PHYLOGEOGRAPHY AND POPULATION GENETICS OF *CAREX MACROCEPHALA,* AND THE MOLECULAR EVOLUTION OF *CAREX* SUBGENUS *VIGNEA*.

By

MATTHEW GEORGE KING

A dissertation/thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

DECEMBER 2007

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation/thesis of MATTHEW GEORGE KING find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

ACKNOWLEDGMENTS

I want to thank Marisa Olson, for all of her support. I would like to acknowledge everyone who has help and contributed to my success as a graduate student. This includes John Clark, Matthew Horning, Larry Hufford, Chuck Cody, Glenn Miller, Lynn Kintner, Josh Brokaw, Marlowe, Joy Mastroguiseppe, and the Web-schwa-stor'f'ords. I would like to thank Timothy Carey, Danielle Davis, and Kristina Bearden for lab assistance. Additionally, I would like to thank Andrew and Becky Friske for transportation throughout the islands of southeast Alaska for collections. I also thank Erica Wheeler, Chris Kissinger (Ministry of Environment, British Columbia), John McIntosh (Pacific Rim National Park), Robert DeVelice (Chugach National Forest), Mary Stensvold (Tongass National Forest), Jay Schleier (Oregon Department of Natural Resources), and Deaydra Wise (Washington State Parks) for help with collection permits and transportation of collections. I would also like to thank staff from the Rancho Santa Ana Botanical Garden Herbarium [RSA], Arnold Arboretum Herbarium at Harvard [A], Missouri Botanical Garden Herbarium [MO], and Marion Ownbey Herbarium at Washington State University [WS]. This work was supported by funding from the Native Plant Society of Oregon, Betty Higinbotham Research Award, Noe Higinbotham Award, Hardman Native Plant Research Award, and WSU School of Biological Sciences. I also thank Karol Marlowe and all reviewers for comments on versions of this manuscript.

PHYLOGEOGRAPHY AND POPULATION GENETICS OF *CAREX MACROCEPHALA,* AND THE MOLECULAR EVOLUTION OF *CAREX* SUBGENUS *VIGNEA.*

Abstract

by Matthew G. King, Ph.D.
Washington State University
December 2007

Chair: Eric H. Roalson

Nuclear ribosomal DNA (nrDNA) has been used for more than a decade in species level phylogenetic analyses. While nrDNA can often be a powerful phylogenetic marker, intra-individual polymorphisms of the internal and external transcribed spacers (ITS, ETS) can lead to problems in their use for phylogeny reconstruction. Through comprehensive cloning we identified high levels of intra-individual polymorphisms and in many cases this led to the polyphyly of individuals of *Carex* subgenus *Vignea*. We suggest that nrDNA contains multiple paralogs in many species and clades within *Vignea* which greatly complicates its use for phylogenetic inference and future studies in *Carex* need to take this into account.

A drawback to phylogenetic based phylogeographic analyses is that they do not account for stochastic lineage sorting that occurs between gene divergence and lineage divergence. Gene divergence (D) begins prior to the split of the lineages (t), and the difference between D and t are not accounted for in phylogenetic analyses. In contrast, coalescent-based statistical methods have been developed which can account for the stochastic forces which drive population divergence, and can account for the lineage

sorting that occurs prior to lineage divergence. We fit a molecular dataset, consisting of the *rpL16* marker and 8 microsatellite loci, to the isolation with migration model as implemented in IMa to test the well-supported phylogenetic hypothesis of relationships within the *Carex macrocephala* species complex (Cyperaceae). By comparing the relative divergence time of the three main lineages within this group, Asian *C. macrocephala*, North American *C. macrocephala*, and *C. kobomugi*, we concluded the phylogenetic hypothesis is incorrect, and the divergence between these lineages occurred during the late Pleistocene epoch.

Population genetic analyses of the *rpL16* marker and 11 microsatellite loci suggest a high level of inbreeding, but also high levels of migration across the west coast of North America (NA). The standardized Gst value is low at 0.032, and AMOVA results show significant amount of variation across all grouping levels. A principal coordinate analysis (PCoA) suggests panmixia across the NA coast, however the high levels of inbreeding suggest this may be an artifact of metapopulation dynamics.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Dedication

I would like to dedicate this dissertation to my advisor, Eric Roalson.
Without his daily support, it would have been difficult to finish,
and it would not have been enjoyable.

CHAPTER 1

EXPLORING EVOLUTIONARY DYNAMICS OF NRDNA IN *CAREX* SUBGENUS *VIGNEA* (CYPERACEAE)

The nuclear ribosomal DNA (nrDNA) internal and external transcribed spacer regions (ITS and ETS) have been a popular source of molecular datasets for species level phylogenetic reconstruction of angiosperm lineages (Baldwin et al. 1995; Hershkovitz et al. 1999; Linder et al. 2000; Álvarez and Wendel 2003; Bailey et al. 2003). Most often used are the first and second internal transcribed spacers (ITS1 and ITS2) and the intervening 5.8S ribosomal subunit, combined to form the entire ITS sequence region (Baldwin 1992; Álvarez and Wendel 2003). Although angiosperm universal primers are not available for the 5′ end of ETS, it is often utilized for phylogenetic research as well (Baldwin and Markos 1998; Linder et al. 2000; Starr et al. 2003). ITS and ETS are often used for resolving species level phylogenies, and have the added benefit of being readily amplifiable from herbarium material.

Nuclear ribosomal DNA is arranged into multiple tandem repeats of the 26S subunit, ITS1, 5.8S subunit, ITS2, 18S subunit, and the large intergenic spacer (IGS) which contains the 5’ and 3’ ends of ETS (Fig. 1). The entire region is repeated hundreds to thousands of times into multi-gene arrays of the nucleolar organizer regions (NOR; Hamby and Zimmer 1992; Baldwin et al. 1995; Cronn et al. 1996; Wendel et al. 1995a; Bailey et al. 2003). It is assumed for the purposes of species level phylogenetic reconstructions that all the copies of nrDNA within an individual concertedly evolve, thereby maintaining each repeat as an identical copy (Hillis et al. 1991; Small et al. 2004). Concerted evolution is accomplished through gene conversion and unequal crossing-over of the NOR’s, and several studies have suggested that concerted evolution is complete or near complete in many lineages where polymorphism within an individual is rare (Zimmer et al. 1980; Hillis et al. 1991; Baldwin et al. 1995; Fuertes-Aguilar et al.

1999; Linder et al. 2000). However, several studies have found paralogous or nonfunctional (pseudogene) copies sometimes coexisting within individuals. In some cases this leads to polyphyly of individuals and species when these markers are used for phylogeny reconstruction (Buckler et al. 1997; Campbell et al. 1997; Gernandt and Liston 1999; Yang et al. 1999; Denduanboripant and Cronk 2000; Kita and Ito 2000; Hartmann et al. 2001; Mayol and Rosselló 2001; Muir et al. 2001; Hughes et al. 2002; Li et al. 2004; Razafimandimbison et al. 2004; Ruggiero and Procaccin 2004; Wei and Wang 2004).

Persistence of non-homogenized paralogous copies within individuals or species reduces the utility of nrDNA spacers in species level phylogenetic analyses because it can lead to incorrect inferences of relationships, particularly if only one of the paralogs is sequenced, as is common in many ITS phylogenetic studies (Sanderson and Doyle 1992; Doyle and Davis 1998; Bailey et al. 2003; Small et al. 2004). If incomplete concerted evolution occurs, this can compound problems caused by evolutionary dynamics of DNA sequences, such as incomplete lineage sorting, hybridization, and recombination (chimerization) within the spacer region, and these processes create difficulties for accurate inference of phylogenetic relationships (McDade 1992; Sanderson and Doyle 1992; Wendel et al. 1995b; Buckler et al. 1997). The persistence of paralogous copies may naturally lead to the existence of non-functional pseudogenes. nrDNA pseudogenes have been amplified and used in phylogenetic reconstruction of Naucleeae (Rubiaceae; Razafimandimbison et al. 2004). Bailey et al. (2003) review methods of detecting nrDNA pseudogenes, and they discuss the function of using these copies to determine the level of polymorphism within an individual. They also propose phylogenetic and statistical

methods of determining whether the sequences within an analysis belong to a pseudogene based on a weighted ratio of substitution rates of the ITS sequence and 5.8S subunit along a given branch. This method makes use of non-parametric bootstrapping to establish the robustness of each statistic used in the calculations.

Carex (Cyperaceae) comprises approximately 2000 species and is the largest genus in the sedge family (Reznicek 1990). Previous molecular phylogenetic studies suggest the genus is not monophyletic (Starr et al. 1999; Yen and Olmstead 2000; Roalson et al. 2001; Starr et al. 2004). *Carex* subgenus *Vignea* includes ca. 300 species, and is characterized by a false-sutured peryginium that is typically lenticular, non-sheathing or absent subtending inflorescence bracts, and usually two (sometimes three) stigmas (Reznicek 1990). As supported by previous molecular studies, *Vignea* is considered to be one of the best supported clades within *Carex* (Yen and Olmstead 2000; Roalson et al. 2001).

Vignea is often divided into as many as 28 sections (Ford et al. 2006), and previous phylogenetic analyses have suggested many of these are not monophyletic (Hendrichs et al. 2004; Starr et al. 2004; Ford et al. 2006). There is some question, however, as to whether the lack of monophyly of *Vignea* sections is due to poor taxonomy or problems with using nrDNA in reconstructing relationships in the group. Ford et al. (2006) suggested the presence of paralogs within two species of *Carex* subgenus *Vignea* when betaine was not included in PCR amplification. Ford et al. noted that these sequences deviated in G-C percentage from the other sequences of *Carex* as well as the presence of mutations within the highly conserved 5.8S ribosomal subunit. Given these characteristics, it is expected that these two sequences represent non-

functional (pseudogene) copies of nrDNA (Buckler and Holtsford 1996; Buckler et al. 1997; Bailey et al. 2003). It is possible for paralogous copies of nrDNA to remain functional and under selective pressure over long periods of evolutionary time (Buckler et al. 1997). The term pseudogenes refers to nonfunctional paralogs, which may show significant deviations in G-C percentage or mutations within the 5.8S subunit which would lead to a change in the relative free energy of the subunit.

Recently, there have been several studies published using ITS and/or ETS in phylogenetic analyses of *Vignea*. Roalson et al. (2001) used datasets comprising nrDNA and chloroplast markers to investigate broader relationships within the Cariceae. This study established a well-supported monophyletic clade of *Vignea*, and also concluded that many of the taxonomic groups within the Cariceae are not monophyletic, specifically, *Carex* sensu stricto. Analyses by Hendrichs et al. (2004) and Ford et al. (2006) concluded many sections within *Vignea* are not monophyletic, based on ITS datasets analyzed with parsimony.

We have created large ITS and ETS sequence data sets with a number of the species represented by numerous cloned sequences. We investigate the level of persistence of paralogous or pseudogene copies of nrDNA within a subset of species in *Carex* subgenus *Vignea*, thereby gaining insight into the impact of using nrDNA sequences in reconstructing species phylogenies in *Vignea*. We also investigate how incomplete concerted evolution can exacerbate phylogenetic issues including shallow and deep incomplete lineage sorting and possible hybridization. This paper has been accepted for publication in Systematic Botany, and has been formatted as such.

MATERIALS AND METHODS

Taxon Sampling. Original sequences of ITS were obtained from 87 individuals representing 73 or 24% of the species of *Carex* subgenus *Vignea* (Appendix). These species were sampled from across most of the androgynous sections as well as some unispicate and gynecandrous species. Species identifications were verified by the authors. Of these 87 individuals, two individuals of *C. cephalophora*, three individuals of *C. gibba*, and one individual of *C. diandra*, *C. rosea*, and *C. macrocephala* were chosen to clone based on preliminary analyses suggesting either polyphyly of species or the possible amplification of a pseudogene. Eighty-four unique cloned sequences were obtained from these eight individuals and included in our phylogenetic analyses. Two species, *C. polystachya* and *C. aphylla*, were chosen as outgroups for both the ITS and ETS analyses based on previous phylogenetic analyses (Hendrichs et al. 2004).

In most cases individuals sequenced for ITS were also successfully amplified and sequenced for ETS. A few individuals were not sequenced for both markers (Appendix). Eighty-three individuals representing 67 species of *Vignea* and the same two outgroups were successfully sequenced for ETS. ETS samples cloned were the same as those for ITS with the exception of individual *C. gibba*1 for which ETS amplification failed (Appendix).

DNA Sequencing and Cloning. DNA was isolated from fresh material (field collected and grown in the greenhouse) or 30mg of herbarium material using a modified 2x CTAB protocol (Roalson et al. 2001). Polymerase chain reaction (PCR) was used to amplify the continuous ITS region using primers ITS5i and ITS4i (Baldwin 1992; Baldwin et al. 1995; Roalson et al. 2001). The 5′ end of ETS was amplified with primers

ETS-1F and 18S-R (Starr et al. 2003). PCR mixtures contained 5μl of 10x PCR buffer; 2.5μl of a 4mM stock solution of a combined set of all four dNTP′s; 3μl of DMSO; 3 units of *Taq*; 2.5μl of 10 pmol/μl of ITS5i and ITS4i; 3μl of 25mM $MgCl_2$; 10-60ng of DNA; and water to a final volume of 50μl for each reaction. These reaction conditions were not modified and those samples that were not amplifiable at these conditions were not included in this study. Thermocycler settings were set to 96°C for 1 min; 96°C for 1 min, 48°C for 1 min, 72°C for 1 min repeated 30 times; and 72°C for 5 min. ETS amplification used the same reaction conditions as ITS, except an annealing temperature of 55°C was used for 1 min.

One to three individuals of the five species chosen for cloning were selected employing standard procedures due to their placement in preliminary phylogenetic analyses (based on clade distribution and potential non-monophyly of initial sequences), and were cloned employing standard procedures. A concern raised by Alvarez and Wendel (2003) was the potential for contaminants to influence the phylogenetic analyses. Due to this concern, DNA was re-extracted from these individuals from leaf tissue that was adjacent to the culm of the inflorescence used in identification. Amplified PCR products of ITS and ETS for each individual were gel purified using the Wizard Gel Purification System under the manufacture's recommended protocol (Promega Corporation). Purified PCR products were ligated into the pGEM-T Easy vector and transformed into JM109 competent *E. coli* cells (Promega Corporation). Clones were directly amplified using standard T7 and SP6 primers under the same PCR conditions as described for ITS with the exception of a 50^0C annealing temperature (Promega Corporation). Initial cloning of the *C. gibba* collections did not result in a copy of the *C.*

gibba 2 specimen in the clade sister to the rest of the subgenus (see results). Further amplifications were conducted, varying DMSO content up to 16%, in order to determine if the copy was absent from that collection.

PCR products were electrophoresed on agarose gels to confirm a single band, and purified using PEG precipitation (Johnson and Soltis 1995). Sequencing was performed for both strands using an Applied Biosystems 377 or 3730 Automated DNA Sequencer. Cycle sequencing of purified PCR products followed the manufacture's protocol for Big Dye Terminators version 2.0 or version 3.

DNA sequences were assembled into contigs and edited using Sequencher version 4.2 (Gene Codes Corporation Inc.). Sequences were aligned in ClustalX version 1.81 (Thompson et al. 1997) using fast pairwise comparisons, and then manually re-aligned in Se-Al version 2.0.9 (Rambaut 1996).

Phylogenetic Analyses. Model selection for maximum likelihood (ML) analyses was completed using DT-ModSel Perl script (Minin et al. 2003), which uses a combination of the Bayesian Information Criterion (BIC) and decision theory to choose an appropriate model, with score and tree files obtained in PAUP* version 4.0b10 (Swofford 2001) with the initial parameters determined for a neighbor joining (NJ) tree. The initial ML trees were obtained by heuristic searching in PAUP* beginning with a NJ tree followed by tree bisection-reconnection (TBR) branch swapping. The initial ML tree was then used to re-optimize model parameters for 10 random addition heuristic searches with TBR branch swapping. This was followed by another round of re-optimizing parameters and another 10 random addition heuristic searches. The topology of the ML tree was identical between each set of heuristic searches.

Nodal Support. Nodal support was determined in PAUP* for the ML trees using 200 ML bootstrap replicates each with 10 random addition replicates. Model parameters were identical to those used in the final heuristic search.

Bayesian posterior probabilities (BPP) of nodes were determined in MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Alteka et al. 2004). Two independent Metropolis-coupled Markov chain Monte Carlo (MC^3) analyses were run with 20 million generations and sampling every 1000 generations. Parameters of the MC^3 were kept at default values except for the number of substitutions (set to 6), number of chains (8) and the heating parameter (temp) was kept at 0.2 for two separate runs and changed to 0.02 for two additional runs. The purpose of this was to determine if changing the heating parameter would change the sampling of the posterior probability distribution.

Statistical Tests. Constraint tests were used to determine if the monophyly of polyphyletic species and individuals could statistically be rejected by the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999) and BPP's. Likelihood scores were obtained for constrained and unconstrained ML trees. The ML tree was determined in PAUP* for each dataset under a likelihood criterion as described above. BPP's were determined for the same constraints of species monophyly.

Chi-square goodness of fit tests were performed on each sequence to determine if there were any significant deviations from the expected base frequencies. These were performed to identify putative pseudogenes, as a small number of point mutations within the 5.8S subunit are not necessarily indicative of a pseudogene, as the free energy of the

subunit may not change significantly. Observed and expected base frequencies were calculated in PAUP* and the chi-square tests were performed manually for each sequence.

Minimum-Energy of Secondary Structures. The minimum-energy of the secondary structure of the 5.8S ribosomal subunit was determined with the online program MFold (Markham and Zucker 2005; Santa Lucia 1998). This program is used to determine the optimal and suboptimal secondary structures of RNA and DNA. Razafimandimbison et al. (2004) used a correlation between free-energy and G-C percentage to screen for putative pseudogenes. In this study we used MFold to determine if sequences with novel mutations in 5.8S had a significant change in free-energy, and we could correlate this with having a significant change in G-C percentage.

RESULTS

Pseudogene Determination. G-C percentage of the 5.8S varied from 54.1% to 57.0% with no significant change in any sequence (lowest p-value = 0.68). Additionally, none of the 5.8S sequences had a significant change in free energy (lowest p-value = 0.55).

Model Selection. The general-time-reversible model (GTR) plus a gamma shape parameter was selected for the ITS dataset, and the Tamura-Nei 3 plus gamma shape distribution for the ETS dataset. Progressive rounds of parameter optimization and heuristic tree searching yielded one maximum likelihood tree for each of the ITS and ETS datasets (Figs. 2-5). The model parameters for the ITS dataset were: base frequencies A: 0.19566424, C: 0.29515080, G: 0.31654936, T: 0.1926356; relative rate

matrix A-C: 1.09121444, A-G: 2.16631056, A-T: 1.81088842, C-G: 0.56985201, C-T: 5.28848493, G-T: 1.0000; and a gamma shape parameter of 0.578663. The model parameters for the ETS dataset were: base frequencies A: 0.21001105, C: 0.20914621, G: 0.27257732, T: 0.30826542; relative rate matrix A-C: 1.00000000, A-G: 3.19039418, A-T: 1.00000000, C-G: 1.00000000, C-T: 5.48624019, G-T: 1.00000000; and a gamma shape parameter of 0.977082.

ITS and ETS Phylogenies. ITS and ETS ML analyses each resulted in a single tree (ITS: -lnL=8381.4991; ETS: -lnL=6878.48453). Patterns within both the ITS and ETS phylogenies suggest that many of the included sections do not appear to be monophyletic. Support throughout the tree appears to be low, with only a fraction of the clades having high support values for both Bayesian posterior probabilities and non-parametric bootstrap percentages. Support for *C. aphylla* and *C. polystachya* as an outgroup is low, and may be a result of the lower genetic distance between the early diverging *C. gibba* clade and *C. aphylla* and *C. polystachya*.

The ETS ML phylogeny appears largely congruent with the ITS ML phylogeny. The ETS phylogeny has a grade of clades containing individuals from several sections of *Vignea*, most of which are not monophyletic. As with the ITS phylogeny the support values for clades on the ETS phylogeny appear to be low for both BPP and ML bootstrap, although the support and resolution of the ETS phylogeny is higher than that of the ITS phylogeny. Levels of intraindividual and intraspecies polymorphism for both ITS and ETS appear to be similar.

Within the ITS phylogeny sequences from *C. cephalophora* are polyphyletic within clade A (Fig. 3). This clade is not well supported with either BPP or bootstrap

percentages. Sequences from *C. gibba* 2 are paraphyletic with sequences from *C. diandra*, however, no duplicate sequences were recovered between these two individuals. *Carex rosea*2 cloned sequences do not form a monophyletic clade, however the *C. rosea* species complex forms a well-supported monophyletic clade. Forty-eight clones were sequenced from *C. macrocephala* and only five different sequences were identified and formed a monophyletic group.

Levels of intra-individual polymorphism are also high for cloned sequences of ETS. Similarly to that of ITS, ETS cloned sequences of *C. cephalophora* are polyphyletic within clade A (Fig. 5). Likewise, *C. gibba* is also found to be polyphyletic with some sequences from one individual paraphyletic with *C. diandra*, and a clade of cloned sequences sister to the rest of *Vignea*.

Section *Macrocephalae*, represented by clade C (Fig. 4) forms a well-supported monophyletic clade. As with ITS, ETS supports a monophyletic clade of *C. macrocephala* cloned sequences, well-supported as sister to *C. kobomugi*. The *Carex rosea* species complex forms a well-supported monophyletic clade (D; Fig. 4). Like ITS, ETS sequences of *C. rosea* are polyphyletic within this clade.

Initial amplification and cloning of ITS and ETS from *C. gibba* resulted in clones from *C. gibba* 1 & 3 forming a clade sister to the rest of *Vignea*, while the sequences from *C. gibba* 2 formed a clade with sample from *C. diandra* (Figs. 2-5). Further amplification of the *C. gibba* 2 sample using a broad range of DMSO concentrations resulted in one clone, *C. gibba* 2-1 placed within the clade of clones from *C. gibba* 1 & 3 (clade E; Figs. 2 & 4v), and was amplified using 12% DMSO. Other samples from this test (*C. gibba* 2-2 [8% DMSO], *C. gibba* 2-3 [12% DMSO], and *C. gibba* 2-4

(8%DMSO]) all grouped with the sampled from the lower DMSO concentration cloning efforts and *C. diandra* (clade B; Figs. 3 & 5).

DISCUSSION

The results of our phylogenetic analyses suggest a complex evolutionary history of nrDNA within lineages of *Carex* subgenus *Vignea*. Incomplete concerted evolution has been demonstrated to occur within *Vignea* through the extensive intra-individual polymorphisms obtained from cloning. Studies have identified intra-individual polymorphims in other groups (Buckler and Holtsford 1996; Campbell et al. 1997; Denduangboripant and Cronk 2000; Hartmann et al. 2001), but typically have not seen the level of polyphyly identified in *Vignea*. Results of the MFold tests and G-C content significance test suggest no correlation between G-C content and free-energy, as well as no correlation between pairwise distance from the most common 5.8S sequence and free-energy. This suggests that sequences with mutations in the 5.8 subunit are not necessarily pseudogenes.

