Molecular Ecology (2006) **15**, 2735–2751 doi: 10.1111/j.1365-294X.2006.02970.x

Phylogeny and biogeography of *Croton alabamensis* (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data

BENJAMIN W. VAN EE,* NICOLAS JELINSKI,† PAUL E. BERRY‡ and ANDREW L. HIPP§
*University of Wisconsin-Madison, Department of Botany, 430 Lincoln Drive, Madison, Wisconsin 53706, USA, †University of Wisconsin-Madison, Center for Sustainability and the Global Environment, Nelson Institute for Environmental Studies, 1710 University Avenue, Madison, Wisconsin 53726, USA, ‡The University of Michigan, Department of Ecology and Evolutionary Biology, Natural History Building (Kraus), 830 North University, Ann Arbor, Michigan 48109-1048, USA, §The Morton Arboretum, 4100 Illinois Route 53, Lisle, Illinois 60532, USA

Abstract

***Croton alabamensis* (Euphorbiaceae *s.s.*) is a rare plant species known from several populations in Texas and Alabama that have been assigned to var. *texensis* and var. *alabamensis*, respectively. We performed maximum parsimony, maximum likelihood, and Bayesian analyses of DNA sequences from the nuclear ribosomal internal transcribed spacer (ITS) and 5.8S regions and chloroplast *trn*L-*trn*F regions from collections of the two varieties of *C. alabamensis* and from outgroup taxa. *C. alabamensis* emerges alone on a long branch that is sister to *Croton* section *Corylocroton* and the Cuban endemic genus *Moacroton*. Molecular clock analysis estimates the split of *C. alabamensis* from its closest relatives in sect. *Corylocroton* at 41 million years ago, whereas the split of the two varieties of *C. alabamensis* occurred sometime in the Quaternary. Amplified fragment length polymorphism (AFLP) analyses were performed using two selective primer pairs on a larger sampling of accessions (22 from Texas, 17 from Alabama) to further discriminate phylogenetic structure and quantify genetic diversity. Using both neighbour joining and minimum evolution, the populations from the Cahaba and Black Warrior watersheds in Alabama form two well-separated groups, and in Texas, geographically distinct populations are recovered from Fort Hood, Balcones Canyonlands, and Pace Bend Park. Most of the molecular variance is accounted for by variance within populations. Approximately equal variance is found among populations within states and between states (varieties). Genetic distance between the Texas populations is significantly less than genetic distance between the Alabama populations. Both sequence and AFLP data support the same relationships between the varieties of *C. alabamensis* and their outgroup, while the AFLP data provide better resolution among the different geographical regions where *C. alabamensis* occurs. The conservation implications of these findings are discussed.**

Keywords: AFLP, *Croton alabamensis*, disjunction, genetic diversity, ITS, *trn*L-*trn*F

Received 15 November 2005; revision received 10 February 2006; accepted 17 March 2006

Introduction

Croton alabamensis E. A. Smith ex Chapman (Euphorbiaceae *s.s.*) is a rare shrub known from a few locally clustered populations in Texas and Alabama, where it occurs above the coastal plains in limestone, shale, or dolomitic outcrops and adjacent outwashes (Farmer & Thomas 1969; Ginzbarg 1992; Fig. 1). It was first discovered in 1877 in the Cahaba River watershed in the Alabama Valley and Ridge physiographic province (Chapman 1883; Farmer & Thomas 1969), and in 1905 it was found further west in the Black Warrior River watershed in the Cumberland Plateau north of Tuscaloosa, Alabama. Farmer (1962) and Farmer & Thomas (1969) made extensive searches for new localities and documented

Correspondence: Benjamin W. van Ee, Fax: (608) 262 7509; E-mail: bvanee@wisc.edu

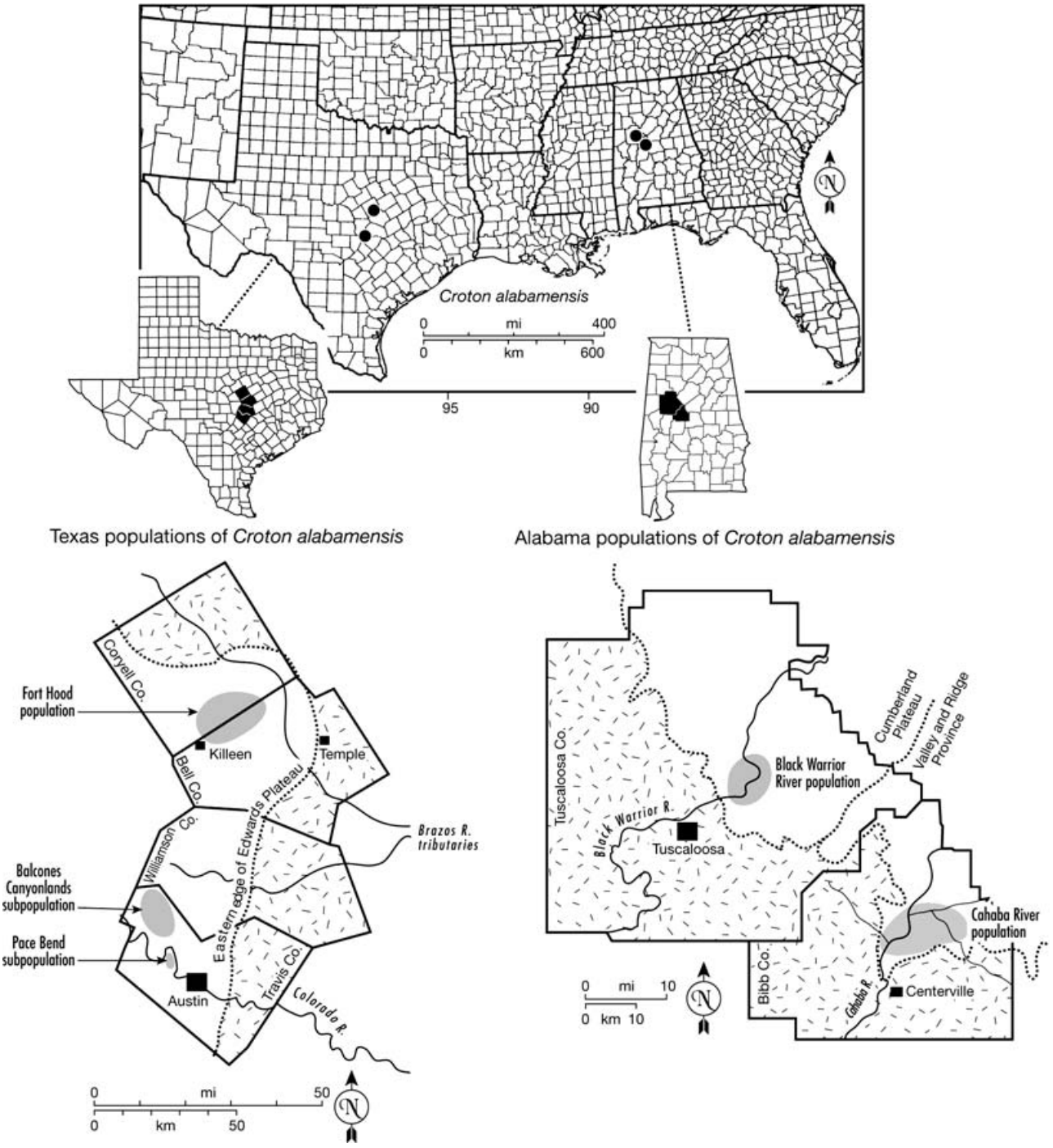

Fig. 1 Distribution maps of populations of *Croton alabamensis*. On the lower set of maps, the hatching indicates coastal plain areas, and the unhatched areas are adjacent piedmont.

additional populations in the Cahaba River watershed. They concluded that there were just two small areas of occurrence of the species about 40 km apart in adjacent Tuscaloosa and Bibb counties (Fig. 1), and they described morphological and habitat differences between plants from the two different watersheds. In a surprising discovery in 1989, *C. alabamensis* was found in central Texas, more than 1000 km from the Alabama populations (Aplet *et al*. 1994). After studying plants from both states, Ginzbarg (1992) described the Texas plants as a new variety, *C. alabamensis* var. *texensis* Ginzbarg. In the decade following, additional populations of var. *texensis* were discovered along the eastern edge of the Edwards Plateau in Bell, Coryell, and Travis counties (Fig. 1).

The current disjunction between Texas and Alabama could be the result of past vicariance events, with the separation of a formerly continuous distribution and the subsequent extinction of geographically intermediate populations. Alternatively, *C. alabamensis* could owe its current distribution to a long-distance dispersal event from one locality to the other. A third possibility is that this species colonized the two parts of its current range from a former refugium, after experiencing a severe genetic bottleneck (Watson *et al*. 2002). The taxonomic status of *C. alabamensis* — whether it should be considered a single species with two varieties or two distinct, more localized species — has obvious conservation consequences and implications for legal protection status.

Webster (1993) placed *C. alabamensis* in section *Lamprocroton* in his sectional conspectus of *Croton,* based on its lepidote indumentum, eglandular leaves, and the presence of petals in the pistillate flowers. However, a careful observation of young leaves reveals a pair of rudimentary glands at the base of the leaf blade that are later obscured by the peltate trichomes, and the simple stigmas are atypical of sect. *Lamprocroton*. Likewise, a molecular survey of *Croton* and tribe *Crotoneae* using the nuclear ribosomal internal transcribed spacer (ITS) and 5.8S regions and the chloroplast *trn*L-*trn*F region (Berry *et al*. 2005), failed to place *C. alabamensis* close to any other member sampled from sect. *Lamprocroton*, which is most diverse in southeastern Brazil. Instead, *C. alabamensis* emerged in the phylogeny near the root of the main *Croton* lineage in an isolated position sister to the Cuban endemic genus *Moacroton* and members of *Croton* sect. *Corylocroton*. *C. alabamensis* is the northernmost woody, perennial species in the genus in eastern North America, with the closest other woody species of *Croton* occurring in mostly frost-free areas of southern Florida and southern coastal Texas.

