Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, Ceroxylon echinulatum (Areceae)

PHILIPP TRÉNEL,* MICHAEL M. HANSEN,† SIGNE NORMAND* and FINN BORCHSENIIUS*
*Department of Biological Sciences, University of Aarhus, Ny Munkegade, Building 1540, DK-8000 Aarhus C, Denmark, †Technical University of Denmark, National Institute of Aquatic Resources, Vejlsøvej 39, DK-8600 Silkeborg, Denmark

Abstract
Knowledge of the role of landscapes in shaping genetic connectivity and divergence is essential for understanding patterns of biogeography and diversity. This is particularly relevant for the Andes region, a major biodiversity hotspot of relatively recent origin. We examined the phylogeography and landscape genetics of the Andean wax palm Ceroxylon echinulatum (Areceae) that occurs in two narrow bands of montane forests on each side of the Andes in Ecuador and northeastern Peru. First, we tested the hypothesis of C. echinulatum being a geographic cline species crossing the Andes in the Amotape–Huancabamba zone (AHZ) of southern Ecuador/northern Peru, as indicated by observations on fruit morphology. Second, we assessed the timeframe of cross-Andean divergence, and third, we investigated the impact of contemporary and historical landscape features on observed spatio-genetic patterns. Individual-based Bayesian clustering (BC) identified a northeastern, southeastern, southwestern, and northwestern cluster, with areas of genetic discontinuity coinciding with the Andes and the Giron–Paute deflection. F-statistics derived from BC suggested an east-to-west dispersal history. Population-based analyses revealed strong genetic structuring at both small and large geographic scales. Interpopulation relationships and Mantel tests strongly supported the cline model with cross-Andean dispersal in the AHZ. Along the cline, gene flow measured as $F_{ST}$ was mainly limited by distance, with less but significant impact of climatic friction. Coalescent analysis revealed that cross-Andean divergence took place during the Quaternary. Significant historical isolation ($R_{ST} > F_{ST}$) was found in the southwestern population. The current study illustrates a joint effect of founder dynamics, divergence by distance and historical isolation on patterns of Andean diversity and distribution.

Keywords: Amotape–Huancabamba zone, Andean phylogeography, clinal isolation, landscape genetics, Palmae, vicariance

Received 23 February 2008; revision accepted 30 May 2008

Introduction
Describing how landscapes and historical changes in landscape properties control dispersal, gene flow, and population divergence is essential for understanding patterns of present-day diversity and distribution (Storfer et al. 2007). Landscapes affect a species’ realized distribution via dispersal constraints in the form of the geographic distribution and geometric properties of suitable areas and the degree of friction to dispersal by the areas intermittent to suitable patches (Graves 1988; McRae 2006). For example, dispersal across a topographically elevated area may be more restricted than dispersal across a flat area of the same size due to the joint impact of spatial and ecological constraints on gene flow along elevational gradients. In addition to the spatio-ecological properties of landscapes, the temporal aspect of landscape change is a second important factor in interpreting extant patterns of genetic variation (Poissant et al. 2005). Landscape rearrangements often produce rapid changes of niche availability, resulting in changes in overall distribution.

Correspondence: Finn Borchsenius, Fax: +45 89 42 27 22; E-mail: finn.borchsenius@biology.au.dk

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd
and metapopulation structure. A major challenge is to disentangle the effects of short-term historical processes operating under the current landscape setting from more ancient historical events associated to past landscapes (Balloux & Lugon-Moulin 2002; Poissant et al. 2005).

The high degree of species richness and endemism of the tropical Andes (Myers et al. 2000) has frequently been attributed to the spatio-ecological properties of the Andean landscape and/or its dynamic history of landscape rearrangements (Gentry 1982; Graves 1988; Moritz et al. 2000). The Andes is characterized by steep environmental gradients (Ghalambor et al. 2006) and often narrow, sinuous distributions of habitat types along altitudinal isoclines, often found in disjunct bands from opposite sides of the Andean versants (Graves 1988). Major landscape rearrangements occurred during phases of drastic surface uplift from Miocene times onward and during climatic oscillations of the Pleistocene (Simpson 1975; Gregory-Wodzicki 2000). Cross-Andean disjunctions have been of particular interest in the study of Andean biogeography and diversity (Moritz et al. 2000). Two main hypotheses have been put forward to explain phylogeographic breaks across the Andes. The Andean uplift vicariance hypothesis holds that disjunct distributions are the result of vicariance caused by the uplift of the Andean cordilleras (Chapman 1926). The across-Andes dispersal hypothesis states that such allopatric patterns were derived after the uplift of the Andes via dispersal across the Andes barrier in specified areas of lowered topography (e.g. Chapman 1926; Haffer 1967). Haffer (1967) noted that cross-Andean dispersal may have been facilitated by favourable climatic conditions during the Pleistocene, when dispersal corridors may have existed or barrier strength may have been diminished.

The south Ecuadorian/north Peruvian Andes represents an area of lowered altitude. South of the Giron–Paute deflection (Fig. 1), the Andes breaks down into a number of smaller mountain ridges, with the 3000-m altitudinal contour line getting discontinuous (Fig. 1).
This interrupted Andean zone has been referred to as the Amotape–Huncabamba zone (AHZ) and has previously been suggested to have facilitated cross-Andean dispersal in lowland and mid-elevation plant and animal species (Weigend 2002).

