Journal of Biogeography (J. Biogeogr.) (2012)

SPECIAL ISSUE

The importance of environmental heterogeneity and spatial distance in generating phylogeographic structure in edaphic specialist and generalist tree species of *Protium* (Burseraceae) across the Amazon Basin

Paul V. A. Fine[1]*, Felipe Zapata[1], Douglas C. Daly[2], Italo Mesones[1], Tracy M. Misiewicz[1], Hillary F. Cooper[1,3] and C. E. A. Barbosa[4]

[1]*Department of Integrative Biology, University of California, Berkeley, 1005 Valley Life Sciences Bldg #3140, Berkeley, CA 94720-3140, USA,* [2]*The New York Botanical Garden, 200th Street and Kazimiroff Boulevard, Bronx, NY 10458-5126, USA,* [3]*Department of Biology, Northern Arizona University, Flagstaff, AZ 86011-5640, USA,* [4]*Instituto Nacional de Pesquisas da Amazônia – INPA, Coordenação de Pesquisas em Ecologia – CPEC, CP 478, 69011-970 Manaus, AM, Brazil*

*Correspondence: Paul V. A. Fine, Department of Integrative Biology, University of California, Berkeley, 1005 Valley Life Sciences Bldg #3140, Berkeley, CA 94720-3140, USA.
E-mail: paulfine@berkeley.edu

ABSTRACT

Aim Edaphic heterogeneity may be an important driver of population differentiation in the Amazon but remains to be investigated in trees. We compared the phylogeographic structure across the geographic distribution of two *Protium* (Burseraceae) species with different degrees of edaphic specialization: *Protium alvarezianum*, an edaphic specialist of white-sand habitat islands; and *Protium subserratum*, an edaphic generalist found in white sand as well as in more widespread soil types. We predicted that in the edaphic specialist, geographic distance would structure populations more strongly than in the edaphic generalist, and that soil type would not structure populations in the edaphic generalist unless habitat acts as a barrier promoting population differentiation.

Location Tropical rain forests of the Peruvian and Brazilian Amazon, Guyana and French Guiana.

Methods We sequenced 1209–1211 bp of non-coding nuclear ribosomal DNA (internal transcribed spacer and external transcribed spacer) and a neutral low-copy nuclear gene (phytochrome C) from *P. subserratum* ($n = 65$, 10 populations) and *P. alvarezianum* ($n = 19$, three populations). We conducted a Bayesian phylogenetic analysis, constructed maximum parsimony haplotype networks and assessed population differentiation among groups (soil type or geographic locality) using analysis of molecular variance and spatial analysis of molecular variance.

Results The edaphic specialist exhibited considerable genetic differentiation among geographically distant populations. The edaphic generalist showed significant genetic differentiation between the Guianan and Amazon Basin populations. Within Peru, soil type and not geographic distance explained most of the variation among populations. Non-white-sand populations in Peru exhibited lower haplotype/nucleotide diversity than white-sand populations, were each other's close relatives, and formed an unresolved clade derived from within the white-sand populations.

Main conclusions Geographic distance is a stronger driver of population differentiation in the edaphic specialist than in the generalist. However, this difference did not appear to be related to edaphic generalism per se as adjacent populations from both soil types in the edaphic generalist did not share many haplotypes. Populations of the edaphic generalist in white-sand habitats exhibited

Journal of Biogeography

high haplotype diversity and shared haplotypes with distant white-sand habitat islands, indicating that they have either efficient long-distance dispersal and/or larger ancestral effective population sizes and thus retain ancestral polymorphisms. These results highlight the importance of edaphic heterogeneity in promoting population differentiation in tropical trees.

Keywords

AMOVA, edaphic gradients, habitat specialization, SAMOVA, South America, trees, tropical rain forest, white sand.

INTRODUCTION

The lowlands of the Amazon Basin contain rain forests with the highest diversity of trees in the world (Gentry, 1988; Fine & Ree, 2006; Hubbell *et al.*, 2008). While many studies have explored mechanisms by which high tree diversity is maintained (reviewed in Wright, 2002), comparatively few studies have evaluated the factors underlying tree species diversification in the Amazon (Pennington & Dick, 2010). Given that the same processes that cause divergence in populations within species are likely to lead to speciation, studying the phylogeography of tree species at large scales may give insight into the processes most important for species formation (Avise, 2000; Knowles & Maddison, 2002; Hickerson *et al.*, 2010). Changes in population-level gene frequencies can result from a variety of processes, but the two considered to be the most important are disruption in gene flow with geographic distance (i.e. tree populations become spatially isolated due to dispersal limitation promoting differentiation due to genetic drift) and habitat specialization related to environmental variation (i.e. trees adapt and specialize to local conditions, e.g. local soil types) (Linhart & Grant, 1996; Tuomisto *et al.*, 2003; Levin, 2004).

Very few Amazonian tree species have been the subject of phylogeographic analyses, and available studies from three unrelated plant taxa indicate an overall low phylogeographic structure. For example, *Ceiba pentandra* (Malvaceae) populations from across the Amazon exhibited only three internal transcribed spacer (ITS) haplotypes (Dick *et al.*, 2007). Similarly, *Symphonia globulifera* (Clusiaceae) populations from French Guiana to Ecuador to Bolivia shared one single ITS haplotype, although a higher diversity of chloroplast DNA haplotypes was recovered (Dick *et al.*, 2003; Dick & Heuertz, 2008). A microsatellite study of *Swietenia macrophylla* (Meliaceae) reported lower than expected geographic variation, even across more than 2000 km (Lemes *et al.*, 2003). In contrast, phylogeographic analyses of Central American trees have revealed stronger phylogeographic structure and more complex haplotype genealogies (Caron *et al.*, 2000; Cavers *et al.*, 2003; Dick *et al.*, 2003; Novick *et al.*, 2003; Dick & Heuertz, 2008). Whereas the topographic heterogeneity that characterizes Central American landscapes may largely explain these phylogeographic patterns, Amazonian landscapes are comparatively flat yet include strong edaphic, flooding and climate gradients (ter Steege *et al.*, 2006; Baraloto *et al.*, 2011). These environmental gradients have been hypothesized to contribute to diversification in the Amazon biota (Gentry, 1986; Moritz *et al.*, 2000). Yet, to date, no studies have investigated the role of environmental heterogeneity (including soils) as a potential factor influencing phylogeographic structure in widespread Amazonian trees.

An ideal study system with which to address this is the flora occurring on the complex mosaic of white-sand and clay soils widely distributed throughout the Amazon Basin. The ancient white-sand deposits are habitat islands, surrounded by other terra firme forests on more fertile soils and they harbour mostly edaphic endemic tree species that often have close relatives on neighbouring soil types (Fine *et al.*, 2010; Fine & Kembel, 2011). There are few known species that commonly occur in both the nutrient-poor white-sand soils and the neighbouring more fertile soil types, suggesting that success in these habitats is associated with suites of habitat-specific traits, including nutrient conservation and anti-herbivore defence among others (Janzen, 1974; Anderson, 1981; Fine *et al.*, 2006).

A partial phylogeny has been reconstructed for one clade of tropical trees, Protieae (including the genera *Protium*, *Tetragastris* and *Crepidospermum* from the family Burseraceae), a species-rich group that includes multiple representatives of both white-sand specialists as well as species restricted to other soil types in terrace and clay forests (Fine *et al.*, 2005). Twenty-six out of the 35 species of Protieae in the Peruvian Amazon were found to be restricted to only one of the three soil types, with edaphic specialization evolving repeatedly in this clade (Fine *et al.*, 2005). One hypothesis for this pattern is that ecological tradeoffs in allocation to growth rate and defence, due to the strong differences in resource availability among the three soil types, may promote habitat specialization in these species, thus leading to repeated habitat-mediated speciation (Fine *et al.*, 2004, 2006). To further investigate the patterns of habitat specialization in Protieae, here we employ a comparative phylogeographic approach between a generalist species, which occurs in white sand as well as other soil types, and a sister clade that includes specialist species, which occurs only in white-sand forests, to evaluate how geographic distance and

environmental heterogeneity influence the phylogeographic structure of species with different degrees of edaphic specialization.

Because white-sand habitats are spatially isolated from one another in the Amazon, we predict that neutral molecular variation should be structured by geographic distance more strongly in the edaphic specialist species than in the edaphic generalist. In the edaphic generalist species, we predict that neighbouring populations should be closely related and freely exchange haplotypes irrespective of habitat type, provided the species is a true habitat generalist with broad environmental tolerances. In this case, neutral molecular variation should also be structured by geographic distance. However, if edaphic heterogeneity acts as a barrier promoting ecological divergence, we expect populations from divergent habitats to be distantly related irrespective of geographic distance and thus they will not exchange haplotypes very often across habitats. In this case, neutral molecular variation should be structured by habitat type and the generalist species may in fact be a complex of cryptic species. Thus, a study of the phylogeographic structure of multiple populations of both an edaphic specialist and a generalist tree species across their geographic ranges provides an important first step in an investigation of the potential mechanisms underlying diversification in Amazonian tree species.

