



**HAL**  
open science

## **Phylogeography and niche modelling of the relict plant *Amborella trichopoda* (Amborellaceae) reveal multiple Pleistocene refugia in New Caledonia**

Valérie Poncet, François Munoz, Jérôme Munzinger, Yohan Pillon, Céline Gomez,  
Marie Couderc, Christine Tranchant-Dubreuil, Serge Hamon, Alexandre de  
Kochko

### ► To cite this version:

Valérie Poncet, François Munoz, Jérôme Munzinger, Yohan Pillon, Céline Gomez, et al.. Phylogeography and niche modelling of the relict plant *Amborella trichopoda* (Amborellaceae) reveal multiple Pleistocene refugia in New Caledonia. *Molecular Ecology*, 2013, 22 (24), pp.6163-6178. <10.1111/mec.12554>. <ird-04996197>

**HAL Id: ird-04996197**

**<https://ird.hal.science/ird-04996197v1>**

Submitted on 18 Mar 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**Phylogeography and niche modelling of the relict plant *Amborella trichopoda* (Amborellaceae)  
reveal multiple Pleistocene refugia in New Caledonia**

Valérie Poncet<sup>1</sup>, François Munoz<sup>2</sup>, Jérôme Munzinger<sup>2,3</sup>, Yohan Pillon<sup>4,5</sup>, Céline Gomez<sup>1,4</sup>, Marie Couderc<sup>1</sup>,  
Christine Tranchant-Dubreuil<sup>1</sup>, Serge Hamon<sup>1</sup>, Alexandre de Kochko<sup>1</sup>

<sup>1</sup> IRD, UMR DIADE, BP 64501, 34394 Montpellier cedex 5, France

<sup>2</sup> UM2 and IRD, UMR AMAP, Boulevard de la Lironde, PS 2, 34398 Montpellier, France

<sup>3</sup> IRD, UMR AMAP, Herbarium NOU, 98848 Nouméa, New Caledonia

<sup>4</sup> IRD, UMR DIADE, 98848 Nouméa, New Caledonia

<sup>5</sup> Tropical Conservation Biology and Environmental Science, University of Hawai'i at Hilo, 200 West Kawili St.,  
Hilo, Hawai'i 96720, USA

*Keywords:* *Amborella trichopoda*, niche modelling, genetic variation, habitat suitability, refugia, last glacial  
maximum

Corresponding author: Valérie Poncet, IRD, UMR DIADE, BP 64501, 34394 Montpellier cedex 5, France. UMR  
DIADE, Tel: +33 (0)467416251, Fax: +33 (0)467416222, valerie.poncet@ird.fr

Running title: *Amborella* phylogeography and LGM refugia

## Abstract

*Amborella trichopoda* Baill. (Amborellaceae, Amborellales), the sole living member of the sister group to all other extant Angiosperms, is endemic to New Caledonia. We addressed the intraspecific phylogeography of *Amborella* by investigating whether its present population genetic structure could be related to its current and past habitats.

We found moderate range-wide genetic diversity based on nuclear microsatellite data, and detected four well-differentiated, geographically distinct genetic groups using Bayesian clustering analyses.

We modeled the ecological niche of *Amborella* based on current climatic and environmental conditions. The predictive ability of the model was very good throughout the Central East mainland zone, but *Amborella* was predicted in the northern part of the island where this plant has not been reported. Furthermore, no significant barrier was detected based on habitat suitability that could explain the genetic differentiation across the area. Conversely, we found that the main genetic clusters could be related to the distribution of suitable habitat at the last glacial maximum (LGM, ca. 21 000 years BP), when *Amborella* experienced a dramatic 96.5% reduction in suitable area. At least two lineages survived in distinct putative refugia located in the Massif des Lèvres and in the vicinity of Mount Aoupinié.

Our findings finally confirmed the importance of LGM rainforest refugia in shaping the current intra- and inter-specific diversity in New Caledonian plants, and revealed the possibility of an as-yet unreported refugium. The combination of niche modelling and population genetics thereby offered novel insight into the biogeographic history of an emblematic taxon.

## Background

Combining insight from population genetics with analyses of species distribution and landscape history leads to a better understanding of how historical processes have shaped species genetic diversity overall (Cordellier, Pfenninger, 2009; McKay *et al.*, 2010; Rebernig *et al.*, 2010). Various ecological processes, such as local habitat suitability (Wellenreuther *et al.*, 2011) and dispersal limitation between populations in fragmented environments, determine the distribution of genetic diversity over large geographic areas. These processes influence species distributions over time and drive their evolutionary history.

Past climatic cycles, particularly the last glacial maximum (LGM) around 23 000-18 000 years BP, have had a major impact on present day levels and distribution of genetic diversity of many species (Hewitt, 2004). Most previous surveys have focused on temperate terrestrial flora and fauna in North America and Europe, where many species survived within refugia during glacial maxima, usually at lower latitudes (Provan, Bennett, 2008; Taberlet *et al.*, 1998). Extensive palaeoecological records and phylogeographic studies in temperate regions have yielded knowledge on refugia locations and colonization histories of many taxa (Feliner, 2011). Conversely, the intensities and impact of Pleistocene climate oscillations on the structure of biodiversity in tropical regions have only been the focus of recent interest (e.g. Cavers, Dick, 2013; Haffer, Prance, 2001; Hardy *et al.*, 2013). The Pleistocene refuge hypothesis in tropical regions mainly implies that, in a globally drier and cooler setting, small areas benefited from locally wetter conditions due to topographical heterogeneity, which sheltered relictual rainforests (Haffer, 1969; Haffer, Prance, 2001). In Africa, for instance, most of the postulated glacial forest refugia distribution areas are assumed to be associated with mountainous or riparian areas (Maley, 1996; Robbrecht, 1996), while a more complex situation has been described in South America, with refugia areas also located in lowland Amazonia (Haffer, 1997; Haffer, Prance, 2001). However, the palaeoenvironmental context of most tropical Pacific islands is still poorly understood (Keppel *et al.*, 2012), although studies of terrestrial pollen diagrams have indicated that tropical rainforests also experienced episodes of range contraction, while the humid vegetation later expanded at the Pleistocene-Holocene

transition (Hope *et al.*, 2004) until the arrival of humans ca. 3 000 years and the expansion of secondary vegetation (Stevenson *et al.*, 2001).