Incomplete concerted evolution is only part of the explanation for the level of polyphyly seen within these lineages. As described previously, issues such as hybridization and lineage sorting can compound issues arising due to incomplete concerted evolution. We are able to statistically reject monophyly of the three cloned individuals of *C. gibba* as well *C. cephalophora* for both sequences of ITS and ETS. Ford et al. (2006) supported *C. gibba* as being sister to the rest of *Vignea*. However, our increased sampling has demonstrated deep time incomplete lineage sorting may have led to the phylogenetic patterns we see in our analyses (Figs. 2 & 4). *Carex gibba* (A) was

sampled from western China, while *C. gibba* (B) was sampled from northern Japan. Alternatively, concerted evolution of introgressed nrDNA would explain the paraphyletic individuals of *C. gibba*.

Carex cephalophora individuals are also polyphyletic in these analyses. The cloned sequences of ITS and ETS display a different pattern from that of *C. gibba* in that there are several species that are sister to the paralogs within *C. cephalophora* (Figs. 3 & 5). We conclude that incomplete concerted evolution as well as shallow incomplete lineage sorting may have caused the pattern shown in these phylogenetic analyses. Hybrids of *C. cephalophora* are not known to occur (Cayouette and Catling 1992).

Several species groups within *Vignea* are strongly supported as monophyletic. As a means to investigate the influence of nrDNA paralogy and incomplete concerted evolution on the relationships of species among and within theses species groups we chose to clone *C. rosea* and *C. macrocephala*. *Carex rosea* is a member of the *Carex rosea* species complex comprising six North American species (Weber and Ball 1984). This species complex is a well-supported monophyletic group, and the cloned sequences of ITS and ETS all fall within this monophyletic group (Figs. 2 & 4). However, the cloned sequences of ITS and ETS of *C. rosea* are not monophyletic. This pattern is indicative of incomplete lineage sorting within this species group as introgression is not thought to occur between species of the *C. rosea* species complex due to shifts in chromosome number (Webber and Ball 1984; Cayouette and Catling 1992).

Carex macrocephala and *C. kobomugi* are the only species of section *Macrocephalae*. It was suggested by Ford et al. (2006) that the sequences of ITS amplified from *C. kobomugi* were non-functional paralogs because of the deviation in G-

C percentage and the mutations present in the conserved 5.8S ribosomal subunit. Of the five different cloned *C. macrocephala* sequences, none significantly deviated in G-C percentage and all sequences contain the mutation found by Ford et al. (2006) in putative pseudogenes at the beginning of the 5.8S ribosomal subunit. The ITS and ETS cloned sequences of *C. macrocephala* form a polytomy sister to the sequences amplified from *C. kobomugi* (Figs. 2 & 4). *Carex macrocephala* is a monoecious species and *C. kobomugi* is dioecious. Hybridization between these two species is apparently rare as populations of female-only individuals of *C. kobomugi* do not appear to be fertilized by pollen coming from sympatric *C. macrocephala* individuals (M. King, pers. observ.).

Deep paralogy levels within *Vignea* is possibly a result of long term and repeated hybridization events between species. Hybridization within *Vignea* is not well documented, and hybrids of the cloned species were not noted in the review of Cayouette and Catling (1992). Although studies have suggested that concerted evolution may occur quickly following a hybridization event (Hillis et al. 1991), this may not be the case for sedges. To test for broadscale introgression through hybridization datasets consisting of variable cpDNA markers, single copy nuclear gene sequences, and possibly population level codominant markers such as microsatellites or allozymes would be needed (Dobes et al. 2004).

Bailey et al. (2003) described differing levels of paralogy seen in nrDNA, and the significance of the impact on phylogenetic reconstruction of incomplete lineage sorting increases the deeper the coalescence events. In this study the amount of incomplete lineage sorting is extensive and paralogous copies coalesce deeper in the tree than typically found (Mayol and Rosselló 2001; Li et al. 2004). Notably, if the polyphyly of *C.*

gibba is due to incomplete lineage sorting this would require a coalescence of nrDNA sequences prior to diversification of major lineages within *Vignea.*

In general, the topologies found here are very similar to those found by Hendrichs et al. (2004) and Ford et al. (2006). It seems likely that the lack of sectional monophyly found by the previous authors in *Vignea* is not entirely due to the paralogy issues noted here. Further, in many of the cases of non-monophyly of species samples (such as found in the *C. rosea* species complex), these may reflect the very close relationship of the species and presumed recency of divergence. It is plausible that the results presented here for the *C. rosea* complex reflect a need to reevaluate the species circumscription of *C. rosea*.

Alternatively, in some lineages such as *Carex* section *Ovales* (Hipp et al. 2006), few paralogs have been found, suggesting ITS and ETS may be appropriate markers for phylogenetic inference. It should be noted, however, this was only made clear by cloning samples in that study. The overall similarity of topologies found in this study and previous phylogenetic analyses of nrDNA in *Vignea* does not ameliorate the fact that ITS and ETS do not appear to be able to unequivocally place species within clades. Further, while at this point there is only evidence for deep paralogy (or hybridization) problems with *C. gibba*, without other evidence there is no way to know whether deep paralogy or hybridization are influencing some of the other species placements that suggest non-monophyletic sections.

Incomplete concerted evolution has reduced the phylogenetic utility of nrDNA in many groups of plants, including *Carex* subgenus *Vignea*. It remains unclear as to whether the phylogenetic patterns we are detecting in *Carex* are due to just incomplete

lineage sorting or the combination of it with interspecific hybridization. Interspecific hybridization has not been shown to be prevalent within the sampled species of *Vignea*. If it occurred early enough within a lineage, it is possible that chimeric arrays of nrDNA could occur within species. It is also possible that arrays of nrDNA have been copied to non-homologous chromosomes, thereby limiting the amount of concerted evolution that can occur to homogenize the copies of nrDNA. With the extensive intra-individual polymorphisms identified within species of *Carex* subgenus *Vignea*, future use of ITS and ETS in *Carex* needs to include rigorous cloning to determine the potentially confounding effects of multiple paralogous copies on phylogenetic inference of relationships.

LITERATURE CITED

ALTEKAR, G., S. DWARKADAS, J. P. HUELSENBECK, and F. RONQUIST. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20: 407-415.

BALDWIN, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* 1: 3–16.

— and S. MARKOS. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S 26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution* 10: 449 463.

—, M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, and M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.

BUCKLER, E. S. and T. P. HOLTSFORD. 1996. *Zea* ribosomal repeat evolution and substitution patterns. *Molecular Biology and Evolution* 13: 623–632.

—, A. IPPOLITO, and T. P. HOLTSFORD. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145: 821–832.

CAMPBELL, C. S., M. F. WOJCIECHOWSKI, B. G. BALDWIN, L. A. ALICE, and M. J. DONOGHUE. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier agamic* complex (Rosaceae). *Molecular Biology and Evolution* 14: 81–90.

CAYOUETTE, J. and P. M. CATLING. 1992. Hybridization in the Genus *Carex* with Special Reference to North America. *The Botanical Review* 58: 351-440.

CRONN, R. C., X. ZHAO, A. H. PATERSON, and J. F. WENDEL. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution* 42: 685 705.

DENDUANGBORIPANT, J. and Q. C. B. CRONK. 2000. High intraindividual variation in internal transcribed spacer sequences in *Aeschynanthus* (Gesneriaceae) implications for phylogenetics. *Proceedings of the Royal Society London B* 267: 1407–1415.

DOBES, C., T. MITCHELL-OLDS, and M. A. KOCH. 2004. Intraspecific diversification in North American *Boechera* stricta (= *Arabis drummondii*), *Boechera xdivaricarpa,* and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers - An integrative approach. *American Journal of Botany* 91: 2087-2101.

DOYLE, J. J. and J. I. DAVIS. 1998. Homology in molecular phylogenetics: a parsimony perspective. Pp. 101–131 in:, Molecular Systematics of Plants II: DNA Sequencing, ed. D. E. Soltis, P. S. Soltis, and J. J. Doyle Boston: Kluwer Academic.

FORD, B. A., M. IRANPOUR, F. C. NACZI, J. R. STARR, and C. A. JEROME. 2006. Phylogeny of *Carex* subg. *Vignea* (Cyperaceae) Based on Non-coding nrDNA Sequence Data. *Systematic Botany* 31: 70–82.

FUERTES-AGUILAR, J., J. A. ROSSELLO, and G. NIETO FELINER. 1999. Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Molecular Ecology* 8: 1341–1346.

GERNANDT, D. S. and A. LISTON. 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *American Journal of Botany* 86: 711–723.

HAMBY, R. K. and E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (eds.), Molecular Systematics of Plants. Chapman & Hall, New York, NY, p. 50–91.

HARTMANN, S., J. D. NASON, and D. BHATTACHARYA. 2001. Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. *Journal of Molecular Evolution* 53: 124–134.

HENDRICHS, M., S. MICHALSKI, D. BEGEROW, F. OBERWINKLER, and F. H. HELLWIG. 2004. Phylogenetic relationships in *Carex*, subgenus *Vignea* (Cyperaceae), based on ITS sequences. *Plant Systematics and Evolution* 246: 109–125.

HERSHKOVITZ, M. A., E. A. ZIMMER, and W. J. HAHN. 1999. Ribosomal DNA sequences and angiosperm systematics. Pp. 268–326 in Molecular Systematics and Plant Evolution, ed. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall. London: Taylor & Francis.

HILLIS, D. M. and M. T. DIXON. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quarterly Review of Biology* 66: 411– 453.

—, C. MORITZ, C. A. PORTER, and R. J. BAKER. 1991. Evidence of biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251: 308–310.

HIPP, A. L., A. A. REZNICEK, P. E. ROTHROCK, and J. A. WEBER. 2006. Phylogeny and classification of *Carex* section *Ovales* (Cyperaceae). *International Journal of Plant Sciences* 167: 1029-1048.

HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, and J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.

HUGHES, C. E., C. D. BAILEY, and S. A. HARRIS. 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. *American Journal of Botany* 89. 1057–1073.

JOHNSON, L. A. and D. E. SOLTIS. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using matK sequences. *Annals of the Missouri Botanical Garden* 82. 149 175.

KITA, Y. and M. ITO. 2000. Nuclear ribosomal ITS sequences and phylogeny in East Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae), with special reference to extensive polymorphism in individual plants. *Plant Systematics and Evolution* 225: 1–13.

LI, J., J. LEDGER, T. WARD, and P. DEL TREDICI. 2004. Phylogenetics of Calycanthaceae based on molecular and morphological data, with a special reference to divergent paralogues of the nrDNA ITS region. *Harvard Papers in Botany* 9: 69 82.

LINDER, C. R., L. R. GOERTZEN, B. V. HEUVEL, J. FRANCISCO-ORTEGA, and R. K. JANSEN. 2000. The complete external transcribed spacer of 18S–26S rDNA:

amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Molecular Phylogenetics and Evolution* 14: 285–303.

MARKHAM, N. R. and M. ZUKER. 2005. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research* 33: W577-W581.

MAYOL, M. and J. A. ROSSELLO. 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution* 19: 167–176.

MCDADE, L. A. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. *Evolution* 46: 1329–1346.

MININ V., Z. ABDO, P. JOYCE, and J. SULLIVAN. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* 52: 1-10.

MUIR, G., C. C. FLEMING, and C. SCHLEOTTERER. 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Molecular Biology and Evolution* 18: 112–119.

RAMBAUT, A. 1996. Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/.

RAZAFIMandIMBISON, S. G., E. A. KELLOGG, and B. BREMER. 2004. Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: A case study from Naucleeae (Rubiaceae). *Systematic Biology* 53: 177 192.

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

ROALSON, E. H., J. T. COLUMBUS, and E. A. FRIAR. 2001 Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and trnT-L-F (cpDNA) region

sequences: assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. *Systematic Botany* 26: 318-341.

RONQUIST F. and J. P. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

RUGGIERO, M. V. and G. PROCACCINI. 2004. The rDNA ITS region in the Lessepsian marine angiosperm *Halophila stipulacea* (Forssk.) Aschers. (Hydrocharitaceae): Intragenomic variability and putative pseudogenic sequences. *Journal of Molecular Evolution* 58: 115-121.

SANDERSON, M. T. and J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy and confidence. *Systematic Biology* 41: 4–17.

SANTALUCIA, J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Science USA* 95: 1460-1465.

SHIMODAIRA, H. and M. HASEGAWA. 1999. Multiple comparisons of log likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114-1116.

SMALL, R. L., R. C. CRONN, and J. F. WENDEL. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145-170.

STARR, J. R., R. J. BAYER, and B. A. FORD. 1999. The phylogenetic position of *Carex* section *Phyllostachys* and its implications for phylogeny and subgeneric circumscription in *Carex* (Cyperaceae). *American Journal of Botany* 86: 563 577.

—, S. A. HARRIS, and D. A. SIMPSON. 2003. Potential of the 5' and 3' ends of the intergenic spacer (IGS) of rDNA in the Cyperaceae: New sequences for lower level phylogenies in sedges with an example from *Uncinia*. *International Journal of Plant Sciences* 164: 213 227.

SWOFFORD, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0 beta 10. Sunderland: Sinauer Associates.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.

WEBBER, J. M. and P. W. BALL. 1984. The taxonomy of the *Carex rosea* group in eastern Canada. *Canadian Journal of Botany* 62: 2058-2073.

WEI, X. X. and X. Q. WANG. 2004. Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. *Molecular Ecology* 13: 3115-3123.

WENDEL, J. F. 1995b. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Science USA* 92: 280–284.

—, A. SCHNABEL, and T. SEELANAN. 1995a. An Unusual Ribosomal DNA Sequence from *Gossypium gossypioides* Reveals Ancient, Cryptic, Intergenomic Introgression. *Molecular Phylogenetics and Evolution* 4: 298-313.

YANG, Y. W., K. N. LAI, P. Y. TAI, D. P. MA, and W. H. LI. 1999. Molecular phylogenetic studies of *Brassica, Rorippa, Arabidopsis* and allied genera based

on the internal transcribed spacer region of 18S–25S rDNA. *Molecular Phylogenetics and Evolution* 13: 455–462.

YEN, A. C. and R. G. OLMSTEAD. 2000. Molecular systematics of Cyperaceae tribe Cariceae based on two chloroplast DNA regions: ndhF and trnL intron-intergenic spacer. *Systematic Botany* 25: 479–494.

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLEY, Y. W. KAN, and A. C. WILSON. 1980. Rapid duplication and loss of genes coding for the alphachains of hemoglobin. *Proceedings of the National Academy of Science USA* 77: 2158–2162.

TABLE 1. Results for Shimodaira-Hasegawa and Bayesian posterior probability tests of monophyletic constraints. Monophyletic constraints for each set of clones and for all individuals in the species.

Monophyletic constraint	-lnL	Diff	p-value	BPP
ITS ML Tree	8592.57711	0		
*cephalophora*1	8640.47182	47.89471	0.064	<0.001
*cephalophora*2	8656.99855	64.42144	0.018*	<0.001
All *cephalophora* individuals	8673.93302	81.35590	0.003*	<0.001
diandra	8634.41151	41.83439	0.088	<0.001
*gibba*1	8663.20824	70.63112	0.01*	<0.001
*gibba*2	8751.03364	60.47239	0.043*	<0.001
*gibba*3	8636.60811	44.03100	0.062	<0.001
All *gibba* individuals	8653.04950	158.45653	<0.0001*	<0.001
*rosea*2	8635.00890	42.43179	0.121	<0.001
All *rosea* individuals	8636.73366	44.15654	0.087	<0.001
Monophyletic constraint	-lnL	Diff	p-value	BPP
ETS ML Tree	6819.60218	0		
*cephalophora*1	6884.54951	64.94733	0.048*	<0.001
*cephalophora*2	6893.11797	73.51579	0.048*	<0.001
All *cephalophora* individuals	6951.92036	132.31818	0.013*	<0.001

diandra	6864.87391	45.27173	0.067	<0.001
*gibba*2	6853.94299	34.34081	0.134	<0.001
All *gibba* individuals	6968.46820	148.86602	<0.0001*	<0.001
*rosea*2	6839.76921	20.16703	0.195	<0.001
All *rosea* individuals	6899.92920	80.32702	0.039*	<0.001

APPENDIX. List of voucher specimens used in analyses of *Carex* subgenus *Vignea*. Specimens are arranged alphabetically order by section. GenBank accession numbers are in the order ITS, ETS1f. Sectional classification follows those used in Ford et al. (2006).

Sect. *Ammoglochin* Dumort.: ***Carex arenaria*** L. 1, Netherlands, Brakman 1877 (WS) EU000926, EU001090; ***Carex arenaria*** L. 2, U.S.A.: New Hanover Co., NC, Godfrey 51166 (WS) EU000927, EU001091; ***Carex brizoides*** L. Amoen., Czech Republic, Vasak RSA 580597 (RSA) EU000930; ***Carex ligerica*** J.Gay., Sweden: Scania, Snogerup 4122 (RSA) EU001026; ***Carex praecox*** Schreb., Crimea, Korzhenervsky (RSA) EU001048, EU001197; ***Carex siccata*** Dewey, U.S.A.: Coconino Co., AZ, Morefield 2898 (RSA) EU001064, EU001218.

Sect. *Bracteosae* Pax: ***Carex bonariensis*** Desf., Uruguay: San Pedro, Aragone et al. 380 (RSA) EU000929, EU001093.

Sect. *Chordorrhizae* (Heuff.) Meinsh.: ***Carex chordorrhiza*** Ehrh. ex L.f., U.S.A.: Coos Co., OR, Zika 13217 (WS) EU000954, EU001114.

Sect. *Dispermae* Ohwi: ***Carex disperma*** Dewey, Finland: Karkkila, Nurmi 84-31 (WS) EU000976, EU001133.

Sect. *Divisae* H. Christ ex Kük. in Engl.: ***Carex curaica*** Kunth, Russia: Southern Siberia, Elias et al. 7620 (RSA) EU000957; ***Carex divisa*** Hudson, Turkmenistan, Kurbanov 586 (MO) EU000977; ***Carex douglasii*** Boott, U.S.A.: Asotin Co., WA, Fishbein 3341 (WS) EU000978, EU001134; ***Carex duriuscula*** C.A.Mey., U.S.A.: Mono Co, CA, Taylor 9129 (RSA) EU000979, EU001135; ***Carex enervis*** C.A.Mey. 1, China, Ho 1173 (A) EU000980, EU001136; ***Carex enervis*** C.A.Mey. 2, China: Xinjiang Prov,

Morefield 5047 (RSA) EU000981; ***Carex gayana*** Desv., Argentina: Santa Cruz, Roivainen 2357 (RSA) EU000982, EU001137; ***Carex macrorrhiza*** Boeck., Argentina: Santa Cruz, Roivainen 2630 (RSA) EU001032, EU001181; ***Carex pansa*** L.H.Bailey, U.S.A.: Pacific Co., WA, Arnot 352 (WS) EU001043, EU001194; ***Carex praegracilis*** W.Boott, U.S.A.: Adams Co., WA, Fishbein 3891 (WS) EU001049, EU001198; ***Carex simulata*** Mack., U.S.A.: Plumas Co., CA, Janeway 3316 (WS) EU001065, EU001219; ***Carex stenophylla*** Wahlenb., Russia, MO 04981469 (MO) EU001070, EU001224.

Sect. *Foetidae* (Tuck. ex L.H.Bailey) Kük. in Engl.: ***Carex incurviformis*** Mack., U.S.A.: Mono Co., CA, Morefield 4831 (WS) EU001020, EU001169; ***Carex maritima*** Lightfoot 1, U.S.A.: Park Co., CO, Weber & Randolph 17402 (WS) EU001033, EU001182; ***Carex maritima*** Lightfoot 2, Tierra Del Fuego, Goodall 963 (RSA) EU001034; ***Carex perglobosa*** Mack., U.S.A.: Park Co., CO, Weber 7914 (WS) EU001047, EU001196; ***Carex vernacula*** L.H.Bailey, U.S.A.: Wheeler Peak, CA, Bell 1459 (WS) EU001077.

Sect. *Gibbae* Kük. in Engl.: ***Carex gibba*** Wahlenb. 1, Japan: Honshu Prov., Tsugaru 24765 (MO) **ITS Sequences (1A** EU000983; **1B** EU000984; **1C** EU000985; **1D** EU000986; **1E** EU000987; **1F** EU000988; **1G** EU000989; **1H** EU000990; **1I** EU000991; **1J** EU000992; **1K** EU000993; **1L** EU000994; **1M** EU000995**)**; ***Carex gibba*** Wahlenb. 2, Japan: Honshu Prov., Tsugaru 17908 (MO) **ITS Sequences (2A** EU000996; **2B** EU000997; **2C** EU000998; **2D** EU000999; **2E** EU001000; **2F** EU001001; **2G** EU001002; **2H** EU001003; **2I** EU001004; **2J** EU001005**), ETSf Sequences (2A** EU001138; **2B** EU001139; **2C** EU001140; **2D** EU001141; **2E** EU001142; **2F** EU001143; **2G** EU001144; **2H** EU001145; **2I** EU001146; **2J** EU001147; **2K**

EU001148)**;** **ITS Sequences (2-1** EU001079; **2-3** EU001080; **2-4** EU001081; **2-2** EU001082**)** **,** **ETSf Sequences (2-1** EU001232; **2-3** EU001233; **2-4** EU001234); ***Carex gibba*** Wahlenb. 3, China: Hunan Prov., Chong-Chun 1380 (A) **ITS Sequences (3A** EU001006; **3B** EU001007; **3C** EU001008; **3D** EU001009; **3E** EU001010; **3F** EU001011; **3G** EU001012; **3H** EU001013; **3I** EU001014; **3J** EU001015; **3K** EU001016; **3L** EU001017**)**, **ETSf Sequences (3A** EU001149; **3B** EU001150; **3C** EU001151; **3D** EU001152; **3E** EU001153; **3F** EU001154; **3G** EU001155; **3H** EU001156; **3I** EU001157; **3J** EU001158; **3K** EU001159; **3L** EU001160; **3M** EU001161; **3N** EU001162; **3O** EU001163; **3P** EU001164; **3Q** EU001165; **3R** EU001166).

