MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 2543–2558 doi: 10.1111/mec.12746

Evidence for ecological divergence across a mosaic of soil types in an Amazonian tropical tree: *Protium subserratum* (Burseraceae)

TRACY M. MISIEWICZ and PAUL V. A. FINE

Department of Integrative Biology, University of California, Berkeley, CA 94720, USA

Abstract

Soil heterogeneity is an important driver of divergent natural selection in plants. Neotropical forests have the highest tree diversity on earth, and frequently, soil specialist congeners are distributed parapatrically. While the role of edaphic heterogeneity in the origin and maintenance of tropical tree diversity is unknown, it has been posited that natural selection across the patchwork of soils in the Amazon rainforest is important in driving and maintaining tree diversity. We examined genetic and morphological differentiation among populations of the tropical tree *Protium subserratum* growing parapatrically on the mosaic of white-sand, brown-sand and clay soils found throughout western Amazonia. Nuclear microsatellites and leaf morphology were used to (i) quantify the extent of phenotypic and genetic divergence across habitat types, (ii) assess the importance of natural selection vs. drift in population divergence, (iii) determine the extent of hybridization and introgression across habitat types, (iv) estimate migration rates among populations. We found significant morphological variation correlated with soil type. Higher levels of genetic differentiation and lower migration rates were observed between adjacent populations found on different soil types than between geographically distant populations on the same soil type. P_{ST}-F_{ST} comparisons indicate a role for natural selection in population divergence among soil types. A small number of hybrids were detected suggesting that gene flow among soil specialist populations may occur at low frequencies. Our results suggest that edaphic specialization has occurred multiple times in *P. subserratum* and that divergent natural selection across edaphic boundaries may be a general mechanism promoting and maintaining Amazonian tree diversity.

Keywords: Amazon, hybrid, microsatellite, soil specialist, white-sand forest

Received 27 November 2013; revision received 20 March 2014; accepted 28 March 2014

Introduction

The role of gene flow and natural selection in the origin and maintenance of species diversity has been a long-standing point of contention in the field of evolutionary biology (Haldane 1948; Ehrlich & Raven 1969; Endler 1977). Traditionally, theory has posited that gene flow among populations will act to homogenize them, limiting local adaptation and maintaining species cohesion (Haldane 1948; Hendry *et al.* 2001). Therefore, prevention of gene flow through geographical isolation is thought to be the first step in species divergence, with reproductive isolation among populations eventually developing by genetic drift (Mayr 1963). More recently however, there has been an increasing acknowledgement that divergent adaptation across heterogeneous environments can also lead to lineage divergence. Population divergence driven by natural selection can occur in the absence of gene flow if geographical barriers are present or in the face of gene flow if natural selection itself is strong enough to cause significant fitness differences across habitat boundaries. As long as gene flow does not swamp population differentiation, barriers to reproduction can develop, further

Correspondence: Tracy M. Misiewicz
E-mail: tmisiewicz@berkeley.edu

promoting population divergence and eventually leading to irreversible reproductive isolation (Rundle & Nosil 2005; Nosil 2012).

Genetic differentiation across heterogeneous environments has been well documented in temperate plant populations with some of the best-known examples stemming from local adaption to variable soils (Antonovics & Bradshaw 1970; Rosenthal *et al.* 2005; Anacker *et al.* 2011). This suggests that specialization across edaphically heterogeneous environments could result from selective pressures that are strong enough to promote population-level divergence even when plant populations are not geographically isolated. Nevertheless, gene flow can become limited among populations for many other reasons, and a variety of evolutionary processes can result in genetic heterogeneity across populations (Latta 2004). The processes of genetic drift and natural selection, however, are predicted to leave markedly different genetic signatures across populations. Drift is expected to generate stochastic population genetic structure, which will correlate with geographical distance when it is sufficient to limit migration and gene flow between populations (Latta 2004). Alternatively, if natural selection is important in population divergence, then we expect to see a pattern of 'isolation by adaptation' (IBA) (Nosil *et al.* 2007, 2009). The expected signal for IBA is similar to that of isolation by distance except that the level of population differentiation is dictated by environmental similarity as opposed to geographical proximity. Theory predicts that when differential adaptation across ecological gradients leads to a reduction in gene flow, populations found in differing habitats will experience lower effective migration rates and greater genetic differentiation than populations found in ecologically similar habitats. Reproductive isolation, either driven or reinforced by natural selection over a long period of time, will eventually have a genome-wide effect on allele frequencies (Grahame *et al.* 2006; Nosil *et al.* 2009; Thibert-Plante & Hendry 2010). Therefore, we can begin to discern the relative importance of these two forces in driving and maintaining differentiation by comparing patterns of genetic divergence relative to variables proximate to genetic drift (Euclidean distance) and natural selection (morphological and environmental variables) (Hendry & Taylor 2004; Nosil *et al.* 2007).

The lowland Amazon rainforest has the highest tree diversity in the world and is also characterized by steep soil gradients yet explanations for the origin of Amazonian species have typically revolved around hypotheses of nonecologically based divergence by genetic drift (reviewed in Haffer 2008). More recently, an alternative hypothesis of tropical tree speciation that is driven and/or maintained by natural selection across ecological boundaries has also been gaining attention in part due to the observation that congeners across diverse plant families are often occur with parapatric distributions on different soil types (Gentry 1988; Tuomisto *et al.* 1995; Fine *et al.* 2005, 2010).

Amazonian soils can strongly differ with regard to abiotic factors such as particle size, nutrient content and moisture level, and these differences can have a direct effect on trees found in different edaphic environments (Pregitzer *et al.* 2010; Smith *et al.* 2011). Moreover, physical and microclimatic variation among edaphically differentiated forest patches may influence the spatial organization of other organisms (Sääksjärvi *et al.* 2004; Álvarez Alonso *et al.* 2013). As a result, biotic interactions in the form of seed dispersal, herbivory, pollination services or pathogen prevalence may interact with abiotic differences to further magnify natural selection across ecological gradients for tropical trees. Yet, while correlations between edaphic shifts and lineage divergence have been demonstrated in a broader phylogenetic context (Fine *et al.* 2005), very few studies have investigated the importance of edaphic heterogeneity in tropical tree diversification at finer taxonomic scales.

Peru's Amazonian lowland rainforest provides an ideal system to examine the role of divergent natural selection in tree diversification because it contains a patchwork of dramatically different soil types each home to distinct tree communities (Fine *et al.* 2005, 2010).

The genus *Protium* (Burseraceae) is known to occur on a variety of soil types throughout the Neotropics (Daly 1987). It is a diverse clade of approximately 140 species of trees with around 100 species found in the lowland Amazon Basin, many of which are soil specialists (Fine *et al.* 2005). Edaphic specialization has evolved independently multiple times within the genus, and species adapted to particular soil types exhibit slower growth and higher mortality when grown outside their respective habitats (Fine *et al.* 2006). This supports the idea that physiological trade-offs accompany adaptations to alternative soil types (Fine *et al.* 2004). *Protium subserratum* (Engl.) is one of the few soil generalists in the genus, although the taxon is more accurately described as an incipient species complex with genetically and morphologically differentiated populations endemic to white-sand and non-white-sand soils (Daly & Fine 2011).

A recent phylogeographical study suggested that populations of *P. subserratum* found on white-sand forests that were separated by over 200 km were more genetically similar to one another than to populations found on brown-sand and clay soils and that some haplotypes were shared between populations through ongoing gene flow or incomplete lineage sorting (Fine

et al. 2013a). However, these results were based on DNA sequences from only three nuclear loci with few phylogenetically informative sites and low sample sizes. While these data were valuable for inferring broader evolutionary histories over a large geographical range, they were not able to provide insight regarding population dynamics.