New Caledonia is a small archipelago in the South West Pacific covering 19 110 km<sup>2</sup> and is considered as a terrestrial biodiversity hotspot because of its rich endemic and threatened flora and fauna (Myers *et al.*, 2000). The topography of the main island (20-23° S and 164-167° E) is complex (e.g. Gomez *et al.*, 2010) and the tropical to subtropical climate is influenced by prevailing southeast trade winds. The archipelago underwent complex geological and environmental changes. The main island, as part of Zealandia, separated from Australia and drifted northeastwards in Late Cretaceous. It then underwent obduction under the Loyalty Arc and re-emerged only ca. 37 Mya, as evidenced by geological data (Neall, Trewick, 2008; Pelletier, 2006) in concordance with phylogenetic data (Cruaud *et al.*, 2012; Grandcolas *et al.*, 2008; Pillon, 2012). The emblematic rainforest shrub *Amborella trichopoda*—the sole member of a lineage dating back from over 230 My—probably colonized New Caledonia later by dispersal or could have survived in ephemeral nearby areas until New Caledonia's emersion. More recent climatic variations due to glacial episodes have probably further shaped the current biodiversity patterns. However, while much emphasis has been put on Cretaceous and Tertiary events, little attention has been paid to the effects of Pleistocene climatic fluctuations on the New Caledonian flora. An analysis of New Caledonian palms suggested rain forest refugia in the wetter and elevated east coast and southern areas during the last glacial maximum (Pintaud *et al.*, 2001), but no recent study has further confirmed this observation.

The effect of past environmental drivers on the current intra-specific diversity in New Caledonian species remains underexplored, although the high species endemism on the island suggests a combination of specialized ecological niches, refugia formation and long periods of isolation. Intraspecific genetic variation patterns could thereby help distinguish the effect of ancient events after re-emergence and of glacial refugia from the constraints imposed by the contemporary environment. Apart from current island biogeography knowledge, species distribution models (SDM, Elith, Leathwick, 2009) are useful for characterizing habitat preferences that could influence the demography and ability of the species to disperse throughout the

available space. Moreover, paleoclimate models allow SDM projection under past climatic contexts and assessment of possible changes in species distributions over time. Landscape genetics further incorporate knowledge on habitat suitability and diversity to assess the impact of ecological processes in the current genetic structure (Holderegger *et al.*, 2010). The combination of genetic and ecological data has recently received increased interest to identify refugia during the late-Pleistocene last glacial maximum (LGM) (Bisconti *et al.*, 2013; Qi *et al.*, 2012; Qiu *et al.*, 2009).

Few studies have investigated the genetic structure of New Caledonian native plants. Kettle *et al.* (2007) used nuclear microsatellites to analyze the population genetics of the rare micro-endemic *Araucaria nemorosa* de Laub. and the coastal *A. columnaris* Hook. Bottin *et al.* (2005) also used microsatellites to survey the heavily logged New Caledonian sandalwood (*Santalum austrocaledonicum* Vieill.), which is almost exclusively restricted to dry forests and limestone substrates. Finally, Kurata *et al.* (2008) used plastid markers to study populations of the carnivorous plant *Nepenthes vieillardii* Hook.f., which is confined to ultramafic substrates (serpentine). These studies are biased towards rare species and none of them addressed rainforest plants.

*Amborella trichopoda* Baill. is a New Caledonian endemic plant which became emblematic after it was found to be the only current member of the sister group to all other living flowering plants (Mathews, Donoghue, 1999; Moore *et al.*, 2007; Qiu *et al.*, 1999). It is a rainforest shrub that is found throughout most of the central part of the main island on non-ultramafic substrates. There has been little focus on understanding the genetic diversity and structure of *Amborella trichopoda* over its natural distribution area (Pillon, Munzinger, 2005).

Here we investigated the phylogeographical history of *Amborella* with a multidisciplinary approach. We genotyped populations with 10 microsatellite markers (Poncet *et al.*, 2012) and investigated the potential *Amborella* ecological niche. Our sampling encompassed several elevations and geographical regions across the *Amborella* distribution range to address the following questions: (i) How diverse and differentiated are *Amborella* populations? (ii) What is the current geographical range in New Caledonia associated with a good habitat suitability for *Amborella*? (iii) Whether and how have recent glacial epochs and present-day environmental conditions shaped the current genetic diversity pattern?

## Material and Methods

### *Amborella* biology

*Amborella trichopoda* is a large evergreen shrub or small tree that grows to 7(-12) m height in the understorey. It is mainly found on slopes in rainforests on acidic substrates in the central part of the island, at 100–900 m elevation (Jeremie, 1982). This part of the island is hilly and the populations appear to be isolated and consist of few dozens of individuals with an estimated density of 433 tree per ha (Col d'Amieu) (from Thien *et al.*, 2003). *Amborella* is diploid ( $2n=26$ , Oginuma *et al.*, 2000) and dioecious (Endress, Igersheim, 2000). Flowering is synchronous within a population, and flowers are both insect- and wind-pollinated. A wide variety of insects pollinates the flowers but more particularly beetles (*Neoadelium fauveli*) dwelling in the forest litter (Thien *et al.*, 2003). The small (5 x 7 mm) drupaceous fruit has a fleshy mesocarp with red juice that could potentially be eaten by a wide range of small animals such as birds or reptiles (Endress, Igersheim, 2000).