Sect. *Heleoglochin* Dumort.: ***Carex albata*** Boot, Japan: Honshu, Tateishi and Hoshi 10439 (A) EU001085; ***Carex appressa*** R.Br., Australia: Tazmania, Orchard 5310 (RSA) EU000925, EU001089; ***Carex cusickii*** Mack. ex Piper and Beattie 1, U.S.A.: Stevens Co., WA, Björk 1060 (WS) EU001117; ***Carex cusickii*** Mack. ex Piper and Beattie 2, U.S.A.: Skamania Co., WA, Schuller 922 (WS) EU000958, EU001118; ***Carex decomposita*** Muhl., U.S.A.: Washington Co., MS, Bryson 3502 (WS) EU000961, EU001119; ***Carex diandra*** Schrank, U.S.A.: Okanogan Co., WA, Naas 5461 (WS) **ITS Sequences (A** EU000964; **B** EU000965; **C** EU000966; **D** EU000967; **E** EU000968; **F** EU000969; **G** EU000970; **H** EU000971; **I** EU000972; **J** EU000973; **K** EU000974**)**, **ETSf Sequences (**EU001131; **A** EU001122; **B** EU001123; **C** EU001124; **D** EU001125; **F** EU001126; **G** EU001127; **J** EU001128; **K** EU001129; **L** EU001130); ***Carex paniculata*** L., Macedonia: Mt. Levkasia,, Christiansen 8290 (RSA) EU001042, EU001193; ***Carex paradoxa*** Willd., Austria: Halstatt, Morton 6011 (RSA) EU001044;

Carex prairea Dewey ex Wood, U.S.A.: Caledonia Co., UT, Bouffod 22980 (WS) EU001050, EU001199.

Sect. *Holarrhenae* (Döll) Pax in Engl. & Prantl: ***Carex sartwellii*** Dewey, U.S.A.: Hillsdale Co., MI, Fritsch 603 (RSA) EU001217; ***Carex unisexualis*** C.B.Clarke, China, Lai 3015 (A) EU001230.

Sect. *Leptocephalae* L.H.Bailey: ***Carex aphylla*** Kunth, AY242014, AY242015

Sect. *Macrocephalae* Kük. in Engl.: ***Carex kobomugi*** Ohwi, Japan, Tsugaru 20118 (A) EU001022, EU001171; ***Carex macrocephala*** Willd. ex Spreng., U.S.A.: San Juan Co., WA, King 408 (WS) **ITS Sequences (A** EU001027; **B** EU001028; **C** EU001029; **D** EU001030; **E** EU001031), **ETSf Sequences (A** EU001176; **B** EU001177; **C** EU001178; **H** EU001179; **K** EU001180).

Sect. *Multiflorae* (J.Carey) Kük. in Engl.: ***Carex alma*** L.H.Bailey, U.S.A.: Tulare Co., CA, Ertter 6879 (WS) EU000923, EU001086; ***Carex annectens*** Bicknell, U.S.A.: Yell Co., AR, Demaree 64960 (WS) EU001088; ***Carex chihuahensis*** Mack., Mexico: Chihuahua, Steinman 881 (RSA) EU000953, EU001113; ***Carex densa*** L.H.Bailey 1, U.S.A.: Butte Co, CA, Janeway 3219 (WS) EU000962, EU001120; ***Carex densa*** L.H.Bailey 2, U.S.A.: Benton Co., OR, Halse 2219 (RSA) EU000963, EU001121; ***Carex leiorhyncha*** C.A.Mey., China, He 80075 (MO) EU001025, EU001175; ***Carex neurocarpa*** Maxim. 1, Japan: Osaka, Seto 28488 (A) EU001037, EU001185; ***Carex neurocarpa*** Maxim. 2, China: Beijing, He 80072 (A) EU001186; ***Carex nubigena*** D.Don 1, China: Xinjiang Prov., Sino-American Exp. 870 (A) EU001039, EU001188; ***Carex nubigena*** D.Don 2, China: Xinjiang Prov., Sino-American Exp. 225 (A) EU001040, EU001189; ***Carex nubigena*** D.Don 3, China: Xinjiang Prov., Sino-American

Exp. 872 (A) EU001190; ***Carex triangularis*** Boeck., U.S.A.: Cass Co., TX, Cory 55920 (WS) EU001073, EU001227; ***Carex vulpinoidea*** Michx., U.S.A.: Shasta Co., CA, Janeway 3928 (WS) EU001078, EU001231.

Sect. *Ovales* Kunth: ***Carex bicknellii*** Britt. AF285039; ***Carex ovalis*** Goodenough, AF285002.

Sect. *Phaestoglochin* Dumort.: ***Carex aggregata*** Mack. 1, U.S.A.: Allegheny Co., PA, Brighton 15247 (WS) EU000921, EU001083; ***Carex aggregata*** Mack. 2, U.S.A.: Ashland Co., OR, Cusick 10832 (RSA) EU000922, EU001084; ***Carex austrina*** Mack., U.S.A.: Oktibbeha Co., MS, Bryson 8601 (WS) EU000928, EU001092; ***Carex cephaloidea*** (Dewey) Dewey 1, U.S.A.: St Louis Co., MN, Chase 13410 (WS) EU001094; ***Carex cephaloidea*** (Dewey) Dewey 2, U.S.A.: Clearwater Co., MN, Thorne 30772 (RSA) EU000931, EU001095; ***Carex cephalophora*** Muhl. ex Willd. 1, U.S.A.: Washington Co., MS, Bryson 3741(WS) **ITS Sequences (1A** EU000932; **1B** EU000933; **1C** EU000934; **1D** EU000935; **1E** EU000936; **1F** EU000937; **1G** EU000938; **1H** EU000939; **1I** EU000940; **1J** EU000941; **1K** EU000942), **ETSf Sequences (1A** EU001096; **1B** EU001097; **1C** EU001098; **1D** EU001099; **1E** EU001100; **1F** EU001101; **1G** EU001102; **1H** EU001103; **1O** EU001104); ***Carex cephalophora*** Muhl. ex Willd. 2, U.S.A.: Hampden Co., MA, Zebryk 4663 (RSA) **ITS Sequences (2A** EU000943; **2B** EU000944; **2C** EU000945; **2D** EU000946; **2E** EU000947; **2F** EU000948; **2G** EU000949; **2H** EU000950; **2I** EU000951; **2J** EU000952**)**, **ETSf Sequences (2A** EU001105; **2B** EU001106; **2C** EU001107; **2D** EU001108; **2E** EU001109; **2F** EU001110; **2G** EU001111; **2O** EU001112); ***Carex gravida*** L.H.Bailey, U.S.A.: Shannon Co., MO, Castaner & Bevard 3664 (WS) EU001018, EU001167;

Carex leavenworthii Dewey 1, U.S.A.: Knox Co., IL, Solomon 726 (WS) EU001024, EU001173; ***Carex leavenworthii*** Dewey 2, U.S.A.: Clinton Co., MO, Croat 17091 (RSA) EU001174; ***Carex muehlenbergii*** Schkuhr ex Spreng., U.S.A.: Grant Parrish, LA, Hyatt 8450 (WS) EU001035, EU001183; ***Carex occidentalis*** L.H.Bailey, U.S.A.: Cache Co., UT, Piep 19 (WS) EU001041, EU001191; ***Carex radiata*** (Wahlenb.) Small, U.S.A.: Sullivan Co., PA, Wahl 1074 (WS) EU001051, EU001200; ***Carex retroflexa*** Muhl. ex Willd. 1, U.S.A.: Butte Co., CA, Janeway 3215 (WS) EU001052, EU001201; ***Carex retroflexa*** Muhl. ex Willd. 2, U.S.A.: St Louis Co., MO, Solomon 3754 (RSA) EU001053, EU001202; ***Carex rosea*** Schkuhr ex Willd. 1, U.S.A.: Chisago Co., MN, Clements 666 (WS) EU001054, EU001203; ***Carex rosea*** Schkuhr ex Willd. 2, U.S.A.: Murray Co., GA, Boufford 22858 (WS) EU001055, EU001204 **ITS Sequences (2A** EU001056; **2B** EU001057; **2C** EU001058; **2D** EU001059; **2E** EU001060; **2F** EU001061; **2G** EU001062; **2H** EU001063**)**, **ETSf Sequences (2B** EU001205; **2C** EU001206; **2D** EU001207; **2E** EU001208; **2G** EU001209; **2H** EU001210; **2I** EU001211; **2J** EU001212; **2K** EU001213; **2L** EU001214; **2M** EU001215; **2N** EU001216); ***Carex socialis*** Mohlenb. and Schwegman, U.S.A.: Oktibbeha Co., MS, Bryson 5394 (WS) EU001066, EU001220; ***Carex sparganioides*** Muhl. ex Willd. 1, Canada: Essex Co., ONT, Calder 15845 (WS) EU001067, EU001221; ***Carex sparganioides*** Muhl. ex Willd. 2, U.S.A.: Allamakee Co., IA, Hartley 6643 (RSA) EU001068, EU001222; ***Carex sparganioides*** Muhl. ex Willd. 3, U.S.A.: Hillsdale Co., MI, Fritsch 644 (RSA) EU001069, EU001223; ***Carex texensis*** Torr. ex L.H.Bailey, U.S.A.: Alachua Co., FL, Thorne 44385 (RSA) EU001072, EU001226; ***Carex tumulicola*** Mack. 1, U.S.A.: Alameda Co., CA, Ertter 10146 (WS) EU001074, EU001228; ***Carex tumulicola*** Mack.

2, U.S.A.: Alameda Co., CA, Ertter 10147 (WS) EU001075, EU001229; ***Carex vallicola*** Dewey, U.S.A.: Coconino Co., AZ, David 681966 (WS) EU001076.

Sect. *Physoglochin* Dumort.: ***Carex davalliana*** Smith, Austria, Mackechnie (RSA) EU000959; ***Carex dioica*** L., Finalnd: Inari, Alho 1980 (RSA) EU000975, EU001132; ***Carex gynocrates*** Wormskjöld, Russia, Kharkevisch 954 (WS) EU001019, EU001168; ***Carex parallela*** Sommerfelt 1, Sweden: Lake Tornetrask, Alm 1380 (WS) EU001045, EU001195; ***Carex parallela*** Sommerfelt 2, Sweden: Lake Tornetrask, Alm 1381 (RSA) EU001046.

Sect *Phyllostachyae* (J.Carey) L.H.Bailey: ***Carex polystachya*** Sw., AF285014, AY241998.

Sect. *Vulpinae* (Heuff.) H.Christ: ***Carex alopecoidea*** Tuckerm., U.S.A.: Allegheny Co., PA, Bright 15242 (WS) EU000924, EU001087; ***Carex conjuncta*** Boott, U.S.A.: Story Co., IA, Lang & Brashear 116 (RSA) EU000955, EU001115; ***Carex crus-corvi*** Shuttlew., U.S.A.: Washington Co., MS, Bryson 3481 (WS) EU000956, EU001116; ***Carex declinata*** Boott, Australia: Queensland, Clemens 43743 (RSA) EU000960; ***Carex jonesii*** L.H.Bailey, U.S.A.: Okanogan Co., WA, Wooten 1326 (WS) EU001021, EU001170; ***Carex laevivaginata*** (Kükenth.) Mack., U.S.A.: Bradley Co., AR, Hyatt 8383 (WS) EU001023, EU001172; ***Carex nervina*** L.H.Bailey, U.S.A.: Sierra Co., NV, Janeway 3577 (WS) EU001036, EU001184; ***Carex neurophora*** Mack., U.S.A.: Trinity Co., CA, Ertter 7330 (WS) EU001038, EU001187; ***Carex oklahomensis*** Mack., U.S.A.: Cleburne Co., AR, Hyatt 9418 (WS) EU001192; ***Carex stipata*** Muhl. ex Willd., U.S.A.: Chelan Co., WA, Fishbein 3503 (WS) EU001071, EU001225.

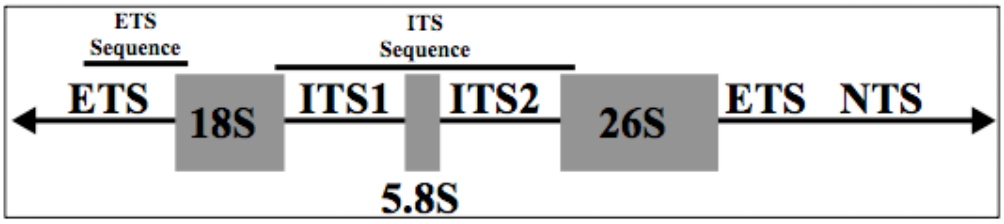
FIG. 1. Schematic representation of nuclear ribosomal DNA organizing region. From left to right: 5' external transcribed spacer. 18S ribosomal subunit, internal transcribed spacer 1, 5.8S ribosomal subunit, internal transcribed spacer 2, 26S ribosomal subunit, 3' external transcribed spacer, non-transcribed spacer.

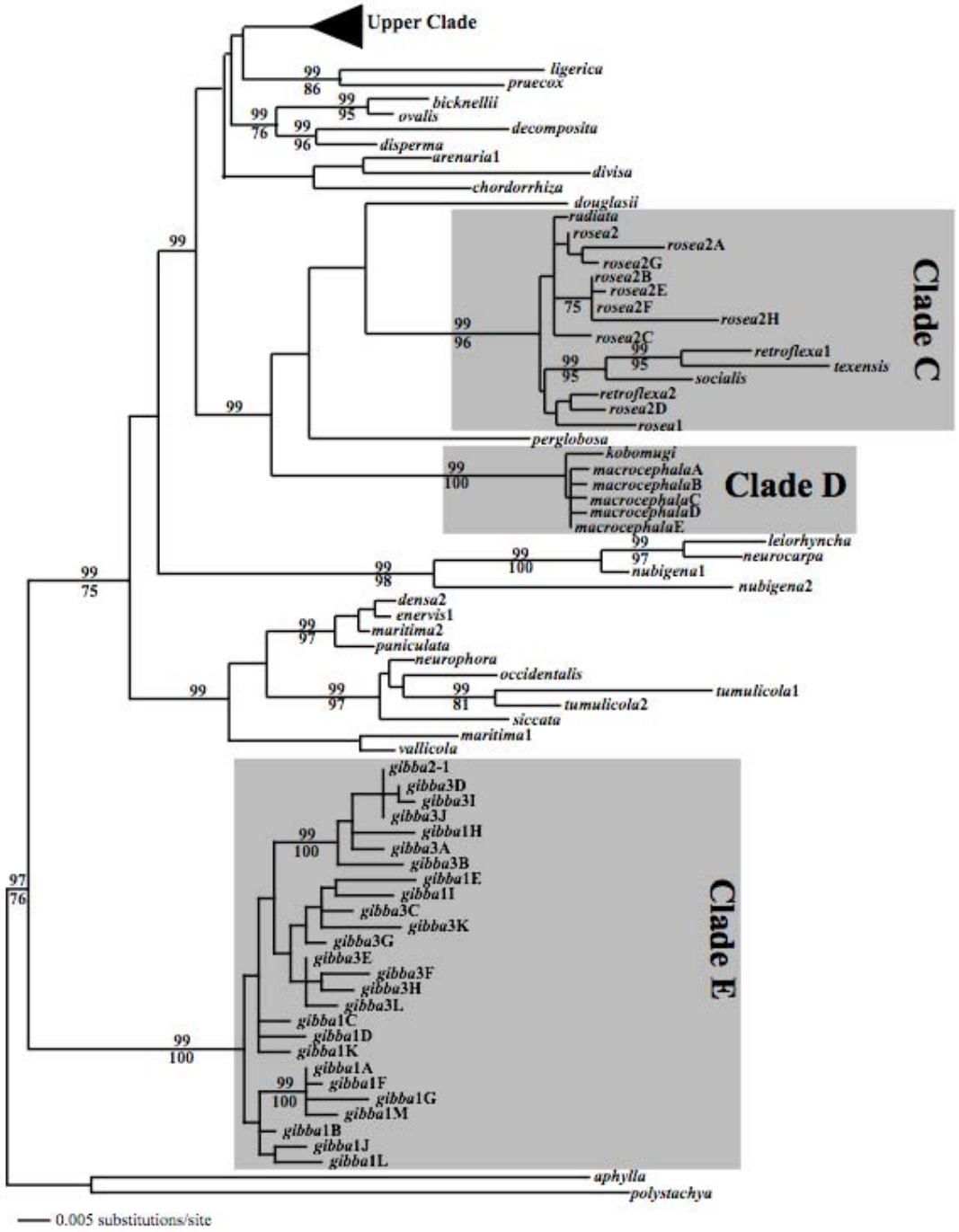
FIG. 2. The maximum likelihood phylogram of the lower grade of the complete ITS sequence (-lnL= 8381.4991). Numbers above branches are Bayesian posterior probabilities, and numbers below branches are non-parametric bootstrap percentages.

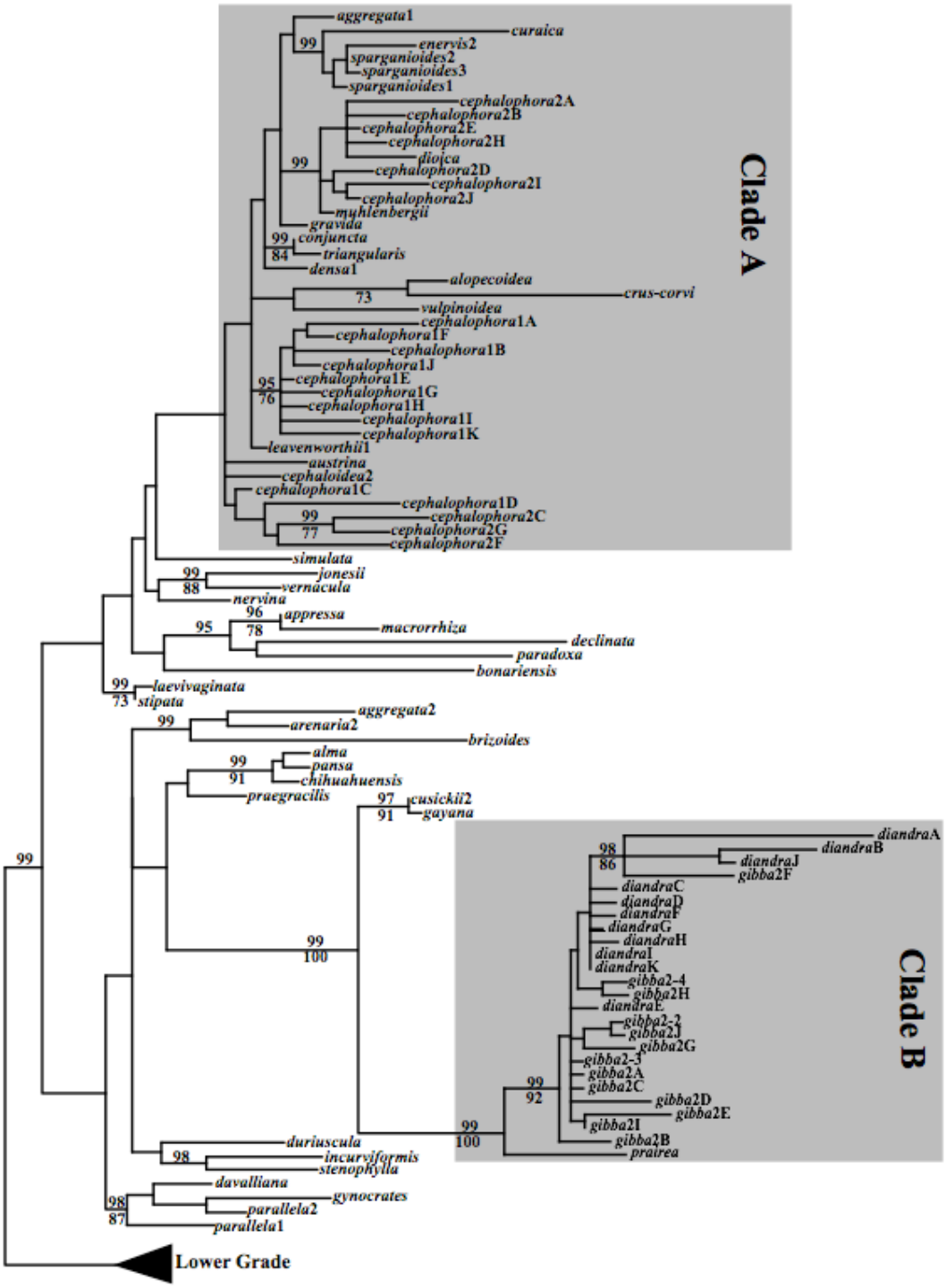
FIG. 3. The maximum likelihood phylogram of the upper clade of the complete ITS sequence (-lnL= 8381.4991). Numbers above branches are Bayesian posterior probabilities, and numbers below branches are non-parametric bootstrap percentages.

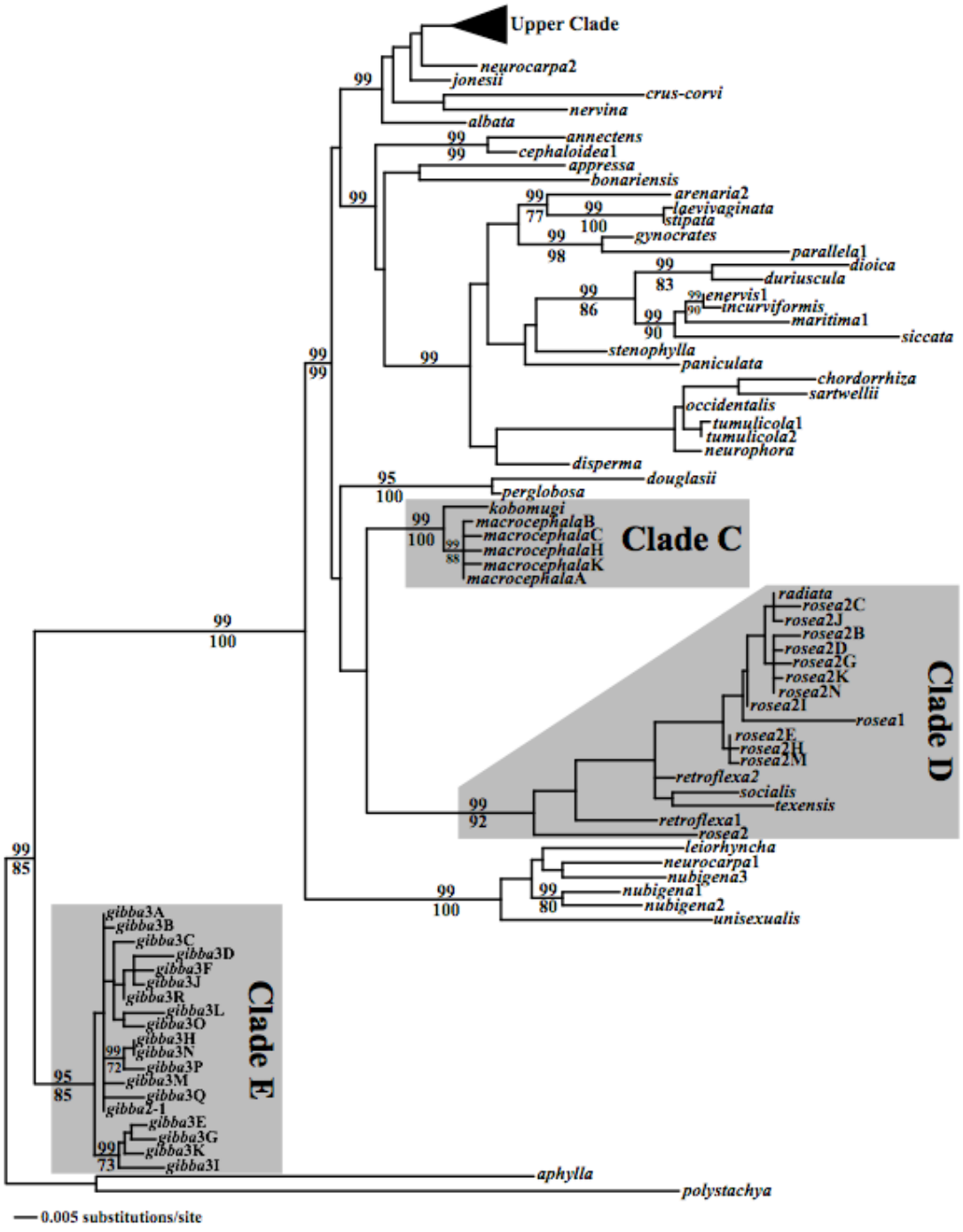
FIG. 4. The maximum likelihood phylogram of the lower grade of the ETS1f sequence (-lnL=6878.48453). Numbers above branches are Bayesian posterior probabilities, and numbers below branches are non-parametric bootstrap percentages.

