF points to BONAP points. Both methods of quantifying sampling effort assume that BONAP data appropriately represent the extent of species' distributions. Although BONAP data are not perfectly correct, they are presence-absence distribution data, unlike occurrence data from the GBIF data repository. This means that when no record exists in a species' distribution for a particular geographic location, one can confidently assume that the species not does exist there. The data are designated presence-absence because

BONAP data managers have actively searched herbaria, institutions and experts for distribution information. Therefore, by comparing GBIF to BONAP data, the degree of missing data (or sampling effort) of GBIF data can be quantified.

To calculate county coverage of GBIF data, a list of counties GBIF points occupy was generated per species (see Fig. 2.2A for sample calculation). Then, a coverage percentage was calculated (GBIF counties/BONAP counties).

The function *mcp* in the R package *adehabitatHR* was used to compare geographic spread of GBIF localities to BONAP centroids. Specifically, *mcp* calculates the minimum convex polygon for a set of spatial points. The user may define a percentage of coordinate data to utilize in the calculation, to account for outliers. Ninety-five percent of GBIF occurrences were used (to account for outliers) and 100% of BONAP centroids were used to compute area polygons. Outliers are only expected to be found in GBIF data. Finally, the relative area of GBIF polygons to BONAP polygons was calculated (see Fig. 2.2C for sample calculation). Species with fewer than five GBIF occurrences were excluded, as polygons calculated from too few coordinates may result in inaccurate area estimates. Furthermore, after examination of the structure of relative area data, it was determined that species with relative areas more than 1.5 times the interquartile range of the data (obtained via *boxplot* in R) likely possessed erroneous GBIF data and were removed.

Elevation ranges were obtained from the Flora of North America (FNA) species treatments online (www.efloras.org) and the mean was calculated per species (e.g., *Carex cephaloidea* 20-300m, mean=160m). Despite their low accuracy and qualitative nature, FNA elevation mean data were preferred over GBIF-enabled WorldClim elevation estimates for consistency and predictive power (i.e., GBIF elevation estimates are of unknown accuracy and

precision). For example, if a user wishes to predict the quality of the climatic niche estimate from GBIF or BONAP data, they can search the FNA for the elevation range of a given species, and use it alongside the regressions from Figure 2.3 to inform their methods decisions.

The function *cor.test* in the *statistics* package in R was used to test for significant correlations between niche differences and sampling effort or elevation. From the same R package, *lm* was used to fit linear regressions when significant correlations were recovered.

2.2.7 Phylogeny

To evaluate how differences in climatic niche estimates affect phylogenetic comparative analyses, a time-calibrated Bayesian *Carex* phylogeny was inferred and ancestral climatic niche PC1 states were reconstructed on the tree. Details on phylogenetic methods can be found in the Methods section of Chapter 3.

2.2.8 Ancestral state reconstruction

Evolutionary biologists use phylogenetic comparative methods to understand how species' traits affect speciation and extinction. Increasingly, ecological niche is treated by researchers as an evolving trait (Vieites *et al.,* 2009) which may be conserved (phylogenetic niche conservatism; Harvey & Pagel; 1991) or may rapidly change. Ancestral state reconstruction is one method used to understand how traits have changed (Cunningham *et al.,* 1998).

The *contMap* function in the *phytools* package for R (Revell, 2012) was used to visualize ancestral states. This function estimates states at internal nodes using maximum likelihood via the *fastAnc* function and uses Felsenstein's contrasts algorithm to infer states along edges

(Revell, 2013). The *fastAnc* function was further used to extract ancestral climatic niche values, which were evaluated using three strategies: (1) r^2 and slope from the relationship of GBIF-inferred ancestral states to BONAP-inferred ancestral states, (2) a paired t-test and (3) sum of squared differences calculations. By calculating r^2, the similarity of each set of reconstructed states are to each other was evaluated. Furthermore, the slope of the relationship was used to determine if one set was consistently predicting higher or lower ancestral values. Using a paired t-test, means were tested for a significant difference. Finally, by calculating the sum of squared differences for each set and the sum of squared differences between sets, levels of variance between and within sets were evaluated.

2.2.9 Additional analyses

To determine if BONAP data were consistently underestimating the elevation niche of *Carex* species, a supplementary regression analysis was conducted. For each species, elevation inferred via BONAP data was plotted against elevation inferred using GBIF occurrence points. A regression line was plotted and compared to a line with slope of 1.

Uncertainty in ancestral state reconstructions was assessed using the *fancyTree* function in the *phytools* package. Ancestral states were plotted and the 95% confidence interval of state estimates was visualized using shading.

While the eastern United States are well divided into small counties, the western United States (west of the Mississippi river) are characterized by fewer and larger counties. To determine if a relationship exists between geography and difference in climatic niche estimates, species were split into eastern and western species (widespread species were not considered). Their differences in climatic niche space were plotted using a boxplot.

2.3 RESULTS

2.3.1 Principal component analysis

The first two principal component axes explained 74% of variance (PC1=40%, PC2=34%) and the first three principal component axes explained 91% of BIOCLIM data variance (PC3=18%). Additional components explained less than one variable's worth of information (1/19 variables = 5.3% per variable; PC4=5%, PC5=2%, PC6=1%). Temperature range and seasonality contributed most to PC1, followed closely by temperature measurements of dry and cold periods and precipitation measurements of cold and wet periods. Precipitation of warm and dry quarters and temperature of warm and wet quarters contributed most to PC2. All PC1 contributions were positive and PC2 contributions were negative. Lastly, diurnal range and maximum temperature of warmest month contributed strongly to PC3. Contributions were both positive and negative. Details on principal component loadings for the first three PC axes can be found in Table 2.1.

Only the first two principal components were considered in calculation of climatic niche because these components contain the majority of the BIOCLIM information (74% of variability, 10 of 12 BIOCLIM variables loaded). Although PC3 explains more variability than would be expected if all variables contributed equally (18%>5.3%), its inclusion would impede interpretation and complicate computation.

The principal component biplot obtained by plotting climate PC1 against climate PC2 (Fig. 2.1) shows 'curvilinear distortion' (Minchin, 1987) or a 'horseshoe artefact'. This is a recognized phenomenon in ecology, resulting from principal component transformation of long-distance ecological gradient data (Gauch *et al.,* 1977; Hill & Gauch, 1980). Species on the ends

of the gradient experience climates so extreme, they begin to resemble each other after variables are transformed. Biological interpretation should not be drawn from the arch pattern. Furthermore, it should be noted that extremes of the data may appear more related than expected.

In Figure 2.1, each species is represented by one GBIF climate estimate and one BONAP climate estimate, connected by a change vector. From climate PC2=2 to 6, GBIF positions are well dispersed, while BONAP positions cluster around PC2=3. This results in a group of vectors pointing downwards. For estimates where PC2<0, GBIF and BONAP positions track each other closely. Very few points occur from PC2=0 to 2. Along PC1, estimates are distributed evenly. GBIF estimates tend to occur at higher PC1 values than BONAP estimates, resulting in vectors directed towards the left of the plot. The outlier on the far right of Figure 2.1 (PC1=~10, PC2=0) is *Carex anthoxanthea*, a West Coast endemic, found from Alaska to Oregon.

2.3.2 Sampling effort influencing error

The negative correlation between GBIF county coverage and differences in GBIF and BONAP climatic niche estimates was weak and only slightly significant at the 5% level (Pearson's linear coefficient correlation ρ=-0.13, p=0.009). County coverage explained only 1.8% of the climatic niche estimate differences (r^2=0.018). With low coverage (<20%), differences were highly variable (from 0 to 8 PC units). However, with high county coverage (>80%), differences in climatic niche space were less variable (0 to 2 PC units).

The degree of geographic spread in GBIF locality data had no significant relationship with differences in GBIF and BONAP climatic niche estimates. Errors were distributed similarly regardless of the relative size of occurrence minimum convex polygons.

Outliers in Figure 2.2B and D were *Carex livida, C. salina, C. micropoda* and *C. viridula*. These species had few characteristics in common. They ranged from hybrid to strictly defined species, and from specialists to widespread in distribution.

2.3.3 Elevation influencing error

Elevation niche and climatic niche estimate differences were significantly correlated (ρ=0.36, p=1.405x10^{-12}). Elevation explained 13% of differences in climatic niche space (r^2=0.13). Higher elevation species were significantly biased in one direction (~200°-300°; ρ=0.23, p=1.316x10^{-5}, r^2=0.05), which further analyses revealed to be downwards in altitude (Fig. 2.5). Outliers seen in Figure 2.3 were *Carex livida, C. salina* and *C. micropoda*. As previously described, no common identifying features of these species could be determined.

2.3.4 Ancestral state reconstruction

Overall patterns in the *contMap* visual representations of ancestral state reconstructions were similar between the GBIF and BONAP datasets. Nodes deeper in the tree, although more subject to uncertainty, had comparable inferred ancestral values. The regression of BONAP states to GBIF states showed that BONAP inferred states were consistently lower than GBIF states (slope=0.902; r^2=0.829). Means between ancestral states sets were significantly different (paired t-test p=2.2x10^{-16}). However, variance between sets of ancestral values was lower than within sum of squared differences (GBIF SSD=698; BONAP SSD=685; SS differences=218).

2.3.5 Additional analyses

Figure 2.5 showed that BONAP data consistently underestimated the elevation niche of *Carex* species (i.e., the GIBF:BONAP regression line was more shallow than the line of elevation equivalence). Elevation was most underestimated when species' GBIF inferred elevation approached 500m, at the intersection of the regression line and the elevation equivalence line.

Uncertainty in ancestral state reconstructions was large, and increased from the tips of the tree to the root (Fig. 2.6).

When comparing the magnitude of differences in niche estimates between western and eastern *Carex* species, substantial overlap in interquartile ranges was seen (Fig. 2.7).

2.4 DISCUSSION

GBIF data are commonly used to estimate species' climatic niche in contemporary ecology and evolutionary biology (>150 scientific papers returned in a search of 'climatic niche' in the GBIF Mendeley Public Library). Using a phyloclimatic perspective, this study evaluates GBIF occurrence data against county-level species distributions in the US. Specifically, climatic inferences from GBIF data and BONAP distribution data are compared, and the influence of sampling effort and elevation on differences in estimated niche is evaluated. Finally, the degree to which differences in climatic niche estimates alter ancestral state reconstruction is assessed.

This research is timely for three reasons. Firstly, relatively little attention has been paid to how biases in GBIF data affect comparative phylogenetic analyses as most authors have focused on the effects of biases on bioclimatic modeling (e.g., Kadmon *et al.,* 2004; Newbold, 2010). Secondly, the integration of geographic, environmental and phylogenetic data has been identified as an important step for future research addressing questions about the role of geography and ecology in diversification (Peterson *et al.,* 2010). Lastly, the climatic niche estimates obtained from GBIF in this work will be used in a further study (Chapter 3), investigating the relationship between the rate of climatic niche evolution in *Carex* and species richness. An understanding of the limits and errors of GBIF data is critical for conclusions to be drawn from trait evolution analyses. This study is also the first to evaluate the utility of GBIF data for climatic niche estimation on a broad geographic scale with over 360 species.

A negative relationship between the magnitude of differences and sampling effort (captured via county coverage and geographic spread) was expected. This is because as sampling effort is reduced, the accuracy of GBIF inferences should decrease as a result of spatial bias or poor representation of species distributions. Greater deviations of GBIF from BONAP climatic

estimates (and vice versa as sampling effort increases) should ensue. This is an intuitive expectation, which has been supported by a study evaluating the spatial bias of dung beetle museum records (SCAMAD database) in Madrid and its effects on environmental niche estimates (Hortal *et al.,* 2008). Yet, other studies investigating how spatial biases in museum data affect niche estimates in Israel (Kadmon *et al.,* 2004) and Egypt (BioMAP project; Newbold, 2010) have shown that spatial bias of occurrence data does not result in substantial environmental bias. The analysis presented here aligns with studies showing that bias does not affect niche estimates (Kadmon *et al.,* 2004; Newbold, 2010) and does not align with the study described above (Hortal *et al,*. 2008), showing bias significantly affects niche estimates. A strong relationship between sampling effort and niche differences was not recovered (Fig. 2.2). This means that for a large number of species in the genus *Carex,* GBIF data with poor coverage of the known range and geographically restricted occurrence records (i.e., low GBIF sampling effort) predict climatic niche as accurately as county centroids. The conflicting results between Hortal *et al*. (2008) and the results presented here may be due to geographic scale. Hortal *et al.* (2008) focused on Madrid and its surroundings, a limited geographic scope leading to more vulnerability in climatic niche estimation error. Furthermore, the city of Madrid may experience an Urban Island Effect, where climate within the city boundaries differs substantially from elsewhere (Ferguson & Woodbury, 2007). Lastly, it should be noted that these three studies (Hortal *et al.,* 2008; Kadmon *et al.,* 2004; Newbold, 2010) examined spatial bias rather than sampling effort. This may account for differences in results.

Rather than studying spatial bias directly, this study examined sampling effort. Sampling effort is more easily quantified on a broad geographic scale and more readily estimated by potential users of GBIF data. Therefore, it may be a more practical GBIF data assessment metric.

However, species with similar levels of sampling effort may have varying degrees of spatial, environmental, taxonomic and temporal bias. To measure spatial bias in our dataset, GBIF estimates should be evaluated against a rigorously verified fine-resolution occurrence dataset. An understanding of the relationship between spatial bias and sampling effort would enhance this work. Unfortunately, collecting high quality occurrence records on this scale would require substantial time and effort.

Although the relationship between county coverage and differences was only weakly significant, a trend in variance of differences can be observed (Fig. 2.2B). With low coverage (<20%), differences were highly variable, while with high coverage (>80%) differences were consistently under 2 PC units. This suggests that while improving sampling effort does not consistently reduce errors, it does decrease the frequency of unreliable estimates when many species are considered. This finding is important for potential users of GBIF data; at low distribution coverage levels, risk of inaccurate niche estimates is high, and action should be taken to ensure high quality niche data (see Recommendations section below).

The lack of relationship between geographic spread and niche differences, and the presence of several outliers (Fig. 2.2D), suggests that *Carex* GBIF data may contain misidentifications or other erroneous records, despite GBIF data cleaning conducted pre-analysis (removal of duplicates, imprecise records, and exclusion of points outside of US continental borders). These records are undetectable without manual verification all data points, which is unfeasible. Sedges are known for their reduced morphology and challenging identification (Starr & Ford, 2009). A dissecting microscope, a detailed flora, and a knowledge of sedge-specific terminology are all require to identify a species. Even with the necessary knowledge and equipment, it is common for non-sedge experts to misidentify *Carex* species. These specimens

and their identifications are rarely reviewed or revisited once stored in collections. Other sources of erroneous GBIF points include incorrectly entered label information or mistakenly georeferenced localities, usually the result of difficult-to-read labels from older specimens. Unfortunately, a few erroneous GBIF records in key locations can be detrimental to the accuracy of climatic niche estimates and will go overlooked unless manually verified. In this study, incorrect GBIF records may hide inside minimum convex polygons and will yield identical relative polygon areas yet significantly different niche estimates. This was observed in Fig. 2.2D, and underscores a limit of GBIF data; the persistent risk of incorrect specimen records.

While the results suggest that undetectable and incorrect GBIF records exist in the data, it is also important to note the number of detectable and incorrect records that were removed pre-analysis. The restriction of GBIF data to the continental US using the *pt.in.poly* function revealed 3086 erroneous records of 60215 total (4.88%). In a study geographically validating GBIF data using country borders for 6147 Fabaceae (legumes) species (i.e., a methodology identical to this study), Yesson *et al.* (2007) found 16% of GBIF records to be geographically incorrect. Thus, the percentage of geographically incorrect points in this study is low in comparison. This may be due to differences in data providers or geographic extent (American versus global GBIF data). The United States and other western countries tend to submit more data to GBIF (Yesson *et al.,* 2007) and have more resources to verify data. However, the nature of the incorrect localities aligned well with the findings of Yesson *et al.* (2007). Often, occurrence points were located just off land. A handful of points were placed in the middle of the ocean or in incorrect countries. Several records were missing negative signs on longitude data or had latitude and longitude records reversed. Although these points can be identified by comparing coordinates to text locality descriptors (continent, country, locality) or by verifying points' membership of country

polygons (e.g., by using TDWG4 areas as implemented by Yesson *et al.,* 2007), substantial effort is required to set up this data verification pipeline. Furthermore, once points are identified, deletion or manual verification, if the quantity of GBIF data is low, is required. Unfortunately, near valid points will go undetected by these strategies and will continue to bias niche estimates.

BONAP county data represent known species distributions at a varying resolutions. County sizes range from 59.13 km^2 (New York County, New York) to 51,947.24 km^2 (San Bernardino County, California) and are on average 2 584 km^2 (United States Census Bureau, 2010; Alaska is not divided into counties and is not included). As a result, the spatial resolution of the BONAP centroid data varies. This creates unequal point accuracies and regional representations. Additionally, county size varies by state. Counties east of the Mississippi River are on average smaller than counties to the west (average east=1558.9 km^2, average west=5543 km^2; United States Census Bureau, 2010). This may have effects on niche estimation (climate in the east is predicted at a finer scale) and county coverage measurements (Fig. 2.7). The transformation of BONAP data into point data via centroids may also incur inaccuracies associated with habitat. For example, *Carex scoparia* is a meadow species. However, the centroid data calculated using BONAP maps may fall in forests, wetlands, mountains or other types of unrealistic habitat types. Most notably, elevation plays a large role in determining microclimate. Elevation on mountain terrain drops off drastically, such that a drop from 2,000 m elevation to 500 m within kilometres is possible. At higher elevations, climate is cooler and precipitation levels are affected.

BONAP data have the potential to supplement GBIF distributions in situations where GBIF data are particularly skewed or incomplete. However, BONAP data may be biased for species occupying high-elevation niches (restricted areas with drastic differences in climate) due

to the coarse county-level scale of climate estimation. This study tested how differences in climatic niche estimates from GBIF and BONAP data vary with elevation. Species inhabiting high elevations should be poorly estimated by BONAP county centroids, thus producing an increase in the magnitude of differences. A highly significant relationship was found ($p=1.405x10^{-12}$) where 13% of the variability in differences was explained by elevation (Fig. 2.3A). Furthermore, high elevation species were found to be biased in one direction (~200°-300°), and low elevation species were found to avoid this direction (Fig. 2.3B). BONAP centroids underestimate elevation niche, as expected (Fig. 2.5). These results indicate that county data become less accurate as species inhabit higher elevations, and that a large proportion of differences between GBIF and BONAP estimates can be attributed to this effect. Specifically, a split of species into two groups with different directions of bias appears in Figure 2.3B at ~1500m. Those that are biased at 200-300° occur at ~1500m and above, and those that are not occur at ~1500m and below. Therefore, it is recommended that species occupying niches at 1000m or above should not be supplemented with county data (see Recommendations section). The elevation 1000m is suggested to remain conservative.

To determine how differences in niche estimates affect phylogenetic comparative analyses, ancestral states were reconstructed throughout a *Carex* species phylogeny using the two climatic niche PC1 datasets (Fig. 2.4). In general, ancestral states constructed from BONAP data were lower than GBIF inferred states (Fig. 2.4 regression). This may be due to the elevation bias of BONAP data (bias inwards of GBIF to BONAP points in niche space in Fig. 2.1; directional bias in Fig. 2.3B). However, the downwards trend of BONAP ancestral states only explains 82.9% of the differences ($r^2=0.829$). Remaining error may be due to inaccuracies of GIBF data.

Means between the two sets of inferred ancestral states were significantly different (paired t-test $p<2.2x10^{-16}$). This may be explained by the uniform shift downwards of BONAP states. Although the means of the sets were significantly different, general patterns appeared similar. Therefore, it is likely that largely similar conclusions would be drawn from visualizations of patterns in *contMap.*

When the variance in each set and between sets was calculated using the sum of squared differences, the variance within sets was three times greater than the variance between sets. Differences between ancestral estimates were less pronounced deeper in the tree, and this was accompanied by larger uncertainty in inferred states (Fig. 2.6). Precise estimates of reconstructed states deeper in phylogenies have been shown to be less reliable than near the tips (Oakley & Cunningham 2000). If error within sets of estimated ancestral states is greater than in the raw data, and uncertainty in deep nodes is large, attaining precise and accurate climatic niche estimates may be less critical than one might assume (with the exception of severely incorrect GBIF data, e.g. points in the middle of the ocean). Furthermore, uncertainty in phylogenetic comparative methods are accompanied by uncertainty in phylogenetic trees. All reconstruction methods result in phylogenies with some degree of inaccuracy and uncertainty. The results presented here suggest that climatic niche evolution research may not be as affected by inaccuracies in distribution data as the uncertainties of tree topology and phylogenetic models.

2.4.1 Recommendations

Although only a weak relationship between county coverage and niche differences was recovered, at low county coverages (<20%) differences in niche space were variable. To ensure good quality climatic niche estimates, GBIF users are encouraged to supplement their GBIF data from other occurrence data sources when coverage of the known distribution is less than 20%. The Biota of North American Program (BONAP) may be a valuable source of occurrence data for plant species occurring at low elevations. However, GBIF users are advised to avoid these data if the mean value of the species' elevation range from the Flora of North America is above 1000m.

For large scale studies, researchers should consider using BONAP data rather than GBIF-enabled data for climatic niche estimation. This is because results obtained from BONAP data will be similar and conclusions may be identical. Perhaps most importantly, a substantial amount of time and effort will be saved. GBIF data require cleaning to remove duplicates and imprecise records, to verify validity of records, and to remove erroneous points. Furthermore, GBIF data contain misidentifications and misapplications, old records since identified as false, and show errant reports (Kartesz, pers. comm.). BONAP is a smaller, cleaner dataset. Unfortunately, BONAP data are not publicly accessible at this time. Communication with the data providers (J. Kartesz at the Taxonomic Data Center, Chapel Hill, NC) is required to obtain permission to use BONAP data. Improved mobilization of the data should be prioritized to allow for future broad scale use.

Lastly, users of GBIF-enabled climatic niche estimates in phylogenetic comparative analyses should note that different distribution datasets and niche estimates will yield different results, although similar conclusions can still be made despite uncertainties.

2.5 TABLES AND FIGURES

Table 2.1 Loadings from principal component analysis of 19 WorldClim BIOCLIM variables estimated from GBIF and BONAP species localities. Only those principal components explaining a larger than expected amount of variation (=5.3%) are included (PC1=40%, PC2=34%, PC3=18%). Loadings are bolded if they are important contributors. This was assessed by comparing against the expected loading value if all variables contributed equally (square root of 1/19=0.229415). BIOCLIM variables are sorted according to contributions to PC1, PC2 and lastly PC3.

BIOCLIM Variable	PC1	PC2	PC3
BIO7 = Temperature Annual Range (BIO5-BIO6)	**0.347006**	-0.02357	0.065644
BIO4 = Temperature Seasonality (standard deviation *100)	**0.333774**	-0.11308	-0.04759
BIO9 = Mean Temperature of Driest Quarter	**0.318571**	-0.00301	0.149076
BIO6 = Min Temperature of Coldest Month	**0.313137**	-0.15464	0.167593
BIO19 = Precipitation of Coldest Quarter	**0.307952**	0.043257	**-0.25993**
BIO13 = Precipitation of Wettest Month	**0.306882**	-0.02487	**-0.24156**
BIO16 = Precipitation of Wettest Quarter	**0.29966**	-0.01965	**-0.26236**
BIO11 = Mean Temperature of Coldest Quarter	**0.288006**	-0.17444	0.227529
BIO3 = Isothermality (BIO2/BIO7) (* 100)	**0.286909**	0.156302	0.204817
BIO18 = Precipitation of Warmest Quarter	-0.07676	**-0.35767**	-0.07555
BIO17 = Precipitation of Driest Quarter	-0.02394	**-0.34506**	-0.19596
BIO10 = Mean Temperature of Warmest Quarter	0.034194	**-0.3438**	**0.242406**
BIO14 = Precipitation of Driest Month	-0.04488	**-0.3421**	-0.18639
BIO8 = Mean Temperature of Wettest Quarter	-0.09437	**-0.30209**	0.212193
BIO1 = Annual Mean Temperature	0.174711	**-0.29373**	**0.249478**
BIO15 = Precipitation Seasonality (Coefficient of Variation)	0.180441	**0.283397**	0.146895
BIO5 = Max Temperature of Warmest Month	0.031265	**-0.27722**	**0.354277**
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))	-0.01996	0.192329	**0.401859**
BIO12 = Annual Precipitation	0.203849	-0.22602	**-0.3181**

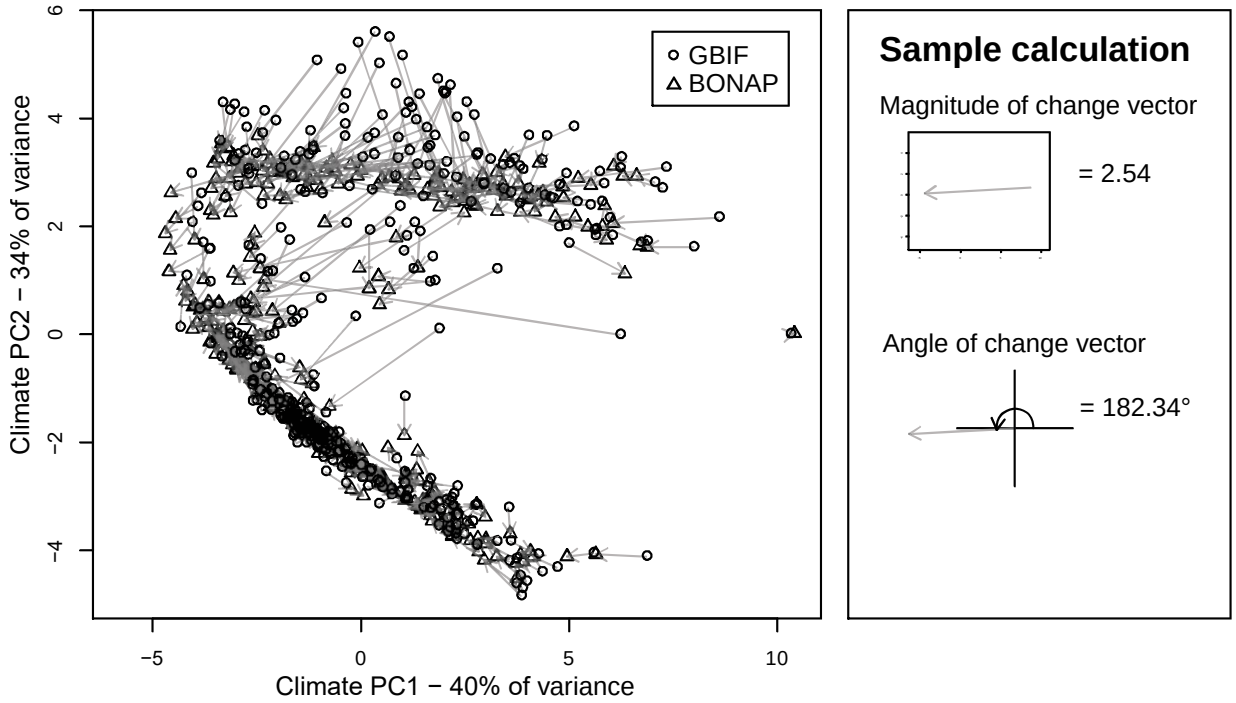


Fig. 2.1 Principal component (PC) climatic niche space. PCA was used to condense 19 WorldClim BIOCLIM variables into two PC axes (PC1=40% and PC2=34%) explaining the majority (74%) of bioclimatic variance. Species' GBIF estimates in niche space are represented by circles and BONAP points are represented by triangles. Grey vectors connect species' GBIF estimates to BONAP estimates, from which magnitude and direction are calculated. A sample calcuation in provided on the right. N=378.

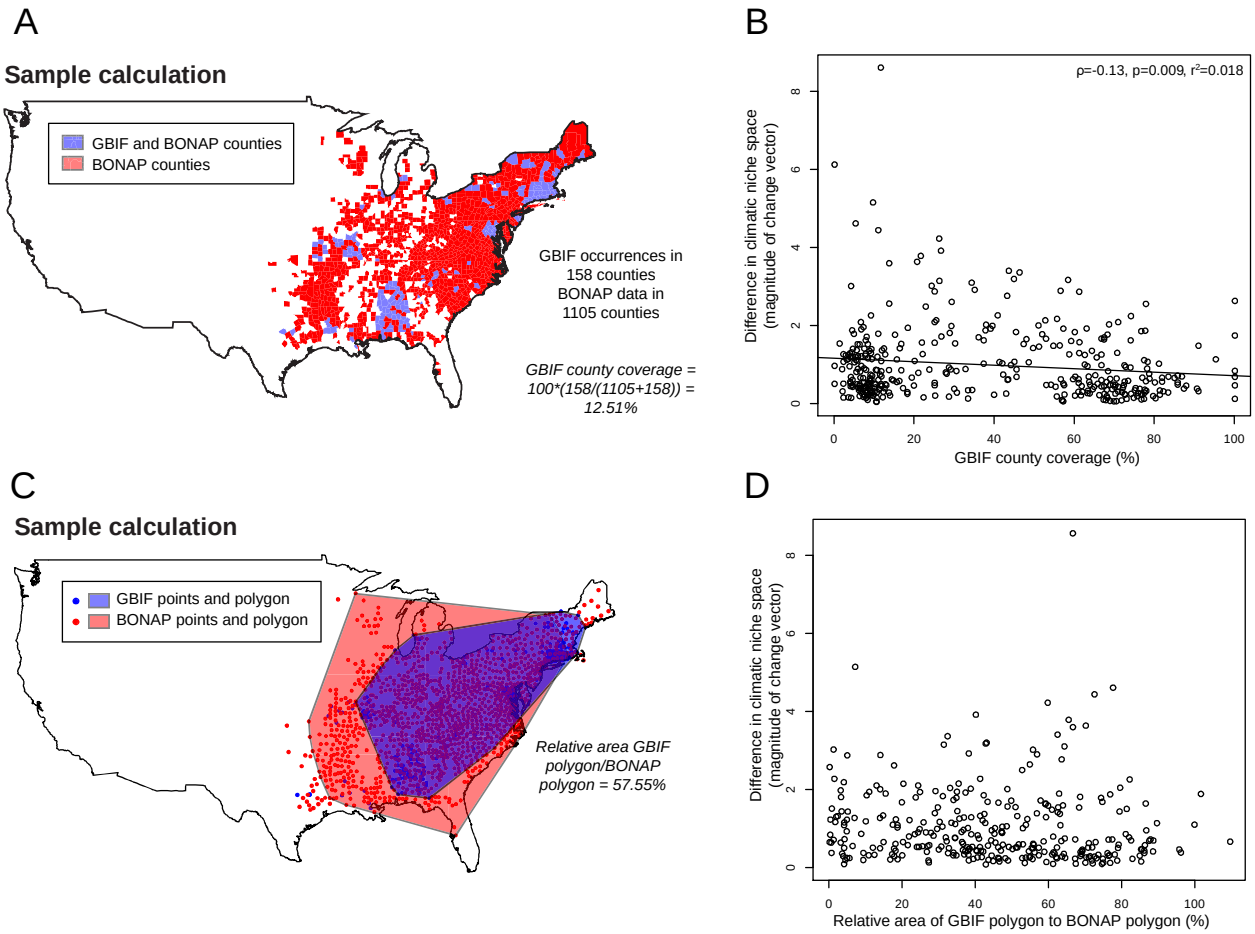


Fig. 2.2 The relationship between sampling effort and magnitude of differences between GBIF and BONAP climatic niche estimates. Regression lines are presented if a significant correlation was found. Correlation coefficients (ρ), p-values and r^2 values are indicated in the top right-hand corner of plots with regression lines. (A) Sample calculation of GBIF county coverage for *Carex lurida*. Counties included in *C. lurida*'s GBIF data distribution are in blue and *C. lurida* BONAP county data are in red. (B) The relationship between GBIF county coverage and differences in climatic niche space. Each circle is one *Carex* species. N=363. (C) Sample calculation of relative geographic spread for *C. lurida*. GBIF points are in blue and BONAP points are in red. Minimum convex polygons calculated from each dataset are visualized. (D) The relationship between relative geographic spread and differences in climatic niche space. Each circle is one *Carex* species. N=321.