The purpose of this study is to reconstruct the phylogenetic relationships of *Croton alabamensis* and its most closely related taxa using DNA sequence data (nuclear ribosomal ITS1, 5.8S, and ITS2 — collectively known as ITS, and chloroplast *trn*L intron and *trn*L-*trn*F intergenic spacer — collectively known as *trn*L-*trn*F), and then to use amplified fragment length polymorphism (AFLP) data to evaluate population structure and genetic diversity within the species and test hypotheses that explain the evolutionary history and disjunct distribution of the species. The use of AFLP markers to infer phylogeny is still somewhat controversial, given the difficulty in assessing homology and independence among bands, as well as the inability to distinguish if there are differences in phylogenetic signal from different parts of the genome. However, several recent studies have successfully used AFLP data — mostly in conjunction with DNA sequence data — to resolve phylogenetic relationships between closely related taxa at the species and intraspecific levels (Hodkinson *et al*. 2000; Koopman *et al*. 2001; Xu & Sun 2001; Zhang *et al*. 2001; Beardsley *et al*. 2003; Després *et al*. 2003; Spooner *et al*. 2005).

Materials and methods

Taxon sampling

We visited nearly all known populations of *Croton alabamensis* in both Texas and Alabama during their flowering season in early spring (mid March) of 2003. Leaf samples and herbarium vouchers were collected from individuals throughout the range of the species. In Alabama, individuals were collected in the Cahaba River watershed in Bibb County at the following localities: Glades (A-GL; 33°03′N, 87°02′W), Highway 219 (A-HW; 33°00′N, 87°08′W), Pratt's Ferry (A-PF; 33°01′N, 87°04′W), and in Tuscaloosa County in the Black Warrior River watershed along Holt Reservoir (A-BW; 33°16′N, 87°25′W). In Texas, individuals were collected from the Fort Hood military reservation in Coryell and Bell counties (T-FH; 31°14′N, 97°34–37′W), and in Travis County in the Balcones Canyonlands National Wildlife Refuge (T-BC; 30°32–36′N, 97°58–59′W), and Pace Bend County Park (T-PB; 31°27′N, 98°00′W). Twenty-two accessions of *C. alabamensis*, 10 of variety *texensis* and 12 of variety *alabamensis*, were sequenced for ITS and *trn*L-*trn*F. Three species each from *Croton* sect. *Corylocroton* and the Cuban endemic genus *Moacroton* were included as the next closest sister taxa based on results in Berry *et al*. (2005). Three placeholder taxa from each of the three main clades of the core *Croton* clade (clades C-2 through C-11 *sensu* Berry *et al*. 2005) were included in the analyses along with an accession of *Brasiliocroton mamoninha* as an additional outgroup. Thirty-eight accessions of *C. alabamensis*, including 21 of var. *texensis* and 17 of var. *alabamensis*, were scored for AFLPs. Five additional taxa, three species of *Moacroton* and two of *Croton* sect. *Corylocroton*, were also scored for bands and included in the AFLP analyses.

Molecular methods

Total genomic DNA was extracted from silica-dried tissue of single individuals (Chase & Hillis 1991) using QIAGEN DNeasy plant kits following the manufacturer's protocol, including a 1% RNase treatment during cell lysis. Extracted DNA was suspended in QIAGEN elution buffer at maximum concentrations and stored at –20 °C. Herbarium voucher specimens, collection locality, labels, and GenBank Accession nos are listed in Table 1. For the majority of accessions, a herbarium voucher was made from the plant from which leaf material was sampled, but in some cases where multiple individuals were sampled from a close-knit population, only leaves in silica were collected.

Table 1 Taxa, vouchers, localities, labels, and GenBank Accession nos. The sectional affiliations of outgroup *Croton* taxa are given in parentheses. The accessions of *Croton alabamensis* from which DNA sequences were used to produce Fig. 2(A, B) are indicated by an asterisk

Taxa	Distribution	Label	Voucher	ITS	*trn*L-*trn*F
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-GL1	van Ee *et al*. 374 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-GL2	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-GL3*	only silica	DQ227512	DQ227544
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-HW1	van Ee *et al*. 369 (WIS)	DQ227513	DQ227545
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-HW2	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-HW3	van Ee *et al*. 370 (WIS)	DQ227514	DQ227546
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	—	Wurdack 088 (US)	AY971177	AY794692
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-PF1	van Ee *et al*. 371 (WIS)	DQ227515	DQ227547
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-PF2	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-PF3	only silica	DQ227516	DQ227548
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Bibb Co., Alabama	A-PF4	van Ee *et al*. 373 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW1*	van Ee *et al*. 363 (WIS)	DQ227506	DQ227538
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW2	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW3	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW4	van Ee *et al*. 366 (WIS)	DQ227507	DQ227539
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW5	van Ee *et al*. 366B (WIS)	DQ227508	DQ227540
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW6	only silica	DQ227509	DQ227541
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW7	van Ee *et al*. 368 (WIS)	DQ227510	DQ227542
Croton alabamensis E.A.Sm. ex Chapm. var. *alabamensis*	Tuscaloosa Co., Alabama	A-BW8	only silica	DQ227511	DQ227543
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Bell Co., Texas	T-FH10*	van Ee *et al*. 349 (WIS)	DQ227520	DQ227552
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Bell Co., Texas	T-FH11	van Ee *et al*. 350 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Bell Co., Texas	T-FH9	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH1	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH12	van Ee *et al*. 341 (WIS)	DQ227521	DQ227553
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH13	van Ee *et al*. 346 (WIS)	DQ227522	DQ227554
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH2	van Ee *et al*. 340 (WIS)	DQ227523	DQ227555
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH3	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH4	van Ee *et al*. 343 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH5	van Ee *et al*. 344 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH6	van Ee *et al*. 345 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH7	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	T-FH8	van Ee *et al*. 347 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Coryell Co., Texas	—	Carr 17733 (BRIT)	AY971178	AY971269
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC1	van Ee *et al*. 352 (WIS)	DQ227517	DQ227549
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC2	van Ee *et al*. 353 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC3	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC4*	only silica	DQ227518	DQ227550
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC5	van Ee *et al*. 356 (WIS)	DQ227519	DQ227551
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC6	van Ee *et al*. 357 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC7	van Ee *et al*. 359 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-BC8	van Ee *et al*. 360 (WIS)	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	—	Nesom 7850 (NY)	AY971179	AY971270
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-PB1	van Ee *et al*. 361 (WIS)	DQ227524	DQ227556
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-PB2	only silica	NA	NA
Croton alabamensis E.A.Sm. ex Chapm. var. *texensis* Ginzbarg	Travis Co., Texas	T-PB3	van Ee *et al*. 362 (WIS)	NA	NA
Brasiliocroton mamoninha P.E. Berry & I. Cordeiro	Espírito Santo, Brazil	NA	Pirani 3411 (NY)	AY971175	AY971267
Croton caracasanus Pittier (section *Corylocroton*)	Distrito Federal, Venezuela	NA	Riina 1288 (WIS)	DQ227525	DQ227557
Croton caudatus Geiseler (Old World *Croton*)	Palawan, Philippines	NA	Soejarto 7728 (MO)	AY971192	AY971283
Croton corylifolius Lam. (section *Corylocroton*)	Pinar del Río, Cuba	NA	HAJB 81975 (WIS)	DQ227526	DQ227558
Croton craspedotrichus Griseb. (section *Cascarilla*)	Pinar del Río, Cuba	NA	HAJB 81991 (WIS)	DQ227532	DQ227564
Croton daphniphyllum ined. Radcl.-Sm. (Old World *Croton*)	Madagascar	NA	McPherson 18310 (MO)	DQ227531	DQ227563
Croton draco Schltdl. & Cham. (section *Cyclostigma*)	Trujillo, Venezuela	NA	Riina 1261 (WIS)	DQ227533	DQ227565
Croton lundellii Standl. (section *Corylocroton*)	Yucatán, Mexico	NA	van Ee 118 (WIS)	DQ227527	DQ227559
Croton mexicanus Müll.Arg. (section *Corylocroton*)	Puntarenas, Costa Rica	NA	Haber 10714 (MO)	NA	NA
Croton nephrophyllus Urb. & Ekman (section *Cascarilla*)	Guantánamo, Cuba	NA	HAJB 81945 (WIS)	DQ227534	DQ227566
Croton niveus Jacq. (section *Eluteria*)	Oaxaca, Mexico	NA	Berry 7596 (WIS)	DQ227535	DQ227567
Croton varelae V.W. Steinm. (section *Geiseleria*)	Nayarit, Mexico	NA	Steinmann 1063 (WIS)	DQ227536	DQ227568
Croton yucatanensis Lundell (section *Argyroglossum*)	Yucatán, Mexico	NA	van Ee 121 (WIS)	DQ227537	DQ227569
Croton zambesicus Müll.Arg. (Old World *Croton*)	Songwe Gorge, Zambia	NA	Zimba 901 (MO)	AY971260	AY971341
Moacroton ekmanii (Urb.) Croizat	Holguín, Cuba	NA	van Ee 393 (WIS)	DQ227528	DQ227560
Moacroton revolutus Alain	Cult. ex Matanzas, Cuba	NA	van Ee 405 (WIS)	DQ227529	DQ227561
Moacroton trigonocarpus (Griseb.) Croizat	Pinar del Río, Cuba	NA	van Ee 380 (WIS)	DQ227530	DQ227562

Table 2 AFLP statistics for 39 accessions of *Croton alabamensis* with and without outgroup taxa (*Moacroton trigonocarpus*, *Moacroton revolutus*, *Moacroton ekmanii*, *Croton lundellii*, *Croton mexicanus*)

	Primer combination		
	1 *Eco*RI + ACG/ *Mse*I + CCA	**2** *Eco*RI + ACT/ *Mse*I + CTC	Total
No. of total bands			
without outgroup	119	142	261
with outgroup	163	189	352
No. of variable bands			
without outgroup	72	97	169
with outgroup	163	189	352
% polymorphic bands			
without outgroup	60%	68%	65%
with outgroup	100%	100%	100%

The nuclear ribosomal internal transcribed spacer (ITS) and 5.8S regions along with the cpDNA *trn*L intron and *trn*L-*trn*F spacer were amplified using published primers (White *et al*. 1990; Taberlet *et al*. 1991; Baldwin *et al*. 1995; Urbatsch *et al*. 2000). Polymerase chain reaction (PCR) products were cleaned using the AMPure magnetic bead method and sequenced at the University of Wisconsin Biotechnology Center with an ABI 3100 automated DNA sequencer. Sequences were edited and assembled in SEQUENCHER 3.0 (GeneCodes Co. 1991–1995), then aligned manually in MACCLADE 4.0 (Maddison & Maddison 2000).