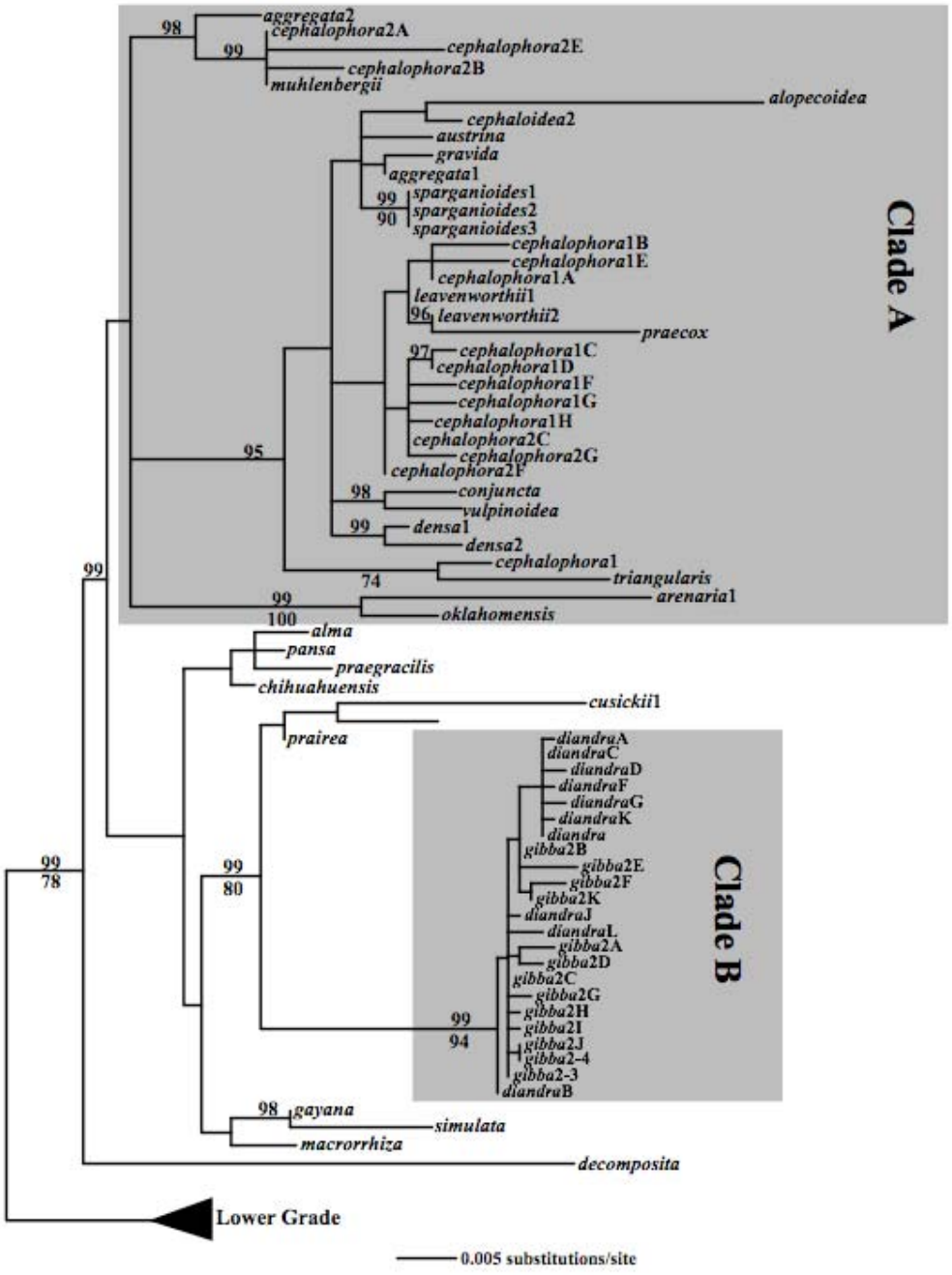
FIG. 5. The maximum likelihood phylogram of the upper clade of the ETS1f sequence (-lnL=6878.48453). Numbers above branches are Bayesian posterior probabilities, and numbers below branches are non-parametric bootstrap percentages.


ETS
Sequence
ITS
Sequence
ETS
18S
ITS1
ITS2
26S
ETS
NTS
5.8S


Upper Clade
ligerica
praecox
bicknellii
ovalis
decomposita
disperma
arenaria1
divisa
chordorrhiza
douglasii
radiata
rosea2
rosea2A
rosea2G
rosea2B
rosea2E
rosea2F
rosea2H
rosea2C
retroflexa1
texensis
socialis
retroflexa2
rosea2D
rosea1
Clade C
perglobosa
kobomugi
macrocephalaA
macrocephalaB
macrocephalaC
macrocephalaD
macrocephalaE
Clade D
leiorhyncha
neurocarpa
nubigena1
nubigena2
densa2
enervis1
maritima2
paniculata
neurophora
occidentalis
tumulicola1
tumulicola2
siccata
maritima1
vallicola
gibba2-1
gibba3D
gibba3I
gibba3J
gibba1H
gibba3A
gibba3B
gibba1E
gibba1I
gibba3C
gibba3K
gibba3G
gibba3E
gibba3F
gibba3H
gibba3L
gibba1C
gibba1D
gibba1K
gibba1A
gibba1F
gibba1G
gibba1M
gibba1B
gibba1J
gibba1L
Clade E
aphylla
polystachya
0.005 substitutions/site


Clade A
Clade B
Lower Grade
aggregata1
curaica
enervis2
sparganioides2
sparganioides3
sparganioides1
cephalophora2A
cephalophora2B
cephalophora2E
cephalophora2H
diojca
cephalophora2D
cephalophora2I
cephalophora2J
muhlenbergii
gravida
conjuncta
triangularis
densa1
alopecoidea
crus-corvi
vulpinoidea
cephalophora1A
cephalophora1F
cephalophora1B
cephalophora1J
cephalophora1E
cephalophora1G
cephalophora1H
cephalophora1I
cephalophora1K
leavenworthii1
austrina
cephaloidea2
cephalophora1C
cephalophora1D
cephalophora2C
cephalophora2G
cephalophora2F
simulata
jonesii
vernacula
nervina
appressa
macrorrhiza
declinata
paradoxa
bonariensis
laevivaginata
stipata
aggregata2
arenaria2
brizoides
alma
pansa
chihuahuensis
praegracilis
cusickii2
gayana
diandraA
diandraB
diandraJ
gibba2F
diandraC
diandraD
diandraF
diandraG
diandraH
diandraI
diandraK
gibba2-4
gibba2H
diandraE
gibba2-2
gibba2J
gibba2G
gibba2-3
gibba2A
gibba2C
gibba2D
gibba2E
gibba2I
gibba2B
prairea
duriuscula
incurviformis
stenophylla
davalliana
gynocrates
parallela2
parallela1


Upper Clade
neurocarpa2
jonesii
crus-corvi
nervina
albata
annectens
cephaloidea1
appressa
bonariensis
arenaria2
laevivaginata
stipata
gynocrates
parallela1
dioica
duriuscula
enervis1
incurviformis
maritima1
siccata
stenophylla
paniculata
chordorrhiza
sartwellii
occidentalis
tumulicola1
tumulicola2
neurophora
disperma
douglasii
perglobosa
kobomugi
macrocephalaB
macrocephalaC
macrocephalaH
macrocephalaK
macrocephalaA
Clade C
radiata
rosea2C
rosea2J
rosea2B
rosea2D
rosea2G
rosea2K
rosea2N
rosea2I
rosea1
rosea2E
rosea2L
rosea2M
retroflexa2
socialis
texensis
retroflexa1
rosea2
Clade D
leiorhyncha
neurocarpa1
nubigena3
nubigena1
nubigena2
unisexualis
gibba3A
gibba3B
gibba3C
gibba3D
gibba3F
gibba3J
gibba3R
gibba3L
gibba3O
gibba3H
gibba3N
gibba3P
gibba3M
gibba3Q
gibba2-1
gibba3E
gibba3G
gibba3K
gibba3I
Clade E
aphylla
polystachya
0.005 substitutions/site


Clade A
Clade B
Lower Grade
0.005 substitutions/site
aggregata2
cephalophora2A
cephalophora2E
cephalophora2B
muhlenbergii
alopecoidea
cephaloidea2
austrina
gravida
aggregata1
sparganioides1
sparganioides2
sparganioides3
cephalophora1B
cephalophora1E
cephalophora1A
leavenworthii1
leavenworthii2
praecox
cephalophora1C
cephalophora1D
cephalophora1F
cephalophora1G
cephalophora1H
cephalophora2C
cephalophora2G
cephalophora2F
conjuncta
vulpinoidea
densa1
densa2
cephalophora1
triangularis
arenaria1
oklahomensis
alma
pansa
praegracilis
chihuahuensis
cusickii1
prairea
diandruA
diandraC
diandraD
diandraF
diandraG
diandraK
diandra
gibba2B
gibba2E
gibba2F
gibba2K
diandraJ
diandraL
gibba2A
gibba2D
gibba2C
gibba2G
gibba2H
gibba2I
gibba2J
gibba2-4
gibba2-3
diandraB
gayana
simulata
macrorrhiza
decomposita

CHAPTER 2

DISCORDANCE BETWEEN PHYLOGENETICS AND COALESCENT-BASED DIVERGENCE MODELING: EXPLORING PHYLOGEOGRAPHIC PATTERNS OF SPECIATION IN THE *CAREX MACROCEPHALA* SPECIES COMPLEX

Phylogeographic analyses that rely on phylogenetic based methods often take for granted the certainty of the relationships within that phylogenetic hypothesis. Sister relationships of well-supported clades are often unquestioned especially when taxonomic sampling within those clades is high, and why not? When multiple individuals of each taxonomic unit are supporting the same congruent explanation of species/population evolution, then there may be no need to further question the results garnered from these phylogenetic based analyses. There is an exhaustive amount of literature on error rate associated with phylogenetic hypotheses, as well as the flaws associated with each phylogenetic criterion (Felsenstein 2004; Sullivan 2005). Researchers are free to pick the phylogenetic method of their choosing along with an appropriate method to determine support. In the recent past many researchers would have supported the idea that if multiple phylogenetic methods suggest the same relationship then we cannot reject that hypothesis. Additionally, statistical means have been employed to test such phylogenetic and phylogeographic relationships which has lent credence to the argument that these methods must be analytically sound (Knowles and Maddison 2002; Knowles 2004; Templeton 2004).

Congruence between the population tree and the gene tree is assumed when using the phylogenetic based methods for reconstructing the evolutionary history within or between closely related species. A further assumption is that shared alleles between divergent populations are the result of incomplete lineage sorting or persistent migration (Maddison and Knowles 2006). When there are no shared alleles between diverging populations it is important to establish whether the alleles in each population belong to reciprocally monophyletic clades in a phylogenetic tree. If the alleles in each population

belong to a separate reciprocally monophyletic clade it can be assumed that migration has ceased for a period of time long enough for lineage sorting to complete for a given gene within each population. Reciprocally monophyletic populations are useful to researchers conducting phylogeographic analyses, as the evolutionary history of the species can be inferred from a single phylogenetic hypothesis (Brunsfeld et al. 2001; Nielson et al. 2001; Demboski and Sullivan 2003; Carstens et al. 2005).

Phylogeographic methodologies typically do not undercut a phylogenetic hypothesis with well-resolved clades that correspond to geographically delimited populations, especially when comparative analyses are undertaken (Brunsfeld et al. 2001; Carstens et al. 2005). However, we must begin to question these methods if we are to understand the forces that influence the divergence process. The inability of phylogenetics to consider the stochastic forces at work within populations is a drawback of the phylogeographic analyses that rely on phylogenetics alone. As an alternative to these phylogenetic based methods researchers are able to employ coalescent methodologies which can account for some of these stochastic forces driving population divergence. These coalescent methodologies are able to accomplish this through simulations by treating the genealogy as a nuisance parameter, one where the topology of the gene tree is not the most important aspect of the estimation; and treating the population parameters as quantities that can be estimated (Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey and Nielsen 2007). The degree of population divergence is influenced by, but not limited to, the time since the divergence began (t), the effective population size (N_e), and migration (m), and these have been the focal parameters of divergence estimates (Dolman and Moritz 2006; Knowles and Carstens 2007).

It is not only necessary to consider estimate population forces when conducting tests of divergence time, but to also consider variance. By accounting for the coalescent variance we are accounting for the stochastic nature of lineage sorting not only in the descendant populations, but in the ancestral population as well. The coalescent variance coupled with the mutational variance, and stochastic variances in sampling of individuals and genes incurs the variance upon each parameter estimate. For this reason, researchers have proposed a multilocus approach to reduce error associated with lineage sorting of a single locus (Beerli and Felsenstein 2001; Hey and Machado 2003; Hey and Nielsen 2004; Dolman and Moritz 2006; Knowles and Carstens 2007).. Furthermore the variance in coalescent times between loci are informative and useful when modeling population divergence (Edwards and Beerli 2000; Hey and Nielsen 2004). We have fit the isolation with migration model of lineage divergence to a molecular dataset consisting of the *rpL16* spacer and 8 microsatellite loci. This will allow use to reconstruct the phylogeographic history of the *Carex macrocephala* species complex (Cyperaceae) and investigate the applicability of using phylogenetic-based phylogeographic analyses. A phylogenetic hypothesis suggests a complex history among the main lineages within the species complex (Fig. 1), and it is this hypothesis we a testing.

Carex macrocephala is restricted to coastal sand dunes and sandy beaches, and its native distribution ranges from mid-Oregon to Alaska and the Aleutian Islands, across the Bering Sea to the Kamchatka Peninsula, south to Sakhalin Island, the Kuril Islands, Russian coast in the Sea of Okhotsk, and the northern portion of Hokkaido Island of Japan (Fig. 2). The closely related *Carex kobomugi* is also restricted to coastal sand dunes and sandy beaches, but has a smaller native range of the islands of Japan, Korean coast,

and the Russian coast of the Sea of Japan (Mastogiuseppe 2002). Phylogenetic analyses using the *rpL16* spacer define three lineages within this complex, North American *C. macrocephala* (NACM), Asian *C. macrocephala* (ACM), and *C. kobomugi* (CK) (Fig. 1). According to the phylogenetic hypothesis, the Asian lineage of *C. macrocephala* is sister to a clade composed of a monophyletic lineage of North American *C. macrocephala* and a monophyletic *C. kobomugi*. Not only does this infer that *C. macrocephala* is paraphyletic with respect to *C. kobomugi*, but the two lineages that are geographically the most distant and that have no sympatric populations share the most recent common ancestor of the three lineages. The phylogenetic hypothesis suggests either complex evolutionary history of these lineages, or the hypothesis is incorrect and further analyses which account for population level processes are needed.

To test the relationships between the three main lineages of the *Carex macrocephala* species complex we will employ the 'Isolation with Migration' (IM) model of population divergence (Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey and Nielsen 2007). Marginal and joint estimates of divergence time (*t*) are possible through this model while simultaneously estimating migration and effective population size. We will attempt to fit the isolation with migration model to our cpDNA *rpL16* and microsatellite dataset. By doing so we can make pairwise comparisons of the main lineages within the *Carex macrocephala* species complex which will allow us to test the following: 1) determine the accuracy of the cpDNA *rpL16*-based phylogenetic hypothesis by comparing the relative divergence times of each pairwise analysis; 2) estimate the divergence time, in years (T), of each of the main lineages of the *Carex macrocephala* species complex to determine the approximate times of divergence between the

populations in Asia and North America; and 3) determine relative ancestral effective population size (N_{eA}) estimates versus the extant lineages, as these estimates can help infer the method of divergence between the lineages (i.e., dispersal versus vicariance).

METHODS

Sampling

We sampled 227 individuals from 40 localities of *C. macrocephala* from North America (Fig. 2; App. A, B) for the *rpL16* intergenic spacer. Of these 227 individuals, 96 were chosen from across the range to genotype for the microsatellite loci. These samples were collected from locations across the North American range, and immediately stored in silica gel. Samples were selected at random within populations, but spread as far away from each other as possible to minimize the possibility of sampling identical genets. Fourteen individuals of *Carex macrocephala* were sampled from across the Asian population range. These samples were obtained from herbarium specimens obtained on loan from Hokkaido Univeristy [HAK], Kyoto University [KYO], Univeristy of Tokyo [TH, TI, TOFO], National Science Museum of Japan [TNS], Tokushima Prefectural Museum [TKPM] and University of Washington [WTU], and the species identity verified. Fourteen specimens of *C. kobomugi* were sampled from across the native range from herbarium sheets as with the samples of *C. macrocephala*. A further three species of *Carex* were sampled to use as outgroups *C. oklahomensis*, *C. crus-corvi*, and *C. rosea*. These species were selected based on their relative genetic distance from the *C. macrocephala* clade in previous phylogenetic analyses (King and Roalson *in press*).

Molecular Markers

DNA was extracted from 30mg of silica gel dried or herbarium material using a modified 2x CTAB protocol (Roalson et al. 2001). Using protocols described in Shaw et al. (2005), we were able to sequence all the cpDNA markers in that study for 25 specimens of *C. macrocephala* and 5 specimens of *C. kobomugi*. Of the 21 markers outlined in Shaw et al., only the *rpL16* spacer showed polymorphism within *C. macrocephala* and between *C. macrocephala* and *C. kobomugi*. PCR protocols for *rpL16* can be found in Shaw et al. (2005), and PCR products were readied for cycle sequencing by incubating with 2 units of Exonuclease I (New England Biolabs, Inc.) and 3 units of Antarctic phosphotase (New England Biolabs, Inc.) at 37°C for 1 hour followed by a 15 minute heat step at 75°C to inactivate the enzymes. Sequencing was performed using an Applied Biosystems 3730 Automated DNA Sequencer. Cycle sequencing of PCR products followed ABI's protocol for Big Dye Terminators version 3. Sequences were obtained for the 3' and 5' strands of the PCR products followed by their assembly into contigs and edited and visually verified using Sequencher v4.6 (GeneCodes Corp.). Sequences are deposited in GenBank and phylogenetic tree datasets are deposited in TreeBase.

Microsatellite loci Cko1-12, Cko1-68, Cko1-78, and Cko1-134 were described by Ohsako and Yamane (2007) and isolated from C. kobomugi; and loci CM01, CM07, CM27, and CM39 were described by King and Roalson (in prep) and isolated from *C. macrocephala*. The loci were chosen because they are all dinucleotide repeats, that show a relatively high level of diversity, and are all in linkage equilibrium for each lineage.

Amplification of these markers followed the protocols outlined in the respective papers, and amplified products were multiplexed at a final dilution ratio of 1:40. Fragment sizes were obtained on an Applied Biosystems 3730 and scored using GENEMAPPER v. 3.7 (Applied Biosystems).

Phylogenetic analyses

All *rpL16* sequences were manually aligned in Se-Al version 2.0.9 (Rambaut 1996). A model of nucleotide substitution was selected using the Perl script DT-ModSel (Minin et al., 2003) and this model was used to determine the maximum likelihood phylogeny using PAUP* 4.0b10 (Swofford 2003). As a means to explore whether different models of phylogenetic substitution would change the topology, we also ran phylogenetic analyses using all 56 common models of nucleotide substitution (Minin et al., 2003). Furthermore, parsimony and neighbor joining analyses were also conducted using PAUP* to determine whether different phylogenetic criteria would alter the topology. A network was constructed using TCS (Clement et al. 2000) to determine whether the statistical parsimony method would reconstruct a network with connections that are different from those found in the maximum likelihood analysis. Nonparametric bootstrap analyses were conducted to assess nodal support using the maximum likelihood criterion with the model chosen by DT-ModSel. We ran 500 bootstrap replicates each with 2 random addition replicates saving 1 tree from each bootstrap replicate. Further support was assessed under a Bayesian framework with MrBayes (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003; Altekar et al. 2004) to generate posterior probabilities for each node. Two separate 25 million generation, 8 chain analyses were

run to assess convergence and mixing and a total of 10 000 topologies were saved from each run. MrBayes run commands of sumt and sump assessed that these runs were reaching stationarity with proper convergence.

Population genetic indices

Summary population genetic indices for the *rpL16* dataset were computed using Arlequin v. 3.11 (Excoffier et al. 2005). The number of segregating sites (S_{NACM}, S_{ACM}, and S_{CK}) within each lineage and for the overall dataset (S_{total}), the mean number of pairwise nucleotide differences (π) for each lineage and the entire dataset (π_{NACM}, π_{ACM}, π_{CK}, π_{total}), and θ based on the number of segregating sites (θ_S) were also computed for each lineage and the overall sample. Arlequin was also used to calculate an Fst value for the *rpL16* dataset in North America. Using the microsatellite dataset we calculated an overall Fst value for North America using Arlequin and pairwise Fst values for each of the main lineages. We also computed linkage disequilibrium values for all microsatellite loci.

Isolation with migration model of population divergence

To determine whether the phylogenetic hypothesis is accurate, and to estimate *t* of the main lineages of the *Carex macrocephala* species complex we used the implementation of the isolation with migration model present in the program IMa (Hey and Nielsen 2007). IMa can obtain marginal posterior probability densities of the population parameters of divergence time, *t*, where $t = t\mu$, and where t is time in years and μ is the mutation rate; population differentiation indices of the ancestral population and the two extant lineages, θ_A, θ_1, and θ_2, respectively, where $\theta_x = N_{ex}\mu$ for uniparentally inherited loci, where N_e is the effective population size; and the migration

rate for each population, m_1 and m_2, where $m_x = m_x/\mu$, where $Mx = N_{ex}m_x = \theta_x m_x$. Pairwise analyses were conducted for each of the three main lineages to test whether divergence times and population parameter estimates changed significantly from what would be expected according to the phylogenetic hypothesis.