MATERIALS AND METHODS

Study species

The section *Papilloprotium* (Daly & Fine, 2011) was chosen to be the focus of this study because it is a small well-supported monophyletic group of four taxa that includes both edaphic generalist and specialist taxa. Within this section, *Protium alvarezianum* Daly & P. Fine and *Protium subserratum* (Engl.) Engl. were sampled for multiple individuals from multiple populations in Peru, Brazil, Guyana and French Guiana (Fig. 1, Table 1). *Protium alvarezianum* is known only from white-sand forests in Peru and the Rio Negro Basin of Brazil and Venezuela and is sister to *Protium reticulatum* Engl., a white-sand specialist known only from a few different white-sand forests in the Rio Negro Basin of Venezuela and Brazil. These two taxa are themselves sister to *Protium subserratum*, which occurs from Ecuador and Peru to French Guiana across northern South America (Daly & Fine, 2011). The fruits of *P. alvarezianum* and *P. subserratum* are thought to be dispersed by a large variety of birds and monkeys (like most red-fruited *Protium*; Daly, 1987) although direct observations in these species have never been published. Flowers of both species are extremely small (< 5 mm), white, fragrant and contain nectar.

Protium subserratum occurs in white-sand forest, clay and brown-sand (terrace) soils (Fine *et al.*, 2005). This taxon presents substantial intra-specific morphological variation in leaf characters, some of which is related to the soil type in which the plants occur. Briefly, in the white-sand forests of Peru and Brazil, specimens have significantly thicker leaflets, entire leaflet margins and are densely pubescent on the abaxial surface of the leaflets; conversely, individuals from other soil types have thinner glabrous leaflets with strongly toothed margins (Fig. 2, Table 2). No consistent morphological differences were observed in any flower or fruit characters that corresponded with the leaf character differences in the various populations of *P. subserratum* that we sampled, nor did there emerge support for independent lineages corresponding to the different morphotypes in a molecular phylogenetic analysis with two chloroplast and three nuclear genes (Daly & Fine, 2011).

We located populations of *P. subserratum* from 11 localities (one in Guyana, two in French Guiana, one in Manaus, Brazil and seven in the Peruvian Amazon; see Fig. 1). In Peru, we located three river basins where we were able to collect one white-sand population and at least one other population on clay or terrace soils: the Nanay River in and around the Allpahuayo-Mishana National Reserve near Iquitos, the Ucayali River near Jenaro Herrera, and the Blanco River in the Matsés National Reserve (Fig. 1). For *P. alvarezianum* we located populations from all three localities in the Peruvian Amazon where they are known to occur (Fig. 1). From each population of both species, we collected voucher specimens of 12–15 individual trees with a pole pruner, including leaves, branches and flowers and/or fruits (if reproductive); a small amount of leaf material was preserved in silica gel for molecular analysis. GPS coordinates were taken at each tree, and a small soil sample (the first few centimetres below the organic layer) was collected within a metre of the trunk of the tree.

Soils

The major habitat types in the lowland Amazon can be differentiated by soil type and by flooding regime (Duivenvoorden & Duque, 2010). Soils are complex, but there are some broad categories that are readily distinguished by very different geological histories and nutrient availability (Hoorn, 1993; Frasier *et al.*, 2008). For example, in the Peruvian Amazon, Fine *et al.* (2005) divided the terra firme habitats into three main categories: white-sand forests, terrace (brown-sand) forests and clay forests. White-sand soils are perhaps the most obviously distinct soil type in the lowland Amazon. These soils are composed mainly of quartz, derived from Precambrian sediments that make up the Guiana and Brazilian shields, which were deposited by rivers draining westward prior to the Andean uplift (Hoorn, 1993, 1994). The trees growing on these soils in the western Amazon have stunted canopies, from 3 to 25 m, thought to be related to the lower cation exchange capacity of white-sand soils (Anderson, 1981). Indeed, measurements of available nitrogen in white-sand soils are of an order of magnitude lower than neighbouring terrace and clay soils that support rain forests with canopy heights of > 40 m (Fine *et al.*, 2005). In Peru, white-sand forests occur on the landscape as habitat islands with areas ranging from a few hectares to a few square kilometres. They are often arranged as archipelagos that are geographically separated from other white-sand forests by tens to hundreds of kilometres, and they

Figure 1 Map of the localities of *Protium subserratum* and *Protium alvarezianum* populations in the present study. See Table 1 for more information about the localities. Localities S8–S12 are shown on the top panel (a) and the Peruvian localities (A1–A3; S1–S7) are shown in (b). Open circles denote white-sand forest localities.

Table 1 Localities of *Protium alvarezianum* and *Protium subserratum* populations from the present study, with the average soil texture of soil samples collected within 1 m from each individual tree. No soil sample was collected for the individual at S9. See Fig. 1 for map of localities. Significant differences (Tukey post hoc tests) among group means are noted with different superscripts.

Population (location)	Habitat	Latitude	Longitude	% Sand	% Silt	% Clay
A1 *alvarezianum*	White sand	−4.320	−77.260	74.4^{b}	22.9^{b}	2.7^{a}
A2 *alvarezianum*	White sand	−4.901	−73.629	74.8^{b}	23.2^{b}	2.0^{a}
A3 *alvarezianum*	White sand	−5.854	−73.629	64.0^{b}	34.1^{b}	1.8^{a}
S1 *subserratum* (Nanay)	Brown sand	−3.978	−73.426	88.9^{b}	7.4^{a}	3.6^{a}
S2 *subserratum* (Nanay)	White sand	−3.973	−73.426	93.7^{b}	6.1^{a}	0.2^{a}
S3 *subserratum* (Nanay)	Clay	−3.833	−73.596	13.3^{a}	48.1^{b}	38.7^{b}
S4 *subserratum* (Jenaro Herrera)	White sand	−4.896	−73.644	66.3^{b}	19.2^{b}	14.6^{b}
S5 *subserratum* (Jenaro Herrera)	Brown sand	−4.862	−73.616	75.3^{b}	23.4^{b}	1.3^{a}
S6 *subserratum* (Rio Blanco)	White sand	−5.855	−73.770	64.3^{b}	34.9^{b}	0.8^{a}
S7 *subserratum* (Rio Blanco)	Clay	−6.304	−73.618	36.0^{a}	45.0^{b}	19.0^{b}
S8 *subserratum* (Manaus – Reserva Ducke)	Clay, sand	−2.5871	−59.558	68.18^{b}	2.60^{a}	29.18^{b}
S9 *subserratum* (Manaus – WS)	White sand	−2.8491	−60.2413			
S10 *subserratum* (Potara River, Guyana)	Brown sand	5.375	−59.577	65.2^{b}	15.5^{b}	19.3^{b}
S11 *subserratum* (Paracou French Guiana)	Brown sand	4.050	−52.040	73.8^{b}	9.6^{ab}	16.6^{b}
S12 *subserratum* (Regina St. Virginie, FG)	Brown sand	5.264	−52.931	69.4^{b}	16.7^{b}	13.9^{b}

Figure 2 Photos of leaflets of *Protium subserratum* from Peru, including close-ups. On the top is an individual from terrace soil forest, and on the bottom is an individual from white-sand forest. Both leaflets were collected from adult trees of similar size and light environment. These two individuals hail from populations that are 5 km distant from one another. Note the differences in leaf teeth and pubescence.

support an endemic flora of about 135 species of specialist and facultative specialist trees (Fine *et al.*, 2010). There are strong dominance patterns in these forests, and a few species generally make up more than half the number of stems; 0.1-ha plots in white-sand forest averaged only 41.5 species of trees ≥ 5 cm diameter at breast height (Fine *et al.*, 2010).

Clay soils (derived from the Solimões or Pebas Formation in the western Amazon) originated from the erosion of Cretaceous metamorphic rocks that became exposed during Andean uplift beginning in the Miocene (Hoorn, 1993). These soils have relatively high nutrient availabilities, and are found only in the western Amazon, generally within a few hundred kilometres of the Andes. These forests are characterized by high canopies and very high species diversities at the 1-ha plot level, commonly exceeding 250 tree species (Pitman *et al.*, 2008). Other clay soils found in the central Amazon and the Guianas are much older and more weathered and have lower nutrient availabilities (Hammond, 2005).

Terrace soils derive from sand and gravel of Andean origin deposited by Pliocene and Pleistocene rivers (Hoorn, 1993, 1994). These soils are intermediate in soil fertility between white-sand and Pebas clay soils. Non-white-sand forests in Brazil, Guyana and French Guiana in this study are similar to Peruvian terrace soils in terms of nutrient availability (Baraloto *et al.*, 2011). These forests are similar in structure and diversity to those on clay soils, although species composition can be markedly different (Pitman *et al.*, 2008).