AFLP protocols are modified from Vos *et al*. (1995), following protocols by M. Berres (Berres 2001; Hipp *et al*. in press). The restriction enzymes *Eco*RI and *Mse*I were used for digestion, and selective amplification employed three selective nucleotides on each primer. The *Eco*RI primer was labelled with 6-FAM fluorescent dye at the 5′ end. Sixteen primer pairs were screened for variability, from which two combinations were chosen: *Eco*RI + ACG/*Mse*I + CCA and *Eco*RI + ACT/*Mse*I + CTC (Table 2). Final PCR products were cleaned using CleanSeq beads (Amersham) and analysed on an ABI 3100 capillary electrophoresis machine with an ROX-labelled internal lane standard, with fragments at 25-bp intervals from 50 to 625 bp.

Sequence analyses

Incongruence between the ITS and *trn*L-*trn*F sequence partitions was estimated using the incongruence length difference (ILD) test (Farris *et al*. 1994), implemented as the partition homogeneity test in PAUP* version 4.0b10 (Swofford 2002), using simple taxon addition tree-bisection–reconnection (TBR) searches holding 10 trees at each step, and without limiting the maximum number of trees saved. Maximum parsimony (MP) and maximum likelihood (ML) analyses were performed in PAUP* version 4.0b10 (Swofford 2002). The MP heuristic searches used 1000 random taxon addition replicates (holding one tree at each step) and TBR branch swapping. All characters were equally weighted, and gaps were scored as either present (1) or absent (0). Bootstrap percentages were obtained from 10 000 bootstrap replicates with simple taxon addition (holding one tree at each step), TBR branch swapping, and saving no more than 1000 trees in order to speed up the process. The best-fit likelihood model was selected from among the 56 models evaluated in MODELTEST version 3.06 (Posada & Crandall 1998), using the hierarchical likelihood ratio test (hLRT) at $\alpha = 0.01$. The ML heuristic search was carried out starting with one of the MP trees and TBR branch swapping.

Bayesian phylogenetic analyses were performed on the combined data set in MRBAYES version 3.0 (Huelsenbeck & Ronquist 2001). The most suitable model of nucleotide substitution was selected by hLRT in MRMODELTEST version 2.2 (Nylander 2004), which is a version of MODELTEST modified to compare 24 instead of 56 models of nucleotide substitution, all of which can be implemented in MRBAYES version 3.0. The selected model was the same as that selected by MODELTEST for the likelihood analysis. Three Bayesian Markov chain Monte Carlo (MCMC) runs of four linked chains (temp = 0.08) were run for 1 000 000 generations each. In each run, every 100th generation was sampled resulting in a total of 10 000 trees from each run. Likelihood-by-generation plots were examined to determine how many generations it took to reach stability.

The combined sequence dataset was tested for a molecular clock using the likelihood ratio test by comparing twice the difference of the likelihood score of a tree assuming a clock and the likelihood score of the same tree without a clock to a χ^2 distribution with $n - 2$ degrees of freedom, where n = number of taxa.

Divergence time estimates

Divergence times were estimated on the ML sequence tree (Fig. 2A) using penalized likelihood (PL) in the program R8S version 1.7 (Sanderson 2003). PL averages local differences in the rate of DNA evolution on different branches, taking into account the topology of branching (Sanderson 2002). PL differs from nonparametric rate-smoothing (NPRS; Sanderson 1997) in that it assigns a penalty for rate changes among branches that are too rapid or frequent, based on a smoothness parameter. If the smoothness parameter is large, then PL approaches a clock-like model of molecular evolution; if the smoothness parameter is small, then PL approaches NPRS, which allows for varying rates of DNA substitution across lineages. The optimal smoothing parameter was determined

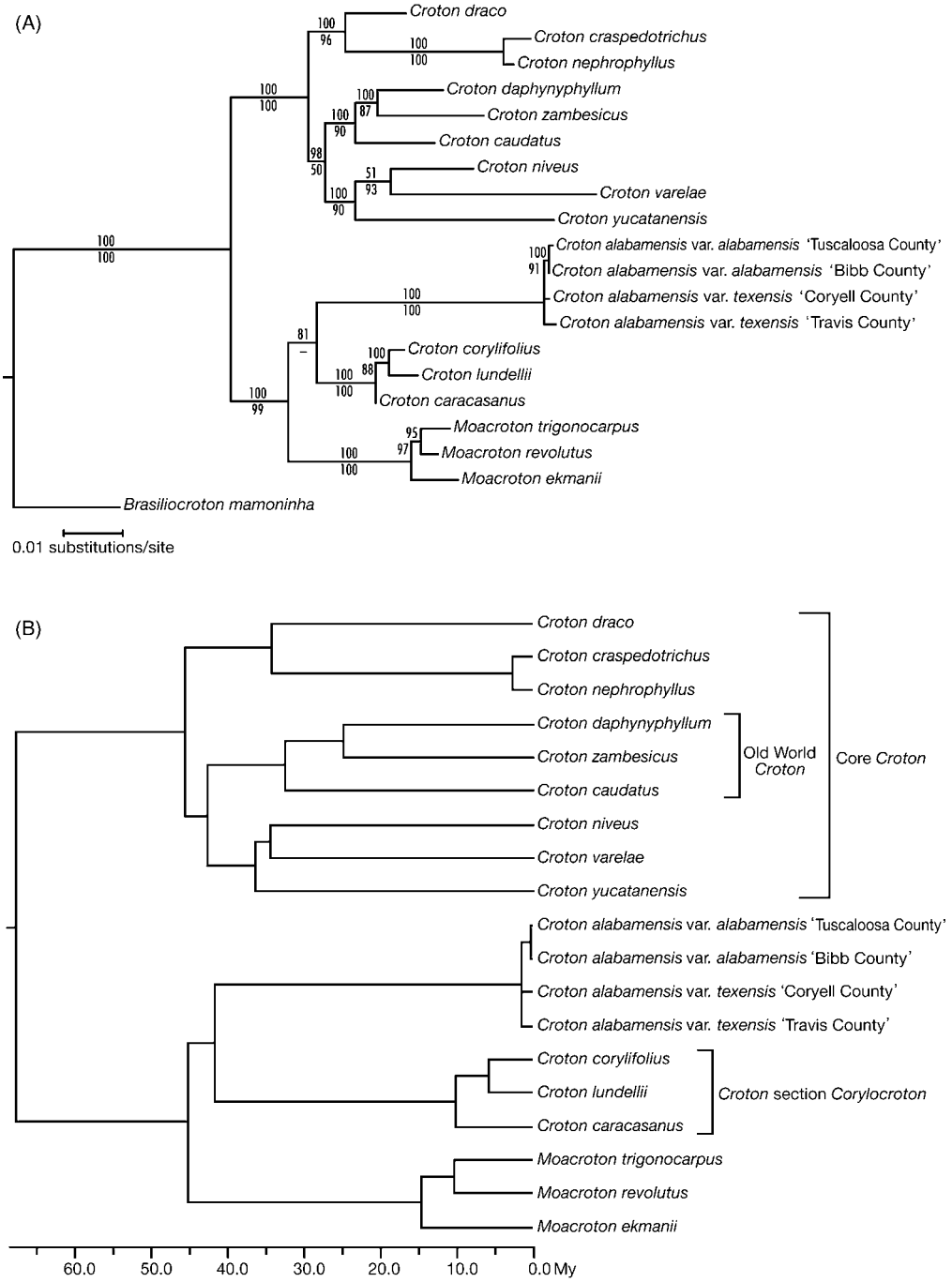

Fig. 2 (A) Maximum-likelihood phylogram of ITS and *trn*L-*trn*F sequence data. Bayesian clade credibility values are shown above the branches and maximum-parsimony bootstrap percentages are shown below the branches. (B) The maximum-likelihood phylogram from Fig. 2A converted to ultrametric form using penalized likelihood.

using cross-validation (Sanderson 2002). This was carried out after pruning off *Brasiliocroton mamoninha* in order to overcome the problem of PAUP* placing the root of the tree on the branch connecting *B. mamoninha* (the furthest-out outgroup) to the rest of the taxa and arbitrarily assigning all of that branch's length on one or the other side of the root. The inclusion of those two branches, one artificially long and the other zero in length, is detrimental to estimating divergence times in R8S (Sanderson 2003). This is why *B. mamoninha* does not appear on the PL-smoothed tree (Fig. 2B).