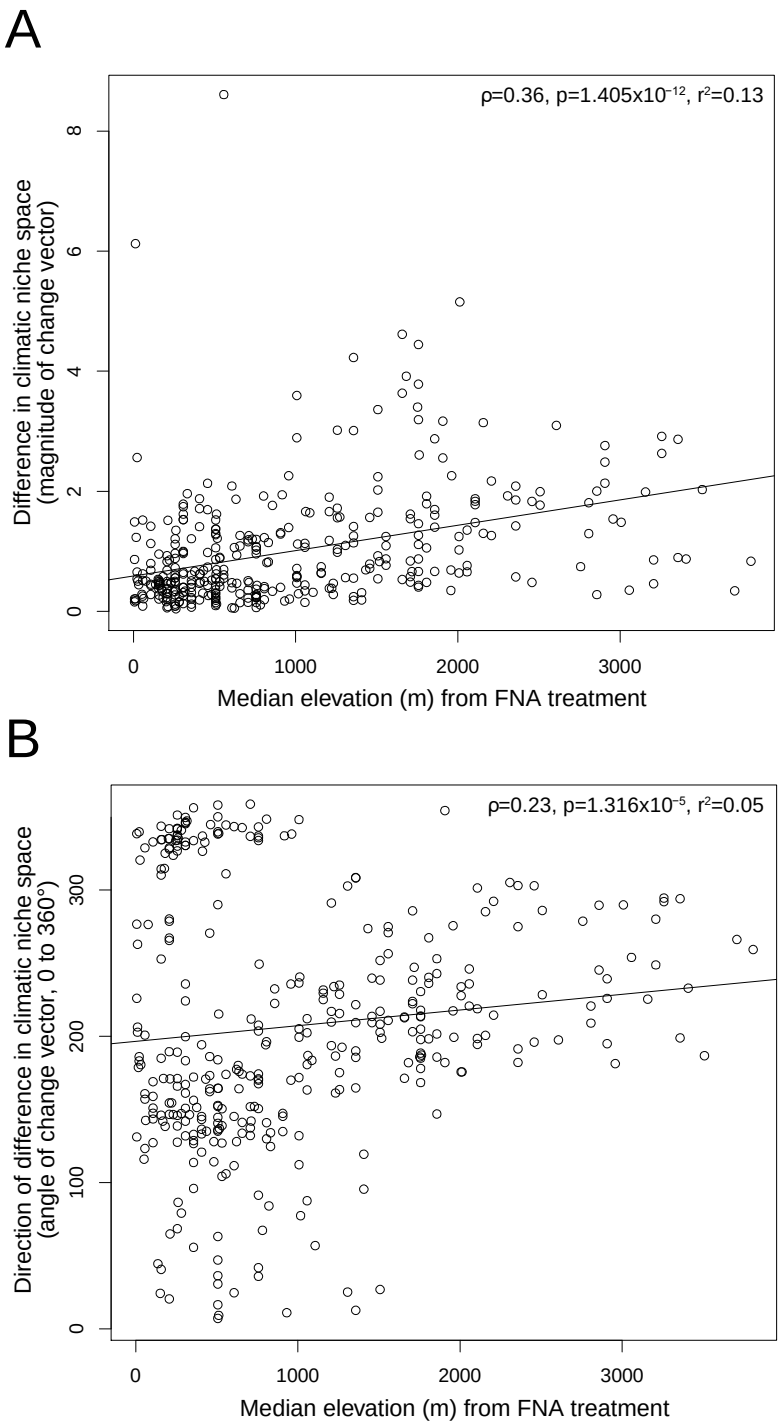
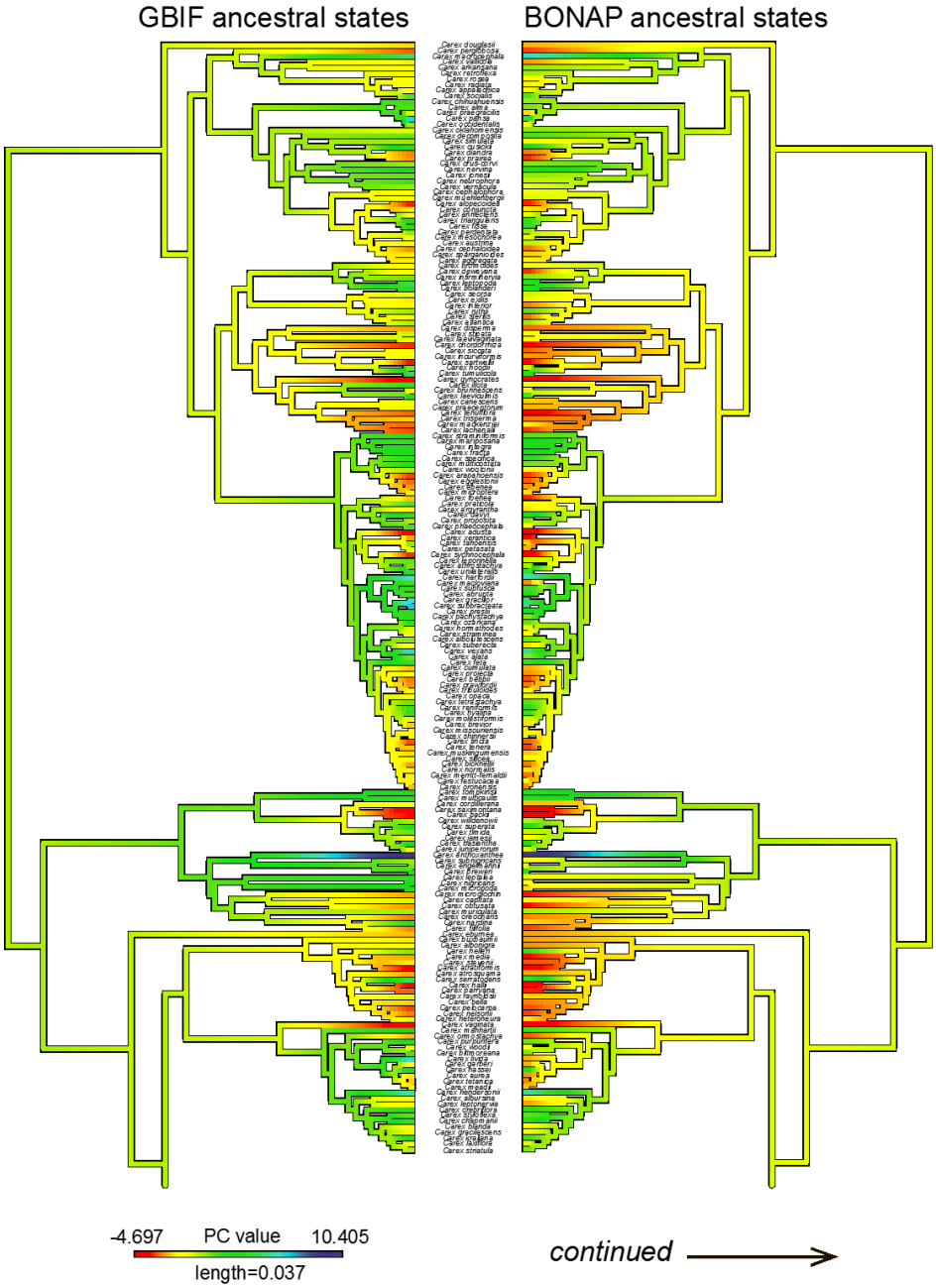


Fig. 2.3 The relationship between elevation niche and differences in climatic niche space. Regression lines are presented if a significant correlation was found. Correlation coefficients (ρ), p-values and r^2 values are indicated in the top right-hand corner of plots with regression lines. Each circle is one *Carex* species. N=363. (A) Elevation plotted against magnitude of differences in PC climatic niche space. (B) Elevation plotted against direction of differences in PC climatic niche space. Direction is in degrees.


GBIF ancestral states
BONAP ancestral states
-4.697
PC value
10.405
length=0.037
continued

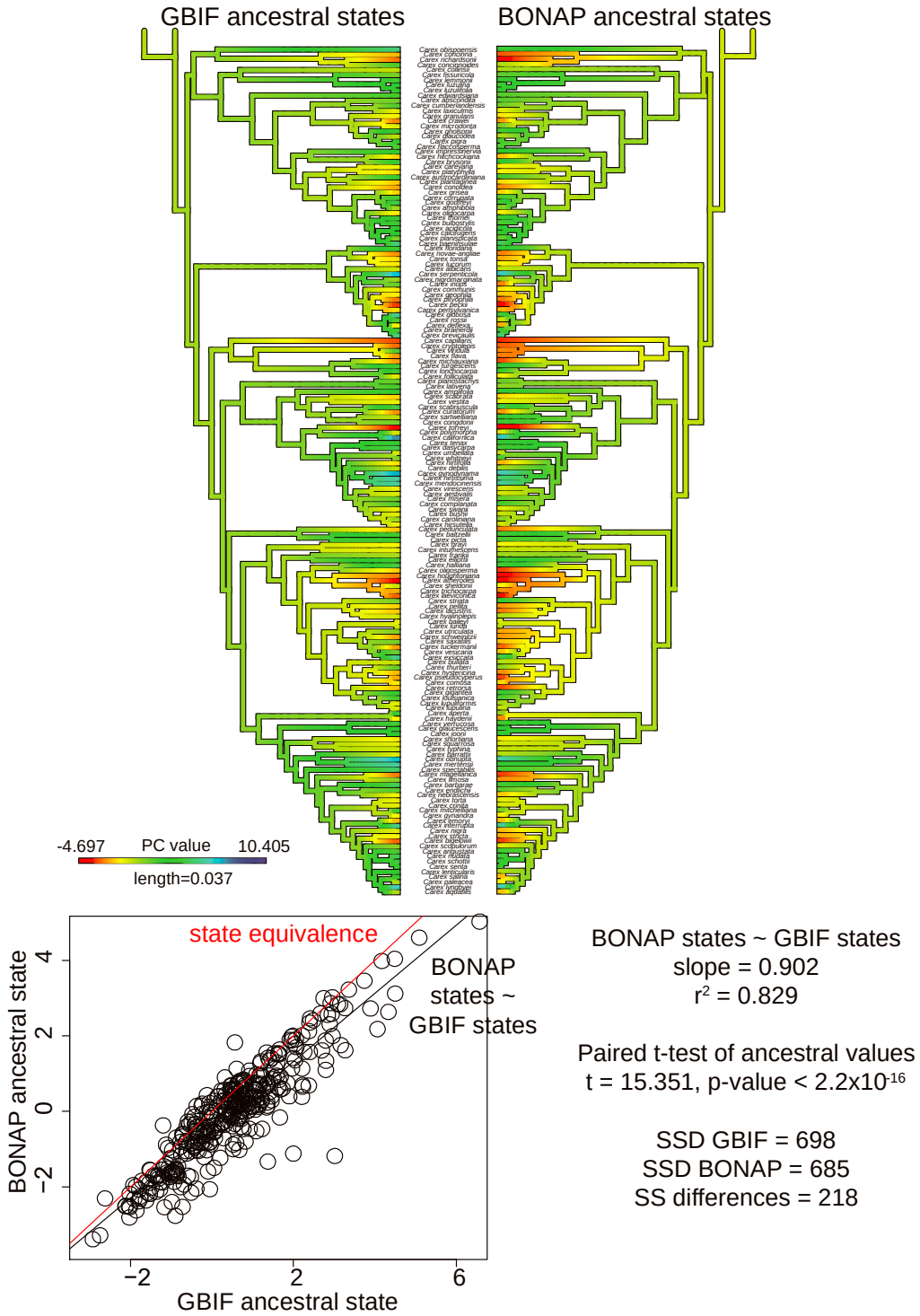


Fig. 2.4 Ancestral climatic niche PC1 reconstruction using GBIF estimates (left) and BONAP estimates (right). Colours represent trait values. The reconstruction was visualized on a *Carex* phylogeny using the *contMap* function in R. A plot of BONAP ancestral states versus GBIF ancestral states is presented at the bottom, showing a slight offset from state equivalence. Further statistics evaluating ancestral state differences are shown (see Methods for details).

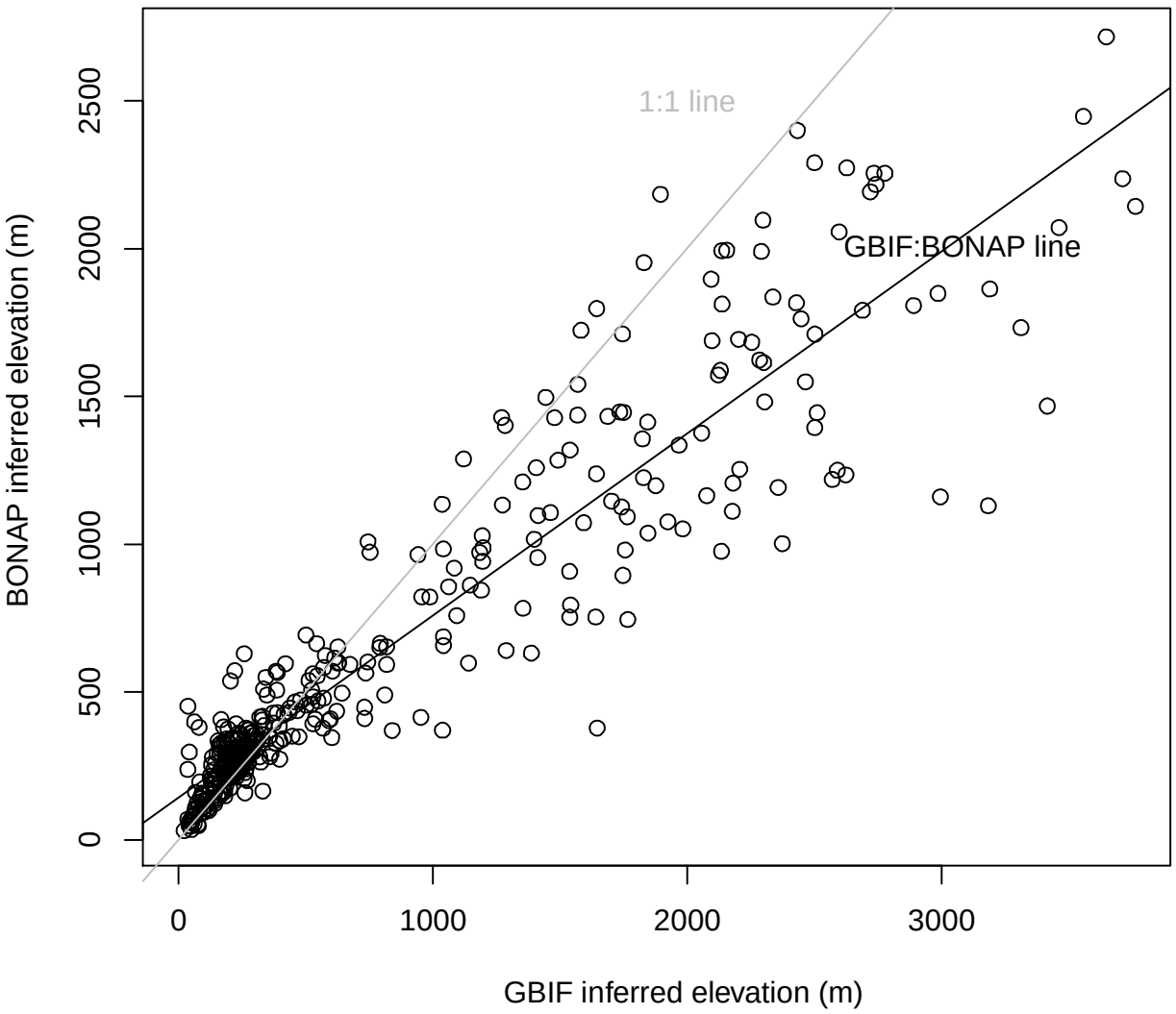


Fig. 2.5 The relationship between elevation predicted from GBIF occurrence data and elevation predicted from BONAP county data. A regression line is presented in black and a 1:1 elevation equivalence line is in grey. Elevation data were obtained from WorldClim at a 5km resolution and were combined with locality data to calculate elevation estimates. Each circle is one *Carex* species. N=378.

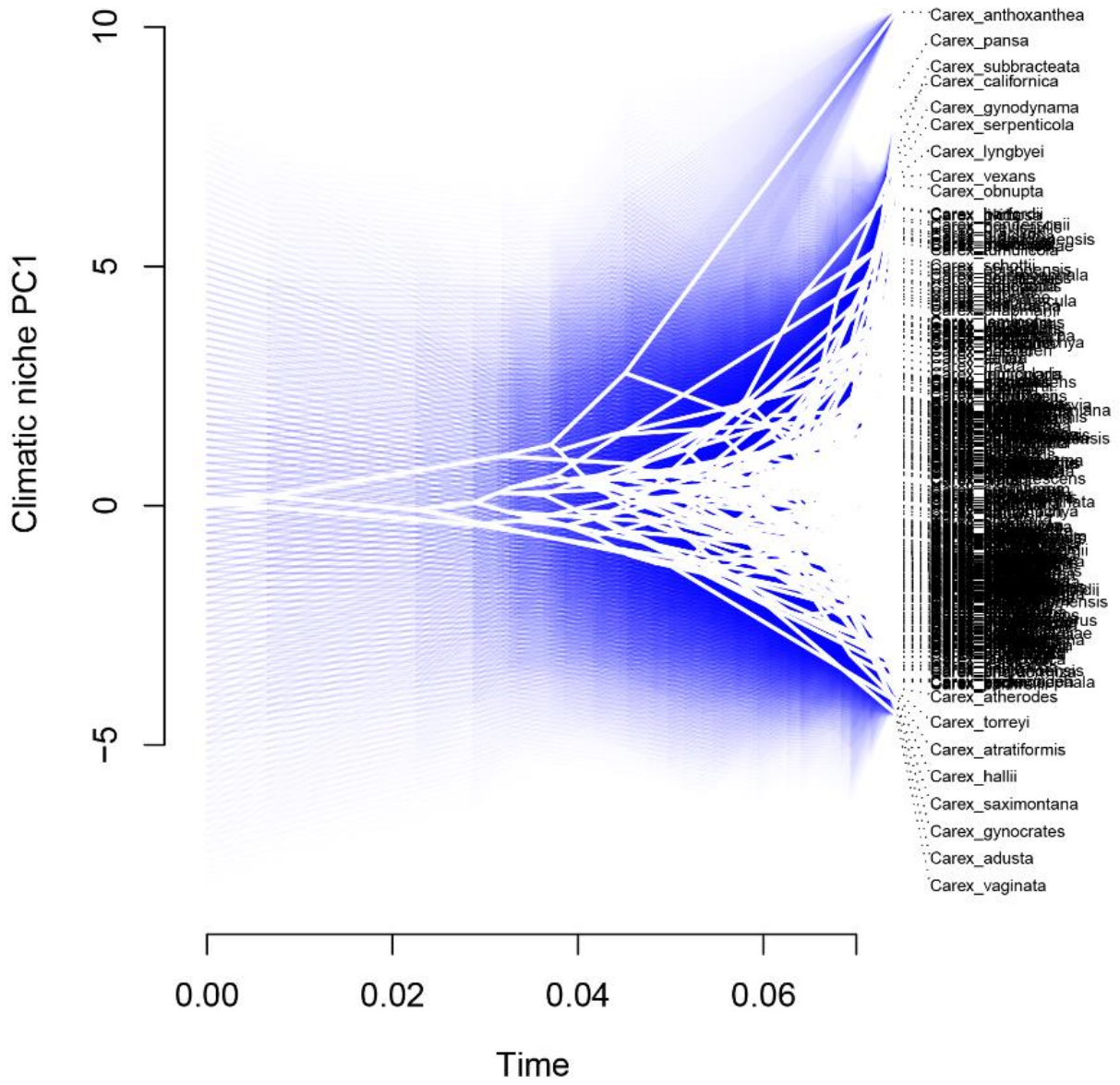


Fig. 2.6 Uncertainty in ancestral state reconstruction from GBIF climatic niche PC1 estimates. Evolutionary time is on the x-axis and trait values on the y-axis. Blue transparent lines extend outwards showing 95% confidence limits. Plot was generated using the *fancyTree* function in the phytools package for R. Uncertainty from reconstruction of BONAP climatic niche estimates appeared qualitatively identical and is therefore not presented. N=363.

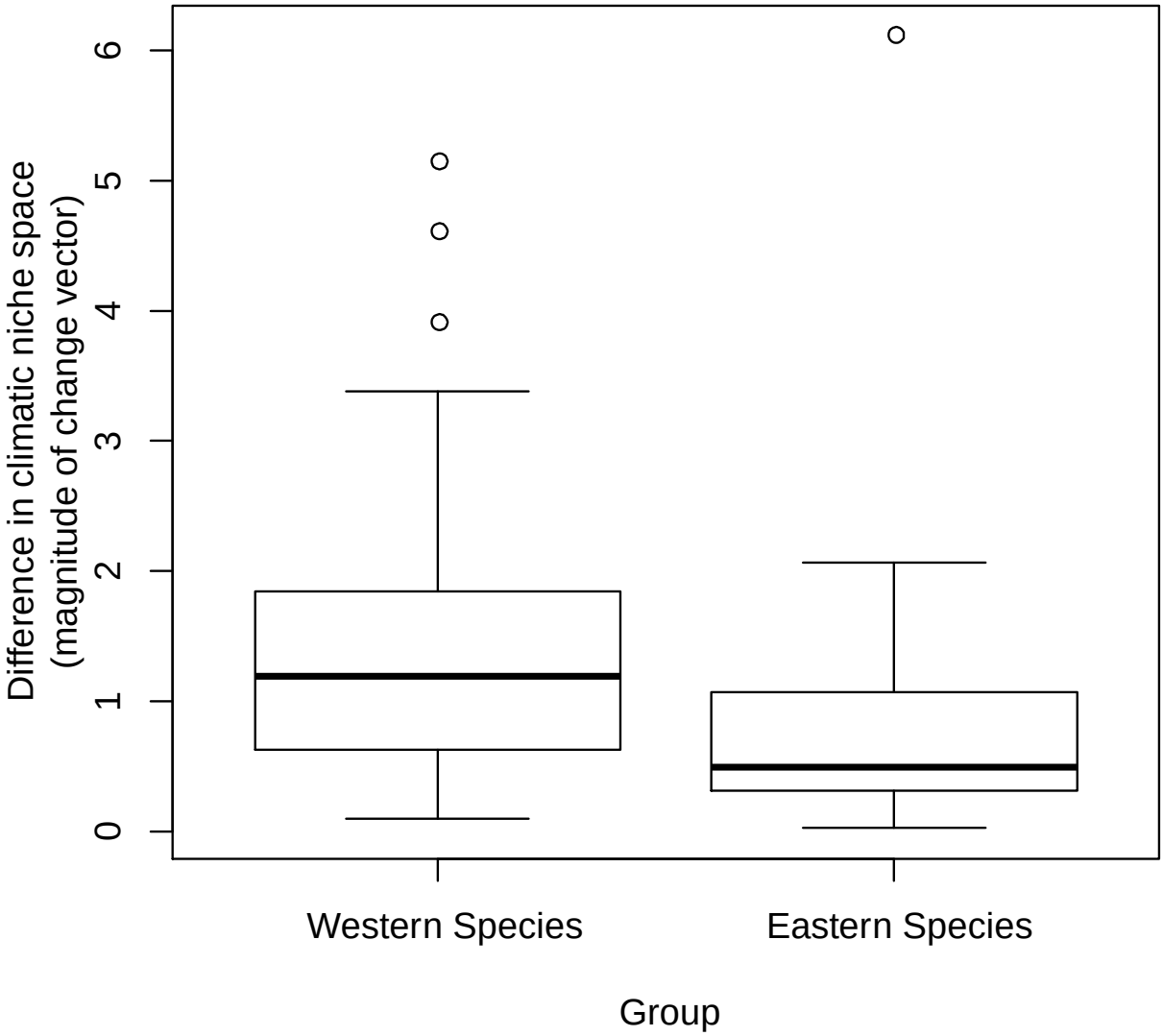


Fig. 2.7 Analysis of the influence of geography on differences in climatic niche estimates. Counties in the eastern US are on average smaller than in the west. Species were separated into strictly eastern and strictly western species. Widespread species were not included. N=179.

Chapter 3: Regional species richness and trait evolution in the genus *Carex* (Cyperaceae)

3.1 INTRODUCTION

Species richness is not distributed evenly across the tree of life. While some lineages have little morphological diversity and are species-poor (e.g., Darwin's 'living fossils'; Stanley, 1979), others undergo rapid and remarkable diversification (e.g., African cichlids, Andean lupines; Meyer *et al.* 1990; Hughes & Eastwood, 2006). Surprisingly, a clear understanding of this prominent macroevolutionary pattern and the processes shaping it, is still lacking.

Exceptional species richness may arise in multiple ways. Firstly, a group may achieve high species richness via its longevity (McPeek & Brown, 2007; Rabosky, 2009; Rabosky *et al.,* 2012). Older clades have had more time to accumulate species, and are expected to be more species-rich than younger clades. This age-species richness relationship has not been consistently observed across taxonomic groups, and researchers speculate diversity carrying capacities may exist (Ricklefs, 2008; Wiens, 2011; Rabosky *et al.,* 2012). Despite this, clade age is still considered to be a valuable explanatory variable of differences in diversity (McPeek & Brown, 2007).

Alternatively, a clade may attain high species richness through high rates of diversification. Diversification is a result of two processes, extinction and speciation. The alteration of either extinction or speciation may increase or decrease diversification rates. Specifically, a decrease in extinction rates and or an increase in speciation rates will result in increased diversification rates, and vice versa to reduce diversification rates. Uniform-rate

speciation models poorly describe phylogenetic trees (Heard, 1992; Ricklefs, 2003), and evidence for differential speciation and extinction through time is mounting (Jetz *et al.,* 2012, Rabosky *et al.,* 2013). Accordingly, several explanations for species richness patterns presuming differences in macroevolutionary rates have been proposed and tested.

Lineage-specific traits or 'key innovations' have frequently been invoked to explain species richness patterns (e.g., nectar spurs, hypocones, sexual dichromatism; Hodges & Arnold, 1995; Hunter, 1998; Seddon *et al.,* 2008). For example, Isaac *et al.* (2005) found that species richness in mammals was significantly correlated with length of gestation period and litter size, where lineages with short gestation periods and large litters were more species-rich. The authors speculated that these 'fast' life histories reduced the likelihood of extinction or increased the rate of speciation by means of a higher capacity to adapt and higher rates of evolution. However, they concluded that it would be difficult to unambiguously test this hypothesized mechanism.

Researchers have also proposed that some clades possess high species richness as a result of 'ecological opportunity' (Simpson, 1953; Schluter, 2000; Glor, 2010; Yoder *et al.,* 2010). Ecological opportunity arises when resources unused by other taxa become available, for example, after the arrival of a new type of resource, or after the colonization of an isolated area with a depauperate biota. For example, Hughes and Eastwood (2006) suggested that the plant genus *Lupinus* reached rapid diversification rates because of ecological opportunities provided by the recent uplift of the Andean mountains.

Most recently, workers have proposed that the capacity of lineages to evolve morphological or ecological novelty may regulate species richness (Lovette *et al.,* 2002; Adamowicz *et al.,* 2008; Pigliucci 2008, Rabosky *et al.,* 2013). This stems from the observation that many lineages have failed to diversify after the arrival of new ecological opportunities, or in

other situations where adaptive radiations are predicted (Lovette *et al.,* 2002). In fact, increased morphological or ecological flexibility is implicitly assumed by other well-supported hypotheses (e.g., key traits of organisms that increase diversity via increased rates of evolution). If species richness is influenced by a group's ability to evolve trait novelty, a coupling of high rates of trait evolution and high species numbers is predicted on a phylogenetic tree. With the rise in availability of time-calibrated phylogenies, hypotheses regarding diversity-trait evolution dynamics are more easily testable.

A small number of studies have investigated the relationship between morphological evolution and species richness in vertebrates (Adams *et al.,* 2009; Venditti *et al.,* 2011; Rabosky & Adams, 2012; Rabosky *et al.,* 2013; Rabosky *et al.,* 2014; Zelditch *et al.,* 2015). Using diversification modelling and phylogenetic comparative methods, these studies tested the relationship between rates of phenotypic evolution and speciation. They found conflicting results, where in some groups (plethodontid salamanders, mammals, Scincid lizards, squirrels; Adams *et al.,* 2009; Venditti *et al.,* 2011; Rabosky *et al.,* 2014, Zelditch *et al.,* 2015) rapid diversification has occurred with little morphological change (and vice versa) and in others rapid morphological change was correlated with species richness (plethodontid salamanders, ray-finned fishes; Rabosky & Adams, 2012; Rabosky *et al.,* 2013). More research on the relationship between morphological trait evolution and diversity is required, to determine if morphological lability plays an important role in diversification. Specifically, studies on plant genera are lacking.

Hypotheses linking trait evolution and species richness are not limited to morphological traits. The ability of species to rapidly adapt to new environmental conditions (i.e., a novel climate or habitat) may similarly be involved in diversification. In particular, Kozak & Wiens

(2007) proposed that rapid rates of climatic niche evolution found in tropical clades may be responsible for the high biodiversity observed in the tropics. High rates of climatic niche evolution indicate lineages are able to track changing environments or occupy unfilled niche space. These groups should experience reduced competition and increased ecological opportunities, resulting in reduced extinction or increased speciation (Kozak & Wiens, 2010). Further studies are required to determine if this mechanism is widespread across regions (e.g., in non-tropical zones) or taxonomic groups.

To understand diversity differences, it is most useful to look to groups that have unusual taxonomic richness. The genus *Carex* (sedges; family Cyperaceae) is characterized by a remarkable number of species (ca. 2100 spp.; Govaerts *et al.,* 2014; Global *Carex* Group, 2015), representing almost 40% of diversity in Cyperaceae. This stands in contrast to its sister clade Scirpeae, containing only 98 species in 9 genera (Léveillé-Bourret *et al.,* 2014). *Carex* is distributed globally, except Antarctica, and shows high ecological diversity (deserts to rainforests; Naczi & Ford, 2008). It is especially diverse in the Northern Hemisphere, where over 20 species can be easily found within a few hectares of a temperate forest (Ball & Reznicek, 2002). While it is highly diverse in temperate zones, it shows low diversity in tropical zones (i.e., it possesses a reverse diversity latitudinal gradient). Therefore, the genus *Carex* presents an opportunity for study of temperate diversity, which has received less attention than tropical diversity. Notably, it possesses the longest aneuploid chromosome sequence in living organisms (n=6-56; Roalson, 2008) due to holocentric chromosomes, a factor which may be involved in sedge diversification (Escudero *et al.,* 2012b).

Historically, classification of the genus *Carex* was difficult due to its reduced morphology (Starr & Ford, 2009). Although its potential as a study taxon for evolutionary

studies has been emphasized, due to high species richness and broad ecological breadth (Starr & Ford, 2009; Waterway *et al.,* 2009), confusing taxonomy remained a barrier. However, modern molecular tools have enabled resolution of phylogenetic relationships and revisions of *Carex* classifications (Starr & Ford, 2009). Concurrent methodological advances in computational biology (e.g., ease of diversification rate reconstruction; Garamszegi, 2014), and new phylogenetic research have enabled study of the evolutionary process in *Carex*, previously not possible (Starr & Ford, 2009).

The high species richness in *Carex* may be a result of either its age or of high rates of diversification. To determine if diversification processes are at play, one must first rule out the possibility that clade age accounts for its diversity. The first study attempting to reconstruct diversification rates in *Carex* employed a molecular phylogeny and modelled evolution as a stochastic, time-homogeneous birth-and-death process (Escudero *et al.,* 2012a). It estimated the age of divergence of its family Cyperaceae at 83.67 My, of the genus *Carex* at 42.19 My and of its immediate sister clade Scirpeae at 62.0 My. Escudero *et al.* (2012a) also found a 1.5 times increase in the net rate of diversification at the crown node of *Carex* over the background rate for Cyperaceae. This rate was higher than average for angiosperms (Valente *et al.,* 2010). In a subsequent study, Escudero & Hipp (2013) found that within Cyperaceae, clade age explains a significant proportion of the disparity in species richness (ca. 33% of among-clade variation in species richness). However, the authors emphasized the threefold increase in net diversification rate in the Cariceae clade (which includes *Carex*) over the background rate for Cyperaceae is not congruent with differences in clade age alone. When species numbers and ages are considered throughout Cyperaceae (diverging in the Eocene; clades at the tribe level), clades attain modest species numbers of 1 to 550, with the exception of tribe Cypereae, the second most diverse tribe

at 1117 species, about 700 of which belong to the genus *Cyperus* (Escudero & Hipp, 2013). *Carex* stands out with species numbers more than double that of other genera its age. These diversification rate reconstruction results and species number comparisons (Escudero *et al.,* 2012a; Escudero & Hipp, 2013) suggest that the diversity of *Carex* cannot be attributed to its age.