Here, we conduct fine-scale analyses using nuclear microsatellites, morphological measurements and a sampling scheme that includes dense sampling of multiple parapatric population pairs found on the mosaic of white-sand, brown-sand and clay soil types distributed at varying distances across 100 km in the Peruvian Amazon. This experimental design represents a powerful natural experiment to simultaneously test the relative importance of habitat and geographical distance in limiting gene flow and driving divergence among edaphic specialist populations of Amazonian trees. If divergent natural selection across habitat boundaries is strong enough to impede gene flow, we expect to find stronger patterns of phenotypic and genetic differentiation in populations found on different habitat types and lower levels of gene flow across habitat boundaries compared to geographically isolated populations found in the same habitat type. Alternatively, if genetic drift has led to barriers to gene flow among populations, we would expect to see little to no phenotypic variation among populations, and strong patterns of isolation by distance, where populations that are more geographically distant from one another are also more genetically distinct than populations in close geographical proximity to each other regardless of habitat type. Moreover, because neutral forces such as genetic drift are expected to drive variation at putatively neutral loci and natural selection is expected to drive phenotypic variation, comparisons of phenotypic variation and neutral genetic variation across habitat boundaries can provide even further insight into the role of non-neutral evolutionary forces as drivers of divergence (Merilä & Crnokrak 2001; Leinonen *et al.* 2006). If stochastic mechanisms are driving population divergence, we expect levels of phenotypic variation among habitats to equal that of neutral genetic variation. If divergent natural selection across habitat boundaries is driving divergence, then phenotypic variation is expected to be greater than neutral genetic variation (Merilä & Crnokrak 2001; Leinonen *et al.* 2006).

We specifically addressed four major questions in this study: (i) Are populations phenotypically and genetically differentiated across all three habitat types? (ii) Can we detect a signature of natural selection over drift divergence among populations on different soil types? (iii) What is the extent of hybridization and introgression across habitat types? (iv) What is the relative importance of spatial distance and soil type in structuring *P. subserratum* tree populations and influencing migration rates among populations?

Materials and methods

Study system

Habitat types. Nonflooded forests found in the Peruvian Amazon have been classified into three broad categories based on soil type (Fine *et al.* 2005), which can be differentiated largely by their nutrient availability and geologic history (a more thorough discussion can be found in Hoorn 1993; Fine *et al.* 2005 and Frasier *et al.* 2008): (i) white-sand soils, which are extremely nutrient poor, include stunted canopies and exist today as geographically isolated habitat islands, often covering only a few square hectares; (ii) clay soils, which represent the most nutrient-rich soils in the western Amazon, have the highest water retention and tallest canopies; (iii) brown-sand soils have significantly higher nutrient availability than white-sand soils (Fine *et al.* 2005) and the height of their forest canopies represents an intermediate between white-sand and clay soil forests.

Focal plant taxa. Protium subserratum is part of the section *Papilloprotium*, which is comprised of four taxa that include both edaphic specialists and generalists (Daly & Fine 2011). It is sister to two white-sand specialist taxa, *P. alvarezianum* and *P. reticulatum,* found in white-sand forest patches in the Rio Negro Basin of Venezuela and Brazil (Daly & Fine 2011). *Protium subserratum* is a soil generalist that is common and widespread across the lowland Amazon Basin. It has small (approximately 4 mm length) fragrant white flowers, which are nectariferous and relatively large (approximately 1.5 cm diameter) red fruits (Misiewicz, personal observation). Little quantitative data regarding the reproductive biology of *P. subserratum* exist, but a variety of stingless bees have frequently been observed visiting flowers (Misiewicz 2014) and monkeys and large birds are hypothesized to be potential seed dispersers (Daly 1987). While no consistent morphological differences have been observed in flower or fruit characters among populations, significant differences in vegetative characters do exist (Daly & Fine 2011; Fine *et al.* 2013a). Within *P. subserratum*, vegetative morphological variation has been noted both across the range and within localized populations (Daly & Fine 2011). Daly and Fine (2011) grouped individuals into four distinct morphotypes based primarily on leaf morphology. Two of these morphotypes have very restricted geographical distributions: Morphotype 1 is restricted to non-white-sand forests in French Guiana, and Morphotype 4 is

restricted to Colombia's Caquetá. The other two have more widespread distributions: Morphotype 2 is associated with clay and brown-sand soils of the central and western Amazon, whereas Morphotype 3 is consistently associated with white-sand soils in the western Amazon (Daly & Fine 2011). Phylogenetic analysis by Daly and Fine (2011) demonstrated that these morphotypes do not form monophyletic clades and should continue to be considered one taxonomic species.

Further phylogeographical analysis by Fine *et al.* (2013a) included populations found on white-sand and non-white-sand soils sampled throughout the Amazon. They found two well-supported clades, one composed of northern Amazonian individuals from non-white-sand soils in Guyana and French Guiana and the other composed of western Amazonian individuals and a single individual from Guyana. Within the western clade, they found groupings composed of Peruvian white-sand individuals, Brazilian non-white-sand individuals and Peruvian non-white-sand individuals; however, none of these had posterior probability support higher than 0.75. These results are consistent with the previous classification of *P. subserratum* as a single species, and no evidence of differentiation between individuals collected on brown-sand or clay soil was detected.

While *P. subserratum* morphotypes are not monophyletic, a two-year reciprocal transplant study demonstrated that morphological variation observed between seedlings associated with white-sand and non-white-sand habitats is not completely due to plasticity (Fine *et al.* 2013b). Seedlings from three white-sand populations and three non-white-sand Peruvian populations (including both clay and brown-sand soils) were collected and transplanted in white-sand and clay soil habitats. Results indicated that seedlings initially collected from white-sand habitat grew slower and produced fewer leaves in both habitat types than seedlings initially collected from non-white-sand habitats. They also demonstrated that that leaf pubescence on new leaf growth was not influenced by habitat type. Leaflet thickness on the other hand did appear to be plastic. Additionally, there are clear quantitative and qualitative differences in secondary chemical compounds between populations found in white-sand and non-white-sand soil habitats (Fine *et al.* 2013b).

Study sites and sampling

Five study sites containing a total of eight populations (n = 5–54) of *P. subserratum* growing on white-sand, brown-sand and clay soil types were established in Loreto, Peru (Fig. 1). Individuals found in white-sand habitats corresponded to Morphotype 3 in Daly and Fine (2011) and the Peruvian white-sand morphotype from populations sampled in Fine *et al.* (2013a). Individuals found on brown-sand and clay soil types corresponded to Morphotype 2 described by Daly and Fine (2011) and the Peruvian non-white-sand morphotype from Fine *et al.* (2013a). Furthermore, seedlings from four of the eight populations sampled in this study [Clay-A, WS-B, BS-B, and WS-C (Fig. 1)] were included in the reciprocal transplant experiment published by Fine *et al.* (2013b). Adult individuals of *P. subserratum* from eight populations were tagged, mapped and

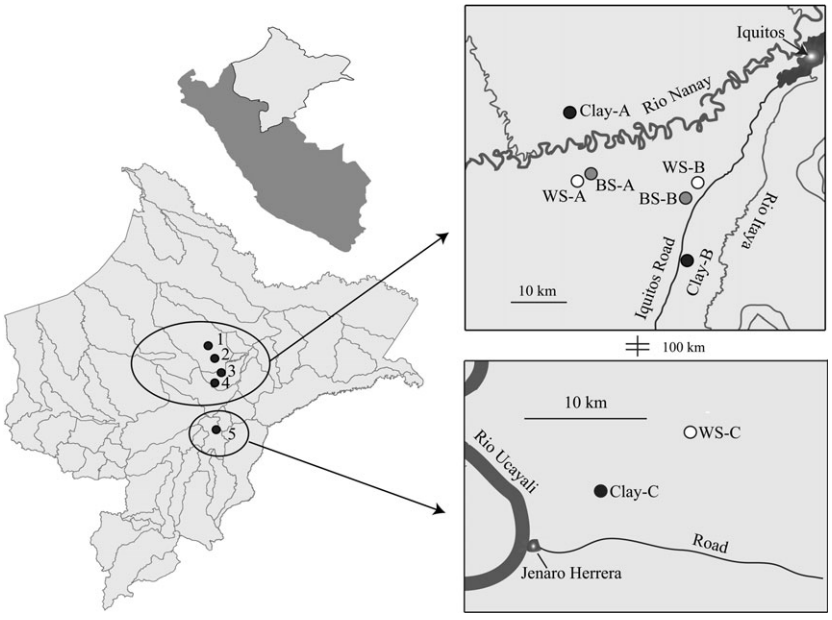

Fig. 1 Sample sites and soil types for populations of *P. subserratum* in the region of Loreto, Peru. Numbered points represent the five sites where populations were found. Each individual population is displayed in the inset. Black circles represent populations found on clay soil, grey circles represent populations found on brown-sand soil, and white circles represent populations found on white-sand soils.

collected in silica for DNA extraction ($N = 201$). Voucher specimens for each population were deposited in the Herbarium Amazonense at the Universidad Nacional de la Amazonía Peruana in Iquitos, Peru (AMAZ), and the University Herbarium at the University of California, Berkeley (UC). To avoid variation in leaf morphology based on age and canopy position, samples used for morphological measurements were only collected from individuals 10 m or taller ($N = 163$). Three leaves were collected from each individual, pressed and dried for later processing. All of these individuals were also included in the genetic sampling.

