Phylogenetic, Morphological, and Chemotaxonomic Incongruence in the North American Endemic Genus Echinacea

Lex Flagel
Iowa State University

Ryan A. Rapp
Iowa State University

Corrinne E. Grover
Iowa State University, corrinne@iastate.edu

Mark P. Widrlechner
United States Department of Agriculture, isumw@iastate.edu

Jennifer S. Hawkins
University of Georgia

Follow this and additional works at: http://lib.dr.iastate.edu/eeob_ag_pubs

Part of the Botany Commons, Ecology and Evolutionary Biology Commons, Molecular Genetics Commons, and the Plant Breeding and Genetics Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/eeob_ag_pubs/60. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Phylogenetic, Morphological, and Chemotaxonomic Incongruence in the North American Endemic Genus Echinacea

Abstract
The study of recently formed species is important because it can help us to better understand organismal divergence and the speciation process. However, these species often present difficult challenges in the field of molecular phylogenetics because the processes that drive molecular divergence can lag behind phenotypic divergence. In the current study we show that species of the recently diverged North American endemic genus of purple coneflower, Echinacea, have low levels of molecular divergence. Data from three nuclear loci and two plastid loci provide neither resolved topologies nor congruent hypotheses about species-level relationships. This lack of phylogenetic resolution is likely due to the combined effects of incomplete lineage sorting, hybridization, and backcrossing following secondary contact. The poor resolution provided by molecular markers contrasts previous studies that found well-resolved and taxonomically supported relationships from metabolic and morphological data. These results suggest that phenotypic canalization, resulting in identifiable morphological species, has occurred rapidly within Echinacea. Conversely, molecular signals have been distorted by gene flow and incomplete lineage sorting. Here we explore the impact of natural history on the genetic organization and phylogenetic relationships of Echinacea.

Keywords
Asteraceae, chloroplast DNA, Echinacea, incomplete lineage sorting, phylogenetics, single-copy nuclear DNA

Disciplines
Botany | Ecology and Evolutionary Biology | Molecular Genetics | Plant Breeding and Genetics

Comments

Rights
Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.

Authors
Phylogenetic, morphological, and chemotaxonomic incongruence in the North American endemic genus Echinacea

Author(s): Lex E. Flagel, Ryan A. Rapp, Corrinne E. Grover, Mark P. Widrlechner, Jennifer Hawkins, Jessie L. Grafenberg, Inés Álvarez, Gyu Young Chung and Jonathan F. Wendel

Source: *American Journal of Botany*, Vol. 95, No. 6 (June 2008), pp. 756-765

Published by: Botanical Society of America, Inc.

Stable URL: http://www.jstor.org/stable/27793066

Accessed: 07-10-2015 20:14 UTC
PHYLOGENETIC, MORPHOLOGICAL, AND CHEMOTAXONOMIC INCONGRUENCE IN THE NORTH AMERICAN ENDEMIC GENUS ECHINACEA

LEX E. FLAGEL, RYAN A. RAPP, CORRINNE E. GROVER, MARK P. WIDRLECHNER, JENNIFER HAWKINS, JESSIE L. GRAFFENBERG, INÉS ÁLVAREZ, GYU YOUNG CHUNG, AND JONATHAN F. WENDEL

Species of the genus Echinacea are geographically circumscribed within a region of North America that has undergone repeated rounds of glaciation (Clayton and Moran, 1982), with the last such round, the Wisconsinan, ending roughly 10,000 yr before the present. Presently, the genus ranges from southern Alberta, Canada to the coast of the Gulf of Mexico in Texas and Louisiana and from the oak savannas of Ohio, the glades of Tennessee, and open habitats in the Carolinas west to the foothills of the Rocky Mountains (Urbatsch et al., 2006). Much of this range was under ice during the last glacial epoch, signifying that Echinacea survived in southerly refugia. Despite the expansive aggregate range of the genus, much of this range has been converted into agricultural production, resulting in an extremely fragmented modern population structure. This distributional history has many potential implications for the genetic architecture of a perennial plant species, most importantly, the disruption of natural processes of intraspecific and interspecific gene flow and the attendant increase in population fragmentation and genetic bottlenecks.

Taxonomically, Echinacea is delimited into nine species (Table 1), including two, E. angustifolia DC and E. paradoxa (Norton) Britton, that each are further divided into two varietals (McGregor, 1968; Flora of North America Editorial Committee, 1993+; McKeown, 1999). These species are all diploid with the exception of E. pallida, which is putatively a polyploid (Mechanda et al., 2004). This taxonomic treatment was devised by McGregor (1968), who spent 15 years studying the genus while making controlled, common-garden crosses, noting that many hybrids have high levels of stability, fecundity, and viability in parent backcrosses. In a recent morphological study, four species with eight subspecies were proposed (Binnis et al., 2002, 2004), but McGregor’s classification continues to be widely used by botanists and herbalists (see discussion in Blumenthal and Urbatsch [2006]) and serves as the basis for the recent Flora of North America treatment (Urbatsch et al., 2006).

McGregor’s results regarding the ease of formation and fertility of interspecific hybrids suggest that Echinacea may either be a young genus in which rapid speciation has occurred (McKeown, 2004) or one in which, for reasons other than relative youth, genetic barriers have incompletely formed. In either case, gene flow between species has been historically common; McGregor noted hybrid swarms in natural sympatric settings, and in more recent molecular work, Mechanda et al. (2004) found evidence of natural hybrid individuals.