In the absence of any fossils assignable to any of the clades in this study, or as sister to any of the clades, dates for the major nodes were calculated using the estimated

age of Antillean endemic lineages (Graham 2003b), such as *Moacroton*, as the minimum age of the node where these taxa diverge from their non-Antillean sister groups (i.e. the node separating *Moacroton* from their sister *Croton* taxa). Although proto-Antillean islands may have existed before the Middle Eocene, due to repeated transgression, subsidence, and the mega-tsunamis created by the K/T meteor impact, it is believed that none existed as continuous subaerial islands until after the Middle Eocene (~45 million years ago; Iturralde-Vinent & MacPhee 1999; Graham 2003a). The stem age of *Moacroton* was therefore fixed at 45 million years ago (Ma).

AFLP analyses

ABI chromatographs were analysed in GENESCAN version 3.7.1 for Windows (Applied Biosystems), with a cut-off peak-height value of 50 and using default curve-fitting options for the size standard. Tabular data were then analysed manually with reference to the chromatographs. All bands below 50 bp in length were eliminated along with ambiguous and nondistinct bands, e.g. bands that could not be confidently determined to be either present or absent in all samples. Unambiguous bands were scored for presence (1) or absence (0), and a binary matrix was constructed for all taxa.

Ordinations were conducted using nonmetric multidimensional scaling (NMS) in PC-ORD version 4.0 (McCune & Mefford 1999). Multidimensional scaling has been demonstrated to be effective at recovering both hierarchical patterns and nonhierarchical patterns in multilocus data (Lessa 1990) and is therefore an appropriate method for evaluating species boundaries without assuming a hierarchical evolutionary pattern, which should not be assumed at the outset of such a project. Nonmetric multidimensional scaling also avoids the assumption of linearity among variables and permits the use of a variety of distance measures (McCune & Grace 2002). Jaccard's (1908) index was used in ordination because it has been shown to perform more consistently at recovering systematic relationships from molecular data than other similarity measures when small amounts of data are available (Landry & Lapointe 1996). Analysis was conducted in the 'autopilot', 'slow and thorough' mode of PC-ORD, which uses a maximum of 400 iterations per 40 runs with real data and 50 runs with randomized data to identify the number of axes, followed by 400 iterations using the optimal starting configuration from the first set of iterations. Accession T-BC3 was not included in the ordination because it was missing data from one primer pair.

With the aim of obtaining a better phylogenetic signal within and among the Texas and Alabama populations of *C. alabamensis*, AFLP data for individual accessions were analysed phylogenetically, using closely related outgroups from *Croton* sect. *Corylocroton* and *Moacroton* to root the tree. A minimum evolution (ME) tree was inferred rooted based on a pairwise distance matrix calculated using the restriction site distance of Nei & Li (1979), which calculates genetic distance based on the probability that two individuals inherited a shared restriction site from a common ancestor. Although Nei and Li's distance is based on the relatively simple Jukes–Cantor model of nucleotide substitution (Jukes & Cantor 1969) and is typically not tailored to the large number of nucleotides that comprise the recognition site of a typical AFLP marker (16 bp in this study and most other published AFLP studies), it recovers relationships similar to those of the more complex distance metric implemented in the RESTDIST program of PHYLIP version 3.6 (Felsenstein 1989), and it may outperform the latter metric as evaluated using maximum likelihood (Hipp *et al*. in press). The model underlying this distance metric, and the ME tree-selection criterion, disregard the effects of reticulation, and consequently they would be ill suited for assessing relationships among populations in the presence of ongoing gene flow. However, given the substantial distance between the Alabama and Texas populations, and the evidence of genetic divergence based on ordinations and analyses in STRUCTURE (below), tree building seems an appropriate method of evaluating whether the root of the species lies among the Texas or the Alabama accessions. Trees were recovered heuristically in PAUP* 4.0b10 (Swofford 2002) with a neighbour-joining starting tree and TBR branch swapping. Branch support was assessed using 1000 heuristic-search nonparametric bootstrap pseudoreplicates. Relationships were also inferred using UPGMA, which presupposes an ultrametric topology, probably a reasonable assumption in assessing relationships among populations within species (Felsenstein 2004).

Genetic structure and diversity

To evaluate whether phylogenetic analyses conducted on individuals of *C. alabamensis* provide a robust estimate of gene flow among the geographic areas from which plants were sampled, population structure was analysed using the Bayesian clustering method of Pritchard *et al*. (2000), implemented in STRUCTURE version 2.0 (http://pritch.bsd.uchicago.edu). Accession T-BC3 was not included in this analysis because data was missing for one primer pair. The method uses MCMC to estimate allele frequencies and assign individuals or populations to clusters probabilistically, under the assumption that populations are at Hardy–Weinberg equilibrium and linkage equilibrium. Because we were using a dominant marker, each locus was coded as known for one copy and unknown (–9) for the other, and the 'no admixture' ancestral model assumed as recommended in the program documentation. The parameter for distribution of allele frequencies (λ) was

estimated in five initial runs with $K = 1$, then set at a constant $\lambda = 0.6572$ for all populations in remaining simulations. The number of populations (K) was estimated under the correlated alleles model. Five independent runs were carried out for each value of K (= number of clusters assumed) between 2 and 7, with parameters and model likelihood estimated over 200 000 MCMC generations following a burn-in period of 50 000 generations. Prior geographic information was not employed.

Expected heterozygosity (Nei's gene diversity, H_E), pairwise F_{ST} (variance among individuals within populations), and estimates of within vs. between population gene diversity were estimated using Lynch & Milligan's (1994) method of recovering unbiased population genetic statistics from dominant markers as implemented in AFLP-SURV version 1.0 (Vekemans 2002). Allele frequencies were estimated using the Bayesian method of Zhivotovsky (1999) assuming a nonuniform prior. Confidence intervals on genetic distances were estimated using 1000 bootstrap replicates. Pairwise genetic divergence values were considered to be significantly different at the 0.05 level if the 95% confidence interval for one pairwise genetic distance excluded the pairwise genetic distance inferred for the other. Estimates were conducted on the data partitioned into the four populations identified using STRUCTURE, assuming Hardy–Weinberg equilibrium and, for comparison, at $F_{IS} = 0.1–0.2$. Additional analyses were performed in AFLP-SURV, assuming five and seven populations. Under all scenarios tested, H_W (gene diversity within populations) is more than twice as large as H_B (gene diversity between populations). Because these alternative scenarios are not supported by the results from STRUCTURE, and do not change our conclusions, we report only the four-population analyses. To further ensure that results were not biased towards recovering higher diversity in the Texas populations, analyses were conducted both with and without a Texas geographic subpopulation of three individuals (Pace Bend) that clusters separately in STRUCTURE under the assumption of $K > 4$ but with a larger Texas subpopulation (Balcones Canyonlands) at $K = 4$.

Hierarchical analysis of molecular variance (AMOVA; Excoffier *et al*. 1992) as implemented in ARLEQUIN version 3.01 (Schneider *et al*. 2000) was used to partition molecular variance at three levels: (i) between states (varieties) within the species, (ii) among populations within states (varieties), and (iii) among individuals within populations. The corresponding fixation indices (F_{CT}, F_{SC}, and F_{ST} respectively) calculated in ARLEQUIN following the method of Excoffier *et al*. (1992) are reported in Table 4, despite the fact that ARLEQUIN does not correct these statistics for marker dominance (see previous paragraph for methods of analyses using dominant markers). This fact notwithstanding, the P values, estimated based on 10 000 permutations, are appropriate for evaluating whether the partitioning of variance components was significant. Although within-population variance is actually $1 - F_{ST}$ rather than F_{ST}, the significance is the same. Therefore, following Excoffier *et al*. (1992) we report F_{ST} in conjunction with the variation among individuals within populations. Data were coded as restriction fragment length polymorphism (RFLP) haplotypic data for purposes of analysis.

Results

Sequence analyses

The aligned ITS and *trn*L-*trn*F sequences are 689 and 1029 bp long, of which 242 and 120 are variable and 163 and 40 were informative across the three genera (*Croton*, *Moacroton*, and *Brasiliocroton*), respectively. Within *Croton alabamensis*, three ITS characters (all substitutions) and six *trn*L-*trn*F (three substitutions and three indels) were informative. All accessions of *C. alabamensis* var. *alabamensis* had identical ITS sequences, as did all accessions of *C. alabamensis* var. *texensis*. Three informative ITS characters separate the two varieties into reciprocally monophyletic groups within *C. alabamensis*. The six informative *trn*L-*trn*F characters resolve the 22 accessions of *C. alabamensis* into more geographically structured groups; however they do so inconsistently (homoplasiously), with no two of the six characters shared by the same group of accessions, and only by invoking parallelisms or reversals do any of these characters exclusively group a single population from a given geographic area.

All analyses were first performed on a full data set, including the 22 accessions of *C. alabamensis* for which both ITS and *trn*L-*trn*F sequences are available. This data set was then pared down to include only four *C. alabamensis* accessions, two of var. *alabamensis* and two of var. *texensis*. All of the tests performed on the data set reported here were performed on both the full data set, including multiple identical sequences, and on the reduced data set. The results of the tests for incongruence between ITS and *trn*L-*trn*F, clock-like behaviour, divergence time estimates, likelihood models, and tree topologies were the same for the two data sets. Bootstrap values and Bayesian clade credibility values were comparable. Therefore, the sequence analyses results reported here are those performed on the data set that includes four accessions of *C. alabamensis* (A-BW1, A-GL3, T-FH10, and T-BC4) in addition to the 16 accessions from the core *Croton* clade, *Croton* sect. *Corylocroton*, *Moacroton*, and *Brasiliocroton*.

Separate analyses of the ITS and *trn*L-*trn*F sequence partitions revealed no strongly supported contradictory clades. The incongruence length difference (ILD) test failed to reject ($P = 0.781$) the hypothesis of no meaningful conflict between the partitions, therefore the two gene regions were analysed together.