Characterization of phenotypic variation

In order to investigate the extent to which phenotypic differences in leaf morphology were correlated with white-sand, brown-sand and clay soil habitats, we characterized variation in eight leaf morphological characters from 163 adult individuals across all eight populations. Measurements were taken three times and averaged. They included leaf length, number of leaflets per leaf, leaflet length, leaflet width, leaflet thickness, number of margin serrations per leaflet, pubescence percentage coverage on the abaxial side of the leaflet blade and the leaflet midrib. Pubescence percentage coverage of abaxial leaflet blade and leaflet midrib was visually estimated using a dissecting microscope to the nearest 10% within a haphazardly placed three by three millimetre square.

Morphological differentiation among individuals found on white-sand, brown-sand and clay soil was assessed in R (R Development Core Team 2008) using principal components analysis (PCA) and multivariate analysis of variance (MANOVA).

Microsatellite genotyping

Genomic DNA was extracted from the leaf material of all adult individuals, and extractions were carried out using a Qiagen DNeasy Plant Mini Kit (Valencia, CA, USA). Genotypes were determined using thirteen nuclear microsatellite markers developed for *P. subserratum* (prot13, prot28, prot29, prot67, prot70, prot78, prot83, prot97, prot99, prot100, prot101, prot102 and prot104), following the protocols described in Misiewicz *et al.* (2012).

Genetic variation, Hardy–Weinberg equilibrium, null alleles and linkage disequilibrium

Summary statistics including number of alleles (A) and observed, expected and unbiased expected heterozygosities (Ho He and UHe) were estimated for each population using GenAlEx version 6.4 (Peakall & Smouse 2006). We used all loci equalized to a sample size of five individuals, the number of individuals in our smallest population, to calculate rarefied allelic richness (A_R) and private allele richness (A_P) using the allelic diversity analyzer ADZE (Szpiech *et al.* 2008). Values were calculated as averages across all loci for each population, and additional A_P values were calculated for combinations of populations grouped by soil type. The inbreeding coefficient (F_{IS}) for each population was calculated across all loci, as were deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD), using GENEPOP version 4.0 (Raymond & Rousset 1995). Deviations from HWE and LD were calculated using the probability test with 10^4 demorizations, 10^4 batches and 10^4 iterations per batch. Significance values were adjusted using sequential Bonferroni corrections (Rice 1989). All data were screened for genotyping errors due to stutter, large allele dropout and the presence of null alleles, using the software Micro-Checker (van Oosterhout *et al.* 2004).

Population genetic structure and differentiation across soil types

If natural selection is more important than genetic drift in driving population differentiation, then population genetic structure should be more strongly correlated with turnover in habitat type than with geographical distance. Hence, we sought to understand how populations are genetically structured using a combination of descriptive statistics including *F*-statistics, analysis of molecular variance (AMOVA) and Bayesian clustering analysis. Additionally, we explicitly explored the extent to which genetic variation among populations could be explained by geographical distance and soil type using Mantel and partial Mantel tests. Genetic differentiation among populations was assessed using population-level pairwise comparisons of θ, analogous to Wright's F_{ST} (Weir & Cockerham 1984), calculated with and without correction for null alleles with FreeNA (Chapuis & Estoup 2007). Analysis of molecular variance (AMOVA), in which populations are grouped hierarchically to explore how groupings affect the partitioning of genetic variation (Excoffier *et al.* 1992), was calculated among populations grouped according to (i) soil type (white-sand, brown-sand and clay), (ii) brown-sand and clay soils (excluding white-sand populations) and (iii) geographical location (sites 1, 2, 3, 4 and 5) using 10^3 permutations.

Population genetic structuring was determined using Bayesian Markov chain Monte Carlo clustering analysis implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000).

STRUCTURE was run using the admixture model and assuming uncorrelated allele frequencies. In order to estimate the 'true' K, a burn-in period of 500 000 generations was followed by 10^5 Markov chain Monte Carlo generations for each value of $K = 1–8$. Simulations were repeated twenty times for each value of K. We used Structure Harvester (Earl & vonHoldt 2012) to interpret the output as described by Evanno *et al.* (2005) and Pritchard *et al.* (2000). Admixture proportions were averaged over all runs using CLUMPP (Jakobsson & Rosenberg 2007), and the final matrix was visualized with DISTRUCT 1.1 (Rosenberg 2004).

Mantel and partial Mantel tests were performed in the program IBDWS version 3.15 (Jensen *et al.* 2005), using 10 000 randomizations. Tests were performed for all populations, brown-sand and clay populations only and populations grouped by soil type. All partial Mantel tests were performed for comparisons of genetic distance and habitat type while controlling for geographical distance.

Selection vs. drift

We assessed the potential role of divergent natural selection and genetic drift as mechanisms underlying population differentiation by comparing phenotypic differentiation and neutral genetic differentiation for groupings of individuals across populations based on soil type. While these statistical tests are generally performed using an index of divergence for genes linked to quantitative traits (Q_{ST}), the degree of phenotypic divergence (P_{ST}) can be used as a surrogate when Q_{ST} cannot be estimated (Leinonen *et al.* 2006). If the value of P_{ST} significantly exceeds F_{ST}, then a hypothesis of divergent selection is supported, whereas P_{ST} and F_{ST} values that do not differ significantly are consistent with neutral genetic differentiation (Merilä & Crnokrak 2001).

P_{ST} values were calculated as described by Leinonen *et al.* (2006). Analyses of variance (ANOVA) were used to estimate the components of phenotypic variances, and calculations were carried out with heritability (h^2) defined as 0.5, where half of phenotypic variation is due to environmental and nonadditive effects. Because all morphological traits measured in this study were leaf traits, and therefore nonindependent, P_{ST} values for morphology were estimated from our first three principal components (PC). Confidence intervals (CI) were estimated using 1000 bootstrap replicates over individuals in R (http://www.r-project.org/). P_{ST} values and multilocus FST values were calculated and compared for individuals grouped by soil type. Comparisons were considered significantly different when their 95% CI did not overlap.

Hybridization and introgression

We tested for hybridization and introgression between edaphically differentiated population pairs of *P. subserratum* that contained individuals showing significant admixture in our population structure analysis with the model-based clustering method of NewHybrid version 1.1b (Anderson & Thompson 2002). We restricted our analysis to two generations; this resulted in six possible hybrid classes (two parental classes, F_1 and F_2 hybrids and two backcrosses between F_1 hybrids and the parental types). Prior information about parental species from the Structure results was included in the analysis. All data were analysed multiple times from overdispersed starting values, over 2^6 sweeps after a burn-in of 500 000 sweeps, following the recommendations of Anderson (2002). All simulations performed equally well in the discrimination of individuals based on soil type and resulted in congruent estimations of posterior probabilities for individual hybrid class assignments. Individuals assigned to a hybrid class with a posterior probability (pp) >90% were considered to be hybrids with parental contributions from individuals growing in different soil types.