In the present study, we examined the phylogeographic history and landscape genetics of the Andean-endemic wax palm Ceroxylon echinulatum Galeano (Arecaceae) from Ecuador. The objectives were: (i) to test the hypothesis of C. echinulatum being a geographic cline species crossing the Andes in the AHZ of southern Ecuador/northern Peru, as indicated by observations on fruit morphology (U-cline hypothesis); (ii) to assess the timeframe of cross-Andean divergence, thereby providing a test of the Andean uplift vicariance hypothesis and across-Andes dispersal hypothesis; and (iii) to investigate the impact of contemporary and historical landscape features of the Andes on the spatio-genetic structure in C. echinulatum. The results provide compelling evidence for clinal differentiation and Quaternary cross-Andean dispersal in the AHZ, with strong effects on the genetic variation in C. echinulatum of the current landscape setting, colonization history and past population isolation during Pleistocene climatic oscillations.

Materials and methods

Study organism

Ceroxylon echinulatum is a dioecious canopy palm tree up to 35 m tall. With their whitish-waxy stems marked with black leaf scars, C. echinulatum forms a conspicuous element of the lower montane forests (1000–2000 m above sea level) of the eastern and western Andean versants of Ecuador (Henderson et al. 1995), and eastern versants of northeastern Peru (Jean-Christophe Pinteaud, IRD, Montpellier, France, personal communication.). The palm is locally known for its leaves that are used as ceremonial baskets during Palm Sunday parades (Borchsenius et al. 1998). Recent analyses in amplified fragment length polymorphism and nuclear ITS sequence data conducted at the genus level (Philipp Trénel, unpublished data) found strong evidence for an extended concept of C. echinulatum. In these analyses, C. alpinum ssp. ecuadorense Galeano, one of two subspecies of another, but morphologically similar species of Ceroxylon (C. alpinum Bonpland ex DC; Galeano 1995), was consistently found to be nested inside C. echinulatum, while only distantly related to the second subspecies of C. alpinum (C. alpinum ssp. alpinum). Hence, C. alpinum ssp. ecuadorense is here considered part of C. echinulatum. Observations on fruit morphology indicate that fruit surface sculpture changes gradually in Ecuador along a northeastern–southeastern–southwestern–northwestern cline. In the northeastern province of Napo (type locality of C. echinulatum; Galeano 1995), fruits are strongly warty; in the northwestern province of Pichincha (type locality of C. alpinum ssp. ecuadorense; Galeano 1995) fruits are minutely warty; in south Ecuadorian populations, fruit sculpturing is intermediate to the northern populations. Based on observations in C. echinulatum (Ruiz 1993) and congeners with similar fruit morphology (Galeano & Bernal 2005), seed dispersal vectors of these palms are mammals (fruit bats, agoutis, squirrels) and birds (guans, toucans, motmots, quetzals, jays, parrots, macaws, thrushes, and toucans). Pollination in C. echinulatum may occur by bees and/or beetles (Ruiz 1993; Knudsen et al. 2001).

Sampling and laboratory procedures

Prior to field collection, species distribution modelling was used to identify climatically suitable areas based on geo-referenced herbarium specimens of C. echinulatum using the MAXENT method (Phillips et al. 2006; further details given below). The obtained map was used to guide sampling efforts during field work in February 2006. Successful sampling occurred at a total of 18 localities, 12 of which represented new locality records for the species. Leaf and root tip material was collected for a total of 121 individuals. Sampling of closely neighbouring individuals was omitted in order to avoid bias caused by sampling of full-siblings.

DNA was extracted using the E.Z.N.A. SP Plant DNA Miniprep Kit (Omega Bio-tek). Initial sequencing of chloroplast (matK and trnD–trnT) microsatellites, i.e. simple sequence repeat (SSR), revealed low levels or no variation. A total of 38 nuclear SSR markers were screened (Gaitán 2003; Billotte et al. 2004a, b). Seven of these provided polymorphic SSR data and were selected for further analysis: mPdCIR015, mBgCIR077, and Ca1, Ca13, Ca17, Cs5 and Cs24 (Table S1, Supplementary material). SSR loci were amplified by polymerase chain reaction (PCR) with one of the primers l-end-labelled with the fluorescent dye CY-5. PCR volumes of 6 μL were prepared using the following reagent concentrations: 0.5 μm primer, 0.2 mM dNTP (GE Health Care), 1 U of Taq polymerase (Ampliqon), 0.6 μL of 10X ammonium reaction buffer and 20–50 ng DNA template. Thermal cycling conditions were as follows: 4 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 40 s at annealing temperature (Tₐ) and 60 s at 72 °C (for Tₐ see Table S1), and a final extension step of 5 min at 72 °C. PCR products were visualized on an ALF Express sequencer using 1 μL of PCR product together with internal standard sizes 100 and 300 and one external standard (50–500 bp; ALF Express). Allele sizes were determined using the software ALFWIN FRAGMENT ANALYZER 1.01 (ALF Express). For each unique allele, the repeat number was verified through sequencing of purified PCR products (QIAquick PCR purification kit, QIAGEN). Sequencing was conducted by Macrogen Inc.
Individual-based analyses

We inferred the hierarchical structure in the data and assigned individuals to clusters by using the Bayesian genotype clustering method of Pritchard et al. (2000) implemented in the program STRUCTURE version 2. The optimal number of genetic clusters $K$ was found by estimating likelihoods for increasingly larger $K$ and choosing the $K$ for which the log likelihood reaches a plateau (Pritchard et al. 2000). We also used Evanno et al.’s (2005) $\Delta K$ statistic, i.e., second order rate of change in the likelihood of $K$ corrected by the standard deviation of the likelihood of $K$ obtained from multiple runs. The modal $\Delta K$ was then used to determine the optimal value of $K$. Three models of ancestry and population intercorrelation were investigated: (i) partial population admixture/correlated allele frequencies; (ii) no-admixture/correlated allele frequencies; and (iii) no-admixture/independent allele frequencies. For each model, two independent Markov chain Monte Carlo (MCMC) runs were initiated per value of $K$ ranging from $K = 1–12$, with each chain consisting of $10^6$ burn-in generations followed by $10^6$ sampling generations, and default settings otherwise.