Soil analyses

Each of the individual soil samples was subjected to particle size analysis. We followed a hydrometer method modified from Day (1965) and Gee & Bauder (1979). The amount of soil used varied according to its composition: sandier soils required more mass to ensure the methods were reproducible. Average amounts were *c.* 20 g for clay soils and 40 g or more for sandy soils. Once all organic matter was removed from the sample, 100 mL of 5% sodium hexametaphosphate (HMP) and 200 mL deionized water were added and shaken on a horizontal shaker for a minimum of 4 h. The final volume was brought to 1 L using distilled water and the mixture was then poured into a 1-L graduated cylinder. In another 1-L graduated cylinder, 100 mL 5% HMP was added to 900 mL water to serve as a blank. After manually shaking the cylinder several times to homogenize the soil mixture, hydrometer (ASTM 152H) readings were measured at 30 s and 1, 3, 10, 30, 60, 90 and 120 min, and compared with the blank each time. From these readings, the percentages of sand, silt and clay were calculated.

Table 2 Leaflet measurements (± SE) for the populations of *Protium alvarezianum* and *Protium subserratum* in the present study. Significant differences (Tukey post hoc tests) among group means are noted with different superscripts.

Population (soil type)	Leaflet length (mm)	Leaflet thickness (mm)	Leaflet margin teeth (no. of teeth cm^{-1})	Leaflet pubescence (no. of hairs cm^{-2})
A1 *alvarezianum*	88.0 ± 3.2	0.11^{a} ± 0.005	1.19^{ab} ± 0.9	0^{a}
A2 *alvarezianum*	92.8 ± 12.9	0.15^{a} ± 0.004	0.55^{a} ± 1.3	0^{a}
A3 *alvarezianum*	120.1 ± 5.1	0.11^{a} ± 0.006	0^{a}	0^{a}
S1 *subserratum* (brown sand)	102.9 ± 10.4	0.15^{a} ± 0.005	0.29^{a} ± 0.4	30.8^{b} ± 6.3
S2 *subserratum* (white sand)	103.9 ± 6.9	0.18^{ab} ± 0.010	0.125^{a} ± 0.13	72.5^{b} ± 4.5
S3 *subserratum* (clay)	172.1 ± 22.6	0.16^{a} ± 0.004	1.89^{b} ± 0.56	0^{a}
S4 *subserratum* (brown sand)	138.3 ± 13.2	0.13^{a} ± 0.006	2.21^{b} ± 1.3	0^{a}
S5 *subserratum* (white sand)	106.2 ± 13.5	0.20^{b} ± 0.004	0^{a}	315^{c} ± 37.9
S6 *subserratum* (white sand)	96.2 ± 5.6	0.21^{b} ± 0.020	0^{a}	207^{c} ± 9.9
S7 *subserratum* (clay)	122.7 ± 11.0	0.14^{a} ± 0.009	1.64^{b} ± 0.64	0^{a}
S8 *subserratum* (Manaus, sand and clay)	137.89 ± 10.3	0.15^{a} ± 0.004	1.01^{b} ± 0.9	0^{a}
S10 *subserratum* (Guyana, brown sand)	128.4 ± 11.4	0.13^{a} ± 0.007	1.63^{b} ± 0.48	0^{a}
S11, S12 *subserratum* (French Guiana, clay)	98.1 ± 13.7	0.13^{a} ± 0.015	0.52^{ab} ± 1.28	0^{a}

Molecular data

DNA extractions

The collection numbers, herbarium where the specimens are deposited, geographic locality and GenBank accession number for each of the samples included in our molecular analyses are listed in Appendix S1 in Supporting Information. From each population of *P. subserratum* and *P. alvarezianum*, we extracted DNA from five to nine individuals (see Table 3). The samples were either dried fresh in silica gel and stored at −80 °C or taken from a herbarium sheet (i.e. *P. reticulatum*). All leaf tissue was ground using a Mini Beadbeater-16 (BioSpec Products Inc., Bartlesville, OK, USA) with 2.3 mm zirconia/silica beads. Extractions were carried out using the DNEasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The protocol for herbarium specimens was modified to include an additional step where 3–5 μL of Proteinase K was added to the sample in the lysis buffer and incubated on a horizontal shaker at 42 °C for 12–24 h.

PCR protocols

We amplified the ribosomal ITS (NY 183: CCT TAT CAT TTA GAG GAA GGA G (F) or ITSleu1: GTC CAC TGA ACC TTA TCA TTT AG (F), and NY 109:GTG ACG CCC AGG CAG ACG T (R) or ITS4: CCT TCC GCT TAT TGA TAT GC (R); Malcomber, 2002; Fine *et al.*, 2005), external transcribed spacer (ETS) (ETS-1: TTC GGT ATC CTG TGT TGC TTA C (F), ETS 18S-IGS: GAG ACA AGC ATA TGA CTA CTG GCA GGA TCA ACC AG (R); Baldwin & Markos, 1998; Weeks *et al.*, 2005) and one nuclear locus, *c.* 700 bp of the phytochrome C (*phy*C) starting *c.* 550 bp downstream in exon 1 (phyC-smbF1: GGC AYT GAA RTC ATA YAA RCT TGC (F), phyC-smbR1: CCR CCC CAC TTG ATC TGY TT (R); Clayton *et al.*, 2007) to assess the phylogenetic relationships within and among the populations of *P. subserratum* and *P. alvarezianum.* Results from a Tajima's *D*-test (Tajima, 1989) revealed that *phy*C evolves neutrally ($P = 0.1$). We included one individual for each of the remaining taxa from *Papilloprotium* (Daly & Fine, 2011) and used *P. ferrugineum* as outgroup because our phylogenetic analyses, with a larger sample of Protieae species, revealed this species to be sister to the other species in *Papilloprotium* with 100% bootstrap support and 1.00 posterior probability (Daly & Fine, 2011).

Polymerase chain reactions (PCR) contained 10 × Bioneer's Accupower reaction buffer, 250 μM deoxynucleotide triphosphates (dNTPs), and 1 unit TLA DNA Polymerase, or 2 × Green GoTaq Promega reaction buffer, 400 μM dNTPs, 3 mM $MgCl_2$ and 1 unit of Taq DNA polymerase, with 0.4 μM of each forward and reverse primer and 2 μL of undiluted template DNA. The amplification products were visualized with ultraviolet (UV) light on 1% TRIS/borate/EDTA (TBE) agarose gels and cleaned using Exonuclease I and Shrimp Alkaline Phosphatase (USB Corporation, Affymetrix Inc., Santa Clara, CA, USA).

Cloning protocols

To assess sequence diversity within the ITS PCR amplicon pool, we cloned PCR products using the pGEM-T vector system (Promega, Madison, WI, USA) following the manufacturer's instructions except that reaction volumes for ligation and transformation were halved. We did not clone ETS and *phy*C PCR products because these did not reveal polymorphisms within individuals with direct sequencing. For blunt-ended PCR products, we carried out a standard A-tailing procedure before ligation. Blue-white colony screening was used to pick at least five positively transformed colonies for each accession, which

Table 3 Indices of molecular diversity per population of *Protium alvarezianum* (A1–A3) and *Protium subserratum* (S1–S11). Sample size, number of haplotypes, number of phylogenetically informative sites and theta (π) are presented for each population. Theta (π) is estimated from the infinite-site equilibrium relationship between the mean number of pairwise differences (π) and theta. These metrics were calculated in Arlequin 3.1, with the combined dataset with a locus separator between *phy*C and ETS–ITS. There were a total of 1209–1211 nucleotide sites for each individual.

Population (soil type or locality)	Sample size	Haplotypes	Phylogenetically informative sites	Theta (π), (SD)
A1	5	5	1	0.20 (0.33)
A2	6	4	2	0.93 (0.85)
A3	8	7	11	2.90 (1.98)
S1 (Peru, Nanay, brown sand)	7	6	2	0.28 (0.39)
S2 (Peru, Nanay, white sand)	6	6	6	3.20 (2.21)
S3 (Peru, Nanay, clay)	6	5	0	0.0 (0.0)
S4 (Peru, Jenaro Herrera, brown sand)	5	1	0	0.0 (0.0)
S5 (Peru, Jenaro Herrera, white sand)	6	6	7	2.4 (1.74)
S6 (Peru, Rio Blanco, white sand)	6	6	11	4.73 (3.11)
S7 (Peru, Rio Blanco, clay)	5	5	4	1.60 (1.32)
S8 (Manaus)	9	9	11	2.52 (1.70)
S10 (Guyana)	6	6	5	2.26 (1.66)
S11-S12 (French Guiana)	8	7	2	0.43 (0.48)

were further verified by PCR using universal M13F/R primers. Plasmids were cleaned following the FastPlasmid Mini Kit protocol (Fisher Scientific, Gaithersburg, MD, USA).