Assessment and maintenance of Echinacea genetic diversity is of interest due to the purported human health benefits from several Echinacea species (Speroni et al., 2002), as well as the cultivation and breeding of the plant as an ornamental...
Echinacea taxa characterized, with U. S. state of origin and USDA Germplasm Resources Information Network Plant Introduction (PI) accession.  

<table>
<thead>
<tr>
<th>Taxon</th>
<th>State</th>
<th>PI Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. angustifolia</td>
<td>OK</td>
<td>PI 631267</td>
</tr>
<tr>
<td>E. angustifolia var. angustifolia</td>
<td>IA</td>
<td>PI 631272</td>
</tr>
<tr>
<td>E. angustifolia var. angustifolia</td>
<td>KS</td>
<td>PI 631318</td>
</tr>
<tr>
<td>E. angustifolia var. strigosa</td>
<td>OK</td>
<td>PI 631266, PI 631220</td>
</tr>
<tr>
<td>E. atrovirens</td>
<td>KS</td>
<td>PI 631255, PI 631299</td>
</tr>
<tr>
<td>E. laevigata</td>
<td>SC</td>
<td>PI 631310, PI 631312</td>
</tr>
<tr>
<td>E. pallida</td>
<td>OK</td>
<td>PI 631275</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>OK</td>
<td>PI 631263, PI 631264, PI 631265</td>
</tr>
<tr>
<td>E. paradoxa var. paradoxa</td>
<td>AR</td>
<td>PI 631292</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>LA</td>
<td>PI 636669</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>MO</td>
<td>PI 631307</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>NC</td>
<td>PI 631313</td>
</tr>
<tr>
<td>E. sanguinea</td>
<td>LA</td>
<td>PI 631257, PI 631258, PI 636372</td>
</tr>
<tr>
<td>E. simulata</td>
<td>KY</td>
<td>PI 631249</td>
</tr>
<tr>
<td>E. tennesseensis</td>
<td>TN</td>
<td>PI 631308</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PI 631250, PI 631234, PI 631235, PI 631236</td>
</tr>
</tbody>
</table>

Notes: AR, Arkansas; IA, Iowa; LA, Louisiana; MO, Missouri; NC, North Carolina; OK, Oklahoma; SC, South Carolina; TN, Tennessee; VA, Virginia

In contrast, a recent study (L. Wu, Iowa State University; P. Dixon, B. Nikolau, G. Kraus, M. Widlechner, and E. Wurtele, unpublished manuscript) of 40 populations of Echinacea, selected to encompass a broad geographical and morphological diversity, examined metabolite profiles generated by HPLC, and reported that patterns of biochemical diversity corresponded well to taxonomic circumscriptions and relationships as conveyed in McGregor’s (1968) monograph. In addition, a morphological study by Binns et al. (2002), although proposing an alternative treatment, used character data to produce a clustering pattern with node support reflecting McGregor’s original treatment.

In this study, we sought to elucidate a phylogenetic framework for Echinacea by using both nuclear and plastid loci. Our goal was to describe genetic relationships among the nine congeners and reveal the parental origin of the polyplody species. The data, however, revealed a history of secondary contact and hybridization, mirroring the glaciation-entwined history and shedding light on some of the processes giving rise to conflicting molecular assessments of phylogenetic relationships.

**MATERIALS AND METHODS**

**Plant material**—We selected 38 accessions of Echinacea (Table 1, Appendix 1) representing the full geographic range of the Echinacea germplasm collection in the U. S. National Plant Germplasm System maintained by the USDA-ARS North Central Regional Plant Introduction Station (NCRPIS), Ames, Iowa (Widlechner and McKeown, 2002). These accessions span the extremes of Echinacea’s geographic distribution and include several accessions from areas where species exist in sympathy (Missouri, Kansas, and Oklahoma). Accessions were keyed to species during the initial regeneration process on the basis of McGregor (1968).

Seed samples of the 38 accessions were soaked for 24 h in a 1 mmol solution of ethephon to overcome dormancy and promote rapid germination (Sari et al., 2001). After soaking, the seeds were transferred to clear plastic boxes with blotters moistened with distilled water. The germination boxes were held at 4°C for 4 weeks and then transferred to germination chambers at a constant 25°C with 14 h of light per day. Three-week-old seedlings were transferred into 20-cm pots in a growing medium consisting of 50% Canadian peat moss, 40% perlite, and 10% mineral soil, and grown under ambient light in a greenhouse at 22–25°C, with daily watering and biweekly fertilizing.

**Tissue preparation and DNA extraction**—Greenhouse grown plants were key to selecting species (McKeown, 1968) at sexual maturity to verify identities as received from the NCRPIS. Young leaves and flower buds were collected for DNA extraction and flash frozen in liquid nitrogen. Samples were ground under liquid nitrogen, and DNA was extracted from 100 mg aliquots using the DNeasy Plant mini DNA extraction kit.