Under model 1, STRUCTURE returns for each of the $K$ populations an $F_{ST}$ analogue that describes the degree of genetic differentiation of population $k$ from a hypothetical ancestral population (Falush et al. 2003). In a unidimensional stepping-stone scenario and assuming constant rates of genetic drift in all $K$ populations, the interpopulation dispersal history can thus be inferred as paths from low to high $F$ values. For each cluster $k$, we calculated a mean $F$ value from a total of $10F$ estimates obtained by running eight additional MCMC chains under model 1.

In order to locate areas of genetic discontinuity (i.e., barriers), we conducted a spatial Bayesian clustering (BC) analysis implemented in the program GENELAND (Guillot et al. 2005). Based on geo-referenced individual multilocus genotype data, GENELAND infers for each cluster $k$ the geographic distribution of posterior probabilities of belonging to that population. Barriers were detected as geographic areas of global low posterior probability of population membership. We ran two independent MCMC chains for a fixed value of $K = 4$ as found in STRUCTURE ($10^6$ generations; sampling frequency 100; $10^5$ generations burn-in), using the allele frequency model of Falush et al. (2003) and otherwise following the instructions of the authors of the program.

Population-based analyses

*Determiniation of sample groups.* In order to define units for population-based analyses, we agglomerated individuals into clusters of samples (sample groups) by making use of information on the proximity of the sampled individuals in an effective distance space. Effective distances, i.e. shortest distances through areas of suitable climatic conditions, were calculated as least-cost distances (LCD) through a friction landscape of climatic suitability (Ray 2005). We then used a simple hierarchical clustering algorithm (UPGMA) to obtain groups of nearest individuals based on rescaled pairwise least-cost similarities, $LCS = 1 - ([LCD]/[max(LCD)])$. This approach is analogous to the one of McRae et al. (2005), who used geographic distances in order to cluster individuals into populations without imposing a priori notions of population boundaries on the data. As pointed out by McRae et al. (2005), sample groups do not necessarily correspond to discrete natural populations, but have to be interpreted as local clusters of samples with allele frequencies representative of those at cluster centroids. LCDs were calculated as follows. A species-specific friction map was obtained by modelling climatically suitable areas with the species distribution modelling method implemented in MAXENT version 2.3 (Phillips et al. 2006), using the $1 \times 1$ km resolution climate data from WORLDCLIM (http://www.worldclim.org) and 58 unique grid cell geo-records of *C. echinulatum* by dividing occurrences into training (75%) and test (25%) data sets, setting the lower threshold for species presence to 13.023, and using default values otherwise. Seven climatic variables of low intercorrelation and presumed importance for the distribution of *C. echinulatum* were identified: temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, temperature annual range, annual precipitation, precipitation seasonality and precipitation of the warmest quarter. A friction map was obtained in ARCGIS 9.2 (ESRI) by inversing the scale, so areas of high suitability had low friction. Then, PATHMATRIX (Ray 2005) implemented in ARCGIS 3.3 (ESRI) was used to find for each pair of samples their LCD in the obtained friction landscape. PATHMATRIX also returns a matrix of along-path distances (APD), i.e. pairwise geographic distances in kilometres along the least-cost path. UPGMA clustering of individuals was conducted in NTSYSpc version 2.11w (Exeter Software), and sample groups were delineated based on 90% LCS. This resulted in the recognition of seven populations and the obliteration of the sole Peruvian sample (SEPE0795) from all subsequent population-based analyses. One sample group had fewer than six individuals (Azuyay). Consequently, all population-based analyses were carried out with and without this sample group included, in the following referred to as the 7-population and 6-population data set analyses, respectively.

Descriptive statistics, within- and among-population differentiation, and historical isolation. For each sample group, allelic richness $A$ was calculated using the program MICROSATellite analyser (MSA) version 4.0 (Dieringer & Schlötterer 2003). Tests for significant deviation from Hardy-Weinberg
equilibrium (HWE) within sample groups \((F_{ST})\) were conducted in **GENEPOP** (Raymond & Rousset 1995). We also used **GENEPOP** to test for linkage equilibrium among all pairs of loci using Fisher’s exact method.

In order to test if deviations from HWE within samples were due to short gene dispersal distances in *C. echinulatum*, we tested for within-population isolation by distance for all populations with \(n > 15\) using Mantel matrix correlation tests implemented in **ARLEQUIN** version 3.0 (Excoffier et al. 2005). Tests were based on matrices of geographic and genetic distances among individuals, the latter estimated as \(D_{PS}\) (proportion of shared alleles; Bowcock et al. 1994) and calculated in **MSA** with 20 000 permutations.

Pairwise population differentiation was estimated using Weir & Cockerham’s (1984) \(\theta_{ST}\), an unbiased estimator of \(F_{ST}\) (hereafter referred to as \(F_{ST}\)). Mean \(F_{ST}\) and associated 95% confidence intervals (CI) were calculated in **MSA** using the one-allele jackknife approach. Significance of pairwise \(F_{ST}\) values was evaluated by shuffling genotypes among sample groups and conducting a total of \(10^4\) permutations.

