

# Discordance between phylogenetics and coalescent-based divergence modelling: exploring phylogeographic patterns of speciation in the *Carex macrocephala* species complex

MATTHEW G. KING\* and ERIC H. ROALSON†

\*University of British Columbia, Department of Botany, Vancouver, BC, Canada V6T1Z4, †Washington State University, School of Biological Sciences, Pullman, WA 99164, USA

## Abstract

We fit a molecular data set, consisting of the *rpL16* cpDNA marker and eight microsatellite loci, to the isolation-with-migration model as implemented in IMA to test a well-supported phylogenetic hypothesis of relationships within the *Carex macrocephala* species complex (Cyperaceae). The phylogenetic hypothesis suggests *C. macrocephala* from North America is reciprocally monophyletic and is sister to a reciprocally monophyletic clade of *C. kobomugi*. The North American *C. macrocephala* and *C. kobomugi* clade form a sister clade with a lineage of Asian *C. macrocephala*, thereby forming a paraphyletic *C. macrocephala* species. Not only does the phylogenetic hypothesis suggest *C. macrocephala* is paraphyletic, but it also suggests that the two lineages which share a partially overlapping distribution, Asian *C. macrocephala* and *C. kobomugi*, are not the most closely related. To test these relationships, we used coalescent-based population genetic models to infer divergence time for each lineage pair within the species complex. The coalescent-based models account for the stochastic forces which drive population divergence, and can account for the lineage sorting that occurs prior to lineage divergence. A drawback to phylogenetic-based phylogeographical analyses is that they do not account for stochastic lineage sorting that occurs between gene divergence and lineage divergence. 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 that the phylogenetic hypothesis is incorrect, and the divergence between these lineages occurred during the Late Pleistocene epoch.

**Keywords:** *Carex*, coalescence, Cyperaceae, isolation with migration, lineage sorting, phylogeography

Received 17 April 2008; revision received 30 October 2008; accepted 6 November 2008

Phylogeographical 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. When multiple individuals of each taxonomic unit support the same congruent phylogeographical hypothesis, 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). 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 phylogeographical relationships, which has lent credence to the argument that these methods must be analytically sound (Knowles & Maddison 2002; Knowles 2004; Templeton 2004).

Phylogeographical methodologies typically do not undercut a phylogenetic hypothesis with well-resolved clades that correspond to geographically delimited populations, especially when comparative analyses are undertaken (Brunsfield *et al.* 2001; Carstens *et al.* 2005). Populations that form well-supported reciprocally monophyletic lineages in phylogenetic analyses are useful to researchers conducting phylogeographical investigations, as the evolutionary

Correspondence: Matthew G. King, Fax: 604-822-6089; E-mail: kingdom@interchange.ubc.ca

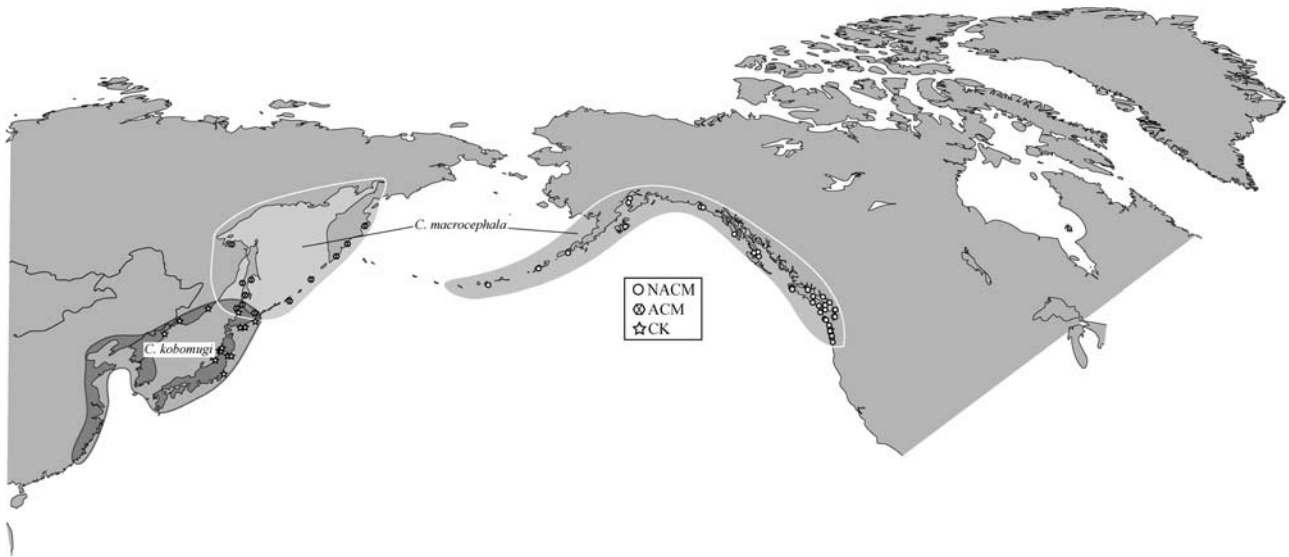


Fig. 1 Map of the northern Pacific Coast showing distribution and sampling localities of *Carex macrocephala* and *C. kobomugi* (NACM, North American *C. macrocephala*; ACM, Asian *C. macrocephala*; CK, *C. kobomugi*).

history of the species can be inferred from the single phylogenetic hypothesis (Brunsfield *et al.* 2001; Nielson *et al.* 2001; Demboski & Sullivan 2003; 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 phylogeographical analyses that rely on phylogenetics alone.

Coalescent-based population genetic models that account for both stochastic lineage sorting and changes in demography can better approximate population-level divergence processes than phylogenetic-based phylogeographical methods. Current models use simulations to account for the coalescent variance to estimate time since divergence, the effective population size ( $N_e$ ); and migration rate ( $m$ ) (Edwards & Beerli 2000; Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). By accounting for the coalescent variance, we can determine uncertainty levels for each parameter estimate. To strengthen our overall assessment, multiple independent loci should be used to minimize the impact of stochastic lineage sorting errors (Beerli & Felsenstein 2001; Hey & Machado 2003; Hey & Nielsen 2004; Dolman & Moritz 2006; Knowles & Carstens 2007b). Inclusion of the variance in coalescent times between loci is informative and useful when modelling population divergence (Edwards & Beerli 2000; Hey & Nielsen 2004). The use of these models is preferable to methods that use a single phylogenetic tree or parsimony network to infer phylogeographical hypotheses because of their ability to minimize the impact of stochastic lineage sorting errors (Beerli & Felsenstein 2001; Hey & Machado 2003).

These issues surrounding phylogeographical analyses based on phylogenetic hypotheses are not new problems.

They have been explored in detail in the theoretical realm as well as empirically tested in several studies. However, many phylogeographical studies within angiosperms ignore these issues, or it could be that the empiricists simply chose not to address these issues and assumed that lineage sorting within well-resolved phylogenies are correct. The development of lineage specific microsatellite loci has helped to alleviate some of these issues, but the focus has been on the hypervariability and codominance of the loci. The ability to use multiple genetic loci in a study that incorporates the coalescent variance for phylogeographical analyses has not been well studied in angiosperm lineages. To illustrate the issues of dependence on a single phylogenetic hypothesis for a phylogeographical analyses, we will use the case of the *Carex macrocephala* species complex.