No apparent key innovations or ecological opportunities exist to explain the high diversity found in *Carex* (Gehrke & Linder, 2011). Escudero *et al.* (2012a) speculated that adaptive radiation following a climatic shift, aided by rapid chromosome evolution, may explain the latitudinal pattern of species richness in *Carex*. Yet, the temporal correspondence of a shift in diversification rates, a global cooling period and a chromosome number transition is limited evidence to explain patterns of sedge species richness. Further investigation of the response of *Carex* to evolving climates, and the effects on its species richness, is required.

The high number of species found in the genus *Carex* has been linked to its ability to occupy a wide range of habitats (Bell *et al.,* 2000; Gehrke & Linder, 2011). In a study using generalized linear models to predict *Carex* species richness in African sky islands (Gehrke & Linder, 2011), habitat heterogeneity was the best predictor (with lineage age and geographic distribution) of lineage size. The authors subsequently posed the question, are ecological or habitat factors influencing sedge diversity? Because of high clade conservatism of habitat and the lack of power of any single habitat factor in the models, they concluded that their results showed no support for ecological speciation (i.e., speciation via filling of different niches). Rather, they suspected that if allopatric or ecological speciation is acting in *Carex*, niche expansion may obscure any evidence. A phylogenetic approach is needed to uncover indications of ecological speciation processes. *Carex* is also notable for its frequent niche segregation along

environmental gradients (Vellend *et al.,* 2000; Dabros & Waterway, 2008; Waterway *et al.,* 2009). The genus' ability to finely partition niche space leads to the hypothesis that niche differentiation (i.e., ecological speciation) and the traversing of niche space via niche evolution may be important to the diversification process in sedges. This positions the group as an ideal candidate for niche evolution research. Further study on broader taxonomic and geographic scales with more complete sampling and more robust methods (e.g., take phylogeny into account) are needed in *Carex*.

The biogeographic and ecological context of diversification processes is important to consider when seeking explanations for diversity outcomes (Harmon & Harrison, 2015). For example, dispersal limitation (Hubbell, 2001; Ozinga *et al.,* 2005), environmental heterogeneity (recall Gehrke & Linder, 2011; Hortal *et al.,* 2009 and references therein) and habitat change influence diversity in space and these factors vary among regions. Furthermore, the size of the region in which a clade diversifies may have an effect on clade diversity (Ricklefs, 2006; Ricklefs *et al.,* 2007).

The existence of ecological limits or diversity carrying capacities on species richness at the continental scale has been the subject of recent debate (Harmon & Harrison, 2015; Rabosky & Hurlbert, 2015). Consistently observed resource-species richness relationships, the stationarity of species numbers throughout time, and the predictable responses of species richness to mass extinctions, key innovations and other disturbances have been used as evidence for limits to diversity at the continental level. If true, the number of *Carex* species on continents of its distribution may have upper limits and speciation processes may be diversity-dependent. This may confound attempts to investigate the generation of species richness patterns on a cross-continental scale.

Conversely, Harmon & Harrison (2015) maintain that no limits exist to diversity at the continental level and that continental communities are highly dynamic, unsaturated entities. Yet, attributes of the continent such as dispersal limitation, resource pulses or other habitat changes, and habitat heterogeneity, nonetheless influence the filling of niche space (Harmon & Harrison, 2015). Therefore, the region in which a clade has diversified may influence its richness. By restricting a study of species richness and its correlates to a singular continental region, any impacts of niche filling on diversification processes are eliminated. Despite the increasing recognition of the importance of the regional-level processes, few authors have considered diversification at this scale (Ricklefs, 2008).

This study focuses on the observed yet unexplained high species richness found in the genus *Carex*. It investigates the relationship between rates of morphological and ecological trait evolution (i.e., trait lability) and species richness. However, it restricts itself to a North American north of Mexico sampling, to avoid the influence of niche filling or other pan-regional factors. It hypothesizes that greater trait lability has led to increased *Carex* diversity, in a North American context. Therefore, it predicts a positive relationship between rates of trait evolution and species numbers, when comparing species groups within *Carex*. A species phylogeny is constructed for almost all North American *Carex* species (477 of 482 excluding subspecies designations, tallied from the FNA online; Ball & Reznicek, 2002) and rates of trait evolution are calculated. These are correlated against species numbers in a non-nested species group comparison framework.

3.2 METHODS

3.2.1 Trait definitions

For the purpose of this study, niche is defined as the set of attributes of the area occupied by a species across its geographic range, influencing its growth rate (the 'Grinnellian' niche; Grinnell, 1917; Soberón, 2007; Peterson, 2011). This definition, however, does not include attributes that are affected by or modified by the presence of the species (e.g., biotic interactions) and is strictly a property of the environment (Peterson, 2011). Furthermore, it does not encompass the set of conditions under which a species *can* persist, only those under which it currently exists (the realized niche; Hutchinson, 1957).

A species' Grinnellian niche differs from morphological traits by its noninteractive, nonlinked quality (Peterson, 2011). Morphological traits are defined as the structures of organisms, which can change in response to their environment and are properties of the species. In this study, trait is defined as a broad term encompassing both species' morphological and ecological attributes.

3.2.2 Flora of North America region

This study is focused on the *Carex* of the Flora of North America region (Canada, USA & Greenland; FNA). This restriction of geographic scope not only reduces the effects of niche filling, dispersal limitation, environmental heterogeneity and habitat change as outlined in the Introduction, but also ensures good data availability and quality, as climatic and habitat data are more extensive and precise for FNA *Carex* than for species found elsewhere on earth (e.g., GBIF and WorldClim, habitat databases). Furthermore, by comparison to *Carex* floras worldwide,

these sedges are taxonomically well known (Ball & Reznicek, 2002), thoroughly collected and precisely geolocated. Therefore, the sedges of the FNA region offer one of the best opportunities to explore the effects of climactic niche and character evolution on the species richness of a major clade.

3.2.3 Phylogeny

3.2.3.1 Taxon sampling

All genomic DNA samples were extracted by Chouinard (2010) from herbarium specimens acquired on loan from various herbaria across North America (see appendix of Chouinard 2010 for voucher details). The FNA was used as a checklist for *Carex* species sampling, with as many as 70 specimens received on loan and 22 specimens sampled per taxon (only one specimen was used per species in the final analysis). A maximum of 3 cm x 1 cm of leaf tissue was extracted from the specimens, preferentially drawn from recently collected plants whose leaves were dried in silica gel. DNA extraction by Chouinard (2010) followed the silica-column method of Alexander *et al.* (2007) with modifications to increase the yield and effectiveness of DNA amplification on old or difficult herbarium specimens.

3.2.3.2 Molecular markers

Phylogenetic relationships for 477 *Carex* species treated in the FNA (Ball & Reznicek, 2002) were reconstructed using the chloroplast markers *matK* (coding) and *rps16* (non-coding intron) and the nuclear ribosomal non-coding sequences ITS and ETS.

Maturase K (*matK*) is a gene positioned in the intron of the lysine *trnK* gene (Zoschke *et al.,* 2009). The protein it encodes may be involved in the splicing of the *trnK* precursor, and

further tasks are hypothesized, as it is retained in several parasitic plants and ferns that have lost *trnK* (Zoschke *et al.,* 2009). It is a highly variable gene and as a result is an effective angiosperm barcoding region (Yu *et al.,* 2011; and in *Carex*; Starr *et al.,* 2009). The *rps16* marker is an intron positioned inside of the ribosomal protein 16 encoding gene. It is often used in plant barcoding and is flanked by two *rps16* exons (Oxelman *et al.,* 1997).

To avoid relying on uniparentally inherited regions, such as the maternally inherited chloroplast, two nuclear markers were selected for amplification and sequencing. The intergenic transcribed spacer (ITS) region of the 18S–26S nuclear ribosomal cistron is a commonly used marker for plant phylogenetic inference (Alvarez & Wendal, 2003). It is composed of the internal transcribed spacer region 1 (ITS1), the 5.8S gene, and the internal transcribed spacer region 2 (ITS2), all three of which were captured by the ITS primers used. The external transcribed spacer (ETS), found in the same rDNA region, is often more phylogenetically informative than the ITS region, but not widely used due to difficulties in primer design (Logacheva *et al.,* 2010). Fortunately, ETS primers have been designed for use in Cyperaceae (Starr *et al.,* 2003).

All four markers have been shown to be effective tools for reconstructing lower-level phylogenies in *Carex* (Starr *et al.,* 2003; Starr *et al.,* 2009). Primers used can be found in Table 3.1.

3.2.3.3 DNA amplification and sequencing

The *matK* region was amplified as detailed in Chouinard (2010). Most *matK* sequencing took place at the Canadian Centre for DNA Barcoding following the Centre's sequencing

protocol (Ivanova & Grainger, 2009) and the remainder was sequenced at the Canadian Museum of Nature (CMN) using the method described in Chouinard (2010).

Sequencing of the *rps16* region was performed at the CMN using the protocol described by Chouinard (2010; additionally described by Gilmour *et al.,* 2013 and Starr *et al.,* 2015). However, the following modifications were made to quantities of PCR reagents used: 0.3mM of each dNTP, 22.5mM $MgCl_2$, 10–50 ng of template DNA, and 0.3 units of HotStart *Taq* DNA polymerase (BioShop Canada Incorporated). Furthermore, the amplification cycle program was modified: 94°C for 120s for initial denaturing; 35 cycles of 94°C for 60s for denaturing, 46°C for 60s for primer annealing and 72°C for 4min for primer extension, followed by a final extension of 72°C for 7min at the end of cycling. Amplification products were purified using an Exonuclease I and Shrimp Alkaline Phosphatase protocol (MJS Biolynx Inc., Canada) and sequenced using an ABI Prism Big Dye Terminator Kit (Applied Biosystems). Sequence data were edited in Sequencher 5.2 (Gene Codes Corporation, Ann Arbor, MI, USA).

Both ITS and ETS regions were amplified and sequenced at the Morton Arboretum, Lisle Illinois USA according to the molecular methods described in Hipp *et al.* (2006). Sequences were obtained on a 3730 DNA Analyzer (ThermoFisher Scientific).

3.2.3.4 Sequence selection

Multiple specimens per species were sequenced by Chouinard (2010; *matK* sequencing) and by the authors of this study for *rps16*, ITS and ETS. However, due to limits of computing power and time on the final 477 species 4 marker phylogenetic analysis and the requirement for a high quality result, the molecular dataset was limited to one specimen per species (excluding subspecies and varieties).

The most representative individuals for each species were determined via a preliminary Maximum Likelihood (ML) analysis as follows. Separate trees were estimated for each marker (four ML trees in total) using all sequences collected. RAxML v7.3.4 (Randomized Axelerated Maximum Likelihood; Stamatakis, 2006) and its rapid bootstrapping algorithm was employed to generate 200 bootstrap trees per analysis using a GTRCAT model of evolution (General Time Reversible model of nucleotide substitution under the Gamma model of rate heterogeneity; GTRCAT is an approximation of GTRGAMMA that reduces computation time; Stamatakis, 2006). RAxML offers fast inference of large phylogenies implemented in a parallel program with low memory usage (Stamatakis, 2006; Stamatakis *et al.,* 2008).

Subsequently, the ML gene trees were used to guide the selection of individuals, based on the following criteria: (1) individuals must have been clustered with other individuals of the same species or closely related species in all four trees (i.e., no questionable individuals or relationships were accepted), (2) individuals with sequences available for all four markers were preferred, (3) when all else was equal (i.e., individuals were found at the same place in all trees and the same amount of DNA existed), selection was random. The ML trees and individuals chosen for final analysis can be found in Appendix 4 and Appendix 2. Finally, the program RogueNaRok (Aberer *et al.,* 2013) was used to ensure no rogue taxa (i.e. possible misidentifications or contaminations) were included in the final molecular dataset. RogueNaRok uses a graph-based algorithm to assess the improvement (consensus support) of removing subsets of taxa from a set of trees. The gene trees were used in the web-based RogueNaRok platform (rnr.h-its.org).

Not all individuals in the analysis could be sequenced for all four markers. Approximately 15% percent of the data was missing in the analysis.

3.2.3.5 Outgroup species

The species *Trichophorum alpinum* and *T. cespitosum* were selected as outgroup taxa as they represent both major lineages discovered within the sister clade to Cariceae (which includes the genus *Carex*) in a recent phylogeny (Léveillé-Bourret *et al.*, 2014). Voucher data for these outgroups are given in Appendix 2. Outgroups were used to ensure certainty of the reconstructed age and position of the ingroup.

3.2.3.6 Sequence alignment

Sequences were aligned with the multiple sequence alignment program MUSCLE v3.5 (Edgar, 2004). MUSCLE uses an iterative refinement algorithm that performs more accurately and more quickly than the commonly used program CLUSTALW (Pais *et al.,* 2014). Although it is less accurate than the consistency-based programs T-Coffee and MAFFT, it has superior speed and memory usage (Edgar & Batzoglous, 2006; Pais *et al.,* 2014) making it the most suitable program for aligning the 479 sequence dataset used in this analysis. Each set of sequences (four sets; *matK, rps16,* ITS, ETS) were aligned separately using the default parameters in MUSCLE.

The aligned sequence datasets were then concatenated into a single matrix. Although some individuals were missing sequences for one or more regions due to amplification or other difficulties (15% data was missing), this likely did not reduce the utility of the alignment. PartitionFinder v1.1.1 (Lanfear *et al*., 2012) was used to determine optimal models of evolution and the optimal partitioning scheme for the supermatrix. BEAST (Bayesian Evolutionary Analysis Sampling Trees; Drummond and Rambaut, 2007) model options were selected and the

AICc (corrected Akaike Information Criterion; Akaike, 1973) was used to evaluate candidate partitioning schemes. Results were used to set parameters of the phylogenetic analysis.

3.2.3.7 Phylogenetic inference

Model-based phylogenetic methods were preferred due to their improved consistency and reliability, over parsimony and distance-based methods (Felsenstein, 2004). Therefore, Maximum Likelihood (ML) and Bayesian approaches were evaluated for suitability.

Phylogenetic comparative methods (e.g., estimation rates of trait evolution) have been developed to operate on ultrametric phylogenies (Garamszegi, 2014). Rather than creating an ultrametric tree from an additive ML tree using post-analysis branch length transformation, a Bayesian analysis was preferred. Using Bayesian inference, a molecular clock can be specified and an ultrametric tree can be estimated directly. Specifically, BEAST (Bayesian Evolutionary Analysis Sampling Trees) was the best choice due to its ability to process large alignments (Drummond and Rambaut, 2007). BEAST is a program that uses Markov chain Monte Carlo (MCMC) algorithms for Bayesian phylogenetic analysis and divergence time dating (Drummond and Rambaut, 2007).

PartitionFinder results recommended a three subset partitioning scheme (ITS, ETS, *matK+rps16*) under GTR+I+G substitution models. Ultrametric phylogenies were generated using BEAST v1.8.2 under an uncorrelated lognormal relaxed clock with a Yule speciation process tree prior, under the above partitioning scheme. Model parameter and statistic priors were left at default values, except the prior distribution of the ingroup *Carex* which was assigned a normal distribution of mean 35 and standard deviation 5. The dating of this prior was based on two fossil-based date estimations from previous studies (*contra.* phylogeny-based date estimates

presented in the introduction): Cyperaceae was estimated to have diverged at 47 My by Iles *et al.* (2015) and *Carex* was estimated to have been widespread by the early Miocene (ca. 23 My; Smith *et al*., 2010). Two chains were run in BEAST, each with 100 million generations and sampling at every 10 000 states. Chain convergence was visually assessed in Tracer v1.6. The first 40 million states were discarded as burn-in and the two runs were combined in TreeAnnotator v1.8.2.

3.2.4 Trait data

Morphological data were extracted from the recent treatment of *Carex* for the Flora of North America online (FNA; Ball & Reznicek, 2002). Authors of FNA treatments describe species using only mature specimens, and encompass the range of morphological plasticity they see in a species by describing characters using ranges of measurements. Numerical data following key words (e.g., 'culm', 'leaves') were parsed in R using custom-built scripts. This strategy yielded more detailed data than another method tested (beta testing of the Taxon Concept Explorer flora tool at etc.cs.umb.edu/etcsite). The data were collapsed for subspecies and varieties to match the molecular dataset. Occasionally, data were only available as relative measurements (e.g., inflorescences 0.46-0.94(-0.99) of culm height 10-55 cm), in which case they were transformed into absolute measurements (4.6-51.7(-54.45) cm). When character data were missing, measurements were extracted from FNA sectional descriptions or from Mackenzie's Flora (1931-1935). Using these resources, data for five characters (culm height, leaf blade width, perigynia length, perigynia width and achene length; Fig. 1.1) were available for over 90% of species. Despite the presence of inflorescence length and spike descriptions in the FNA online, data were not reported consistently by FNA contributors and therefore could not be

used in this study. The data were verified manually in Excel to ensure measurements were reasonable (i.e., to identify outliers).

Water and light tolerance were used as habitat predictors. Water habitat was taken from the updated National Wetland Plant List (NWPL; Lichvar, 2012) which provides wetness scores from 1 to 5 for 350 FNA *Carex* species. Species missing from the NWPL database, as well as light affinity for all species, was scored *de novo* by professionals familiar with North American taxa (J. Starr, University of Ottawa and A. Hipp, Morton Arboretum). Scoring was completed individually by JS and AH with scoring disagreements resolved by averaging. Light preference was scored from 1 to 3, with 3 as shade tolerant and 1 for open canopy.

Georeferenced specimen locality data for all species in the study were collected via the Global Biodiversity Information Facility (GBIF; www.gbif.org) from participating collections databases. Occurrence data can be removed or uploaded to GBIF by participating institutions at any time. Therefore, to ensure the capture of the highest quantity of specimen data possible, data were downloaded at three time points using the *gbif* function in the *dismo* package for R (from 16 May to 28 May 2013, 13 January 2014 and 3 July 2014). These data were then merged, yielding one to 216,475 occurrences per *Carex* species. GBIF data were mapped in R and maps were visually assessed for outliers (e.g., to identify any points in the ocean or lat long reversals). Duplicates and imprecise records were removed (either latitude or longitude had one decimal place or less). Next, climate data were extracted from the WorldClim climate database (worldclim.org/bioclim; Hijmans *et al.,* 2005) using the *getData* function in *dismo*. Each GBIF locality record produced one set of BIOCLIM variables (19 in total). For each species, these BIOCLIM data were averaged to obtain one mean and one standard deviation measure for each of 19 BIOCLIM variables (19 means and 19 SDs for all species).

Principal component analysis (PCA) was used to obtain two variables (PC1, PC2) for 19 BIOCLIM variables and for 5 morphological variables. PCA allows a broader description of morphology and climatic niche in fewer uncorrelated variables. The function *prcomp* in the R *stats* package was used to generate principal components that were rotated, centered and scaled.

3.2.5 Trait evolution

Phylogenetic comparative methods (PCMs) aim to determine correlates and causes of species trait evolution and diversification. They allow for the acknowledgement of non-independence of species due to shared ancestry (Felsenstein, 1985; Harvey & Pagel, 1991). PCMs have been advocated as promising approaches to understanding evolutionary processes at the macroevolutionary scale (Pigliucci, 2008).

Rates of evolution are estimated in PCMs by fitting models of evolutionary process to trees. There are two models of evolution used widely in the literature; the Brownian motion (BM) process (Felsenstein, 1985) and the Ornstein-Uhlenbeck (OU) process (Hansen, 1997; Butler & King, 2004; Hansen *et al.,* 2008). BM models evolution as a random-walk constant variance process, while OU evolves towards an adaptive optimum. Model selection (AIC and likelihood) using the *fitContinuous* function in the R package *geiger* (Harmon *et al.,* 2008) was used to identify which process (BM or OU) fit the *Carex* data best.

Numerous packages and programs exist for the estimation of evolutionary rates. Most notably, the packages *geiger* and *phytools* (Revell, 2012) fit models (BM, OU or others) for continuous character evolution and return parameter estimates (i.e., rates). The package *ape* (Paradis *et al.,* 2004) also provides tools to simulate continuous character evolution. For this study, the functions *extract.clade (ape* package), *paintSubTree* and *brownie.lite* (*phytools*

package) were used to extract rate estimates. The precise usage of these functions is detailed below. When trait data did not match phylogeny tips (e.g., trait data was missing for a particular species), the tip was dropped for the analysis. Therefore, the number of species in each of the six trait evolution analyses (morphology PC1, morphology PC2, climatic niche PC1, climatic niche PC2, wetness habitat, light habitat) was different.

3.2.6 Sister clade comparison

Sister clade contrasts test the effects of a trait on diversification, while eliminating effects of clade age and relatedness (Barraclough *et al.,* 1998). They are used widely in comparative studies of diversity (Rabosky, 2009). Using species richness as a proxy for diversity, a sister taxa analysis was used to identify if a relationship exists between rates of trait evolution and sedge species richness. Non-nested sister clades (or nodes) were compared, so that each species was included in the analysis only once.

Using a custom script in R, sister clade contrasts were automatically selected for use in the analysis. The script operated on parent nodes (the ancestor of sisters to be compared, identified in red in Fig. 3.1). Parent nodes must have met the following criteria to be selected: (1) each descendant node has three species or more, and (2) the node has not already been selected, to maintain non-nestedness of contrasts. Rates of trait evolution could not be calculated with one descendant species, and with two descendant species the error on the rate calculation was unacceptably high. The script tested parent nodes on the tree randomly, starting with zero parent nodes, and adding successive parent nodes as they met the criteria. Once a full set of contrasts was established (i.e., the tree was exhaustively searched for potential contrasts and no more could meet the criteria), the set was saved and the search was repeated. In total, this search was

conducted 10 000 times, each starting with a different and random parent node. Out of the 10 000 potential sets of sister contrasts, the sets with the highest number of parent nodes were identified (maximization of the number of non-nested contrasts). Ten of these sets, rather than one set, were selected for further analysis to increase the power of subsequent analyses. For a visualization of one set of contrasts, see Figure 3.1.

Within each set, *extract.clade* was used at parent nodes to obtain clades containing only the two sister groups of interest (i.e., clades were extracted at the parent node in red Fig. 3.1). Next, *paintSubTree* mapped two different states on the clade, each corresponding to an arbitrary sister group state (either 1 or 2). The function *brownie.lite* was used to estimate rates of Brownian motion trait evolution in states 1 and 2. Subsequently, the ratio of the two rates of trait evolution was calculated and log transformed. The ratio of species richness (i.e., number of species) per sister group was also calculated and log transformed. A sample of these calculations can be seen in Figure 3.1. A regression analysis of the log transformed rates of trait evolution and the number of species per comparison, was performed for each set of ten contrasts. An r^2 and a slope were computed for each, yielding ten r^2 and ten slopes. The average r^2 and slope were calculated.

To determine significance of trait evolution rate and species richness correlations, the regression analysis was conducted with 100 simulated datasets in place of real trait data (i.e., to develop a null distribution). Continuous characters were simulated on the tree using the package *phytools* (Revell, 2012). Namely, the function *fastBM* was used to simulate random character states, according to a specified rate of trait evolution. The rate used was the mean rate of trait evolution from all characters (morphology PC1, morphology PC2, wetness habitat, light habitat, climatic niche PC1, climatic niche PC2). A relationship between trait evolution rate and species

richness was considered significant if its mean r^2 and its mean slope were greater than 95% of simulated data regression statistics.

3.3 RESULTS

3.3.1 Phylogeny

The final four marker (ETS, ITS *matK, rps16*) molecular dataset consisted of 477 *Carex* species, two outgroup species, 3415 aligned base pairs and contained 2044 unique patterns. The final phylogeny ranged in Bayesian posterior support from 0.9 to 1.

Results from previous phylogenetic analyses of *Carex* species have found that three major lineages exist in North American *Carex* (Starr *et al.,* 2015, Global *Carex* Group, 2015), termed the Core *Carex* Clade, the *Vignea* Clade and the Core Unispicate Clade. While the Core *Carex* Clade and the *Vignea* Clade have received strong support from previous reconstructions, the Core Unispicate Clade has been poorly supported (Starr *et al.,* 2015). This study's results show the Core Unispicate Clade group nested within *Vignea* and Core *Carex* (Fig. 3.2), adding to inconsistencies seen among studies as to the position of the three clades (Global *Carex* Group, 2015). While this is an important phylogenetic result, the placement of major lineages has no effect on the results of trait evolution analyses.

This study compares species numbers of sister groups. Therefore, achieving high quality resolution of lower-level (e.g., species or sectional-level) relationships is necessary. With the exception of a select number of species (*Carex heterostachya, C. fuliginosa, C. edwardsiana, C. squarrosa, C. typhina*), species followed sectional classifications and expected relationships (Fig. 3.2).

3.3.2 Trait data

Principal component 1 (PC1) of the BIOCLIM climate variable PCA explained 42% of total variance and PC2 explained 28% of variance (PC1+PC2=70%). PC3 and PC4 explained more than one variable's worth of information (PC3=16%, PC4=7%; >5.3%). However, they were not included further in trait analyses due to the marked drop in the percentage of explained variation (28% to 16%). Temperatures of cold and dry periods contributed most to PC1, followed by contributions from annual mean temperature, annual precipitation and precipitation of wet periods (Table 3.2). PC2 was loaded most strongly by precipitation of the warmest quarter and precipitation seasonality, as well as precipitation of dry periods and temperature seasonality and annual range. Seventeen of 19 BIOCLIM variables contributed strongly to PC1 or PC2, leaving mean diurnal range and maximum temperature of warmest month, both of which contributed to PC3.

Morphology PC1 explained 49% of variance and PC2 explained 19% of variance. No other PC explained a large portion of variance. Reproductive characters (achene and perigynia descriptors) contributed most strongly to PC1 and vegetative characters (culm and leaf descriptors) contributed strongly to PC2 (Table 3.3). Therefore, morphology PC1 can be interpreted as reproductive and morphology PC2 as vegetative.

3.3.3 Trait evolution

The regression between log transformed species richness ratios and log transformed trait evolution ratios for 100 simulated continuous character datasets was consistently weak (mean r^2 of 100 simulated data sets=0.034, median r^2=0.024). The slopes of the same regressions were weak (mean slope of 100 simulated data sets=0.169, median slope= 0.142), but positive in the

majority of cases. Regressions with real trait evolution data can be compared to these null distributions to establish significance (Fig. 3.3).

Significant correlations between rates of trait evolution and species richness numbers (log transformed) between sister groups were recovered for 5 of 6 traits at the 0.05 level. They were, in decreasing order of correlation strength, morphology PC1, climatic niche PC2, climatic niche PC1, light habitat and wetness habitat (Fig. 3.4). Only one trait, morphology PC2, did not show a correlation more extreme than 95% of simulations. Mean r^2 values ranged from 0.07 to 0.24 and slopes from 0.32 to 2.46.

The distribution of trait diversity through time and throughout the *Carex* species phylogeny can be visualized in ancestral state reconstructions in Figures A3.1 to A3.3, found in the Appendices. While these plots do not measure trait evolution directly, or test a specific question, they show how diversity is distributed throughout *Carex* clades in North America, from a phylogenetic perspective. The plots show that in some clades, traits are well conserved (best seen in habitat, Fig. A3.2), while in equally as many, traits are diverse. In morphology and in climatic niche (Figures A3.1 and A3.3), extreme trait values obscure conservatism or divergence patterns.

3.4 DISCUSSION

An explanation for differences in species richness across the tree of life continues to elude evolutionary biologists. Several hypotheses have been proposed to explain this pattern, coining new terms such as 'key innovations' or 'ecological opportunities'. Recently, it has been proposed that increased capacities to produce novel morphologies or to occupy novel ecological niches, measured via higher rates of morphological and ecological trait evolution, may promote speciation. This study examined the relationship between species richness and rates of character evolution in North American *Carex,* north of Mexico (Fig. 2.1). Within the genus, higher rates of reproductive morphology, habitat and climatic niche evolution lead to greater species numbers when comparing sister groups on a logarithmic scale (Fig. 2.4). This supports the hypothesis that greater trait lability has increased *Carex* species diversity in North America.

Only one trait examined, vegetative morphology, did not show a significant correlation with species richness differences (Fig. 2.4). This is not surprising, as vegetative characters in *Carex* show much less variation than reproductive characters (Metcalfe, 1969 & 1971; Reznicek, 1990). Furthermore, it has been suggested that vegetative traits diverge less rapidly than floral traits in angiosperms (Grant, 1949). This hypothesis has been supported by a study showing that the relationship between molecular change and vegetative morphology is weaker than the relationship between molecular change and reproductive morphology (Barraclough & Savolainen, 2001). In *Carex* of eastern Asia, vegetative morphology is more diverse than in North American species (Starr *et al.,* 2015) yet species richness is reduced.

The strongest significant correlation was found between species richness and reproductive morphology (Fig. 2.4). Rates of reproductive morphology evolution explained 24%

of differences in species numbers (or vice versa). The next strongest correlation was found with climatic niche PC2 (mean r^2=0.16). Subsequent significant correlations had similar mean r^2 values (r^2=0.10-0.11). Unfortunately, the processes leading to significant correlations have not been investigated by this study. Therefore, it is only possible to speculate on the reasons for differences in correlation magnitudes. A possible hypothesis for why reproductive morphology is most strongly correlated with species richness, is that it is better able to establish reproductive isolation, leading to more rapid speciation. Reproductive isolation can be accomplished through mechanical incompatibilities of reproductive morphology, and this is often observed in plant genera (Rieseberg & Willis, 2007). Although the reproductive traits examined here (perigynia and achenes) are not responsible for sedge fertilization, they represent a proxy for such structures. Furthermore, the importance of perigynia morphology in sedge identification (perigynium is the most important diagnostic structure; Starr, pers. comm.) lends support to this hypothesis. Diversity of perigynia morphology is a result of the multitude of dispersal mechanisms, including wind, water and zoochory (attachment to animals, ingestion by animals, ants, birds, and mammals; Egorova, 1999; Handel, 1978). Further research would be required to determine if the incompatibility of reproductive morphologies is aiding to increase speciation in *Carex*.

A rate of trait evolution can be interpreted in multiple ways. The most obvious interpretation is that the rate represents the speed at which traits change on a tree. But it is also a phylogenetically corrected measure of trait disparity (O'Meara *et al.,* 2006), sometimes termed the amount of 'morphospace' or 'niche space' occupied (Hutcheon & Garland, 2004). Hence, finding higher rates of trait evolution associated with higher species numbers implies that in North America, *Carex* species groups with greater morphological and ecological diversity are

more species-rich. These results not only bolster the hypothesis for trait flexibility's effect on speciation, but also provide an explanation for diversity differences within genera inhabiting identical geographic regions (i.e., conditions under which one would not expect to find substantial differences in diversity).

As previously described, *Carex* has been found to be frequently involved in niche partitioning along environmental gradients (Waterway *et al.,* 2009) and support has been found for an important role of habitat heterogeneity in the generation of sedge species richness (Gehrke & Linder, 2011). This study addresses ecological diversity and species richness in *Carex* not as previous studies have before (from a phylogenetic descriptive or statistical point of view), but from a phylogenetic data modelling framework. Using a larger sampling than any previous study, this work finds further support for an integral role of ecological disparity in the generation of sedge species richness.

If one considers the results from an evolutionary perspective, as this study was originally designed, a number of alternative hypotheses may explain the results obtained. Because causality cannot be established from correlation, it is possible that trait evolution does not result in increased species richness. Rather, high rates of speciation, resulting in high species richness, may increase trait disparity through time (Rabosky & Adams, 2012). Additionally, a hypothesized phenomenon known as 'punctuated equilibrium' may explain the results, where rapid speciation leads to rapid morphological change (Eldredge & Gould, 1972; Stanley, 1975). A longstanding challenge in the study of extant biodiversity has been how to statistically establish causality. Unfortunately, most phylogenetic comparative methods do not address causality (Paradis, 2014), and those that do are still under development (e.g., the application of path analysis to phylogenetic analyses; Gonzalez-Voyer & von Hardenberg, 2014). To rule out

alternative hypotheses that may explain the results of this study, application of a method establishing causality is required.