Sequencing and alignment

Amplification products were cycle-sequenced using the original PCR primers or the SP6-T7 universal primers for cloned products in 10 μL reactions with the standard BigDye (Applied Biosystems, Foster City, CA, USA) protocol and afterwards cleaned using an EDTA–ethanol precipitation step. Sequencing was performed on an ABI 3730 sequencer (Applied Biosystems). The resulting chromatograms were edited, and forward and reverse sequences were assembled using Geneious Pro 5.3.3 (Drummond *et al.*, 2010). Initial sequence alignments were made using ClustalW (EMBL-EBI, Cambridge, UK) and subsequent manual alignments in MacClade 4.05 (Maddison & Maddison, 2002). For cloned products, we used only identical sequences that were recovered more than once per PCR amplicon pool, and considered sequences differing in fewer than three polymorphisms as Taq error. When necessary, consensus sequences from each ITS amplicon were created for final analyses. No divergent/paralogous copies in ITS were detected.

Data analyses

Phylogenetic analyses

We analysed the concatenated data set using Bayesian inference in MrBayes 3.2 (Ronquist & Huelsenbeck, 2003) with the data partitioned by gene evolution under different stochastic evolutionary models. We used MrModeltest 2.3 (Nylander, 2004) to select the best stochastic model that fitted the dataset under the Akaike information criterion corrected for small sample size (AIC_c) (HKY+I for ETS and *phy*C; GTR for ITS). All model parameters were unlinked among partitions and we let partitions evolve under different rates. We used the default uniform priors for all parameters, and ran a Metropolis-coupled Markov chain Monte Carlo (MC^3) for 10×10^6 generations sampling every 1000 generations with two internal runs and four chains, one cold and three incrementally heated; after preliminary analyses we decided to lower the heating temperature to 0.1 for final analyses to ensure proper chain swapping. Adequate mixing and convergence of the MC^3 to the stationary distribution was confirmed by inspecting a plot of likelihood values per generation and evaluating that the average standard deviation of split frequencies stabilized below 0.01. We summarized the posterior probability distribution of sampled trees with a 50% majority rule consensus tree after discarding the first 25% of sampled trees as 'burn-in'.

Haplotype networks

To visualize the patterns of relatedness in populations of *P. subserratum* generated by each locus, haplotype networks were created in tcs 1.21 using statistical parsimony (Clement *et al.*, 2000), with 95% parsimony connection limit and considering gaps as missing data.

Population structure

We investigated the structure of the populations of the two species to reveal correlations with genetic variation, soil type and geography across the Amazon Basin by using a hierarchical approach implemented through analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992) employing Arlequin 3.1 (Excoffier *et al.*, 2005). We entered our sequence data into Arlequin, and analysed all loci using a locus separator function to separate ETS–ITS from *phy*C. Our population S12 of *P. subserratum* in French Guiana contained only two individuals, so we lumped it together with the six individuals from S11 (Paracou), which is 250 km distant, to form one single sample with eight individuals. Individuals in populations S11 and S12 shared the same haplotypes for both loci.

The AMOVA partitions the total variance observed into covariance components that result from the variance among groups (geographic regions or soil types), among populations within groups and within populations. Next, the covariance components are used to calculate fixation indices, ϕ-statistics, which are analogous to Wright's *F*-statistics (Excoffier *et al.*, 2005). The significance of the fixation indices is tested using a nonparametric permutation approach with 1000 replicates (Excoffier *et al.*, 2005).

We tested the relative influence of geographic distance versus soil type on the phylogeographic structure of *P. subserratum*, specifically whether populations in white-sand localities geographically isolated from one another were more similar to each other than to populations in clay or terrace soil located nearby. To do so, we grouped the populations by location with respect to the factor of interest and compared the ϕ-statistics in sequential analyses. For example, to quantify the effect of cross-Amazon separation, we grouped the Peruvian and Brazilian populations of *P. subserratum* into one group, and the Guyanan and French Guianan populations into a second group. To quantify the relative strength of soil versus geographic distance within Peru, we grouped the white-sand populations into one group and the terrace and clay populations into a second group and compared the ϕ-statistics with a second analysis where white-sand and clay and terrace populations were grouped by location in the three river basins within Peru (Fig. 1). For *P. alvarezianum*, we did not group the populations but simply compared the ϕ-statistics of genetic diversity with the scores of the *P. subserratum* white-sand populations to see whether isolation by distance would be a stronger influence than for white-sand populations of *P. subserratum*, given that *P. alvarezianum* occurs only in white-sand islands while *P. subserratum* may be sharing alleles with non-white-sand populations.

As a further test of phylogeographic structure we conducted a spatial analysis of molecular variance (SAMOVA) using samova 1.0. This method implements an approach to define groups of

populations by maximizing the proportion of total genetic variance due to differences among groups through simulated annealing (Dupanloup *et al.*, 2002). SAMOVA was run with 100 simulated annealing processes and different numbers of groups (K) until the F_{CT} value (the proportion of genetic variance among groups of populations) reached a plateau. SAMOVA requires more than three populations per species, thus we only carried out this analysis for *P. subserratum*.

RESULTS

The total aligned length of the three nuclear loci for *P. alvarezianum* (n = 19) and *P. subserratum* (n = 65) was 1209 to 1211 bp. The combined data set contained 42 polymorphic sites, distributed evenly among the three loci (*phy*C with 15, ITS with 13 and ETS with 14).

In the Bayesian phylogenetic analysis, *P. subserratum* received strong support as a monophyletic group with 0.97 posterior probability, sister to *P. reticulatum* + all *P. alvarezianum* populations, which themselves formed a monophyletic group with 1.0 posterior probability (Fig. 3). *Protium alvarezianum* populations separated into three subclades that mostly corresponded to geographically separated populations: the most westerly population, Morona (A1) was sister to a polytomy that included two well-supported subclades – one that consisted of a group of the Jenaro Herrera (A2) population plus one individual from Rio Blanco (A3_pf23), and one that included all of the remaining Rio Blanco