*Carex macrocephala* Willd. ex Spreng. (Cyperaceae), the large-headed sedge, 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. 1). The closely related *Carex kobomugi* Ohwi, the Japanese sand sedge, 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 (Mastrogueissepe 2002). The phylogenetic analyses we conducted for this study suggest three reciprocally monophyletic lineages within this complex: North American *C. macrocephala* (NACM), Asian *C. macrocephala* (ACM), and *C. kobomugi* (CK). According to our 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 imply 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 there is either a complex evolutionary history of these lineages, or the hypothesis is incorrect and further analyses that account for population-level processes are needed.

By using a coalescent-based population genetic model on a molecular data set compiled from the *C. macrocephala* species complex, we are able to assess the accuracy of the phylogenetic hypothesis. We can use simulations to estimate the divergence time between each lineage in the species complex. If the phylogenetic hypothesis is accurate, the NACM and CK lineages would have the shortest time since divergence, and the ACM-CK or ACM-NACM lineages would have the longest time since divergence. If the patterns observed in the phylogenetic hypothesis are the result of stochastic lineage sorting on a single locus and not a true reflection of the lineages' history, then an alternative hypothesis is possible. Alternative hypotheses may include lineages NACM-ACM sharing the shortest divergence time, as this would support a monophyletic species, or lineages ACM-CK sharing the shortest divergence time, and this would support a relationship based on geographical sympatry in Asia.

We employed coalescent-based population genetic models of lineage divergence to test the relationships between the three main lineages of the *C. macrocephala* species complex (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). We fit the isolation-with-migration (IM) model to our cpDNA *rpL16* and microsatellite data set. By doing so, we made pairwise comparisons of the main lineages within the *C. macrocephala* species complex which allowed us to: (i) determine the accuracy of the cpDNA *rpL16*-based phylogenetic hypothesis by comparing the relative divergence times of each pairwise analysis; (ii) estimate the divergence time, in years ( $T$ ), for each of the main lineages to determine the approximate times of divergence between the populations in Asia and North America; and (iii) determine relative ancestral effective population size ( $N_{eA}$ ) estimates vs. the extant lineages, as these estimates can help infer the method of divergence between the lineages (i.e. dispersal vs. vicariance).

## Methods

### Sampling

We sampled 227 individuals from 40 localities of *Carex macrocephala* from North America (= NACM; Fig. 1; Appendices I, II) 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 in a haphazard manner within populations, but spread as far away from each other as possible to minimize the possibility of sampling identical genets. Fourteen individuals of *C. macrocephala* were sampled from across the Asian population range (= ACM). These samples were obtained from herbarium specimens obtained on loan from Hokkaido University (HAK), Kyoto University (KYO), University 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* (= CK) 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* Mackenzie, *C. crus-corvi* Shuttlw. ex Kunze, and *C. rosea* Schkuhr ex Willd. These species were selected based on their relative genetic distance from the *C. macrocephala* clade in previous phylogenetic analyses (King & Roalson 2008b in press).

### Molecular markers

Whole genomic DNA was extracted from 30 mg of silica gel-dried or herbarium material using a modified 2× cetyltrimethyl ammonium bromide 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 five specimens of *C. kobomugi*. Of the 21 markers outlined in Shaw *et al.* (2005), only the *rpL16* spacer showed polymorphism within *C. macrocephala* and between *C. macrocephala* and *C. kobomugi*. Protocols for amplifying *rpL16* via polymerase chain reaction (PCR) can be found in Shaw *et al.* (2005). Preliminary sequences of the *rpL16* spacer yielded a high level of variability. To help ensure the polymorphisms were not an artefact of PCR amplification, all individuals were amplified using Eppendorf MasterTaq (95414009-1) which has a higher fidelity than regular DNA *Taq* Polymerase (Eppendorf Corp.). PCR products were readied for cycle sequencing by incubating with 2 U of Exonuclease I (New England Biolabs, Inc.) and 3 U of Antarctic phosphatase (New England Biolabs, Inc.) at 37 °C for 1 h followed by a 15-min 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 Terminator version 3. Cycle-sequenced products were cleaned with Edge BioSystems' Ultra gel filtration system DTR plates (55373; Edge Biosystems). Sequences were obtained for the 3' and 5' strands of the PCR products followed by their assembly

into contigs, edited and visually verified using Sequencher version 4.6 (GeneCodes Corp.). All *rpL16* sequences were manually aligned in Se-Al version 2.0.9 (Rambaut 1996). Following the manual alignment, all polymorphisms were visually checked again in Sequencher. Dirty sequences were removed from the analysis. Thirty-seven individuals of North American *C. macrocephala* were resequenced to test for sequencing error rate, and all individuals of *C. macrocephala* from Asia and all *C. kobomugi* individuals were resequenced. Sequences are deposited in GenBank (Accession nos. FJ424626–FJ424702).

Microsatellite loci Cko1–12, Cko1–68, Cko1–78, and Cko1–134 were described by Ohsako & Yamane (2007) and isolated from *C. kobomugi*, and loci CM01, CM07, CM27, and CM39 were described by King & Roalson (2008a) and isolated from *C. macrocephala*. The loci were chosen because they all 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 version 3.7 (Applied Biosystems).

#### Phylogenetic analyses

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 affect the topology, we also ran phylogenetic analyses using all 56 common models of nucleotide substitution (Minin *et al.* 2003). Furthermore, parsimony and neighbour-joining analyses were also conducted using PAUP\* to determine whether different phylogenetic criteria would alter the topology. A network was constructed using the program 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 two random addition replicates saving one tree from each bootstrap replicate. Further support was assessed under a Bayesian framework with MrBayes (Huelsenbeck *et al.* 2001; Ronquist & Huelsenbeck 2003; Altekar *et al.* 2004) to generate posterior probabilities for each node. Two separate 25 million generation, eight chain analyses were run to assess convergence and mixing, and a total of 10 000 topologies were saved from each run. MrBayes runs 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* data set were computed using Arlequin version 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 data set ( $S_{total}$ ), the mean number of pairwise nucleotide differences ( $\pi$ ) for each lineage and the entire data set ( $\pi_{NACM}$ ,  $\pi_{ACM}$ ,  $\pi_{CK}$ ,  $\pi_{total}$ ), and  $\theta$  based on the number of segregating sites ( $\theta_s$ ) were also computed for each lineage and the overall sample. Arlequin was also used to calculate an  $F_{ST}$  value for the *rpL16* data set in North America. Using the microsatellite data set, we calculated an overall  $F_{ST}$  value for North America using Arlequin and pairwise  $F_{ST}$  values for each of the main lineages. One assumption of the isolation-with-migration model is the independent segregation of loci. For this reason, we 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  $\tau$ , time since divergence, 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 & Nielsen 2007). The IM approach allows the estimation of divergence time between only two populations in each analysis. The IM model assumes the two populations in the analysis are the most closely related, and that gene flow is only occurring between those two populations. In our analytical design, we are testing hypotheses of relative divergence time between three populations, whereby one of the analyses will violate the assumption that the two populations in the analysis are the most closely related. The pairwise analysis that violates the IM model will be the analysis that estimates the oldest divergence time, as this will estimate the two most distantly related populations. The two lineages that have the oldest divergence time are not the most closely related, as the third lineage is more closely related to the other two.