Demographics

Migration rates as low as one migrant per generation can swamp population-level differentiation (Wright 1931). Therefore, migration rates were estimated between all population pairs in order to further explore the effect of soil type and geography in limiting migration rates among populations. Migration among populations was assessed using Migrate-N, which samples coalescent genealogies, in order to calculate maximum-likelihood estimates of past migration rates (Beerli 1998; Beerli & Felsenstein 1999, 2001). Migration rate $(M = \theta m/\mu)$ was estimated for all population pairs. All migration rates were initially calculated between pairs of populations under a model where parameter θ, the ratio between immigration rate and mutation rate per generation, is fixed and the parameter for migration rate M is variable in order to allow for asymmetric bias in migration between populations. A maximum of 20 individuals were randomly sampled from each population for analysis with Migrate-N. All analyses were executed under a Brownian mutation model and relative mutation rates were estimated from the data, with starting values for the M and θ estimated from F_{ST} values. We tested multiple chain lengths, replicates and heating schemes to test for convergence across runs. Final analyses consisted of 10 short chains and three long chains; the short chains ran for 50 000 generations with a sampling increment of 100 generations, and the long

chains ran for 500 000 generations with a sampling increment of 100 generations. A total of 5 million genealogies were visited over the short chain runs and 50 million genealogies were visited over long chain runs. No heating scheme was used and the burn-in was set to 250 000. All analyses were then repeated as described above using the maximum-likelihood estimates of parameters θ and M from the previous run as the new parameter start values. Finally, we repeated analyses for a subset of population pairs (WS-A and BS-A; WS-B and BS-B; Clay-A and Clay-C; and WS-A and WS-C) excluding locus Prot100, which tested for null alleles in more than two populations to assess its potential impact on demographic estimates. Gelman's R statistic was used to assess convergence (Beerli & Felsenstein 2001).

Results

Phenotypic variation

The first three principal components accounted for 82% of the variation in the data and were strongly associated with soil type (Fig 2). As expected from previous studies, pubescence on the leaflet midvein and blade, entire leaf margins and thicker leaflets differentiated individuals found on white-sand soil from individuals found on brown-sand and clay soils (Fine *et al*. 2013a). Furthermore, we detected previously undescribed differences between individuals found on clay and brown-sand soils where the abundance of leaflet margin serrations and the number of leaflets per leaf explained most of the variation. Individuals identified as putative older generational hybrids in hybrid assignment analyses did not reflect intermediate phenotypes with regard to leaf morphology.

Analysis of leaf traits with MANOVA showed that leaflet thickness, number of margin serrations per leaflet, pubescence percentage coverage on the abaxial side of the leaflet blade and pubescence percentage coverage on the abaxial side of the leaflet midrib varied significantly among individuals found on all three soil types ($P < 0.05$).

Genetic variation, Hardy–Weinberg equilibrium, null alleles and linkage disequilibrium

All summary statistics for microsatellite data are listed in Table 1. Deviations from HWE were observed in 10 of the 104 total within-population comparisons for each locus, and no particular locus or population showed a consistent pattern of deviation from HWE. We found one pair of loci (prot83 and prot100) in linkage disequilibrium (LD) in one population (WS-C); no loci were found to be in LD in any other population.

Micro-Checker did not detect any genotyping errors due to stutter or large allele dropout. Evidence for the presence of null alleles was found in all populations except Clay-B. No locus consistently showed signs of null alleles across populations. Four populations tested positive for null alleles at one locus (WS-B-prot100, BS-B-prot83, WS-C-prot70 and Clay-C-prot101). Populations Clay-A and BS-B tested positive for null alleles at three loci (Prot99, Prot100, Prot102 and Prot83, Prot99, Prot100, respectively) and population WS-B tested positive at four loci (Prot28, Prot78, Prot100 and Prot102).

Genetic structure and ecotypic variation

Pairwise F_{ST} values revealed strong genetic differentiation across soil type and values differed only slightly when calculated with and without null allele corrections

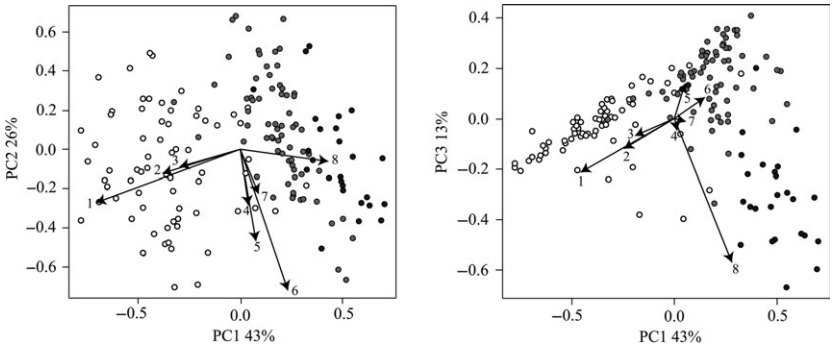

Fig. 2 Principal components analysis (PCA) of leaf morphological characters for *P. subserratum* individuals. White, grey and black circles represent individuals found in white-sand, brown-sand and clay soil habitats, respectively. Axes one through eight, respectively, represent the following characters: percentage pubescence coverage on leaflet midrib, percentage pubescence coverage on abaxial side of leaflet blade, leaflet thickness, leaflet width, number of leaflets per leaf, leaf length, leaflet length and number of margin serrations per leaflet.

Table 1 Populations sampled, region found, soil type, UTM coordinates for each population, number of individuals sampled (*N*), observed heterozygosity (H_O), expected heterozygosity (He), unbiased expected heterozygosity (UHe), average number of alleles (A), rarefied average number of private alleles (A_P), rarefied average number of alleles (A_R) and inbreeding coefficient (F_{IS})

Population	Site	Soil Type	Latitude	Longitude	*N*	Ho	He	UHe	A	A_P	A_R	F_{IS}
Clay-A	1	Clay	−3.830323320	−73.594676100	15	0.05	0.60	0.58	5.3	0.43	3.27	0.08
WS-A	2	White-sand	−3.917752825	−73.551974793	54	0.53	0.56	0.48	5.9	0.14	2.71	0.16
BS-A	2	Brown-sand	−3.909758490	−73.552303691	45	0.44	0.48	0.57	6.0	0.23	3.18	0.01
WS-B	3	White-sand	−3.950526640	−73.408218868	19	0.41	0.43	0.44	3.2	0.13	2.39	0.05
BS-B	3	Brown-sand	−3.976726271	−73.427493203	29	0.51	0.51	0.53	5.2	0.24	2.96	0.05
Clay-B	4	Clay	−4.058777584	−73.432082104	13	0.66	0.65	0.61	5.6	0.46	3.59	0.01
WS-C	5	White-sand	−4.864078785	−73.615967787	21	0.31	0.35	0.36	2.8	0.26	2.03	0.10
Clay-C	5	Clay	−4.887118050	−73.649012166	5	0.61	0.61	0.67	4.5	0.69	4.12	0.11

(Table 2). Accordingly, all subsequent analyses were carried out on the full data set.

AMOVA results confirmed that the majority of genetic variation is explained by soil type and not by geographical locality. Even when white-sand populations were excluded from the analysis, soil type continued to be important in explaining genetic variation among populations found on brown-sand and clay soils (Table 3).

Population structure analysis revealed strong patterns of genetic differentiation by soil type. Using the method of Evanno *et al.* (2005), the most likely model contained two clusters, consistent with populations found on white-sand or non-white-sand. Using the method of Pritchard *et al.* (2000), we found $K = 3$ and $K = 4$ to be the best-supported models, where LnP(D) begins to asymptote (Figs 3 and 4). The identified clusters for $K = 3$ are consistent with populations found on white-sand, brown-sand and clay soils. Clustering for $K = 4$ remained consistent with clustering by soil type but also showed patterns of isolation by distance within soil types with population WS-C distinguished from WS-A and WS-B found 100 km away.

Structure analysis also revealed individuals with genetic contributions from multiple soil types. When $K = 2$, four individuals showed admixture between white-sand and non-white-sand clusters in populations WS-A and BS-A at site 2. When $K = 3$, the same four individuals with admixed genomes remained consistent, exhibiting admixture between white-sand and brown-sand populations and an additional individual found in population Clay-B contained admixture from both clay and brown-sand clusters. No additional individuals were identified as having significant admixture in the model where $K = 4$ (Fig. 3).