**Outgroup selection**—Over 50 members of the Heliantheae and several suspected close allies from Zinnieae and Ecliptineae (Appendix 1) were used in the initial sequencing of the trnG plastid locus to create unrooted trees (Appendix S1, see Supplemental Data with online version of this article). From these analyses, the genus Sonnativula was found to be sister to Echinacea and was treated as the outgroup in subsequent analyses. DNA used in outgroup selection was obtained from previous studies where vouchers have already been deposited (Urbatsch et al., 2000).

**Locus amplification, molecular cloning, and sequencing**—Nuclear loci—Because there is little sequence information available for the genus Echinacea (14 sequences deposited in GenBank as of 11/16/07), nuclear loci were selected that previously have demonstrated high utility in molecular phylogenetic studies. Three nuclear loci were selected: alcohol dehydrogenase (Adh), cellulose synthase (CesA), and glycerolaldheyde 3-phosphate acetyl transferase (GPAT) (Table 2). These loci were selected based on their previous utility in species-level molecular systematic studies (Adh [Sang et al., 1997; Small and Wendel, 2000]; CesA [Cron et al., 2002; Sanchiru et al., 2003]; and GPAT [Tank and Sang, 2001]). Degenerate primers were used to perform preliminary locus amplification, after which Echinacea-specific primers were designed to

---

*All use subject to JSTOR Terms and Conditions*
span exonic and intronic regions such that amplicons of ~800–1100 bp could be generated for each locus (Table 2).

PCR amplification of the three nuclear loci was performed using the following generalized protocol: initial denaturation phase of 95°C for 5 min, 35 cycles of amplification at 95°C for 30 s, primer-specific annealing temperature (Table 2) for 45 s, 72°C elongation for 60 s. After 35 cycles of amplification, a final elongation phase of 72°C for 7 min was used to complete polymerization. PCR reactions were conducted in a 40 µL volume of 1× Taq polymerase buffer, 100–500 ng total genomic DNA, 2.0 mM MgCl₂, 0.4 µM of both forward and reverse primers, 0.25 mM dNTPs, and 2 units of Taq polymerase (Bioline USA, Randolph, Massachusetts, USA).

Individual PCR amplification products were visualized on 1% agarose gels, and products were excised and extracted with the Qiagen QIAEX II Gel Extraction Kit. The purified product was cloned into the pGEM-T Easy vector system (Promega, Madison, Wisconsin, USA), and M13R following the ABI-Prism Big Dye Terminator sequencing method (version 3.1; Applied Biosystems, Foster City, California, USA).

Amplification products were visualized on 1% agarose gel, with T Jacques (1998) or manually. Ambiguous bases from the ends of the were removed manually or using the trimseq program from the EMBOSS software package (Rice et al., 2000) with the following parameter settings (“window = 20” and “percent = 10”). Next, a consensus read was generated from the forward and reverse sequence reads using MUSCLE 3.52 multiple alignment software (Edgar, 2004). Each output alignment (hereafter referred to as a clone sequence) was saved for further analysis.

Clone sequences were imported and manually inspected with BioEdit sequence alignment viewing software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Ambiguous bases in each clone sequence were corrected manually by comparing sequence quality from trace files. Corrected clones were assembled into individual-specific files and aligned with MUSCLE 3.52. Nontarget sequences were visibly detected and removed. Once alllelic variants could be visually identified, consensus sequences were segregated into allele-specific files, which were aligned with MUSCLE 3.52 and condensed into one allelic consensus sequence. All allelic variants from each nuclear locus surveyed were collated into one final file, which was used for all downstream phylogenetic analyses. For each of the three nuclear loci, the allele number and nucleotide diversity (π) (Nei, 1987) (computed with the program DNASP 4.0 [Rozas et al., 2003]) were calculated and tabulated (Tables 3 and 4).

Recombinant sequences can arise naturally through homologous recombination or artifically via PCR strand swapping (Bradley and Hillis, 1997), making detection and removal of recombinant sequences important, because they increase homoplasy and confound interpretation. We used two separate recombination detection algorithms, MaxChi (Smith, 1992) and SiScan (Gibbs et al., 2000), as implemented in the RDP-V2 program (Martin et al., 2005). A P-value of 0.01 was used as a threshold for significance when applied to 1000 parametric bootstrap replicates for both MaxChi and SiScan. Recombinant events detected within individuals, and thus likely arising from PCR strand-swapping, were removed from the analyses. Recombinant events detected between taxa, and thus likely arising naturally through hybridization and homologous recombination, were rare. The few detected events were not significant when step-down, multiple-testing correction was applied. These few sequences, though possibly recombinant, were left in the analyses. Paralogy tests were conducted by first identifying potential paralogous sequences from a phylogenetic tree (e.g., multiple placements for clones from

---

**Table 2. Amplification conditions and primers for nuclear genes used in phylogenetic analysis in *Echinacea*.**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene portion sequenced</th>
<th>Forward nonspecific primer sequence (Forward consensus-specific primer sequence)</th>
<th>Reverse nonspecific primer sequence (Reverse consensus-specific primer sequence)</th>
<th>Annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh</td>
<td>Exon 7–9</td>
<td>CTGCCTGTGCAATGGGARCGACGGAAG (CTGTTTGAACAAATAGCAGC)</td>
<td>GCAACGCACACCCCACACCTG (TCCTGTATCACAATTCAGG)</td>
<td>51</td>
</tr>
<tr>
<td>CesA</td>
<td>Exon 9–12</td>
<td>CTGGCTCTGGAATTAACCCGGCTG (GCTGTATGCCATGAGGATTG)</td>
<td>ATGCTCACACCCGCCTATG (CGTGATGACCAACGCTTG)</td>
<td>59.5</td>
</tr>
<tr>
<td>GPAT</td>
<td>Exon 4–9</td>
<td>TTGGYGGACATTATTATGTCGCC (TGGCTTGTGAGATCAAACTG)</td>
<td>CACCACGAGCCCTGCACCCCTG (CATATTACCTTGACGACCC)</td>
<td>56</td>
</tr>
</tbody>
</table>

* Full-length sequences from *Echinacea* were not obtained; exon/intron position was estimated from orthologous full-length genes in closely related species.