### *Amborella* distribution

We gathered information on georeferenced locations of *Amborella* from a comprehensive range of available data, namely voucher specimens, census in permanent plots and a set of populations sampled for genetic studies (Table S1, Supporting Information). The voucher specimen data for *Amborella trichopoda* were collected from several herbarium databases: Sonnerat (P), Viot (NOU), Tropicos (MO), Bishop Museum (BISH) and GBIF ([www.gbif.org](http://www.gbif.org)). All voucher records with obvious geocoding errors were discarded. When the coordinates were provided but information about the geographical reference system was missing, the most appropriate system was chosen according to the consistency with respect to the map location and locality information. If data had no coordinates but information on the locality was accurate enough, coordinates were attributed using the MacKee collection Gazetteer (<http://www2.mnhn.fr/novcal/>) corrected in WGS84. Conversely, data with ambiguous or vague locality information were excluded from the study. The final sample

included 90 voucher specimens from 70 distinct locations (Table S1A, Supporting Information). We added 29 recent GPS-located points extracted from the NC-PIPPN database (New Caledonian Plant and Inventory Permanent Plot Network) (Table S1B, Supporting Information). Finally, we also selected and sampled 18 populations for genetic analysis (Table S1C, Supporting Information).

#### *Population sampling and microsatellite genotyping*

For population genetic analyses, we first selected 12 locations representative of the *Amborella* distribution (Fig. 1), and then we included, when possible, replicated populations over an elevation gradient for some of these first-level locations (denoted A, B or C after the name of the first-level location, Table 1). Data related to herbarium specimens (IRD, Noumea, New Caledonia) indicated two flowering periods, one from early March to late May and the other in August (Jeremie, 1982). Samples were collected from March to May 2009 but not all shrubs were flowering and/or carrying fruits, so sex information could only be partially recorded.

Leaves from 305 specimens were collected and dried in silica gel. Ten specific expressed sequence tag (EST)-SSR primer pairs developed by Poncet et al. (2012) were used to amplify DNA from all sampled individuals: Amt-23b, Amt-30, Amt-12b, Amt-36, Amt-41, Amt-43, Amt-07b, Amt-13b, Amt-11, and Amt-16 as described in Poncet et al. (2012). In addition, the individual used for complete genome sequencing by the *Amborella* Genome Sequencing Project (*Amborella* Genome Project, submitted), growing at the arboretum of the University of California at Santa Cruz, CA USA, was also genotyped. The collection data for this individual do not precisely place it within a known population from New Caledonia.

Amplified fragments were separated by capillary electrophoresis on an ABI 3130 Genetic Analyzer (Applied Biosystems, Carlsbad, California, USA), and fragment sizes were assessed using GeneMapper 3.21 (Applied Biosystems) according to a size standard GS-500 LIZ (Applied Biosystems). Automated allele calling was visually checked for each sample by two independent scorings (two persons). Unusual peak patterns were considered

as due to missing data. We retained only samples with < 10% missing information (i.e. samples correctly genotyped for at least 9 of the 10 markers). The final number of genotyped samples was 237.

#### *Genetic diversity, population differentiation and Hardy-Weinberg equilibrium*

Genetic diversity within populations was assessed in terms of allele frequencies, expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_o$ ) and allelic richness using the Fstat program (vers. 2.9.3) (Goudet, 2001). For each locus in each population or site,  $F_{IS}$  was calculated and tested for significant deviations from zero using Fstat (vers. 2.9.3) (Goudet, 2001), and significant departures from Hardy–Weinberg equilibrium were tested using PowerMarker version 3.25 (Liu, Muse, 2005). Inferences of population structure may be influenced by departure from neutrality in EST-SSR since expressed sequence tags (ESTs) are sequenced portions of complementary DNA copies of mRNA and represent part of the transcribed portion of the genome in given conditions (Excoffier, Lischer, 2010). We calculated the  $F_{ST}$  coefficients from inferred groups (see below) for each locus to identify outlier loci that could be influenced by directional selection. Loci were detected under selection by examining the joint distribution of  $F_{ST}$  and heterozygosity under a hierarchical island model implemented in Arlequin 3.5 software (Excoffier, Lischer, 2010).

#### *Cluster analyses*

The population structure was assessed via a Bayesian inference model (Structure program vers. 2.3.3) (Pritchard *et al.*, 2000). It was run with the admixture ancestral model, while ignoring population affiliations when assigning individuals, and assuming correlated allele frequencies among populations. The number of clusters  $K$  was inferred using 20 independent runs of 100 000 Markov chain Monte Carlo (MCMC) iterations after a burn-in period of 100 000, implemented for each  $K$  ( $K$  ranging from 1 to 18). The most likely  $K$  value was evaluated following Evanno *et al.* (2005). Once the most likely  $K$  value was determined, the proportion of

membership in each of the K clusters was determined for each individual from the run with the highest posterior probability.

The genetic structure of species populations was also inferred by using the Geneland package (vers. 4.0) (Guillot *et al.*, 2005; Guillot *et al.*, 2008), in R software, for assessing the robustness of the results against different methods. Geneland also determines the number of population subdivisions for multi-locus genotype data using a Bayesian procedure, but further acknowledges spatial proximity when assigning individuals to clusters. Five independent runs were implemented for each analysis using 1 000 000 MCMC iterations with a thinning value of 100. Correlated and null allele model options were activated, and we used the default setting for other parameters. K was made to vary from 1 to 18 clusters. The model with the highest posterior probability was post-processed to obtain the posterior probability of subpopulation membership for each pixel of the spatial domain (114 x 82 pixels) with a burn-in of 100.

An unrooted Neighbor-joining (NJ) tree was built both at individual and population levels using the shared-allele distance in the PowerMarker program (vers. 3.25) (Tamura *et al.*, 2011). A thousand matrices were obtained by bootstrapping, and the consensus tree was provided by Mega 5 software (Tamura *et al.*, 2011).

We assessed the distribution of the genetic variance among and within genetic groups through a hierarchical analysis of molecular variance (AMOVA) using Arlequin 3.5 software (Excoffier, Lischer, 2010). We analyzed the among-group variance on the basis of the identified clusters. Statistical significance of the AMOVA was assessed on the basis of 10 000 permutations.

The degree of genetic differentiation for each population pair, or each inferred cluster pair, was further estimated by calculating multilocus estimates of pairwise  $F_{st}$  and associated p values (10 000 permutations) implemented in Arlequin 3.5 software (Excoffier, Lischer, 2010).