Historical population isolation was investigated by assessing whether mutations contributed significantly to population differentiation. We tested if \(R_{ST}\), an \(F_{ST}\) analogue based on the stepwise mutation model (Slatkin 1995), was significantly larger than \(F_{ST}\) (Hardy et al. 2003). We used **SPAGEDI** version 1.2 (Hardy & Vekemans 2002) to calculate global and pairwise \(R_{ST}\) values and tested whether \(R_{ST} > pR_{ST}\) by permuting allele sizes among allelic states \((10^3\) cycles), where \(pR_{ST}\) approximates the null distribution of \(F_{ST}\).

**Interpopulation phylogeographic relationships.** Under recurrent gene flow, conspecific populations may share a reticulating rather than bifurcating phylogeographic history (Legendre & Makarenkov 2002). Interpopulation relationships were therefore depicted in a neighbour-joining (NJ) reticulogram constructed in **T-Rex** (Makarenkov 2001) using the pairwise population distance measure \(D_{\lambda}\) (Takezaki & Nei 1996) as calculated in **MSA**. Robustness of NJ relationships was evaluated through 10^3 bootstrap replications conducted in **PHYLIP** version 3.65 (Felsenstein 1995). We used step criterion \(Q_2\) (see Legendre & Makarenkov 2002 for further details) owing to its conservative behaviour, generally adding more reticulations to the tree when compared to other stop criteria (Legendre & Makarenkov 2002).

**Isolation by distance and testing the U-cline hypothesis.** In order to test if gene flow patterns in *C. echinulatum* corroborate the U-cline hypothesis, we compared Mantel test statistics obtained from different tests correlating pairwise genetic distances \([F_{ST} / (1 - F_{ST})]\) (Rousset 1997) with one of the following measures of interpopulation spatial and/or bioclimatic distance. (i) U-cline geographic distances which were calculated as the shortest geographic distances along the stepping-stone path: Napo, Tungurahua, Zamora–Chinchipe, Loja–El Oro, Azuay, Bolivar–Cotopaxi, Pichincha (Fig. 1). (ii) U-cline APD and (iii) U-cline LCD distances. Initial inspection of a map of simple LCD paths showed that the shortest LCD paths between northeastern and northwestern sample groups crossed the northern Andes. We therefore recalculated APD and LCD distances by constraining least-cost paths not to exceed an altitudinal limit of 3000 m above sea level and obtained cost distances consistent with the U-cline model (the U-cline APD and U-cline LCD). (iv) We also investigated the effects of the strengths of climatic filters to migration along the U-cline path by conducting a partial Mantel test (Smouse et al. 1986) that assessed how much of the genetic variation could be attributed to the effects of U-cline LCD after controlling for the geographic component in this measure, i.e. the U-cline APD. (v) Using partial Mantel tests, we also investigated whether the U-cline model was a significantly better description of distribution of allele frequencies than the structure identified in individual-based BC analyses (see above) by defining a distance matrix with 0-distances among sample groups belonging to the same BC cluster and distances of 1 among sample groups belonging to different BC clusters. Further alternative hypotheses were investigated by using (vi) Euclidean geographic distances and (vii) simple LCD distances. All population distance measures were calculated as medians of individual distance measures. Mantel tests were conducted in **ARLEQUIN** using 20 000 permutations.

**Coalescent-based demographic estimates of population divergence across the Andes.** An isolation-with-migration model (Hey & Nielsen 2007) was used to estimate the approximate timeframe of cross-Andean divergence between populations from opposite sides of the Andes. We analyzed the population pair Zamora–Chinchipe/Loja–El Oro, as this was found to be the focal area of cross-Andean gene flow (see below), using the program **IMA** (Hey & Nielsen 2007). Initial searches in parameter space using different parameter settings were conducted in order to find appropriate values for prior bounds (-q1 -m1 10 -m2 10 -t 10) and heating parameters to ensure sufficient mixing (-f6 -n8 -g1 0.8 -g2 0.9 -k8). Two independent MCMC runs were initiated and ran for 100 000 burn-in plus 50 million sampling generations; sampling occurred at each 100th generation. Assuming a molecular clock for the inferred timeframe of population isolation, we converted estimates of divergence time scaled by the neutral mutation rate into divergence times in years using published mutation rates commonly observed in microsatellites (10^-6–10^-4; Selkoe & Toonen 2006).
Results

In 91.7% of the samples, complete genotypes were scored. Observed numbers of alleles were 9, 2, 4, 7, 8, 8, and 10, for the markers, mBgClR077, Ca1, Ca13, Ca17, Cs5, and Cs24, respectively, with marker-wise allelic richness $A_r$ ranging from 1.45 to 3.17 (Table S1). Linkage equilibrium could not be rejected for any pair of loci after Bonferroni correction for multiple comparisons.

Individual-based analyses

BC analyses conducted in STRUCTURE revealed that values of log likelihood of $K$ increased steeply from $K = 1$ to $K = 4$, at which point the likelihoods of $K$ levelled off (Fig. S1, Supplementary material). Also, the $\Delta K$ statistic supported an optimal value of $K = 4$ for all three ancestry/allele frequency correlation models (Fig. S1). Inspection of individual assignments revealed that the structure at $K = 4$ corresponds to a northwestern (NW), a southwestern (SW), a southeastern (SE), and a northeastern cluster (NE) (Fig. S2, Supplementary material). Under model 3, a secondary peak in $\Delta K$ was observed at $K = 2$. This structure represented a west/east Andean split. Geographic substructuring was also evident at $K = 5$ and $K = 6$, with a fifth and sixth cluster corresponding to a subdivision of the NE cluster and the off-splitting of a single individual (SEya0810) from the SE cluster, respectively (Fig. S2). The SEya0810 individual showed genetic affinity to both the SE Zamora–Chinchipe cluster and the NE Tungurahua cluster and could represent an un-sampled intermittent population in the SE area.