---

**Table 3. Sequence information for *Echinacea* nuclear loci.**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Individuals sequenced (species)</th>
<th>Total alignment bp (coding length)</th>
<th>No. alleles detected</th>
<th>% Heterozygosity all (diploids)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh</td>
<td>35 (8)</td>
<td>880 (485)</td>
<td>116</td>
<td>94.3 (92.9)</td>
</tr>
<tr>
<td>CesA</td>
<td>71 (9)</td>
<td>950 (644)</td>
<td>141</td>
<td>73.2 (74.19)</td>
</tr>
<tr>
<td>GPAT</td>
<td>84 (9)</td>
<td>1264 (337)</td>
<td>148</td>
<td>67.9 (63.89)</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Phylogenetic gene-tree reconstruction for the CesA, Adh, and GPAT gene loci. These phylogenies represent the Bayesian consensus trees and include node support values based on Bayesian posterior probabilities and branch lengths drawn relative to sequence divergence. The outgroup is indicated by “OG” (Sanvitata frutiosa for Adh and GPAT; Zinnia violacea for CesA). The retention index (RI) and homoplasy index (HI) values are documented above each tree. All *Echinacea* species are classified as in McGregor (1966) and coded as follows: *E. angustifolia* (red), *E. atrorubens* (orange), *E. lavrignata* (light green), *E. pallida* (purple), *E. paradoxa* (yellow), *E. purpurea* (dark blue), *E. sanguinea* (brown), *E. simulata* (light blue), and *E. tennesseense* (dark green).
June 2008] FLAGEL ET AL. — PHYLOGENETIC ANALYSIS OF THE GENUS ECHINACEA 759
combined manually, and the resulting sequences were aligned in the program CLUSTAL X (Thompson et al., 1997). Minimal manual adjustment was necessary because sequence diversity was low.

**Phylogenetic analyses—Nuclear loci**—Three different phylogenetic analyses were applied to the data: (1) distance-based analyses with nonparametric bootstrapping, performed with the program Phylip 3.63 (Felsenstein, 1989); (2) parsimony analyses using PAUP+ 4.0 (Swofford, 2001); and (3) Bayesian likelihood analyses with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). All three analyses gave highly congruent topologies (data not shown), though only Bayesian phylogenies have the advantage of retaining both branch length and the rate model, which was drawn from a gamma distribution while allowing unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model), and the rate model, which was altered to allow unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model), and the rate model, which was altered to allow unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model), and the rate model, which was altered to allow unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model), and the rate model, which was altered to allow unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model).

For all phylogenies, 2–4 MCMC runs were initiated, each with a minimum of 10,000,000 generations (more generations used when needed to reach stationarity). Prior distribution settings were applied to the data: (1) distance-based analyses with nonparametric bootstrapping, performed with the program Phylip 3.63 (Felsenstein, 1989); (2) parsimony analyses using PAUP+ 4.0 (Swofford, 2001); and (3) Bayesian likelihood analyses with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). All three analyses gave highly congruent topologies (data not shown), though only Bayesian phylogenies have the advantage of retaining both branch length and node support; thus, only these phylogenies are shown (Fig. 1).

Baysian phylogenies for each nuclear locus were estimated using a Markov chain Monte Carlo (MCMC) sampler of tree space as implemented by MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). For all phylogenies, 2–4 MCMC runs were initiated, each with a minimum of 10,000,000 generations (more generations used when needed to reach stationarity). Prior distribution settings were left at the default values, with the exceptions of the nucleotide substitution model, which was altered to allow unique rates of substitution among or between all pairs of nucleotides (e.g., the general time reversible [GTR] model), and the rate model, which was chosen using a grid of priors allowing for invariant sites. Runs were started from a random tree and allowed to proceed in parallel while sampling and recording the topology every 100 generations of the MCMC chain. Performance of individual runs was assessed and phylogenies compared between runs. Majority rule (>50%) consensus trees were constructed after removing the “burn-in period” samples (the first 10% of sampled trees). Topologically, Bayesian analyses were highly congruent between runs, indicating that multiple MCMC chains consistently achieved stationarity around the same subset of possible topologies.

Further, as an exploratory tool we used the group-assignment program Structure 2.1 (Pritchard et al., 2000) to give an alternative view of the gene-sequence data. With the goal of understanding genomic levels of gene flow, the Structure algorithm performs best with randomly sampled unlinked markers, such as AFLPs and simple sequence repeats (SSRs); however, it can be used with nuclear sequence data as an exploratory tool, giving a graphical overview of population structure within these data. Analyses were run separately on polymorphic base pairs from the three nuclear gene datasets. For each data set, the number of clusters (K) was incremented from 1 to 12, and the best fit was assessed with a likelihood ratio test. Group assignments were plotted for individuals and species for each value of K (Appendix S2, see Supplemental Data with online version of article).