The correlation of high trait disparity through time and species richness has not been consistently observed across taxonomic groups. In a study of globally distributed squirrels, Zelditch *et al.* (2015) found that patterns of diversification, measured by rates of speciation and species richness, were uncorrelated with morphological disparity. However, the authors state that their finding of young, homogenous clades and older, more morphological disparate clades may be in part a result of the disparate environments they inhabit and their phylogenetic community structure. Alternatively, they hypothesize that phenotypic evolution is governed by ecology and speciation processes by geography. They dispute that trait evolution leads to speciation and state that the coincidence of geographic and adaptive landscapes may explain reported positive correlations (Rabosky & Adams, 2012; Rabosky *et al.,* 2013). When these landscapes do not coincide, the authors explain that results such as theirs will be found. The geographic sampling of this study reduces the effect of geography on speciation. Therefore, the results do not reject the proposed ideas of Zelditch *et al.* (2015). Instead, the findings of this study support the idea that trait lability and trait disparity may be important in generating species richness at the regional (continental) scale. It is possible that at larger than regional scales (e.g., at the global scale), higher operating processes such as the influences of geographic landscapes, influence speciation most.

Carex is more diverse than other temperate angiosperms, and this may be due to higher overall rates of trait evolution. But what is the mechanism underpinning rapid trait evolution in *Carex*? One hypothesis is that its holocentric chromosomes allow for accelerated genetic change. Holocentric chromosomes are a unique property of sedges and few others (Hipp *et al.,* 2010).

Centromeres in holocentric chromosomes are diffuse and kinetochore activity is distributed along the whole chromosome. This allows for extraordinary chromosome number variation (n=6-56; Davies, 1956) and frequent chromosomal rearrangements (Luceño & Castroviejo, 1991; Hipp *et al.,* 2010; Escudero *et al*., 2010; Chung *et al*., 2011). Chromosome number variation can affect total recombination rates (Bell, 1982; Nokkala *et al*., 2004; Escudero *et al*., 2012b) and the rate of gene flow among populations of closely related species (Whitkus, 1988; Hipp *et al*., 2010). Furthermore, reproductive isolation provided by genetic differentiation and chromosome number transitions in *Carex* may lead to speciation (Hipp *et al.,* 2010; Escudero *et al.,* 2012a). These may be adaptive properties, leading to fast colonization of new ecological niches and adaptation of morphologies. Further research should attempt to integrate chromosomal data into diversification analyses.

The results of Chapter 2 demonstrated that acquiring highly accurate distribution data may not be necessary for accurate climatic niche estimates and phylogenetic comparative analyses. Therefore, the methods used in this study (Chapter 3) to estimate climatic niche (i.e., the use of GBIF occurrence data), are justified. While the effects of errors in climatic niche estimates on results are understood, it remains unclear how differences in habitat scoring may affect results. Scoring of traits is a method used widely in ecology (Drew & Perera, 2011), but results in personal biases in the data.

Phylogenetic data and trait data did not perfectly overlap. For example, data could not be downloaded from the GBIF server for seven *Carex* species (*C. echinata, C. lasiocarpa, C. pauciflora, C. podocarpa, C. rostrata, C. scirpoidea,* and *C. triquetra*). In situations with incomplete overlap, tips were dropped from the tree. The author of this thesis acknowledges that this is not the most appropriate solution to deal with missing data. Tip dropping may impact

results through its effect on the tree topology and species numbers. Unfortunately, however, a complete sampling was not possible due to time constraints.

In this study, data analysis methods were developed to suit the precise question posed. This is because no appropriate method exists in the literature. However, before developing a custom analysis framework, a number of potential methods were investigated and ruled out. Firstly, the popular State Speciation Extinction (SSE) toolbox – BiSSE, QuaSSE, GeoSSE, and others – was considered (Maddison *et al*., 2007; FitzJohn et al., 2009; FitzJohn, 2010, 2012). This is a set of likelihood methods and R packages that have been developed to test the effects of character states on speciation or extinction rates on phylogenies. However, despite their widespread use (Maddison *et al.,* 2007 cited 387 times in Google Scholar search), these methods have been shown to infer statistically significant associations with neutral traits (Rabosky & Goldberg, 2015). This is because, among other reasons outlined by Rabosky & Goldberg (2015), the model requires only one concurrent shift in character state and rate of diversification, which may be incorrectly placed. Therefore, SSE was determined as inadequate for this study. A newly developed Bayesian statistical framework (BAMM) models dynamics of speciation, extinction and trait evolution on trees (Rabosky *et al.,* 2013; Rabosky *et al.,* 2014). However, the program requires complete taxon sampling or non-random incomplete taxon sampling for analysis. The data sampling in this study is non-random and incomplete, and therefore it was not possible to implement BAMM methods. Lastly, the newly developed R package EvoRAG (Evolutionary Rates Across Gradients; Lawson & Weir, 2014; Weir & Lawson, 2015), which uses likelihood methods to compare evolutionary rates, was investigated. Unfortunately, an analysis using EvoRAG and species richness numbers is not possible due to the high number of replicates required.

The method developed for this study used non-nested sister groups selected from the tree for comparison, according to a set of requirements (see methods). Subsequently, rates of trait evolution were calculated for each group, compared and plotted against corresponding species richness ratios using custom scripts in R. Although, ad hoc methods suffer because they are unverified; simulation studies have not been conducted to determine their power or accuracy. A further shortfall of the method developed is that a criterion used to automatically detect appropriate contrasts – a minimum of three descendants per node while maximizing the number of contrasts – may harbour bias. If the number of contrasts are maximized, the program will by nature choose smaller contrasts, rather than disparate contrasts (e.g., 30 species compared to 3). It is unclear at this point in time how to correct for this bias.

Macroevolutionary-scale species richness studies have previously recognized the existence of a 'taxonomic sampling' bias (Rabosky *et al.,* 2013). This occurs because taxonomists are more likely to split morphologically divergent species into separate lineages, than morphologically similar species. This may not affect the relationship between ecological trait disparity and species richness, but will exaggerate any relationship between morphology and species richness.

Wetness and light were used as habitat predictors in this study. However, adding a third predictor, pH as a proxy for acidity or basicity tolerance, would improve the characterization of habitat niche. *Carex* is frequently found in wetlands (bogs, fens, and others), but also inhabits similarly wet environments with variable pH levels. Soil pH affects the nutrient status of soils. Consequently, different plant physiologies are required to inhabit different acid or basic environments. This important aspect of niche is currently lacking. pH data were not readily available for all *Carex* species of North America, and therefore a partnership with scientists who

have extensive knowledge of the pH niche of *Carex* species would be required to obtain good data. Alternatively, the literature could be mined for *Carex* species pH tolerance data. However, this would be time consuming. Adding pH as a habitat predictor is an important avenue for further research.

The accuracy of this study's results rely on the tree topology (Fig. 2.2). Therefore, calculations of trait evolution rates across randomly selected trees from the BEAST post burn-in distribution should be considered for further work. Additionally, a bootstrap analysis to determine confidence in tree topologies would be beneficial. Unfortunately, due to time constraints, these analyses were not possible.

3.5 TABLES AND FIGURES

Table 3.1 Primer sequences used in DNA amplification and sequencing.

Type	Region	Forward primer		Reverse primer		Reference
cpDNA	*matK*	matK-2.1F:	5' CCT ATC CAT CTG GAA ATC TTA G 3'	matK-5R:	5' GTT CTA GCA CAA GAA AGT CG 3'	Starr *et al.*, 2009
cpDNA	*rps16*	rps16F:	5’ AAA CGA TGT GGT AGA AAG CAA C 3’	rps16R:	5’ AAC ATC AAT TGC AAC GAT TCG ATA 3’	Shaw *et al*., 2005 but modified to remove ambiguities
Nuclear	ITS	ITS-1:	5’ GTC CAC TGA ACC TTA TCA TTT AG 3’	ITS4:	5’ TCC TCC GCT TAT TGA TAT GC 3’	Urbatsch *et al*., 2000 and White *et al*., 1990
Nuclear	ETS	ETS-1F:	5’ CTG TGG CGT CGC ATG AGT TG 3’	18S-R:	5’ AGA CAA GCA TAT GAC TAC TGG CAG G 3’	Starr *et al*., 2003

Table 3.2 Loadings from principal component analysis of 19 WorldClim BIOCLIM variables estimated from GBIF occurrence data. Only those principal components explaining a larger than expected amount of variation are included (>5.3%: PC1=42%, PC2=28%, PC3=16%, PC4=7%). Loadings are bolded if they are important contributors (>square root of 1/19 variables=0.229415). BIOCLIM variables are sorted according to their contributions to PC1, PC2, PC3 and PC4.

BIOCLIM Variable	PC1	PC2	PC3	PC4
BIO11 = Mean Temperature of Coldest Quarter	**0.333849**	0.01324	-0.10339	**0.229201**
BIO6 = Min Temperature of Coldest Month	**0.329456**	-0.0267	-0.02686	**0.289255**
BIO9 = Mean Temperature of Driest Quarter	**0.309547**	-0.12133	-0.0777	0.048826
BIO1 = Annual Mean Temperature	**0.305478**	0.160053	-0.16681	0.121889
BIO12 = Annual Precipitation	**0.281465**	0.06568	**0.281621**	**-0.26876**
BIO13 = Precipitation of Wettest Month	**0.272504**	-0.16703	0.13843	**-0.29621**
BIO16 = Precipitation of Wettest Quarter	**0.268684**	-0.16567	0.15461	**-0.31028**
BIO3 = Isothermality (BIO2/BIO7) (* 100)	**0.244413**	-0.20902	**-0.25007**	0.008242
BIO19 = Precipitation of Coldest Quarter	**0.236629**	-0.21553	0.198342	**-0.3242**
BIO10 = Mean Temperature of Warmest Quarter	**0.230339**	**0.27629**	-0.21324	-0.005
BIO18 = Precipitation of Warmest Quarter	0.113953	**0.353102**	0.139396	0.060068
BIO15 = Precipitation Seasonality (Coefficient of Variation)	0.008027	**-0.3481**	**-0.26254**	-0.07362
BIO8 = Mean Temperature of Wettest Quarter	0.090206	**0.325107**	-0.20311	0.19157
BIO14 = Precipitation of Driest Month	0.126601	**0.318984**	**0.297723**	-0.04934
BIO17 = Precipitation of Driest Quarter	0.148461	**0.307437**	**0.294806**	-0.0761
BIO4 = Temperature Seasonality (standard deviation *100)	-0.22172	**0.273964**	-0.08859	**-0.33225**
BIO7 = Temperature Annual Range (BIO5-BIO6)	-0.16802	**0.240523**	**-0.25854**	**-0.44086**
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))	0.105125	0.01111	**-0.45553**	**-0.33338**
BIO5 = Max Temperature of Warmest Month	0.220704	0.227562	**-0.31247**	-0.12354

Table 3.3 Loadings from principal component analysis of six morphological characters (perigynia length, perigynia width, achene length, achene width, culm height, leaf width). Important contributors to principal components explaining a larger than expected amount of variation are bolded (PC1=49%, PC2=19%).

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Achene width	**-0.474**	-0.36147	0.106515	0.385384	-0.30025	-0.62821
Perigynia length	**-0.46338**	0.015398	-0.37295	-0.49016	0.56536	-0.29337
Perigynia width	**-0.45533**	0.043531	-0.49912	-0.11885	-0.57944	0.437922
Achene length	**-0.44666**	-0.3602	0.361374	0.247659	0.392534	0.569865
Culm height	-0.23958	**0.76264**	-0.09427	0.560185	0.194292	-0.02324
Leaf width	-0.31044	**0.394773**	0.678943	-0.47113	-0.25026	-0.0472

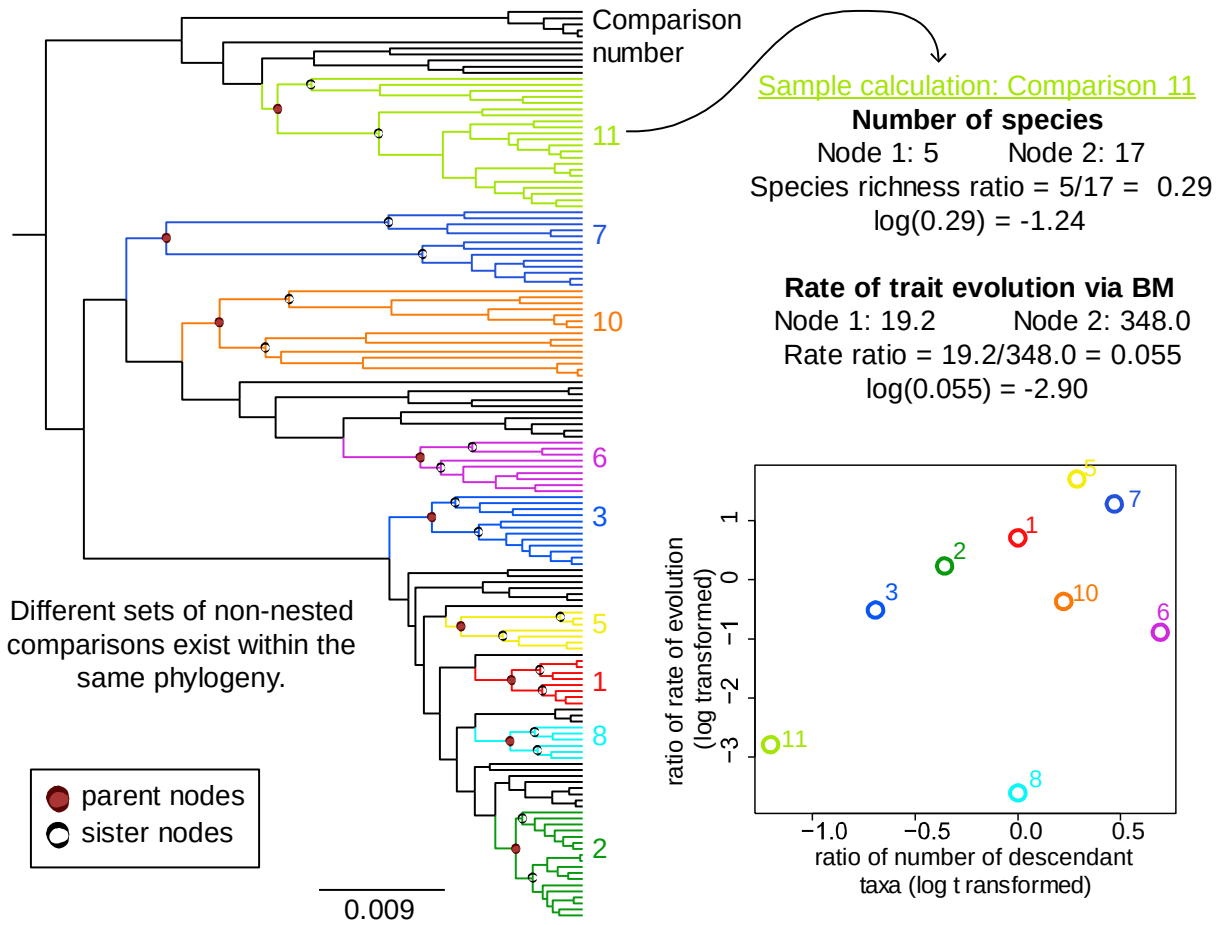
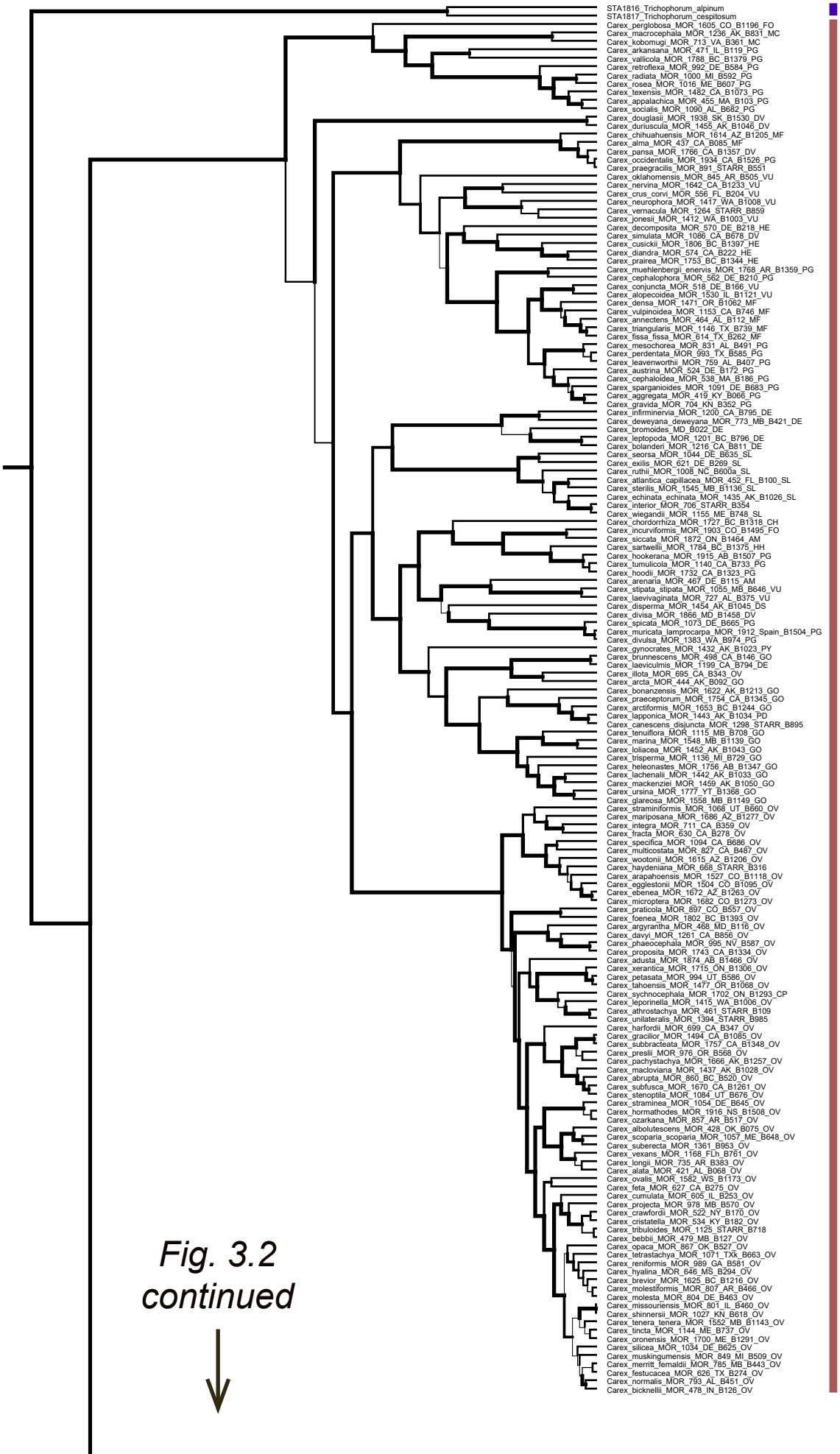


Fig. 3.1 An example of one set of non-nested sister group comparisons. Each sister contrast is shown in a different colour. Parental nodes are circled in red and sister nodes are circled in black. For each sister comparison, the ratio of species richness and ratio of trait evolution is calculated. The data then are log transformed and plotted. This procedure is performed for ten different sets of sister groups.



Trichophorum clade (outgroup)

Vignea clade

Fig. 3.2 continued

Fig. 3.2 continued

Core Unispicate Clade

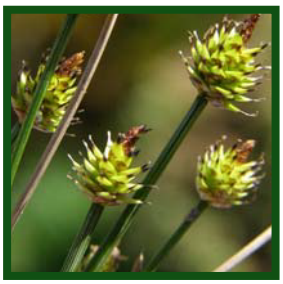


Fig. 3.2 continued

Fig. 3.2 continued

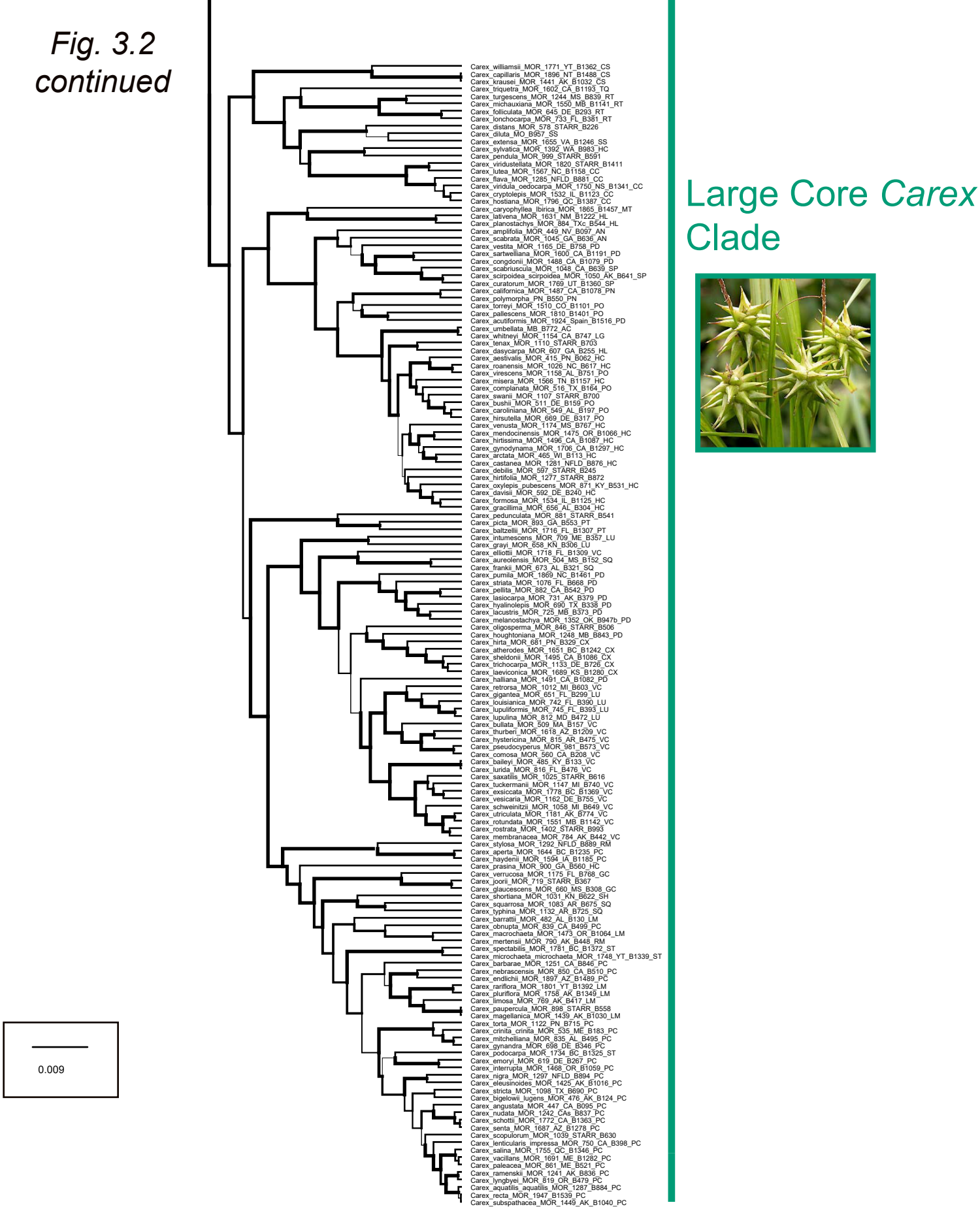


Fig. 3.2 Maximum clade credibility phylogeny summarizing two BEAST runs of 100 million generations each. Two ougroup species and 477 North American *Carex* species are included. Thickness of branches correspond to Bayesian posterior probabilities. The phylogeny was constructed from the four-gene supermatrix (*matK, rps16*, ITS, ETS). Major clades (from Starr *et al.,* 2015) are indicated and are accompanied by a photo of a member species. Time scale is in millions of years. Labels contain sample information (Morton Arboretum identifier, MOR code and Starr Lab identifier, B code), two-letter geography information and section information.

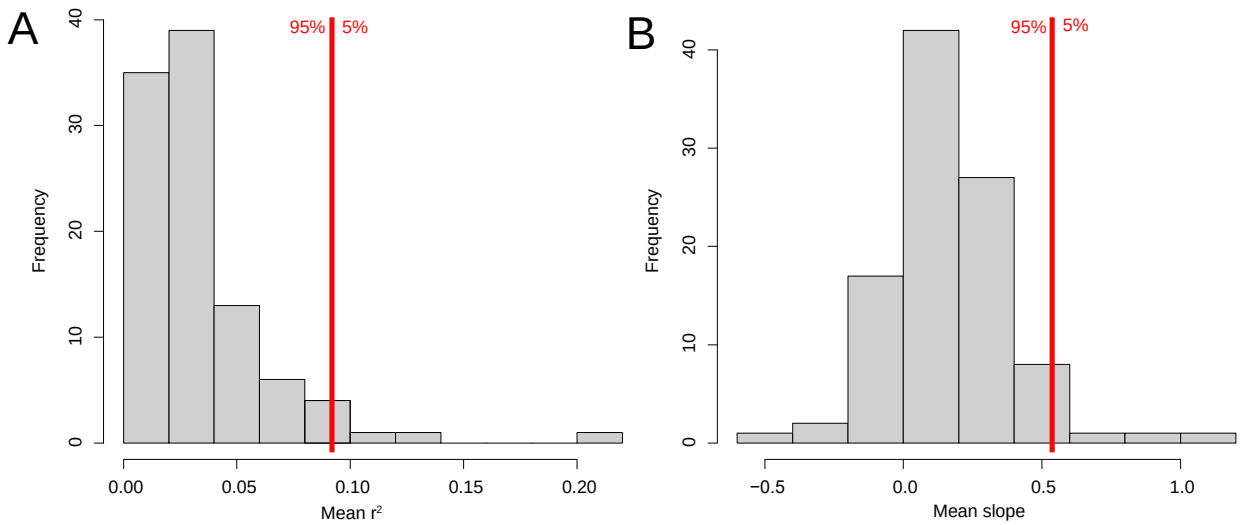


Fig. 3.3 Histogram of (A) mean r^2 values and (B) slopes from regressions of species richness against rates of trait evolution (log transformed) using 100 continuous character evolution simulations. See methods for details on the procedure used to obtain regressions. The red line represents the 0.05 significance level.

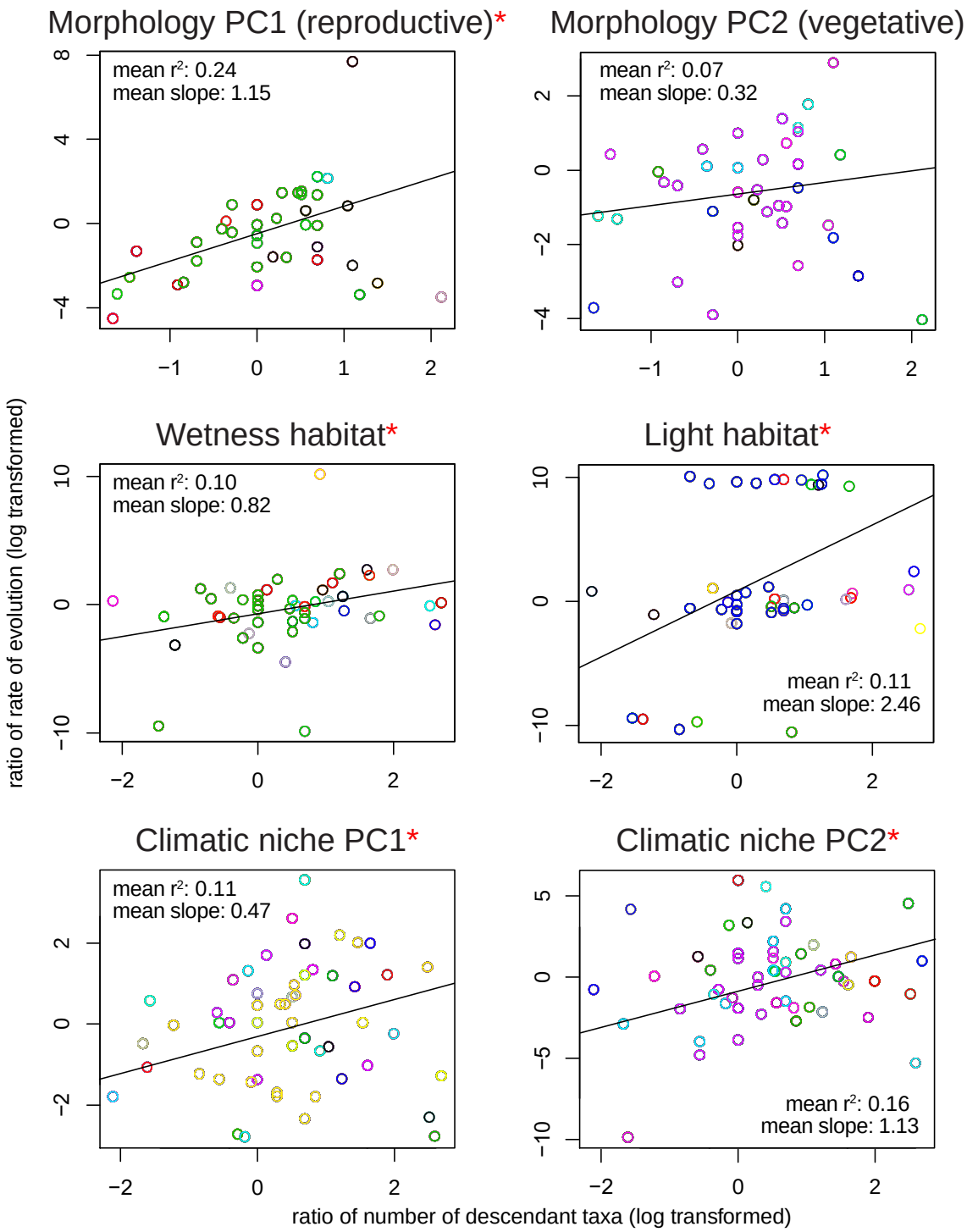


Fig. 3.4 Non-nested sister clade analysis. The ratio of species richness (i.e., the number of descendant taxa) and the ratio of rates of evolution per sister clade contrast are plotted for contrasts satisfying the criterion (= minimum of three descendants per clade). Sets of non-nested sister clades were obtained by maximizing the number of contrasts. Different colours represent different sets of non-nested sister clades, each with the maximum number of contrasts. Mean r^2 and slope values are averaged regression statistics among sets. Asterisks represent significance according to the null distribution of 100 continuous character simulations on the tree.

Chapter 4: Conclusions

This thesis used the exceptionally diverse flowering plant genus *Carex* (Cyperaceae) to test the sensitivity of climatic niche estimates to distribution data sampling (Chapter 2) and to evaluate the relationship between species richness and rates of trait evolution in a geographically comprehensive sample (Chapter 3). It is among the first set of studies to realize the potential of *Carex* as a model evolutionary radiation for ecological and evolutionary research.

In Chapter 2, results showed that sampling effort of GBIF data did not have a strong effect on climatic niche estimates. However, with few counties from species' BONAP distributions covered by GBIF occurrence data, differences in niche space were variable. BONAP data for high elevation species were found to be biased, with elevation explaining 13% of niche differences. Inferred ancestral states for BONAP data were consistently lower than for GBIF (i.e., means were significantly different), although variability between sets was less than within sets.

Consequently, this study recommends supplementing GBIF data with BONAP centroids when coverage of the known distribution is low, to ensure accuracy of niche estimates. However, users should only use BONAP data to supplement species' distribution data when their elevation niche is below 1000m. This can be calculated using the mean from the elevation range in floral descriptions from the Flora of North America. When the scale of the study is large, researchers should consider using only BONAP data, to save time and avoid erroneous inferences (i.e., no data cleaning required). Researchers should further note that different niche estimates will yield different results in phylogenetic analyses, although conclusions that might be drawn are largely

similar. This robustness analysis is informative for researchers seeking to understand the limits of GBIF and BONAP distribution data and their use in climatic niche modeling on a phylogeny.