Priors used in IMa analyses were as follows: NACM-CK: q1= 5, q2= 5, qA= 15, m1= 2, m2= 2, t= 10; ACM-CK: q1= 5, q2= 5, qA= 15, m1= 2, m2= 2, t=10; and NACM-ACM: q1= 5, q2= 5, qA= 15, m1= 10, m2=10, t= 10. These priors were established after running 10 separate 2.5 million steps runs with a single chain for each analysis followed by three separate 15 million steps runs with 2 chains to assess convergence. Final runs were conducted with 4 chains and run for 25 million steps after a burn-in period of 5 million steps saving a total of 250,000 genealogies. We ran these analyses twice to ensure convergence of parameter estimates. IMa analyses were run until the peak location set values were equal for set 1 and set 2. This implies the Markov chains reached convergence after a sufficiently long burn-in period, and where sampling from the appropriate likelihood space. We applied the HKY model of nucleotide substitution for the *rpL16* locus in IMa as this model was chosen for the phylogenetic analysis by DT-ModSel, and the SSM model for the microsatellite loci.

Joint estimates of parameters are possible in the implementation of the isolation with migration model in the program IMa (Hey and Nielsen 2007). IMa generates a series of genealogies on which model parameters are estimated by their likelihood. By averaging over a large enough sample of genealogies an average estimate for each parameter can be obtained by weighting those genealogies by their probability given the data. Although *t* is not part of the joint posterior probability density function, these values

can be thought of as a value derived from the joint posterior probability distribution. By averaging *t* across all the genealogies by their probability we can obtain a weighted estimate of *t*. Choice of prior estimates of *t* are important, and it was for this reason we ran several preliminary analyses to determine the appropriate prior.

We have estimated a mutation rate, $\mu = 9.41 \times 10^{-9}$ site^{-1} year^{-1} $\pm$ 9×10^{-1} for the *rpL16* data set by calibrating to the approximated dates of the formation of the beaches off the Columbia River littoral cell (Twichell and Cross 2002). By identifying haplotypes found only on these beaches we were able to calibrate the possible date of those nodes, and thereby determine a possible mutation rate. This mutation rate falls well within the previously described mutation rates for cpDNA markers (Wolfe et al. 1987; Brunsfeld and Sullivan 2006; Ann et al. 2007). Furthermore we used a mutation rate range of 3×10^{-5} locus^{-1} year^{-1} to 6×10^{-4} locus^{-1} year^{-1} with a mean of 2×10^{-4} locus^{-1} year^{-1} for the microsatellite loci. This range was chosen based on microsatellite mutation rates for dinucleotide repeats in monocots (Thuillet et al. 2002; Vigouroux et al. 2002).

RESULTS

rpL16 phylogenetic hypotheses

Within the 227 sampled individuals of *C. macrocephala* from North America, 57 different haplotypes were sequenced at a total aligned sequence length of 748 base pairs. Each haplotype was assigned an arbitrary two-letter code based on a preliminary placement within a neighbor joining tree. Of the 14 individuals of Asian *C. macrocephala*, 8 haplotypes were recovered, and for the 14 individuals of *C. kobomugi*, 8 haplotypes were also recovered with a total length of 756 base pairs. DT-ModSel selected

the HKY + I + G model of nucleotide substitution with a transition/transversion ratio of 0.83815359, a proportion of invariant sites of 0.76811202, and a gamma shape parameter of 0.665549. Bases frequencies were estimated to be A: 0.42591376, C: 0.13757311, G: 0.15051696, and T: 0.28599617. The maximum likelihood phylogeny is shown in Figure 1 with maximum likelihood bootstrap and Bayesian posterior probabilities shown on branches. The populations of North American *Carex macrocephala* are reciprocally monophyletic and are sister to a reciprocally monophyletic clade of *Carex kobomugi*. This relationship would infer a paraphyletic *C. macrocephala,* with the Asian samples of *C. macrocephala* coming out as sister to the NA *C. macrocephala – C. kobomugi* clade. All these relationships are well supported, and the support of a monophyletic ingroup is high at 98% bootstrap and >99% Bayesian posterior probability. Regardless of the model of nucleotide substitution chosen for the maximum likelihood analysis, and regardless of the phylogenetic criterion used for the reconstruction, the relationships among NACM, CK, and ACM did not change. Nor did this relationship change with the alternate treatment of the gaps within the datasets. Neither treating gaps as missing data or coding gaps a 5^{th} character state in parsimony analysis changed the relationships among the NACM, CK, and ACM clades.

The network constructed using TCS at 98% parsimony is shown in Figure 3. Parsimony reconstruction at the 98% level was necessary due to the number of reconstructions that were possible below that limit. However, an alternate relationship of the CK clade was not made until the parsimony reconstruction limit was set below 93%, and this also allowed 18 alternate placements of the CK clade within the network. This is further evidence that the relationships among the lineages, according to the *rpL16* dataset,

suggest CK and NACM are sister to each other. When gaps were treated as a 5th character state the CK clade was only connected to the network at the 92% reconstruction limit and below. *Carex crus-corvi* failed to connect to the network at these reconstruction limits.

Population genetic summary statistics

There were 59 segregating sites (S_{total}) for the entire *rpL16* dataset, with S_{NACM} = 40, S_{ACM} = 12, S_{CK} = 7. Mean pairwise nucleotide difference values were π_{total} = 5.14052, π_{NACM} = 3.35739, π_{ACM} = 3.74725, π_{CK} = 2.109, and the population diversity indices, θ, based on segregating sites were $\theta_{(S)total}$ = 10.922, $\theta_{(S)NACM}$ =7.7718, $\theta_{(S)ACM}$ = 3.7734 $\theta_{(S)CK}$ = 2.2011. An insignificant mean F_{ST} value of 0.04772 ± 0.056 was computed for the *rpL16* samples from across North America. Fixation indices suggest a similar pattern for the microsatellite datasets with an F_{ST} value of -0.023264 ± 0.031 for all of NACM and pairwise values of $F_{ST(NACM-ACM)}$ = 0.1277 p = 0.003, $F_{ST(ACM-CK)}$ = 0.4286 p < 0.0001, $F_{ST(NACM-CK)}$ = 0.5872 p < 0.0001. None of the loci showed significant linkage disequilibrium with the smallest p-value = 0.242.

Lineage divergence

Multidimensional peak locations of divergence time estimates are $t_{NACM-ACM}$ = 1.542, t_{ACM-CK} = 1.9802, and $t_{NACM-CK}$ = 4.7336 (Table 1, Fig. 4). All multidimensional peak locations fall within the 95% confidence intervals of the marginal posterior density distributions (Table 1, Fig. 4). The geometric mean of mutation rates is 1.23 x10^{-5} locus^{-1} year^{-1}.

Population parameter estimates

Migration rate estimates are effectively zero for migration between NACM-CK and ACM-CK, but migration estimates were higher for the NACM to ACM divergence, $m_{NACM-ACM} = 0.145$ and $m_{ACM-NACM} = 7.0664$ (Table 1). Population parameter estimates obtained from IMa can be given in terms of marginal posterior densities, or from multidimensional peak estimates. Here we provide the multidimensional peak estimates for θ, the population size scaled by mutation rate, from IMa for each of the pairwise analyses: NACM to ACM we obtained a value of $\theta_{NACM} = 1.8133$, $\theta_{ACM} = 1.336$, and $\theta_A = 5.764$; ACM to CK a value of $\theta_{ACM} = 1.7207$, $\theta_{CK} = 1.7052$, and $\theta_A = 8.6269$; and NACM to CK a value of $\theta_{NACM} = 2.5843$, $\theta_{CK} = 3.117$, and $\theta_A = 6.3575$ (Table 1). All multidimensional peak locations fall within the 95% confidence interval of the marginal posterior density distribution.

DISCUSSION

Phylogenetic versus isolation with migration coalescent models

There is strong phylogenetic support using the cpDNA *rpL16* spacer for the North American lineage of *C. macrocephala* and *C. kobomugi* being more closely related to one another than either is to the Asian populations of *C. macrocephala*. According to this hypothesis we might infer that there were complex evolutionary scenarios occurring during the splitting of these lineages. One possible explanation would be an early range expansion of one lineage into a southern Asian distribution, followed quickly by a dispersal and expansion into North America, all the while the main lineage of *C. macrocephala* in Asia was in isolation. Although this is only one possible explanation, it does not consider another possible explanation that stochastic forces at work within these

populations have left an alternate genetic signature. An alternative hypothesis to the phylogenetic hypothesis is that the phylogenetic patterns we see in the *rpL16* spacer are artifacts of lineage sorting, and not a true representation of the relationships within this group. This is most often a case of incomplete lineage sorting, and much effort has been devoted towards overcoming these obstacles (Carstens and Knowles 2007). The results of our coalescent analyses suggest an alternate hypothesis as to the relationships between the lineages within the *C. macrocephala* species complex. According to the IMa results, the oldest divergence times were between NACM and CK, and the most recent divergence time between NACM and ACM. Phylogenetic analyses examine only gene divergence with which researchers infer lineage relationships. In this study we fit our molecular dataset to the isolation with migration model, which does account for lineage divergence by treating the gene tree as a nuisance parameter, and also by estimating parameters that influence divergence time. The discordance between the gene tree and the coalescent hypothesis seen in this example is a unique form of discordance due to lineage sorting, because typically incomplete lineage sorting is the cause of any discordance (for other examples see Maddison 1997).

The power of these coalescent based analyses is that they consider the lineage sorting which began prior to split of the populations. Gene divergence (*D*) begins prior to population divergence, and the coalescent variance generated in this period is important in approximating *t* and the ancestral N_e (Edwards and Beerli 2000). Phylogeographic analyses which rely soley on a phylogenetic tree to make inferences of divergence time ignore the time difference between gene divergence and population divergence. Only when $Ne \approx 0$ or when t/N_e is large can we assume $D = t$ (Edwards and Beerli 2000). This

may be the discrepancy in our phylogenetic and coalescent-based divergence analyses. Population subdivision may affect the results of our coalescent analyses, but based on an overall estimated Fst values for the NACM lineage of 0.04772 for the *rpL16* samples and -0.023264 for the microsatellite dataset there is no apparent population subdivision (Wakeley 2000).

Phylogeographic hypothesis

The divergence between NACM and CK is the oldest lineage divergence inferred by the coalescent analyses. By fitting our dataset to the isolation with migration model we have assumed that gene flow and divergence is only occurring between the two lineages being examined. In actuality this is not the case, as there are quite clearly three lineages diverging from one another through time. This can make inferences difficult unless we make a further assumption. We need to assume that the oldest divergence event, NACM-CK, is an artifact of the other two divergence events, NACM-ACM and CK-ACM. This would infer that the NACM and CK lineages did not directly diverge from each other. This is a direct contradiction to the phylogenetic hypothesis, where we would assume NACM and CK share the most recent common ancestor. The IMa coalescent analyses suggests the most recent common ancestor for both the NACM and CK lineages is shared through ACM and not with each other, and by conducting the pairwise analysis of NACM and CK we violate the assumptions of the isolation with migration model. This is the reason for the large discrepancy in divergence time estimates between the NACM-CK comparison and the ACM-CK and NACM-ACM comparisons. We will therefore only make inferences using the divergences between NACM –ACM and ACM-CK.

If ACM and CK diverge followed by NACM and ACM we can infer a possible evolutionary history of these lineages. We can use the estimates of θ to attempt to reconstruct a possible explanation of early evolution within this species complex. Dolman and Moritz (2006) noted that population subdivision may inflate estimates of θ, and thereby inflate estimates of N_e. An insignificant F_{ST} value of 0.04772 for the *rpL16* dataset and an insignificant F_{ST} value of -0.023264 for the microsatellite data set suggest that very little population subdivision occurs along the west coast of North America. This low F_{ST} may be due to high levels of gene flow among the North American populations. Furthermore, the F_{ST} comparisons suggest stronger fixation between NACM and CK , F_{ST} = 0.5872, than between NACM and ACM, F_{ST} = 0.1277.

In all pairwise comparisons the ancestral effective population sizes were larger than the total current estimated effective population size of the two lineages. This infers a reduction in N_e greater than would be expected for a vicariance event, where in vicariance $N_{e1} + N_{e2} = N_{eA}$. By removing the geometric mean of mutation rates we can estimate the effective population size, $N_{e(NACM)} \approx 147{,}000$, $N_{e(ACM)} \approx 109{,}000$ to $140{,}000$, $N_{e(CK)} \approx 138{,}000$. Based on extensive population observations, this estimate of N_e for NACM seems reasonable.

Migration only persisted for the NACM-ACM lineages and gene flow is mostly occurring from ACM to NACM. Furthermore, there are no shared haplotypes between NACM and ACM, but there are shared microsatellite alleles suggesting that gene flow persisted only through pollen, and not through seeds. However, this could also be an artifact of incomplete lineage sorting within these markers.

By rescaling by the mutation rate from *t* we can estimate divergence time in years, T, for the NACM-ACM divergence as 125,000 years with a 95% confidence interval of 103,000 years to 163,000 years; and 160,000 years for the ACM-CK divergence with a 95% confidence interval of 75,000 years to 296,000 years. This suggests a divergence for all lineages, which corresponds to the last inter-glacial warming period prior to the onset of deep glaciation approximately 110,000 yeas ago (Lisiecki, Raymo, 2007). Additionally, there is paleoclimatic evidence to suggest sea level was at least 6m higher than present levels during the last interglacial, which may have contributed to the subsequent isolation of populations along the northwest North American coast (Brigham-Grette and Hopkins 1995). The $N_{e(A)}$ of the NACM-ACM lineage is 4.83x the $N_{e(NACM)}$ and 6.56x the $N_{e(ACM)}$. This suggests a large reduction in N_e for both lineages and in these cases may have been caused by cycles of Pleistocene glaciation, where populations of *C. macrocephala* and *C. kobomugi* would have been repeatedly reduced followed by expansion caused by the rising and lowering of sea level and movement of the continental ice sheets.

The divergence time estimates obtained from IMa and the placement of the root in the phylogenetic analysis suggest the ancestral lineage existed in Asia. The divergence between *C. macrocephala* and *C. kobomugi* may have occurred through vicariance, as there are populations that occur in sympatry, and the $N_{e(ACM)} \approx N_{e(CK)}$ which also suggest divergence through vicariance. The divergence between ACM and NACM may have occurred through dispersal followed by range expansion along the North American coast. We can make this inference as the $N_{e(NACM)} > N_{e(ACM)}$ and the overall topology of the

genealogy suggests population growth; it is star shaped with short internal branches (Slatkin and Hudson 1991; Wakeley 2003).

Conclusions

In this paper we fit the isolation with migration model of lineage divergence to our molecular dataset to test the *rpL16* phylogeny. The results of our IMa analyses suggest a different pattern of relationships than that suggested by the phylogenetic hypothesis. The oldest divergence date between any of our pairwise comparisons is between the North American *C. macrocephala* and *C. kobomugi.* This is in direct contrast to the phylogenetic result where these two lineages would share the most recent common ancestor. Furthermore, the IMa results suggest lineages within *C. macrocephala* diverged more recently, approximately 125,000 years ago. As all phylogenetic methods employed here estimated the same evolutionary relationships with strong statistical support, we can conclude that this is not merely the result of phylogenetic uncertainty, but caused by underlying stochastic forces such as lineage sorting.

Here we have shown a case that in lieu of complete lineage sorting and a well-supported reciprocally monophyletic phylogenetic hypothesis, the coalescent based divergence time estimates suggest a different phylogeographic hypothesis. The use of coalescent simulations to fit our data to the isolation with migration model of lineage divergence allowed us to develop a less complex phylogeographic hypothesis than that estimated using phylogenetic approaches.

LITERATURE CITED

Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407-415.

Ann, W., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol. Biol. Evol. 24:90–101.

Beerli, P, and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. USA 98:4563-4568.

Brigham-Grette, J., and D.M. Hopkins. 1995. Emergent marine record and paleoclimate of the last interglaciation along the northwest Alaskan coast. Quaternary Research, 43:159-173.

Brunsfeld, S. J., and J. Sullivan. 2006. A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conservation Genetics 6:895-904.

Brunsfeld, S. J., J. Sullivan, D. E. Soltis, and P. S. Soltis. 2001. Comparative phylogeography of Northwestern North America: A synthesis. Pp. 319-339. *in* J. Silvertown and J. Antonovics, eds. Integrating ecological and evolutionary processes in a spatial context. Blackwell Science, Oxford.

Carstens, B. C., J. D. Degenhardt, A. S. Stevenson, and J. Sullivan. 2005. Accounting for coalescent stochasticity in testing phylogeographic hypotheses: Testing models of

Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. Mol. Ecol. 14:255-265.

Carstens, B. C, and L. L. Knowles. 2007. Shifting distributions and speciation: species divergence during rapid climate change. Mol. Ecol. 16:619-27.

Clement, M., D. Posada, and K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657-1660.

Demboski, J. R., and J. Sullivan. 2003. Extensive mtDNA variation within the yellow-pine chipmunk, *Tamias amoenus* (Rodentia: Sciuridae), and phylogeographic inferences for northwest North America. Mol. Phylo. Evol. 26: 389-408.

Dolman, G, and C. Moritz. 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). Evolution 60:573–582.

Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.11: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland.

Hey, J., and C. A. Machado. 2003. The study of structured populations-- new hope for a difficult and divided science. Nat. Rev. Genet. 4:535-543.

Hey, J, and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747-760.

Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA 104:2785–2790.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Knowles, L. L. 2004. The burgeoning field of statistical phylogeography. J. Evol. Biol. 17:1-10.

Knowles, L. L., and B.C. Carstens. 2007. Estimating a geographically explicit model of population divergence for statistical phylogeography. Evolution 61:477-493.

Knowles, L.L. and W. P. Maddison. 2002. Statistical phylogeography. Mol. Ecol. 11:2623-2635.

Lisiecki, L. E. and M. E. Raymo (2007). Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. Quaternary Science Reviews 26: 56-69.

Maddison, W. P. 1997. Gene trees in species trees. Syst Biol. 46:523-536.

Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21-30.

Mastroguiseppe, J. 2002. Flora of North America: North of Mexico. Volume 23 Magnoliophyta: Commelinidae (in part): Cyperaceae. ed. Flora of North America Editorial Committee. Oxford University Press. New York.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52:1-10.

Nielsen, R., and J. W. Wakeley. 2001. Distinguishing Migration from Isolation: an MCMC Approach. Genetics 158:885-896.

Nielson, M., K. Lohman, J. Sullivan. 2001. Phylogeography of the tailed frog (*Ascaphus truei*): Implications for the biogeography of the Pacific Northwest. Evolution 55:147-160.

Nordborg, M. 2001. Coalescent theory. Pp:179–212. *in* D. J. Balding, M. J. Bishop, and C. Cannings, eds, Handbook of Statistical Genetics, John Wiley & Sons, Inc., Chichester.

Ohsako, T., and K. Yamane. Isolation and characterization of polymorphic microsatellite loci in Asiatic Sand Sedge, *Carex kobomugi* Ohwi (Cyperaceae). Molecular Ecology Notes. Online Early Articles.

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/.

Roalson, E. H., J. T. Columbus, and E. A. Friar. 2001. Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and trnT-L-F (cpDNA) region sequences: assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. Syst. Bot. 26:318-341.

Ronquist, F., J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572-1574.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 2005. The Tortoise and the Hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92:142–166.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555-562.

Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogeny of alleles. Genetics 123:603-613.

Sullivan, J. 2005. Maximum-likelihood estimation of phylogeny from DNA sequence data. *in* E. Zimmer and E. Roalson, eds. Molecular Evolution: Producing the Biochemical Data, Part B. Methods in Enzymology. 395:757-779.

Swofford, D. L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland.

Templeton, A. R. (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. Mol. Ecol. 13:789-809.

Thuillet, A. C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille, T. Bataillon. 2002. Direct estimation of mutation rate for 10 microsatellite loci in Durum Wheat, *Triticum turgidum* (L.) Thell. *ssp durum* desf. Mol. Biol. Evol. 19:122-125.

Twichell D. C., and V. A. Cross. 2002. Holocene evolution of the southern Washington and northern Oregon shelf and coast: Geologic discussion and GIS data release. USGS Open-File Report 01-076. http://pubs.usgs.gov/of/of01-076/index.htm

Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis, J. S. C. Smith, and J. Doebley. 2002. Rate and pattern of mutation at microsatellite loci in Maize. Mol. Biol. Evol. 19:1251-1260.

Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. Evolution 54:1092-1101.

Wakeley, J. 2003. Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. Pp. 193-215 *in* R. Singh and M. Uyenoyama, eds. The Evolution of Population Biology. Cambridge University Press, Cambridge.

Wolfe, K.H., W. H. Li, and P.M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84:9054-9058.

TABLE 1. Results of fitting the *rpL16* cpDNA marker and 8 microsatellite loci to the isolation with migration model implemented in IMa to determine divergence times of the *C. macrocephala* species complex. Population size and time since divergence scaled by the mutation rate, θ and *t*, respectively, see Methods. Migration values between ACM-CK and NACM-CK are effectively zero. Peak values are from the Multidimensional peak locations; Mean, 95Low, and 95High represent the mean and 95% confidence limits from the marginal posterior probability distribution.

Pairwise analysis	Estimate	Population Size			Migration		Time
		θ_{NACM}	θ_{ACM}	θ_{A}	$m_{NACM-ACM}$	$m_{ACM-NACM}$	*t*
NACM-ACM	Peak	1.8133	1.336	8.764	0.145	7.0664	1.542
	Mean	1.8139	1.3964	8.6532	0.5542	5.9592	1.571
	95Low	1.384	1.2176	5.3719	0.055	2.385	1.275
	95High	2.4525	1.6527	13.1933	1.515	9.555	2.025
		θ_{ACM}	θ_{CK}	θ_{A}	m_{ACM-CK}	m_{CK-ACM}	*t*
ACM-CK	Peak	1.7207	1.7052	8.6269	0.01271	0.0001	1.9802
	Mean	1.7202	1.6959	8.0499	0.01001	0.0170	1.923
	95Low	1.3837	1.3662	5.3287	0.003	0.007	0.935
	95High	2.3627	2.1527	9.5647	0.0339	0.0557	3.6704
		θ_{NACM}	θ_{CK}	θ_{A}	$m_{NACM-CK}$	$m_{CK-NACM}$	*t*
NACM-CK	Peak	2.5843	3.117	6.3575	0.0001	0.011	4.7336
	Mean	2.6096	2.7972	6.4322	0.0198	0.021	4.8337

95Low	1.7709	1.5287	4.8784	0.001	0.001	2.8725
95High	3.6384	4.5862	7.1092	0.071	0.079	6.4125

Figure Legends.