**Plastid loci—Plastid nucleotide diversity within Echinacea and among outgroup species was low. We applied the same Bayesian phylogeny reconstruction method used for nuclear loci. The Bayesian tree search algorithm was allowed to run for 100,000,000 generations, achieving stationarity and the resulting consensus tree can be found in Fig. 2.**

---

**Table 4.** Nucleotide diversity (π) partitioned between various genomic domains. SD values are the standard deviations associated with each estimate of π.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Overall (SD)</th>
<th>Coding regions (SD)</th>
<th>Noncoding region (SD)</th>
<th>Syn sites</th>
<th>Nonsyn sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh</td>
<td>0.03224 (0.00082)</td>
<td>0.03947 (0.00103)</td>
<td>0.02128 (0.00100)</td>
<td>0.06857</td>
<td>0.0338</td>
</tr>
<tr>
<td>CesA</td>
<td>0.01491 (0.00035)</td>
<td>0.00628 (0.00031)</td>
<td>0.03846 (0.00111)</td>
<td>0.02538</td>
<td>0.0064</td>
</tr>
<tr>
<td>GPAT</td>
<td>0.01166 (0.00046)</td>
<td>0.00457 (0.00038)</td>
<td>0.01545 (0.00063)</td>
<td>0.01401</td>
<td>0.0034</td>
</tr>
</tbody>
</table>

**Notes:** Nonsyn, nonsynonymous sites: nucleotide sites that cause an amino acid replacement if substituted. Syn, synonymous sites: nucleotide sites that do not cause an amino acid replacement if substituted.

---

**RESULTS**

**Nuclear loci sequence characteristics**—In total, approximately 3.1 Mb of Echinacea nuclear DNA were sequenced, including 1 Mb for Adh, 1.2 Mb for CesA, and 0.93 Mb for GPAT. After allelic sequences were processed and identified, the raw data generated approximately 92 kb, 138 kb, and 151 kb, respectively, of total unaligned sequence data for phylogenetic analyses. For GPAT and Adh, we also amplified and sequenced nuclear loci from Sanvitalia fruticosa Hensl. We were unable to amplify the CesA locus in S. fruticosa and instead used Zinnia violacea Cav., a close relative, (GenBank accessions AF323039, AF323040, and AF323041) as an outgroup for this data set.

Observed levels of heterozygosity were between approximately 68 and 94% (Table 3). Alleles were defined strictly by haplotype, thus alleles may differ at a minimum by a single nucleotide polymorphism. We have applied this strict assessment because we have sequenced multiple clones per individual (between 8–12), allowing us the opportunity to remove many PCR and sequencing errors. Heterozygosity values for CesA and GPAT were similar (68% and 73%). However, much higher heterozygosity was found in Adh (~94%), possibly due to an increased substitution rate caused by a loss of purifying selection on pseudogenized sequences at this locus. In all cases, heterozygosity was relatively high, which could result from a reported sporophytic self-incompatibility system in the genus Echinacea (McKeown, 2004; Stephens, in press).

Overall mean values of nucleotide diversity, π, ranged from ~0.012 to ~0.032 for the three nuclear loci (Table 4). We partitioned nucleotide diversities between coding and noncoding regions and synonymous and nonsynonymous sites within coding regions (Table 4). As expected, the levels of π for CesA and GPAT were greater in noncoding than in coding regions (6.12 and 3.38 times greater, respectively); likewise, π in synonymous sites was approximately 39.7 and 7.2 times greater, respectively, than in nonsynonymous sites.

A rather different pattern was observed at the Adh locus, which had the highest overall mean values of π (0.03229). In addition, levels of π were approximately 54% lower at noncoding sites (0.02128) than at coding sites (0.03947), and levels of π were only approximately two times higher when comparing synonymous to nonsynonymous sites (0.06857 to 0.0338). These statistics would be unusual for a functional nuclear gene experiencing neutral evolution (Li, 1997). These factors, along with stop codons and indels in the open-reading frames of several taxa (data not shown), suggest that the Adh locus we sequenced represents either a pseudogene or possibly a nuclear locus with nonfunctional allelic variants. We were unable to isolate orthologous Adh loci.
from some taxa \(E. \textit{atrorubens} (\text{Nutt.} \ \text{Nutt.})\), likely due to the higher rate of loss of pseudogenized genes. For this reason, taxon sampling in the \textit{Adh} data set remains incomplete. We report these findings regarding the limited phylogenetic utility of the \textit{Adh} locus in the hopes that it may be avoided in future studies in \textit{Echinacea}. Also, it serves as an example of one of the pitfalls often encountered when selecting nuclear loci for phylogenetic studies.