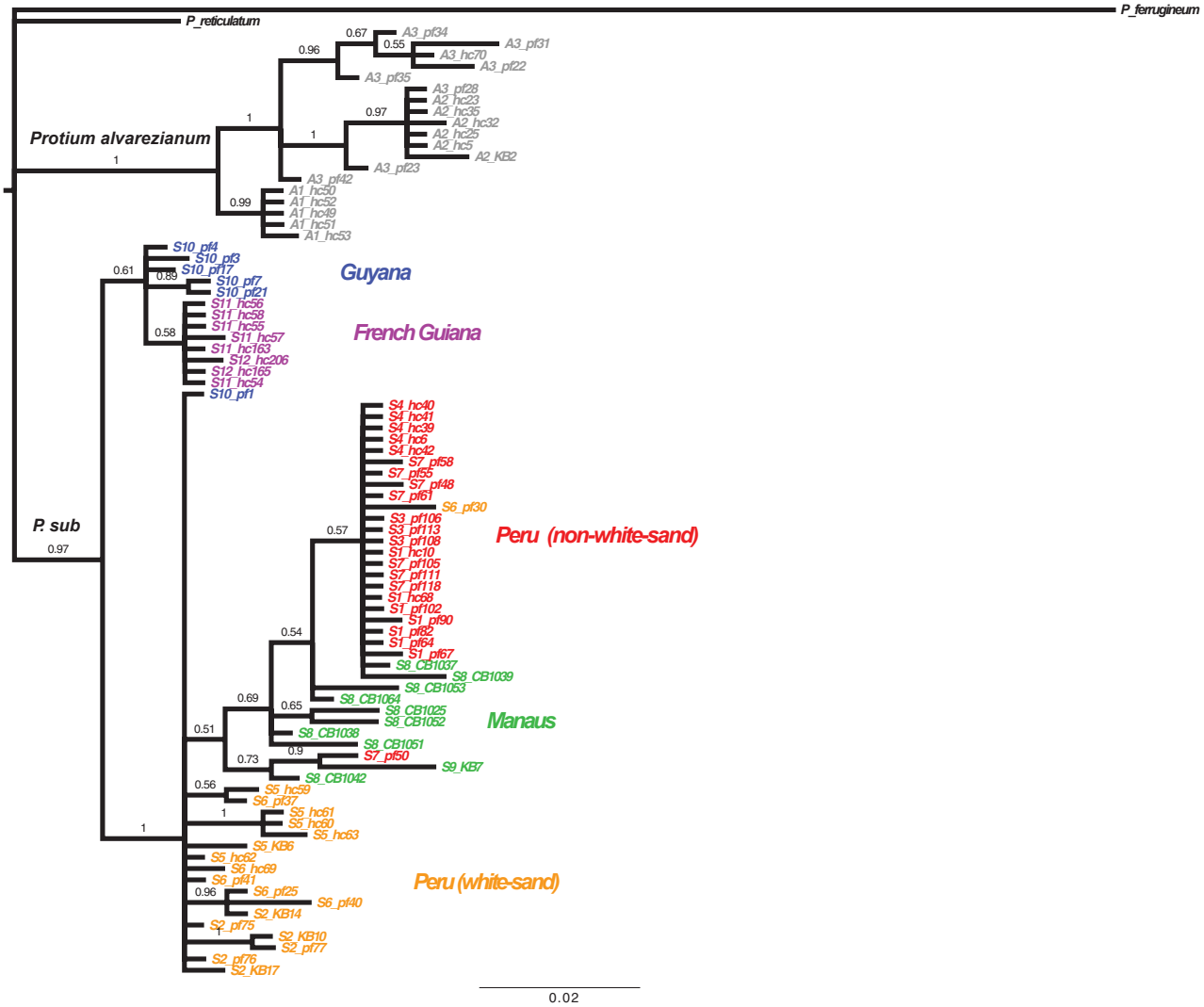

Figure 3 Majority rule (50%) consensus tree of Bayesian inference for concatenated *phy*C, internal transcribed spacer (ITS) and external transcribed spacer (ETS) nuclear genes of populations of *Protium subserratum* and *Protium alvarezianum*. Branch lengths correspond to expected substitutions per site. Posterior probabilities are shown at the nodes. Colours correspond to the morphotypes within *P. subserratum* corresponding to different soil types and/or geographic location: red represents clay and terrace soil morphotypes in Peru, orange represents white-sand morphotypes in Peru, green represents the Manaus population, blue represents Guyanan morphotypes, and purple represents French Guianan morphotypes.

Table 4 Population pairwise ϕ_{ST} statistics for *Protium alvarezianum*. See Table 1 and Fig. 1 for locality information. All three comparisons are significantly different from one another ($P < 0.0001$). Above the diagonal is pairwise geographic distance (km).

	A1	A2	A3
A1	0	407.6	436.8
A2	0.91785	0	106
A3	0.60164	0.55485	0

population (Fig. 3). Pairwise ϕ_{ST} comparisons for *P. alvarezianum* revealed high population structuring among each of the three white-sand populations of this species, with high ϕ_{ST} values among them (ranging from 0.55 to 0.92; Table 4), much higher than the pairwise comparisons within the western Amazon of *P. subserratum* white-sand morphotypes (ranging from 0.008 to 0.375; Table 5).

There is evidence of substantial phylogeographic structure between eastern and western populations of *P. subserratum*, as all western individuals belonged to a clade with 1.0 posterior probability, although one Guyanan individual was a member of this 'western Amazonian' clade. Unlike *P. alvarezianum,* there was no support for any geographic structure by river basin, as individuals from S2, S5 and S6 (white-sand locations) were not differentiated by river basin in the phylogeny, and individuals from S1, S3, S4 and S7 (clay and terrace locations) occurred in the same large clade, not showing any geographic structure. With one exception, all clay and terrace soil populations from all three river basins were found in one clade that is supported with a 0.57 posterior probability, derived from a grade that included almost all white-sand morphotypes and all of the Manaus individuals (Fig. 3). There was one individual from a white-sand population in this same group of clay and terrace individuals. In general, the posterior probabilities within the *P. subserratum* clade are low, consistent with the hypothesis that this lineage contains only a single species, and that different morphotypes are not monophyletic lineages (Daly & Fine, 2011).

The haplotype networks for each locus showed broad congruence that matches the concatenated Bayesian analyses of the three loci (Fig. 4). In general, white-sand populations and Manaus populations of *P. subserratum* had much higher haplotype and nucleotide diversity than clay and terrace populations, even though they had comparable sample sizes (Table 3, Fig. 4). For *phy*C, all but one of the 22 individuals from the three Peruvian clay and terrace populations share one single haplotype. White-sand populations were composed of six different haplotypes, one of which is shared with a Guyanan population more than 2000 km to the east. All French Guianan populations shared the same two haplotypes that differ by one base pair. For ITS and ETS, all French Guianan individuals shared the same haplotype with most Guyanan individuals. As in *phy*C, there was a diversity of haplotypes in the white-sand morphotypes and the Manaus population and very few in the Peruvian clay and terrace morphotypes. Two white-sand individuals shared haplotypes with the clay and terrace morphotypes. Pairwise ϕ_{ST} values mirrored these results, with extremely low ϕ_{ST} values among all clay and terrace populations, as well as between two of the white-sand islands with each other (Table 5).

There was significant population structure among groups, among populations within groups, and within populations when an AMOVA was performed on all *P. subserratum* populations, with most of the explained variation occurring between eastern and western populations (Table 6). For the AMOVA analyses conducted on only Peruvian populations of *P. subserratum*, grouping by soil (white sand versus non-white-sand) explained the highest percentage of the variance, with a significant among-group component (Table 6). When grouping by river basin, most of the variation was explained among populations within groups rather than among groups, consistent with the hypothesis that difference in soil type structures *P. subserratum* populations to a greater degree than spatial distance.

The SAMOVA results corroborated the phylogenetic and AMOVA results. F_{CT} reached a plateau when the number of groups (K) was set to 6 (0.7; Table 7). SAMOVA groups 1 and

Table 5 Population pairwise ϕ_{ST} statistics for *Protium subserratum*. See Table 1 and Fig. 1 for locality information. Numbers in boldface are pairwise comparisons that are not significantly different from each other with $\alpha = 0.05$. Above the diagonal is pairwise geographic distance (km).

		S1	S2	S3	S4	S5	S6	S7	S8	S10	S11
Peru, Nanay, brown sand	S1	–	0.556	24.81	104.9	100.5	212.2	259.5	1547	1857	2538
Peru, Nanay, white sand	S2	0.6899	–	24.95	105.4	101.1	212.7	260.1	1547	1856	2538
Peru, Nanay, clay	S3	**−0.0244**	0.7231	–	118.3	114.4	225.7	274.8	1565	1864	2550
Peru, Jenaro, brown sand	S4	**0.0228**	0.6985	**0.00**	–	4.89	107.5	156.6	1584	1935	2598
Peru, Jenaro, white sand	S5	0.6884	0.3757	0.7172	0.6923	–	111.7	160.3	1580	1930	2593
Peru, Rio Blanco, white sand	S6	0.4246	**0.0087**	0.4381	0.4006	0.2727	–	52.68	1617	2010	2562
Peru, Rio Blanco, clay	S7	**0.0296**	0.5708	**0.040**	**−0.00**	0.5606	0.2970	–	1612	2029	2658
Manaus	S8	0.3843	0.3708	0.3941	0.3643	0.4946	0.2319	0.3172	–	885.3	1115
Guyana	S10	0.8265	0.5484	0.8489	0.8340	0.7000	0.5210	0.7514	0.5898	–	848.1
French Guiana	S11	0.9495	0.8161	0.09731	0.9707	0.8600	0.7638	0.9104	0.7711	0.5167	–

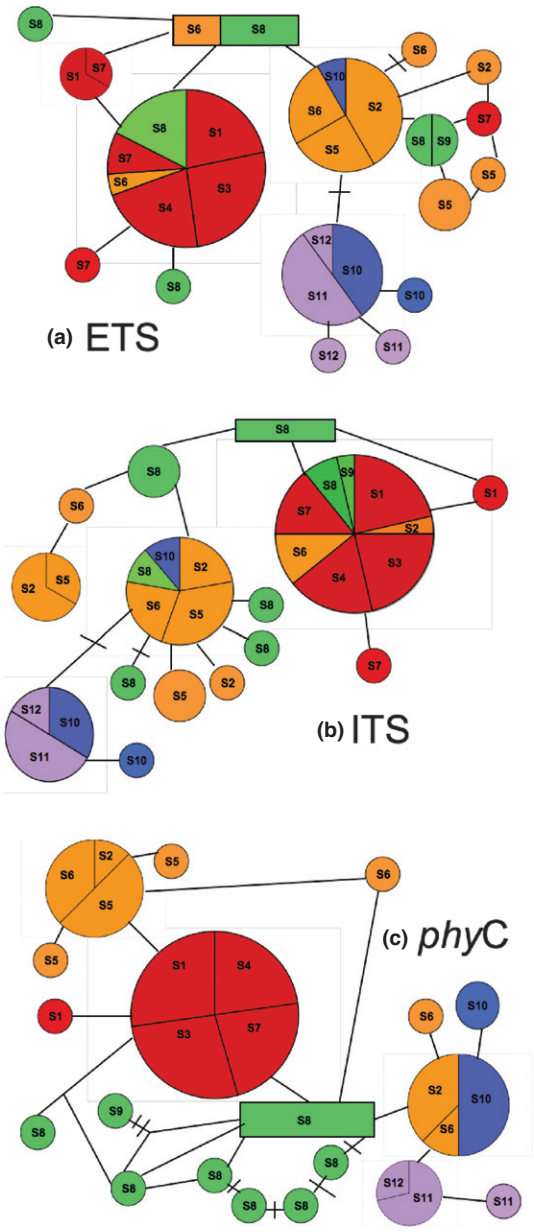

Figure 4 Haplotype networks of *Protium subserratum* for (a) external transcribed spacers (ETS), (b) internal transcribed spacer (ITS) and (c) *phy*C. Purple represents the French Guianan populations, blue represents the Guyanan population, green represents the Manaus population, orange represents Peruvian white-sand populations, and red represents Peruvian clay and terrace populations. The pie charts are labelled with their S1–S12 labels which correspond to the sites in Table 1. The size of the circle approximates the number of individuals for each haplotype. A rectangle means that TCS infers the haplotype to be 'ancestral'. A hatched line means there were two changes between linked haplotypes.