Spatial BC of genotypes was used to depict the posterior probability landscape of population membership for each of the four clusters identified in the STRUCTURE analysis (Fig. 2). Areas of global low posterior probabilities and steep turnover in probabilities, reflecting potential barriers, were found along the higher Andes and along the western Andean slopes at the Giron-Paute deflection.

Under model 1, $F$ values for the clusters SE, NE, SW, and NW were on average $0.3004 \pm 0.0002$, $0.3102 \pm 0.0002$, $0.3761 \pm 0.0003$, and $0.4584 \pm 0.0002$, respectively, suggesting a westward dispersal scenario followed by northward migration on both sides of the Andes.

Population-based analyses

Descriptive statistics, within- and among-population differentiation, and historical isolation. Population-wise mean allelic richness $A$ ranged from 1.60 to 2.17 (Table S2, Supplementary material).
Table 1  Pairwise genetic distances between seven sample groups of Ecuadorian Ceraxylon echinulatum. Upper diagonal: mean $F_{ST}$ estimated using Weir & Cockerham’s (1984) estimator $\theta_{ST}$ and associated 95% confidence intervals calculated in msa (Dieringer & Schlötterer 2003) using the one-allele jackknife procedure. $F_{ST}$ values that are significant at the 5% level after Bonferroni correction for multiple comparisons are shown in bold italics. All values except † are significant under the false discovery rate criterion (Benjamin & Hochberg 1995) with $q = 0.05$. Lower diagonal: Slatkin’s (1995) $R_{ST}$. $R_{ST}$ values that are significantly larger than $F_{ST}$ are indicated in bold. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$

<table>
<thead>
<tr>
<th>Cluster‡</th>
<th>Sample group (sample size)</th>
<th>Pichincha</th>
<th>Bolivar–Cotopaxi</th>
<th>Azuay</th>
<th>Loja–El Oro</th>
<th>Zamora–Chinchipe</th>
<th>Tungurahua</th>
<th>Napo</th>
</tr>
</thead>
<tbody>
<tr>
<td>NW</td>
<td>(1) Pichincha (n = 20)</td>
<td></td>
<td>0.118**</td>
<td>0.277*</td>
<td>0.364***</td>
<td>0.391***</td>
<td>0.570****</td>
<td>0.657****</td>
</tr>
<tr>
<td>NW</td>
<td>(2) Bolivar and Cotopaxi  (n = 18)</td>
<td>0.162</td>
<td>0.061–0.182</td>
<td>0.158–0.413</td>
<td>0.323–0.406</td>
<td>0.346–0.439</td>
<td>0.533–0.608</td>
<td>0.630–0.689</td>
</tr>
<tr>
<td>NW</td>
<td>(3) Azuay (n = 3)</td>
<td></td>
<td>0.186†</td>
<td></td>
<td>0.292****</td>
<td>0.345C</td>
<td>0.567****</td>
<td>0.661****</td>
</tr>
<tr>
<td>SW</td>
<td>(4) Loja–El Oro (n = 42)</td>
<td></td>
<td>0.436</td>
<td>0.002–0.387</td>
<td>0.252–0.335</td>
<td>0.295–0.402</td>
<td>0.538–0.597</td>
<td>0.638–0.689</td>
</tr>
<tr>
<td>SE</td>
<td>(5) Zamora–Chinchipe (n = 24)</td>
<td>0.618</td>
<td>0.530*</td>
<td>0.245</td>
<td>0.353***</td>
<td>0.509***</td>
<td>0.593***</td>
<td>0.624***</td>
</tr>
<tr>
<td>NE</td>
<td>(6) Tungurahua (n = 6)</td>
<td></td>
<td>0.599</td>
<td>0.140</td>
<td>0.286–0.388</td>
<td>0.474–0.544</td>
<td>0.562–0.625</td>
<td></td>
</tr>
<tr>
<td>NE</td>
<td>(7) Napo (n = 7)</td>
<td></td>
<td>0.605</td>
<td>0.482</td>
<td>0.317</td>
<td>0.680*</td>
<td>0.505</td>
<td>0.264***</td>
</tr>
</tbody>
</table>

‡cluster membership as identified in structure analysis, see Fig. 2. NW, northwestern; SW, southwestern; SE, southeastern; NE, northeastern.

material). Within-population monomorphism was observed in nine out of 49 (18%) population/locus combinations. This was restricted to the three markers mPdCIR015, mBgCIR077, and CaI. The northwesternmost (Pichincha) and northeasternmost (Napo) sample groups accounted together for more than half of the observed monomorphic population/locus combinations (43% and 29%, respectively). In all cases of within-population monomorphism, different alleles were fixed on opposite versants of the Andes (Table S2). Southern sample groups (Loja–El Oro and Zamora–Chinchipe) were polymorphic in all analyzed markers. Across all loci, number of private alleles, i.e. alleles occurring exclusively in one population, per number of total alleles were 2/14 (14.3%), 5/20 (25.0%), 2/26 (7.7%), 3/27 (11.1%), 1/15 (6.7%), 0/16 (0%), and 0/17 (0%) for the sample groups Napo, Tungurahua, Zamora–Chinchipe, Loja–El Oro, Azuay, Bolivar–Cotopaxi, and Pichincha, respectively.