**Phylogenetic results—Nuclear loci**—Topologies of the three nuclear gene trees are shown in Fig. 1. Overall, few species form monophyletic groups with respect to these gene trees. The exceptions are \textit{E. laevigata} and \textit{E. tennesseensis}, which both form monophyletic groups in the \textit{GPAT} tree. In addition, we observed no phylogenetic differentiation between varietal groups within either \textit{E. angustifolia} or \textit{E. paradoxa} (data not shown), and thus, we removed varietal designations from Fig. 1. The Bayesian \textit{GPAT} gene tree divides the genus into two clades, with alleles from \textit{E. angustifolia}, \textit{E. atrorubens}, \textit{E. paradoxa}, and \textit{E. tennesseensis} in one clade and alleles from \textit{E. laevigata}, \textit{E. purpurea}, \textit{E. sanguinea} Nutt., and \textit{E. simulata} McGregor in the other. This split was also observed in our parsimony and distance-based trees, with 100% bootstrap support in the latter (data not shown). There are, however, a few exceptions to this division, i.e., one \textit{E. sanguinea} and two \textit{E. simulata} alleles can be found within the \textit{E. angustifolia}, \textit{E. atrorubens}, \textit{E. tennesseensis}, and \textit{E. paradoxa} clade. Such a division has been documented by others (Kim et al., 2004), though these authors found \textit{E. laevigata} sister to \textit{E. tennesseensis}. Barring this exception, the first chronological divergence that takes place in the \textit{GPAT} phylogeny is well-supported by multiple phylogenetic methods and by the AFLP data from Kim et al. (2004). Resolution beyond this initial division in the \textit{GPAT} phylogeny becomes less clear because many of the taxa share alleles with other taxa; the exceptions are \textit{E. tennesseensis} and \textit{E. laevigata} as noted.

The topological patterns of the \textit{Adh} and \textit{CesA} gene phylogenies are more complex than that of \textit{GPAT}. Neither phylogeny has a single monophyletic species; furthermore, there is often reliable node support for polyphyletic associations in both phylogenies. Notably though, in both phylogenies there frequently are small clades of alleles from the same species, although these clades are paraphyletic with regard to species in all cases. An additional complicating factor is a high level of homoplasy found in the \textit{CesA} phylogeny. The most parsimonious class of tree scores (of which there were many equally parsimonious trees) overall had a homoplasy index \((HI) \) of 0.53. Comparatively, the \textit{Adh} and \textit{GPAT} loci had \(HI \) values of 0.16 and 0.22 respectively. All three nuclear phylogenies are populated by both short internal and terminal branches; thus, it is not surprising that there are several unresolved polytomies.

**Plastid loci**—Plastid analysis found the relationship of genera allied with \textit{Echinacea}. The Mexican and Southwest US endemic genus \textit{Sanvitalia} appears as sister to \textit{Echinacea} with good node support (Fig. 2; online Appendix S1). Within the genus nucleotide diversity is extremely low and results in a phylogenetic hypothesis rich in polytomies (Fig. 2). Using this plastid phylogeny of the genus, we compared the genetic distances to geographic distances via a Mantel test as implemented in the program \textit{PASSaGE} (Rosenberg, 2001). This test demonstrated that the genetic structure of these \textit{Echinacea} plastid loci is statistically correlated to their relative geographic distances from one another \((P < 0.05)\) and not to taxon label.

**DISCUSSION**

The primary goals of this study were to generate a large collection of sequence data for diverse populations of \textit{Echinacea} and to use these data to reconstruct a species-level phylogeny. Previous attempts to reconstruct the genetic and evolutionary relationships of \textit{Echinacea} (Kapteyn et al., 2002; Binns et al., 2004; Kim et al., 2004; Mechanda et al., 2004) have provided phylogenetic resolution among particular taxa but have been limited by their depth of population sampling and/or number of phylogenetically useful characters. Our approach to determine the evolutionary history of this genus was to use plastid and nuclear loci, the latter typically offering excellent resolution at the species level due to relatively high rates of sequence divergence when the assumptions of the phylogenetic model are met. Nuclear sequence data, however, may also be phylogenetically problematic or misleading. For example, hybridization, paralogy, incomplete lineage sorting, and secondary contact are all features or processes capable of obfuscating organismal-level relationships in a phylogenetic framework. In addition to these biological features, technical issues, such as PCR recombinants (Small et al., 2004) and automated sequencer base-calling errors, play an increasingly troubling role as nucleotide diversity decreases among the taxa being sampled.

Although no definitive species-level relationships may be formed from the topologies generated, several features of the data set are particularly striking, given the connections of this genus to North American glacial history and geography. In the plastid data, close ties between genetic structure and geographical distribution suggest a prominent role of past rounds of glaciation. First, the cytotypes suggest southerly refugia on either side of the Mississippi River, with both containing a unique cytotype along with other cytotypes present in both refugia. The nuclear data further corroborate the idea of secondary, postglacial, contact between species with incomplete reproductive barriers. Telling aspects include low sequence diversity but a high number of alleles, broad taxonomic distribution of nearly identical alleles, and incongruent topologies between loci.