By comparing non-nested sister groups on a *Carex* species phylogeny, Chapter 3 demonstrated that rates of reproductive morphology, climatic and habitat niche evolution were significantly related to North American *Carex* species richness, in decreasing order of significance. Vegetative morphology (morphology PC2) was not significantly related to species richness. These results support the posed hypothesis that greater trait lability leads to greater diversity in North American *Carex.* Interpreted otherwise, the results suggest that within North American *Carex*, groups that occupy higher amounts of morphospace and niche space through evolutionary time are more successful. This study's findings add to a body of evidence suggesting niche disparity is important in the generation of *Carex* species richness. By restricting the study to North America, the results are exempt from the effects of niche filling or other broad-scale geographic processes (e.g., limits to dispersal, environmental heterogeneity, etc.). However, when other recent works are considered (Zelditch *et al.,* 2015), it is likely that these processes do not extend past the regional scale.

The first study (Chapter 2) is highly relevant to how distribution data can be used to address ecological and evolutionary questions. It enables researchers to make informed decisions on data and methods employed. The second study (Chapter 3) contributes to knowledge of evolutionary patterns in temperate zones. It establishes a relationship between morphological and ecological trait evolution and species richness in a flowering plant radiation.

References

Aberer, A. J., Krompaß, D., & Stamatakis, A. (2011). RogueNaRok: an efficient and exact algorithm for rogue taxon identification. In *Heidelberg Institute for Theoretical Studies: Exelixis-RRDR-2011–10*.

Adamowicz, S. J., Purvis, A., & Wills, M. A. (2008). Increasing morphological complexity in multiple parallel lineages of the Crustacea. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 4786–4791.

Adams, D. C., Berns, C. M., Kozak, K. H., & Wiens, J. J. (2009). Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society B: Biological Sciences*, *276*, 2729–2738.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*, 255 –265.

Alexander, P. J., Rajanikanth, G., Bacon, C. D., & Bailey, C. D. (2007). Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes*, *7*, 5–9.

Álvarez, I., & Wendel, J. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, *29*, 417–434.

Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, *93*, 1527–1539.

Ball, P. W., & Reznicek, A. A. (2002). *Carex Linnaeus*. *Flora of North America*. New York, New York, USA: Oxford University Press.

Barraclough, T. G., Nee, S., & Harvey, P. (1998). Sister-group analysis in identifying correlates of diversification. *Evolutionary Ecology*, *12*, 751–754.

Barraclough, T. G., & Savolainen, V. (2001). Evolutionary Rates and Species Diversity in Flowering Plants. *Group*, *55*, 677–683.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15.

Bell, G. (1982). *The masterpiece of nature: the evolution and genetics of sexuality*. Berkley, California, USA: University of California Press.

Bell, G., Lechowicz, M. J., & Waterway, M. J. (2000). Environmental heterogeneity and species diversity of forest sedges. *Journal of Ecology*, *88*, 67–87.

Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O., & Linder, H. P. (2015). On the complexity of triggering evolutionary radiations. *New Phytologist*, *207*, 313–326.

Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, *164*, 683–695.

Chouinard, B. N. (2010). *DNA Barcodes for the Cariceae (Carex & Kobresia, Cyperaceae) of North America, North of Mexico*. University of Ottawa, Ottawa, Ontario.

Chung, K.-S., Weber, J. a, & Hipp, A. L. (2011). Dynamics of chromosome number and genome size variation in a cytogenetically variable sedge (*Carex scoparia* var. *scoparia*, Cyperaceae). *American Journal of Botany*, *98*, 122–9.

Cunningham, C. W., Omland, K. E., & Oakley, T. H. (1998). Reconstructing character states: a critical reappraisal. *Trends in Ecology & Evolution*, *13*, 361–366.

Dabros, A., Waterway, M. J. (2008). Segregation of sedge species (Cyperaceae) along environmental gradients in fens of the Schefferville region, northern Quebec. In *Sedges uses, diversity and systematics of the Cyperaceae.* Eds. Naczi, R. F. & Ford, B. A. St. Louis, Missouri, USA: Missouri Botanical Garden Press, 145–161.

Davies, E. W. (1956). Cytology, evolution and origin of the aneuploid series in the genus *Carex*. *Hereditas*, *42,* 349-365.

Dennis, R., & Thomas, C. (2000). Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, *4,* 73-77.

Drew, C., & Perera, A. (2011). Expert knowledge as a basis for landscape ecological predictive models. In *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*. Eds Drew, C. A., Wiersma, Y. & Huettmann, F. Heidelberg, New York, Dordrecht, London: Springer Publishing Company, 229–248.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–7.

Edgar, R. C., & Batzoglouu, S. (2006). Multiple sequence alignment. *Current Opinion in Structural Biology*, *16*, 368–373.

Egorova, T. (1999). The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). St. Petersberg., Russia: St. Petersberg. State Chemical Pharmaceutical Academy.

Escudero, M., & Hipp, A. L. (2013). Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). *American Journal of Botany*, *100*, 2403–11.

Escudero, M., Hipp, A. L., & Luceño, M. (2010). Karyotype stability and predictors of chromosome number variation in sedges: a study in *Carex* section Spirostachyae (Cyperaceae). *Molecular Phylogenetics and Evolution*, *57*, 353–63.

Escudero, M., Hipp, A. L., Waterway, M. J., & Valente, L. M. (2012a). Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (*Carex*, Cyperaceae). *Molecular Phylogenetics and Evolution*, *63*, 650–655.

Escudero, M., Hipp, A. L., Hansen, T. F., Voje, K. L., & Luceño, M. (2012b). Selection and inertia in the evolution of holocentric chromosomes in sedges (*Carex*, Cyperaceae). *New Phytologist*, *195*, 237–247.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*, 783–791.

Ferguson, G. & Woodbury, A. (2007). Urban heat island in the subsurface. *Geophysical Research Letters, 34,* L23713.

Fitzjohn, R. G. (2010). Quantitative traits and diversification. *Systematic Biology*, *59*, 619–633.

Fitzjohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*, 1084–1092.

Fitzjohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, *58*, 595–611.

Fritz, S. A., Schnitzler, J., Eronen, J. T., Hof, C., Böhning-Gaese, K., & Graham, C. H. (2013). Diversity in time and space: wanted dead and alive. *Trends in Ecology and Evolution*, *28*, 509–516.

Frodin, D. (2004). History and concepts of big plant genera. *Taxon*, *53*, 881–882.

Gauch Jr., H., Whittaker, R., & Wentworth, T. (1977). A comparative study of reciprocal averaging and other ordination techniques. *The Journal of Ecology*, *65*, 157–174.

Gehrke, B., & Linder, H. P. (2011). Time, space and ecology: why some clades have more species than others. *Journal of Biogeography*, *38*, 1948–1962.

Gilmour, C. N., Starr, J. R., & Naczi, R. F. (2013). *Calliscirpus*, a new genus for two narrow endemics of the California Floristic Province, *C. criniger* and *C. brachythrix* sp. nov.(Cyperaceae). *Kew Bulletin*, *68*, 85–105.

Global *Carex* Group. (2015). Making *Carex* monophyletic: a new broader circumscription. *Botanical Journal of the Linnean Society*, *179*, 1–42.

Glor, R. E. (2010). Phylogenetic insights on adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 251–270.

Gonzalez-Voyer, A., & von Hardenberg, A. (2014). An introduction to phylogenetic path analysis. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology Concepts and Practice.* Ed. Garamszegi, Z. Heidelberg, New York, Dordrecht, London: Springer Publishing Company, 201–230.

Govaerts, R., Koopman, J. & Simpson, D. (2014). World Checklist of Cyperaceae. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. Retrieved September 2015.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, *19*, 497–503.

Grant, V. (1949). Pollination systems as isolating mechanisms in angiosperms. *Evolution*, *3*, 82–97.

Grinnell, J. (1917). The niche-relationships of the California Thrasher. *The Auk*, *34,* 427–433.

Handel, S. (1978). The competitive relationship of three woodland sedges and its bearing on the evolution of ant-dispersal of *Carex pedunculata*. *Evolution, 32,* 156–163.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation, *51*, 1341–1351.

Hansen, T. F., Pienaar, J., & Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, *62*, 1965–77.

Harmon, L. J., & Harrison, S. (2015). Species diversity is dynamic and unbounded at local and –continental scales. *The American Naturalist*, *185,* 584–593.

Harmon, L., Weir, J., Brock, C., Glor, R., & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*, *24*, 129–131.

Harvey, P., & Pagel, M. (1991). *The comparative method in evolutionary biology*. Oxford, UK: Oxford University Press.

Heard, S. B. (1992). Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution, 46,* 1818–1826.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hijmans, R. J., Garrett, K., & Huaman, Z. (2000). Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology*, *14*, 1755–1765.

Hill, M., & Gauch Jr., H. (1980). Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, 47–58.

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

Hipp, A. L., Rothrock, P. E., Whitkus, R., & Weber, J. A. (2010). Chromosomes tell half of the story: the correlation between karyotype rearrangements and genetic diversity in sedges, a group with holocentric chromosomes. *Molecular Ecology*, *19*, 3124–38.

Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, *157*, 331–341.

Hodges, S., & Arnold, M. (1995). Spurring plant diversification: are floral nectar spurs a key innovation? *Proceedings of the Royal Society B: Biological Sciences*, *262*, 343–348.

Hortal, J., Jiménez-Valverde, A., Gómez, J. F., Lobo, M. J., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, *117*, 847–858.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, New Jersey, USA: Princeton University Press.

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10334–10339.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in Ecology and Evolution*, *13*, 31–36.

Hutchinson, G. (1957). *A treatise on limnology: geography, physics, and chemistry. Geography and physics of lakes*. New Jersey, USA: John Wiley & Sons, Inc.

Iles, W. J., Smith, S. Y., Gandolfo, M., & Graham, S. (2015). Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society*, *178*, 346–374.

Isaac, N. J. B., Jones, K. E., Gittleman, J. L., & Purvis, A. (2005). Correlates of species richness in mammals: body size, life history, and ecology. *The American Naturalist*, *165*, 600–607.

Ivanova, N. V, Borisenko, A. V, & Hebert, P. D. (2009). Express barcodes: racing from specimen to identification. *Molecular Ecology Resources*, *9*, 35–41.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*, 444–448.

Jongman, R., Braak, C. T., & Van Tongeren, O. (1995). *Data analysis in community and landscape ecology*. Cambridge, UK: Cambridge University Press.

Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, *14*, 401–413.

Kartesz, J.T. (2014). The Biota of North America Program (BONAP). Taxonomic Data Center. URL at http://www.bonap.net/tdc. Chapel Hill, N.C.

Kozak, K. H., & Wiens, J. J. (2007). Climatic zonation drives latitudinal variation in speciation mechanisms. *Proceedings of the Royal Society B: Biological Sciences*, *274*, 2995–3003.

Kozak, K. H., & Wiens, J. J. (2010). Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters*, *13*, 1378–1389.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701.

Lawson, A. M., & Weir, J. T. (2014). Latitudinal gradients in climatic-niche evolution accelerate trait evolution at high latitudes. *Ecology Letters, 17*, 1427–1436.

Léveillé-Bourret, É., Gilmour, C. N., Starr, J. R., Naczi, R. F., Spalink, D., & Sytsma, K. J. (2014). Searching for the sister to sedges (*Carex*): resolving relationships in the Cariceae-Dulichieae-Scirpeae clade (Cyperaceae). *Botanical Journal of the Linnean Society*, *176*, 1–21.

Lichvar, R.W. 2012. *The National Wetland Plant List*. Hanover, New Hampshire, USA.

Logacheva, M. D., Valiejo-Roman, C. M., Degtjareva, G. V, Stratton, J. M., Downie, S. R., Samigullin, T. H., & Pimenov, M. G. (2010). A comparison of nrDNA ITS and ETS loci for phylogenetic inference in the Umbelliferae: an example from tribe Tordylieae. *Molecular Phylogenetics and Evolution*, *57*, 471–476.

Losos, J. B., & Mahler, D. L. (2010). Adaptive radiation: the interaction of ecological opportunity, adaptation, and speciation. In *Evolution Since Darwin: The First 150 Years*. Eds. Bell, M. A., Futuyma, D. I., Fanes, W. F., Levinton, J. S. Sunderland, Massachusetts, USA: Sinauer Associates, Inc., 381–420.

Lovette, I. J., Bermingham, E., & Ricklefs, R. E. (2002). Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 37–42.

Luceño, M., & Castroviejo, S. (1991). Agmatoploidy in *Carex laevigata* (Cyperaceae). Fusion and fission of chromosomes as the mechanism of cytogenetic evolution in Iberian populations. *Plant Systematics and Evolution*, *177*, 149–159.

Mackenzie, K. K. Poales, Cyperaceae, Cypereae (pars). In *North American Flora*. New York, New York, USA: The New York Botanical Garden.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, *56*, 701–710.

Maguilla, E., Escudero, M., Waterway, M. J., Hipp, A. L., & Luceño, M. (2015). Phylogeny, systematics, and trait evolution of *Carex* section *Glareosae*. *American Journal of Botany*, *102*, 1128–1144.

McPeek, M. A., & Brown, J. M. (2007). Clade age and not diversification rate explains species richness among animal taxa. *The American Naturalist*, *169*, E97–E106.

Metcalfe, C. (1971). *Anatomy of the monocotyledons V: Cyperaceae*. Oxford, UK: Clarendon Press.

Metcalfe, C. (2011). Anatomy as an aid to classifying the Cyperaceae. *American Journal of Botany*, *56*, 782–790.

Meyer, A., Kocher, T., Basasibwaki, P. & Wilson, A. (1990). Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature, 347,* 550–553.

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, *69*, 89–107.

Naczi, R. F., & Ford, B. A. (2008). *Sedges: uses, diversity and systematics of the Cyperaceae.* St. Louis, Missouri, USA: Missouri Botanical Garden Press.

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, *34*, 3–22.

Niles, E., & Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In *Models in Paleobiology.* Ed. Schopf, T. J. M. New York, New York, USA: Doubleday, 82–115.

Nokkala, S. (2004). Holocentric chromosomes in meiosis. I. Restriction of the number of chiasmata in bivalents. *Chromosome Research*, *12*, 733–739.

O'Meara, B., Ané, C., Sanderson, M., & Wainwright, P. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution, 60,* 922–933.

Oakley, T., & Cunningham, C. (2000). Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution*, *54*, 397–405.

Oxelman, B., Lidén, M., & Berglund, D. (1997). Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution*, *206*, 393-410.

Ozinga, W. A., Hennekens, S. M., Schaminée, J. H. J., Bekker, R. M., Prinzing, A., Bonn, S., van Groenendael, J. M. (2005). Assessing the relative importance of dispersal in plant communities using an ecoinformatics approach. *Folia Geobotanica*, *40*, 53–67.

Pais, F. S.-M., Ruy, P. D. C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology*, *9*, 4.

Paradis, E. (2014). An introduction to the phylogenetic comparative method. *In Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology Concepts and Practice*. Ed. Garamszegi, Z. Heidelberg, New York, Dordrecht, London: Springer Publishing Company, 3–18.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.

Peterson, A. T. (2011). Ecological niche conservatism: A time-structured review of evidence. *Journal of Biogeography*, *38*, 817–827.

Peterson, A. T., Knapp, S., Guralnick, R., Soberón, J., & Holder, M. T. (2010). The big questions for biodiversity informatics. *Systematics and Biodiversity*, *8*, 159–168.

Peterson, A. T., Soberón, J. M., & Sanchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science*, *285*, 1265–1267.

Phillips, S. J., Anderson, R. P. & Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190***,** 231–259.

Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews: Genetics*, *9*, 75–82.

Pyron, R. A., Costa, G. C., Patten, M. A., & Burbrink, F. T. (2014). Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biological Reviews*. 90, 1248–1262.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rabosky, D. L. (2009a). Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, *12*, 735–743.

Rabosky, D. L. (2009b). Ecological limits on clade diversification in higher taxa. *The American Naturalist*, *173*(5), 662–674.

Rabosky, D. L., & Adams, D. C. (2012). Rates of morphological evolution are correlated with species richness in salamanders. *Evolution*, *66*, 1807–1818.

Rabosky, D. L., Donnellan, S. C., Grundler, M., & Lovette, I. J. (2014). Analysis and visualization of complex macroevolutionary dynamics: An example from Australian Scincid lizards. *Systematic Biology*, *63*, 610–627.

Rabosky, D. L., & Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology, 64,* 340–355.

Rabosky, D. L., & Hurlbert, A. H. (2015). Species richness at continental scales is dominated by ecological limits. *The American Naturalist*, *185*, 572–583.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, *4*, 1958.

Rabosky, D. L., Slater, G. J., & Alfaro, M. E. (2012). Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biology*, *10*, e1001381.

Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, *30*, 1719–1727.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223.

Revell, L. J. (2013). Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution*, *4*, 754–759.

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

Ricklefs, R. E. (2003). Global diversification rates of passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 2285–2291.

Ricklefs, R. E. (2006). Global Variation in the Diversification Rate of Passerine Birds. *Ecology*, *87*, 2468–2478.

Ricklefs, R. E. (2008). Disintegration of the ecological community. *The American Naturalist*, *172*, 741–750.

Ricklefs, R. E., Losos, J. B., & Townsend, T. M. (2007). Evolutionary diversification of clades of squamate reptiles. *Journal of Evolutionary Biology*, *20*, 1751–1762.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science*, *317*, 910–914.

Roalson, E. H. (2008). A synopsis of chromosome number variation in the Cyperaceae. *Botanical Review*, *74*, 209–393.

Sanderson, M. J., & Donoghue, M. J. (1994). Shifts in diversification rate with the origin of angiosperms. *Science*, *264*, 1590–1593.

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. Oxford, UK: Oxford University Press.

Seddon, N., Merrill, R. M., & Tobias, J. A. (2008). Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. *The American Naturalist*, *171*, 620–631.

Simpson, G. (1955). *Major features of evolution*. New York, New York, USA: Columbia University Press.

Smith, S. Y. (2010). The Cretaceous and Paleogene fossil record of Poales: Review and current Research. In *Diversity, Phylogeny, and Evolution in the Monocotyledons*. Ed. Seberg, O. Aarhus, Denmark: Aarhus University Press, 333–356.

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distribution of species. *Ecology Letters*, *10*, 1115–1123.

Soberón, J. M., Arriaga, L., & Lara, L. (2002). Issues of quality control in large, mixed-origin entomological databases. *Towards a global biological information infrastructure, 70,* 15-22.

Soberón, J. M., Llorente, J., & Oñate, L. (2000). The use of specimen-label databases for conservation purposes: an example using Mexican papilionid and pierid butterflies. *Biodiversity and Conservation Conservation*, *9*(10), 1441–1466.

Spalink, D., Drew, B. T., Pace, M. C., Zaborsky, J. G., Starr, J. R., Cameron, K. M., & Sytsma, K. J. Historical biogeography and diversification of the cosmopolitan Cyperaceae. *Molecular Phylogenetics and Evolution, in press*.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–90.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, *57*, 758–771.

Stanley, S. (1975). A theory of evolution above the species level. *Proceedings of the National Academy of Sciences*, *72*, 646–650.

Stanley, S. (1979). Macroevolution, pattern and process. Oxford, UK: Taylor & Francis, Ltd.

Starr, J. R., & Ford, B. A. (2009). Phylogeny and evolution in cariceae (Cyperaceae): Current knowledge and Future Directions. *The Botanical Review*, *75*, 110–137.

Starr, J. R., Janzen, F. H., & Ford, B. A. (2015). Three new, early diverging *Carex* (Cariceae, Cyperaceae) lineages from East and Southeast Asia with important evolutionary and biogeographic implications. *Molecular Phylogenetics and Evolution*, *88*, 105–120.

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

Todd Streelman, J., & Danley, P. D. (2003). The stages of vertebrate evolutionary radiation. *Trends in Ecology and Evolution*, *18*, 126–131.

United States Census 2010. County by State. United States Census Bureau. Retrieved September 2015.

Valente, L., Savolainen, V., & Vargas, P. (2010). Unparalleled rates of species diversification in Europe. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1489–1496.

Velland, M., Lechowicz, M. J., & Waterway, M. J. (2000). Germination and establishment of forest sedges (*Carex*, Cyperaceae): tests for home-site advantage and effects of leaf litter. *American Journal of Botany*, *87*, 1517–1525.

Venditti, C., Meade, A., & Pagel, M. (2011). Multiple routes to mammalian diversity. *Nature*, *479*, 393–396.

Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8267–8272.

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*, 2868–2883.

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

Weir, J. T., & Lawson, A. (2015). Evolutionary rates across gradients. *Methods in Ecology and Evolution*, *in press.*

Whitkus, R. (1988). Experimental hybridizations among chromosome races of *Carex pachystachya* and the related species *C. macloviana* and *C. preslii* (Cyperaceae). *Systematic Botany*, *13*, 146–153.

Wiens, J. J. (2011). The causes of species richness patterns across space, time, and clades and the role of "ecological limits". *The Quarterly Review of Biology*, *86*, 75–96.

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., Culham, A. (2007). How global is the Global Biodiversity Information Facility? *PLoS ONE*, *2*, e1124.

Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., Roches, S. D. E. S. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, *23*, 1581–96.

Yu, J., Xue, J. H., & Zhou, S. L. (2011). New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution*, *49*, 176–181.

Zelditch, M. L., Li, J., Tran, L. a. P., & Swiderski, D. L. (2015). Relationships of diversity, disparity, and their evolutionary rates in squirrels (Sciuridae). *Evolution*, *69*, 1284–1300.

Zoschke, R., Nakamura, M., Liere, K., Sugiura, M., Börner, T., & Schmitz-Linneweber, C. (2010). An organellar maturase associates with multiple group II introns. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 3245–50.

Appendix 1. Supplementary Figures for Chapter 2

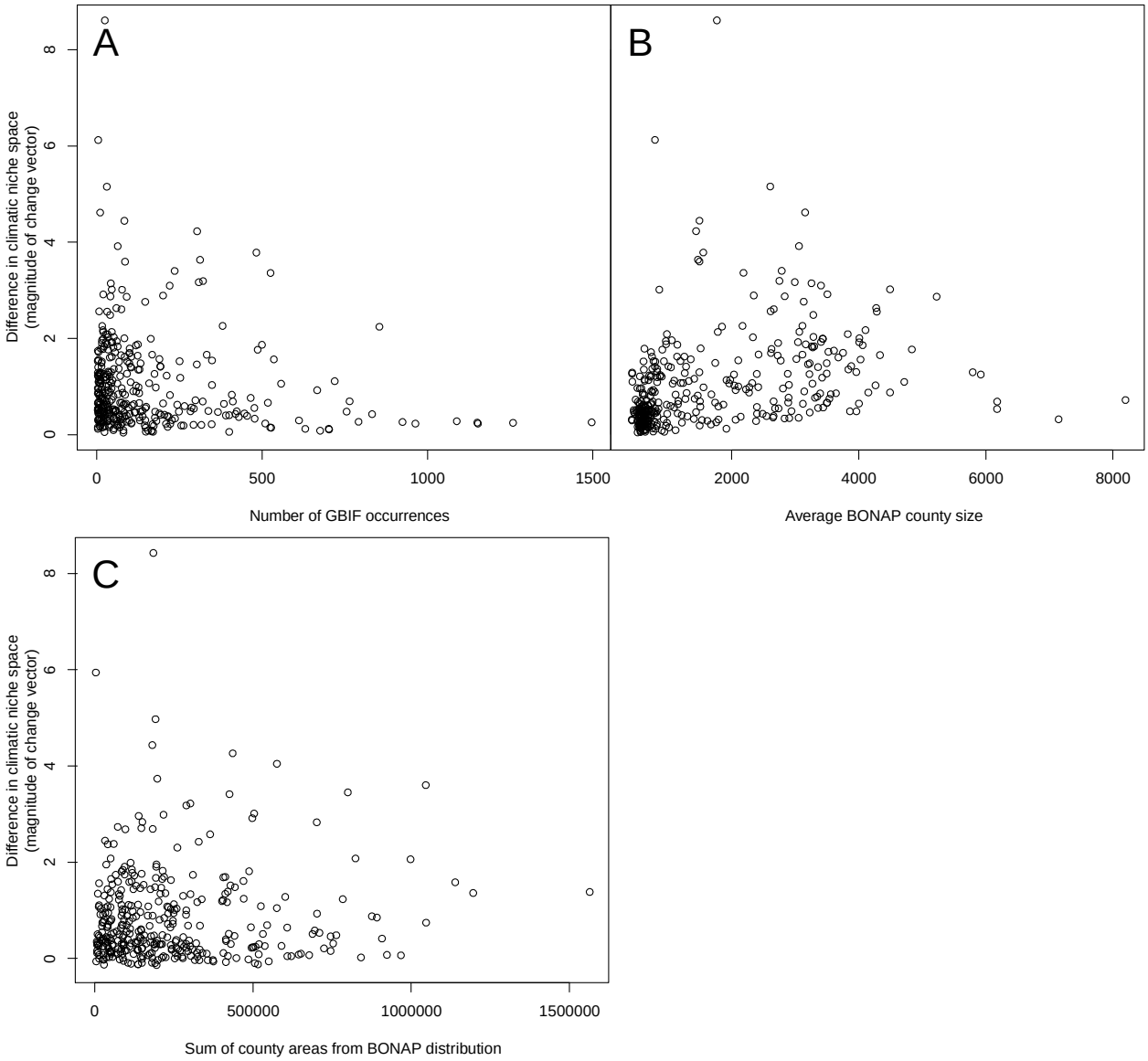


Fig. A1.1 Supplementary analyses searching for correlates of differences in niche estimates. No significant correlations were recovered. Each circle is one *Carex* species. N=363.

Appendix 2. DNA sample data

Table A2.1 *Carex* individuals selected for phylogenetic analysis. Voucher details can be found in Appendix B of Chouinard (2010). Use B numbers for correspondence. Locations where herbarium specimens were collected (column 3) are abbreviated into states or provinces, when possible.

Specific epithet	MOR#	Location	B#	Section
aboriginum	MOR1640	ID	B1231	Racemosae
abrupta	MOR860	BC	B520	Ovales
abscondita	MOR410	DE	B057	Careyanae
acidicola	MOR392	AL	B039	Griseae
acutiformis	MOR1924	Spain	B1516	Paludosae
adelostoma	MOR1421	AK	B1012	Racemosae
adusta	MOR1874	AB	B1466	Ovales
aestivalis	MOR415	PN	B062	Hymenochlaenae
aggregata	MOR419	KY	B066	Phaestoglochin
alata	MOR421	AL	B068	Ovales
albicans	MOR433	MD	B080	Acrocystis
albolutescens	MOR428	OK	B075	Ovales
albonigra	MOR1423	AK	B1014	Racemosae
albursina	MOR1186	KY	B780	Laxiflorae
alma	MOR437	CA	B085	Multiflorae
alopecoidea	MOR1530	IL	B1121	Vulpinae
amphibola	MOR388	AL	B035	Griseae
amplifolia	MOR449	NV	B097	Anomalae
angustata	MOR447	CA	B095	Phacocystis
annectens	MOR464	AL	B112	Multiflorae
anthoxanthea	MOR1370	WA	B961	Circinatae
aperta	MOR1644	BC	B1235	Phacocystis
appalachica	MOR455	MA	B103	Phaestoglochin
aquatilis	MOR1287	LAB	B884	Phacocystis
arapahoensis	MOR1527	CO	B1118	Ovales
arctata	MOR465	WI	B113	Hymenochlaenae
arcta	MOR444	AK	B092	Glareosae
arctiformis	MOR1653	BC	B1244	Glareosae
arenaria	MOR467	DE	B115	Ammoglochin
argyrantha	MOR468	MD	B116	Ovales
arkansana	MOR471	IL	B119	Phaestoglochin
assiniboinensis	MOR456	MB	B104	Hymenochlaenae
atherodes	MOR1651	BC	B1242	Carex
athrostachya	MOR461	CA	B109	Ovales
atlantica	MOR452	FL	B100	Stellulatae
atrata	MOR527	Taiwan	B175	Racemosae
atratiformis	MOR1553	AB	B1144	Racemosae
atrofusca	MOR1456	AK	B1047	Aulocystis
atrosquama	MOR1522	MO	B1113	Racemosae
aurea	MOR501	CA	B149	Bicolores
aureolensis	MOR504	MS	B152	Squarrosae
austrina	MOR524	DE	B172	Phaestoglochin
austrocaroliniana	MOR1531	NC	B1122	Careyanae
backii	MOR1809		B1400	Phyllostachyae
baileyi	MOR485	KY	B133	Vesicariae
baltzellii	MOR1716	FL	B1307	Pictae
barbarae	MOR1251	CA	B846	Phacocystis
barrattii	MOR482	AL	B130	Limosae
basiantha	MOR371	SC	B017	Phyllostachyae
bebbii	MOR479	MB	B127	Ovales
bella	MOR1523	CO	B1114	Racemosae
bicknellii	MOR478	IN	B126	Ovales
bicolor	MOR1424	AK	B1015	Bicolores