The IMA program can obtain (i) marginal posterior probability densities of the population parameters of divergence time,  $\tau$ , where  $\tau = T\mu$ , and where  $T$  is time in years and  $\mu$  is the mutation rate; (ii) population differentiation indices of the ancestral population and the two extant lineages,  $\theta_A$ ,  $\theta_1$ , and  $\theta_2$ , respectively, where  $\theta = N_e\mu$  for uniparentally inherited loci or  $\theta = 4N_e\mu$  for diploid autosomal loci, where  $N_e$  is the effective population size; and (iii) the migration rate for each population,  $m_1$  and  $m_2$ , where  $m$  is the migration rate per mutation =  $m/\mu$ , and where the population migration rate ( $M$ ) is  $M = 2N_e m = \theta m$ . 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.

Joint estimates of parameters are possible in the implementation of the isolation-with-migration model in the program IMA (Hey & Nielsen 2007). The IMA program 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  $\tau$  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  $\tau$  across all the genealogies by their probability, we can obtain a weighted estimate of  $\tau$ . Choice of prior estimates of  $\tau$  is important, and it was for this reason that we ran several preliminary analyses to determine the appropriate prior.

Prior estimates of  $N_e$  and  $m$  in IMA can be established through pre-analyses, or by estimates of demography. Priors used in IMA analyses were as follows: NACM-CK:  $\theta_1 = 5$ ,  $\theta_2 = 5$ ,  $\theta_A = 15$ ,  $m_1 = 2$ ,  $m_2 = 2$ ,  $\tau = 10$ ; ACM-CK:  $\theta_1 = 5$ ,  $\theta_2 = 5$ ,  $\theta_A = 15$ ,  $m_1 = 2$ ,  $m_2 = 2$ ,  $\tau = 10$ ; and NACM-ACM:  $\theta_1 = 5$ ,  $\theta_2 = 5$ ,  $\theta_A = 15$ ,  $m_1 = 10$ ,  $m_2 = 10$ ,  $\tau = 10$ . These priors were established after running 10 separate 2.5 million-step runs with a single chain for each analysis followed by three separate 15 million-step runs with two chains to assess convergence. Final runs were conducted with four chains and run for 25 million steps after a burn-in period of 5 million steps saving a total of 250 000 genealogies. We left the heating scheme unchanged from the linear default with a heating parameter of 0.05. We ran these analyses twice to ensure convergence of parameter estimates. The output of IMA includes mean values of the parameters for two sets after the burn-in period representing the first and second half of the total generations of the post burn-in run. We ran our analyses until the peak location set values were equal for set 1 and set 2. This, along with the high effective sample size and low autocorrelation estimates implies the Markov chains reached convergence after a sufficiently long burn-in period, and were sampled from the appropriate likelihood space (Appendix III). We applied the Hasegawa-Kishino-Yano (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.

We used a mean mutation rate,  $\mu = 1 \times 10^{-8}$ /site/year, with an interval set from  $1 \times 10^{-7}$ /site/year to  $1 \times 10^{-9}$ /site/year for the *rpL16* data (Wolfe *et al.* 1987; Willyard *et al.* 2007). This mutation rate falls well within the previously described mutation rates for cpDNA markers used in phylogeographical studies (Brunsfield & Sullivan 2006). We used a mutation rate range of  $3 \times 10^{-5}$ /locus/year to  $6 \times 10^{-4}$ /locus/year with a mean of  $2 \times 10^{-4}$ /locus/year for the microsatellite loci. This range was chosen based on

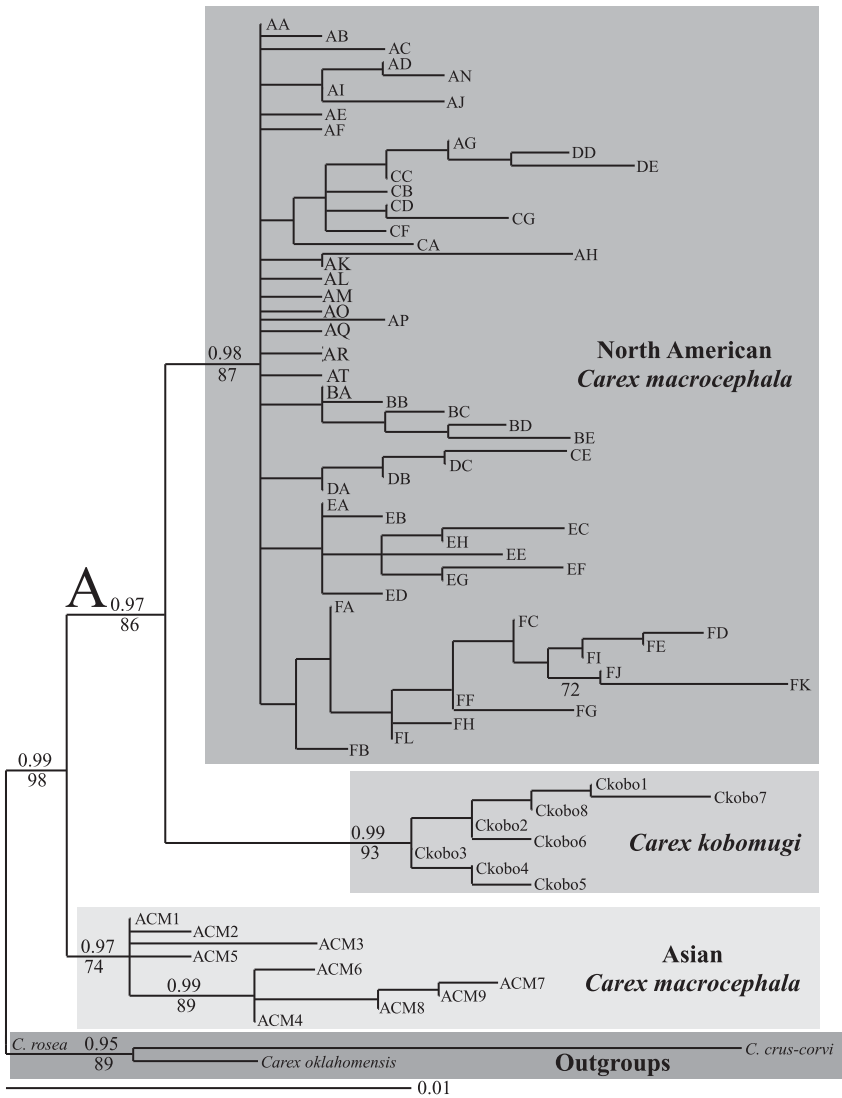
microsatellite mutation rates in monocots (Thuillet *et al.* 2002; Vigouroux *et al.* 2002). The program IMA is able to compute the geometric mean of mutation rates when provided a mutation rate and confidence interval for each locus.