2 corresponded to locations S10 and S11 from eastern Amazon in Guyana and French Guiana, respectively (Fig. 1a). SAMOVA group 3 corresponded to location S8 from Manaus (Fig. 1a). SAMOVA groups 4, 5 and 6 corresponded to locations from western Amazon; group 4 included locations S1, S3, S4 and S7 (clay and terrace locations), and group 5 and 6 included locations S6 plus S2 and S5 (white-sand locations), respectively (Fig. 1b).

DISCUSSION

Phylogeographic patterns in specialist versus generalist tree species

White-sand forests in the Peruvian Amazon occur as widely separated habitat islands that can be quite distant from each other. Thus it was predicted that white-sand specialist trees would exhibit strong patterns of neutral genetic differentiation among populations when compared with edaphic generalist trees. This is because an edaphic generalist that occurs on both white-sand and non-white sand soils could potentially interbreed across soil types reducing the effect of genetic drift among widely separated habitats. Consistent with our expectation, the edaphic specialist *P. alvarezianum* and the edaphic generalist *P. subserratum* presented different patterns of phylogeographic structure, with *P. alvarezianum* exhibiting much higher ϕ_{ST} measures than *P. subserratum*.

Unexpectedly, results for the generalist *P. subserratum* suggested very low overall sharing of haplotypes across different habitat types from the same river basins, with the white-sand populations showing relatively higher haplotypic and nucleotide diversity, and exhibiting lower ϕ_{ST} measures among widely separated white-sand habitat islands. These results suggest that white-sand and non-white-sand populations of *P. subserratum* act effectively as distinct gene pools, with the white-sand populations perhaps having more efficient dispersal mechanisms and/or larger ancestral effective population sizes than the specialist *P. alvarezianum*.

It is possible that differences in pollinator and dispersal assemblages in these two species could be related to the average dispersal distances of pollen or seeds, which would directly influence ϕ_{ST} measures. Both the specialist *P. alvarezianum* and the generalist *P. subserratum* are probably pollinated by small bees (T.M.M., pers. obs.), which are unlikely to be able to fly the 100 + km between the geographically separated white-sand habitats in Fig. 1. Future population genetic studies at a finer scale are under way to tease apart pollen versus seed-mediated gene flow and their relative importance in inter-population gene flow among white-sand habitat islands as well as between habitat types, and how differences in demographic parameters (e.g. bottlenecks, effective population sizes) may affect patterns of genetic differentiation.

Evaluating the relative importance of edaphic heterogeneity and geographic distance in the generalist *P. subserratum*

Soil type and geographic distance both play a role in generating population structure in *P. subserratum*, depending on the spatial scale of analysis (Table 6). At the level of the continent, geographic distance was correlated with neutral genetic

Table 6 Analysis of molecular variance (AMOVA) for populations of *Protium subserratum* grouped by geography (Guianas versus Amazon), river basin (Peru only), or soil type (white sand versus non-white sand, Peru only).

Guyana and French Guiana versus Peru and Manaus			
AMOVA	d.f.	Variance components	% explained
Among groups	1	2.57*	59.55
Among populations, within groups	8	0.78***	18.14
Within populations	54	0.96***	22.31
ϕ_{ST}		0.77	

*$P < 0.05$, ***$P < 0.001$.

	River Basin, within Peru			Soil type, within Peru		
AMOVA	d.f.	Variance components	% explained	d.f.	Variance components	% explained
Among groups	2	−0.34	−22.34	1	1.09*	52.31
Among populations, within groups	4	1.07***	69.84	5	0.19***	9.00
Within populations	34	0.81***	52.50	34	0.81***	38.69
ϕ_{ST}		0.47			0.55	

*$P < 0.005$, ***$P < 0.001$.

Table 7 Comparison of fixation indices corresponding to the groups of populations detected by spatial analysis of molecular variance (SAMOVA) in *Protium subserratum* based on two nuclear DNA regions: internal transcribed spacer–external transcribed spacer (ITS–ETS) and *phy*C. K = 6 groups is highlighted in boldface because it corresponds to the best supported model.

No. of groups (*K*)	Groups	F_{SC}	F_{ST}	F_{CT}
2	[S10, S11] and [S1, S2, S3, S4, S5, S6, S7, S8]	0.44	0.77	0.59
3	[S10, S11], [S1, S3, S4, S7, S8] and [S2, S5, S6]	0.30	0.70	0.59
4	[S10, S11], [S1, S3, S4, S7], [S8] and [S2, S5, S6]	0.20	0.68	0.61
5	[S10, S11], [S1, S3, S4, S7], [S8], [S2, S6] and [S5]	0.10	0.68	0.65
6	**[S10], [S11], [S1, S3, S4, S7], [S8], [S2, S6] and [S5]**	**−0.05**	**0.67**	**0.69**
7	[S10], [S11], [S1, S3, S4, S7], [S8], [S2], [S6] and [S5]	−0.10	0.67	0.71
8	[S10], [S11], [S1, S3, S4], [S7], [S8], [S2], [S6] and [S5]	−0.11	0.66	0.70
9	[S10], [S11], [S1], [S3, S4], [S7], [S8], [S2], [S6] and [S5]	−0.20	0.65	0.70

F_{SC}, proportion of genetic variance between populations within groups.
F_{ST}, proportion of genetic variance between groups and populations overall.
F_{CT}, proportion of genetic variance among groups.

divergences between Peruvian, Brazilian and Guianan populations. Within Peru, geographic distance among the three river basins did not appear to be a significant barrier for populations of *P. subserratum* within the same habitat type (Tables 5 & 6, Fig. 4). On the other hand, there was significant population structure between white-sand and non-white-sand soils, even within the same river basin (Tables 5 & 6, Fig. 4). This suggests that at the scale of the northern Peruvian Amazon, geographic distance is not a strong factor shaping population genetic structure while soil type is extremely important, consistent with the idea that adaptation to soil type may play a fundamental role in lineage divergence in these trees. Whereas soil type was expected to be very important given the morphological differences in leaflets between white-sand and non-white-sand populations, the complete lack of geographic structure among the three Peruvian river basins for both white-sand and non-white-sand populations is surprising. Because white-sand islands are small habitats compared with terrace and clay habitats in the Peruvian Amazon, as well as geographically isolated from one another, we predicted low levels of haplotype sharing among white-sand populations due to population bottlenecks, which does not appear to be the case. Instead, the neutral differentiation imposed by geographic distance is being balanced by significant amounts of gene flow or the populations are maintaining high levels of ancestral polymorphisms.

White-sand populations of *P. subserratum* have a large number of haplotypes compared with clay and terrace populations (Table 3, Fig. 4). In a study of two edaphic races of Jeffrey pine trees (*Pinus jeffreyi*) inhabiting serpentine and non-serpentine soils, the more extreme and island-like serpentine habitat specialist population had less genetic diversity than populations on non-serpentine soils (Furnier & Adams, 1986). Population size is a strong correlate of genetic diversity (Linhart & Grant, 1996), and Jeffrey pine trees occur at low densities in serpentine areas, compared to non-serpentine areas where they often exhibit high densities in a much larger geographic area. *Protium subserratum*, however, is found in much higher densities in white-sand forest than clay and terrace soils (Fine *et al.*, 2005, 2010). Even though terrace and clay soils cover much more relative area in the Peruvian Amazon, the flora found on these soils is much more diverse, and *P. subserratum* is neither extremely common nor found at high densities where present (Fine *et al.*, 2005; Fine & Kembel, 2011). Populations on clay and terrace soils therefore may be inbred or perhaps clay and terrace populations are very recent migrants from other source populations (perhaps representing a genetic bottleneck). To further test these different possibilities, we are currently collecting larger sample sizes of the populations from different soil types and developing microsatellites. This will allow us to assess migration more quantitatively and estimate population contraction versus expansion (Kuhner, 2009), allowing us to test hypotheses about the relative size and migration rates of different *P. subserratum* populations in different habitats.