The Mantel test comparing all population pairs showed no correlation between genetic distance and geographical distance among populations ($r = 0.00$, $P > 0.1$) and a strong positive correlation between genetic distance and habitat type ($r = 0.72$, $P < 0.01$). When controlling for geographical distance in partial Mantel analysis, genetic distance remained positively associated with habitat type ($r = 0.72$, $P < 0.01$). When comparing only brown-sand and clay soil populations in the analysis, the Mantel test still showed no significant correlation between genetic distance and geographical distance between populations ($r = 0.14$, $P > 0.1$) and a positive correlation between genetic distance and habitat type ($r = 0.60$, $P < 0.01$). When controlling for geographical distance in the partial Mantel test, genetic

Table 2 Pairwise F_{ST} values for all population pairs. Values below the diagonal are estimated without using corrections for null alleles. Values above the diagonal are estimated using corrections for null alleles

Pop.	Clay-A	WS-A	BS-A	WS-B	BS-B	Clay-B	WS-C	Clay-C
Clay-A	0	0.315	0.266	0.336	0.282	0.150	0.398	0.159
WS-A	0.334	0	0.340	0.037	0.360	0.295	0.187	0.279
BS-A	0.277	0.358	0	0.367	0.022	0.234	0.411	0.233
WS-B	0.341	0.035	0.377	0	0.389	0.307	0.176	0.298
BS-B	0.295	0.379	0.024	0.400	0	0.249	0.445	0.245
Clay-B	0.155	0.313	0.242	0.317	0.258	0	0.339	0.107
WS-C	0.408	0.186	0.426	0.173	0.460	0.353	0	0.377
Clay-C	0.151	0.290	0.238	0.296	0.250	0.107	0.382	0

Table 3 Analysis of molecular variance (AMOVA) for populations of *Protium subserratum* grouped by geography (sites 1, 2, 3, 4 and 5), by soil type (white-sand, brown-sand and clay soil) and by soil type excluding white-sand (brown-sand and clay soil)

AMOVA	d.f.	Sum of squares	Variance of components	% explained
Soil type (white-sand, brown-sand and clay)				
Among groups	2	539.66	1.07*	25.74
Among populations, within groups	5	96.62	0.31*	7.34
Within populations	722	2016.23	2.79*	66.91
Soil type (brown-sand and clay)				
Among groups	1	115.13	0.83	19.93
Among populations, within groups	3	53.04	0.32*	7.68
Within populations	377	1138.52	3.02*	72.39
Geographical location				
Among groups	4	173.91	−0.36	−9.45
Among populations, within groups	3	462.36	1.39*	36.38
Within populations	722	2016.23	2.79*	73.06

*$P < 0.01$.

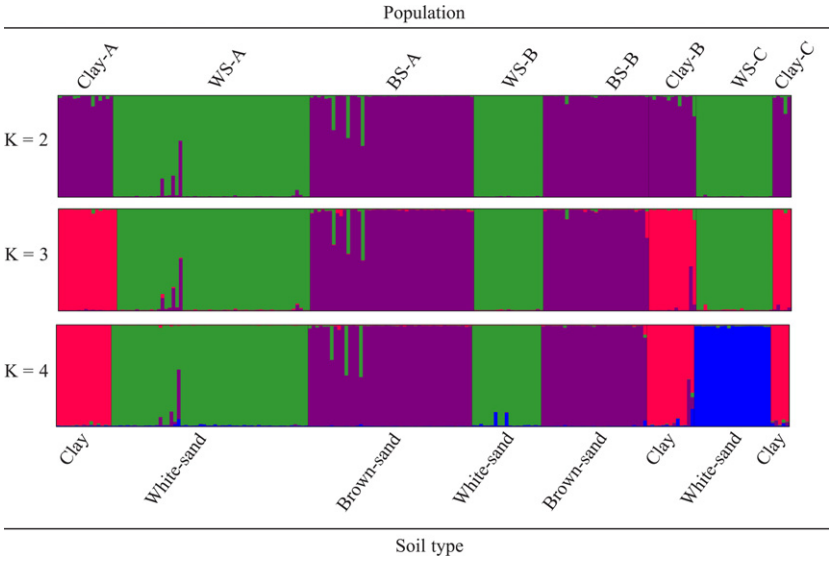

Fig. 3 Evolutionary clusters (K = 2–4) inferred from STRUCTURE analysis of 201 *P. subserratum* individuals from populations of white-sand, brown-sand and clay soil habitats. Each colour represents an inferred character, and each individual is represented by a vertical line shaded according to its probability of assignment to a given population.

distance remained positively associated with habitat type ($r = 0.65$, $P = 0.1$). Mantel tests comparing only populations found on the same soil type showed strong correlations between genetic distance and geographical distance (clay populations: $r = 0.84$, $P = 0.33$; white-sand populations: $r = 0.98$, $P = 0.16$). The insignificant p-values are most likely being driven by the small number of comparisons ($n = 3$, d.f. = 2 for white-sand and clay populations).

Selection vs. drift

With the exception of PC2 in the white-sand and clay comparison, all comparisons of phenotypic differentiation and neutral genetic variation (P_{ST}-F_{ST}) among soil types demonstrated much higher P_{ST} values compared to F_{ST} values, supporting a hypothesis of divergent natural selection for all comparisons except for PC2 for comparison between white-sand and clay populations (Fig. 5).

Hybridization and introgression

Results from the analysis of hybridization and introgression using NewHybrid corresponded with the Structure results and demonstrated low levels of introgression among populations found on different soil types. In the analysis of populations WS-A and BS-A,

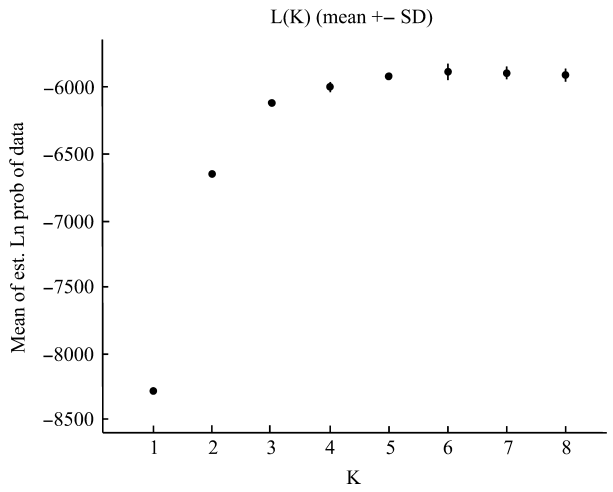

Fig. 4 Average log probabilities for K = 1–8.

NewHybrid assigned 52 individuals as pure parental white-sand individuals and 42 individuals as pure parental non-white-sand individuals. One individual could not be classified with confidence as white-sand parental or F_2 hybrid (PP = 0.80 and PP = 0.20, respectively). The four individuals, identified as admixed in Structure, were assigned with confidence as F_2 hybrids (PP > 0.9). No individuals were identified as F_1 hybrids.

Analysis between population pair Clay-B and BS-B also corroborated our Structure results where all individuals were identified as parental types with the exception of the one admixed individual. NewHybrid confidently assigned this individual as having a hybrid origin however, was unable to distinguish it as an F1 hybrid or F2 hybrid (PP = 0.75 and PP = 0.13).

Population demographics

Exclusion of locus Prot100 did not substantially change migration estimates, and thus, we have reported results from our analysis using the complete data set. Overall, inferred historical migration rates were consistently higher among populations found on the same soil type than among populations found on different soil types even when those populations were over 100 km apart (Table 4, Fig. 6a,b). We found that migration rates were higher between brown-sand and clay populations than among populations found on either soil type or white-sand populations. Furthermore, we found higher migration rates from white-sand populations into brown-sand and clay populations than from either non-white-sand population into white-sand habitat (Table 4). Overall migration rates of less than one migrant per generation were observed when M values were averaged among populations found on different soil types, and a rate of greater than one migrant per generation was observed when M values were averaged among populations found on the same soil type (Fig. 6b).