Appendix 2. DNA sample data

bigelowii	MOR476	AK	B124	Phacocystis	*chordorrhiza*	MOR1727	BC	B1318	Chordorrhizae
biltmoreana	MOR1870	NC	B1462	Paniceae	*circinata*	MOR1779	AK	B1370	Circinatae
blanda	MOR1621	KS	B1212	Laxiflorae	*collinsii*	MOR561	DE	B209	Collinsiae
bolanderi	MOR1216	CA	B811	Deweyanae	*communis*	MOR1724	GA	B1315	Acrocystis
bonanzensis	MOR1622	AK	B1213	Glareosae	*comosa*	MOR560	CA	B208	Vesicariae
brainerdii	MOR494	CA	B142	Acrocystis	*complanata*	MOR516	TX	B164	Porocystis
brevicaulis	MOR1375	WA	B966	Acrocystis	*concinna*	MOR1499	CO	B1090	Clandestinae
brevior	MOR1625	BC	B1216	Ovales	*concinnoides*	MOR1720	CA	B1311	Clandestinae
breweri	MOR1378	WA	B969	Inflatae	*congdonii*	MOR1488	CA	B1079	Paludosae
bromoides		MD	B022	Deweyanae	*conjuncta*	MOR518	DE	B166	Vulpinae
brunnescens	MOR498	CA	B146	Glareosae	*conoidea*	MOR1230	DE	B825	Griseae
brysonii	MOR389	AL	B036	Griseae	*cordillerana*	MOR1346	BC	B943	Phyllostachyae
bulbostylis	MOR382	MS	B029	Griseae	*corrugata*	MOR404	AL	B051	Griseae
bullata	MOR509	MA	B157	Vesicariae	*crawei*	MOR1273	NFLD	B868	Granulares
bushii	MOR511	DE	B159	Porocystis	*crawfordii*	MOR522	NY	B170	Ovales
buxbaumii	MOR499	AK	B147	Racemosae	*crebriflora*	MOR777		B425	#N/A
calcifugens	MOR402	FL	B049	Griseae	*crinita*	MOR535	ME	B183	Phacocystis
californica	MOR1487	CA	B1078	Paniceae	*cristatella*	MOR534	KY	B182	Ovales
canescens	MOR1298	LAB	B895	Glareosae	*crus-corvi*	MOR556	FL	B204	Vulpinae
capillaris	MOR1896	NT	B1488	Chlorostachyae	*cryptolepis*	MOR1532	IL	B1123	Ceratocystis
capitata	MOR1838		B1429	Capituligerae	*cumberlandensis*	MOR551	AR	B199	Careyanae
careyana	MOR1203	KY	B798	Careyanae	*cumulata*	MOR605	IL	B253	Ovales
caroliniana	MOR549	AL	B197	Porocystis	*curatorum*	MOR1769	UT	B1360	Scirpinae
caryophyllea	MOR1865	lbirica	B1457	Mitratae	*cusickii*	MOR1806	BC	B1397	Heleoglochin
castanea	MOR1281	NFLD	B876	Hymenochlaenae	*dasycarpa*	MOR607	GA	B255	Hallerianae
cephaloidea	MOR538	MA	B186	Phaestoglochin	*davisii*	MOR592	DE	B240	Hymenochlaenae
cephalophora	MOR562	DE	B210	Phaestoglochin	*davyi*	MOR1261	CA	B856	Ovales
chalciolepis	MOR1931	CO	B1523	Racemosae	*debilis*	MOR597	OK	B245	Hymenochlaenae
chapmanii	MOR566	FL	B214	Laxiflorae	*decomposita*	MOR570	DE	B218	Heleoglochin
cherokeensis	MOR564	FL	B212	Hymenochlaenae	*deflexa*	MOR1927	BC	B1519	Acrocystis
chihuahuensis	MOR1614	AZ	B1205	Multiflorae	*densa*	MOR1471	OR	B1062	Multiflorae

deweyana	MOR773	MB	B421	Deweyanae
diandra	MOR574	CA	B222	Heleoglochin
diluta	MOR1366	MO	B957	Spirostachyae
disperma	MOR1454	AK	B1045	Dispermae
distans	MOR578	Greece	B226	Spirostachyae
divisa	MOR1866	MD	B1458	Divisae
divulsa	MOR1383	WA	B974	Phaestoglochin
douglasii	MOR1938	SK	B1530	Divisae
duriuscula	MOR1455	AK	B1046	Divisae
ebenea	MOR1672	AZ	B1263	Ovales
eburnea	MOR576	MI	B224	Albae
echinata	MOR1435	AK	B1026	Stellulatae
edwardsiana	MOR1944	TX	B1536	Griseae
egglestonii	MOR1504	CO	B1095	Ovales
eleusinoides	MOR1425	AK	B1016	Phacocystis
elliottii	MOR1718	FL	B1309	Vesicariae
elynoides	MOR599	NV	B247	Filifoliae
emoryi	MOR619	DE	B267	Phacocystis
endlichii	MOR1897	AZ	B1489	Phacocystis
engelmannii	MOR1933	BC	B1525	Inflatae
epapillosa	MOR620	UT	B268	Racemosae
exilis	MOR621	DE	B269	Stellulatae
exsiccata	MOR1778	BC	B1369	Vesicariae
extensa	MOR1655	VA	B1246	Spirostachyae
festucacea	MOR626	TX	B274	Ovales
feta	MOR627	CA	B275	Ovales
filifolia	MOR1234	CA	B829	Filifoliae
fissa	MOR614	TX	B262	Multiflorae
fissuricola	MOR1919	ID	B1511	Aulocystis
flacca	MOR1611	MI	B1202	Thuringiaca
flaccosperma	MOR405	LO	B052	Griseae
flava	MOR1285	NFLD	B881	Ceratocystis
floridana	MOR640	FL	B288	Acrocystis
foenea	MOR1802	BC	B1393	Ovales
folliculata	MOR645	DE	B293	Rostrales
formosa	MOR1534	IL	B1125	Hymenochlaenae
fracta	MOR630	CA	B278	Ovales
frankii	MOR673	AL	B321	Squarrosae
fraserianus	MOR686	MD	B334	#N/A
fuliginosa	MOR1429	AK	B1020	Aulocystis
garberi	MOR1426	AK	B1017	Bicolores
geophila	MOR674	TX	B322	Acrocystis
geyeri	MOR676	AB	B324	Firmiculmes
gholsonii	MOR687	AL	B335	Granulares
gigantea	MOR651	FL	B299	Lupulinae
glacialis	MOR1290	NFLD	B887	Lamprochlaenae
glareosa	MOR1558	MB	B1149	Glareosae
glaucescens	MOR660	MS	B308	Glaucescentes
glaucodea	MOR1327	IN	B924	Griseae
globosa	MOR654	CA	B302	Acrocystis
gmelinii	MOR1431	AK	B1022	Racemosae
godfreyi	MOR1589	AL	B1180	Griseae
gracilescens	MOR661	KY	B309	Laxiflorae
gracilior	MOR1494	CA	B1085	Ovales
gracillima	MOR656	AL	B304	Hymenochlaenae
granularis	MOR1209	MB	B804	Granulares
gravida	MOR704	KN	B352	Phaestoglochin
grayi	MOR658	KN	B306	Lupulinae
grisea	MOR390	KY	B037	Griseae
gynandra	MOR698	DE	B346	Phacocystis
gynocrates	MOR1432	AK	B1023	Physoglochin
gynodynama	MOR1706	CA	B1297	Hymenochlaenae

Appendix 2. DNA sample data

halliana	MOR1491	CA	B1082	Paludosae	*integra*	MOR711	CA	B359	Ovales
hallii	MOR1556	MB	B1147	Racemosae	*interior*	MOR706	MI	B354	Stellulatae
harfordii	MOR699	CA	B347	Ovales	*interrupta*	MOR1468	OR	B1059	Phacocystis
hassei	MOR1669	CA	B1260	Bicolores	*intumescens*	MOR709	ME	B357	Lupulinae
haydeniana	MOR668	NV	B316	Phacocystis	*jamesii*	MOR364	AL	B010	Phyllostachyae
haydenii	MOR1594	IA	B1185	Phacocystis	*jonesii*	MOR1412	WA	B1003	Vulpinae
heleonastes	MOR1756	AB	B1347	Glareosae	*joorii*	MOR719	FL	B367	Glaucescentes
helleri	MOR1266	NV	B861	Racemosae	*juniperorum*	MOR365	KY	B0011	Phyllostachyae
hendersonii	MOR1749	BC	B1340	Laxiflorae	*klamathensis*	MOR1871	OR	B1463	Paniceae
heteroneura	MOR659	NV	B307	Racemosae	*kobomugi*	MOR713	VA	B361	Macrocephalae
heterostachya	MOR1824	YT	B1415	Paludosae	*kraliana*	MOR715	AR	B363	Laxiflorae
hirsutella	MOR669	DE	B317	Porocystis	*krausei*	MOR1441	AK	B1032	Chlorostachyae
hirta	MOR681	PN	B329	Carex	*lachenalii*	MOR1442	AK	B1033	Glareosae
hirtifolia	MOR1277	QC	B872	Hirtifoliae	*lacustris*	MOR725	MB	B373	Paludosae
hirtissima	MOR1496	CA	B1087	Hymenochlaenae	*laeviconica*	MOR1689	KS	B1280	Carex
hitchcockiana	MOR391	KY	B038	Griseae	*laeviculmis*	MOR1199	CA	B794	Deweyanae
holostoma	MOR1433	AK	B1024	Racemosae	*laevivaginata*	MOR727	AL	B375	Vulpinae
hoodii	MOR1732	CA	B1323	Phaestoglochin	*lapponica*	MOR1443	AK	B1034	Paludosae
hookerana	MOR1915	AB	B1507	Phaestoglochin	*lasiocarpa*	MOR731	AK	B379	Paludosae
hormathodes	MOR1916	NS	B1508	Ovales	*latebracteata*	MOR1185	AR	B779	Phyllostachyae
hostiana	MOR1796	QC	B1387	Ceratocystis	*lativena*	MOR1631	NM	B1222	Hallerianae
houghtoniana	MOR1248	MB	B843	Paludosae	*laxa*	MOR1445	AK	B1036	Paniceae
hyalina	MOR646	MS	B294	Ovales	*laxiculmis*	MOR1205	PN	B800	Careyanae
hyalinolepis	MOR690	TX	B338	Paludosae	*laxiflora*	MOR757	MI	B405	Laxiflorae
hystericina	MOR815	AR	B475	Vesicariae	*leavenworthii*	MOR759	AL	B407	Phaestoglochin
idahoa	MOR1365	MO	B956	Racemosae	*lemmonii*	MOR763	CA	B411	Aulocystis
illota	MOR695	CA	B343	Ovales	*lenticularis*	MOR750	CA	B398	Phacocystis
impressinervia	MOR1586	MS	B1177	Griseae	*leporinella*	MOR1415	WA	B1006	Ovales
incurviformis	MOR1903	CO	B1495	Foetidae	*leptalea*	MOR766	MI	B414	Leptocephalae
infirminervia	MOR1200	CA	B795	Deweyanae	*leptonervia*	MOR1196	MB	B790	Laxiflorae
inops	MOR1803	BC	B1394	Acrocystis	*leptopoda*	MOR1201	BC	B796	Deweyanae

limosa	MOR769	AK	B417	Limosae	*microchaeta*	MOR1748	YT	B1339	Scitae
livida	MOR1212	NJ	B807	Paniceae	*microdonta*	MOR1222	MS	B817	Granulares
loliacea	MOR1452	AK	B1043	Glareosae	*microglochin*	MOR1507	Stordalen	B1098	Leucoglochin
lonchocarpa	MOR733	FL	B381	Rostrales	*micropoda*	MOR1434	AK	B1025	Dornera
longii	MOR735	AR	B383	Ovales	*microptera*	MOR1682	CO	B1273	Ovales
louisianica	MOR742	FL	B390	Lupulinae	*misera*	MOR1566	TN	B1157	Hymenochlaenae
lucorum	MOR1538	NC	B1129	Acrocystis	*missouriensis*	MOR801	IL	B460	Ovales
lupuliformis	MOR745	FL	B393	Lupulinae	*mitchelliana*	MOR835	AL	B495	Phacocystis
lupulina	MOR812	MD	B472	Lupulinae	*molesta*	MOR804	DE	B463	Ovales
lurida	MOR816	FL	B476	Vesicariae	*molestiformis*	MOR807	AR	B466	Ovales
lutea	MOR1567	NC	B1158	Ceratocystis	*muehlenbergii*	MOR1768	AR	B1359	Phaestoglochin
luzulifolia	MOR762	CA	B410	Aulocystis	*multicaulis*	MOR826	AR	B486	Ovales
luzulina	MOR1710	CA	B1301	Aulocystis	*multicostata*	MOR827	CA	B487	Ovales
lyngbyei	MOR819	OR	B479	Phacocystis	*muricata*	MOR1912	Spain	B1504	Phaestoglochin
mackenziei	MOR1459	AK	B1050	Glareosae	*muriculata*	MOR1612	NM	B1203	Phaestoglochin
macloviana	MOR1437	AK	B1028	Ovales	*muskingumensis*	MOR849	MI	B509	Ovales
macrocephala	MOR1236	AK	B831	Macrocephalae	*nardina*	MOR1312	QC	B909	Nardinae
macrochaeta	MOR1473	OR	B1064	Limosae	*nebrascensis*	MOR850	CA	B510	Phacocystis
magellanica	MOR1439	AK	B1030	Limosae	*nelsonii*	MOR1634	MO	B1225	Racemosae
manhartii	MOR1563	GE	B1154	Laxiflorae	*nervina*	MOR1642	CA	B1233	Vulpinae
marina	MOR1548	MB	B1139	Glareosae	*neurophora*	MOR1417	WA	B1008	Vulpinae
mariposana	MOR1686	AZ	B1277	Ovales	*nigra*	MOR1297	NFLD	B894	Phacocystis
meadii	MOR1544	MB	B1135	Paniceae	*nigricans*	MOR1440	AK	B1031	Dornera
media	MOR1574	NB	B1165	Racemosae	*nigromarginata*	MOR855	AL	B515	Acrocystis
melanostachya	MOR1352	OK	B947b	Paludosae	*normalis*	MOR793	AL	B451	Ovales
membranacea	MOR784	AK	B442	Vesicariae	*norvegica*	MOR1418	OR	B1009	Racemosae
mendocinensis	MOR1475	OR	B1066	Hymenochlaenae	*novae-angliae*	MOR795	ME	B453	Acrocystis
merritt-fernaldii	MOR785	MB	B443	Ovales	*nova*	MOR434	CO	B081	Racemosae
mertensii	MOR790	AK	B448	Racemosae	*nudata*	MOR1242	CAs	B837	Phacocystis
mesochorea	MOR831	AL	B491	Phaestoglochin	*obispoensis*	MOR1599	CA	B1190	Hymenochlaenae
michauxiana	MOR1550	MB	B1141	Rostrales	*obnupta*	MOR839	CA	B499	Phacocystis

obtusata	MOR841	MB	B501	Obtusatae	*petricosa*	MOR1797	QC	B1388	Aulocystis
occidentalis	MOR1934	CA	B1526	Phaestoglochin	*phaeocephala*	MOR995	NV	B587	Ovales
oklahomensis	MOR845	AR	B505	Vulpinae	*picta*	MOR893	GA	B553	Pictae
oligocarpa	MOR1334	AR	B931	Griseae	*pigra*	MOR1587	MS	B1178	Griseae
oligosperma	MOR846	NH	B506	Vesicariae	*pityophila*	MOR1632	NM	B1223	Acrocystis
opaca	MOR867	OK	B527	Ovales	*planispicata*	MOR400	MD	B047	Griseae
oreocharis	MOR1678	AZ	B1269	Filifoliae	*planostachys*	MOR884	TXc	B544	Hallerianae
ormostachya	MOR868	MI	B528	Laxiflorae	*plantaginea*	MOR1695	TN	B1286	Careyanae
oronensis	MOR1700	ME	B1291	Ovales	*platyphylla*	MOR886	MD	B546	Careyanae
ouachitana	MOR1340	AR	B937	Griseae	*pluriflora*	MOR1758	AK	B1349	Limosae
ovalis	MOR1582	WS	B1173	Ovales	*podocarpa*	MOR1734	BC	B1325	Scitae
oxylepis	MOR871	KY	B531	Hymenochlaenae	*polymorpha*	MOR890	PN	B550	Paniceae
ozarkana	MOR857	AR	B517	Ovales	*praeceptorum*	MOR1754	CA	B1345	Glareosae
pachystachya	MOR1666	AK	B1257	Ovales	*praegracilis*	MOR891	CA	B551	Divisae
paeninsulae	MOR1247	FLc	B842	Griseae	*prairea*	MOR1753	BC	B1344	Heleoglochin
paleacea	MOR861	ME	B521	Phacocystis	*prasina*	MOR900	GA	B560	Hymenochlaenae
pallescens	MOR1810		B1401	Porocystis	*praticola*	MOR897	CO	B557	Ovales
panicea	MOR1708	ME	B1299	Paniceae	*preslii*	MOR976	OR	B568	Ovales
pansa	MOR1766	CA	B1357	Divisae	*projecta*	MOR978	MB	B570	Ovales
parryana	MOR1436	AK	B1027	Racemosae	*proposita*	MOR1743	CA	B1334	Ovales
pauciflora	MOR876	AK	B536	Leucoglochin	*pseudocyperus*	MOR981		B573	Vesicariae
paupercula	MOR898		B558	Limosae	*pumila*	MOR1869	NC	B1461	Paludosae
peckii	MOR1427	AK	B1018	Acrocystis	*purpurifera*	MOR1272	KY	B867	Laxiflorae
pedunculata	MOR881	MA	B541	Clandestinae	*pyrenaica*	MOR1400	WA	B991	Dornera
pellita	MOR882	CA	B542	Paludosae	*radfordii*	MOR1240	SC	B835	Laxiflorae
pelocarpa	MOR997	NV	B589	Racemosae	*radiata*	MOR1000	MI	B592	Phaestoglochin
pendula	MOR999	Austria	B591	Rhynchocystis	*ramenskii*	MOR1241	AK	B836	Phacocystis
pensylvanica	MOR973	MB	B565	Acrocystis	*rariflora*	MOR1801	YT	B1392	Limosae
perdentata	MOR993	TX	B585	Phaestoglochin	*raynoldsii*	MOR986	CA	B578	Racemosae
perglobosa	MOR1605	CO	B1196	Foetidae	*recta*	MOR1947		B1539	Phacocystis
petasata	MOR994	UT	B586	Ovales	*reniformis*	MOR989	GA	B581	Ovales

Appendix 2. DNA sample data

retroflexa	MOR992	DE	B584	Phaestoglochin	*siccata*	MOR1872	ON	B1464	Ammoglochin
retrorsa	MOR1012	MI	B603	Vesicariae	*silicea*	MOR1034	DE	B625	Ovales
reznicekii	MOR1014	AL	B605	Acrocystis	*simulata*	MOR1086	CA	B678	Divisae
richardsonii	MOR1022	MD	B613	Clandestinae	*socialis*	MOR1090	AL	B682	Phaestoglochin
roanensis	MOR1026	NC	B617	Hymenochlaenae	*sparganioides*	MOR1091	DE	B683	Phaestoglochin
rosea	MOR1016	ME	B607	Phaestoglochin	*specifica*	MOR1094	CA	B686	Ovales
rossii	MOR1003	CA	B595	Acrocystis	*spectabilis*	MOR1781	BC	B1372	Scitae
rostrata	MOR1402	WA	B993	Vesicariae	*specuicola*	MOR1616	AZ	B1207	Racemosae
rotundata	MOR1551	MB	B1142	Vesicariae	*spicata*	MOR1073	DE	B665	Phaestoglochin
rupestris	MOR1923	Spain	B1515	Rupestres	*spissa*	MOR1617	AZ	B1208	Hispidae
ruthii	MOR1008	NC	B600a	Stellulatae	*sprengelii*	MOR1568	NB	B1159	Hymenochlaenae
sabulosa	MOR1448	YT	B1039	Racemosae	*squarrosa*	MOR1083	AR	B675	Squarrosae
salina	MOR1755	QC	B1346	Phacocystis	*stenoptila*	MOR1084	UT	B676	Ovales
sartwelliana	MOR1600	CA	B1191	Paludosae	*sterilis*	MOR1545	MB	B1136	Stellulatae
sartwellii	MOR1784	BC	B1375	Holarrhenae	*stevenii*	MOR1512	CO	B1103	Racemosae
saxatilis	MOR1025		B616	Vesicariae	*stipata*	MOR1055	MB	B646	Vulpinae
saximontana	MOR1529	CO	B1120	Phyllostachyae	*straminea*	MOR1054	DE	B645	Ovales
scabrata	MOR1045	GA	B636	Anomalae	*straminiformis*	MOR1068	UT	B660	Ovales
scabriuscula	MOR1048	CA	B639	Scirpinae	*striata*	MOR1076	FL	B668	Paludosae
schottii	MOR1772	CA	B1363	Phacocystis	*striatula*	MOR1066	DE	B658	Laxiflorae
schweinitzii	MOR1058	MI	B649	Vesicariae	*stricta*	MOR1098	TX	B690	Phacocystis
scirpoidea	MOR1050	AK	B641	Scirpinae	*styloflexa*	MOR1101	GA	B693	Laxiflorae
scoparia	MOR1057	ME	B648	Ovales	*stylosa*	MOR1292	NFLD	B889	Racemosae
scopulorum	MOR1039	CA	B630	Phacocystis	*subbracteata*	MOR1757	CA	B1348	Ovales
senta	MOR1687	AZ	B1278	Phacocystis	*suberecta*	MOR1361	AR	B953	Ovales
seorsa	MOR1044	DE	B635	Stellulatae	*subfusca*	MOR1670	CA	B1261	Ovales
serpenticola	MOR1763	CA	B1354	Acrocystis	*subnigricans*	MOR1476	OR	B1067	Inflatae
serratodens	MOR1028	CA	B619	Racemosae	*subspathacea*	MOR1449	AK	B1040	Phacocystis
sheldonii	MOR1495	CA	B1086	Carex	*superata*	MOR366	KY	B012	Phyllostachyae
shinnersii	MOR1027	KN	B618	Ovales	*supina*	MOR1596	AK	B1187	Lamprochlaenae
shortiana	MOR1031	KN	B622	Shortianae	*swanii*	MOR1107		B700	Porocystis

sychnocephala	MOR1702	ON	B1293	Cyperoideae
sylvatica	MOR1392	WA	B983	Hymenochlaenae
tahoensis	MOR1477	OR	B1068	Ovales
tenax	MOR1110	LO	B703	Hallerianae
tenera	MOR1552	MB	B1143	Ovales
tenuiflora	MOR1115	MB	B708	Glareosae
tetanica	MOR1116	MB	B709	Paniceae
tetrastachya	MOR1071	TXk	B663	Ovales
texensis	MOR1482	CA	B1073	Phaestoglochin
thornei	MOR1542	GA	B1133	Griseae
thurberi	MOR1618	AZ	B1209	Vesicariae
timida	MOR359	AL	B005	Phyllostachyae
tincta	MOR1144	ME	B737	Ovales
tompkinsii	MOR1485	CA	B1076	Firmiculmes
tonsa	MOR1119		B712	Acrocystis
torreyi	MOR1510	CO	B1101	Porocystis
torta	MOR1122	PN	B715	Phacocystis
triangularis	MOR1146	TX	B739	Multiflorae
tribuloides	MOR1125		B718	Ovales
trichocarpa	MOR1133	DE	B726	Carex
triquetra	MOR1602	CA	B1193	Triquetrae
trisperma	MOR1136	MI	B729	Glareosae
tuckermanii	MOR1147	MI	B740	Vesicariae
tumulicola	MOR1140	CA	B733	Phaestoglochin
turbinata	MOR1679	AZ	B1270	Acrocystis
turgescens	MOR1244	MS	B839	Rostrales
typhina	MOR1132	AR	B725	Squarrosae
umbellata	MOR1179	MB	B772	Acrocystis
unilateralis	MOR1394	WA	B985	Ovales
ursina	MOR1777	YT	B1368	Glareosae
utriculata	MOR1181	AK	B774	Vesicariae
vacillans	MOR1691	ME	B1282	Phacocystis
vaginata	MOR1776	YT	B1367	Paniceae
vallicola	MOR1788	BC	B1379	Phaestoglochin
venusta	MOR1174	MS	B767	Hymenochlaenae
vernacula	MOR1264	CA	B859	Foetidae
verrucosa	MOR1175	FL	B768	Glaucescentes
vesicaria	MOR1162	DE	B755	Vesicariae
vestita	MOR1165	DE	B758	Paludosae
vexans	MOR1168	FLh	B761	Ovales
virescens	MOR1158	AL	B751	Porocystis
viridula	MOR1750	NS	B1341	Ceratocystis
viridustellata	MOR1820		B1411	Ceratocystis
vulpinoidea	MOR1153	CA	B746	Multiflorae
whitneyi	MOR1154	CA	B747	Leucoglochin
wiegandii	MOR1155	ME	B748	Stellulatae
willdenowii	MOR358	KY	B003	Phyllostachyae
williamsii	MOR1771	YT	B1362	Chlorostachyae
woodii	MOR1592	IA	B1183	Paniceae
wootonii	MOR1615	AZ	B1206	Ovales
xerantica	MOR1715	ON	B1306	Ovales

Table A2.2 Voucher information for outgroups used in this study.

DNA number	Species name	Collection number	Herbarium	Accession number	Collector(s)	Country	Province
STA1816	*Trichophourm alpinum*	861	CAN	312672	Porsild, R T	Canada	Yukon
STA1817	*Trichophorum cespitosum*	1219	CAN	589883	Saarela, J M; Percy, D M	Canada	British Columbia

Appendix 3. Supplementary Figures for Chapter 3

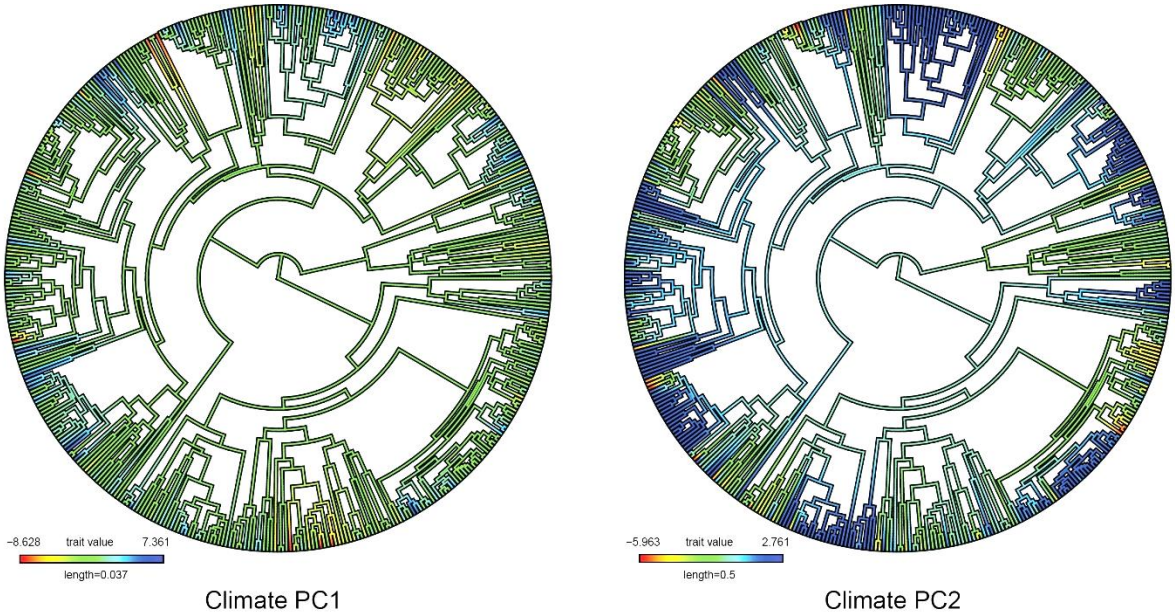


Fig. A3.1 Climate data plotted on the *Carex* spp. ultrametric phylogeny. Climate variables were taken from the BIOCLIM WorldClim database. Principal component analysis (PCA) was used to reduce the means of 19 BIOCLIM variables (*prcomp* in R) into climate PC1 and PC2. The map was generated using the *contMap* function in the *phytools* package for R. N=445.

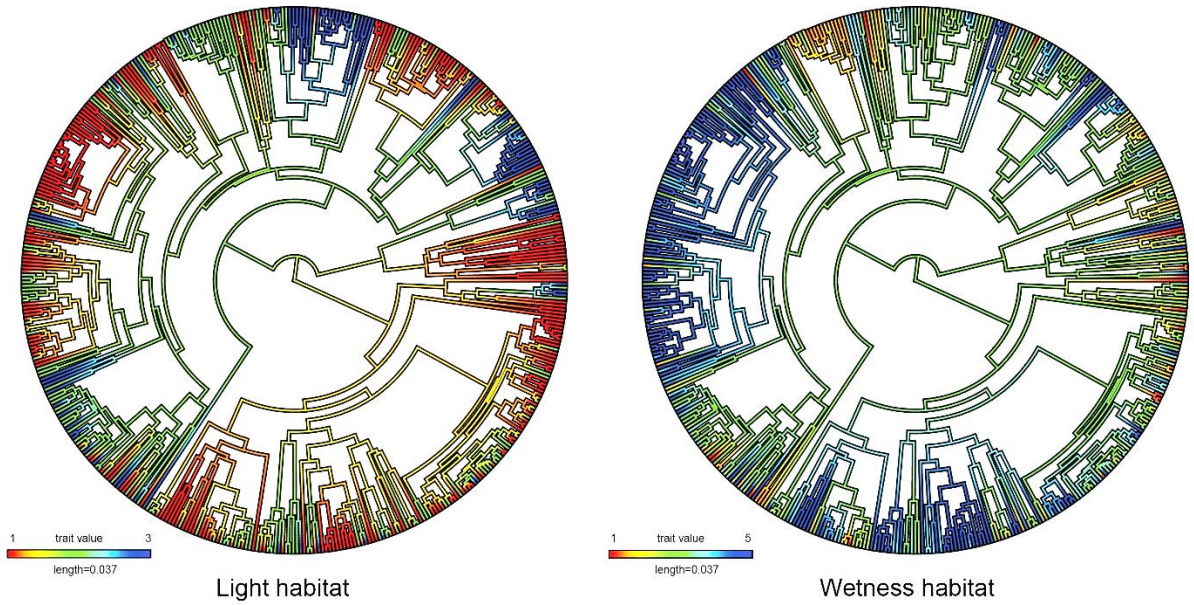


Fig. A3.2 Habitat data plotted on the *Carex* spp. ultrametric phylogeny. Light habitat was scored by experts. Wetness habitat was taken from the National Wetland Plant List (NWPL; Lichvar, 2012). Species missing from the NWPL were scored by experts for wetness habitat. The map was generated using the *contMap* function in the *phytools* package for R. Light habitat N=463, wetness habitat N=460.

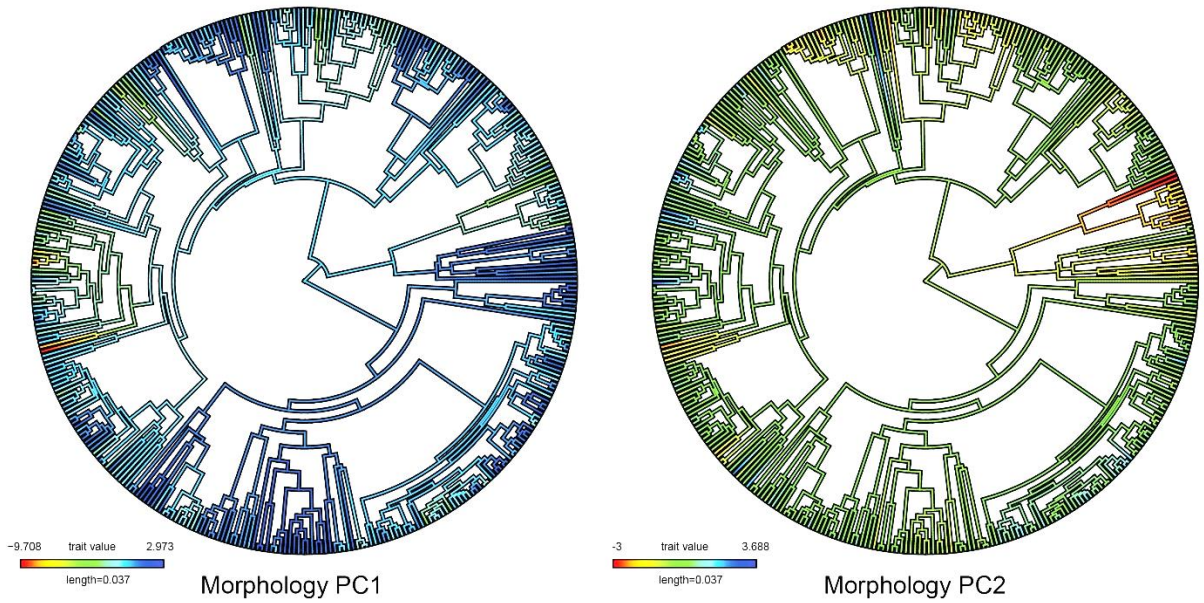


Fig. A3.3 Morphological data for those individuals with complete data sampling (data for four characters available) plotted on the *Carex* spp. ultrametric phylogeny. Morphological data was extracted from the Flora of North America (2002). Principal component analysis was used to reduce the number of morphological variables. Only PC1 and PC2 were analysed. The map was generated using the *contMap* function in the *phytools* package for R. N=371.

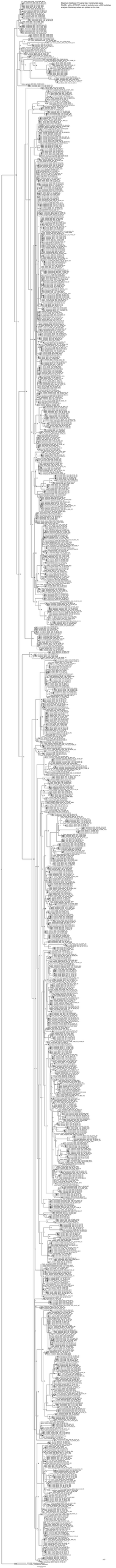
Appendix 4: Maximum Likelihood gene trees

Attached are the gene trees used to select sequences for final analysis. Unfortunately, they cannot be viewed in the printed version of this thesis. In the PDF file, tree topology and label information can be seen by zooming in.