Parsimony analyses of the combined data set resulted in nine most parsimonious trees of 459 steps, with a consistency index (CI) of 0.678 and a retention index (RI) of 0.831. The hierarchical likelihood ratio test implemented in MODELTEST 3.06 (Posada & Crandall 1998) indicated that GTR + I + Γ was the best fitting likelihood model for the combined data. Maximum likelihood analyses using this model produced a topology very similar to that obtained by parsimony. The only difference was that parsimony places *C. alabamensis* sister to *Moacroton* plus *Croton* sect. *Corylocroton*, albeit with low support (bootstrap = 51%), whereas likelihood places *C. alabamensis* sister to *Croton* sect. *Corylocroton* alone (Fig. 2A). This appears to be a case of inconsistency, also known as long-branch attraction, in the parsimony analysis (Felsenstein 1978).

The Bayesian MCMC runs resulted in three sets of 10 000 trees each. Examination of the likelihood-by-generation plots revealed that stability was reached by approximately 100 000 generations (1000 retained trees). To be conservative, we considered the first 5000 trees as the burn-in. All three runs yielded identical 50% majority rule post burn-in consensus trees. This suggests adequate mixing, therefore we pooled the remaining 15 000 trees to estimate the posterior clade credibility values. The topology of the Bayesian tree is identical to that of the likelihood tree in that it places *C. alabamensis* sister to *Croton* sect. *Corylocroton*.

Divergence time estimates

Using the GTR + I + Γ model the combined ITS and *trn*L-*trn*F data set does not evolve in a clock-like fashion ($P < 0.05$), so a simple molecular clock cannot be used to date nodes within the phylogeny. The optimal smoothing parameter (3.2×10^7) was found in R8S using cross-validation in which the stem age of *Moacroton* was fixed at 45 Ma (Sanderson 2002). The high value of this smoothing parameter for Penalized Likelihood suggests that the data are not entirely un-clock-like (Sanderson 2002). Figure 2B shows the ML tree (Fig. 2A) converted into a chronogram using PL. By fixing the stem age of *Moacroton* at 45 Ma, the stem age of the core *Croton* clade is estimated to be approximately 68 Ma, and the stem age of *C. alabamensis* is estimated at 41 Ma. This is concordant with the Davis *et al*. (2005) fossil-based estimate of 65 Ma as the divergence time of *Croton* from the next closest taxon in their analysis of 124 taxa representing all families of Malpighiales (Angiosperm Phylogeny Group 2003), using *atp*B, *rbc*L, 18S, and *nad1*B-C.

AFLP analyses

The two AFLP primer combinations yielded a total of 352 unambiguously scorable bands for the entire data set (38 accessions of *C. alabamensis*, plus five outgroup accessions consisting of *Croton mexicanus*, *Croton lundelli*, *Moacroton ekmanii*, *Moacroton trigonocarpus*, and *Moacroton revolutus*), 163 of which were generated by primer combination 1, and 189 of which were generated by primer combination 2. No markers were monomorphic across the entire data set (including the outgroup), while 99 markers (28%) were monomorphic within *C. alabamensis*. Within the 38 accessions of *C. alabamensis*, a total of 261 unambiguously scorable bands were generated, 119 from primer combination 1 and 142 from primer combination 2 (Table 2). Approximately twice as many fragments are fixed in Texas as in Alabama (Table 3).

Populations defined using STRUCTURE all correlate with geographic populations identified in the field, but not all populations could be separated using this method. The four-population model has by far the highest posterior probability [Pr ($K = 4$) $\gg 0.999$]. Two populations are recovered in Alabama (corresponding to the Black Warrior

Table 3 Genetic diversity within populations and geographic regions of *Croton alabamensis*. Polymorphism, expected heterozygosity (H_E), and variance components were calculated using the method of Lynch & Milligan (1994), which provides unbiased estimates for dominant markers such as AFLPs. Abbreviations: SE, standard error; Var, variance. Accession T-BC3 was not included in analysis due to missing data

Populations	*n*	Fragments fixed	Fragments unique	% loci polymorphic	H_E	SE (H_E)	Var (H_E)	% Var: individuals	%Var: loci
Alabama	**17**	**14**	**35**						
Alabama 1: Black Warrior River watershed	8	39	9	43.8%	0.09996	0.00761	0.000058	48.9%	51.1%
Alabama 2: Cahaba River watershed	9	23	11	48.3%	0.11855	0.00793	0.000063	44.2%	55.8%
Subpopulation Highway 219 ($n = 2$)		35	0						
Subpopulation Glades ($n = 3$)		37	3						
Subpopulation Pratt's Ferry ($n = 4$)		26	11						
Texas	**21**	**25**	**65**						
Texas 1: Travis County	10	31	20	58.0%	0.15636	0.00852	0.000073	42.2%	57.8%
Subpopulation Balcones Canyonlands ($n = 7$)		33	17						
Subpopulation Pace Bend ($n = 3$)		60	4						
Texas 2: Fort Hood	11	48	14	48.9%	0.11159	0.00799	0.000064	37.5%	62.5%

River watershed as one population, and the Cahaba River watershed — comprising the Highway 219, Glades, and Pratt's Ferry subpopulations — as the other; and two in Texas, corresponding to the Fort Hood accessions in one population, and the Balcones Canyonlands and Pace Bend accessions in the other). All individuals are assigned with 100% confidence to one of the four populations, with the exception of a Pratt's Ferry individual assigned with Pr = 0.999 to the Cahaba River watershed and with Pr = 0.001 to the Balcones Canyonlands/Pace Bend population. Assuming five populations ($K = 5$) causes the Pace Bend (TX) population to separate out. At the same time, all $K = 5$ reconstructions result in the joint assignment of several Balcones Canyonlands individuals to both the Balcones Canyonlands and Pace Bend populations. Assuming $K = 6–7$ results in the assignment of many individuals from the Balcones Canyonlands (TX) and Pratt's Ferry (AL) populations jointly to additional populations, but no new population areas are recognized. In no analyses are the three Alabama sites within the Cahaba River watershed distinguished from one another. There is some difficulty in distinguishing the Alabama populations from one another, as numerous suboptimal reconstructions for all $K > 2$ apportioned all population structure to Texas, recovering Alabama as a single population. These suboptimal reconstructions all differ by ~80–170 in log likelihood from the clusters of reconstructions in which two Alabama populations are recovered.

Over half of the molecular variance (54.77%) is due to variation among individuals within populations, with the remainder of the variance apportioned almost evenly to variance between states (21.08%) and variance between populations within states (24.15%). F_{SC} and F_{ST} are significant ($P < 0.001$) while F_{CT} is not ($P = 0.3334$) (Table 4). This is almost certainly due to the low power of the permutation test in our study. The null distribution of F_{CT} is obtained by assuming that populations are real but that groups are not, leaving the population composition intact but permuting populations with regard to group, and keeping the number of populations per group fixed between permutations (Excoffier *et al*. 1992). There are only three ways of allocating four populations to each of two different groups with two populations each [$4!/(2!2^2) = 3$]. Thus, the most significant P value in this design, that is, the lowest possible probability of finding F_{CT} in the null distribution ≥ the estimated F_{CT}, is one-third.

Pairwise distances between populations (Table 5) demonstrate a significantly higher degree of differentiation between the two Alabama populations than between the two Texas populations. All Texas–Alabama comparisons demonstrate greater differentiation than within-state population comparison, but only one interstate (intervarietal) genetic distance is significantly higher in both Nei's D and F_{ST} than the genetic distance between the two Alabama populations: the comparison between populations at Fort Hood (Texas 2) and Black Warrior River watershed (Alabama 1).

Table 4 Hierarchical AMOVA within *Croton alabamensis*. For purposes of analysis, data were scored as RFLP haplotypes, and as such F-statistics cannot be directly compared with analyses that correct for the fact that AFLPs are dominant markers. Molecular variance is significant at two levels: between populations within states (F_{SC}, variance component b) and among individuals within populations (F_{ST}, variance component c). The lack of significance for F_{CT} (variance component a) is presumably due at least in part to a lack of power for the permutation test with our sampling design (see Discussion)

Source of variation	d.f.	Sum of squares	Variance components	% variance	Fixation indices	P value (10 000 permutations)
Between states (varieties) within the species	**1**	204.655	Va = 6.359	21.08%	$F_{CT} = 0.211$	0.3334
Between populations within states (varieties)	**2**	171.089	Vb = 7.286	24.15%	$F_{SC} = 0.306$	< 0.001
Among individuals within populations	**34**	561.835	Vc = 16.525	54.77%	$F_{ST} = 0.453$	< 0.001
Total	**37**	937.579	30.169			

Table 5 Pairwise distance matrix between populations: pairwise F_{ST} above the diagonal, Nei's D below the diagonal, 95% confidence intervals in parentheses based on 1000 bootstrap pseudoreplicates. Distances are calculated using the method of Lynch & Milligan (1994), which is designed to give unbiased estimates for dominant markers such as AFLPs. Accession T-BC3 was not included in analysis due to missing data

	Alabama 1	Alabama 2	Texas 1	Texas 2
Alabama 1	—	0.2823 (0.1977–0.3581)	0.2835 (0.2183–0.3405)	0.3738 (0.2984–0.4409)
Alabama 2	0.0493 (0.0302–0.0715)	—	0.2842 (0.2177–0.3417)	0.3482 (0.2773–0.4130)
Texas 1	0.0594 (0.0403–0.0793)	0.0651 (0.0443–0.0874)	—	0.1187 (0.0799–0.1640)
Texas 2	0.0731 (0.0508–0.0974)	0.0719 (0.0505–0.0970)	0.0207 (0.0127–0.0311)	—

Effects on estimates of H_J (expected heterozygosity) and allelic diversity of assuming F_{IS} (within population inbreeding) > 0 were negligible. Consequently, only results assuming Hardy–Weinberg equilibrium are presented (Tables 3 and 5). Effects of excluding the Pace Bend (Texas) population from analyses were also negligible and were disregarded in the remainder of this study except as explicitly indicated. Under the assumption of four populations (as recovered in STRUCTURE), 59.7% of the fragments segregate. Average expected heterozygosity (H_E) in the Alabama populations is 0.1093, somewhat lower than average H_E in the Texas populations (0.1340 if all individuals are analysed, or 0.1337 if the Pace Bend population is excluded from analysis). Total gene diversity (H_T) is apportioned primarily to diversity within populations ($H_W = 0.1216$), with little among-population diversity ($H_B = 0.0485$); F_{ST} for the four populations is 0.2832 ($P < 0.001$, based on 1000 permutations of individuals among populations). A very similar result is found when states are treated as the units of analysis ($H_W = 0.1338$, $H_B = 0.0386$ at $F_{IS} = 0$; $H_W = 0.1260$, $H_B = 0.0421$ at $F_{IS} = 0.25$), although expected divergence from Hardy–Weinberg equilibrium, when treating each state as a population, makes interpretation of this result less straightforward.