## Results

### *rpL16* phylogenetic hypotheses

Within the 227 sampled individuals of *Carex macrocephala* from North America, 57 different haplotypes were sequenced at a total aligned sequence length of 748 base pairs. The entire aligned sequence was within the noncoding intergenic spacer region. Each haplotype was assigned an arbitrary two-letter code based on a preliminary placement within a neighbour-joining tree. Of the 14 individuals of Asian *C. macrocephala*, nine haplotypes were recovered, and for the 14 individuals of *Carex kobomugi*, eight haplotypes were also recovered with a total length of 756 base pairs. The DT-ModSel algorithm selected the HKY + I + G model of nucleotide substitution with a transition/transversion ratio of 0.8382, a proportion of invariant sites of 0.7681, and a gamma shape parameter of 0.6655. Bases frequencies were estimated to be A: 0.4259, C: 0.1376, G: 0.1505, and T: 0.2850. The maximum-likelihood phylogeny is shown in Fig. 2 with maximum-likelihood bootstrap and Bayesian posterior probabilities shown on branches. The populations of North American *C. macrocephala* are monophyletic and are sister to a monophyletic *C. kobomugi*. This relationship would infer a paraphyletic *C. macrocephala*, with the Asian samples of *C. macrocephala* coming out as sister to the North American *C. macrocephala*-*C. kobomugi* clade. All these relationships are well supported, and the support of a monophyletic ingroup is high at 98% bootstrap and > 0.99 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 gaps within the data sets. Neither did treating gaps as missing data or coding gaps a fifth 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 Fig. 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* data set, suggest CK and NACM are sister to each other. When gaps were treated as



**Fig. 2** Maximum-likelihood phylogram of the *Carex macrocephala* species complex ( $-\ln L = 1846.17868$ ) with three outgroup species *C. oklahomensis*, *C. crus-corvi*, and *C. rosea*. Main lineages are comprised of Asian *C. macrocephala*, *C. kobomugi*, and North American *C. macrocephala*. Haplotypes were arbitrarily assigned two-letter codes based on placement within a neighbour joining tree. Numbers above branches are Bayesian posterior probabilities and numbers below branches are maximum-likelihood bootstrap percentages.

a fifth 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.

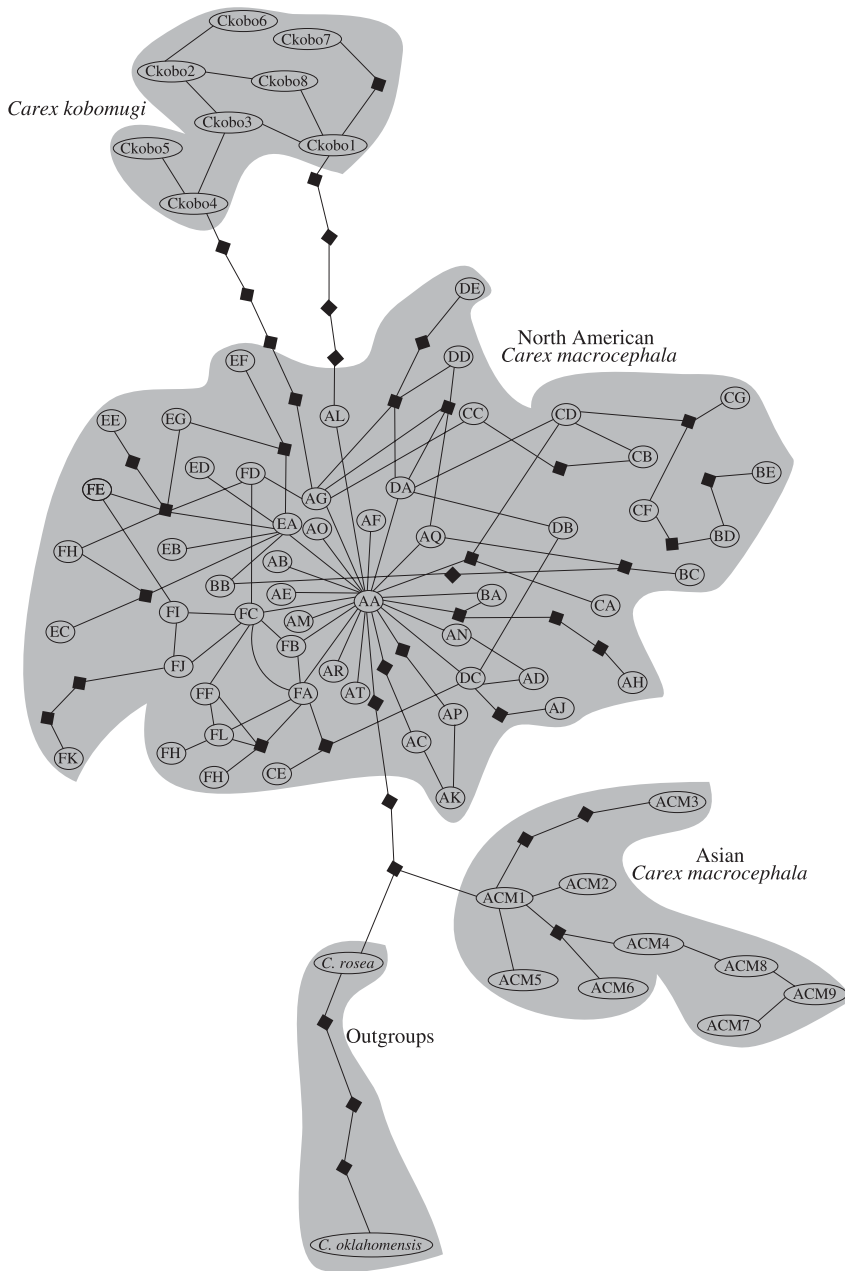
*Population genetic summary statistics*

There were 59 segregating sites ( $S_{total}$ ) for the entire *rpL16* data set, with  $S_{NACM} = 40$ ,  $S_{ACM} = 12$ ,  $S_{CK} = 7$ . Mean pairwise nucleotide difference values were  $\pi_{total} = 5.141$ ,  $\pi_{NACM} = 3.357$ ,  $\pi_{ACM} = 3.747$ ,  $\pi_{CK} = 2.109$ , and the population diversity indices,  $\theta$ , based on segregating sites were  $\theta_{(S)total} = 10.922$ ,  $\theta_{(S)NACM} = 7.772$ ,  $\theta_{(S)ACM} = 3.773$ ,  $\theta_{(S)CK} = 2.201$ . An insignificant mean  $F_{ST}$  value of  $0.048 \pm 0.056$  was computed for the *rpL16* samples from across North America. Fixation indices suggest a similar pattern for the microsatellite data sets with an  $F_{ST}$  value of  $-0.023 \pm 0.031$  for all of NACM and pairwise values of  $F_{ST(NACM-ACM)} =$

$0.127$   $P = 0.003$ ,  $F_{ST(ACM-CK)} = 0.429$   $P < 0.0001$ ,  $F_{ST(NACM-CK)} = 0.587$   $P < 0.0001$ . None of the loci showed significant linkage disequilibrium with the smallest  $P$  value = 0.242.

*Lineage divergence*

The IMA analyses were run until the effective sample sizes were at least 100 and the autocorrelation suggests we have adequate convergence and mixing of the Markov chains. Joint estimates of divergence time are given as the multidimensional peak locations of divergence time by the IMA program. Those estimates of divergence time for each pairwise analysis are  $\tau_{NACM-ACM} = 1.542$ ,  $\tau_{ACM-CK} = 1.980$ , and  $\tau_{NACM-CK} = 4.734$  (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 IMA results provided a geometric mean of mutation rates as  $1.23 \times 10^{-5}$ /locus/year.



**Fig. 3** Parsimony network of *Carex macrocephala* species complex (tcs program, 98% parsimony) 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.

### 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,  $m_{\text{NACM-ACM}} = 0.145$  and  $m_{\text{ACM-NACM}} = 7.066$  (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  $\theta$ , the population size scaled by mutation rate, for each of the pairwise analyses. For the NACM to ACM estimate, we obtained a value of  $\theta_{\text{NACM}} = 1.813$ ,  $\theta_{\text{ACM}} = 1.336$ , and  $\theta_{\text{A}} = 5.764$ ; ACM to CK a

value of  $\theta_{\text{ACM}} = 1.721$ ,  $\theta_{\text{CK}} = 1.705$ , and  $\theta_{\text{A}} = 8.627$ ; and NACM to CK a value of  $\theta_{\text{NACM}} = 2.584$ ,  $\theta_{\text{CK}} = 3.117$ , and  $\theta_{\text{A}} = 6.358$  (Table 1). All multidimensional peak locations fall within the 95% confidence interval of the marginal posterior density distribution.