### *Environmental data and species distribution modelling*

We performed species distribution modelling (SDM) using MAXENT software (version 3.3.3k, Phillips *et al.*, 2006; Phillips, Dudik, 2008) in order to assess the response of *Amborella* to environmental variation, and to deduce both its present and past (last glacial maximum, LGM, ~21 000 BP) potential distributions. SDM was based on current occurrences of *Amborella* and on environmental variables at these locations. Compared to other presence-only models (such as ENFA, GARP and Bioclim), Maxent shows good predictive accuracy (Elith, Leathwick, 2009; Phillips *et al.*, 2006). We obtained geographic information system (GIS) layers displaying environmental information on morphopedology (Podwojewski, Beaudou, 1987), elevation (digital elevation model, DEM, from the DTSI governmental institution in New Caledonia) and bioclimate (Hijmans *et al.*, 2005). These variables are known to be important drivers of plant ecology in tropical forests of New Caledonia. We considered seven morphopedology classes bearing specific vegetation types (out of the 44 original classes) and mapped them at 1:200000 scale. Ultramafic substrates were divided into peridotite-laterite (1) and serpentinite (2); whereas volcano-sedimentary soils were divided into limestone (3), basalt (4), phtanite (5), and other volcano-sedimentary substrates (6); finally a miscellaneous class (7) included other substrates, mostly marine salted sediments. We used the Spatial Analyst tool of ArcGIS 9.3 to derive the slope and aspect from the DEM. The original angular measure of aspect was transformed into two complementary quantitative variables of eastness (sine of aspect) and northness (cosine of aspect). Annual rainfall information was provided by Météo France (Météo France, 2007), and eight classes were considered, 389-1000 mm, 1000-1300 mm, 1300-1600 mm, 1600-2000 mm, 2000-2500 mm, 2500-3000 mm, 3000-3500 mm and above 3500 mm. The bioclimatic variables related to temperatures were extracted from WorldClim (<http://www.worldclim.org/>) at 30 arc-second resolution (1 km<sup>2</sup>), and a principal component analysis was performed to assess the variation in bioclimatic variables over a random sample of 4 000 points spread over the mainland and the Loyalty Islands, using the “ade4” library in R software (Chessel *et al.*, 2004). We identified three major principal components that described 98.5% of the overall variation. The first (PCA1) represented an overall gradient in mean annual and quarter temperatures. The second (PCA2) highlighted a

gradient in temperature seasonality. The third (PCA3) contrasted the temperature seasonality during the year and the mean diurnal variation. We used these principal components as synthetic bioclimatic GIS layers for subsequent ecological analyses. This dimensionality reduction acknowledged the strong correlation between bioclimatic variables and thus prevented potential SDM overfitting (Phillips *et al.*, 2006).

We used the 70 herbarium recorded presences (Table S1A, Supporting Information) as calibration data of the Maxent model, and the 47 other presence data (Table S1B, S1C, Supporting Information) as validation data so as to check the model consistency when applied to an independent dataset. The predictor environmental variables were morphopedology, rainfall (both categorical), bioclimatic PCA1, PCA2 and PCA3 variables, elevation, slope, northness and eastness. We included linear, quadratic, hinge and binary features in the model. The convergence threshold was set at  $10^{-5}$ , the maximum number of iterations at 1 000 and the regularization value  $\beta$  at  $10^{-4}$  (Phillips *et al.*, 2006). Ten replicate runs of the model were performed based on random subsampling of the calibration dataset. The Maxent model yielded a logistic probability of presence for each pixel of the study area. The predictive power of the model was evaluated by threshold-dependent binomial omission tests and the area under the (receiver operating characteristic; ROC) curve (AUC), based on calibration and test presence points and on 1 000 random background points representing putative absences.

We then mapped the binary potential distribution of *Amborella* under current environmental conditions, based on the assumption that the species was absent below and present above a given threshold of the logistic probability of presence. We chose the sensitivity-specificity equality threshold provided by Maxent, which is not biased toward better prediction of presence or absence, and proved to yield reliable results (Liu *et al.*, 2005). We also considered the maximum sensitivity plus specificity threshold, which provides a larger predicted distribution possibly, including suboptimal margins of the suitable habitat. We finally compared the potential distribution with the map of current vegetation types (Jaffré *et al.*, 2012).

The logistic probability of presence was also applied as an index of local habitat suitability. The objective was twofold, to assess whether: (i) the local habitat suitability affected the local genetic structure, and (ii) the heterogeneity of habitat conditions between populations could explain their current genetic differentiation.

We calculated the cumulated habitat distances, as the sum of the probabilities of presence in pixels between each pair of populations, to represent potential ecological barriers. We analyzed the relative contribution of geographic distances and habitat distances to genetic differentiation between populations using the `mantel.rtest` function in the R software package.

In order to investigate the putative refugia of *Amborella* during the last glacial maximum (LGM, ~21 000 BP), we mapped its potential distribution under LGM environmental conditions. Topographical information (elevation and slope) and morphopedology (based on geology) were assumed identical to present. LGM bioclimatic layers derived from the PMIP2 database (<http://pmip2.lsce.ipsl.fr/>) (Braconnot *et al.*, 2007) were combined to get the maps corresponding to the three PCA axes presented above.

## Results

### *Genetic variability over the species range*

All loci were polymorphic at the species level with a mean of 11 alleles per locus and a polymorphic information content (PIC) of 0.62 per locus (Table S2, Supporting Information).

Clustering analyses showed that the 18 populations could be merged into 12 groups corresponding to the 12 sites we chose (see above). Although the relatively small sizes of some populations probably affected some genetic parameters, most of the following results were consistent with those previously reported on larger populations with the same markers (Poncet *et al.*, 2012). The presence of null alleles across loci was very low (Fig. S2C, Supporting Information), and no deviation in the inbreeding coefficient  $F_{IS}$  was observed across loci and sites, except for Amt-43 in Tonine (Table 2). Although some loci were fixed in some populations and sites, no significant outlier locus was detected (Table S3, Supporting Information). No significant heterozygote deficit or excess was detected within any of the sites and, apart from few loci, no significant deviation from the Hardy-Weinberg equilibrium at the site level was detected (Table 2). Allelic richness based on sample-size correction ranged from 2.2 (Nakada) to 4.7 (Tonine).