Phylogenetic and ordination analyses

Phylogenetic analysis of the AFLP data (using ME and UPGMA) supports the monophyly of *C. alabamensis*, the distinctness of the two varieties, and two major populations within each state (Fig. 3A,B). Genetic structure implied by this analysis closely mirrors the populations recovered using STRUCTURE, with the exception that the Pace Bend (TX) accessions cluster with the Fort Hood (TX) population in the ME and UPGMA trees, while Pace Bend (TX) and Balcones Canyonlands (TX) cluster together in the STRUCTURE analyses. The root of the species based on the ME analysis, which does not presuppose ultrametricity, falls among the Texas accessions, but with very low support (< 50%) (Fig. 3A). This same rooting is recovered with high support (100% MP bootstrap and 100% posterior probability) in the sequence analysis. UPGMA analysis (Fig. 3B) roots the species at the midpoint of the branch between the Texas and Alabama populations. Although this rooting is more highly supported (bootstraps of 90% and 94% for the Texas and Alabama clades respectively), assuming that the data are clock-like may not be valid, in which case ME has a higher probability of recovering the correct tree (Felsenstein 2004).

The results of NMS ordination of the *C. alabamensis* taxa in two dimensions closely mirror the AFLP trees (ME and UPGMA), which is not surprising given that the ordination is based on a very similar pairwise distance measure (Jaccard's distance and Nei and Li's restriction site distance). Figure 4 shows the distinct separation between the Alabama individuals and the Texas individuals. The Alabama individuals were further grouped within watershed boundaries, namely the Black Warrior River watershed and the Cahaba River watershed (the later containing the Highway 219 (A-HW), Glades (A-GL), and Pratt's Ferry (A-PF) subpopulations). The Texas individuals were more disparate, with the Fort Hood population clustering closely together, close to individuals from the Pace Bend (T-PB) subpopulation. The Balcones Canyonlands (T-BC) grouping shows quite a few outliers, which were quite distinct from the rest of the Texas individuals.

Discussion

Croton alabamensis is a highly diverged lineage within *Croton*, and any attempt to reconstruct phylogenetic relationships must overcome the problem presented by the long branch leading up to it. What is remarkable, however, is that it has no closely related species sampled to date, and the closest relatives are either taxa endemic to the island of Cuba (*Moacroton*) or else members of section *Corylocroton* that occur in Mexico, Central America, the Caribbean, and South America. The placement of *C. alabamensis* by Webster (1993) in section *Lamprocroton* was clearly erroneous, and if sectional distinctions are to be maintained in the future, it may need to be placed in its own distinct section of the genus.

Within *C. alabamensis*, both DNA sequence and AFLP data are unclear in placing the outgroup root inside either of the two varieties, making it difficult to determine if either of them is derived from the other. The analyses of the combined sequence data and the analyses of the AFLP data consistently resolve the two varieties of *C. alabamensis* as genetically distinct, with no ambiguity in the placement of accessions. The ITS portion of the sequence data includes three characters which consistently separate the two varieties. The *trn*L-*trn*F portion of the sequence data was more variable and more geographically structured, being able to resolve the accessions from Fort Hood, Texas and the Black Warrior River watershed in Alabama. This is consistent with what is expected, given that chloroplast DNA is transmitted from a single parent. The *trn*L-*trn*F region was more homoplasious compared to the ITS region, and when analysed alone accessions of *C. alabamensis* were not recovered as a monophyletic group. The AFLP data provided the best resolution among the terminal clades, resolving not only both regions but the groups within each region as well. The unambiguous genetic differentiation between these two areas, along with the negligible probability of natural migration or genetic exchange between the Texas and Alabama populations, which are over 1000 km from one another, supports Ginzbarg's (1992) recognition of the Texas and Alabama populations as separate varieties

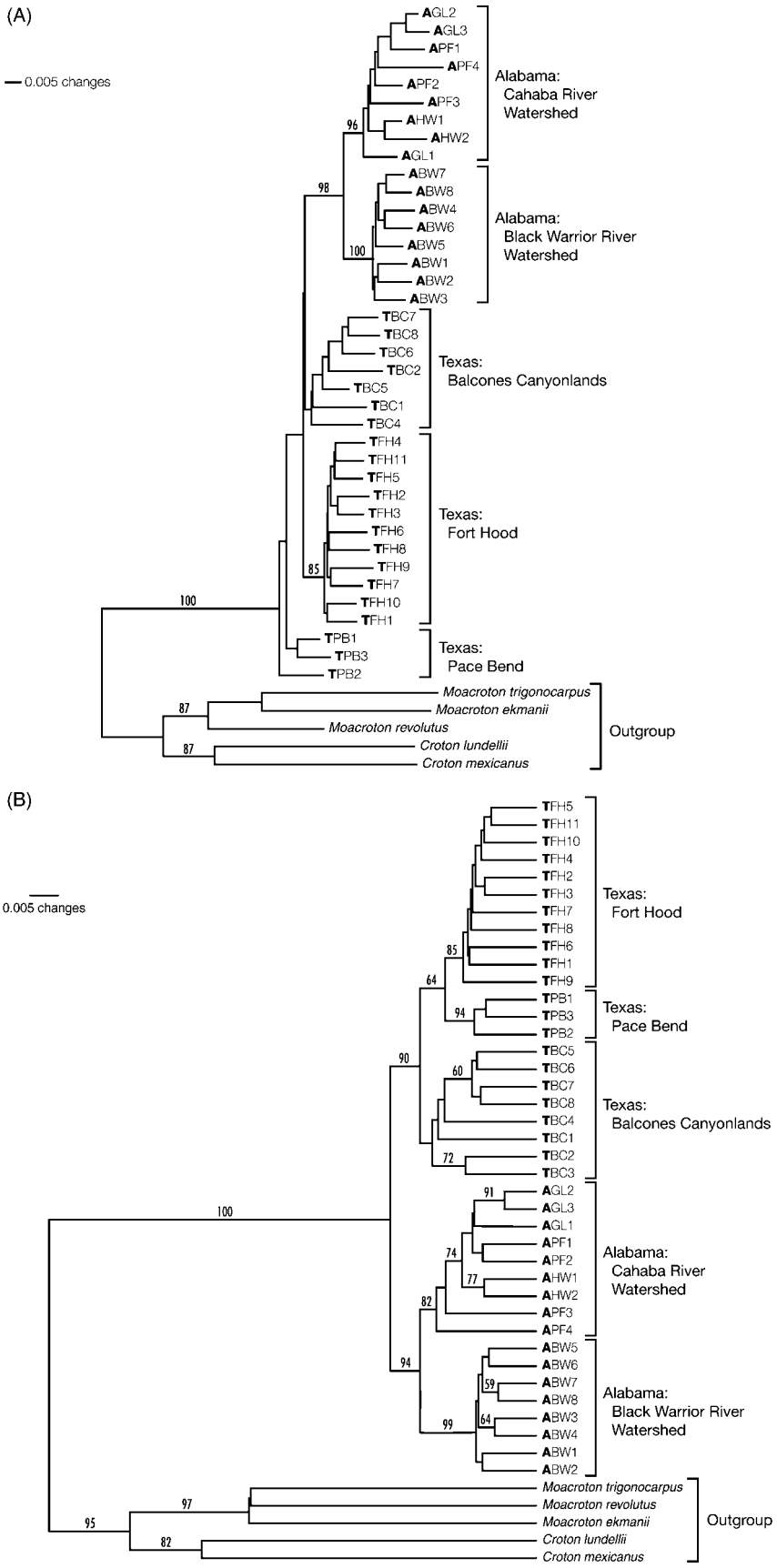

Fig. 3 (A) AFLP minimum evolution (ME) tree, based on a pairwise distance matrix calculated using Nei & Li's (1979) restriction site distance. Accession T-BC3 was not included in this analysis due to missing data. (B) AFLP UPGMA tree, based on a pairwise distance matrix calculated using Nei & Li's (1979) restriction site distance.