Notwithstanding the general absence of species-level monophyly in the trees generated from nuclear loci, these trees do offer some insight into species origins and history. One example concerns the origins of the polyploid species \textit{E. pallida}, which appears interspersed throughout the diploid phylogeny. The placement of \textit{E. pallida} with \textit{E. angustifolia}, \textit{E. atrorubens}, and \textit{E. laevigata} alleles is frequent, though \textit{E. pallida} can be found sister to other taxa as well (Fig. 1). Additionally, based on our population-structure analysis, designations for \textit{E. pallida} tended to be assorted evenly among several groups (online Appendix S2B, D, and F). These results indicate that the polyploid \textit{E. pallida} was either formed more than once from different parental origins or that the formation of \textit{E. pallida} took place at a basal level in the genus and the observed patterns at the tips of the trees are artifacts of subsequent hybridization and incomplete lineage sorting (Wendel and Doyle, 1998; Small et al., 2004). Given the shallow nature of these trees, which yield poor basal resolution, it is difficult to determine which hypothesis is correct.

In stark contrast to the sequence data reported here, a recent chemotaxonomic study by Wu et al. (L. Wu, Iowa State University; P. Dixon, B. Nikolau, G. Kraus, M. Widrlechner, and E. Wurtele, unpublished manuscript) finds strong support for McGregor’s (1968) taxonomic treatment when individuals are clustered based on their metabolite profiles. Additionally, a
recent morphological analysis by Binns et al. (2002) also contains node support for McGregor’s classification. It is interesting to note that although our analysis of neutral gene variation suggests supraspecific, landscape-scale processes at work, the physiology and morphology are consistent with well-differentiated and adapted species, perhaps reflecting specific ecological niches. It is possible that the high degree of physiological and morphological integrity has been maintained by selection on relatively few loci, which were not sampled during this study. Alternatively, in the absence of selection it is possible that neutral processes fixed these traits during glacial maxima, when population sizes were presumably small.

In either case, future studies seeking to elucidate genetic relationships within this genus should probably employ marker technology that has broad genomic coverage, such as AFLPs, as the phylogenetic signal within the nuclear and plastid genome appears to be extremely weak. This study has also highlighted the importance of including geographically representative individuals from all species, as using a subset of these data could easily generate an incorrect yet well-supported topology.

LITERATURE CITED


ROSENBERG, M. S. 2001. PASSAGE: Pattern analysis, spatial statistics, and geographic exegesis. Department of Biology, Arizona State University, Tempe, Arizona, USA.


Appendix I. The taxa analyzed in this study, collection locality, and GenBank accession numbers have been shortened with a hyphen; e.g., EU423454-5 indicates accessions EU423454, EU423455, and EU423456. Nonsyntactical GenBank accessions are separated with a comma. All nonapplicable values are indicated with a dash (—).

**Taxon**—Collection locality; PI accession; Voucher; GenBank accession codes; Abb: Cesa; GPAP; trnM; trnS.
Parthenium

Echinacea purpurea; LSU; et al. 92; EU440192, EU440202, EU440216, EU440224. Echinacea purpurea; Louisiana; PI 631307; EU423526-7; EU423652-5; EU423812-4; EU440168, EU440176-7. EU440230, EU440244, EU440287. Echinacea purpurea; Missouri; PI 631307; Rapp et al. 76; ISC; EU423528-32; EU423656-7; EU423815-7; EU440114, EU440182; EU440185; EU440299, EU440249. Echinacea purpurea; North Carolina; PI 631133; Rapp et al. 71; ISC; EU423553-7; EU423658-60; EU423818-20; EU440082, EU440105; EU440234, EU440249.

Echinacea sanguinea; LSU; PI 631215; Rapp et al. 94; ISC; EU423538-9; EU423661-2; EU423821-2; EU440424, EU440425, EU440427. Echinacea sanguinea; LSU; PI 631258; EU423540-2; EU423663-6; EU423823-7; EU440087, EU440102, EU440112, EU440200, EU440215, EU440260, EU440280. Echinacea sanguinea; LSU; PI 635572; EU423543-4; EU423667-71; EU423829-30; . Echinacea simulata; Kentucky; PI 631249; Rapp et al. 92; ISC; EU423672-5; EU423831-2; EU440099, EU440104; EU440257, EU440279, EU440288. Echinacea simulata; Missouri; PI 631304; EU423676-80; EU423833-4; EU440091, EU440141; EU440235, EU440282. Echinacea simulata; Tennessee; PI 631308; Rapp et al. 89; ISC; EU423545-7; EU423681-87; EU423835-6; EU440113, EU440183-4; EU440261, EU440270. Echinacea tennesseensis; Tennessee; PI 631250; Rapp et al. 84; ISC; EU423548-51; EU423688-91; EU423837-40; EU440155, EU440163, EU440195, EU440204, EU440248. Echinacea tennesseensis; Tennessee; PI 631324; Rapp et al. 81; ISC; EU423552-4; EU423692-5; EU440281-4; EU440106, EU440137; EU440226, EU440233, EU440240. Echinacea tennesseensis; Tennessee; PI 631326; Rapp et al. 85; ISC; EU423558-61; EU423698-702; EU423849-52; EU440188; EU440220, EU440225, EU440253. Eclipta prostrata (L.) L.; Louisiana; -. Leyland s.n.; LSU; -.; -. EU586869; Encelia farinosa A.Gray ex Torr.; Arizona; -. Urbatsch 7099; LSU; -.; -. EU586875; -.