### *Gene flow within sites*

The Neighbor joining analysis at the individual level exhibited a geographically structured distribution of accessions. However, intermixed clustering of individuals from adjacent populations within a site was observed for all 12 sites (Fig. S1, Supporting Information), suggesting close genetic relationships at this scale, and no strong effect of the elevation gradients on within-site populations. This was confirmed by Bayesian-based population clustering. The posterior distributions of the cluster number (K) from Geneland displayed a clear mode at K = 12 for each of the five replicates performed (Fig. S2, Supporting Information). This suggested that adjacent populations located at the same site could be considered as belonging to the same cluster corresponding, here, to the sites. Moreover, the genetic differentiation between adjacent populations was low, as shown by the low pairwise  $F_{ST}$  values (mean  $F_{ST}$  [min, max] = 0.03 [-0.01, 0.06]) (Table S4, Supporting Information), suggesting high gene flow at the local scale.

To estimate genetic diversity among the 12 sites, all pairwise  $F_{ST}$  values were calculated and showed highly significant values for all pairs (Table S5, Supporting Information). The highest differentiation was observed between Nakada and the other sites. Several rare or specific alleles were found within some sites (Table 2). In particular, Tonine presented the highest fraction of private alleles. Although private allele presence was probably overestimated owing to limited sample sizes of some populations, their distribution was consistent with the genetic diversity pattern. The highest level of genetic diversity was found in Tonine, Aoupinié and Ba (Table 2).

An AMOVA based on sampled populations and sites (Table 3, (a)) highlighted that 33.3% ( $P < 0.0001$ ) of the genetic variance occurred among sites, while only 2.3% ( $P < 0.0001$ ) was assigned among populations within sites. This confirmed the genetic homogeneity within sites and the high differentiation between sites ( $F_{ST} = 0.36$ ).

### *Species genetic structure*

The Bayesian clustering analysis with Structure identified four well-differentiated clusters. Indeed, the method described by Evanno (2005), showed the highest peak at  $K = 4$  (Fig. 2). The four clusters were congruent with four geographic groups: group I in the north, group II in the center, group III in western center, and group IV in the south (Nakada). Interestingly, Nakada was a very distinct cluster that was clearly separated as soon as  $K = 2$  and did not split when  $K$  was increased (Fig. S3A, Supporting Information). Clustering levels found at  $K = 11$  and  $K = 12$  confirmed the genetic homogeneity at the site level, with clear merging of adjacent populations, including Me Fomechawa and Me Ori which were 5.3 km apart. Separate clustering analysis of group I and group II revealed secondary splits within each of the two groups. With an optimal value at  $K = 2$ , group I could further be subdivided into Pwicate on one side and Tonine, Ponandou and Tchamba on the other (Fig. S3B, Supporting Information). Likewise, Dogny was separated from the other group II populations, but Amieu displayed some level of admixture.

The NJ tree (Fig. 3 (B)) further illustrates the extreme position of Nakada populations, and shows the relative positions of populations within each group. Since the exact origin of the *Amborella* specimen used for whole genome sequencing was unknown, this accession was added to the NJ tree analysis and turned out to cluster with the Amieu population (Fig. 3 (B)).

Partitioning the genetic variance using AMOVA with sites grouped into the four groups assigned with Structure analysis (Table 3, (b)) revealed highly significant ( $P < 0.0001$ ) genetic differentiation among groups (19.55% of the variance,  $F_{ST} = 0.37$ ), while 18.4% ( $P < 0.0001$ ) was assigned among sites within groups.

### *Amborella current ecological requirements and distribution*

Under current climatic conditions, the *Amborella* SDM showed a very good predictive ability, which proved to be consistent based on training and test data (area under the ROC curve, AUC training = 0.933, AUC test = 0.916). *Amborella* occurred only on volcano-sedimentary soils, was more likely found at mid-elevation

(between 750-1 000 m), in moderately steep settings (peak around 10% slope) and in rather rainy areas (rainfall 2 000-3 000 mm) (Fig. S4, Supporting Information). It was positively associated with PCA2 and negatively with PCA3, indicating that it grows in less seasonal contexts regarding monthly temperatures. Finally, there was no preference found regarding the aspect (eastness and northness variables). The jackknife tests of variable importance (Fig. S5, Supporting Information) highlighted the prominent effect of morphopedology, followed by PCA1, rainfall and elevation, when each variable is taken alone. The morphopedology, and to a lesser extent rainfall, improved the model when added to the other variables.

We generated a binary presence-absence map of the potential distribution, based on the probability of presence predicted by the model (SDM) on the mainland, to which we applied a threshold separating the predicted presences and absences (Fig. 4). We considered the sensitivity-specificity equality threshold at  $p = 0.379$ , and the maximum sensitivity and specificity threshold at 0.293. The first threshold represents a balanced predictive ability of presences and absences, while the second enlarges the suitable habitat to include margins where the species can be more sparsely found, thereby increasing the prediction of presences but not that of absences (Table S6, Supporting Information).

Under current climatic conditions, *Amborella* could be potentially distributed throughout the central eastern mainland zone. The area of the predicted potential distribution was 1 512 km<sup>2</sup>, i.e. about 9% of the New Caledonian mainland. According to the SDM, 59% of the potential habitat of *Amborella* would be found in areas that are currently covered with forests (Jaffré *et al.*, 2012), while 27% would be found in savannah, and 10.6% in secondary and thicket vegetation.