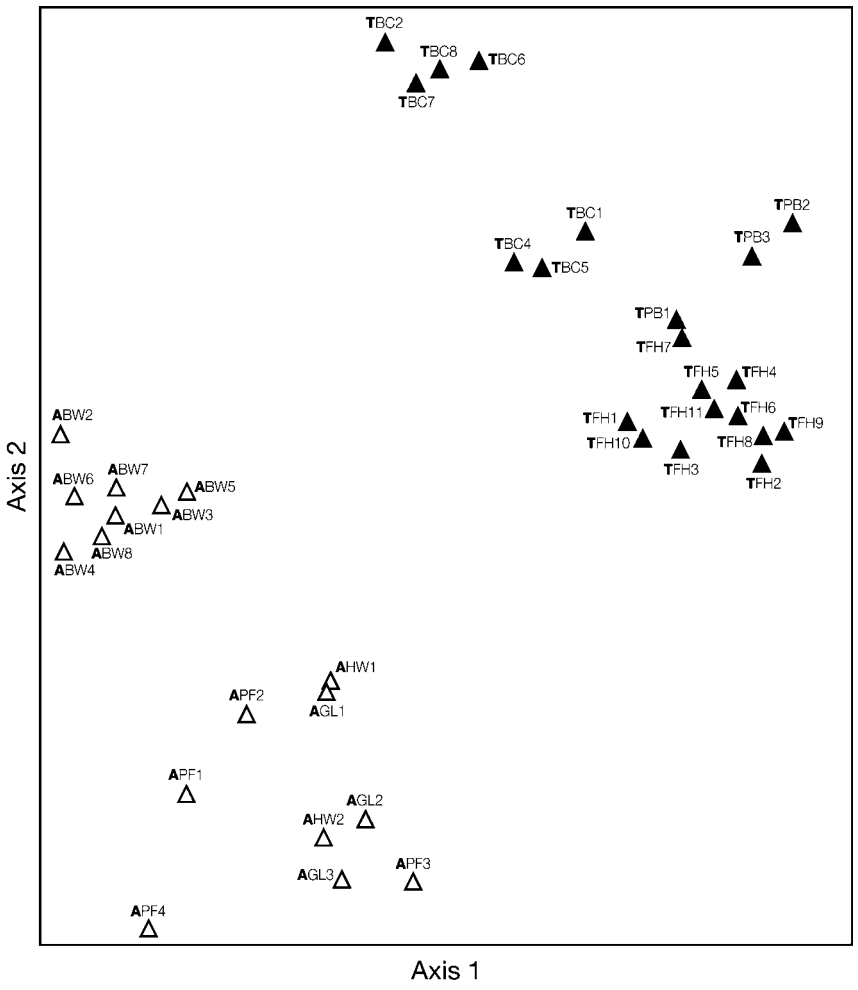

Fig. 4 Two-dimensional NMS ordination of AFLP data. Location abbreviations: A, Alabama; **T**, Texas; BC, Balcones Canyonlands, BW, Black Warrior River; GL, Glades; HW, Highway 219; PB, Pace Bend; PF, Pratt's Ferry. Accession T-BC3 was not included due to missing data.

(although one could also argue for their recognition at the level of subspecies, given their geographic differentiation).

Within Alabama, the AFLP data clearly resolved the Black Warrior River watershed accessions as one unit and the Cahaba River watershed accessions as another, somewhat more disparate unit. This is consistent with topographical boundaries between the watersheds that could inhibit gene flow between these two areas.

Croton alabamensis is monoecious, with separate staminate and pistillate flowers on the same plant. Pistillate flowers generally open before the staminate flowers, but there is usually overlap in flowering of the two sexes before all flowers on an inflorescence wither (Farmer 1962). In both field and garden studies, Farmer (1962) found that pistillate flowers failed to develop into fruits in the absence of pollen, but he indicated that *C. alabamensis* is self-fertile, given his observation of heavy seed crops from single plants isolated from other known individuals of *C. alabamensis*. This suggests that *C. alabamensis* has a mixed mating system involving both out crossing and geitonogamous selfing. Farmer (1962) suggested that the species is primarily wind-pollinated, but we made numerous observations of bees and other insects visiting male and female flowers in Texas and Alabama, as well as the production of nectar at the base of the petals. We conclude that the species is modally insect-pollinated, although it may also have a mixed pollination system involving some degree of pollen transport by wind.

Biogeographical history

The molecular clock analysis reveals that the divergence of the two varieties of *C. alabamensis* was relatively recent, although we can only estimate this as having occurred some time during the Quaternary. Long-distance dispersal was suggested by Ginzbarg (1992) as a means for the establishment of the Texas populations of *C. alabamensis*, but this hypothesis was most probably influenced by the timing of the discovery of the Texas populations, nearly 90 years after the initial discovery and description of the species in Alabama. Most species of *Croton* have seeds that are dispersed by the explosive dehiscence of their capsules,

but *Croton* seeds are also a favourite food of doves (Johnson 1956), and 'doveweed' is the common name for several species of *Croton* in the United States. Ginzbarg (1992) suggested that passenger pigeons (*Ectopistes migratorius*), relatives of doves that until recently were one of the most abundant birds on earth, could have been the agent of long-distance dispersal of *C. alabamensis* seeds from Alabama to Texas during their former fall migration southwards. Passenger pigeons were reported from the counties in Texas where *C. alabamensis* now occurs, which presumably were on the western end of the birds' winter range (Bent 1932; Oberholser 1974). However, fruit maturation and dispersal in *C. alabamensis* typically takes place in May and June (Farmer 1962; personal observation), so it may actually have been more plausible for dispersal to have occurred in the opposite direction during the birds' spring migration to the north and east. Passenger pigeons were actually very variable in their foraging and migration patterns, but their sheer numbers — populations estimated at between 3 and 5 billion individuals in eastern North America several centuries ago (Schorger 1955) — and their extreme mobility both make this argument plausible.

The genetic diversity of *C. alabamensis* within both Texas and Alabama is not consistent with the expectations of a relatively recent long-distance dispersal, namely that one of the two populations would be genetically much more depauperate than the other. As evidenced by the ordination of the AFLP data, there is no gradient from either region to the other. Genetically, they appear to be two well-separated groups, each mutually distinct from the other. Lower divergence between the Texas populations (pairwise $F_{ST} = 0.1187$, $D = 0.0207$) than the Alabama populations ($F_{ST} = 0.2823$, $D = 0.0493$) is probably not an effect of the distance between populations, as the Texas populations are nearly twice as far from one another as the Alabama populations are. The greater divergence within Alabama may reflect a longer period of divergence between the Alabama populations. However, the Texas populations may have occupied a larger and more contiguous range that was only recently reduced to more separate populations, due to human disturbance. Also, access to private lands in Texas is notoriously restricted, and there may be undiscovered geographically intermediate populations. Weak polarization of both the sequence and AFLP data suggest that the Texas population was the progenitor of the Alabama population; however, the data are not strong enough, nor entirely appropriate, to support this phylogenetic conclusion. Additionally, allelic diversity within *C. alabamensis* is comparable within the two geographic regions, suggesting that recent long-distance dispersal is not a likely explanation for the origin of either variety.

A different biogeographical scenario was put forth by Watson *et al*. (2002) for similarly disjunct populations of *Eriocaulon koernickianum* (Eriocaulaceae) on either side of the Mississippi. In this species, there was virtually no genetic variability among disjunct populations, as measured by isozyme markers, and the authors hypothesized that the modern populations had experienced a genetic bottleneck and had recolonized their current habitats after severe fragmentation of their former range. This is clearly not the case in *C. alabamensis*, which shows much more structured genetic variation within local populations, between geographically separated populations within each state, and finally between Alabama and Texas populations.

The biogeographical hypothesis most consistent with our findings at this point is that the current disjunction between Texas and Alabama is the result of past vicariance events, with the separation of a formerly more continuous distribution and the extinction of geographically intermediate populations over time. However, the areas where *C. alabamensis* now occurs were separated first by Cretaceous seas and then by the Mississippi Embayment as sedimentation steadily filled in the valley (Guccione & Zachary 2004), and none of the current piedmont and outcrop habitats ever bridged the gap. During the Pleistocene Epoch, large-scale shifts in the vegetation of the southern United States occurred as a result of the glaciation cycles farther to the north, and species such as *C. alabamensis* would presumably have needed to migrate to warmer areas farther south. However, *C. alabamensis* is already at the southern and western most extremes of the fall line, below which lies the Gulf coastal plain, an area which was periodically inundated during glacial times. Assuming that *C. alabamensis* was unable to colonize the flatter, less rocky, and possibly less wooded habitats on the Gulf coastal plain, it may have been forced as far southwards as its habitat requirements would allow, and then for reasons unknown, it has not been able to expand farther north after the retreat of the glaciers. Another shrubby endemic species with a similarly restricted habitat in the southern Appalachians is *Neviusia alabamensis* (Rosaceae). It was until recently considered the only species in the genus, but in 1992, a second species was discovered in northern California (Shevock *et al*. 1992).

Our confidence in the actual dates of nodes on Fig. 2B is dependent on the calibration point used to obtain them, which is inferred from geological events in the Caribbean. Given this, it would be inappropriate to place too much emphasis on the actual dates, although they are consistent with more rigorously derived estimates of Davis *et al*. (2005). Nonetheless, both the topology and the age estimates of the chronogram show a clear pattern. The crown ages of the core *Croton* clade and the clade containing *C. alabamensis*, *Croton* sect. *Corylocroton* and *Moacroton* are shown to be of approximately the same age. This is a conservative estimate, and possibly an artefact of the sampling, given that on the ML tree which was smoothed

to produce the chronogram the branch leading from the most recent common ancestor (MRCA) of these two clades (essentially the root node of the chronogram) to the *C. alabamensis*/*Moacroton*/sect. *Corylocroton* clade is roughly two-thirds as long as the branch from the MRCA to the core *Croton* clade. In an expanded study with much broader sampling in *Croton* than here, and using 65 Ma as the fixed stem age of *Croton s.l.* (i.e. including *Moacroton*), van Ee and Berry (in prep.) estimate the age of the *C. alabamensis*/*Moacroton*/sect. *Corylocroton* clade to be 20% older than that of the core *Croton* clade. Based on this, it seems reasonable that the crown age of the *C. alabamensis*/*Moacroton*/sect. *Corylocroton* clade is older than the crown age of the core *Croton* clade. Thus, *Croton alabamensis* constitutes one of the oldest lineages within the genus, having diverged from its nearest relatives as long ago, or longer ago, than the three major clades of the core *Croton* clade, which include over 1000 species, as well as the Old World/New World split in the genus. In contrast, the divergence between the two modern varieties of *C. alabamensis* is a much more recent (Quaternary) event.