FIG 1. Maximum likelihood phylogram of the *Carex macrocephala* species complex (-lnL = 1846.17868) with three outgroup species *C. oklahomensis*, *C. crus-corvi*, and *C. rosea*. Main lineages are comprised of Asian *C. macrocephala* (ACM), *C. kobomugi* (CK), and North American *C. macrocephala* (NACM). Haplotypes were arbitrarily assigned two letter codes based on placement within a neighbor joining tree. Numbers above branches are Bayesian posterior probabilities and numbers below branches are maximum likelihood bootstrap percentages.

FIG 2. Map of the northern Pacific Coast showing distribution and sampling localities of *C. macrocephala* and *C. kobomugi*. NACM designates locations sampled for North American *C. macrocephala,* ACM for Asian *C. macrocephala* , and CK for *C. kobomugi*.

FIG 3. TCS constructed 98% parsimony network of *C. macrocpehala* species complex and two outgroups *C. oklahomensis* and *C. rosea*. Gaps were treated as missing data, and shaded areas show the main lineages defined by the maximum likelihood phylogenetic topology. Bars represent one-step differences between haplotypes, and black diamonds represent unsampled haplotypes or those missing from the lineages. The third outgroup, *C. crus-corvi,* failed to connect at the 98% parsimony level.

FIG 4. Marginal histograms and the multidimensional peak estimates of time since divergence from our IMa analyses. Boxes represent the 95% marginal density distribution obtained from IMa of bin values for each pairwise comparison; NACM = North

American *C. macrocephala*, ACM = Asian *C. macrocephala*, and CK = *C. kobomugi*. Solid line indicates the mean bin value for each marginal distribution, and the triangles represent the multi-dimensional peak estimates obtained from IMa.

APPENDIX A. Sampling locations from North America with the number of individuals sampled at each location.

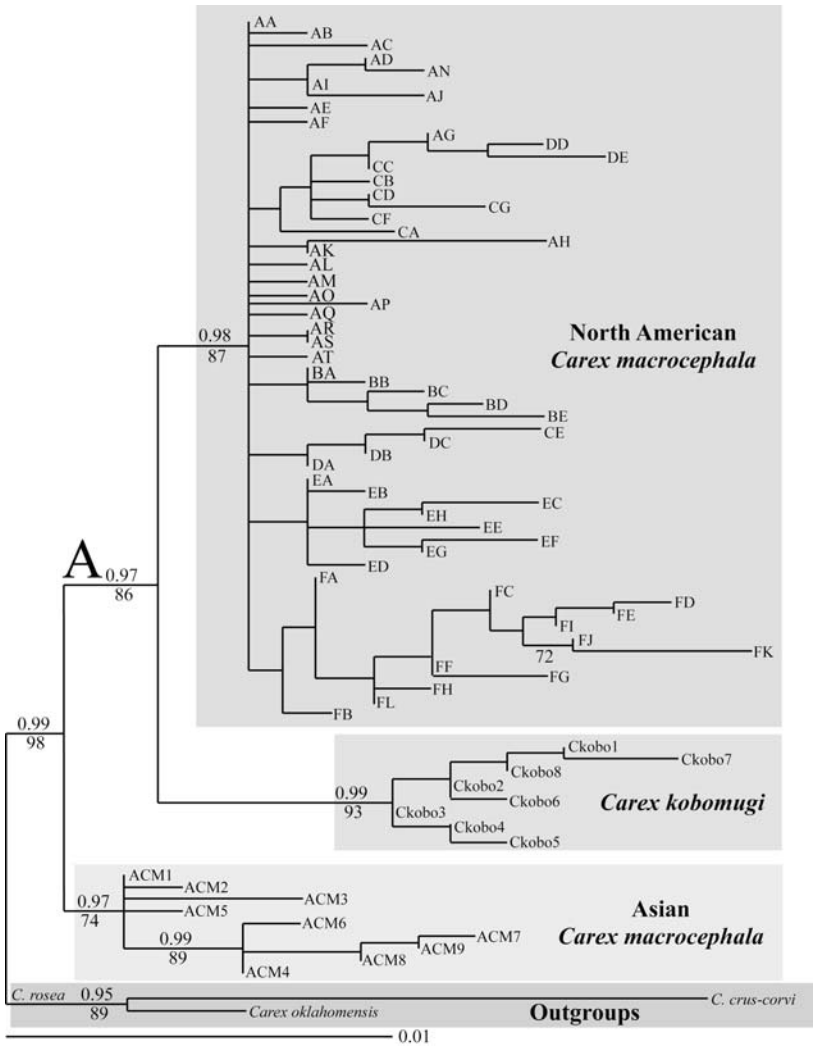
Location Number	Location	Latitude N	Longitude W	Sample size
1	Mouth of Moose Creek, OR	44^{0}21.6259998'	124^{0}53.83002'	4
2	Mouth of Beaver Creek, OR	44^{0}31.4290002'	124^{0}44.17002'	6
3	Taft, OR	44^{0}55.7470002'	124^{0}7.05'	3
4	Siletz Bay, OR	44^{0}55.674'	124^{0}14.17002'	2
5	Neskowin Beach, OR	45^{0}6'	123^{0}58.2'	8
6	Whalen Island South, OR	45^{0}16.3960002'	123^{0}57.013002'	6
7	Whalen Island North, OR	45^{0}16.719'	123^{0}57.019998'	5
8	Camp Magruder, OR	45^{0}34.9810002'	123^{0}57.889998'	5
9	Rockaway Beach, OR	45^{0}37.56'	123^{0}56.611998'	6
10	Del Rey Beach, OR	46^{0}28.840002'	123^{0}55.857'	3
11	Camp Rilea Beach, OR	46^{0}6.8590002'	123^{0}56.695998'	5
12	Leadbetter Point State Park 1, WA	46^{0}31.6480002'	124^{0}27.93'	4
13	Leadbetter Point State Park 2, WA	46^{0}36.429'	124^{0}25.90998'	8
14	Midway Beach, WA	46^{0}46.1449998'	124^{0}56.50998'	6
15	Westport Lighthouse State Park, WA	46^{0}53.2369998'	124^{0}7.408998'	6
16	La Push Beach, WA	47^{0}54.9490002'	124^{0}38.541'	6
17	Crescent Beach, WA	48^{0}9.7369998'	123^{0}42.42'	7
18	Fort Worden State Park, WA	48^{0}8.883'	122^{0}45.678'	7
19	Deception Pass State Park, WA	48^{0}23.478'	122^{0}38.848002'	3
20	Spencer Spit State Park, WA	48^{0}32.4220002'	122^{0}51.318'	3
21	Port Renfrew, BC	48^{0}32.16'	124^{0}24.649998'	6
22	Long Beach 1 PRNP, BC	49^{0}43.47'	125^{0}46.012002'	6
23	Long Beach 2 PRNP, BC	49^{0}43.47'	125^{0}46.012002'	7
24	Mackenzie Beach PRNP, BC	49^{0}7.9930002'	125^{0}54.157002'	6

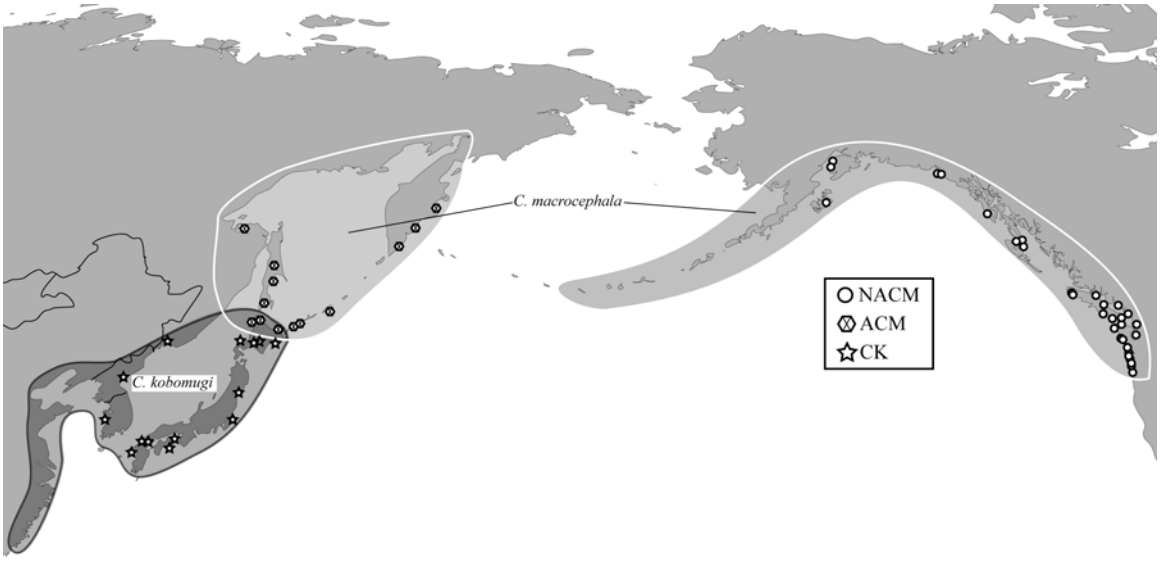
25	Wickaninish Beach PRNP, BC	49^{0}11.46'	125^{0}40.384998'	6
26	Oyster Beach, BC	49^{0}53.73'	125^{0}8.811'	6
27	San Josef Beach 1, BC	50^{0}40.4680002'	128^{0}16.606002'	6
28	San Josef Beach 2, BC	50^{0}40.4680002'	128^{0}16.606002'	6
29	Rose Spit, BC	54^{0}10.179'	131^{0}39.396'	5
30	Naikoon Provincial Park, BC	54^{0}6.762'	131^{0}42.873'	10
31	Tow Hill, BC	54^{0}43.35'	131^{0}47.463'	8
32	Tlell Beach, BC	53^{0}34.7299998'	131^{0}55.908'	6
33	Pasagshak Beach Kodiak Island, AK	57^{0}27.5140002'	152^{0}27.021'	6
34	Shelikof Beach Kruzof Island, AK	57^{0}10.1860002'	135^{0}45.355998'	6
35	Shelikof Beach 2 Kruzof Island, AK	57^{0}9.981'	135^{0}45.361998'	3
36	Yakutat Canon Beach, AK	59^{0}29.5690002'	139^{0}43.630998'	9
37	Yakutat Coast Guard Beach, AK	59^{0}30.597'	139^{0}46.441998'	5
38	Mouth of Kenai River, AK	60^{0}34.2'	151^{0}15'	9
39	Kalifornsky Beach, AK	61^{0}31.4299998'	151^{0}16.168002'	2
40	Kasilof Beach, AK	60^{0}23.3539998'	151^{0}17.79'	6

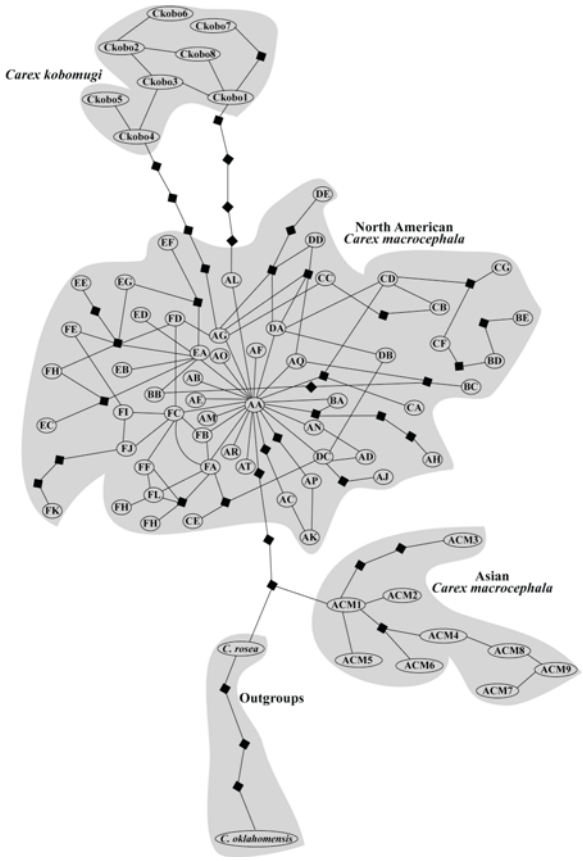
APPENDIX B. Specimens sampled from herbarium sheets for samples collected in Asia.

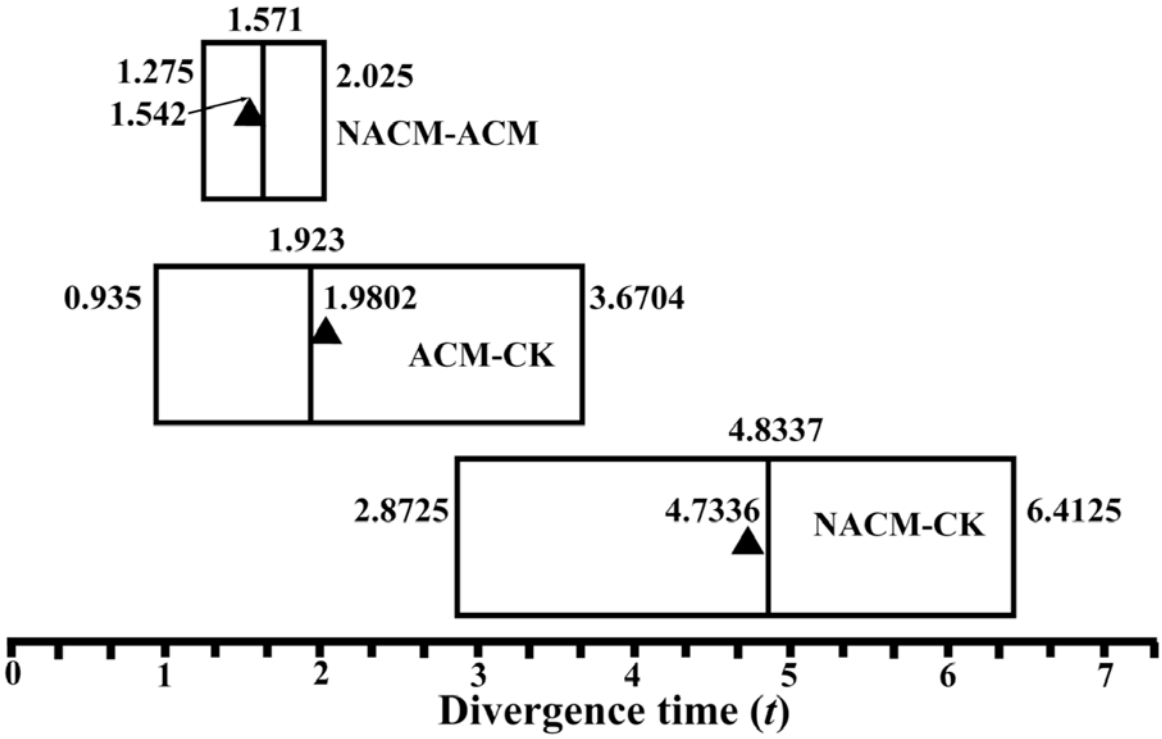
	Location	Collection
	Asian *C. macrocephala*	
1	Kunashir Island Kurils	Gage 1677 [WTU]
2	Urup Island Otkyti Bay, Kurils	Gage 1121 [WTU]
3	Etorofu Island, Kurils	Kondo 7927 [TH]
4	Sakhalin	Honda and Kimura 17-8-1940 [TH]
5	Sakhalin Sakaehama?	Saito 728129 [TH]
6	Sakhalin	Komatsu 1387 [TH]
7	Ulban Bay, Russia	s.n. [TNS]
8	Hokkaido, Cape Soya, Japan	Yamazaki 4627 [TH]
9	Hokkaido, Cape Soya, Asachino, Japan	Hara 21334 [TH]
10	Hokkaido, Abashiri, Japan	Shimizu 80:220 [TH]
11	Zhupanova, Kamchatka, Russia	s.n. [TI]
12	Khalaktyrka, Kamchatka, Russia	s.n. [HAK]
13	Kamchatka Peninsula	s.n. [WTU]
	C. kobomugi	
1	Honshu, Kadzusu-ichinomigu	Togashi 8703 [TH]
2	Yamaguchi Yoshiki-gun, Japan	Oka 32671 [TH]
3	Fukuura-mura, Honshu Island, Japan	Ohashi 6657 [TH]
4	Kagoshima Satsuma Peninsula, Japan	Kondo 2229 [TOFO]
5	Nakamura, Japan	Amano 1138 [TH]
6	Kunsan, Korea	Tyson 4938 [TH]

7	Korea	Urata 7531 [TH]
8	1946 Russian Manchuria	Hura [TH]
9	Yamaguchi, Japan	Oka 28373 [TNS]
10	Hokkaido Island, Nemura-shi, Japan	s.n. [TNS]
11	Hokkaido Island, Hamakoshimizu, Japan	s.n. [TNS]
12	Otaru City, Japan	Abe 56954 [TKPM]
13	Daikokujima, Japan	s.n. [TOFO]
14	Tokushima, Japan	Abe 42618 [TKPM]


AA
AB
AC
AD
AN
AI
AJ
AE
AF
AG
DD
DE
CC
CB
CD
CG
CF
CA
AH
AK
AL
AM
AO
AP
AQ
AR
AS
AT
BA
BB
BC
BD
BE
CE
DC
DB
DA
EA
EB
EC
EH
EE
EF
EG
ED
FA
FC
FD
FE
FI
FJ
FK
72
FF
FG
FH
FL
FB
North American
Carex macrocephala
0.98
87
A
0.97
86
0.99
98
Ckobo1
Ckobo7
Ckobo8
Ckobo2
Ckobo6
0.99
93
Ckobo3
Ckobo4
Ckobo5
Carex kobomugi
ACM1
ACM2
ACM3
0.97
74
ACM5
ACM6
0.99
89
ACM7
ACM9
ACM8
ACM4
Asian
Carex macrocephala
C. rosea
0.95
89
C. crus-corvi
Carex oklahomensis
Outgroups
0.01


C. macrocephala
C. kobomugi
NACM
ACM
CK


Ckobo6
Ckobo7
Ckobo2
Ckobo8
Carex kobomugi
Ckobo3
Ckobo1
Ckobo5
Ckobo4
DE
North American
Carex macrocephala
EF
DD
CG
EE
EG
AL
CC
CD
CB
ED
FD
BE
FE
AG
DA
CF
FA
AO
AF
AQ
DB
BD
FH
EB
AB
BC
BB
AE
BA
AA
CA
EC
FI
FC
AM
AN
FB
FJ
AR
DC
AD
AH
FF
FA
AT
AP
AJ
FL
AC
FK
FH
CE
AK
ACM3
Asian
Carex macrocephala
ACM2
ACM1
ACM4
ACM8
ACM9
ACM5
ACM6
ACM7
C. rosea
Outgroups
C. oklahomensis


1.571
1.275
1.542
2.025
NACM-ACM
1.923
0.935
1.9802
3.6704
ACM-CK
4.8337
2.8725
4.7336
NACM-CK
6.4125
0
1
2
3
4
5
6
7
Divergence time (t)

CHAPTER 3

CAREX MACROCEPHALA MAY HAVE BEEN UNRESTRICTED TO REFUGIA DURING THE LAST GLACIAL MAXIMUM

The northwestern portion of North America has been popular region for phylogeographic studies, and many of the researchers aims have been to identify potential glacial refugia. Several of these projects focus on species with large biogeographic disjuncts (Brunsfeld et al. 2001; Nielson et al. 2001; Demboski and Sullivan 2003; Carstens et al. 2004) or on species native to montane regions of the Cascade and Rocky Mountains (Demboski and Cook 2001; Good et al. 2003; Carstens et al. 2005). Potential refugia have been located in areas south of the continental ice sheet front (Brunsfeld et al 2001; Demboski and Sullivan 2003; Steele and Storfer 2006), but some studies have suggested refugia north of the continental ice sheet maximum (Brunsfeld et al. 2001; Fleming and Cook 2002; Wreckworth et al 2005). These studies have suggested the potential refugia were located along the Pacific coast, however these taxa are not obligate coastal species and therefore speculation remains as to the precise location of the coastal refuge. Non-obligate coastal species may have existed in inland refugia and recolonized coastal areas immediately following the retreat of the ice sheet. By using an obligate coastal and terrestrial species as our model organism we can attempt to infer locations along the northwest coast of North America that may have been coastal refugia.

For this study we have chosen *Carex macrocephala*, the large-headed sedge, as our model organism. *Carex macrocephala* is a low-lying sedge confined to sandy beaches and sand dunes distributed along the west coast of North America (NA) from mid-Oregon north through Alaska, the Aleutian Islands, across the Bering Sea to the Kamchatka Peninsula, and south along the east coast of Asia to Hokkaido Island of Japan (Figure 1). It spreads across unstabilized sand dunes by extensive rhizome growth

(Mastrogiuseppe 2002). *Carex macrocephala* has monoecious individuals, although typically the shoots are unisexual, and this is referred to as paradioecy (Standley 1985; Pannell 2006). There are no reports of self-incompatibility in *Carex*, and selfing rates are thought to be high. A previous study by King and Roalson (in review) using statistical phylogeographic coalescent analyses suggests that *C. macrocephala* has existed within North America for at least 125,000 years ± 20,000 years.

Given the assumption that *C. macrocephala* has persisted in NA since before the last glacial maximum (LGM), we can infer that there must have been populations harboring genetic and allelic diversity during that time (Widmer and Lexer 2001), approximately 18,000 years ago (Barendregt and Duk-Rodkin 2004). It is assumed that the genetic diversity present in the refugial populations would leave identifiable patterns that persisted for several millennia, which will allow us to locate those refugia (Hewitt 1996; Widmer and Lexer 2001; Hewitt 2004). We assume, as most glacial refugial hypotheses do, that the range of *C. macrocephala* would have been restricted, and the genetic diversity would have been reduced due to bottleneck events (Abbott et al. 2000).

The population genetic structure sampled from across the range is the result of old and recent population level forces such as drift, migration, etc. Often, shallow genetic processes masks the deep time population genetic signal and this is particularly true when examining the genetic structure of northern populations that are the descendants of Pleistocene glacial refuge populations (Riddle 1996). When researching potential glacial refuge locations we often assume that current genetic diversity within those locations is higher than in the non-refuge populations due to founder effects (Hewitt 1996; Widmer and Lexer 2001). It is also important to consider areas colonized from multiple refuge

populations can mislead our attempts to identify those refugia (Widmer and Lexer 2001; Abbott and Brochmann 2003). If our ability to identify refugia locations is limited to modern day population genetic structure we assume that the structure is the direct result of the glaciation process. When modern day population genetic forces are masking the genetic structure we are investigating we may still able to infer potential glacial refugia. High levels of gene flow, inbreeding, long-term shifting distributions and shifting population sizes all affect the genetic patterns we are able to sample. By using multilocus datasets of codominant markers it may be possible to infer population genetic structure and assignment to demes (Pritchard et al. 2000; Huelsenbeck and Andolfatto submitted). Even then there are several assumptions that must hold, such as having neutral loci in Hardy-Weinberg equilibrium (HWE) and linkage equilibrium. Higher than expected levels of inbreeding will cause non-HWE and linkage disequilibrium, and can reduce allelic diversity within subpopulations.