#### *No significant ecological barrier to gene flow*

SDM probabilities were recorded at the population locations used for the genetic studies, so as to provide an index of the local habitat suitability under current climatic conditions. We found no significant relationship between the local probabilities of presence predicted by the model (SDM) and the local genetic parameters

(allele richness, heterozygoty,  $F_{IS}$ ; all ANOVA p-values above 0.33). This result indicated that the local habitat suitability, as estimated through the SDM, did not influence the local population genetic structure. To assess the impact of isolation by distance on the genetic structure, we performed a Mantel test of the geographical distances and the Rousset's (1997) genetic distances between populations ( $F_{ST} / (1 - F_{ST})$ ). Within the whole species distribution range, the correlation between geographic and genetic distances proved significant ( $R_{xy} = 0.313$ ,  $P < 0.001$ ). When the two Nakada populations were removed from the analysis, this correlation even increased to  $R_{xy} = 0.422$  ( $P < 0.001$ ). Indeed, although geographically close to populations of group II (ca. 20 km between Nakada and Dogny), the Nakada populations were genetically highly differentiated ( $F_{ST}$  Nakada-Dogny = 0.49, pairwise permutation  $P < 0.0001$ ). Within the two largest clusters, groups I (north, 6 populations) and II (center, 8 populations), we observed an even greater increase in genetic differentiation among individuals with geographic distance (Mantel  $R_{xy} = 0.816$   $P = 0.009$  and  $R_{xy} = 0.543$   $P = 0.006$  for groups I and II, respectively). We further performed a partial Mantel test to assess whether part of the genetic differentiation could be explained by habitat distances in addition to the isolation-by-distance effect. Because there was a strong correlation between geographic and cumulated habitat distances (Mantel  $R_{xy} = 0.996$ ,  $P < 0.001$ ), we obtained comparable results when relating genetic distances and cumulated habitat distances (overall partial Mantel test  $P = 0.49$ ).

#### *Amborella refugia under LGM*

The projection of the distribution model (SDM) under LGM bioclimatic conditions (ca. 21 000 years BP) showed that the suitable area for *Amborella* was much smaller than the current suitable area (52.5 km<sup>2</sup> instead of 1 512 km<sup>2</sup>, hence 96.5% less). It was located to the northeastern zone of its current distribution (Fig. 4). Most of the suitable habitat was restricted to two mountainous regions: an area on the northeastern coast corresponding to the Massif des Lèvres refugium and a smaller isolated refugium located more inland and to the south in the central mountain chain, close to Mount Aoupinié. The reduced level of rainfall was the prominent factor explaining the absence of suitable area in the southern part of the distribution. The Aoupinié

refugium could have extended more toward the southwest than illustrated here because the LGM climatic layer shows a sharp change in rainfall in this area, which limits the extent of suitable habitat. Interestingly, the genetic group I superimposes the northern refugium, and group II includes the second refugium (Fig. 4).

## Discussion

Both frequency-based and Bayesian clustering analyses revealed four highly-differentiated genetic groups of *Amborella* populations over its natural range. We here discuss dispersal and demographic constraints that may explain this pattern within the potential distribution modelled under current environmental conditions. By projecting this potential distribution under the climate of the last glacial maximum (LGM), we finally assess the impact of the contraction and expansion history on the current genetic structure of *Amborella*.

### *How do the current ecological constraints shape the distribution of Amborella?*

Niche modelling based on current climatic conditions showed that the suitable *Amborella* habitat is typical of mid-elevation wet evergreen forests, and driven mainly by bioclimatic factors. The model performed well and we obtained consistent predictions on the basis of both herbarium data and accurate GPS-located data, which confirms the value of naturalist and herbarium data for predictive ecological surveys, when properly analyzed and controlled (Bebber *et al.*, 2010). The potential distribution of *Amborella* covers 30% of the current extent of wet tropical forests, and is mostly included within the eastern-central phytogeographic sector defined by Jaffré and Veillon (1989) for palms, and by Jaffré (1995) for conifers, with a slight extension toward the western-central sector (Fig. S6, Supporting Information). The western and eastern-central sectors include humid forests of the central chain and are limited in the south by ultramafic substrates (mainly peridotites) and in the north by the Mt. Panié range. Interestingly, although *Amborella* has never been encountered in the

Panié range or at Roches Ouaième, we found this area to be suitable, without any clear geological barrier. This absence of populations in this region suggests that constraints other than the current bioclimate may have limited the distribution over time (Fig. 4).

Apart from hampering widespread distribution, habitat quality is known to affect local population persistence and performance in combination with genetic and demographic processes (Vergeer *et al.*, 2003). However, we did not find any correlation between local habitat suitability and the local genetic diversity statistics of *Amborella* populations. Our results suggest, conversely, that the loci themselves behave neutrally according to the habitat variation and/or that habitat variation does not significantly influence local population dynamics. Furthermore, least-cost habitat distances, as an index of the potential ecological barriers between populations, did not explain the genetic differentiation between populations better than geographic distances. Although ecophysiological constraints of ultramafic areas prevented the installation of populations, habitat heterogeneity in general did not represent a strong enough barrier to gene flow within the species distribution range. On the other hand, the isolation-by-distance patterns between populations indicated that the dispersal abilities are still limited. Pollen dispersion is both wind- and insect-mediated, but wind is limited in the understory and insects are mainly beetles dwelling in the forest litter, suggesting limited pollen dispersal capacities (Endress, Igersheim, 2000). Seed dispersal agents are unknown, but fruit scattering is also probably limited in the understory. However, within populations we found a genetic shuffle consistent with the dioecy of *Amborella* (Endress, Igersheim, 2000) and the presence of one seed per fruit also limits the proximity of related individuals in the population. However, pure dispersal limitation is not the only main factors involved the genetic distribution since the pattern of isolation by distance was clearer within genetic clusters than between clusters, suggesting that other processes have constrained the dynamics of *Amborella* over time.

#### *Genetic differentiation explained by biogeography and history*

Since neither current ecological barriers nor dispersal limitation could explain the high differentiation between the four main genetic clusters, further past ecological processes were probably involved. We hypothesized