Significant deviation from within-population HWE across all loci was found in all except in one sample group (Pichincha) at the 5% level and in two sample groups after Bonferroni correction (Bolivar–Cotopaxi, Zamora–Chinchipe; Table S2). Within-population analyses revealed significant associations between genetic and geographic distances (Pichincha: $r_{ln} = 0.17$, $P_{ln} = 0.012$; $r_{ln} = 0.21$, $P_{ln} = 0.003$; Bolivar–Cotopaxi: $r_{ln} = 0.36$, $P_{ln} = 0.001$; $r_{ln} = 0.48$, $P_{ln} < 0.001$; Loja–El Oro: $r_{ln} = 0.14$, $P_{ln} = 0.010$; $r_{ln} = 0.11$, $P_{ln} = 0.001$; Zamora–Chinchipe: $r_{ln} = 0.58$, $P_{ln} = 0.012$; $r_{ln} = 0.45$, $P_{ln} < 0.001$), pointing to a within-population Wahlund effect rather than null alleles as a major determinant of the observed population-wise deviations from HWE.

Among-population comparisons also revealed strong and significant genetic structuring in Ceraxylon echinulatum. Pairwise $F_{ST}$ values ranged from 0.12 to 0.66 (Table 1). Globally and across all loci, $R_{ST}$ was found to be significantly larger than $F_{ST}$ ($R_{ST} = 0.54$, $pR_{ST} = 0.41$, $P = 0.037$). Pairwise $R_{ST}$ values ranged from 0.13 to 0.62 (Table 1). Two population pairs displayed $R_{ST}$ values significantly larger than $F_{ST}$ at the 5% level, and a third population pair was marginally significant (Table 1), in all three cases involving the Loja–El Oro sample group. Furthermore, we observed a tendency of increased $R_{ST}$ values among population pairs including the Loja–El Oro sample group (mean $R_{ST} = 0.55 \pm 0.03$, 1 SE, $n = 6$) when compared to the remaining sample group pairs (mean $R_{ST} = 0.41 \pm 0.01$, 1 SE, $n = 15$; $P = 0.018$, Wilcoxon rank sum test), indicating a phylogeographically isolated position of the Loja–El Oro sample group.

Interpopulation phylogeographic relationships. Branching and reticulation patterns in the unrooted NJ reticulogram using the 7-population data set (Fig. 3) were consistent with the hypothesized U-shaped gene flow cline in C. echinulatum with crossing of the Andes in the south of Ecuador. Relationships were strongly supported in terms of bootstrap percentages (> 93%). Three significant reticulations were recognized, two of which crosses the Andes. Despite approximately equal geographic distances among populations in the NW and NE regions, northeastern sample
groups appeared to be more differentiated from each other than northwestern populations, as indicated by longer internodes between northeastern sample groups when compared to northwestern sample groups. All of these properties were reflected when analyzing the 6-population data set (data not shown), indicating that the inclusion of the Azuay population did not considerably bias our results.

Isolation by distance, effective distance and U-cline distances. Mantel and partial Mantel tests strongly supported the U-cline hypothesis for *C. echinulatum* in both the 6- and 7-population data sets (Table 2, Fig. 4). Correlations were nonsignificant between genetic distances and spatial distances that did not take the Andes barrier explicitly into account, e.g. Euclidean geographic distances and LCD effective distances. When forcing geographic LCD and APD distance paths to follow the hypothesized U-shaped cline in gene flow, however, strong correlations between genetic and geographic/effective distances appeared ($r \geq 0.90, P \leq 0.003$; Table 2). This demonstrates that the northern and central Andes of Ecuador form a strong barrier to gene flow in *C. echinulatum*. In this system, gene flow is mainly limited by geographic distance along the U-cline path ($r = 0.90, P \leq 0.001$) with minor but significant effects of bioclimatic friction, as indicated by an increase in $r$ when using LCD effective distances ($r = 0.92, P \leq 0.003$, Fig. 4) and by this relationship remaining significant also after controlling for APD using a partial Mantel test ($P \leq 0.05$). A partial Mantel test using a distance matrix based on the four clusters derived from BC analysis indicate that the U-cline model is superior to a four-cluster model (Table 2).

Cross-Andean divergence times. A single peak in the marginal posterior probability distribution of divergence times was recovered at 0.175 (Supplementary material, Fig. S3), and the 95% credibility interval ranged from 0.045 to 0.615. Thus, based on published mutation rates in microsatellites, divergence across the Andes can be inferred to have occurred before 615 000 years ago, clearly indicating that cross-Andean disjunction in *C. echinulatum* cannot be ascribed to Late Tertiary mountain-building processes.

Discussion

The U-cline hypothesis and cross-Andean dispersal in the AHZ

The current study presents strong evidence for *Ceroxylon echinulatum* being a cline species that has surpassed the Andes barrier in the AHZ by means of Quaternary dispersal. In summary, complementary individual- and population-based analyses showed strong genetic differentiation in *C. echinulatum* (Table 1), limited gene flow across the northern and central Andes, and cross-Andean gene flow in the AHZ (Fig. 3). Gene flow was found to follow the U-cline model, and was mainly limited by distance along this path (Fig. 4), with minor but significant effects of habitat friction (Table 2). A westward migration history from a southeast Andean source area could be inferred, and an instance of past population isolation was evident in the western AHZ area. In combination, these results indicate a severe barrier effect of the Andean mountain chain on a mid-elevation species, but also that this barrier is surmounted in a specific area of diminished barrier strength in southern Ecuador/northern Peru (Figs 1–3).

We note that the distribution of *C. echinulatum* along the eastern slopes of the Peruvian Andes to date is poorly understood, and hence, we cannot exclude the possibility that the observed U-cline might be nested in a Y-shaped overall distribution extending farther south into Peru.