Discussion

Our analysis of population-level genetic and morphological differentiation identified three clearly diverged groups in *P. subserratum*, each associated with white-sand, brown-sand or clay soil habitats. Additionally, our results suggest that *P. subserratum* populations

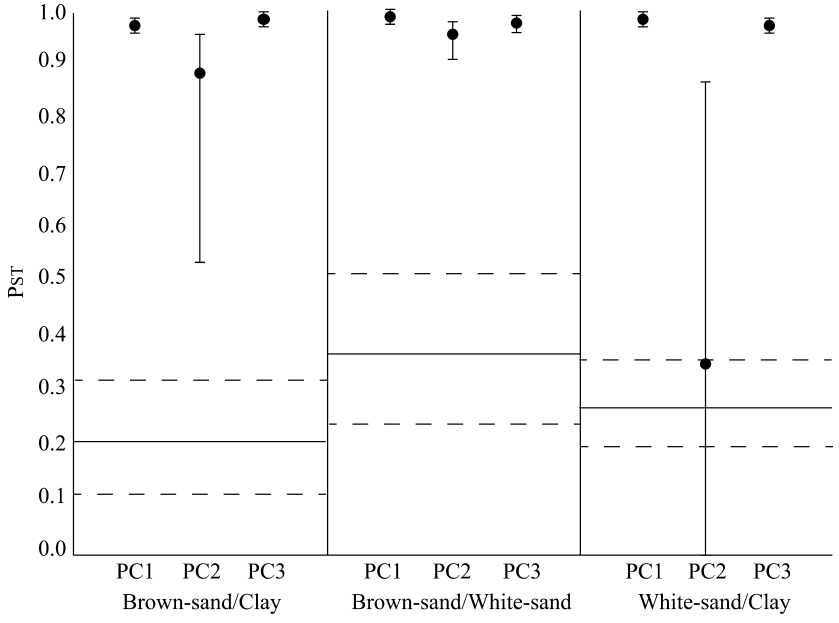

Fig. 5 Comparison of P_{ST} and F_{ST} values and their 95% confidence intervals. P_{ST} values were calculated from principal components for individuals found on clay, white-sand and brown-sand soil types. Solid horizontal lines represent the multilocus F_{ST} value, and dotted lines represent the 95% CI of the F_{ST} value.

Table 4 Maximum-likelihood (ML) migration (*M*) estimates and their 95% confidence intervals for all population pairs

Source population	Recipient population	M MLE	M 5%	M 95%
Clay-A	Clay-B	0.58	0.47	0.72
Clay-C	Clay-B	1.57	1.32	1.81
WS-A	Clay-B	0.22	0.17	0.29
WS-B	Clay-B	0.24	0.19	0.31
WS-C	Clay-B	0.43	0.35	0.52
BS-A	Clay-B	0.38	0.31	0.46
BS-B	Clay-B	1.20	1.05	1.36
Clay-B	Clay-A	1.44	1.22	1.68
Clay-C	Clay-A	3.30	2.92	3.71
WS-A	Clay-A	0.17	0.12	0.23
WS-B	Clay-A	0.80	0.67	0.95
WS-C	Clay-A	0.38	0.31	0.46
BS-A	Clay-A	1.7	1.49	1.94
BS-B	Clay-A	0.74	0.62	0.88
Clay-A	Clay-C	1.36	1.80	1.57
Clay-B	Clay-C	2.30	2.03	2.57
WS-A	Clay-C	0.26	0.20	0.34
WS-B	Clay-C	0.29	0.23	0.36
WS-C	Clay-C	0.24	0.16	0.31
BS-A	Clay-C	0.68	0.59	0.79
BS-B	Clay-C	0.43	0.35	0.53
Clay-A	WS-A	0.12	0.08	0.17
Clay-B	WS-A	0.06	0.04	0.10
Clay-C	WS-A	0.57	0.43	0.74
WS-B	WS-A	3.98	3.63	4.38
WS-C	WS-A	3.19	2.92	3.48
BS-A	WS-A	0.13	0.09	0.18
BS-B	WS-A	0.18	0.12	0.25
Clay-A	WS-B	0.35	0.26	0.46
Clay-B	WS-B	0.07	0.04	0.13
Clay-C	WS-B	0.07	0.04	0.12
WS-C	WS-B	3.73	3.31	4.20
WS-A	WS-B	5.37	4.71	6.10
BS-A	WS-B	0.16	0.11	0.23
BS-B	WS-B	0.07	0.04	0.11
Clay-A	WS-C	0.95	0.74	1.19
Clay-B	WS-C	0.40	0.28	0.54
Clay-C	WS-C	0.14	0.08	0.23
WS-A	WS-C	1.90	1.63	2.21
WS-B	WS-C	1.81	1.44	2.28
BS-A	WS-C	0.06	0.03	0.11
BS-B	WS-C	0.08	0.04	0.15
Clay-A	BS-A	1.27	1.09	1.45
Clay-B	BS-A	0.68	0.56	0.84
Clay-C	BS-A	1.97	1.74	2.25
WS-A	BS-A	0.73	0.61	0.88
WS-B	BS-A	0.55	0.45	0.67
WS-C	BS-A	0.26	0.19	0.34
BS-B	BS-A	5.12	4.74	5.51
Clay-A	BS-B	0.42	0.33	0.53
Clay-B	BS-B	1.20	1.03	1.40
Clay-C	BS-B	0.99	0.80	1.21
WS-A	BS-B	0.46	0.37	0.57
WS-B	BS-B	0.47	0.39	0.57
WS-C	BS-B	0.28	0.19	0.39
BS-A	BS-B	3.65	3.30	4.00

found on brown-sand and clay soils are experiencing a significant, and likely more recent, ecological divergence. We detected morphological differentiation among individuals found on white-sand, brown-sand and clay soils. Populations of *P. subserratum* across all three soil types were found to be more genetically similar to geographically distant populations found on the same soil type than to nearby populations found on different soil types. The average degree of phenotypic variation was much greater than the overall degree of neutral genetic differentiation for all comparisons among soil types, suggesting that natural selection may play a more important role than drift in driving divergence among these populations. A small number of hybrid individuals were detected between brown-sand and white-sand groups and between clay and brown-sand groups, suggesting that gene flow among populations on different soil types does occur at low frequency. Finally, migration rates were found to be higher between geographically distant populations found on the same soil type than they were between adjacent populations on different soil types. Taken together, we observe a signature of fine-scale ecological specialization across multiple soil boundaries. While we cannot discern whether initial divergence took place in allopatry or parapatry, our results provide evidence consistent with the hypothesis that natural selection plays an important role in maintaining the diversity in parapatric populations within the *P. subserratum* species complex.

Population differentiation across habitat types

We found high levels of genetic differentiation among populations found on all three soil types, even when those populations were directly adjacent to one another, presenting a clear pattern of 'isolation by adaptation'. Similar results across a wide range of taxa including passerines (Smith *et al.* 1997), anolis lizards (Ogden & Thorpe 2002), stickleback fish (Berner *et al.* 2009) and dolphins (Mendez *et al.* 2010) have been used to support the idea that natural selection is an important driver of diversification.

While Fine *et al.* (2013a) did not detect genetic or significant morphological differentiation between individuals found on brown-sand and clay soil, our results were

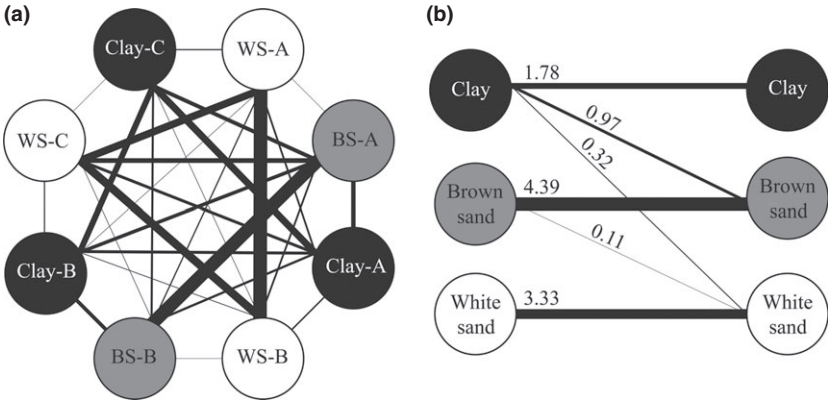

Fig. 6 Gene flow (as migrants per generation, $M = \theta m/\mu$) estimated in Migrate-N. (a) ML migration rates between population pairs were averaged for one overall estimate to which lines between population pairs are proportional. (b) ML estimates of migration across all soil types were averaged. Lines are proportional to the average migration rate between populations found on each soil type.

consistent with their overall findings that populations associated with white-sand habitats are distinct from those found in non-white-sand habitats and that soil type as opposed to geographical distance is more important in the structuring of genetic variation between white-sand and non-white-sand populations. However, estimates of genetic diversity within populations strongly differed between the two studies. Fine *et al.* (2013a) reported higher haplotypic and genetic diversity in white-sand populations than in clay and brown-sand populations. Here, we found the opposite pattern, with white-sand populations exhibiting lower levels of allelic diversity when compared to clay and brown-sand populations. Our measures of lower genetic diversity in white-sand populations may be expected given the small size and fragmented nature of white-sand habitat islands; higher levels of allelic diversity in brown-sand and clay populations accord well with the wide distribution of these more common soil types.