Helianthus quinquenervis (Hook.) A. Gray; Colorado; -. Barkhalter 2674; LSU; -.; -. EU586870; -. Helianthus uniflorus Torr. & Gray; Wyoming; -. Urbatsch 7094; LSU; -.; -. EU440130; -. Heliposis fitzfolia S. Watson; Mexico; -. Wendt et al. 1896; TEX; -.; -. EU586846; -. Heliposis helianthoides Sweet; Missouri; -. Urbatsch s.n.; LSU; -.; -. EU586878; -. Heliposis oppositifolia (L.) Druce; Texas; -. Butterwick 3833; TEX; -.; -. EU586845; -. Heliposis parvifolia A.Gray; Mexico; -. Wendt et al. 1662; TEX; -.; -. EU586848; -. Partenium hysterophorus L.; Louisiana; -. Urbatsch s.n.; LSU; -.; -. EU440122; -. Podacaenium pachyphylum (Klotzsch) R.K.Jansen, N.A.Harriman & Urbatsch; Mexico; -. Panero 2780; TEX; -.; -. EU586854; -. Philactis nelsoni S.F.Blake; Mexico; -. Breadlove 28267; TEX; -.; -. EU586885; EU440323. Philactis zinnioides Serad.; Mexico; -. Hartman & Funk 4184; TEX; -.; -. EU586853; -. Ratibida columnifera (Nutt.) Wooton & Standl.; Louisiana; -. Richardson 60; LSU; -.; -. EU440161; -. Ratibida mexicana (S.Watson) Sharp; Mexico; -. Panero 2257; TEX; -.; -. EU440117; -. Ratibida pinnata Barnh.; Tennessee; -. Urbatsch s.n.; LSU; -.; -. EU586880; -. Rudbeckia alpilcola Piper; Washington; -. Urbatsch s.n.; LSU; -.; -. EU586881; -. Rudbeckia grandiflora (Sweet) C.C.Gmel. ex DC.; Louisiana; -. Urbatsch s.n.; LSU; -.; -. EU586877; EU440299, EU440307. Rudbeckia laciniata L.; Texas; -. Urbatsch s.n.; LSU; -.; -. EU440118; -. Rudbeckia nitida Nutt.; Florida; -. Urbatsch s.n.; LSU; -.; -. EU586879; -. Rudbeckia triloba L.; Virginia; -. Urbatsch s.n.; LSU; -.; -. EU440142; -. Salmea oligocephala Hessm.; Mexico; -. Escobedo 1861; TEX; -.; -. EU586860; -. Salmea palmeri S.Watson; Mexico; -. Solis 644; TEX; -.; -. EU586858; -. Salmea scandens DC.; Texas; -. Diaz-Luna 19824; TEX; -.; -. EU586861; -. Sanvitalia angustifolia Engelm. ex A.Gray; Mexico; -. Chiang et al. 8820; TEX; -.; -. EU586841; -. Sanvitalia fruticosa Hessm.; Mexico; -. Medrano 1446; TEX; -.; -. EU586843; -. Sanvitalia fruticosa ; Mexico; -. Valient-Bauzer 701; TEX; -.; -. EU440183; -. EU440108; -. Sclerocephalus divaricatus Bentl. & Hook.f. ex Hessm.; Mexico; -. Feddeana 1728; TEX; -.; -. EU586844; -. Siphium compositum Michx.; Florida; -. Urbatsch s.n.; LSU; -.; -. EU440133; -. Spilanthes nervosa Chodat; Brazil; -. Irvin 1532; TEX; -.; -. EU586857; -. Spilanthes urens Jacq.; Mexico; -. Kei 15103; TEX; -.; -. EU586855; -. Wyeethia angustifolia Nutt.; California; -. Baldwin s.n.; UC/JPES; -.; -. EU440119; -. Zaluzania megacephala Sch.Bip.; Mexico; -. Patterson 6378; TEX; -.; -. EU586888; -. Zaluzania mollissima A.Gray; Mexico; -. Hendrickson 17544; TEX; -.; -. EU586889; EU440295, EU440317. Zaluzania pringlei Greenm.; Mexico; -. Koch et al. 7921; TEX; -.; -. EU586890; EU440312, EU440318. Zaluzania triloba (Ortega) Pers.; Mexico; -. Johnson 5857; TEX; -.; -. EU586891; EU440304. Zinnia acerosa A.Gray; Mexico; -. Neson 6204; TEX; -.; -. EU586830; -. Zinnia acerosa; Arizona; -. Urbatsch 7100; LSU; -.; -. EU586876; -. Zinnia flavicoma (DC.) Olorode & A.M.Torres; Mexico; -. Hartman & Funk 4192; TEX; -.; -. EU586881; EU440141. Zinnia grandiflora Nutt.; Texas; -. Sites & Smith 647; TEX; -.; -. EU586829; -. Zinnia grandiflora; Arizona; -. Urbatsch 7102; LSU; -.; -. EU440124; -. Zinnia maritima Kunth; Mexico; -. Spooner et al. 2551; TEX; -.; -. EU586832; -. Zinnia oligantha L.M.Johnst.; Mexico; -. Hendrickson 14163; TEX; -.; -. EU586833; EU440291, EU440316. Zinnia peruviana L.; Mexico; -. Mayfield et al. s.n.; TEX; -.; -. EU586896; -. Zinnia peruviana; Bolivia; -. Spooner 6549; TEX; -.; -. EU586887; EU440303. 