Conservation implications

Croton alabamensis should be considered a high-priority taxon for conservation because of its phylogenetic uniqueness (Hunter 2004), based on its isolated position on a long branch within the species-poor sister group to the rest of *Croton* (Berry *et al*. 2005), and the absence of any closely related species. Our results show that both the Alabama and the Texas populations of *C. alabamensis* share similar levels of genetic diversity and neither variety shows any overt signs of genetic bottlenecks or inbreeding depression, as evidenced by the fact that most of the molecular variance is found within populations (54.77%). A comparable amount of molecular variance is accounted for by divergence between populations within each variety and by divergence between the two varieties (24.15% and 21.08%, respectively). This supports the need to protect both the two varieties as well as all remaining populations of each variety, which is feasible with a rare species such as *C. alabamensis*, given the small number of discrete populations. *C. alabamensis* is locally abundant and reasonably protected in the places where it is known to occur. In Texas, most populations of *C. alabamensis* are known from government land where they are actively being protected, such as on the Fort Hood military reservation in Coryell and Bell counties, and in the Balcones Canyonlands National Wildlife Refuge and Pace Bend County Park in Travis County. The population at Pace Bend Park appears to be at greatest risk due to high human impact (it is on prime lake frontage and surrounded by suburban developments). In Alabama, the populations in Bibb County occur in lands administered by the Nature Conservancy, and in Tuscaloosa County, most of the plants are found bordering the Black Warrior River on steep terrain surrounding Holt Reservoir. The two populations with the lowest amount of genetic diversity are those at Fort Hood (Texas) and the Black Warrior River watershed (Alabama), indicating these two areas in particular may need special protection.

Acknowledgements

We wish to thank Laura Sánchez and Charles Pekins of Fort Hood Natural Resources, Charles Sexton and Bill Reiner of Balcones Canyonlands National Wildlife Refuge, Paul Fushille of Travis County Transportation and Natural Resources, Steve Ginzbarg of the University of Alabama Herbaria, Jean Nance, and Richard and Nancy Cobb for their assistance with collecting plants in Texas and Alabama. Kandis Elliot of the Multimedia Facilities of the University of Wisconsin Botany Department prepared the graphics for this paper. This research is based upon work supported by the National Science Foundation under Grant no. DEB-0212481 and a Davis Fund research grant from the University of Wisconsin-Madison Department of Botany.

References

Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.

Aplet GH, Laven RD, Falkner MB, Shaw RB (1994) Population and site characteristics of a recently discovered disjunct population of *Croton alabamensis* (Euphorbiaceae). *Sida*, **16**, 37–55.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **82**, 247–277.

Beardsley PM, Yen A, Olmstead RG (2003) AFLP phylogeny of *Mimulus* section *Erythranthe and* the evolution of hummingbird pollination. *Evolution*, **57**, 1397–1410.

Bent AC (1932) Life histories of North American gallinaceous birds; orders Galliformes and Columbiformes. Smithsonian Institution. *United States National Museum*. Bulletin, 162.

Berres M (2001) *General fluorescent AFLP (fAFLP) protocols*. University of Wisconsin Department of Zoology, Madison, Wisconsin. Available at http://ravel.zoology.wisc.edu/sgaap/AFLP_html/fAFLP_protocols.htm.

Berry PE, Hipp AL, Wurdack KJ, Van Ee B, Riina R (2005) Molecular phylogenetics of the giant genus *Croton* and tribe Crotoneae (Euphorbiaceae sensu stricto) using ITS and *trn*L-*trn*F DNA sequence data. *American Journal of Botany*, **92**, 1520–1534.

Chapman AW (1883) *Flora of the Southern United States*, 2nd edn. Ivison, Blakeman, Taylor and Co., New York.

Chase MW, Hillis HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon*, **40**, 215–220.

Davis CC, Webb OC, Wurdack KJ, Jaramillo CA, Donoghue MJ (2005) Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. *American Naturalist*, **165**, E36–E65.

Després L, Gielly L, Redoutet W, Taberlet P (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution*, **27**, 185–196.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Farmer JA (1962) *An ecological life history of* Croton alabamensis *EA Smith ex Chapman*. PhD Thesis, University of Alabama.

Farmer JA, Thomas JL (1969) Disjunction and endemism in *Croton alabamensis*. *Rhodora*, **71**, 94–103.

Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance on incongruence. *Cladistics*, **10**, 315–319.

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, **28**, 49–62.

Felsenstein J (1989) PHYLIP (version 3.2): phylogeny inference package. *Cladistics*, **5**, 164–166.

Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

Ginzbarg S (1992) A new disjunct variety of *Croton alabamensis* (Euphorbiaceae) from Texas. *Sida*, **15**, 41–52.

Graham A (2003a) Historical phytogeography of the Greater Antilles. *Brittonia*, **55**, 357–383.

Graham A (2003b) Geohistory models and Cenozoic paleoenvironments of the Caribbean region. *Systematic Botany*, **28**, 378–386.

Guccione MJ, Zachary DL (2004) *Geologic History of the Southeastern United States and Its Effects on Soils of the Region*. Southern Cooperative Series Bulletin # 395, Southern Association of Agricultural Experiment Station Directors, Raleigh North Carolina.

Hipp AL, Rothrock PE, Reznicek AA, Berry PE (in press) Changes in chromosome number associated with speciation in sedges: a phylogenetic study in *Carex* section *Ovales* (Cyperaceae) using AFLP data. In: *Monocots: Comparative Biology and Evolution* (eds Columbus JT, Friar EA, Hamilton CW, Porter JM, Prince LM, Simpson MG). Rancho Santa Ana Botanic Garden, Claremont, California.

Hodkinson TR, Renvoize SA, Chonghaile GN, Stapleton CMA, Chase MW (2000) A comparison of ITS nuclear rDNA sequence data and AFLP markers for phylogenetic studies in *Phyllostachys* (Bambusoideae, Poaceae). *Journal of Plant Research*, **113**, 259–269.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: bayesian inference of phylogenetic trees. *Biometrics*, **17**, 754–755.

Hunter ML (2004) *Fundamentals of Conservation Biology*, 2nd edn. Blackwell Scientific, Medford, Massachusetts.

Iturralde-Vinent MA, MacPhee RDE (1999) Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History*, **238**, 1–95.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudense de Sciences Naturelles*, **44**, 223–270.

Johnson JR (1956) *The Last Passenger*. Macmillan, New York.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro MN), pp. 21–132. Academic Press, New York.

Koopman WJM, Zevenbergen MJ, van den Berg RG (2001) Species relationships in *Lactuca* s1 (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany*, **88**, 1881–1887.

Landry PA, Lapointe FJ (1996) RAPD problems in phylogenetics. *Zoologica Scripta*, **25**, 283–290.

Lessa E (1990) Multidimensional analysis of geographic genetic structure. *Systematic Zoology*, **39**, 242–252.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Maddison DR, Maddison WP (2000) *MACCLADE 4: analysis of phylogeny and character evolution*. Sinauer Associates, Sunderland, Massachusetts.

McCune B, Grace JB (2002) *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, Oregon.

McCune B, Mefford MJ (1999) *PC-ORD, version 4.0: Multivariate Analysis of Ecological Data*. MjM Software Design, Gleneden Beach, Oregon.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.

Nylander JAA (2004) *MRMODELTEST version 2*. Program distributed by the author. Department of Systematic Zoology, Uppsala University, Uppsala, Sweden.

Oberholser HC (1974) *The Bird Life of Texas*. University of Texas Press, Austin, Texas.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, **14**, 1218–1231.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, **19**, 101–109.

Sanderson MJ (2003) R8S: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics Applications Note*, **19**, 301–302.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: A software for population genetics data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Schorger AW (1955) *The Passenger Pigeon*. University of Wisconsin Press, Madison, Wisconsin.

Shevock JR, Ertter B, Taylor DW (1992) *Neviusia cliftonii* (Rosaceae: Kerrieae), an intriguing new relict species from California. *Novon*, **2**, 285–289.

Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]. *Taxon*, **54**, 43–61.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) version 4*. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Urbatsch LE, Baldwin BG, Donoghue MJ (2000) Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. *Systematic Botany*, **25**, 539–565.

Vekemans X (2002) *AFLP-SURV, version 1.0*. Distributed by the Author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Watson LE, Kornkven AB, Miller CR, Allison JR, McCarty NB, Unwin MM (2002) Morphometric and genetic variation in

Eriocaulon koernickianum Van Heurck & Muller-Argoviensis (Eriocaulaceae): a disjunct species of the southeastern United States. *Castanea*, **67**, 416–426.

Webster GL (1993) A provisional synopsis of the sections of the genus *Croton* (Euphorbiaceae). *Taxon*, **42**, 793–823.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols. A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315–324. Academic Press, San Diego.

Xu F, Sun M (2001) Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; Amaranthaceae) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. *Molecular Phylogenetics and Evolution*, **21**, 372–387.

Zhang L-B, Comes HP, Kadereit JW (2001) Phylogeny and quaternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. *American Journal of Botany*, **88**, 2331–2345.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

This project is part of a larger, ongoing project to reconstruct the phylogeny and biogeography of *Croton* — one of the largest genera of flowering plants. This research is a portion of Benjamin van Ee's PhD thesis work in the laboratory of Paul Berry. Nicolas Jelinski worked on the AFLP portion of the study as an undergraduate REU fellow in the Botany Department at UW-Madison. He is currently a Masters student at UW-Madison's Nelson Institute for Environmental Studies. Andrew Hipp is the herbarium curator and plant systematist at The Morton Arboretum; his current research focuses on systematics and evolution in oaks and sedges. Paul Berry is currently professor in the Ecology and Evolutionary Biology Department and Herbarium Director at the University of Michigan; he continues to work on taxonomy, systematics, and evolutionary questions in the Euphorbiaceae.