An explanation for the discrepancy in measurements of genetic diversity between these two studies could be that nuclear DNA sequences and nuclear microsatellite markers represent different temporal depths in the evolutionary history of these diverging groups. Microsatellite markers, with fast rates of evolution, represent more recent evolutionary events, while nuclear DNA sequences, exhibiting lower rates of polymorphism, provide insights into evolutionary events that took place deeper in time. As seen in the Hawaiian silversword alliance, when demographic factors change over time, this variation may be reflected in the genetic signatures of different molecular markers (Friar *et al.* 2007; Remington & Robichaux 2007 and discussed in Lawton-Rauh *et al.* 2007).

One possible demographic hypothesis that reconciles the results of these two studies and which may also shed light on the divergence histories of these populations relates to the geologic history of the Amazon. Prior to the Andean uplift, white-sand soils in the western Amazon were likely much more widespread than they are today, while brown-sand and clay soil types were less common (Hoorn 1993; Frasier *et al.* 2008). If white-sand specialist populations of *P. subserratum* were previously larger and brown-sand and clay populations were much smaller, the pattern of genetic diversity would correspond to those observed in nuclear sequence data. Microsatellite data, meanwhile, would reflect patterns of genetic diversity under the more recent history of the extent of these soil types. This hypothesis would also make sense with what appears to be a primary diversification from white-sand soils onto brown-sand soils with a secondary and more recent diversification from brown-sand soils onto clay soils, a pattern previously undetected by sequence data by Fine *et al.* (2013a).

Hybridization and introgression

While low levels of interspecific gene flow were observed between edaphically differentiated populations, we found no evidence for substantial admixture between any edaphically differentiated population pairs. Low levels of introgression were observed between population pairs of *P. subserratum* found on white-sand and brown-sand soils and brown-sand and clay soils. This pattern is consistent with that of a bimodal hybrid zone which is characterized by parental forms that retain their genetic integrity in spite of gene flow and hybrid zones that are composed of F_2 and backcross hybrids but very few intermediate F_1 hybrids. Bimodal hybrid zones are often attributed to the presence of premating barriers to reproduction in the form of assortative mating (Jiggins & Mallet 2000); they have been particularly well described in *Heliconius* butterflies (Arias *et al.* 2008), ground crickets (Howard *et al.* 1998) and sunflowers (Rieseberg *et al.* 1999). Alternatively, bimodal hybrid zones can also be maintained by particularly strong natural selection if F_1 hybrids fail to survive and reproduce in alternative habitats (Arnold & Bennett 1993; Ross & Harrison 2002).

Edaphic heterogeneity and reproductive isolation

Edaphic differences play an important role in driving the development of both pre- and postzygotic barriers to reproduction among adaptively differentiated plant populations. Changes in flowering phenology are associated with soil preference in two parapatric species of palms on Lord Howe Island (Savolainen *et al.* 2006). High rates of seed abortion were demonstrated between copper-tolerant and intolerant populations of *Mimulus guttatus* (Searcy & Macnair 1990), and strong natural selection against migrants was demonstrated between species of *Lasthenia* growing on and off serpentine soils (Yost *et al.* 2012).

Prezygotic barriers to reproduction may be present in soil specialist populations of *P. subserratum*. While flowering times do not differ among populations found on different soil types (Misiewicz 2014), edaphic differences may still play an indirect role in limiting pollen movement among populations. Initial pollinator surveys suggest that bee communities found in nearby brown-sand and white-sand forest types significantly differ from one another (Misiewicz 2014). If insect pollinators exhibit habitat preferences, they may indirectly decrease gene flow across habitat boundaries.

As was the case for population differentiation across habitat types, migration rates were higher between distant populations occurring on the same soil type than between adjacent populations found on different soil types. Particularly noteworthy is the result that when migration rates were averaged across pairs of edaphically differentiated populations, the values are less than one. Theory predicts that when values are lower than one, populations should be strongly differentiated. Migration rates averaged across population pairs found on the same soil type are greater than one, the point at which migration is predicted to overcome population-level differentiation (Fig. 6b; Zhang *et al.* 2011). However, these results should be interpreted with caution. Demographic models implemented in Migrate do not take into account the possibility of retained ancestral polymorphism and assume that all genetic similarity among populations is the result of gene flow. While all migration rates among differing soil types were low, we found that rates of migration from white-sand populations into clay and brown-sand populations and from brown-sand populations into clay populations were higher than in the reverse direction. Patterns of asymmetric gene flow among edaphically differentiated populations could be a reflection of earlier divergence if ancestral alleles are retained in the more recently diverged populations. Alternatively, the same pattern could be the result of asymmetric strength of reproductive barriers.

Edaphically driven postzygotic barriers to reproduction may also play a role in reducing successful gene flow among soil specialist populations of *P. subserratum*. White-sand, brown-sand and clay soil habitats represent a mosaic of microclimatic differences, and the correlated variation in *P. subserratum* vegetative morphology with soil habitat type may reflect adaption to different selective regimes. Moreover, herbivore communities differ markedly between non-white-sand and white-sand habitats, as does the defensive chemistry *P. subserratum* individuals associated with these habitat types (Fine *et al.* 2013b). In the genus *Protium*, trade-offs in growth rate and defence investment have been found in closely related sister species specialized on different soil types (Fine *et al.* 2006). Physiological trade-offs between growth and defence in soil specialist populations of *P. subserratum* could lead to strong natural selection against phenotypic intermediates in both soil types, even when the potential for gene flow among habitats is high.

Divergence by drift

Our results suggest that natural selection across heterogeneous soil types plays an important role in driving and maintaining diversity in soil specialist populations of *P. subserratum*. However, high levels of neutral genetic differentiation among habitats and low levels of gene flow are consistent with the late stages of speciation and soil specialist populations could represent cryptic species, which diverged in allopatry and have since come into secondary contact. Furthermore, a hypothesis of initial divergence by genetic drift cannot be excluded. While specialization onto different soil types has clearly occurred multiple times within the species, we cannot disentangle the initial drivers of reproductive isolation without a more complete understanding of the phylogeographical history of the species. Nevertheless, it is worth noting that there are dozens of tree species that are endemic to white-sand forests in Peru with congeners that are associated with parapatric clay and/or brown-sand habitats (Fine *et al.* 2010). Some of these species belong to extremely recently derived groups like *Inga* (Fabaceae) (Richardson *et al.* 2001), while others are thought to be much older (*Mauritia*, Arecaceae) (Couvreur *et al.* 2011). Although it cannot be ruled out, a hypothesis of allopatric divergence by drift with subsequent edaphic specialization followed by migration to the current parapatric distributions for so many tropical tree lineages seems less parsimonious. In either case, we believe that edaphic specialization appears to be a general mechanism that promotes and maintains Amazonian tree diversity.

Acknowledgements

T.M.M. and P.V.A.F. gratefully acknowledge the Ministerio del Ambiente of Peru for providing research and export permits, Carlos Rivera of SERNANP-Allpahuayo-Mishana and the Instituto de Investigaciones de la Amazonía Peruana (IIAP) for institutional support. This research was financially supported by National Science Foundation Grants DEB-0909567 (P.V.A.F.) and DEB-1311117 (T.M.M and P.V.A.F.) as well as the Junior Faculty Research Grant funded by the Academic Senate of UC Berkeley. The authors thank Julio Sanchez and Italo Mesones for assistance in the field. The manuscript benefited greatly from the insightful comments of three anonymous reviewers, Benjamin Carter, Seth Kauppinen and Sonal Singhal.

References

Álvarez Alonso J, Metz MR, Fine PVA (2013) Habitat specialization by birds in western Amazonian white-sand forests. *Biotropica*, **45**, 365–372.

Anacker B, Whittal J, Goldberg E, Harrison S (2011) Origins and consequences of serpentine endemism in the California flora. *Evolution*, **65**, 365–376.