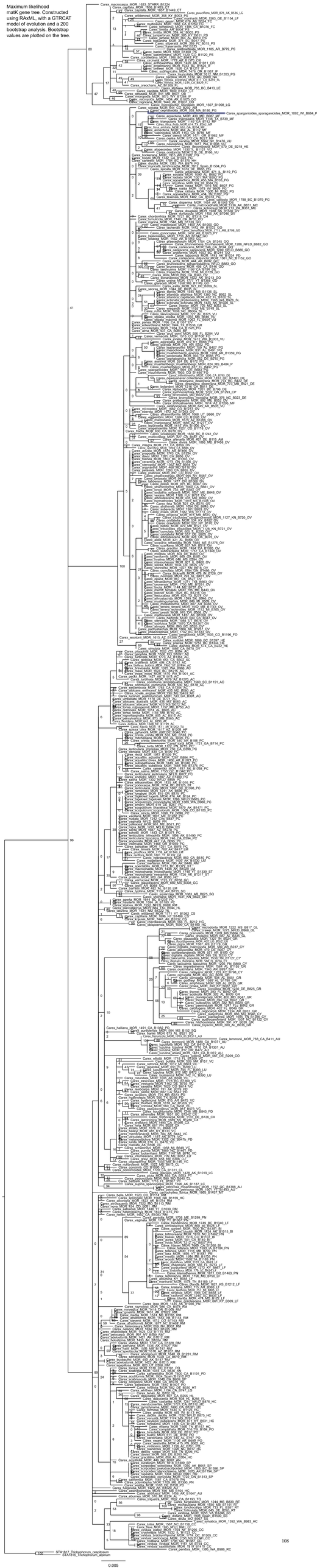
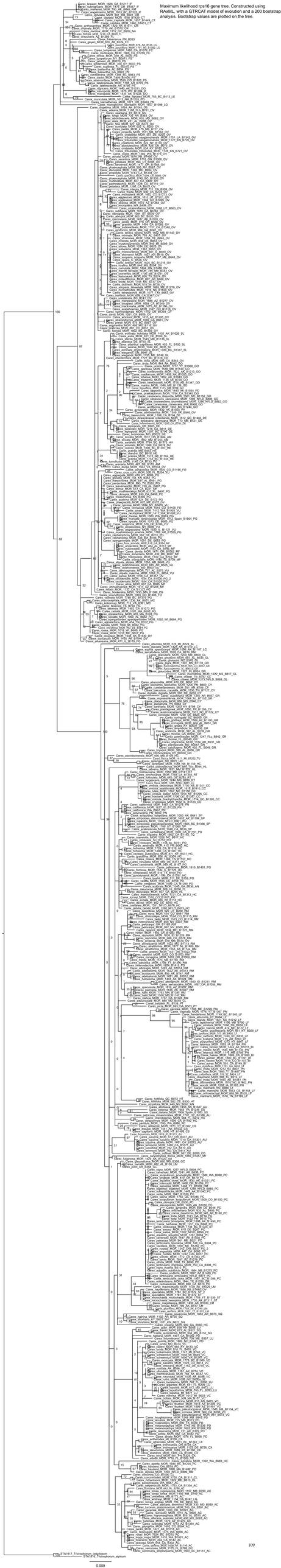
Maximum likelihood ETR gene tree. Concatenated using
Phylomerge. Tip labels are prefixed with species codes. Tree
simplified. Bootstrap values are printed on the tree.

Ccr_ETR_Ccr_ETR_Ccr_ETR_1
Ccr_ETR_Ccr_ETR_Ccr_ETR_2
Ccr_ETR_Ccr_ETR_Ccr_ETR_3
Ccr_ETR_Ccr_ETR_Ccr_ETR_4
Ccr_ETR_Ccr_ETR_Ccr_ETR_5
Ccr_ETR_Ccr_ETR_Ccr_ETR_6
Ccr_ETR_Ccr_ETR_Ccr_ETR_7
Ccr_ETR_Ccr_ETR_Ccr_ETR_8
Ccr_ETR_Ccr_ETR_Ccr_ETR_9
Ccr_ETR_Ccr_ETR_Ccr_ETR_10
Ccr_ETR_Ccr_ETR_Ccr_ETR_11
Ccr_ETR_Ccr_ETR_Ccr_ETR_12
Ccr_ETR_Ccr_ETR_Ccr_ETR_13
Ccr_ETR_Ccr_ETR_Ccr_ETR_14
Ccr_ETR_Ccr_ETR_Ccr_ETR_15
Ccr_ETR_Ccr_ETR_Ccr_ETR_16
Ccr_ETR_Ccr_ETR_Ccr_ETR_17
Ccr_ETR_Ccr_ETR_Ccr_ETR_18
Ccr_ETR_Ccr_ETR_Ccr_ETR_19
Ccr_ETR_Ccr_ETR_Ccr_ETR_20
Ccr_ETR_Ccr_ETR_Ccr_ETR_21
Ccr_ETR_Ccr_ETR_Ccr_ETR_22
Ccr_ETR_Ccr_ETR_Ccr_ETR_23
Ccr_ETR_Ccr_ETR_Ccr_ETR_24
Ccr_ETR_Ccr_ETR_Ccr_ETR_25
Ccr_ETR_Ccr_ETR_Ccr_ETR_26
Ccr_ETR_Ccr_ETR_Ccr_ETR_27
Ccr_ETR_Ccr_ETR_Ccr_ETR_28
Ccr_ETR_Ccr_ETR_Ccr_ETR_29
Ccr_ETR_Ccr_ETR_Ccr_ETR_30
Ccr_ETR_Ccr_ETR_Ccr_ETR_31
Ccr_ETR_Ccr_ETR_Ccr_ETR_32
Ccr_ETR_Ccr_ETR_Ccr_ETR_33
Ccr_ETR_Ccr_ETR_Ccr_ETR_34
Ccr_ETR_Ccr_ETR_Ccr_ETR_35
Ccr_ETR_Ccr_ETR_Ccr_ETR_36
Ccr_ETR_Ccr_ETR_Ccr_ETR_37
Ccr_ETR_Ccr_ETR_Ccr_ETR_38
Ccr_ETR_Ccr_ETR_Ccr_ETR_39
Ccr_ETR_Ccr_ETR_Ccr_ETR_40
Ccr_ETR_Ccr_ETR_Ccr_ETR_41

Maximum likelihood TD gene tree. Constructed using IQ-TREE v1.6.12 under the model TIM3+F+I+R4 bootstrapping analysis. Bootstrap values are printed on the tree.


Ecoli_ybiT_P76386_
Ecoli_ycfM_P76387_
Ecoli_ycfN_P76388_
Ecoli_ycfP_P76389_
Ecoli_ycfQ_P76390_
Ecoli_ycfR_P76391_
Ecoli_ycfS_P76392_
Ecoli_ycfT_P76393_
Ecoli_ycfU_P76394_
Ecoli_ycfV_P76395_
Ecoli_ycfW_P76396_
Ecoli_ycfX_P76397_
Ecoli_ycfY_P76398_
Ecoli_ycfZ_P76399_
Ecoli_yceA_P76400_
Ecoli_yceB_P76401_
Ecoli_yceC_P76402_
Ecoli_yceD_P76403_
Ecoli_yceE_P76404_
Ecoli_yceF_P76405_
Ecoli_yceG_P76406_
Ecoli_yceH_P76407_
Ecoli_yceI_P76408_
Ecoli_yceJ_P76409_
Ecoli_yceK_P76410_
Ecoli_yceL_P76411_
Ecoli_yceM_P76412_
Ecoli_yceN_P76413_
Ecoli_yceO_P76414_
Ecoli_yceP_P76415_
Ecoli_yceQ_P76416_
Ecoli_yceR_P76417_
Ecoli_yceS_P76418_
Ecoli_yceT_P76419_
Ecoli_yceU_P76420_
Ecoli_yceV_P76421_
Ecoli_yceW_P76422_
Ecoli_yceX_P76423_
Ecoli_yceY_P76424_
Ecoli_yceZ_P76425_
Ecoli_ydaA_P76426_
Ecoli_ydaB_P76427_
Ecoli_ydaC_P76428_
Ecoli_ydaD_P76429_
Ecoli_ydaE_P76430_
Ecoli_ydaF_P76431_
Ecoli_ydaG_P76432_
Ecoli_ydaH_P76433_
Ecoli_ydaI_P76434_
Ecoli_ydaJ_P76435_
Ecoli_ydaK_P76436_
Ecoli_ydaL_P76437_
Ecoli_ydaM_P76438_
Ecoli_ydaN_P76439_
Ecoli_ydaO_P76440_
Ecoli_ydaP_P76441_
Ecoli_ydaQ_P76442_
Ecoli_ydaR_P76443_
Ecoli_ydaS_P76444_
Ecoli_ydaT_P76445_
Ecoli_ydaU_P76446_
Ecoli_ydaV_P76447_
Ecoli_ydaW_P76448_
Ecoli_ydaX_P76449_
Ecoli_ydaY_P76450_
Ecoli_ydaZ_P76451_
Ecoli_yeaA_P76452_
Ecoli_yeaB_P76453_
Ecoli_yeaC_P76454_
Ecoli_yeaD_P76455_
Ecoli_yeaE_P76456_
Ecoli_yeaF_P76457_
Ecoli_yeaG_P76458_
Ecoli_yeaH_P76459_
Ecoli_yeaI_P76460_
Ecoli_yeaJ_P76461_
Ecoli_yeaK_P76462_
Ecoli_yeaL_P76463_
Ecoli_yeaM_P76464_
Ecoli_yeaN_P76465_
Ecoli_yeaO_P76466_
Ecoli_yeaP_P76467_
Ecoli_yeaQ_P76468_
Ecoli_yeaR_P76469_
Ecoli_yeaS_P76470_
Ecoli_yeaT_P76471_
Ecoli_yeaU_P76472_
Ecoli_yeaV_P76473_
Ecoli_yeaW_P76474_
Ecoli_yeaX_P76475_
Ecoli_yeaY_P76476_
Ecoli_yeaZ_P76477_
Ecoli_yebA_P76478_
Ecoli_yebB_P76479_
Ecoli_yebC_P76480_
Ecoli_yebD_P76481_
Ecoli_yebE_P76482_
Ecoli_yebF_P76483_
Ecoli_yebG_P76484_
Ecoli_yebH_P76485_
Ecoli_yebI_P76486_
Ecoli_yebJ_P76487_
Ecoli_yebK_P76488_
Ecoli_yebL_P76489_
Ecoli_yebM_P76490_
Ecoli_yebN_P76491_
Ecoli_yebO_P76492_
Ecoli_yebP_P76493_
Ecoli_yebQ_P76494_
Ecoli_yebR_P76495_
Ecoli_yebS_P76496_
Ecoli_yebT_P76497_
Ecoli_yebU_P76498_
Ecoli_yebV_P76499_
Ecoli_yebW_P76500_
Ecoli_yebX_P76501_
Ecoli_yebY_P76502_
Ecoli_yebZ_P76503_
Ecoli_yecA_P76504_
Ecoli_yecB_P76505_
Ecoli_yecC_P76506_
Ecoli_yecD_P76507_
Ecoli_yecE_P76508_
Ecoli_yecF_P76509_
Ecoli_yecG_P76510_
Ecoli_yecH_P76511_
Ecoli_yecI_P76512_
Ecoli_yecJ_P76513_
Ecoli_yecK_P76514_
Ecoli_yecL_P76515_
Ecoli_yecM_P76516_
Ecoli_yecN_P76517_
Ecoli_yecO_P76518_
Ecoli_yecP_P76519_
Ecoli_yecQ_P76520_
Ecoli_yecR_P76521_
Ecoli_yecS_P76522_
Ecoli_yecT_P76523_
Ecoli_yecU_P76524_
Ecoli_yecV_P76525_
Ecoli_yecW_P76526_
Ecoli_yecX_P76527_
Ecoli_yecY_P76528_
Ecoli_yecZ_P76529_
Ecoli_yedA_P76530_
Ecoli_yedB_P76531_
Ecoli_yedC_P76532_
Ecoli_yedD_P76533_
Ecoli_yedE_P76534_
Ecoli_yedF_P76535_
Ecoli_yedG_P76536_
Ecoli_yedH_P76537_
Ecoli_yedI_P76538_
Ecoli_yedJ_P76539_
Ecoli_yedK_P76540_
Ecoli_yedL_P76541_
Ecoli_yedM_P76542_
Ecoli_yedN_P76543_
Ecoli_yedO_P76544_
Ecoli_yedP_P76545_
Ecoli_yedQ_P76546_
Ecoli_yedR_P76547_
Ecoli_yedS_P76548_
Ecoli_yedT_P76549_
Ecoli_yedU_P76550_
Ecoli_yedV_P76551_
Ecoli_yedW_P76552_
Ecoli_yedX_P76553_
Ecoli_yedY_P76554_
Ecoli_yedZ_P76555_
Ecoli_yeeA_P76556_
Ecoli_yeeB_P76557_
Ecoli_yeeC_P76558_
Ecoli_yeeD_P76559_
Ecoli_yeeE_P76560_
Ecoli_yeeF_P76561_
Ecoli_yeeG_P76562_
Ecoli_yeeH_P76563_
Ecoli_yeeI_P76564_
Ecoli_yeeJ_P76565_
Ecoli_yeeK_P76566_
Ecoli_yeeL_P76567_
Ecoli_yeeM_P76568_
Ecoli_yeeN_P76569_
Ecoli_yeeO_P76570_
Ecoli_yeeP_P76571_
Ecoli_yeeQ_P76572_
Ecoli_yeeR_P76573_
Ecoli_yeeS_P76574_
Ecoli_yeeT_P76575_
Ecoli_yeeU_P76576_
Ecoli_yeeV_P76577_
Ecoli_yeeW_P76578_
Ecoli_yeeX_P76579_
Ecoli_yeeY_P76580_
Ecoli_yeeZ_P76581_
Ecoli_yefA_P76582_
Ecoli_yefB_P76583_
Ecoli_yefC_P76584_
Ecoli_yefD_P76585_
Ecoli_yefE_P76586_
Ecoli_yefF_P76587_
Ecoli_yefG_P76588_
Ecoli_yefH_P76589_
Ecoli_yefI_P76590_
Ecoli_yefJ_P76591_
Ecoli_yefK_P76592_
Ecoli_yefL_P76593_
Ecoli_yefM_P76594_
Ecoli_yefN_P76595_
Ecoli_yefO_P76596_
Ecoli_yefP_P76597_
Ecoli_yefQ_P76598_
Ecoli_yefR_P76599_
Ecoli_yefS_P76600_
Ecoli_yefT_P76601_
Ecoli_yefU_P76602_
Ecoli_yefV_P76603_
Ecoli_yefW_P76604_
Ecoli_yefX_P76605_
Ecoli_yefY_P76606_
Ecoli_yefZ_P76607_
Ecoli_yegA_P76608_
Ecoli_yegB_P76609_
Ecoli_yegC_P76610_
Ecoli_yegD_P76611_
Ecoli_yegE_P76612_
Ecoli_yegF_P76613_
Ecoli_yegG_P76614_
Ecoli_yegH_P76615_
Ecoli_yegI_P76616_
Ecoli_yegJ_P76617_
Ecoli_yegK_P76618_
Ecoli_yegL_P76619_
Ecoli_yegM_P76620_
Ecoli_yegN_P76621_
Ecoli_yegO_P76622_
Ecoli_yegP_P76623_
Ecoli_yegQ_P76624_
Ecoli_yegR_P76625_
Ecoli_yegS_P76626_
Ecoli_yegT_P76627_
Ecoli_yegU_P76628_
Ecoli_yegV_P76629_
Ecoli_yegW_P76630_
Ecoli_yegX_P76631_
Ecoli_yegY_P76632_
Ecoli_yegZ_P76633_
Ecoli_yehA_P76634_
Ecoli_yehB_P76635_
Ecoli_yehC_P76636_
Ecoli_yehD_P76637_
Ecoli_yehE_P76638_
Ecoli_yehF_P76639_
Ecoli_yehG_P76640_
Ecoli_yehH_P76641_
Ecoli_yehI_P76642_
Ecoli_yehJ_P76643_
Ecoli_yehK_P76644_
Ecoli_yehL_P76645_
Ecoli_yehM_P76646_
Ecoli_yehN_P76647_
Ecoli_yehO_P76648_
Ecoli_yehP_P76649_
Ecoli_yehQ_P76650_
Ecoli_yehR_P76651_
Ecoli_yehS_P76652_
Ecoli_yehT_P76653_
Ecoli_yehU_P76654_
Ecoli_yehV_P76655_
Ecoli_yehW_P76656_
Ecoli_yehX_P76657_
Ecoli_yehY_P76658_
Ecoli_yehZ_P76659_
Ecoli_yeiA_P76660_
Ecoli_yeiB_P76661_
Ecoli_yeiC_P76662_
Ecoli_yeiD_P76663_
Ecoli_yeiE_P76664_
Ecoli_yeiF_P76665_
Ecoli_yeiG_P76666_
Ecoli_yeiH_P76667_
Ecoli_yeiI_P76668_
Ecoli_yeiJ_P76669_
Ecoli_yeiK_P76670_
Ecoli_yeiL_P76671_
Ecoli_yeiM_P76672_
Ecoli_yeiN_P76673_
Ecoli_yeiO_P76674_
Ecoli_yeiP_P76675_
Ecoli_yeiQ_P76676_
Ecoli_yeiR_P76677_
Ecoli_yeiS_P76678_
Ecoli_yeiT_P76679_
Ecoli_yeiU_P76680_
Ecoli_yeiV_P76681_
Ecoli_yeiW_P76682_
Ecoli_yeiX_P76683_
Ecoli_yeiY_P76684_
Ecoli_yeiZ_P76685_
Ecoli_yejA_P76686_
Ecoli_yejB_P76687_
Ecoli_yejC_P76688_
Ecoli_yejD_P76689_
Ecoli_yejE_P76690_
Ecoli_yejF_P76691_
Ecoli_yejG_P76692_
Ecoli_yejH_P76693_
Ecoli_yejI_P76694_
Ecoli_yejJ_P76695_
Ecoli_yejK_P76696_
Ecoli_yejL_P76697_
Ecoli_yejM_P76698_
Ecoli_yejN_P76699_
Ecoli_yejO_P76700_
Ecoli_yejP_P76701_
Ecoli_yejQ_P76702_
Ecoli_yejR_P76703_
Ecoli_yejS_P76704_
Ecoli_yejT_P76705_
Ecoli_yejU_P76706_
Ecoli_yejV_P76707_
Ecoli_yejW_P76708_
Ecoli_yejX_P76709_
Ecoli_yejY_P76710_
Ecoli_yejZ_P76711_
Ecoli_yekA_P76712_
Ecoli_yekB_P76713_
Ecoli_yekC_P76714_
Ecoli_yekD_P76715_
Ecoli_yekE_P76716_
Ecoli_yekF_P76717_
Ecoli_yekG_P76718_
Ecoli_yekH_P76719_
Ecoli_yekI_P76720_
Ecoli_yekJ_P76721_
Ecoli_yekK_P76722_
Ecoli_yekL_P76723_
Ecoli_yekM_P76724_
Ecoli_yekN_P76725_
Ecoli_yekO_P76726_
Ecoli_yekP_P76727_
Ecoli_yekQ_P76728_
Ecoli_yekR_P76729_
Ecoli_yekS_P76730_
Ecoli_yekT_P76731_
Ecoli_yekU_P76732_
Ecoli_yekV_P76733_
Ecoli_yekW_P76734_
Ecoli_yekX_P76735_
Ecoli_yekY_P76736_
Ecoli_yekZ_P76737_
Ecoli_yelA_P76738_
Ecoli_yelB_P76739_
Ecoli_yelC_P76740_
Ecoli_yelD_P76741_
Ecoli_yelE_P76742_
Ecoli_yelF_P76743_
Ecoli_yelG_P76744_
Ecoli_yelH_P76745_
Ecoli_yelI_P76746_
Ecoli_yelJ_P76747_
Ecoli_yelK_P76748_
Ecoli_yelL_P76749_
Ecoli_yelM_P76750_
Ecoli_yelN_P76751_
Ecoli_yelO_P76752_
Ecoli_yelP_P76753_
Ecoli_yelQ_P76754_
Ecoli_yelR_P76755_
Ecoli_yelS_P76756_
Ecoli_yelT_P76757_
Ecoli_yelU_P76758_
Ecoli_yelV_P76759_
Ecoli_yelW_P76760_
Ecoli_yelX_P76761_
Ecoli_yelY_P76762_
Ecoli_yelZ_P76763_
Ecoli_yemA_P76764_
Ecoli_yemB_P76765_
Ecoli_yemC_P76766_
Ecoli_yemD_P76767_
Ecoli_yemE_P76768_
Ecoli_yemF_P76769_
Ecoli_yemG_P76770_
Ecoli_yemH_P76771_
Ecoli_yemI_P76772_
Ecoli_yemJ_P76773_
Ecoli_yemK_P76774_
Ecoli_yemL_P76775_
Ecoli_yemM_P76776_
Ecoli_yemN_P76777_
Ecoli_yemO_P76778_
Ecoli_yemP_P76779_
Ecoli_yemQ_P76780_
Ecoli_yemR_P76781_
Ecoli_yemS_P76782_
Ecoli_yemT_P76783_
Ecoli_yemU_P76784_
Ecoli_yemV_P76785_
Ecoli_yemW_P76786_
Ecoli_yemX_P76787_
Ecoli_yemY_P76788_
Ecoli_yemZ_P76789_
Ecoli_yenA_P76790_
Ecoli_yenB_P76791_
Ecoli_yenC_P76792_
Ecoli_yenD_P76793_
Ecoli_yenE_P76794_
Ecoli_yenF_P76795_
Ecoli_yenG_P76796_
Ecoli_yenH_P76797_
Ecoli_yenI_P76798_
Ecoli_yenJ_P76799_
Ecoli_yenK_P76800_
Ecoli_yenL_P76801_
Ecoli_yenM_P76802_
Ecoli_yenN_P76803_
Ecoli_yenO_P76804_
Ecoli_yenP_P76805_
Ecoli_yenQ_P76806_
Ecoli_yenR_P76807_
Ecoli_yenS_P76808_
Ecoli_yenT_P76809_
Ecoli_yenU_P76810_
Ecoli_yenV_P76811_
Ecoli_yenW_P76812_
Ecoli_yenX_P76813_
Ecoli_yenY_P76814_
Ecoli_yenZ_P76815_
Ecoli_yeoA_P76816_
Ecoli_yeoB_P76817_
Ecoli_yeoC_P76818_
Ecoli_yeoD_P76819_
Ecoli_yeoE_P76820_
Ecoli_yeoF_P76821_
Ecoli_yeoG_P76822_
Ecoli_yeoH_P76823_
Ecoli_yeoI_P76824_
Ecoli_yeoJ_P76825_
Ecoli_yeoK_P76826_
Ecoli_yeoL_P76827_
Ecoli_yeoM_P76828_
Ecoli_yeoN_P76829_
Ecoli_yeoO_P76830_
Ecoli_yeoP_P76831_
Ecoli_yeoQ_P76832_
Ecoli_yeoR_P76833_
Ecoli_yeoS_P76834_
Ecoli_yeoT_P76835_
Ecoli_yeoU_P76836_
Ecoli_yeoV_P76837_
Ecoli_yeoW_P76838_
Ecoli_yeoX_P76839_
Ecoli_yeoY_P76840_
Ecoli_yeoZ_P76841_
Ecoli_yepA_P76842_
Ecoli_yepB_P76843_
Ecoli_yepC_P76844_
Ecoli_yepD_P76845_
Ecoli_yepE_P76846_
Ecoli_yepF_P76847_
Ecoli_yepG_P76848_
Ecoli_yepH_P76849_
Ecoli_yepI_P76850_
Ecoli_yepJ_P76851_
Ecoli_yepK_P76852_
Ecoli_yepL_P76853_
Ecoli_yepM_P76854_
Ecoli_yepN_P76855_
Ecoli_yepO_P76856_
Ecoli_yepP_P76857_
Ecoli_yepQ_P76858_
Ecoli_yepR_P76859_
Ecoli_yepS_P76860_
Ecoli_yepT_P76861_
Ecoli_yepU_P76862_
Ecoli_yepV_P76863_
Ecoli_yepW_P76864_
Ecoli_yepX_P76865_
Ecoli_yepY_P76866_
Ecoli_yepZ_P76867_
Ecoli_yeqA_P76868_
Ecoli_yeqB_P76869_
Ecoli_yeqC_P76870_
Ecoli_yeqD_P76871_
Ecoli_yeqE_P76872_
Ecoli_yeqF_P76873_
Ecoli_yeqG_P76874_
Ecoli_yeqH_P76875_
Ecoli_yeqI_P76876_
Ecoli_yeqJ_P76877_
Ecoli_yeqK_P76878_
Ecoli_yeqL_P76879_
Ecoli_yeqM_P76880_
Ecoli_yeqN_P76881_
Ecoli_yeqO_P76882_
Ecoli_yeqP_P76883_
Ecoli_yeqQ_P76884_
Ecoli_yeqR_P76885_
Ecoli_yeqS_P76886_
Ecoli_yeqT_P76887_
Ecoli_yeqU_P76888_
Ecoli_yeqV_P76889_
Ecoli_yeqW_P76890_
Ecoli_yeqX_P76891_
Ecoli_yeqY_P76892_
Ecoli_yeqZ_P76893_
Ecoli_yerA_P76894_
Ecoli_yerB_P76895_
Ecoli_yerC_P76896_
Ecoli_yerD_P76897_
Ecoli_yerE_P76898_
Ecoli_yerF_P76899_
Ecoli_yerG_P76900_
Ecoli_yerH_P76901_
Ecoli_yerI_P76902_
Ecoli_yerJ_P76903_
Ecoli_yerK_P76904_
Ecoli_yerL_P76905_
Ecoli_yerM_P76906_
Ecoli_yerN_P76907_
Ecoli_yerO_P76908_
Ecoli_yerP_P76909_
Ecoli_yerQ_P76910_
Ecoli_yerR_P76911_
Ecoli_yerS_P76912_
Ecoli_yerT_P76913_
Ecoli_yerU_P76914_
Ecoli_yerV_P76915_
Ecoli_yerW_P76916_
Ecoli_yerX_P76917_
Ecoli_yerY_P76918_
Ecoli_yerZ_P76919_
Ecoli_yesA_P76920_
Ecoli_yesB_P76921_
Ecoli_yesC_P76922_
Ecoli_yesD_P76923_
Ecoli_yesE_P76924_
Ecoli_yesF_P76925_
Ecoli_yesG_P76926_
Ecoli_yesH_P76927_
Ecoli_yesI_P76928_
Ecoli_yesJ_P76929_
Ecoli_yesK_P76930_
Ecoli_yesL_P76931_
Ecoli_yesM_P76932_
Ecoli_yesN_P76933_
Ecoli_yesO_P76934_
Ecoli_yesP_P76935_
Ecoli_yesQ_P76936_
Ecoli_yesR_P76937_
Ecoli_yesS_P76938_
Ecoli_yesT_P76939_
Ecoli_yesU_P76940_
Ecoli_yesV_P76941_
Ecoli_yesW_P76942_
Ecoli_yesX_P76943_
Ecoli_yesY_P76944_
Ecoli_yesZ_P76945_
Ecoli_yetA_P76946_
Ecoli_yetB_P76947_
Ecoli_yetC_P76948_
Ecoli_yetD_P76949_
Ecoli_yetE_P76950_
Ecoli_yetF_P76951_
Ecoli_yetG_P76952_
Ecoli_yetH_P76953_
Ecoli_yetI_P76954_
Ecoli_yetJ_P76955_
Ecoli_yetK_P76956_
Ecoli_yetL_P76957_
Ecoli_yetM_P76958_
Ecoli_yetN_P76959_
Ecoli_yetO_P76960_
Ecoli_yetP_P76961_
Ecoli_yetQ_P76962_
Ecoli_yetR_P76963_
Ecoli_yetS_P76964_
Ecoli_yetT_P76965_
Ecoli_yetU_P76966_
Ecoli_yetV_P76967_
Ecoli_yetW_P76968_
Ecoli_yetX_P76969_
Ecoli_yetY_P76970_
Ecoli_yetZ_P76971_
Ecoli_yeuA_P76972_
Ecoli_yeuB_P76973_
Ecoli_yeuC_P76974_
Ecoli_yeuD_P76975_
Ecoli_yeuE_P76976_
Ecoli_yeuF_P76977_
Ecoli_yeuG_P76978_
Ecoli_yeuH_P76979_
Ecoli_yeuI_P76980_
Ecoli_yeuJ_P76981_
Ecoli_yeuK_P76982_
Ecoli_yeuL_P76983_
Ecoli_yeuM_P76984_
Ecoli_yeuN_P76985_
Ecoli_yeuO_P76986_
Ecoli_yeuP_P76987_
Ecoli_yeuQ_P76988_
Ecoli_yeuR_P76989_
Ecoli_yeuS_P76990_
Ecoli_yeuT_P76991_
Ecoli_yeuU_P76992_
Ecoli_yeuV_P76993_
Ecoli_yeuW_P76994_
Ecoli_yeuX_P76995_
Ecoli_yeuY_P76996_
Ecoli_yeuZ_P76997_
Ecoli_yevA_P76998_
Ecoli_yevB_P76999_
Ecoli_yevC_P77000_
Ecoli_yevD_P77001_
Ecoli_yevE_P77002_
Ecoli_yevF_P77003_
Ecoli_yevG_P77004_
Ecoli_yevH_P77005_
Ecoli_yevI_P77006_
Ecoli_yevJ_P77007_
Ecoli_yevK_P77008_
Ecoli_yevL_P77009_
Ecoli_yevM_P77010_
Ecoli_yevN_P77011_
Ecoli_yevO_P77012_
Ecoli_yevP_P77013_
Ecoli_yevQ_P77014_
Ecoli_yevR_P77015_
Ecoli_yevS_P77016_
Ecoli_yevT_P77017_
Ecoli_yevU_P77018_
Ecoli_yevV_P77019_
Ecoli_yevW_P77020_
Ecoli_yevX_P77021_
Ecoli_yevY_P77022_
Ecoli_yevZ_P77023_
Ecoli_yewA_P77024_
Ecoli_yewB_P77025_
Ecoli_yewC_P77026_
Ecoli_yewD_P77027_
Ecoli_yewE_P77028_
Ecoli_yewF_P77029_
Ecoli_yewG_P77030_
Ecoli_yewH_P77031_
Ecoli_yewI_P77032_
Ecoli_yewJ_P77033_
Ecoli_yewK_P77034_
Ecoli_yewL_P77035_
Ecoli_yewM_P77036_
Ecoli_yewN_P77037_
Ecoli_yewO_P77038_
Ecoli_yewP_P77039_
Ecoli_yewQ_P77040_
Ecoli_yewR_P77041_
Ecoli_yewS_P77042_
Ecoli_yewT_P77043_
Ecoli_yewU_P77044_
Ecoli_yewV_P77045_
Ecoli_yewW_P77046_
Ecoli_yewX_P77047_
Ecoli_yewY_P77048_
Ecoli_yewZ_P77049_
Ecoli_yexA_P77050_
Ecoli_yexB_P77051_
Ecoli_yexC_P77052_
Ecoli_yexD_P77053_
Ecoli_yexE_P77054_
Ecoli_yexF_P77055_
Ecoli_yexG_P77056_
Ecoli_yexH_P77057_
Ecoli_yexI_P77058_
Ecoli_yexJ_P77059_
Ecoli_yexK_P77060_
Ecoli_yexL_P77061_
Ecoli_yexM_P77062_
Ecoli_yexN_P77063_
Ecoli_yexO_P77064_
Ecoli_yexP_P77065_
Ecoli_yexQ_P77066_
Ecoli_yexR_P77067_
Ecoli_yexS_P77068_
Ecoli_yexT_P77069_
Ecoli_yexU_P77070_
Ecoli_yexV_P77071_
Ecoli_yexW_P77072_
Ecoli_yexX_P77073_
Ecoli_yexY_P77074_
Ecoli_yexZ_P77075_
Ecoli_yeyA_P77076_
Ecoli_yeyB_P77077_
Ecoli_yeyC_P77078_
Ecoli_yeyD_P77079_
Ecoli_yeyE_P77080_
Ecoli_yeyF_P77081_
Ecoli_yeyG_P77082_
Ecoli_yeyH_P77083_
Ecoli_yeyI_P77084_
Ecoli_yeyJ_P77085_
Ecoli_yeyK_P77086_
Ecoli_yeyL_P77087_
Ecoli_yeyM_P77088_
Ecoli_yeyN_P77089_
Ecoli_yeyO_P77090_
Ecoli_yeyP_P77091_
Ecoli_yeyQ_P77092_
Ecoli_yeyR_P77093_
Ecoli_yeyS_P77094_
Ecoli_yeyT_P77095_
Ecoli_yeyU_P77096_
Ecoli_yeyV_P77097_
Ecoli_yeyW_P77098_
Ecoli_yeyX_P77099_
Ecoli_yeyY_P77100_
Ecoli_yeyZ_P77101_
Ecoli_yezA_P77102_
Ecoli_yezB_P77103_
Ecoli_yezC_P77104_
Ecoli_yezD_P77105_
Ecoli_yezE_P77106_
Ecoli_yezF_P77107_
Ecoli_yezG_P77108_
Ecoli_yezH_P77109_
Ecoli_yezI_P77110_
Ecoli_yezJ_P77111_
Ecoli_yezK_P77112_
Ecoli_yezL_P77113_
Ecoli_yezM_P77114_
Ecoli_yezN_P77115_
Ecoli_yezO_P77116_
Ecoli_yezP_P77117_
Ecoli_yezQ_P77118_
Ecoli_yezR_P77119_
Ecoli_yezS_P77120_
Ecoli_yezT_P77121_
Ecoli_yezU_P77122_
Ecoli_yezV_P77123_
Ecoli_yezW_P77124_
Ecoli_yezX_P77125_
Ecoli_yezY_P77126_
Ecoli_yezZ_P77127_
Ecoli_yfaA_P77128_
Ecoli_yfaB_P77129_
Ecoli_yfaC_P77130_
Ecoli_yfaD_P77131_
Ecoli_yfaE_P77132_
Ecoli_yfaF_P77133_
Ecoli_yfaG_P77134_
Ecoli_yfaH_P77135_
Ecoli_yfaI_P77136_
Ecoli_yfaJ_P77137_
Ecoli_yfaK_P77138_
Ecoli_yfaL_P77139_
Ecoli_yfaM_P77140_
Ecoli_yfaN_P77141_
Ecoli_yfaO_P77142_
Ecoli_yfaP_P77143_
Ecoli_yfaQ_P77144_
Ecoli_yfaR_P77145_
Ecoli_yfaS_P77146_
Ecoli_yfaT_P77147_
Ecoli_yfaU_P77148_
Ecoli_yfaV_P77149_
Ecoli_yfaW_P77150_
Ecoli_yfaX_P77151_
Ecoli_yfaY_P77152_
Ecoli_yfaZ_P77153_
Ecoli_yfbA_P77154_
Ecoli_yfbB_P77155_
Ecoli_yfbC_P77156_
Ecoli_yfbD_P77157_
Ecoli_yfbE_P77158_
Ecoli_yfbF_P77159_
Ecoli_yfbG_P77160_
Ecoli_yfbH_P77161_
Ecoli_yfbI_P77162_
Ecoli_yfbJ_P77163_
Ecoli_yfbK_P77164_
Ecoli_yfbL_P77165_
Ecoli_yfbM_P77166_
Ecoli_yfbN_P77167_
Ecoli_yfbO_P77168_
Ecoli_yfbP_P77169_
Ecoli_yfbQ_P77170_
Ecoli_yfbR_P77171_
Ecoli_yfbS_P77172_
Ecoli_yfbT_P77173_
Ecoli_yfbU_P77174_
Ecoli_yfbV_P77175_
Ecoli_yfbW_P77176_
Ecoli_yfbX_P77177_
Ecoli_yfbY_P77178_
Ecoli_yfbZ_P77179_
Ecoli_yfcA_P77180_
Ecoli_yfcB_P77181_
Ecoli_yfcC_P77182_
Ecoli_yfcD_P77183_
Ecoli_yfcE_P77184_
Ecoli_yfcF_P77185_
Ecoli_yfcG_P77186_
Ecoli_yfcH_P77187_
Ecoli_yfcI_P77188_
Ecoli_yfcJ_P77189_
Ecoli_yfcK_P77190_
Ecoli_yfcL_P77191_
Ecoli_yfcM_P77192_
Ecoli_yfcN_P77193_
Ecoli_yfcO_P77194_
Ecoli_yfcP_P77195_
Ecoli_yfcQ_P77196_
Ecoli_yfcR_P77197_
Ecoli_yfcS_P77198_
Ecoli_yfcT_P77199_
Ecoli_yfcU_P77200_
Ecoli_yfcV_P77201_
Ecoli_yfcW_P77202_
Ecoli_yfcX_P77203_
Ecoli_yfcY_P77204_
Ecoli_yfcZ_P77205_
Ecoli_yfdA_P77206_
Ecoli_yfdB_P77207_
Ecoli_yfdC_P77208_
Ecoli_yfdD_P77209_
Ecoli_yfdE_P77210_
Ecoli_yfdF_P77211_
Ecoli_yfdG_P77212_
Ecoli_yfdH_P77213_
Ecoli_yfdI_P77214_
Ecoli_yfdJ_P77215_
Ecoli_yfdK_P77216_
Ecoli_yfdL_P77217_
Ecoli_yfdM_P77218_
Ecoli_yfdN_P77219_
Ecoli_yfdO_P77220_
Ecoli_yfdP_P77221_
Ecoli_yfdQ_P77222_
Ecoli_yfdR_P77223_
Ecoli_yfdS_P77224_
Ecoli_yfdT_P77225_
Ecoli_yfdU_P77226_
Ecoli_yfdV_P77227_
Ecoli_yfdW_P77228_
Ecoli_yfdX_P77229_
Ecoli_yfdY_P77230_
Ecoli_yfdZ_P77231_
Ecoli_yfeA_P77232_
Ecoli_yfeB_P77233_
Ecoli_yfeC_P77234_
Ecoli_yfeD_P77235_
Ecoli_yfeE_P77236_
Ecoli_yfeF_P77237_
Ecoli_yfeG_P77238_
Ecoli_yfeH_P77239_
Ecoli_yfeI_P77240_
Ecoli_yfeJ_P77241_
Ecoli_yfeK_P77242_
Ecoli_yfeL_P77243_
Ecoli_yfeM_P77244_
Ecoli_yfeN_P77245_
Ecoli_yfeO_P77246_
Ecoli_yfeP_P77247_
Ecoli_yfeQ_P77248_
Ecoli_yfeR_P77249_
Ecoli_yfeS_P77250_
Ecoli_yfeT_P77251_
Ecoli_yfeU_P77252_
Ecoli_yfeV_P77253_
Ecoli_yfeW_P77254_
Ecoli_yfeX_P77255_
Ecoli_yfeY_P77256_
Ecoli_yfeZ_P77257_
Ecoli_yffA_P77258_
Ecoli_yffB_P77259_
Ecoli_yffC_P77260_
Ecoli_yffD_P77261_
Ecoli_yffE_P77262_
Ecoli_yffF_P77263_
Ecoli_yffG_P77264_
Ecoli_yffH_P77265_
Ecoli_yffI_P77266_
Ecoli_yffJ_P77267_
Ecoli_yffK_P77268_
Ecoli_yffL_P77269_
Ecoli_yffM_P77270_
Ecoli_yffN_P77271_
Ecoli_yffO_P77272_
Ecoli_yffP_P77273_
Ecoli_yffQ_P77274_
Ecoli_yffR_P77275_
Ecoli_yffS_P77276_
Ecoli_yffT_P77277_
Ecoli_yffU_P77278_
Ecoli_yffV_P77279_
Ecoli_yffW_P77280_
Ecoli_yffX_P77281_
Ecoli_yffY_P77282_
Ecoli_yffZ_P77283_
Ecoli_yfgA_P77284_
Ecoli_yfgB_P77285_
Ecoli_yfgC_P77286_
Ecoli_yfgD_P77287_
Ecoli_yfgE_P77288_
Ecoli_yfgF_P77289_
Ecoli_yfgG_P77290_
Ecoli_yfgH_P77291_
Ecoli_yfgI_P77292_
Ecoli_yfgJ_P77293_
Ecoli_yfgK_P77294_
Ecoli_yfgL_P77295_
Ecoli_yfgM_P77296_
Ecoli_yfgN_P77297_
Ecoli_yfgO_P77298_
Ecoli_yfgP_P77299_
Ecoli_yfgQ_P77300_
Ecoli_yfgR_P77301_
Ecoli_yfgS_P77302_
Ecoli_yfgT_P77303_
Ecoli_yfgU_P77304_
Ecoli_yfgV_P77305_
Ecoli_yfgW_P77306_
Ecoli_yfgX_P77307_
Ecoli_yfgY_P77308_
Ecoli_yfgZ_P77309_
Ecoli_yfhA_P77310_
Ecoli_yfhB_P77311_
Ecoli_yfhC_P77312_
Ecoli_yfhD_P77313_
Ecoli_yfhE_P77314_
Ecoli_yfhF_P77315_
Ecoli_yfhG_P77316_
Ecoli_yfhH_P77317_
Ecoli_yfhI_P77318_
Ecoli_yfhJ_P77319_
Ecoli_yfhK_P77320_
Ecoli_yfhL_P77321_
Ecoli_yfhM_P77322_
Ecoli_yfhN_P77323_
Ecoli_yfhO_P77324_
Ecoli_yfhP_P77325_
Ecoli_yfhQ_P77326_
Ecoli_yfhR_P77327_
Ecoli_yfhS_P77328_
Ecoli_yfhT_P77329_
Ecoli_yfhU_P77330_
Ecoli_yfhV_P77331_
Ecoli_yfhW_P77332_
Ecoli_yfhX_P77333_
Ecoli_yfhY_P77334_
Ecoli_yfhZ_P77335_
Ecoli_yfiA_P77336_
Ecoli_yfiB_P77337_
Ecoli_yfiC_P77338_
Ecoli_yfiD_P77339_
Ecoli_yfiE_P77340_
Ecoli_yfiF_P77341_
Ecoli_yfiG_P77342_
Ecoli_yfiH_P77343_
Ecoli_yfiI_P77344_
Ecoli_yfiJ_P77345_
Ecoli_yfiK_P77346_
Ecoli_yfiL_P77347_
Ecoli_yfiM_P77348_
Ecoli_yfiN_P77349_
Ecoli_yfiO_P77350_
Ecoli_yfiP_P77351_
Ecoli_yfiQ_P77352_
Ecoli_yfiR_P77353_
Ecoli_yfiS_P77354_
Ecoli_yfiT_P77355_
Ecoli_yfiU_P77356_
Ecoli_yfiV_P77357_
Ecoli_yfiW_P77358_
Ecoli_yfiX_P77359_
Ecoli_yfiY_P77360_
Ecoli_yfiZ_P77361_
Ecoli_yfjA_P77362_
Ecoli_yfjB_P77363_
Ecoli_yfjC_P77364_
Ecoli_yfjD_P77365_
Ecoli_yfjE_P77366_
Ecoli_yfjF_P77367_
Ecoli_yfjG_P77368_
Ecoli_yfjH_P77369_
Ecoli_yfjI_P77370_
Ecoli_yfjJ_P77371_
Ecoli_yfjK_P77372_
Ecoli_yfjL_P77373_
Ecoli_yfjM_P77374_
Ecoli_yfjN_P77375_
Ecoli_yfjO_P77376_
Ecoli_yfjP_P77377_
Ecoli_yfjQ_P77378_
Ecoli_yfjR_P77379_
Ecoli_yfjS_P77380_
Ecoli_yfjT_P77381_
Ecoli_yfjU_P77382_
Ecoli_yfjV_P77383_
Ecoli_yfjW_P77384_
Ecoli_yfjX_P77385_
Ecoli_yfjY_P77386_
Ecoli_yfjZ_P77387_
Ecoli_yfkA_P77388_
Ecoli_yfkB_P77389_
Ecoli_yfkC_P77390_
Ecoli_yfkD_P77391_
Ecoli_yfkE_P77392_
Ecoli_yfkF_P77393_
Ecoli_yfkG_P77394_
Ecoli_yfkH_P77395_
Ecoli_yfkI_P77396_
Ecoli_yfkJ_P77397_
Ecoli_yfkK_P77398_
Ecoli_yfkL_P77399_
Ecoli_yfkM_P77400_
Ecoli_yfkN_P77401_
Ecoli_yfkO_P77402_
Ecoli_yfkP_P77