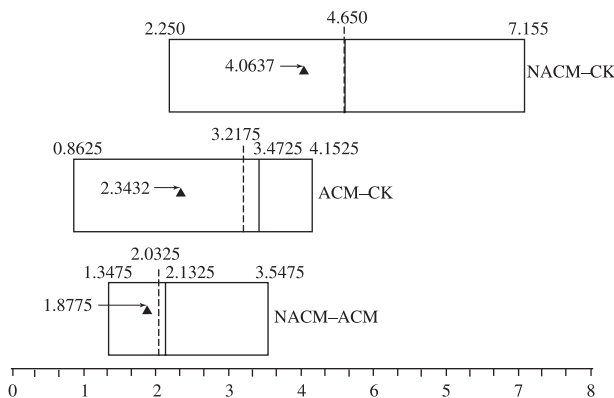
### Discussion

#### *Phylogenetic vs. coalescent-based hypotheses*

The level of variation observed in the *Carex macrocephala* species complex for the *rpL16* spacer is quite high for a cpDNA marker. This spacer region has not been utilized as

**Table 1** Results of fitting the *rpl16* cpDNA marker and eight microsatellite loci to the isolation-with-migration model implemented in IMA to determine divergence times of the *Carex macrocephala* species complex. Parameter estimates are presented for each pairwise analysis where the effective population size,  $\theta$ , divergence time,  $\tau$ , and migration rate,  $m$ , are each scaled by the mutation rate, 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 $\tau$
		$\theta_{NACM}$	$\theta_{ACM}$	$\theta_A$	$m_{NACM-ACM}$	$m_{ACM-NACM}$	
NACM-ACM	Peak	1.813	1.336	8.764	0.145	7.066	1.542
	Mean	1.814	1.396	8.653	0.554	5.959	1.571
	95Low	1.384	1.218	5.372	0.055	2.385	1.275
	95High	2.453	1.653	13.193	1.515	9.555	2.025
ACM-CK	$\theta_{ACM}$		$\theta_{CK}$	$\theta_A$	$m_{ACM-CK}$	$m_{CK-ACM}$	$\tau$
	Peak	1.721	1.705	8.627	0.013	0.001	1.980
	Mean	1.720	1.696	8.050	0.010	0.017	1.923
	95Low	1.384	1.366	5.329	0.003	0.007	0.935
NACM-CK	$\theta_{NACM}$		$\theta_{CK}$	$\theta_A$	$m_{NACM-CK}$	$m_{CK-NACM}$	$\tau$
	Peak	2.584	3.117	6.358	0.001	0.011	4.734
	Mean	2.600	2.797	6.432	0.010	0.021	4.834
	95Low	1.771	1.529	4.878	0.001	0.001	2.873
	95High	3.638	4.586	7.109	0.071	0.079	6.413



**Fig. 4** Marginal boxplots 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*; CK, *C. kobomugi*). Solid line indicates the mean bin value for each marginal distribution, and the triangles represent the multidimensional peak estimates obtained from IMA.

often as the *trnL-trnF* region or the *psbA-trnH* region, and it may be difficult to determine the level of variation possible. We took the necessary steps to ensure the polymorphisms observed in this data were real and not an artefact of amplification or sequencing. A study using additional microsatellite and cpDNA loci as well as additional individuals suggests *C. macrocephala* in North America has not gone through a significant genetic bottleneck in the past 100 000 years (M.G. King *et al.*

unpublished data). Other groups have begun finding higher levels of variation, and with the rapid employment of whole chloroplast genome, sequencing many more hypervariable regions will become available (Shaw & Small 2005; Watanabe *et al.* 2006; Shaw *et al.* 2007).

There is strong phylogenetic support for a sister relationship between the North American lineage of NACM and *C. kobomugi* CK when the *rpl16* spacer is used to construct the phylogeny. 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 a range expansion of an early lineage into a southern Asian distribution, followed quickly by a dispersal and expansion into North America, all the while a separate lineage of *C. macrocephala* in Asia was in isolation. Alternatively, early gene flow between the ancestral NACM and CK lineages may lead to a similar pattern found in the *rpl16* data. However, the results of our coalescent analyses suggest the phylogenetic patterns observed in the *rpl16* spacer may be artefacts of lineage sorting, and not a true representation of the relationships within this group. According to the IMA results, the oldest divergence times were between NACM and CK, and the most recent divergence time was between NACM and ACM. Typically incomplete lineage sorting, not reciprocal monophyly, is the cause of misinterpretation in phylogeographical analyses (for other examples see Maddison 1997).

If we had included single-copy nuclear markers in the original phylogeny reconstruction, the issue of a paraphyletic *C. macrocephala* may have been avoided. Unfortunately, most developed single copy nuclear markers for plants

show little to no variation within *Carex* and nuclear ribosomal DNA markers show high levels of incomplete lineage sorting within *C. macrocephala* (King & Roalson 2008b). We chose to develop microsatellites *de novo* to address these issues. King & Roalson (2008a) and Ohsako & Yamane (2007) were the first studies to publish microsatellites for any species in *Carex*, and as such were the logical choice as nuclear markers for our coalescent analyses. By including microsatellite loci in the analyses not only did fixed differences between the *C. macrocephala* lineages disappear, but the relationships became clear. It is interesting to note that the cpDNA loci showed no shared haplotypes between any of the *C. macrocephala* and *C. kobomugi* lineages, but that the microsatellite loci did. A population genetic explanation of this could be asymmetric pollen flow between the lineages. The benefits of nuclear sequences in coalescent analyses are higher than the microsatellites but only at adequate variability levels. However, simply including additional nuclear loci in a phylogenetic analysis does not resolve issues surrounding the lack of statistical support in phylogeographical inferences.

The power of coalescent-based models is their ability to 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  $\tau$  and the ancestral  $N_e$  (Edwards & Beerli 2000). Phylogeographical analyses which rely solely on a phylogenetic tree or haplotype network to make inferences of divergence time ignore the coalescent variance and ignore the time difference between gene divergence and population divergence. Only when  $N_e \approx 0$  or when  $\tau/N_e$  is large can we assume  $D = t$  (Edwards & 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 the insignificant overall estimates of  $F_{ST}$  values for the NACM lineage of 0.0477 for the *rpL16* samples and  $-0.0233$  for the microsatellite data set, there is no apparent population subdivision (Wakeley 2000). The results of our pairwise  $F_{ST}$  analyses alone suggest a different set of relationships than the *rpL16* data set, where the lowest  $F_{ST}$  value (0.1277) is between NACM and ACM lineages and the highest  $F_{ST}$  value (0.5872) is between the NACM and CK lineages.