Anderson EC (2002) *User's guide to the program NewHybrids version 1.0*. Available at: http://ib.berkeley.edu/labs/slatkin/eriq/software/new_hybs_doc.pdf.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Antonovics J, Bradshaw AD (1970) Evolution in closely adjacent plant populations VIII. Clinal patterns at a mine boundary. *Heredity*, **25**, 349–362.

Arias CF, Munoz AG, Jiggins CD, Mavarez J, Bermingham E, Linares M (2008) A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. *Molecular Ecology*, **17**, 4699–4712.

Arnold ML, Bennett BD (1993) Natural hybridization in Louisiana irises: genetic variation and ecological determinants. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 115–139. Oxford University Press, New York.

Beerli P (1998) Estimation of migration rates and population sizes in geographically structured populations. In: *In Advances in Molecular Ecology* (ed. Carvalho G), pp. 39–53. NATO-ASI workshop series. IOS Press, Amsterdam.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum-likelihood estimation of migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, **98**, 4563–4568.

Berner D, Grandchamp AC, Hendry AP (2009) Variable progress toward ecological speciation in parapatry:stickleback across eight lake-stream transitions. *Evolution*, **63**, 1740–1753.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Couvreur TLP, Forest F, Baker WJ (2011) Origin and global diversification patterns of tropical forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology*, **9**, 44.

Daly DC (1987) A taxonomic revision of *Protium* (Burseraceae) in eastern Amazonia and the Guianas. PhD Dissertation, City University of New York, New York.

Daly DC, Fine PVA (2011) An ecologically significant new section of *Protium* (Burseraceae) from Amazonia. Studies in neotropical Burseraceae XVI. *Systematic Botany*, **36**, 939–949.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Ehrlich PR, Raven PH (1969) Differentiation of populations. *Science*, **165**, 1228–1232.

Endler JA (1977) *Geographic variation, clines, and speciation*. Princeton University Press, Princeton, New Jersey.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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.

Fine PVA, Mesones I, Coley PD (2004) Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, **305**, 663–665.

Fine PVA, Daly DC, Munoz GV, Mesones I, Cameron KM (2005) The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution*, **59**, 1464–1478.

Fine PVA, Miller ZJ, Mesones I *et al.* (2006) The growth-defense trade-off and habitat specialization in plants in Amazonian forests. *Ecology*, **87**, S150–S162.

Fine PVA, Garcia-Villacorta R, Pitman NCA, Mesones I, Kembel SW (2010) A floristic study of the white-sand forests of Peru. *Annals of the Missouri Botanical Garden*, **97**, 283–305.

Fine PVA, Zapata F, Daly DC, Mesones I, Misiewicz TM, Cooper HF (2013a) Phylogeography of edaphic specialist and generalist species of *Protium* (Burseraceae): the relative importance of geographic distance and environmental heterogeneity in generating phylogeographic structure across the Amazon basin. *Journal of Biogeography*, **40**, 646–661.

Fine PVA, Metz MR, Lokvam J *et al.* (2013b) Insect herbivores, chemical innovation and the evolution of habitat specialization in Amazonian trees. *Ecology*, **94**, 1764–1775.

Frasier CL, Albert VA, Struwe L (2008) Amazonian lowland, white-sand areas as ancestral regions for South American biodiversity: biogeographic and phylogenetic patterns in *Potalia* (Angiospermae: Gentianaceae). *Organisms Diversity & Evolution*, **8**, 44–57.

Friar EA, Cruse-Sanders JM, McGlaughlin ME (2007) Gene flow in *Dubautia arborea* and *D. cilolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing gene hybridization. *Molecular Ecology*, **16**, 4028–4038.

Gentry AH (1988) Patterns of plant community diversity and floristic composition on environmental and geographical gradients. *Annals of the Missouri Botanical Garden*, **75**, 1–34.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution*, **60**, 268–278.

Haffer J (2008) Hypotheses to explain the origin of species in Amazonia. *Brazilian Journal of Biology*, **68**, 917–947.

Haldane JBS (1948) The theory of a cline. *Journal of Genetics*, **48**, 277–284.

Hendry AP, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. *Evolution*, **58**, 2319–2331.

Hendry AP, Day T, Taylor EB (2001) Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution*, **55**, 459–466.

Hoorn C (1993) Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: results of a palynostratigraphic study. *Paleogeography Paleoclimatology Paleoecology*, **112**, 187–238.

Howard DJ, Gregory PG, Chu J, Cain ML (1998) Conspecific sperm precedence is an effective barrier to hybridization between closely related species. *Evolution*, **52**, 511–516.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics*, **6**, 6–13.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250–255.

Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist*, **161**, 51–58.

Lawton-Rauh A, Friar EA, Remington DL (2007) Coordinated temporal patterns of differentiation among molecular markers in recently-derived species of the Hawaiian silversword alliance adaptive radiation (Heliantheae, Asteraceae). *Molecular Ecology*, **16**, 3993–3994.

Leinonen T, Cano JM, Makinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, **19**, 1803–1812.

Mayr E (1963) *Animal species and evolution*. Harvard University Press, Cambridge, Massachusetts.

Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C, Bordino P (2010) Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range. *Molecular Ecology*, **19**, 2212–2228.

Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.

Misiewicz TM (2014) Ecological divergence and reproductive isolation in an Amazonian tropical tree: *Protium subserratum* (Burseraceae). PhD Dissertation, University of California Berkeley, California.

Misiewicz TM, Barbosa CEA, Fine PVA (2012) Microsatellite Primers for an Amazonian lowland tropical tree, *Protium subserratum* (Burseraceae). *AJB Primer Notes and Protocols in the Plant Sciences*, **99**, E465–E467.

Nosil P (2012) *Ecological speciation*. Oxford University Press, Oxford.

Nosil P, Egan SP, Funk DJ (2007) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of the Sciences*, **99**, 13612–13615.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

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

Pregitzer CC, Bailey JK, Hart SC, Schweitzer JA (2010) Soils as agents of selection: feedbacks between plants and soils alter seedling survival and performance. *Evolutionary Ecology*, **24**, 1045–1059.

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and Ecumenicism. *Journal of Heredity*, **86**, 248–249.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org.

Remington DL, Robichaux RH (2007) Influences of gene flow on adaptive speciation in the *Dubautia arborea–D. ciliolata* complex. *Molecular Ecology*, **16**, 4014–4027.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Richardson JE, Pennington TR, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of neotropical rainforest trees. *Science*, **293**, 2242–2245.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenthal DM, Ludwig F, Donovan LA (2005) Plant responses to an edaphic gradient across an active sand dune/desert boundary in the great basin desert. *International Journal of Plant Sciences*, **166**, 247–255.

Ross CL, Harrison RG (2002) A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution*, **56**, 2296–2312.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Sääksjärvi IE, Haataja S, Neuvonen S *et al.* (2004) High local species richness of parasitic wasps (Hymenoptera: Ichneumonidae; Pimplinae and Rhyssinae) from the lowland rainforests of Peruvian Amazonia. *Ecological Entomology*, **29**, 735–743.

Savolainen V, Anstett MC, Lexer C *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Science*, **441**, 210–213.

Searcy KB, Macnair MR (1990) Differential seed production in *Mimulus gattatus* in response to increasing concentrations of

copper in the pistil by pollen from copper tolerant and sensitive sources. *Evolution*, **44**, 1424–1435.
Smith TB, Wayne RK, Giman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science*, **276**, 1855–1857.
Smith SS, Schweitzer JA, Turk P *et al.* (2011) Soil-mediated local adaptation alters seedling survival and performance. *Plant and Soil*, **352**, 243–251.
Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.
Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.
Tuomisto H, Ruokolainen K, Linna A, Danjoy W, Rodriguez Z (1995) Dissecting Amazonian biodiversity. *Science*, **299**, 241–244.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–259.
Yost JM, Barry T, Kay KM, Rajakaruna N (2012) Edaphic adaptation maintains the coexistence of two cryptic species on serpentine soils. *American Journal of Botany*, **99**, 890–897.
Zhang C, Zhang D, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747–761.

T.M.M. and P.V.A.F. designed the research and wrote the paper. T.M.M. performed the research and analysed the data.

Data accessibility

Sample locations; microsatellite data; morphological data; Structure, NewHybrid and Migrate-N input files: Dryad entry DOI:10.5061/dryad.S3k04.