In an attempt to strengthen the coastal glacial refuge hypothesis, we will use a multilocus approach to attempt to identify refugia of *C. macrocephala* along the northwest coast of North America (NA). We will test whether current levels of inbreeding and gene flow affect our ability to make these inferences, and whether these forces have lead to other effects of population genetic structuring.

Materials and Methods

Sample collection

Six hundred individuals of *C. macrocephala* were collected from 41 localities across the coast of NA (Figure 1, Table 1). Leaf specimens were collected from across the range of each locality in a manner that would minimize the probability of sampling the same genet. Many times this would involve sampling ramets from several miles of beaches, and in other cases it was possible to follow the rhizome spread across the beach, and shoots were sampled as far apart as possible to ensure separate genet sampling. If populations were small and confined, leaf specimens were sampled at random. Leaf tissue was immediately stored in individual silica gel cellophane bags.

Molecular techniques

DNA was extracted following a modified 2x CTAB protocol using 30mg of dried leaf material and pulverized using liquid nitrogen and micropestels (Roalson et al. 2001). We used a cpDNA marker, the *rpL16* intergenic spacer, to identify if patterns changed between the nuclear and chloroplast genomes. The cpDNA marker has the added benefit of being uniparentally inherited, which reduces the effective population size (N_e) of the marker. This marker has also been shown to be variable within populations of *C. macrocephala* along the NA coast (King and Roalson in review). Amplification protocols of the *rpL16* spacer can be found in Shaw et al (2005), and sequencing was conducted on an Applied Biosystems 3730 automated DNA sequencer following ABI's protocol for Big Dye Terminators version 3.1. Forward and reverse strand templates were sequenced, contiged and edited using Sequencher version 4.6 (GeneCodes Corp.). Sequences were manually aligned in Se-Al version 2.0 (Rambaut 1996).

Microsatelite loci were characterized in King and Roalson (in prep), and all loci are comprised of dinucleotide repeats except for CM16, which is comprised of

trinucleotide repeats. All these loci were developed from an individual of *C. macrocephala* from the coast of Washington state. Amplification of the fragments followed the protocols outlined in King and Roalson (in prep), and amplified products were multiplexed with three other loci each with a different fluorescent dye. Final dilution ratios of 1:40 were needed and fragments were visualized on an Applied Biosystems 3730 and scored using GENEMAPPER version 3.7 (Applied Biosystems). To minimize the chance of null alleles all individuals for all loci were scored manually. Three localities, 10, 28, and 32, were re-extracted and genotyped to assess genotyping error, and ascertain clonal identity (Bonin et al. 2004; Arnaud-Haond et al. 2007).

Population genetic analyses

We first analyzed the microsatellite dataset to identify general population genetic patterns and to identify whether the loci are in Hardy-Weinberg equilibrium (HWE) and linkage equilibrium. FSTAT (Goudet et al. 1996) was used to calculate allelic diversity (A), Nei's heterozygosity (Ho), Nei's genetic differentiation (Gst), HWE, and levels of linkage equilibrium. We also used FSTAT (Goudet et al. 1996) to obtain summary population genetic indices of Wright's F-statistics, *F*, *f*, and Θ, following Weir and Cockerham (1984), where *F* is an estimate of inbreeding across all subpopulations, *f* is an estimate of inbreeding within each subpopulation, and Θ is an estimate of differentiation across all subpopulations (Wright 1951; Nei 1973). Values for G_{ST} are highly dependent on the level of variability, so we used RECODEDATA (Meirmans 2006) to obtain a dataset that would estimate the maximum Gst value possible and scaled our Gst value to this maximum Gst value, as recommended by Hedrick (2005).

Arlequin v3.11 was used to test two hypotheses of genetic structuring using both the *rpL16* and microsatellite datasets (Excoffier et al. 2005) by running AMOVA analyses using two different groupings of the different 41 localities (Excoffier et al. 1992). The first structural grouping was based on the littoral cells off the coast of North America and the major currents off the mainland coast. Littoral currents are those currents that are nearest to the shore, and can be broken into the major littoral cells (Twitchell and Cross 2002). We hypothesized that most of the movement of individuals would take place within the major current systems, and therefore grouped localities into a mid-Oregon (populations 1-5), northern Oregon (6-9), Columbia River littoral cell (10-14), Washington coast (15-16), San Juan Islands (17-20), Vancouver Island (21-29), Queen Charlotte Islands (QCI) (30-33), southeast Alaska (34-37), and south-central Alaska (38-41). A second AMOVA was conducted where the populations were lumped together into larger groups of the southern coast (1-17), San Juan and Vancouver Islands (18-29), QCI (30-33), and Alaska (34-41).

To determine the level of population partitioning we used the program STRUCTURE, which uses a probabilistic approach to assigning individuals to a set prior number of demes (Pritchard et al. 2000; Falush et al. 2007). STRUCTURE is able to calculate a likelihood score based on a fixed prior number of demes, K. Assignment into demes would help us infer the location of ancestral populations, and those subpopulations which are descendant from them. This approach assumes that departures of HWE and linkage equilibrium are a result of the inbreeding caused by population structuring and not a result of inbreeding due to non-random mating within a subpopulation. STRUCTURE analyses were run for 2 million generations with a 200k generation burn-

in. We did not assume a correlated allele frequency across the subpopulations. We analyzed K values from 1 to 41 with 5 independent runs each to asses that the likelihood scores were not significantly different between runs. Although the model assumptions are similar, we also used the program STRUCTURAMA, which assigns individuals to populations using a Dirichlet process prior (Huelsenbeck et al. submitted; Huelsenbeck and Andolfatto submitted). The advantage to this method is that it treats K as a random variable instead of fixing it as in STRUCTURE. STRUCTURAMA analyses were run for 5 million generation with a 500K generation burn-in with 4 chains. Two independent runs confirmed convergence of the chains.

Statistically, it may be difficult to infer the level of genetic structuring using population genetic techniques that require loci to be in linkage equilibrium and HWE. As a means of demonstrating any level of population structuring beyond panmixia we conducted a principal coordinate analysis (PCoA) on the microsatellite dataset. The advantage to PCoA analyses is that they do not rely on stable loci that are in HWE or linkage equilibrium. PCoA reduces the dimensionality of the dataset while maintaining the variance and covariance between the samples. We account for the variance in each level of the principal coordinates where the first principal coordinate always has the most variance and the second principal component has the second most variance, etc. Individuals will group together according to their position within each of the principal coordinates. PCoA analyses were conducted using GenAlEx version 6 (Peakall and Smouse 2006), and the first three principal axes were simultaneously visualized using GNUPLOT (Williams and Kelley 2004).

To determine whether genetic diversity was reduced during the last glacial maximum we used the program SWEEP-BOTT to test for a bottleneck event (Galtier et al 2000). SWEEP-BOTT uses a coalescent-based likelihood approach to determine the time and strength of bottleneck events for sequence data under the infinite sites model. Although we are using only a single-locus to estimate the bottleneck event the power of this method is still higher than using Tajima's D (Galtier et al. 2000). We are also able to statistically test whether a non-founder event or bottleneck event was more likely. We removed two haplotypes, each with frequencies of less than 1%, to fit the dataset to the infinite sites model. We conducted initial runs of 10,000 iterations for the first phase and 50,000 iterations for the second phase to assess the prior range of parameters for theta, $\theta = N_e\mu$ for uniparentally inherited loci; time of bottleneck, *T* in units of $2N_e$ generations; and strength of bottleneck, *S.* After the initial runs, prior range for *T* was set from 0.3 to 1 and the range of S from .01 to .2. Final runs of 100,000 steps were conducted to assess the maximum likelihood of *T, S,* and to conduct the likelihood ratio test.

Results

Individuals from locations 10, 28, and 32, showed low levels of clone duplication, and no reproducible allelic dropout (Bonin et al. 2004). As such, duplicate genotypes within localities were removed leading to a final overall sample size of 548 (Arnaud-Haond et al. 2007). All microsatellite loci differed significantly from HWE after Bonferroni corrections, with all p-values less than 0.00010 (Rice 1989). All loci were in significant LD with at least two other loci, and we would need to prune our dataset down to two loci to remove significant LD. The microsatellites dataset had an insignificant $\Theta =$

0.013 ± .015 across all loci and populations with significant values of f = 0.956 ±.04 and F = 0.969 ±.03. By standardizing the Gst value of the overall dataset by the maximum Gst calculated from the RECODEDATA dataset we obtained a standardized estimate of Gst = 0.032. The *rpL16* dataset had a slightly significant Fst value = 0.042 ± .022. Most pairwise Fst values of the microsatellite dataset showed a significant level of fixation between populations (Table 2.). Observed heterozygosity (Ho) levels ranged from 0.000 to 0.127 with an Ho = 0.041 across all microsatellite loci, and are summarized in Table 3. The inbreeding estimate, *f,* within each subpopulations ranged from 0.69 in locality 38, to an *f* of 1 in several localities.

Results of the AMOVA analyses show that a significant amount of the variance is accounted for within subpopulations, between subpopulations within groups, and between groups for both models of genetic structuring, and for both the microsatellite and *rpL16* datasets (Table 4). However, the STRUCUTRE analyses show the highest likelihood score when the parameter of the number of subpopulations, K, was set to 15. Likewise, the results of the STRUCTURAMA analyses suggests a high K value = 18 with a posterior probability = 0.3712.

The first three principal coordinates of the PCoA account for 59.22% of the variation within the dataset with principal axis 1 (PA1) accounting for 25.52%, PA2 = 17.34%, PA3 = 16.35%, PA4 = 15.14%, PA5 = 13.31%, PA6 = 12.33% of the variation. As the first three principal coordinates account for a majority of the variation we graph the principal coordinates for each individual at the first three principal coordinates axis (Figure 3).

The non-founder event model had a –ln likelihood score of -152.8412 and the bottleneck event model had a –ln likelihood score of -172.2840. The likelihood ratio test statistic is 38.8856 with a χ^2 df =2 results in a p-value = 3.5983×10^{-9}. Therefore, SWEEP-BOTT and the likelihood ratio test strongly suggest the most appropriate model is the non-founder event model. The non-founder event model assumes there has not been a measurable decline in genetic diversity. However, if a bottleneck did occur the maximum likelihood estimates of the timing of the bottleneck event, $T = 0.63699$ with the strength of the bottleneck, $S = 0.112$.Using the $N_e = 147{,}000$ obtained from a multilocus estimate from King and Roalson (in review) we can estimate the time and strength of the bottleneck event at approximately 186,000 years ago and reduced by 16,000 individuals.

Discussion

The overall levels of *f* and *F* indicate high inbreeding levels within the subpopulations, while the low standardized Gst of 0.032 indicates a low fixation level across North America. Our overall estimates of $F = 0.979$ and $f = 0.956$ indicate very high levels of inbreeding, and these patterns suggest that the levels of inbreeding within each subpopulation are significantly high to cause the non-HWE and LD beyond that of which is normally expected for subdivided populations. The inbreeding levels are higher than previously observed in clonal sedges (Stenström et al. 2001; Tyler 2002). If inbreeding is occurring at a relatively high rate, then the number of population structures, K, is overestimated (Pritchard et al. 2000). When assigning individuals to demes with STRUCTURE we assume the amount of LD present in the dataset is a result of

inbreeding due to population substructuring and not a result of direct inbreeding, such as self-pollination. Likewise, STRUCTURAMA has similar assumptions, and can also overestimate K when inbreeding levels are higher than expected. In consideration of the high inbreeding values of $F = 0.979$ and $f = 0.956$ we suggest the respective K values of 15 and 18 for the STRUCTURE and STRUCTURAMA analyses to be overestimates, and due to the high levels of inbreeding we are not able to correlate these K values to an actual estimate (Pritchard et al. 2000)

We assume the high levels of homozygosity are a result of inbreeding occurring within each subpopulation. Nonrandom mating within each subpopulation leads to departures of HWE, but deviations at this level are typically the result of high levels of selfing (Fenster et al. 2003; Balloux et al. 2004). Furthermore the levels of linkage disequilibrium between these loci are typical when selection is acting upon the loci, or when inbreeding levels are high (Charlesworth 2003; Glemin et al. 2006). The high levels of inbreeding may have led to the reduction in allelic and genetic diversity within some of the subpopulations. Ingvarsson (2002) discussed a means of reducing genetic variation in self pollinating species through the recurrent extirpation and colonization of subpopulations in a metapopulation.

The results of our AMOVA analyses support our hypothesis that inbreeding is resulting in overestimated K values, and that migrant flow is homogenizing the populations across NA. By partitioning the sampling localities into groups according to littoral cells and currents off the northwest coast, the percent of variation associated with the between group category was at a significant amount of 5.5% for the microsatellite dataset and 7.2% for the *rpL16* dataset. By lumping localities further according to general

geographic region we reduced the between group variance to 3.3% for the microsatellite dataset and 4.8% for the *rpL16* dataset, which are both still significant amounts of variation. These are large geographic regions that are covering, in some cases, more than a thousand miles of distance. Additionally between group Φ_{CT} values ranged from 0.018 to 0.037 indicating low fixation even across large geographic areas. The higher levels of Φ_{ST} are indicative of the significant pairwise F_{ST} values, which can result due to the high levels of inbreeding. This suggests that migration is high enough to homogenize the genetic variation among these regions.

In an attempt to ameliorate our findings and show population structuring regardless of the LD and non-HWE, we performed the PCoA. We expect individuals to group together according to their pairwise genetic distance. When those genetic distances are homogenized and the lineage exists in a metapopulation there should be no sub-grouping within the PCoA. According to the first three principal axes, which show a majority of the variation, there is a single large group centered around the origin (Figure 2). This strongly suggests migration is high enough to overcome drift and inbreeding in the NA population of *C. macrocephala*.

Carex macrocephala is a paradioecious, rhizomatously spreading, clonal sedge. A single genet can send out multiple male and female inflorescences across a relatively large area. This can lead to a high rate of self-pollination, and as such a direct form of inbreeding in a sexually reproducing plant. If migration were not homogenizing the populations along the northwest coast we would expect drift to have a larger impact on the allele frequencies (Abbott et al. 2000). By comparing the F_{ST} values between the *rpL16* and microsatellite datasets we determined that gene flow is not limited to pollen,

and that drift is not having a large impact on the subpopulations. If the levels of gene flow were not able to overcome drift within each subpopulation, or if gene flow was limited to pollen dispersal we would expect to see population sub-structuring in the *rpL16* dataset. As it is, the AMOVA and Fst value for the *rpL16* marker suggest high gene flow across the west coast of North America.

According to results of the SWEEP-BOTT analysis the *rpL16* marker did not significantly bottleneck during the LGM. The likelihood ratio test suggests the most appropriate model for this dataset is the non-founder event model, which assumes there has not been a significant reduction in genetic diversity in the history of the population. Taken alone the bottleneck model suggests the time to the bottleneck event was approximately 186,000 years ago, and the strength of the bottleneck was weak (Galtier et al. 2000). Given the error associated with using a single marker for estimating bottleneck events under a coalescent model, this may coincide with the divergence of *C. macrocephala* in Asia from *C. macrocephala* in NA approximately 125,000 years ago (King and Roalson in review).

Our initial expectation was that a bottleneck event would have reduced the genetic diversity along the NA coast during the LGM, and that following the retreat of the continental ice sheet and the rising of sea level, populations of *C. macrocephala* would have expanded from those refugia colonizing along the NA coast. The newly colonized populations would have a reduced level of genetic diversity, and we could then measure those levels to determine the location of refuge areas during the LGM. Even though our estimates of current gene flow are high, we would still have expected a reduction of genetic diversity if *C. macrocephala* had been restricted during the LGM. Based on the

results from the SWEEP-BOTT analysis, PCoA, AMOVA, and F-statistics we conclude *C. macrocephala* was not restricted in its range during the LGM, but that it had a large distribution and was not confined to a few glacial refugia.

Identifying the locations of suitable habitat for *C. macrocephala* during the Pleistocene is difficult when sea level was lower than current levels. Offshore drilling records indicate ice-free areas may have existed around the Queen Charlotte Islands, Vancouver Island, and northern Washington coast. Other evidence suggests that ice-free areas were common along the coast, and that the ice sheet only extended to the coast in the form of glaciers along the valleys (Barrie and Conway 1999; Fulton et al. 2004; Kaufman and Manley 2004; Porter 2004). Sediment deposits along the continental shelf suggest these areas may have contained sandy beaches, and it would not be unreasonable to assume that these beaches formed quickly. Following the retreat of the Cordilleran Ice Sheet there was a rapid rise in eustatic sea level of more than 100m occurring over 5000 years. It then took an additional 8000 years for sea level to rise another 20m to its current elevation (Twichell and Cross 2002). Twichell and Cross (2002) used core drillings off the coast of North America to imply that the sandy beaches along the northwest coast of North America have only been in their current position for less than 3000 years (Barrie and Conway 1999; Twichell and Cross 2002). This further strengthens our argument that *C. macrocephala* would have taken advantage of any newly formed beach system, and the ability to migrate quickly along the coast was essential for its propagation.

The analytical evidence suggests *C. macrocephala* was not restricted during the LGM, but existed as a large metapopulation population across the North American range. Other studies on species that usually inhabit areas north of the continental ice sheet

suggest reduction of the populations into glacial refugia either north or south of the ice sheet (Dorken and Barrett 2004; Soltis et al. 2006; Edh et al. 2007; Hodgins and Barrett 2007; Michalski and Durka 2007; Naciri and Gaudeul 2007; Magri et al. 2008). *Carex macrocephala* has the advantage of oceanic dispersal, and the ability to follow the retreat of sea level, and the subsequent rise. Studies on species inhabiting sky island populations in the American southwest indicate the LGM may have been responsible for population migration throughout the mountain ranges by providing suitable habitat between the ranges, and allowing these species to have a wider distribution (Masta 2000; Knowles 2001; DeChaine and Martin 2004). These studies suggest climate change due to Pleistocene glaciation was responsible for species having a widespread distribution during the LGM.

Conclusions

The goals of our study were to identify the potential glacial refugia of C. macrocephala along the coast of North America. The results of the AMOVA and PCoA suggest gene flow is homogenizing the subpopulations along the coast into a single metapopulation. Furthermore, the genetic fixation level across NA is low, and inbreeding values are high. Although we expected a bottleneck to have occurred during the LGM, the SWEEP-BOTT analysis suggests the most appropriate model is a non-founder event model, and that the bottleneck event would have occurred long before the LGM. Paleogeologic evidence suggests there could have been a large number of ice-free areas along the coast which may have had sandy beaches. These pieces of evidence and analytical results suggest that *C. macrocephala* was wide-spread during the LGM, rather than being

isolated into only a few glacial refugia. Thus, the picture for this obligate coastal species differs substantially from that of other PNW terrestrial plants that underwent substantial bottlenecks in glacial refugia during the Pleistocene (Brunsfeld et al. 2001; Brunsfeld and Sullivan 2006)

References

Abbott RJ, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Molecular Ecology* **12**, 299-313.

Abbott R, Smith LC, Milne RL, Crawford RM, Wolff K, Balfour J (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science*, **289**, 1343-1346.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences* **103**, 12447-12450.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, online early access.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021-3031.

Barendregt RW, Duk-Rodkin A (2004) Chronology and extent of Late Cenozoic ice sheets in North America: a magnetostratigraphic assessment. in: Quaternary Glaciations – Extent and Chronology; Part II: North America. (eds. J Ehlers and PL Gibbard), pp. 1-8. Elsevier. Amsterdam.

Barrie, JV Conway, KW (1999) Late Quaternary glaciation and postglacial stratigraphy of the northern Pacific margin of Canada. *Quaternary Research*, **51**, 113-123.

Bonin A, Bellemain E, Bronken Eidesen P, et al. (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Brunsfeld, SJ, Sullivan J (2006) A multi-compartmented glacial refugium in the northern

Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). *Conservation Genetics*, **6**, 895-904.

Brunsfeld SJ, Sullivan J, Soltis DE, and Soltis PS (2001) Comparative phylogeography of Northwestern North America: A synthesis. In: Integrating ecological and evolutionary processes in a spatial context. (eds. J. Silvertown and J. Antonovics), pp. 319-339. Blackwell Science, Oxford.

Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J (2004) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology*, **53**, 781-792.

Carstens BC, Degenhardt JD, Stevenson AL, Sullivan J (2005) Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. *Molecular Ecology*, **14**, 255-265.

Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**, 1051-1070.

DeChaine EG, Martin AP (2004) Historic cycles of fragmentation and expansion in *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. *Evolution* **58**, 113-127.

Demboski JR, Cook JA (2001) Phylogeography of the dusky shrew, *Sorex monticolus* (Insectivora, Soricidae): Insight into deep and shallow history in northwestern North America. *Molecular Ecology*, **10**, 1227-1240.

Demboski JR, Sullivan J, (2003) Extensive mtDNA variation within the yellow-pine

chipmunk, *Tamias amoenus* (Rodentia: Sciuridae), and phylogeographic inferences for northwest North America. *Molecular Phylogenetics and Evolution*, **26**, 389-408.

Dorken ME, Barrett SCH (2004) Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. *Molecular Ecology* **13**, 2699-2707.

Edh K, Widen B, Ceplitis ALF (2007) Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic *Brassica cretica* (Brassicaceae). *Molecular Ecology* **16**, 4972-4983.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary *Bioinformatics Online* **1**, 47-50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**, 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164**, 1567-1587.

Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution* **57**, 995-1007.

Fleming MA, Cook JA (2002) Phylogeography of endemic ermine (*Mustela erminea*) in southeast Alaska. *Molecular Ecology*, **11**, 795-807.

Fulton RJ, Ryder JM, Tsang S (2004) The Quaternary glacial record of British Columbia,

Canada. in: Quaternary Glaciations – Extent and Chronology; Part II: North America. (eds. J Ehlers and PL Gibbard), pp. 39-50. Elsevier. Amsterdam.

Galtier N, Depaulis F, Barton NH (2000) Detecting Bottlenecks and Selective Sweeps From DNA Sequence Polymorphism. *Genetics* **155**, 981-987.

Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences* **273**, 3011-3019.

Good JM, Demboski JR, Nagorsen DW, Sullivan J (2003) Phylogeography and introgressive hybridization: Chipmunks (genus *Tamias*) in the northern Rocky Mountains. *Evolution*, **57**, 1900-1916.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions: Biological Sciences* **359**, 183-195.

Hodgins KA, Barrett SCH (2007) Population structure and genetic diversity in tristylous *Narcissus triandrus*: insights from microsatellite and chloroplast DNA variation. *Molecular Ecology* **16**, 2317-2332.