The influence of lineage sorting on phylogenetic and coalescent analyses has been well discussed in the theoretical and empirical literature (Doyle 1992; Maddison 1997; Maddison & Knowles 2006; Carstens & Knowles 2007; Eidesen *et al.* 2007; Knowles & Carstens 2007a; Rosenberg & Tao 2008). Plant systematic analyses have discussed the benefits of multiple loci for phylogeny reconstruction, and several studies have used microsatellite loci for this purpose when hypervariability is needed to distinguish between species in very closely related taxa. This has not been well

investigated in plant phylogeographical analyses where incomplete lineage sorting can be prevalent or where phylogenies with well-supported reciprocally monophyletic lineages are taken for granted (Golden & Bain 2000; Comes & Abbott 2001; Holderegger & Abbott 2003; Smitsen *et al.* 2004; Bittkau & Comes 2005; Shaw & Small 2005; Jakob & Blattner 2006; Chung *et al.* 2007; Liston *et al.* 2007; Syring *et al.* 2007; Pucas *et al.* 2008; Yuan *et al.* 2008). It is difficult to determine whether these issues have a large impact on plant phylogeographical inferences, but given the variability in the usefulness of well-established sequencing loci, we could guess that errors associated with complete lineage sorting in plant phylogeography could be severe.

#### *Violation of the isolation-with-migration model*

As stated in the Methods section, there are two key assumptions of the isolation-with-migration (IM) model of divergence implemented in IMA. First, the two populations in the analysis are the most closely related, and second, that gene flow is only occurring between these two populations. In our study, there are three lineages diverging from each other, and so one of the analyses violates the assumption of IM model. This can make inferences difficult unless we make a further assumption. We need to assume that the oldest divergence event, NACM–CK, is an artefact of the other two divergence events, NACM–ACM and CK–ACM. This would infer that the NACM and CK lineages were not the most closely related, and that they were independently diverging from each other because each lineage was diverging from ACM. This is a direct contradiction to the phylogenetic hypothesis, where we would hypothesize that NACM and CK share the most recent common ancestor. Likewise, the analysis comparing the ACM and CK populations may also violate these assumptions, but due to the low levels of migration, this would be a minor violation. 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 IM 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 phylogeographical inferences using the divergences between NACM–ACM and ACM–CK.

#### *Phylogeographical inferences*

The inferences described above are statistically supported by the values obtained in the IMA analyses and from the population genetic indices, and we can use the estimates of  $\theta$  to attempt to reconstruct a possible explanation of early evolution within this species complex. In all pairwise

comparisons, the ancestral effective population sizes were larger than the total current estimated effective population size of the two lineages. This implies a reduction in  $N_e$  greater than would be expected for a vicariance event, where in vicariance  $N_{e1} + N_{e2} = N_{eA}$ . By dividing out the geometric mean of mutation rates from  $\theta$ , we can estimate the effective population size of each lineage as  $N_{e(NACM)} \approx 147\,000$ ,  $N_{e(ACM)} \approx 109\,000$ – $140\,000$ ,  $N_{e(CK)} \approx 38\,000$ . Based on extensive population observations, this estimate of  $N_e$  for NACM seems reasonable.

Estimating divergence time through coalescent simulations has an advantage over estimating divergence time from a phylogeny by accounting for the coalescent variance. Because the comparison between NACM and CK violates the IMA model, this would suggest that the ancestral lineage existed in Asia. The divergence time estimates would be nonsensical if the ancestral lineage existed in North America. The divergence between *C. macrocephala* and *C. kobomugi* may have occurred through vicariance, which is supported by the near-equal effective population size estimates of these lineages. 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 & Hudson 1991; Wakeley 2003).

Migration only persisted for the NACM–ACM lineages and gene flow is mostly occurring from ACM to NACM. Furthermore, there are no shared chloroplast 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 artefact of incomplete lineage sorting or homoplasy within these markers.

By rescaling by the mutation rate from  $\tau$ , 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 to 163 000 years; and 160 000 years for the ACM–CK divergence with a 95% confidence interval of 75 000 to 296 000 years. The added benefit of using a multilocus approach with a geometric mean of mutation rates is that the mutation rate error for each individual locus will not have a large impact on our estimates of  $N_e$  or divergence time in years. For example, if we were off by our estimates of mutation rate for the *rpL16* spacer by an order of magnitude, our estimate of divergence between NACM and ACM would change to 86 660 years and 153 000 years. Contrast that with a true change of an order of magnitude of each estimate if we only used a single locus.

The divergence times of 125 000 years and 160 000 years before present suggests a divergence for all lineages corresponding to the last interglacial warming period prior to the onset of deep glaciation approximately 110 000 years

ago (Lisiecki & Raymo 2007). Additionally, there is palaeoclimatic evidence to suggest sea level was at least 6 m higher than present levels during the last interglacial period, which may have contributed to the subsequent isolation of populations along the northwest North American coast (Brigham-Grette & Hopkins 1995). The  $N_{e(A)}$  of the NACM–ACM lineage is 4.83 $\times$  the  $N_{e(NACM)}$  and 6.56 $\times$  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 by the movement of the continental ice sheets.

## Conclusions

In this study, we fit the isolation-with-migration model of lineage divergence to our molecular data set 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 a direct contradiction 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 phylogeographical 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 phylogeographical hypothesis than that estimated using phylogenetic approaches of a single cpDNA marker.

## Acknowledgements

We extend our gracious thanks to Timothy Carey, Danielle Davis, and Kristina Bearden for all their help in laboratory work. We also wish to acknowledge the help of Andrew Friske, Erica Wheeler, Chris Kissinger (Ministry of Environment, British Columbia), John McIntosh (Pacific Rim National Park), President Guujaaw and the Haida Nation, 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. We would also like to thank Michael Webster, Richard Gomulkeiwicz, Julianno Sambatti, Rose Andrew, and two anonymous

referees for reviewing this manuscript. This research was financially supported by the Native Plant Society of Oregon, Betty Higinbotham Trust, and Washington State University.

## References

- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F (2004) Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, **20**, 407–415.
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.
- Bittkau C, Comes HP (2005) Evolutionary processes in a continental island system: molecular phylogeography of the Aegean *Nigella arvensis* alliance (Ranunculaceae) inferred from chloroplast DNA. *Molecular Ecology*, **14**, 4065–4083.
- Brigham-Grette J, Hopkins DM (1995) Emergent marine record and paleoclimate of the last interglaciation along the northwest Alaskan coast. *Quaternary Research*, **43**, 159–173.
- Brunsfeld S, Sullivan J (2005) 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, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. In: *Integrating Ecological and Evolutionary Processes in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 319–339. Blackwell Science, Oxford, UK.
- 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.
- Carstens BC, Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology*, **56**, 400–411.
- Chung JD, Lin TP, Chen YL, Cheng YP, Hwang SY (2007) Phylogeographic study reveals the origin and evolutionary history of a *Rhododendron* species complex in Taiwan. *Molecular Phylogenetics and Evolution*, **42**, 14–24.
- Clement M, Posada D, Crandall KA (2000) tcs: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
- Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution*, **55**, 1943–1962.
- 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.
- Dolman G, Moritz C (2006) A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*, **60**, 573–582.
- Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144–163.
- Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.
- Eidosen PB, Alsos IG, Popp M, Stensrud O, Suda J, Brochmann C (2007) Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species. *Molecular Ecology*, **16**, 3902–3925.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (Version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- Golden JL, Bain JF (2000) Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packera* (Asteraceae) species in southwestern Alberta. *Evolution*, **54**, 1566–1579.
- Hey J, Machado CA (2003) The study of structured populations — new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 535–543.
- Hey J, Nielsen R (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, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.
- Holderegger R, Abbott RJ (2003) Phylogeography of the Arctic-Alpine *Saxifraga oppositifolia* (Saxifragaceae) and some related taxa based on cpDNA and ITS sequence variation. *American Journal of Botany*, **90**, 931–936.
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.
- Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution*, **23**, 1602–1612.
- King MG, Roalson EH (2008a) Isolation and characterization of 11 microsatellite loci from *Carex macrocephala* (Cyperaceae). *Conservation Genetics*. DOI 10.1007/s10592-008-9558-5.
- King MG, Roalson EH (2008b) Exploring evolutionary dynamic of nrDNA in *Carex* subgenus *vignea*. (Cyperaceae). *Systematic Botany*, **33**, 514–524.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.
- Knowles LL, Carstens BC (2007a) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.
- Knowles LL, Carstens BC (2007b) Estimating a geographically explicit model of population divergence. *Evolution*, **61**, 477–493.
- Lisiecki LE, Raymo ME (2007) Plio–Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, **26**, 56–69.
- Liston A, Parker-Defeniks M, Syring JV, Willyard A, Cronn R (2007) Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in *Pinus lambertiana*. *Molecular Ecology*, **16**, 3926–3937.
- Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.
- Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.
- Mastrogiuseppe J (2002) *Flora of North America: North of Mexico*, Vol. 23. *Magnoliophyta: Commelinidae (in Part): Cyperaceae* (eds Flora of North America Editorial Committee). Oxford University Press, New York.

- Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Nielsen 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.
- Ohsako T, Yamane K (2007) Isolation and characterization of polymorphic microsatellite loci in Asiatic sand sedge, *Carex kobomugi* Ohwi (Cyperaceae). *Molecular Ecology Notes*, **7**, 1023–1025.
- Pucas M, Choler P, Tribsch A *et al.* (2008) Post-glacial history of the dominant alpine sedge *Carex curvula* in the European Alpine System inferred from nuclear and chloroplast markers. *Molecular Ecology*, **17**, 2417–2429.
- Rambaut A (1996) *Se-Al: Sequence Alignment*, ed. available at <http://evolve.zoo.ox.ac.uk/>.
- Roalson EH, Columbus JT, Friar EA (2001) Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and trnT-L-F (cpDNA) region sequences: assessment of subgeneric and sectional relationships in *Carex* with emphasis on section Acrocystis. *Systematic Botany*, **26**, 318–341.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rosenberg NA, Tao R (2008) Discordance of species trees with their most likely gene trees: the case of five taxa. *Systematic Biology*, **57**, 131–140.
- Shaw J, Small RL (2005) Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany*, **92**, 2011–2030.
- Shaw J, Lickey EB, Beck JT *et al.* (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.
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.
- Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.
- Smitsen RD, Breitwieser I, Ward JM (2004) Phylogenetic implications of trans-specific chloroplast DNA sequence polymorphism in New Zealand Gnaphalieae (Asteraceae). *Plant Systematics and Evolution*, **249**, 37–53.
- Sullivan J (2005) Maximum-likelihood estimation of phylogeny from DNA sequence data. In: *Molecular Evolution: Producing the Biochemical Data, Part B. Methods in Enzymology* (eds Zimmer E, Roalson EH), pp. 757–779. Elsevier Inc., New York.
- Swofford DL (2003) *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Syring J, Farrell K, Businsky R, Cronn R, Liston A (2007) Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology*, **56**, 163–181.
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.
- Thuillet A-C, Bru D, David J *et al.* (2002) Direct estimation of mutation rate for 10 microsatellite loci in Durum wheat, *Triticum turgidum* (L.) Thell. ssp. *durum* desf. *Molecular Biology and Evolution*, **19**, 122–125.
- Vigouroux Y, Jaqueth JS, Matsuoka Y *et al.* (2002) Rate and pattern of mutation at microsatellite loci in *Maize*. *Molecular Biology and Evolution*, **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. In: *The Evolution of Population Biology* (eds Singh R, Uyenoyama M), pp. 193–215. Cambridge University Press, Cambridge, UK.
- Watanabe K, Kajita T, Murata J (2006) Chloroplast DNA variation and geographical structure of the *Aristolochia kaempferi* group (Aristolochiaceae). *American Journal of Botany*, **93**, 442–453.
- Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007) Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution*, **24**, 90–101.
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA*, **84**, 9054–9058.
- Yuan QJ, Zhang ZY, Peng H, Ge S (2008) Chloroplast phylogeography of *Dipentodon* (Dipentodontaceae) in southwest China and northern Vietnam. *Molecular Ecology*, **17**, 1054–1065.

---

This research is being submitted in partial fulfillment of the Ph.D. research for MGK conducted under the supervision of EHR. MGK research interests are focused on population genomics and the use of coalescent theory to understand important events in the evolution of organisms. MGK is currently working as a postdoctoral research fellow at the University of British Columbia, and is researching intragenic recombination in homoploid hybrid sunflowers. EHR is an associate professor at Washington State University where his interests include phylogenetics and systematics of sedges and gesneriads. EHR also works closely with the Centre for Integrated Biotechnology at WSU, and is currently working on projects in *Eleocharis*, *Carex*, *Cyrtandra*, and *Achemines*.

---

## Appendix I

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

Location no.	Location	Latitude N	Longitude W	Sample size
1	Mouth of Moose Creek, OR	44°21.626'	124°53.83'	4
2	Mouth of Beaver Creek, OR	44°31.429'	124°44.17'	6
3	Taft, OR	44°55.747'	124°07.05'	3
4	Siletz Bay, OR	44°55.674'	124°14.17'	2
5	Neskowin Beach, OR	45°6'	123°58.2'	8
6	Whalen Island South, OR	45°16.396'	123°57.013'	6
7	Whalen Island North, OR	45°16.719'	123°57.02'	5
8	Camp Magruder, OR	45°34.981'	123°57.89'	5
9	Rockaway Beach, OR	45°37.56'	123°56.612'	6
10	Del Rey Beach, OR	46°28.84'	123°55.857'	3
11	Camp Rilea Beach, OR	46°6.859'	123°56.696'	5
12	Leadbetter Point State Park 1, WA	46°31.648'	124°27.93'	4
13	Leadbetter Point State Park 2, WA	46°36.429'	124°25.91'	8
14	Midway Beach, WA	46°46.145'	124°56.51'	6
15	Westport Lighthouse State Park, WA	46°53.237'	124°07.409'	6
16	La Push Beach, WA	47°54.95'	124°38.541'	6
17	Crescent Beach, WA	48°9.737'	123°42.42'	7
18	Fort Worden State Park, WA	48°8.883'	122°45.678'	7
19	Deception Pass State Park, WA	48°23.478'	122°38.848'	3
20	Spencer Spit State Park, WA	48°32.422'	122°51.318'	3
21	Port Renfrew, BC	48°32.16'	124°24.65'	6
22	Long Beach 1 PRNP, BC	49°43.47'	125°46.012'	6
23	Long Beach 2 PRNP, BC	49°43.47'	125°46.012'	7
24	Mackenzie Beach PRNP, BC	49°7.993'	125°54.157'	6
25	Wickaninnish Beach PRNP, BC	49°11.46'	125°40.385'	6
26	Oyster Beach, BC	49°53.73'	125°8.811'	6
27	San Josef Beach 1, BC	50°40.468'	128°16.606'	6
28	San Josef Beach 2, BC	50°40.468'	128°16.606'	6
29	Rose Spit, BC	54°10.179'	131°39.396'	5
30	Naikoon Provincial Park, BC	54°6.762'	131°42.873'	10
31	Tow Hill, BC	54°43.35'	131°47.463'	8
32	Tlell Beach, BC	53°34.73'	131°55.908'	6
33	Pasagshak Beach Kodiak Island, AK	57°27.514'	152°27.021'	6
34	Shelikof Beach Kruzof Island, AK	57°10.186'	135°45.356'	6
35	Shelikof Beach 2 Kruzof Island, AK	57°9.981'	135°45.362'	3
36	Yakutat Canon Beach, AK	59°29.569'	139°43.631'	9
37	Yakutat Coast Guard Beach, AK	59°30.597'	139°46.442'	5
38	Mouth of Kenai River, AK	60°34.2'	151°15'	9
39	Kalifornsky Beach, AK	61°31.43'	151°16.168'	2
40	Kasilof Beach, AK	60°23.354'	151°17.79'	6

OR, Oregon; WA, Washington; BC, British Columbia; AK, Alaska.