Floristic studies and studies on individual plant and animal groups have previously pointed to the AHZ as a region of biotic exchange between east- and west-Andean lowland and mid-elevation biotas (Simpson 1975;
In line with these findings, we found evidence of cross-Andean dispersal in the area of the AHZ. However, the precise location of cross-Andean connectivity in the AHZ cannot be inferred from the current data. In our study, only a single Peruvian sample was included. Hence, we cannot exclude the possibility that the crossing of the Andes would have occurred farther south in the north Peruvian region. More sampling in northern Peru as well as in the under-sampled area north of the Zamora–Chinchipe sample group is necessary in order to determine the fine-scale patterns of cross-Andean dispersal in the AHZ.

The strong genetic structure observed at small geographic scales is suggestive of dispersal limitation in *C. echinulatum*. The sample groups examined for within-population isolation by distance in the current study covered distances of 4 km (Zamora-Chinchipe), 6 km (Pichincha), 66 km (Bolivar–Cotopaxi), and 89 km (Loja–El Oro); in all four cases, significant isolation by distance and deviation from HWE at the 5% level were observed. The lack of panmixia in the Zamora–Chinchipe sample group covering a distance of only 4.3 km may be due to a combined effect of short gene flow ranges and the narrow, linear distribution of this population at its locality. The fact that *C. echinulatum* displays dispersal limitation at small spatial scales highlights the importance of rare long-distance dispersal events for historical processes of colonization and range expansion at the metapopulation level (Nathan & Muller-Landau 2000), especially in a highly structured landscape where short-distance dispersion is hindered by the presence of barriers and habitat friction (Spiegel & Nathan 2007).

Strong genetic structuring was also observed at the interpopulation level, with $F_{ST}$ values ranging from 0.12 to 0.66 (Table 1). We observed an increasing degree of marker-wise monomorphism on each side of the Andes when proceeding from southern to northern sample groups, with the highest monomorphism on western and eastern sides found in the Pichincha and Napo sample groups, respectively. $F$ values derived from the *structure* analysis also indicated that northern clusters had undergone greater genetic drift following subdivision of the ancestral population when compared to their southern counterparts.
This is indicative of a northward migration history on both sides of the Andes. Colonization dynamics are known to increase genetic variation among populations, with drift within the newly founded populations being the source of the enhanced differentiation (Giles & Goudet 1997).

Narrow, linear habitats are highly prone to fragmentation (Graves 1988), resulting in a stepping-stone metapopulation system with subpopulations of small effective sizes and, hence, high rates of drift. Thus, the particular narrow and sinuous shape of the distribution of suitable habitat in *C. echinulatum* (Fig. 1) could further explain the great genetic differentiation observed. Low rates of migration between subpopulations and the dioecious breeding system of *C. echinulatum*, reducing the effective population sizes, would further speed up differentiation via drift.

**Recent gene flow and phylogeographic history**

Comparison of estimates of genetic differentiation derived from measures neglecting mutational processes (\(F_{ST}\)) vs. measures assuming an explicit mutation model (\(R_{ST}\)) may provide information on the relative importance of short- and long-term processes generating the observed genetic structure. Our results indicate that both types of demographic processes have been important for genetic differentiation in *C. echinulatum*. \(F_{ST}\)-based estimates of pairwise population differentiation correlated strongly with models incorporating present-day landscape features of the Andes (Table 2), including a spatial component (distance), a bioclimatic component (habitat friction) and a biological component (the Andes barrier, U-cline models). We found evidence for all three landscape elements shaping gene flow in *C. echinulatum*, with the Andes barrier having the strongest impact, followed by distance and habitat friction (Table 2, Fig. 4). In this, our results contribute to an increasing amount of studies that have found landscapes to be primary determinants of genetic variation in neutral genetic markers. For example, profound effects of altitudinal relief and mountain ridges were reported by Giordano et al. (2007), whereas other studies found significant associations between genetic differentiation and ecological parameters (e.g. Jørgensen et al. 2005) or least-cost distances (studies listed in McRae 2003).

However, long-term historical processes also contributed to the genetic variation in *C. echinulatum*. With longer periods of geographic separation, the contribution of mutations to the genetic differentiation of populations becomes non-negligible compared to migration and drift processes (Hardy et al. 2003). Interpretation of \(F_{ST}\) and \(R_{ST}\) values per se may be problematic due to analytical (Balloux & Lugon-Moulin 2002) and conceptual problems related to their underlying assumptions hardly ever being met in natural systems (Lugon-Moulin et al. 1999). However, the complementary and comparative use of \(F\) and \(R\) statistics has been suggested to explore the potential signatures of short- and long-term demographic processes in microsatellite data (Lugon-Moulin et al. 1999; Hardy et al. 2003). Under the presumption of similar mutation rates in the analyzed sample groups, we inferred significant population separation in the Loja–El Oro sample group (= SW cluster in BC analyses) relative to the mutation rate. The isolated position of this sample group was also evident in the reticulumogram analysis that revealed little genetic exchange between this and other sample groups (Fig. 3), raising the question of what might have caused the historical separation of this population.

The Loja–El Oro sample group is physically delineated by the Andes barrier, the Girón–Paute deflection and by the dry coastal habitats of Ecuador and northwestern Peru (Fig. 1). Based on the distribution patterns of birds (Poulsen & Krabbe 1998) and high-Andean plants (Jørgensen et al. 1995), the Andes barrier and Girón–Paute deflection have been suggested to constitute areas of once much stronger filter barriers for Andean species (but see García-Moreno et al. 1999). During the cold and dry conditions of the Pleistocene glacial periods, distributional ranges of montane forest were displaced down-slope in the order of several hundreds of altitudinal metres (Burnham & Graham 1999), resulting in elevated resistance to cross-Andean gene flow. At the same time, dry valleys as the Girón–Paute became even drier and, hence, stronger filter barriers to moist montane forest species (Jørgensen et al. 1995). Furthermore, on the basis of floristic, paleoecological and paleoclimatic data, Simpson (1975) argued for the presence of a once more extensive area of moist montane forests along the west-Andean slopes of Peru and southwestern Ecuador during the glacial periods of the Pleistocene. These forests could, indeed, have functioned as a glacial refuge for SW *C. echinulatum*, and, together with the increased barrier effect of the Andes and Girón–Paute, explain the inferred signal of historical isolation in the Loja–El Oro sample group.