that the alternance of contraction and expansion of rainforests due to climatic variations during the Pleistocene led to bottlenecks in the past genetic structure, whose signature can be detected in the present-day genetic diversity. Projecting ecological niche models (ENMs) onto historical landscapes provided a spatial context for testing this phylogeographic scenario. Such approaches have been widely used to identify refugia during the late-Pleistocene LGM (e.g.: Beatty, Provan, 2011; Bisconti *et al.*, 2013; McKay *et al.*, 2010; Svenning *et al.*, 2011). Two suitable spatially distinct areas were found in the LGM range of *Amborella*, near the Massif des Lèvres and Mt. Aoupinié. They correspond to current populations, respectively from groups I and II, with typical signatures of former refuge populations (Hampe, Petit, 2005), i.e. higher levels of regional genetic diversity and private alleles. This thereby supports the scenario of survival and expansion of at least two lineages from these putative refugia, creating the current groups I and II or part of them. Moreover, this scenario is congruent with a parallel study based on the genomic analysis of 12 individuals, one from each collecting site (*Amborella* Genome Project, submitted). Using the pairwise sequentially Markovian coalescent (PSMC, Li, Durbin, 2011) model, they concluded that *Amborella* experienced a strong bottleneck around the LGM, and showed that at that time there were two distinct sublineages. Finally, the small range and the low allelic richness of the two remaining spatially restricted genetic groups III and IV suggest that they probably differentiated from the two LGM lineages through a founder effect (Hampe, Petit, 2005). The Nakada group IV consists of the geographically most extreme populations in the southern and leading part of the distribution. Moreover, these populations grow on schistose soils on the southern slope of Mt. Nakada, while the summit of Nakada is ultramafic, thus partly isolating them ecologically from the other populations. The reasons for the significant differentiation of group III, Mé Foméchwawa - Mé Ori populations, geographically encompassed within Group II, are unclear and might be the result of a founder effect on the distribution fringe. Additional collections in areas of transition between Nakada or Me Ori and Group II, where *Amborella* is known to occur according to herbarium specimens, might be necessary.

The dramatic decrease in suitable habitat area during the LGM (52.5 km<sup>2</sup>, hence 96.5% less) is also consistent with paleoenvironmental findings, indicating that wet tropical forests of southern New Caledonia regressed in

drier periods (Stevenson *et al.*, 2001; Stevenson, Hope, 2005). Latham (1986) further estimated that at ca. 25 000 BP in western New Caledonia rainfall was 30% lower than present-day levels. Increases in moisture and wet tropical forest expansion subsequently took place between ca. 13.5 ka and ca. 11 ka BP (Hope *et al.*, 2004; Stevenson *et al.*, 2001). A former study indicated that the wetter Massif des Lèvres could have acted as a Pleistocene refugia for palm species during these drier periods (Pintaud *et al.*, 2001), as also suggested by the presence of many other micro-endemic species. Pillon *et al.* (2009) also observed that non-hybrid species of the endemic *Codia* genus are mainly restricted to these putative rainforest refugia. Our work provides evidence on the role of this region as refugia in shaping current intraspecific diversity patterns. Furthermore, we outlined another previously undetected putative refugium in the Aoupinié massif. This new refuge is consistent with the high level of local endemism previously observed on this mount (e.g. *Syzygium* species, Dawson, 1999). This refugium is original and would represent a new isolated central mountain refugium, thus extending the hypothesis of eastern refugia (Pintaud *et al.*, 2001), based on a refined ecological model. Concordance across multiple intraspecific studies would confirm a broader role of this refugium.

The postglacial colonization from Massif des Lèvres and Mt Aoupinié refugia towards southern areas could have been facilitated by the apparent absence of substantial ecological barriers once ecological conditions improved. However, the better southward than northward expansion of *Amborella* suggests that habitat suitability differences may have persisted between these two directions. Rainforest colonization may have been quicker in the southern part because of the overall higher rainfall in this part. Conversely, the northern part could have remained unsuitable for a longer period. This scenario could have concerned some other species since a floristic break was observed between Massif des Lèvres and Mt Panié for palms and conifers, indicating the presence of two different phytogeographic sectors (Jaffré, Veillon, 1989; Pintaud *et al.*, 2001) (Fig. S6, Supporting Information). Although accurately incorporating dispersal into predictive models is not straightforward (Svenning *et al.*, 2011), our data highlight that taking historical dispersal events into account is essential for gaining insight into species distributions.

### *Conservation implications*

Despite its restricted distribution on world and regional scales, *Amborella* exhibits a moderate but structured diversity level and a broad and almost continuous distribution in New Caledonia, encompassing 36.9% of current wet tropical forests on volcano-sedimentary soils. However, the species is threatened by habitat destruction due to overgrazing, fire and urban expansion. Indeed, another important insight gained through the species distribution model is that 27% of the potential distribution currently overlaps savannah areas. The regression of wet tropical forest and its replacement by savannah was shown to be related to human activities (Ibanez et al., 2013) and coincided with the arrival of humans, i.e. the Melanesians around 3000 years BP and Europeans in the mid-1800s (Stevenson, 2004). Conservation efforts in New Caledonia should focus on refugia as key areas in a global warming setting (Keppel et al. 2012). While Aoupinié mountain is already a nature reserve, Massif des Lèvres has no protection status and little is known about this area.

### *Conclusion*

Our findings evidenced the presence of a minimum of two LGM rainforest refugia in New Caledonia and confirmed the importance of refugia in shaping the current intra- and inter-specific diversity in New Caledonian plants. Pleistocene refuge hypothesis in tropical regions has been addressed for tropical trees in the Neotropics (e.g. Cavers, Dick, 2013) and in Africa (e.g. Hardy *et al.*, 2013). But for most islands in the tropical Pacific, the paleoenvironmental context of vegetation history is still poorly understood (Hope et al., 2004). In New Caledonia, while many phylogenetic studies (Cruaud et al., 2012; Grandcolas et al., 2008; Pillon, 2012) have shed some light on Cretaceous and Tertiary events, little attention had been paid to the effects of Pleistocene climatic fluctuations on its flora. Indeed, very few population genetics studies have been carried out on New Caledonian native plants (Bottin et al., 2005; Kettle et al., 2007; Kurata et al., 2008; Verhaegen et al., 2013) and none of them addressed rainforest plants. *Amborella trichopoda*, as a widespread rainforest shrub growing in non-ultramafic soils, was therefore a choice group to reveal LGM refugia and to understand the biogeography of this biodiversity hotspot.

## **Acknowledgements**

The authors wish to thank Jean-Christophe Pintaud, Adeline Barnaud and Tanguy Jaffré for insightful comments. We thank the people of the Ateu, Ba, Katrikoin, Kouare and Paola tribes for their hospitality and guidance through the forest and the members of the botanical laboratory at IRD Noumea for their assistance. We are also grateful to Camille Barré for technical help on GIS analysis. The authors thank Sylvain Santoni and the INRA genomic platform in Montpellier. This work was partly supported by IRD funding (Action Incitative) “AMBRES”.