Maximum likelihood
matK gene tree. Constructed
using RAxML, with a GTRCAT
model of evolution and a 200
bootstrap analysis. Bootstrap
values are plotted on the tree.


Carex_macrocarpa_MOR_1633_STARR_01429_CV
Carex_cigillata_MOR_1936_B1429_CV
Carex_capillaris_MOR_1557_B1449_CT
Carex_willdenowii_MOR_358_KP_CV
Carex_pauciflora_MOR_876_AK_B536_LG
Carex_multicaulis_MOR_1084_B1450_CV
Carex_bolipitensis_MOR_1563_GE_B1154_LF
Carex_bolipitensis_MOR_1564_CA_B1076_FC
Carex_simulata_MOR_1565_AL_B1077_FC
Carex_turnerpilorum_MOR_921_CV_B511_PB
Carex_tumulicola_MOR_1566_CA_B1078_FC
Carex_supaltiensis_MOR_971_BC_B511_PB
Carex_esperata_MOR_809_FC_B515_PB
Carex_tumulicola_MOR_1567_CA_B1079_FC
Carex_backii_MOR_1569_B1493_PB
Carex_occidentalis_MOR_1570_AL_B1120_PB
Carex_orcuttiana_MOR_1348_BC_B843_PB
Carex_orcuttiana_MOR_1119_AK_B1075_CB
Carex_engelmannii_MOR_1571_BC_B1021_CF
Carex_forbesii_MOR_1572_BC_B1021_CF
Carex_olnyensis_MOR_1573_BC_B1203_PG
Carex_martinii_MOR_1574_BC_B1203_PG
Carex_mariposana_MOR_1012_WI_CA_B541_L
Carex_triquetra_MOR_124_CA_B635_CA
Carex_oreophila_AZ_1575_STARR_01429_CV
Carex_eleocharis_MOR_1576_BC_B1421_CT
Carex_eleocharis_MOR_1577_BC_B1421_CT
Carex_microgyna_MOR_1578_AZ_B1082_F
Carex_microgyna_MOR_1579_AZ_B1082_F
Carex_nigricans_MOR_1580_BC_B1021_CF
Carex_ecusta_MOR_1581_BC_B1021_CF
Carex_sparganioides_sparganioides_MOR_1507_B1098_LG
Carex_sparganioides_sparganioides_MOR_1508_WI_B684_P
Carex_gynocrates_MOR_439_MD_B087_MF
Carex_lime_MOR_1582_CA_B1080_FC
Carex_triangularis_MOR_1583_CA_B1081_FC
Carex_hoodii_MOR_1584_CA_B1082_FC
Carex_hoodii_MOR_1585_CA_B1083_FC
Carex_amnicola_MOR_464_IC_B1137_MF
Carex_vulpinoidea_MOR_1586_CA_B1084_FC
Carex_davisii_MOR_1587_CA_B1085_FC
Carex_deflata_MOR_572_CA_B1087_MF
Carex_davisii_MOR_1588_CA_B1086_FC
Carex_neurophysa_MOR_1477_WA_B1006_CV
Carex_alopecoidea_MOR_1589_BC_B1021_CF
Carex_nudicrista_MOR_1590_CA_B1088_FC
Carex_hookeriana_MOR_1570_AZ_B1087_PG
Carex_hoodii_MOR_1591_CA_B1089_FC
Carex_hoodii_MOR_1592_CA_B1090_FC
Carex_divulsa_MOR_1383_WA_B529_PG
Carex_spicata_MOR_1593_CA_B1091_FC
Carex_spicata_MOR_1070_CA_B600_CB
Carex_obnupta_MOR_1594_MD_B1121_PG
Carex_hassei_MOR_1595_CA_B1092_FC
Carex_hassei_MOR_1596_CA_B1093_FC
Carex_hassei_MOR_1597_CA_B1094_FC
Carex_hassei_MOR_1598_CA_B1095_FC
Carex_hassei_MOR_1599_CA_B1096_FC
Carex_hassei_MOR_1600_CA_B1097_FC
Carex_hassei_MOR_1601_CA_B1098_FC
Carex_hassei_MOR_1602_CA_B1099_FC
Carex_hassei_MOR_1603_CA_B1100_FC
Carex_hassei_MOR_1604_CA_B1101_FC
Carex_hassei_MOR_1605_CA_B1102_FC
Carex_hassei_MOR_1606_CA_B1103_FC
Carex_hassei_MOR_1607_CA_B1104_FC
Carex_hassei_MOR_1608_CA_B1105_FC
Carex_hassei_MOR_1609_CA_B1106_FC
Carex_hassei_MOR_1610_CA_B1107_FC
Carex_hassei_MOR_1611_CA_B1108_FC
Carex_hassei_MOR_1612_CA_B1109_FC
Carex_hassei_MOR_1613_CA_B1110_FC
Carex_hassei_MOR_1614_CA_B1111_FC
Carex_hassei_MOR_1615_CA_B1112_FC
Carex_hassei_MOR_1616_CA_B1113_FC
Carex_hassei_MOR_1617_CA_B1114_FC
Carex_hassei_MOR_1618_CA_B1115_FC
Carex_hassei_MOR_1619_CA_B1116_FC
Carex_hassei_MOR_1620_CA_B1117_FC
Carex_hassei_MOR_1621_CA_B1118_FC
Carex_hassei_MOR_1622_CA_B1119_FC
Carex_hassei_MOR_1623_CA_B1120_FC
Carex_hassei_MOR_1624_CA_B1121_FC
Carex_hassei_MOR_1625_CA_B1122_FC
Carex_hassei_MOR_1626_CA_B1123_FC
Carex_hassei_MOR_1627_CA_B1124_FC
Carex_hassei_MOR_1628_CA_B1125_FC
Carex_hassei_MOR_1629_CA_B1126_FC
Carex_hassei_MOR_1630_CA_B1127_FC
Carex_hassei_MOR_1631_CA_B1128_FC
Carex_hassei_MOR_1632_CA_B1129_FC
Carex_hassei_MOR_1633_CA_B1130_FC
Carex_hassei_MOR_1634_CA_B1131_FC
Carex_hassei_MOR_1635_CA_B1132_FC
Carex_hassei_MOR_1636_CA_B1133_FC
Carex_hassei_MOR_1637_CA_B1134_FC
Carex_hassei_MOR_1638_CA_B1135_FC
Carex_hassei_MOR_1639_CA_B1136_FC
Carex_hassei_MOR_1640_CA_B1137_FC
Carex_hassei_MOR_1641_CA_B1138_FC
Carex_hassei_MOR_1642_CA_B1139_FC
Carex_hassei_MOR_1643_CA_B1140_FC
Carex_hassei_MOR_1644_CA_B1141_FC
Carex_hassei_MOR_1645_CA_B1142_FC
Carex_hassei_MOR_1646_CA_B1143_FC
Carex_hassei_MOR_1647_CA_B1144_FC
Carex_hassei_MOR_1648_CA_B1145_FC
Carex_hassei_MOR_1649_CA_B1146_FC
Carex_hassei_MOR_1650_CA_B1147_FC
Carex_hassei_MOR_1651_CA_B1148_FC
Carex_hassei_MOR_1652_CA_B1149_FC
Carex_hassei_MOR_1653_CA_B1150_FC
Carex_hassei_MOR_1654_CA_B1151_FC
Carex_hassei_MOR_1655_CA_B1152_FC
Carex_hassei_MOR_1656_CA_B1153_FC
Carex_hassei_MOR_1657_CA_B1154_FC
Carex_hassei_MOR_1658_CA_B1155_FC
Carex_hassei_MOR_1659_CA_B1156_FC
Carex_hassei_MOR_1660_CA_B1157_FC
Carex_hassei_MOR_1661_CA_B1158_FC
Carex_hassei_MOR_1662_CA_B1159_FC
Carex_hassei_MOR_1663_CA_B1160_FC
Carex_hassei_MOR_1664_CA_B1161_FC
Carex_hassei_MOR_1665_CA_B1162_FC
Carex_hassei_MOR_1666_CA_B1163_FC
Carex_hassei_MOR_1667_CA_B1164_FC
Carex_hassei_MOR_1668_CA_B1165_FC
Carex_hassei_MOR_1669_CA_B1166_FC
Carex_hassei_MOR_1670_CA_B1167_FC
Carex_hassei_MOR_1671_CA_B1168_FC
Carex_hassei_MOR_1672_CA_B1169_FC
Carex_hassei_MOR_1673_CA_B1170_FC
Carex_hassei_MOR_1674_CA_B1171_FC
Carex_hassei_MOR_1675_CA_B1172_FC
Carex_hassei_MOR_1676_CA_B1173_FC
Carex_hassei_MOR_1677_CA_B1174_FC
Carex_hassei_MOR_1678_CA_B1175_FC
Carex_hassei_MOR_1679_CA_B1176_FC
Carex_hassei_MOR_1680_CA_B1177_FC
Carex_hassei_MOR_1681_CA_B1178_FC
Carex_hassei_MOR_1682_CA_B1179_FC
Carex_hassei_MOR_1683_CA_B1180_FC
Carex_hassei_MOR_1684_CA_B1181_FC
Carex_hassei_MOR_1685_CA_B1182_FC
Carex_hassei_MOR_1686_CA_B1183_FC
Carex_hassei_MOR_1687_CA_B1184_FC
Carex_hassei_MOR_1688_CA_B1185_FC
Carex_hassei_MOR_1689_CA_B1186_FC
Carex_hassei_MOR_1690_CA_B1187_FC
Carex_hassei_MOR_1691_CA_B1188_FC
Carex_hassei_MOR_1692_CA_B1189_FC
Carex_hassei_MOR_1693_CA_B1190_FC
Carex_hassei_MOR_1694_CA_B1191_FC
Carex_hassei_MOR_1695_CA_B1192_FC
Carex_hassei_MOR_1696_CA_B1193_FC
Carex_hassei_MOR_1697_CA_B1194_FC
Carex_hassei_MOR_1698_CA_B1195_FC
Carex_hassei_MOR_1699_CA_B1196_FC
Carex_hassei_MOR_1700_CA_B1197_FC
Carex_hassei_MOR_1701_CA_B1198_FC
Carex_hassei_MOR_1702_CA_B1199_FC
Carex_hassei_MOR_1703_CA_B1200_FC
Carex_hassei_MOR_1704_CA_B1201_FC
Carex_hassei_MOR_1705_CA_B1202_FC
Carex_hassei_MOR_1706_CA_B1203_FC
Carex_hassei_MOR_1707_CA_B1204_FC
Carex_hassei_MOR_1708_CA_B1205_FC
Carex_hassei_MOR_1709_CA_B1206_FC
Carex_hassei_MOR_1710_CA_B1207_FC
Carex_hassei_MOR_1711_CA_B1208_FC
Carex_hassei_MOR_1712_CA_B1209_FC
Carex_hassei_MOR_1713_CA_B1210_FC
Carex_hassei_MOR_1714_CA_B1211_FC
Carex_hassei_MOR_1715_CA_B1212_FC
Carex_hassei_MOR_1716_CA_B1213_FC
Carex_hassei_MOR_1717_CA_B1214_FC
Carex_hassei_MOR_1718_CA_B1215_FC
Carex_hassei_MOR_1719_CA_B1216_FC
Carex_hassei_MOR_1720_CA_B1217_FC
Carex_hassei_MOR_1721_CA_B1218_FC
Carex_hassei_MOR_1722_CA_B1219_FC
Carex_hassei_MOR_1723_CA_B1220_FC
Carex_hassei_MOR_1724_CA_B1221_FC
Carex_hassei_MOR_1725_CA_B1222_FC
Carex_hassei_MOR_1726_CA_B1223_FC
Carex_hassei_MOR_1727_CA_B1224_FC
Carex_hassei_MOR_1728_CA_B1225_FC
Carex_hassei_MOR_1729_CA_B1226_FC
Carex_hassei_MOR_1730_CA_B1227_FC
Carex_hassei_MOR_1731_CA_B1228_FC
Carex_hassei_MOR_1732_CA_B1229_FC
Carex_hassei_MOR_1733_CA_B1230_FC
Carex_hassei_MOR_1734_CA_B1231_FC
Carex_hassei_MOR_1735_CA_B1232_FC
Carex_hassei_MOR_1736_CA_B1233_FC
Carex_hassei_MOR_1737_CA_B1234_FC
Carex_hassei_MOR_1738_CA_B1235_FC
Carex_hassei_MOR_1739_CA_B1236_FC
Carex_hassei_MOR_1740_CA_B1237_FC
Carex_hassei_MOR_1741_CA_B1238_FC
Carex_hassei_MOR_1742_CA_B1239_FC
Carex_hassei_MOR_1743_CA_B1240_FC
Carex_hassei_MOR_1744_CA_B1241_FC
Carex_hassei_MOR_1745_CA_B1242_FC
Carex_hassei_MOR_1746_CA_B1243_FC
Carex_hassei_MOR_1747_CA_B1244_FC
Carex_hassei_MOR_1748_CA_B1245_FC
Carex_hassei_MOR_1749_CA_B1246_FC
Carex_hassei_MOR_1750_CA_B1247_FC
Carex_hassei_MOR_1751_CA_B1248_FC
Carex_hassei_MOR_1752_CA_B1249_FC
Carex_hassei_MOR_1753_CA_B1250_FC
Carex_hassei_MOR_1754_CA_B1251_FC
Carex_hassei_MOR_1755_CA_B1252_FC
Carex_hassei_MOR_1756_CA_B1253_FC
Carex_hassei_MOR_1757_CA_B1254_FC
Carex_hassei_MOR_1758_CA_B1255_FC
Carex_hassei_MOR_1759_CA_B1256_FC
Carex_hassei_MOR_1760_CA_B1257_FC
Carex_hassei_MOR_1761_CA_B1258_FC
Carex_hassei_MOR_1762_CA_B1259_FC
Carex_hassei_MOR_1763_CA_B1260_FC
Carex_hassei_MOR_1764_CA_B1261_FC
Carex_hassei_MOR_1765_CA_B1262_FC
Carex_hassei_MOR_1766_CA_B1263_FC
Carex_hassei_MOR_1767_CA_B1264_FC
Carex_hassei_MOR_1768_CA_B1265_FC
Carex_hassei_MOR_1769_CA_B1266_FC
Carex_hassei_MOR_1770_CA_B1267_FC
Carex_hassei_MOR_1771_CA_B1268_FC
Carex_hassei_MOR_1772_CA_B1269_FC
Carex_hassei_MOR_1773_CA_B1270_FC
Carex_hassei_MOR_1774_CA_B1271_FC
Carex_hassei_MOR_1775_CA_B1272_FC
Carex_hassei_MOR_1776_CA_B1273_FC
Carex_hassei_MOR_1777_CA_B1274_FC
Carex_hassei_MOR_1778_CA_B1275_FC
Carex_hassei_MOR_1779_CA_B1276_FC
Carex_hassei_MOR_1780_CA_B1277_FC
Carex_hassei_MOR_1781_CA_B1278_FC
Carex_hassei_MOR_1782_CA_B1279_FC
Carex_hassei_MOR_1783_CA_B1280_FC
Carex_hassei_MOR_1784_CA_B1281_FC
Carex_hassei_MOR_1785_CA_B1282_FC
Carex_hassei_MOR_1786_CA_B1283_FC
Carex_hassei_MOR_1787_CA_B1284_FC
Carex_hassei_MOR_1788_CA_B1285_FC
Carex_hassei_MOR_1789_CA_B1286_FC
Carex_hassei_MOR_1790_CA_B1287_FC
Carex_hassei_MOR_1791_CA_B1288_FC
Carex_hassei_MOR_1792_CA_B1289_FC
Carex_hassei_MOR_1793_CA_B1290_FC
Carex_hassei_MOR_1794_CA_B1291_FC
Carex_hassei_MOR_1795_CA_B1292_FC
Carex_hassei_MOR_1796_CA_B1293_FC
Carex_hassei_MOR_1797_CA_B1294_FC
Carex_hassei_MOR_1798_CA_B1295_FC
Carex_hassei_MOR_1799_CA_B1296_FC
Carex_hassei_MOR_1800_CA_B1297_FC
Carex_hassei_MOR_1801_CA_B1298_FC
Carex_hassei_MOR_1802_CA_B1299_FC
Carex_hassei_MOR_1803_CA_B1300_FC
Carex_hassei_MOR_1804_CA_B1301_FC
Carex_hassei_MOR_1805_CA_B1302_FC
Carex_hassei_MOR_1806_CA_B1303_FC
Carex_hassei_MOR_1807_CA_B1304_FC
Carex_hassei_MOR_1808_CA_B1305_FC
Carex_hassei_MOR_1809_CA_B1306_FC
Carex_hassei_MOR_1810_CA_B1307_FC
Carex_hassei_MOR_1811_CA_B1308_FC
Carex_hassei_MOR_1812_CA_B1309_FC
Carex_hassei_MOR_1813_CA_B1310_FC
Carex_hassei_MOR_1814_CA_B1311_FC
Carex_hassei_MOR_1815_AZ_B1308_CV
Carex_woodrowii_MOR_1816_AZ_B1309_CV
Carex_cudioki_MOR_1555_BC_B1397_HE
Carex_cudioki_MOR_1556_BC_B1398_HE
Carex_digitata_MOR_1817_CA_B1222_HE
Carex_pyrenaica_pyrenaica_MOR_1086_CA_B878_DV
Carex_pyrenaica_MOR_1087_CA_B879_DV
Carex_pyrenaica_MOR_1088_CA_B880_DV
Carex_furtiensis_MOR_1121_AZ_B1042_AC
Carex_brunnescens_MOR_1122_AK_B1043_AC
Carex_brainerdii_MOR_1124_CA_B1044_AC
Carex_deflexa_bulonis_MOR_1125_AK_B1045_AC
Carex_deflexa_deflexa_MOR_1126_AK_B1046_AC
Carex_hassei_MOR_1128_BC_B1218_AC
Carex_circinata_MOR_1129_AK_B1047_AC
Carex_circinata_circinata_MOR_1130_AK_B1048_AC
Carex_circinata_MOR_1131_AK_B1049_AC
Carex_lutkensii_MOR_1132_AK_B1050_AC
Carex_lutkensii_MOR_1133_AK_B1051_AC
Carex_lutkensii_MOR_1134_AK_B1052_AC
Carex_lutkensii_MOR_1135_AK_B1053_AC
Carex_lutkensii_MOR_1136_AK_B1054_AC
Carex_lutkensii_MOR_1137_AK_B1055_AC
Carex_lutkensii_MOR_1138_AK_B1056_AC
Carex_lutkensii_MOR_1139_AK_B1057_AC
Carex_lutkensii_MOR_1140_AK_B1058_AC
Carex_lutkensii_MOR_1141_AK_B1059_AC
Carex_lutkensii_MOR_1142_AK_B1060_AC
Carex_lutkensii_MOR_1143_AK_B1061_AC
Carex_lutkensii_MOR_1144_AK_B1062_AC
Carex_lutkensii_MOR_1145_AK_B1063_AC
Carex_lutkensii_MOR_1146_AK_B1064_AC
Carex_lutkensii_MOR_1147_AK_B1065_AC
Carex_lutkensii_MOR_1148_AK_B1066_AC
Carex_lutkensii_MOR_1149_AK_B1067_AC
Carex_lutkensii_MOR_1150_AK_B1068_AC
Carex_lutkensii_MOR_1151_AK_B1069_AC
Carex_lutkensii_MOR_1152_AK_B1070_AC
Carex_lutkensii_MOR_1153_AK_B1071_AC
Carex_lutkensii_MOR_1154_AK_B1072_AC
Carex_lutkensii_MOR_1155_AK_B1073_AC
Carex_lutkensii_MOR_1156_AK_B1074_AC
Carex_lutkensii_MOR_1157_AK_B1075_AC
Carex_lutkensii_MOR_1158_AK_B1076_AC
Carex_lutkensii_MOR_1159_AK_B1077_AC
Carex_lutkensii_MOR_1160_AK_B1078_AC
Carex_lutkensii_MOR_1161_AK_B1079_AC
Carex_lutkensii_MOR_1162_AK_B1080_AC
Carex_lutkensii_MOR_1163_AK_B1081_AC
Carex_lutkensii_MOR_1164_AK_B1082_AC
Carex_lutkensii_MOR_1165_AK_B1083_AC
Carex_lutkensii_MOR_1166_AK_B1084_AC
Carex_lutkensii_MOR_1167_AK_B1085_AC
Carex_lutkensii_MOR_1168_AK_B1086_AC
Carex_lutkensii_MOR_1169_AK_B1087_AC
Carex_lutkensii_MOR_1170_AK_B1088_AC
Carex_lutkensii_MOR_1171_AK_B1089_AC

Maximum likelihood rps16 gene tree. Constructed using RAxML, with a GTRCAT model of evolution and a 200 bootstrap analysis. Bootstrap values are plotted on the tree.