## Appendix II

Specimens sampled from herbarium sheets for samples collected in Asia

	Location	Collection
	<i>Asian Carex macrocephala</i>	
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	Sakaehama Island	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, Russia	s.n. (WTU)
14	Hokkaido, Sakanoshita, Japan	s.n. (HAK)
	<i>Carex kobomugi</i>	
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)

## Appendix III

Autocorrelation and effective sample size (ESS) estimates for each IMA analysis, where TMRCAs are the time to the most recent common ancestor for loci 0–8 (0, rplL16 cpDNA spacer; 1, CM01; 2, CM07; 3, CM27; 4, CM39; 5, Cko1–78; 6, Cko1–12; 7, Cko1–68; 8, Cko1–134; A, NACM-ACM analysis; B, ACM-CK analysis; C, NACM-CK analysis)

A

Steps	$L(P)$	$t$	TMRCa0	TMRCa1	TMRCa2	TMRCa3	TMRCa4	TMRCa5	TMRCa6	TMRCa7	TMRCa8
1	0.9343	0.9215	0.8672	0.8876	0.8619	0.8663	0.9073	0.8778	0.951	0.9164	0.8855
10	0.4989	0.3281	0.4544	0.3091	0.3105	0.3181	0.3224	0.4264	0.5349	0.4248	0.306
50	0.4212	0.2924	0.3803	0.2853	0.1762	0.288	0.2862	0.1938	0.1373	0.2805	0.296
100	0.4082	0.2452	0.3233	0.257	0.1517	0.2743	0.2533	0.1469	0.1185	0.2685	0.2721
500	0.3209	0.2284	0.2906	0.2364	0.1405	0.2007	0.2262	0.1021	0.0956	0.187	0.2095
1000	0.2607	0.1767	0.2367	0.1888	0.1211	0.1842	0.1761	0.0876	0.0531	0.111	0.1782
5000	0.2486	0.1559	0.2031	0.161	0.1272	0.1527	0.1682	0.0669	0.0215	0.101	0.1615
10 000	0.2154	0.1496	0.1942	0.1472	0.0775	0.1406	0.1446	0.0419	0.0117	0.1284	0.1433
50 000	0.2084	0.1293	0.1695	0.1279	0.0685	0.1294	0.1386	0.0228	-0.022	0.134	0.1358
100 000	0.158	0.1155	0.1421	0.098	0.0233	0.1244	0.0906	-0.0259	0.0204	0.1117	0.098
500 000	0.1533	0.0701	0.1235	0.0494	0.0234	0.1036	-0.0203	-0.0352	-0.0471	0.0875	0.0705
1 000 000	0.099	-0.0137	0.0714	-0.015	0.0195	0.0955	-0.0238	-0.0272	-0.0555	0.0701	0.0373
ESS	164	222	375	887	933	634	1071	1195	876	599	958

B

Steps	$L(P)$	$t$	TMRCa0	TMRCa1	TMRCa2	TMRCa3	TMRCa4	TMRCa5	TMRCa6	TMRCa7	TMRCa8
1	0.9484	0.9297	0.936	0.9014	0.8579	0.8864	0.91	0.8898	0.9352	0.8964	0.8539
10	0.5824	0.5363	0.5293	0.3001	0.3421	0.5378	0.6213	0.3762	0.3975	0.6902	0.4636
50	0.4429	0.2797	0.2585	0.2398	0.2951	0.227	0.2561	0.2014	0.2201	0.1767	0.276
100	0.2937	0.1551	0.1547	0.1698	0.199	0.1604	0.1518	0.1604	0.1973	0.132	0.1613
500	0.2678	0.1221	0.1328	0.1585	0.1502	0.1221	0.1219	0.0411	0.1408	0.1009	0.165
1000	0.2023	0.1067	0.1295	0.12	0.1152	0.1119	0.117	0.0237	0.1247	0.0793	0.1287
5000	0.1221	0.0866	0.0763	0.0865	0.055	0.0884	0.0995	0.0233	0.0481	0.0576	0.0694
10 000	0.1408	0.0653	0.0635	0.0762	0.0666	0.0401	0.0548	0.016	0.0563	0.0668	0.0432
50 000	0.1048	0.0346	0.06	0.0516	0.0797	0.0404	0.0431	0.0159	0.0143	0.0328	0.0492
100 000	0.0565	0.0649	0.0134	0.0397	0.0215	0.0383	0.0342	0.018	0.0061	0.0142	0.0365
500 000	0.0114	0.0062	0.0372	-0.0181	0.0221	0.0264	0.0266	0.004	0.0032	0.0068	0.0164
1 000 000	0.0154	0.0371	-0.0106	-0.0436	0.0146	-0.0256	-0.0272	-0.0225	-0.0277	-0.0172	-0.0041
ESS	234	289	412	686	489	459	461	603	447	373	562

C

Steps	$L(P)$	$t$	TMRCa0	TMRCa1	TMRCa2	TMRCa3	TMRCa4	TMRCa5	TMRCa6	TMRCa7	TMRCa8
1	0.947	0.9462	0.9433	0.9469	0.96	0.9271	0.948	0.9638	0.9328	0.9537	0.9195
10	0.8152	0.8114	0.569	0.3344	0.8171	0.4641	0.6325	0.3326	0.3803	0.7635	0.815
50	0.4615	0.4493	0.5476	0.4087	0.5283	0.3764	0.389	0.3677	0.3085	0.3459	0.3443
100	0.1936	0.1926	0.1619	0.1702	0.1527	0.177	0.1884	0.154	0.1563	0.1557	0.1764
500	0.16	0.1595	0.1645	0.1463	0.1313	0.1681	0.1342	0.1522	0.1539	0.1668	0.1379
1000	0.1299	0.1292	0.1204	0.1122	0.1095	0.1126	0.1258	0.1295	0.1119	0.1268	0.1213
5000	0.1055	0.105	0.0668	0.0724	0.0766	0.0981	0.0713	0.1055	0.0751	0.0805	0.1053
10 000	0.0759	0.0583	0.0744	0.0483	0.0647	0.0635	0.0709	0.0601	0.0433	0.0494	0.0698
50 000	0.0745	0.073	0.0672	0.0511	0.0653	0.0363	0.0673	0.0552	0.0347	0.0477	0.0674
100 000	0.0788	0.0744	0.0424	0.0729	0.0579	0.037	0.0456	0.0301	0.0724	0.0438	0.0451
500 000	0.0685	0.0632	0.0355	0.0466	0.0408	0.0202	0.0475	0.0533	0.0209	0.0432	0.0347
1 000 000	0.0425	0.0402	0.0341	0.0031	0.0245	0.0226	0.0097	0.0102	0.0109	0.018	0.0071
ESS	224	276	320	362	374	340	349	291	471	387	310