**Cross-Andean phylogeography and implications for the study of Andean diversification**

Cross-Andean disjunct distributions have been reported in 1438 species (c. 18%) of Ecuador’s lowland vascular plants (Jørgensen & León-Yánez 1999). These patterns have lead to the interpretation that much of the lowland biota has evolved by means of vicariance due to the rise of the Andes (e.g. Raven 1999). Sound support for the Andean uplift vicariance hypothesis has accumulated during the last decade. Several molecular studies of species complexes or populations distributed from either side of the Andean divide, including freshwater fish (Perdices et al. 2002),
rainforest trees (Dick et al. 2003; Trénel et al. 2007), butterflies (Brower 1994), frogs (Weigt et al. 2005), snakes (Zamudio & Greene 1997), birds (Cheviron et al. 2005) and bats (Hoffmann & Baker 2003), found levels of cross-Andean genetic divergence consistent with the timeframe of the major Andean orogeny (see also Moritz et al. 2000).

The temporal framework of cross-Andean divergence in the current study has been estimated to the Quaternary (< ~600 000 years ago). At that time, the current topographic setting of the south Ecuadorian/north Peruvian Andes has largely been in place (Gregory-Wodzicki 2000), and hence, orogeny-driven vicariance must be rejected as an explanation of the disjunct cross-Andean distribution in C. echinulatum. Support for the across-Andean dispersal hypothesis has also been reported in studies of bees (Dick et al. 2004), bats (Hoffmann & Baker 2003) and birds (Garcia-Moreno et al. 1999).

Our results suggest that present-day patterns of distribution and diversity in low-vagility Andean taxa such as C. echinulatum can indeed be the result of Quaternary dispersal dynamics through the high-structure matrix of the Andean landscape. Hence, we caution against a priori attributing cross-Andean disjunct distributions to ancient vicariance events without further reasoning. Our results point to a complex and dynamic interaction of dispersal and vicariance, including founder dynamics, across-barrier dispersal, divergence-by-distance, and secondary isolating mechanisms and population coalescence. The concept of divergence along geographic and ecological clines is well-known from the study of ring species (Coyne & Orr 2004). In Ecuador, eastern and western slopes differ in edaphic composition and seasonal precipitation, with generally younger soils and more pronounced seasonality along the western mountain slopes (Neill & Jorgensen 1999). Seasonality is known to influence flowering time (Whitmore 1998). Although the status of reproductive isolation cannot be inferred on the basis of the current data, it is worth noting that flowering in C. echinulatum has been reported to occur in May in the province of Pichincha (Ruiz 1993), suggestive of temporal isolation and ecological divergence along the path of the cline.

The results of this study provide an example of clinal divergence and Quaternary dispersal in a complex landscape matrix characterized by linear and narrow habitats, ecological gradients, dispersal corridors and historical landscape rearrangements. In this, our study adds to the increasingly complex picture of Andean diversification, and points to the importance of an integrated analysis of spatial, ecological and historical aspects of landscapes for an understanding of the micro-evolutionary processes underlying present-day patterns of diversity and distribution.

Acknowledgements

We gratefully thank Priscilla Muriel, Alvaro Perez, Daniela Cevallos, and Hugo Navarrete for their invaluable help and assistance during field work and Anni Slot, Camilla Håkansson and Jane Frydenberg for assistance in the laboratory. Special thanks go to Jean-Christophe Pintaud for the loan of tissue, and Gloria Galeano, Jens-Christian Svenning, and Henrik Balslev for their advice and support throughout the study. We thank the Carlsberg Foundation and the Danish Natural Science Research Council (grant 272-06-0476) for financial support, and Knud Heigaards Fond, Frimodt Heineke Fonden and HKH Kronprins Frederiks Fond for additional funding of PT’s field work.

References

Billotte N, Couvreur T, Marseilliac N et al. (2004b) A new set of microsatellite markers for the peach palm (Bactris gasipaes Kunth); characterization and across-taxon utility within the tribe Cocoeae. Molecular Ecology Notes, 4, 580–582.


This work was conducted in the context of P.T.’s PhD study on the evolutionary history of the wax palm subfamily Ceroxylidae (Arecaceae). A main motivation was to understand the spatio-temporal and ecogeographical drivers of present-day patterns of diversity and distribution. M.M.H.’s research focuses on estimation of demographic parameters and detection of adaptive divergence in wild populations, particularly fishes, using population genetics and genomics approaches. S.N. studies plant macroecology, with special focus on the distribution and diversity patterns of European plants. F.B.’s research covers plant systematics and evolution in especially Asian Marantaceae and South American palms.

**Supplementary material**

The following supplementary material is available for this article:

**Table S1** Microsatellite marker information, including primer sequence, annealing temperature, allelic range, and observed number of alleles.

**Table S2** Descriptive statistics for seven populations of *Ceroxylon echinulatum*, including allelic richness *A* and within-population and global *F*-statistics.

**Fig. S1** STRUCTURE (Pritchard et al. 2000) results.

**Fig. S2** Assignment of individuals to clusters.

**Fig. S3** The marginal posterior probability distribution of the divergence time *t* for the sample group Loja–El Oro and Zamora–Chinchipe, scaled by the neutral mutation rate.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03853.x

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.