Historical biogeography of *P. subserratum*

The Bayesian phylogeny of *P. subserratum* shows eastern morphotypes as sister to the western and central Amazonian populations (Fig. 3). Within the Peruvian Amazonian clade, clay and terrace populations showed very little genetic variation (Table 3, Fig. 3) and formed an unresolved clade recently derived from within the grade including the white-sand + Manaus populations. This putative recent switch to clay and terrace soils in the western Amazon is consistent with the older age of white-sand habitats and the origin of more fertile soil types that has occurred post-Andean uplift in the mid to late Miocene in the western Amazon (Hoorn, 1993; Frasier *et al.*, 2008). The Manaus population at 'S8' and 'S9' includes *P. subserratum* trees from both sandy and clay soils plus one individual from the extreme white-sand forest at S9 that had similar leaflet morphology to Peruvian white-sand morphotypes. All individuals sampled from S8 appeared similar in their leaflet morphology to *P. subserratum* populations in Peru on terrace and clay soils (Table 2). Together, the Manaus + Peru terrace and clay populations form one single clade with 0.51 posterior probability. However, the Manaus population showed a great diversity of haplotypes, while the Peruvian clay and terrace populations did not. Thus Peruvian clay and terrace populations may be younger and may have arrived only recently from the central Amazon.

One advantage of our study system is that all individuals sampled can be put safely into a broader phylogenetic context, given that we are currently amassing all Protieae samples for a large phylogenetic study (Fine *et al.*, 2005; Daly *et al.*, in press). In addition, the fact that there are multiple fossils in the Burseraceae (including one from the Protieae) means that we will soon be able to use a fossil-calibrated tree for our DNA sequence data to estimate minimum divergence dates among species, and perhaps even population-level divergences within *P. subserratum*. This will provide us with greater insight into the timing of colonization events into different soil types and geographic regions.

Broader implications: can edaphic specialization lead to speciation?

Specialization to divergent soil types has occurred twice in this small clade of four species of *Protium* sect. *Papilloprotium* (Fig. 3). The most recent case is within the edaphic generalist *P. subserratum*, where the evidence presented above suggests that populations on clay and terrace soils diverged from populations on white-sand soils. The morphological differences between white-sand and clay/terrace populations are unlikely to be due to phenotypic plasticity because there are strong differences in population structure between these two morphotypes, even when the edaphically divergent populations occur only a few kilometres apart. Instead, white-sand and non-white-sand lineages in the phylogeny and haplotype networks correspond to morphological differences in leaflet pubescence, margin teeth and thickness (Table 2). That the two morphotypes are not reciprocally monophyletic (Fig. 3) and even share haplotypes (Fig. 4) is consistent with a scenario of incipient parapatric speciation across strong edaphic gradients (similar to Savolainen *et al.*, 2006) with incomplete lineage sorting. Extreme differences in soil fertility and substrate may cause strong natural selection for divergent strategies involving allocation tradeoffs. If the gradient is strong enough, speciation can occur even in the face of gene flow between populations on the opposite soil type (Endler, 1977). The mechanism by which this 'divergence with gene flow' is thought to occur is that intermediate phenotypes generated by crosses between parental populations on divergent habitat types would be less fit in either habitat type (Endler, 1977; Nosil, 2008). In the case of *Protium*, tradeoffs in growth rate and defence investment have been found in congeners specialized to white-sand and clay soils (Fine *et al.*, 2006). These physiological tradeoffs may contribute to habitat-mediated speciation by conferring post-zygotic isolation if intermediate phenotypes did not receive adequate defence against herbivores in white-sand forests and if they cannot compete in high-resource clay and terrace soils. Reciprocal transplant experiments and hand-pollination crosses among white-sand and non-white-sand morphotypes are under way, which will allow us to assess parental and hybrid fitness across habitat types.

Alternatively, speciation could have already happened (with or without allopatry), with subsequent dispersal of edaphic

specialists to currently parapatric distributions on adjacent areas of white-sand, clay and terrace habitats thus allowing introgression between morphotypes (similar to red and silver maples; Saeki *et al.*, 2011). This implies that the populations of *P. subserratum* in the western Amazon currently inhabiting white-sand forests may have had wider distributions during Pleistocene or Pliocene glacial maxima, and their geographic range may previously have included a wide variety of soil types. Warming and the re-expansion of contiguous aseasonal lowland rain forest in the region (Bush, 1994; Haffer & Prance, 2001) may have resulted in more recent competitive exclusion of slower growing, highly defended trees such as the white-sand morphotypes. In addition, the expectation that reciprocal monophyly be present between white-sand and clay morphotypes for them to be considered 'good species' may not be warranted, given that introgression among morphologically divergent, habitat specialist species has been reported for other trees (*Quercus*, Petit *et al.*, 2004; *Betula*, Thórsson *et al.*, 2010).

CONCLUSIONS

We have found substantial phylogeographic structure in both the white-sand specialist *P. alvarezianum* and the edaphic generalist *P. subserratum*. For the specialist *P. alvarezianum*, high ϕ_{ST} values among populations are consistent with the likely difficulty of achieving gene flow among geographically isolated habitat islands. Although geographic distance plays a role in shaping continent-wide phylogeographic patterns in the edaphic generalist species *P. subserratum*, it does not play as important a role at a local geographic scale as it does for the edaphic specialist *P. alvarezianum*. Within Peru, soil type was an important factor differentiating white-sand from non-white-sand populations. Moreover, phylogenetic and phylogeographic analyses suggest that these populations act effectively as distinct gene pools, with little sharing of haplotypes. These results are consistent with the hypothesis that *P. subserratum* may in fact be a complex of unrecognized species. Whether this pattern is due to incipient speciation via habitat specialization or past speciation with secondary contact remains to be fully elucidated. In either case, the edaphic heterogeneity present in the western Amazon as a result of its complicated geological history is strongly implicated in driving and maintaining population differentiation in these lowland rain forest trees. Given the lack of topographic complexity found throughout the lowland Amazon Basin, we speculate that habitat heterogeneity and the spatial arrangement of habitats may be a more general driver of population-level differences than geographic distance per se for influencing genetic structure and speciation in Amazonian species.

ACKNOWLEDGEMENTS

We thank the Peruvian Ministry of Natural Resources (INRENA) for permission to conduct this study, and the Instituto de Investigaciones de la Amazonía Peruana (IIAP) for logistical support and permission to work in and around the Estacíon Allpahuayo. We would like to thank Nallarett Dávila and Julio Sanchez for field assistance, and Krista Beckley for lab assistance. Cristina Martinez-Habibe assisted with maps and Anand Varna took the photograph in Fig. 2. We thank Robin Foster and Corine Vrisendorp and the Field Museum of Natural History in Chicago for facilitating travel to the Reserva Matsés. Krista McGuire arranged fieldwork and logistics in Guyana. Chris Baraloto arranged fieldwork and logistics in French Guiana. We thank Rauri Bowie for advice and help with analyses. Earlier drafts of this manuscript were improved by comments from Ricardo Perreira, Chris Dick, Stephen Cavers, Michael Heads, Andreas Chavez and one anonymous referee. This research was supported by National Science Foundation grants DEB-0919567 (P.V.A.F. and D.C.D.) and OISE-0402061 (P.V.A.F.) as well as the Michigan Society of Fellows and the UC Berkeley Committee on Research Junior Faculty Research Fund.

REFERENCES

Anderson, A.B. (1981) White-sand vegetation of Brazilian Amazonia. *Biotropica*, **13,** 199–210.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Baldwin, B.G. & Markos, S. (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution*, **10,** 449–463.

Baraloto, C., Rabaud, S., Molto, Q., Blanc, L., Fortunel, C., Hérault, B., Dávila, N., Mesones, I., Ríos, M., Valderrama, E. & Fine, P.V.A. (2011) Disentangling stand and environmental correlates of aboveground biomass in Amazonian forests. *Global Change Biology*, **17,** 2677–2688.

Bush, M.B. (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21,** 5–17.

Caron, H., Dumas, S., Marque, G., Messier, C., Beou, E., Petit, R.J. & Kremer, A. (2000) Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology*, **9,** 1089–1098.

Cavers, S., Navarro, C. & Lowe, A.J. (2003) Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology*, **12,** 1451–1460.

Clayton, J.W., Fernando, E.S., Soltis, P.S. & Soltis, D.E. (2007) Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. *International Journal of Plant Sciences*, **168,** 1325–1339.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9,** 1657–1659.

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

Daly, D.C. & Fine, P.V.A. (2011) A new Amazonian section of *Protium* (Burseraceae) including both edaphic specialist and generalist taxa. Studies in Neotropical Burseraceae XVI. *Systematic Botany*, **36,** 939–949.

Daly, D.C., Fine, P.V.A. & Martínez-Habibe, C. (in press) Burseraceae: a model for studying the Amazon flora. *Rodriguésia*.