Huelsenbeck, J. P., and P. Andolfatto. Submitted. Inference of population structure under a Dirichlet process model. *Genetics*.

Huelsenbeck, J. P., E. T. Huelsenbeck, and P. Andolfatto. Submitted. Structurama: Bayesian inference of population structure. *Bioinformatics*.

Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* **56**, 2368-2373.

Kaufman DS, Manley WF (2004) Pleistocene maximum and Late Wisconsinan glacier extents across Alaska, U.S.A. in: Quaternary Glaciations – Extent and Chronology; Part II: North America. (eds. J Ehlers and PL Gibbard), pp. 9-28. Elsevier. Amsterdam.

Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshopprers. *Molecular Ecology* **10**, 691-701.

Lisiecki LE, Raymo ME (2007) Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews* **26**, 56-69.

Magri, D, Fineschi S, Bellarosa R, Buonamici A, Sebastiani F, Schirone B, Simeone MC, Vendramin GG. The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology* **12**, 1-13.

Masta SE (2000) Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? *Evolution* **54**, 1699-1711.

Mastroguiseppe J (2002) Flora of North America: North of Mexico. Volume 23 Magnoliophyta: Commelinidae (in part): Cyperaceae. ed. Flora of North America Editorial Committee. Oxford University Press. New York.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* **60**, 2399-2402.

Michalski SG, Durka W (2007) High selfing and high inbreeding depression in peripheral populations of *Juncus atratus*. *Molecular Ecology* **16**, 4715-4727.

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323.

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **41**, 225-233.

Nielson M, Lohman K, Sullivan J, (2001) Phylogeography of the tailed frog (*Ascaphus truei*): Implications for the biogeography of the Pacific Northwest. *Evolution,* **55**, 147-160.

Ortiz MA, Tremetsberger K, Talavera S, Stuessy T, Garcia-Castano JL (2007) Population structure of Hypochaeris salzmanniana DC. (Asteraceae), an endemic species to the Atlantic coast on both sides of the Strait of Gibraltar, in relation to Quaternary sea level changes. *Molecular Ecology* **16**, 541-552.

Pannell J (1997) The Maintenance of Gynodioecy and Androdioecy in a Metapopulation. *Evolution* **51**, 10-20.

Pannell JR, Verdú M (2006) The evolution of gender specialization from dimorphic hermaphroditism: paths from heterodichogamy to gynodioecy and androdioecy. *Evolution* **60**, 660-673.

Peakall ROD, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Porter SC (2004) Glaciation of western Washington, U.S.A. in: Quaternary Glaciations – Extent and Chronology; Part II: North America. (eds. J Ehlers and PL Gibbard), pp. 289-294. Elsevier. Amsterdam.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.

Ranwell, DS (1972) Ecology of Salt Marshes and Sand Dunes. Chapman and Hall. London.

Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* **43**, 223-225.

Riddle BR (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. *Trends in Ecology and Evolution* **11**, 207-211.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The Tortoise and the Hare II: Relative Utility of 21 Noncoding Chloroplast DNA Sequences for Phylogenetic Analysis. *American Journal of Botany* **92**,142–166.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**, 4261-4293.

Standley LA (1985) Paradioecy and gender ratios in *Carex macrocephala* Cyperaceae. *American Midland Naturalist*, **113**, 283-286.

Steele CA, Storfer A (2006) Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (Dicamptodon tenebrosus). *Molecular Ecology* **15**, 2477-2487.

Stenström A, Jonsson BO, Jónsdóttir IS, Fagerström T, Augner M (2001) Genetic variation and clonal diversity in four clonal sedges (*Carex*) along the Arctic coast of Eurasia. *Molecular Ecology* **10**, 497-513.

Twichell DC, Cross VA (2002) Holocene Evolution of the Southern Washington and Northern Oregon Shelf and Coast: Geologic Discussion and GIS Data Release. USGS Open-File Report 01-076. http://pubs.usgs.gov/of/of01-076/index.htm

Tyler T (2002) Geographical distribution of allozyme variation in relation to post-glacial history in Carex digitata, a widespread European woodland sedge. *Journal of Biogeography* **29**, 919-930.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Weckworth BV, Talbot S, Sage GK, Person DK, Cook JA (2005) A signal for independent coastal and continental histories among North American wolves. *Molecular Ecology*, **14**, 917-931.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Williams T, C Kelley (2004) GNUPLOT 4.2.2 http://www.gnuplot.info/

Widmer A, Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology and Evolution*, **16**, 267-268.

Table 1. Sampling localities of *C. macrocephala* across the northwest coast of North America. N refers to sample size, *f* is the inbreeding coefficient of Weir and Cockerham (1984), Allelic Diveristy is calculated as the number of alleles per individual per locus, and the *rpL16* haplotypic diversity.

Pop	Location	Latitude	Longitude	N	*f*	Allelic Diversity	*rpL16* Diveristy
1	Mouth of Moose Creek, OR	44.36043333	124.0897167	10	0.937	0.12	0
2	Mouth of Beaver Creek, OR	44.52381667	124.0736167	10	0.928	0.31	0.8
3	Taft, OR	44.92911667	124.01175	7	0.924	0.42	0.9524
4	Siletz Bay, OR	44.9279	124.0236167	3	1	0.55	0.6667
5	Neskowin Beach, OR	45.1	123.97	10	0.849	0.24	0.8444
6	Whalen Island South, OR	45.27326667	123.9502167	8	0.938	0.39	1
7	Whalen Island North, OR	45.27865	123.9503333	7	1	0.39	0.9524
8	Camp Magruder, OR	45.58301667	123.9648333	4	0.915	0.57	0.5
9	Rockaway Beach, OR	45.626	123.9435333	5	0.964	0.47	1
10	Del Rey Beach, OR	46.04806667	123.93095	20	0.893	0.24	0.9263
11	Camp Rilea Beach, OR	46.11431667	123.9449333	9	0.945	0.36	0.8056
12	Leadbetter Point State Park 1, WA	46.52746667	124.04655	15	0.851	0.19	0.7619
13	Leadbetter Point State Park 2, WA	46.60715	124.0431833	10	0.892	0.29	0.7778
14	Midway Beach, WA	46.76908333	124.0941833	10	0.904	0.23	0.8444

15	Westport Lighthouse State Park, WA	46.88728333	124.1234833	8	1	0.23	0.8571
16	La Push Beach, WA	47.91581667	124.64235	10	0.941	0.25	0.9333
17	Crescent Beach, WA	48.16228333	123.707	10	0.977	0.25	0.9333
18	Fort Worden State Park, WA	48.14805	122.7613	10	0.833	0.22	0.8889
19	Deception Pass State Park, WA	48.3913	122.6474667	10	0.909	0.15	0.7778
20	Spencer Spit State Park, WA	48.54036667	122.8553	7	0.843	0.26	0.8571
21	Port Renfrew, BC	48.536	124.4108333	29	0.895	0.17	0.8547
22	Bamfiled, BC	48.78685833	125.173225	26	0.826	0.16	0.7631
23	Long Beach 1 PRNP, BC	49.07245	125.7668667	7	0.889	0.47	0.9048
24	Long Beach 2 PRNP, BC	49.07245	125.7668667	20	0.921	0.23	0.7947
25	Mackenzie Beach PRNP, BC	49.13321667	125.9026167	20	0.904	0.21	0.8474
26	Wickaninish Beach PRNP, BC	49.0191	125.6730833	20	0.964	0.20	0.8789
27	Oyster Beach, BC	49.8955	125.14685	29	0.887	0.13	0.5493
28	San Josef Beach 1, BC	50.67446667	128.2767667	30	0.863	0.15	0.9218
29	San Josef Beach 2, BC	50.67446667	128.2767667	20	0.952	0.22	0.9105
30	Rose Spit, BC	54.16965	131.6566	26	0.908	0.18	0.9446
31	Naikoon Provincial Park, BC	54.1127	131.71455	30	0.895	0.16	0.8023
32	Tow Hill, BC	54.07225	131.79105	30	0.905	0.16	0.9126

33	Tlell Beach, BC	53.57883333	131.9318	30	0.94	0.16	0.8897
34	Pasagshak Beach Kodiak Island, AK	57.45856667	152.45035	10	0.917	0.34	0.5111
35	Shelikof Beach Kruzof Island, AK	57.16976667	135.7559333	10	0.947	0.34	0.8
36	Shelikof Beach 2 Kruzof Island, AK	57.16635	135.7560333	15	0.987	0.23	0.9143
37	Yakutat Canon Beach, AK	59.49281667	139.7271833	15	0.935	0.25	0.7714
38	Yakutat Coast Guard Beach, AK	59.50995	139.7740333	10	0.69	0.25	0.3778
39	Mouth of Kenai River, AK	60.57	151.25	10	0.914	0.31	0.7778
40	Kalifornsky Beach, AK	61.52383333	151.2694667	10	0.984	0.31	0.8222
41	Kasilof Beach, AK	60.38923333	151.2965	20	0.945	0.22	0.5731

Table 2. Pairwise Fst significance compairisons, + indicate significant Fst values, while 0 indicates no significance.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1																																							
2	+																																							
3	+	0																																						
4	+	+	+																																					
5	+	+	+	+																																				
6	+	+	0	+	+																																			
7	+	+	+	+	+	+																																		
8	+	+	+	+	+	+	+																																	
9	+	+	+	+	+	+	+	+																																
10	+	+	+	+	+	+	+	+	+																															
11	+	+	+	+	+	+	+	+	+	+																														
12	+	+	+	+	+	+	+	+	+	+	+																													
13	+	+	+	+	+	+	+	+	+	+	+	+																												
14	+	+	+	+	+	+	+	+	+	+	+	+	+																											
15	+	+	+	+	+	+	+	+	+	+	+	+	+	+																										
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																									
17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																								
18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																							
19	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																						
20	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																					
21	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																				
22	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																			
23	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																		
24	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+																	
25	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+																
26	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+															
27	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+														
28	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	+	+													
29	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+												
30	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+											
31	+	0	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+										
32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+									
33	+	+	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
34	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+							
35	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+						
36	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	+	+				
37	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
38	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
39	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	+	+	+	+	
40	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
41	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	+

Table 3. Per locus genetic summaries where Θ, *F,* and *f* are the fixation indices of Weir and Cockerham (1984); Ho is observed heterozyosity, He is expected heterozygosity, Ht is gene diversity, and A is the number of alleles.

Locus	Gst	Θ	*F*	*f*	Ho	He	Ht	A
CM01	0.012	0.009	0.832	0.816	0.083	0.518	0.59	7
CM07	0.009	0.01	0.988	0.876	0.062	0.492	0.545	7
CM12	0.011	0.009	0.958	0.954	0.02	0.564	0.628	8
CM13	0.015	0.012	0.961	0.956	0.025	0.716	0.843	9
CM16	0.016	0.012	1	1	0	0.423	0.506	6
CM25	0.063	0.058	0.996	0.99	0.003	0.157	0.42	3
CM27	0.015	0.011	0.824	0.803	0.127	0.666	0.782	10
CM28	0.012	0.001	0.954	0.95	0.03	0.569	0.648	6
CM35	0.006	0.004	1	1	0	0.237	0.251	4
CM36	0.011	0.008	0.922	0.916	0.049	0.539	0.603	5
CM39	0.019	0.002	0.975	0.914	0.055	0.464	0.603	8
Overall	0.016	0.013	0.979	0.956	0.041	0.459	0.547	73

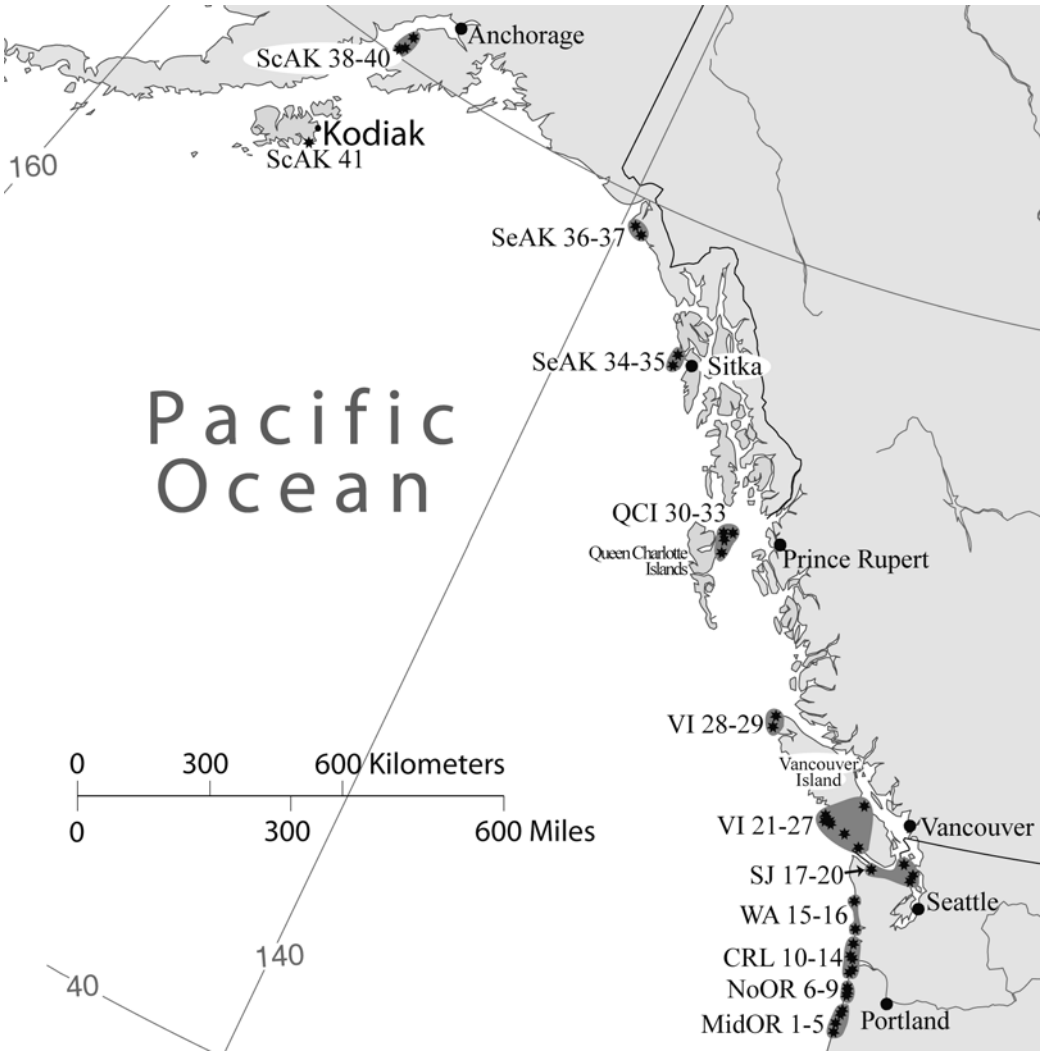
Table 4. Results of the AMOVA analyses. Top row of the comparison indicates the percent variation, while the bottom row indicates the Φ statistic analog. All levels of variation were significant, and all Φ statistics were significant.

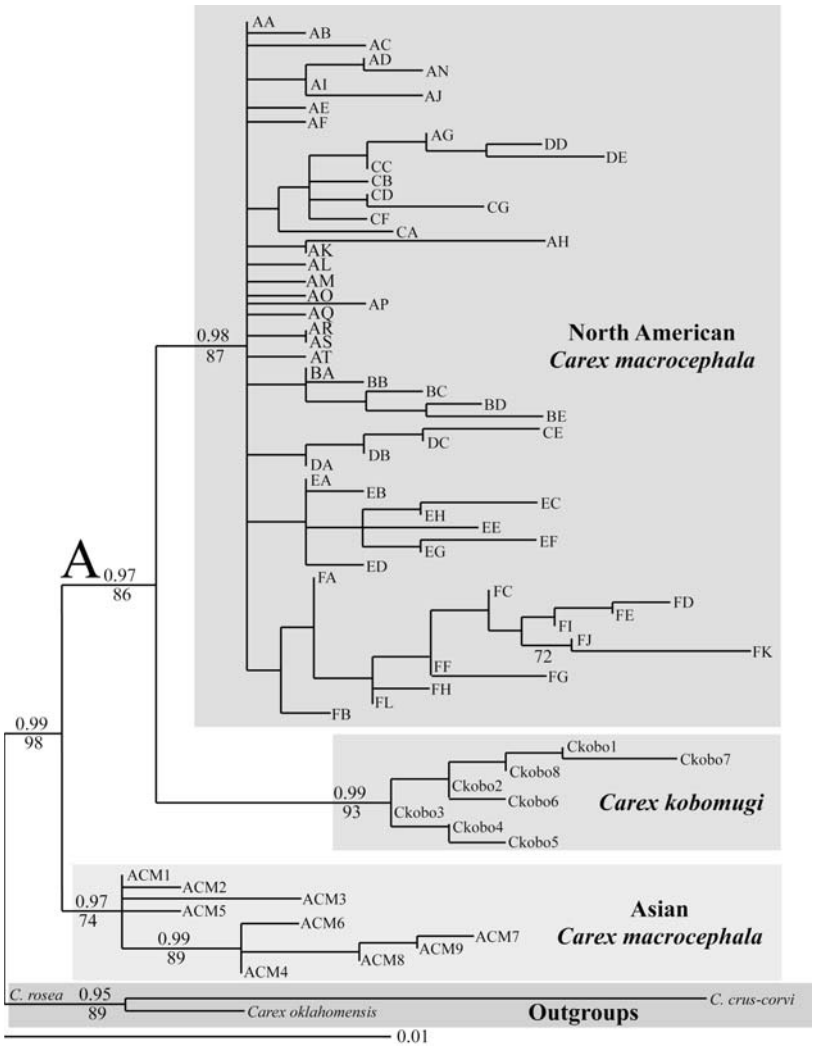
		Between Groups	Between Pops within Groups	Within Pops
Littoral Cell	Microsatellite	5.37%	9.09%	85.54%
		Φ_{CT} = .028	Φ_{SC} = .034	Φ_{ST} = .059
	cpDNA rpl16	7.24%	12.56%	80.2%
		Φ_{CT} = .037	Φ_{SC} = .05	Φ_{ST} = .083
Geographic Regions	Microsatellite	3.25%	11.09%	85.66%
		Φ_{CT} = .018	Φ_{SC} = .036	Φ_{ST} = .053
	cpDNA rpl16	4.77%	12.52%	82.7%
		Φ_{CT} = .03	Φ_{SC} = .036	Φ_{ST} = .065

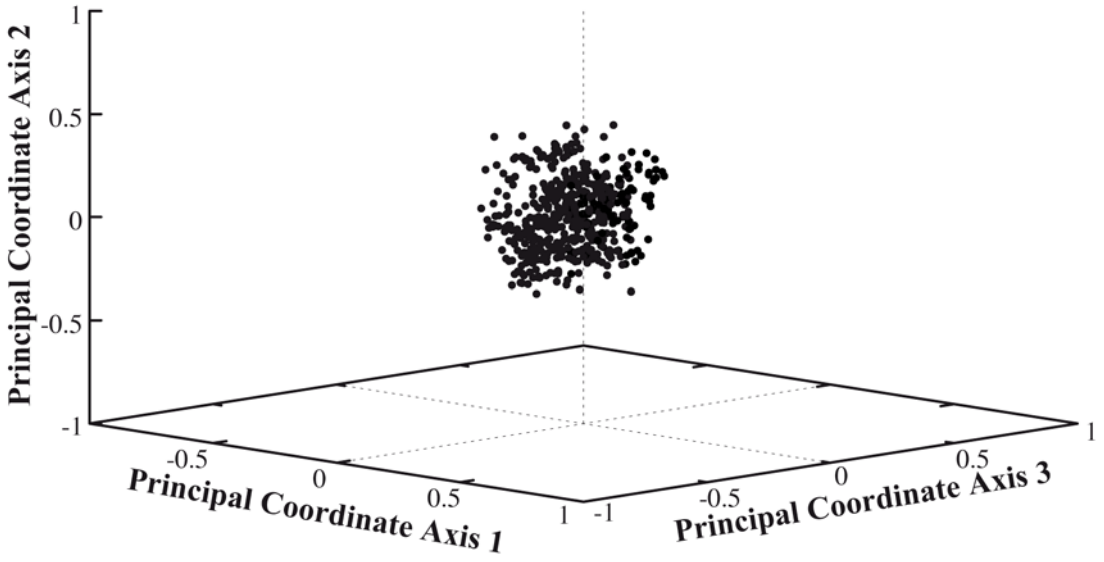
Figure 1. Map of the 41 sampling localities of *C. macrocephala* from the northwest coast of North America. Black stars represent localities and dark grey outlines surrounding localities represent the first AMOVA designations of groups.

Figure 2. Maxmimum likelihood phylogram of *rpL16* cpDNA intergenic spacer from King and Roalson (in review). Numbers above branches are Bayesian postierior probabilities and numbers below branches are bootstrap percentages. Haplotypes were designated arbitrarily based on preliminary assignment to a neighbor joining tree.

Figure 3. Graph of the first three principal coordinates of the principal coordinate analysis based on genetic distance of the eleven microsatellite loci. Principal axis (PA) 1 explains 25.52% of the variation, PA2 explains 17.34% of the variation, and PA3 explains 16.35% of the variation.


Anchorage
ScAK 38-40
Kodiak
ScAK 41
160
SeAK 36-37
SeAK 34-35
Sitka
Pacific
Ocean
QCI 30-33
Queen Charlotte
Islands
Prince Rupert
VI 28-29
Vancouver
Island
0 300 600 Kilometers
0 300 600 Miles
VI 21-27
Vancouver
SJ 17-20
Seattle
WA 15-16
CRL 10-14
NoOR 6-9
MidOR 1-5
Portland
140
40


AA
AB
AC
AD
AN
AI
AJ
AE
AF
AG
DD
DE
CC
CB
CD
CG
CF
CA
AH
AK
AL
AM
AO
AP
AQ
AR
AS
AT
BA
BB
BC
BD
BE
CE
DC
DB
DA
EA
EB
EC
EH
EE
EF
EG
ED
FA
FC
FD
FE
FI
FJ
FK
72
FF
FG
FH
FL
FB
North American
Carex macrocephala
A
0.97
86
0.98
87
0.99
98
Ckobo1
Ckobo7
Ckobo8
Ckobo2
Ckobo6
Ckobo3
Ckobo4
Ckobo5
0.99
93
Carex kobomugi
ACM1
ACM2
ACM3
ACM5
ACM6
ACM7
ACM9
ACM8
ACM4
0.97
74
0.99
89
Asian
Carex macrocephala
C. rosea
0.95
89
C. crus-corvi
Carex oklahomensis
Outgroups
0.01


Principal Coordinate Axis 2
1
0.5
0
-0.5
-1
-0.5
0
0.5
1
Principal Coordinate Axis 1
-1
-0.5
0
0.5
1
Principal Coordinate Axis 3