Day, P.R. (1965) Particle fractionation and particle-size analysis. *Methods of soil analysis, Part 1* (ed. by C.A. Black), pp. 545–567. American Society of Agronomy, Madison, WI.

Dick, C.W. & Heuertz, M. (2008) The complex biogeographical history of a widespread tropical tree species. *Evolution*, **62,** 2760–2774.

Dick, C.W., Abdul-Salim, K. & Bermingham, E. (2003) Molecular systematics reveals cryptic Tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist*, **162,** 691–703.

Dick, C.W., Bermingham, E., Lemes, M.R. & Gribel, R. (2007) Extreme long-distance dispersal of the lowland tropical rainforest tree *Ceiba pentandra* L. (Malvaceae) in Africa and the Neotropics. *Molecular Ecology*, **16,** 3039–3049.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2010) *Geneious v5.3.3*. Biomatters Ltd, Auckland, New Zealand. Available at: http://www.geneious.com.

Duivenvoorden, J.F. & Duque, A.J. (2010) Composition and diversity of northwestern Amazonian rainforests in a geoecological context. *Amazonia, landscape and species evolution: a look into the past* (ed. by C. Hoorn and F.P. Wesselingh), pp. 360–372. Wiley-Blackwell, Chichester, UK.

Dupanloup, I., Schneider, S. & Excoffier, I. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11,** 2571–2581.

Endler, J.A. (1977) *Geographic variation, speciation and clines*. Princeton University Press, Princeton, NJ.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial restriction sites. *Genetics*, **131,** 479–491.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin version 3: an integrated software package for population genetics data analysis. *Evolutionary Biology Online*, **1,** 47–50.

Fine, P.V.A. & Kembel, S.W. (2011) Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography*, **34,** 552–565.

Fine, P.V.A. & Ree, R.H. (2006) Evidence for a time-integrated species–area effect on the latitudinal gradient in tree diversity. *The American Naturalist*, **168,** 796–804.

Fine, P.V.A., Mesones, I. & Coley, P.D. (2004) Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, **305,** 663–665.

Fine, P.V.A., Daly, D.C., Villa, F.G., Mesones, I. & Cameron, K.M. (2005) The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the Western Amazon. *Evolution*, **29,** 1464–1478.

Fine, P.V.A., Miller, Z.J., Mesones, I., Irazuzta, S., Appel, H.M., Stevens, M., Saaksjarvi, I., Schultz, J. & Coley, P.D. (2006) The growth–defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology*, **87,** S150–S162.

Fine, P.V.A., García-Villacorta, R., Pitman, N.C.A., Mesones, I. & Kembel, S.W. (2010) A floristic study of the white-sand forests of Peru. *Annals of the Missouri Botanical Garden*, **97,** 283–305.

Frasier, C.L., Albert, V.A. & 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, and Evolution*, **8,** 44–57.

Furnier, G.R. & Adams, W.T. (1986) Geographic patterns of allozyme variation in Jeffrey pine. *American Journal of Botany*, **73,** 1009–1015.

Gee, G.W. & Bauder, J.W. (1979) Particle size analysis by hydrometer: a simplified method for routine textural analysis and a sensitivity test of measurement parameters. *Soil Science Society of America Journal*, **43,** 1004–1007.

Gentry, A.H. (1986) Endemism in tropical versus temperate plant communities. *Conservation biology: the science of scarcity and diversity* (ed. by M. Soulé), pp. 153–181. Sinauer, Sunderland, MA.

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

Haffer, J. & Prance, G.T. (2001) Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Amazoniana*, **16,** 579–607.

Hammond, D.S. (2005) Biophysical features of the Guiana Shield. *Tropical forests of the Guiana Shield: ancient forests in a modern world* (ed. by D. Hammond), pp. 15–194. CABI Publishing, Wallingford, UK.

Hickerson, M.J., Carstens, B.C., Cavender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F. & Yoder, A.D. (2010) Phylogeography's past, present and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54,** 291–301.

Hoorn, C. (1993) Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: results of a palynostratigraphic study. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **105,** 267–309.

Hoorn, C. (1994) An environmental reconstruction of the palaeo-Amazon river system (Middle–Late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **112,** 187–238.

Hubbell, S.P., He, F., Condit, R., Borda-de-Agua, L., Kellner, J. & ter Steege, H. (2008) How many tree species are there in the Amazon and how many of them will go extinct? *Proceedings of the National Academy of Sciences USA*, **105,** 11498–11504.

Janzen, D.H. (1974) Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica*, **6,** 69–103.

Knowles, L.L. & Maddison, W.P. (2002) Statistical phylogeography. *Molecular Ecology*, **11,** 2623–2635.

Kuhner, M.K. (2009) Coalescent genealogy samplers: windows into population history. *Trends in Ecology and Evolution*, **24,** 86–93.

Lemes, M.R., Gribel, R., Proctor, J. & Grattapaglia, D. (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. *Molecular Ecology*, **12,** 2875–2883.

Levin, D.A. (2004) Ecological speciation: crossing the divide. *Systematic Botany*, **29,** 807–816.

Linhart, Y.B. & Grant, M.C. (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27,** 237–277.

Maddison, D.R. & Maddison, W.P. (2002) *MacClade 4.05: analysis of phylogeny and character evolution. Version 4.05*. Sinauer, Sunderland, MA.

Malcomber, S.T. (2002) Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution*, **56,** 42–57.

Moritz, C., Patton, J., Schneider, C. & Smith, T.B. (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31,** 533–563.

Nosil, P. (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17,** 2103–2106.

Novick, R.R., Dick, C., Lemes, M.R., Navarro, C., Caccone, A. & Bermingham, E. (2003) Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Molecular Ecology*, **12,** 2885–2893.

Nylander, J.A.A. (2004) *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.

Pennington, R.T. & Dick, C.W. (2010) Diversification of the Amazonian flora and its relation to key geological and environmental events: a molecular perspective. *Amazonia, landscape and species evolution: a look into the past* (ed. by C. Hoorn and F.P. Wesselingh), pp. 374–385. Wiley-Blackwell, Chichester, UK.

Petit, R.J., Bodénès, C., Ducousso, A., Roussel, G. & Kremer, A. (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161,** 151–164.

Pitman, N.C.A., Mogollon, H., Dávila, N., Ríos, M., García-Villacorta, R., Guevara, J., Baker, T.R., Monteagudo, A., Phillips, O.L., Vásquez-Martínez, R., Ahuite, M., Aulestia, M., Cardenas, D., Cerón, C.E., Loizeau, P.-A., Neill, D.A., Núñez, P., Palacios, W.A., Spichiger, R. & Valderrama, E. (2008) Tree community change across 700 km of lowland Amazonian forest from the Andean foothills to Brazil. *Biotropica*, **40,** 525–535.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19,** 1572–1574.

Saeki, I., Dick, C.W., Barnes, B.V. & Murakami, N. (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*A. saccharinum* L.): impacts of habitat specialisation, hybridization and glacial history. *Journal of Biogeography*, **38,** 992–1005.

Savolainen, V., Anstett, M.-C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M.V., Powell, M.P., Springate, D., Salamin, N. & Baker, W.J. (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441,** 210–213.

ter Steege, H., Pitman, N.C.A., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., Molino, J.-F., Prévost, M.-F., Spichiger, R., Castellanos, C., von Hildebrand, P. & Vasquez, R. (2006) Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, **443,** 444–447.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphisms. *Genetics*, **123,** 585–595.

Thórsson, Æ.Th., Pálsson, S., Lascoux, M. & Anamthawat-Jónsson, K. (2010) Introgression and phylogeography of *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland inferred from cpDNA haplotype variation. *Journal of Biogeography*, **37,** 2098–2110.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, **299,** 241–244.

Weeks, A., Daly, D.C. & Simpson, B.B. (2005) The phylogenetic history and historical biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution*, **35,** 85–101.

Wright, S.J. (2002) Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, **130,** 1–14.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Collection numbers and GenBank accession numbers for the individuals in the study.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

Paul Fine and his laboratory group study the ecology and evolution of habitat specialization in Amazonian trees, with a special focus on *Protium* (Burseraceae) as a model system.

Author contributions: P.V.A.F. designed the project, I.M., C.E.A.B. and P.V.A.F. conducted the fieldwork, T.M.M., H.F.C., C.E.A.B., F.Z. and P.V.A.F. performed the lab work and data analyses, and P.V.A.F., F.Z., T.M.M. and D.C.D. wrote the manuscript.

Editor: Malte Ebach

This Special Issue was motivated by research undertaken by partners in the SEEDSOURCE consortium (http://www.seedsource.net), a collaborative project funded by the European Commission under FP6. The project aimed to establish the evolutionary basis and geographic distribution of genetic variation in a range of important Neotropical tree species for better sourcing and control of germplasm and forest conservation.