## References

- Amborella Genome Project (submitted) The *Amborella* Genome and the Evolution of Flowering Plants. *Science*.
- Beatty GE, Provan J (2011) Comparative phylogeography of two related plant species with overlapping ranges in Europe, and the potential effects of climate change on their intraspecific genetic diversity. *Bmc Evolutionary Biology* **11**.
- Bebber DP, Carine MA, Wood JR, *et al.* (2010) Herbaria are a major frontier for species discovery. *Proc Natl Acad Sci U S A* **107**, 22169-22171.
- Bisconti R, Canestrelli D, Salvi D, Nascetti G (2013) A geographic mosaic of evolutionary lineages within the insular endemic newt *Euproctus montanus*. *Mol Ecol* **22**, 143-156.
- Bottin L, Verhaegen D, Tassin J, *et al.* (2005) Genetic diversity and population structure of an insular tree, *Santalum austrocaledonicum* in New Caledonian archipelago. *Mol Ecol* **14**, 1979-1989.
- Braconnot P, Otto-Bliesner B, Harrison S, *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features *Climate of the Past* **3**, 261-277.
- Cavers S, Dick CW (2013) Phylogeography of Neotropical trees. *Journal of Biogeography* **40**, 615-617.
- Chessel D, Dufour A-B, Thioulouse J (2004) The ade4 package-I-One-table methods. *R News* **4**, 5-10.
- Cordellier M, Pfenninger M (2009) Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod *Radix balthica* (Pulmonata, Basommatophora). *Molecular Ecology* **18**, 534-544.
- Cruaud A, Jabbour-Zahab R, Genson G, Ungricht S, Rasplus JY (2012) Testing the emergence of New Caledonia: fig wasp mutualism as a case study and a review of evidence. *PLoS One* **7**, e30941.
- Dawson JW (1999) Myrtaceae - Myrtoideae I : *Syzygium*. In: *Flore de la Nouvelle-Calédonie et Dépendances* (ed. Morat P), pp. 1-144. Muséum National d'Histoire Naturelle, Paris.
- Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology Evolution and Systematics* **40**, 677-697.
- Endress PK, Igersheim A (2000) The Reproductive Structures of the Basal Angiosperm *Amborella Trichopoda* (Amborellaceae). *International Journal of Plant Sciences* **161**, 237-248.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620.
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-567.
- Feliner GN (2011) Southern European glacial refugia: A tale of tales. *Taxon* **60**, 365-372.
- Gomez C, Mangeas M, Petit M, *et al.* (2010) Use of high-resolution satellite imagery in an integrated model to predict the distribution of shade coffee tree hybrid zones. *Remote Sensing of Environment* **114**, 2731-2744.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>.
- Grandcolas P, Murienne J, Robillard T, *et al.* (2008) New Caledonia: a very old Darwinian island? *Philos Trans R Soc Lond B Biol Sci* **363**, 3309-3317.
- Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes* **5**, 712-715.
- Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* **24**, 1406-1407.
- Haffer J (1969) Speciation in Amazonian forest birds. *Science* **165**, 99-105.
- Haffer J (1997) Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* **6**, 451-476.

- Haffer J, Prance GT (2001) Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differentiation. *Amazoniana-Limnologia Et Oecologia Regionalis Systemae Fluminis Amazonas* **16**, 579-605.
- Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8**, 461-467.
- Hardy OJ, Born C, Budde K, *et al.* (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes rendus Geoscience* **in press**.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* **359**, 183-195.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.
- Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends Plant Sci* **15**, 675-683.
- Hope G, Kershaw AP, van der Kaars S, *et al.* (2004) History of vegetation and habitat change in the Austral-Asian region. *Quaternary International* **118**, 103-126.
- Jaffré T (1995) Distribution and Ecology of the Conifers of New Caledonia. In: *Conifers of the Southern Hemisphere* (eds. Enright N.J., Hill R.S.), pp. 171-196. Melbourne University Press, Australia.
- Jaffré T, Rigault F, Munzinger J (2012) La végétation. In: *Atlas de la Nouvelle-Calédonie* (eds. Bonvallet J, Gay J-C, Habert É). IRD-Congrès de la Nouvelle-Calédonie, Marseille-Nouméa.
- Jaffré T, Veillon J-M (1989) Morphology, distribution and ecology of palms in New Caledonia. In: *Palms of the South-West Pacific* (ed. Dowe JL), pp. 158-159. Milton: Fund Palm and Cycad Society of Australia.
- Jeremie J (1982) Monimiacées, Amborellacées, Atherospermatacées, Trimeniacées, Chloranthaceae. In: *Flore de la Nouvelle-Calédonie et Dépendances* (ed. Aubréville A), pp. 127-179. Museum National d'Histoire Naturelle, Paris, France.
- Keppel G, Van Niel KP, Wardell-Johnson GW, *et al.* (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography* **21**, 393-404.
- Kettle CJ, Hollingsworth PM, Jaffre T, Moran B, Ennos RA (2007) Identifying the early genetic consequences of habitat degradation in a highly threatened tropical conifer, *Araucaria nemorosa* Laubenfels. *Mol Ecol* **16**, 3581-3591.
- Kurata K, Jaffre T, Setoguchi H (2008) Genetic diversity and geographical structure of the pitcher plant *Nepenthes vieillardii* in New Caledonia: A chloroplast DNA haplotype analysis. *Am J Bot* **95**, 1632-1644.
- Latham M (1986) *Altération et pédogenèse sur roches ultrabasiques en Nouvelle-Calédonie*, Etudes et Thèses edn. Orstom Editions, Paris.
- Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-498.
- Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385-393.
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128-2129.
- Maley J (1996) The African rain forest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh* **104B**, 31-73.
- Mathews S, Donoghue MJ (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947-950.
- McKay BD, Mays HL, Jr., Peng YW, *et al.* (2010) Recent range-wide demographic expansion in a Taiwan endemic montane bird, Steere's Liocichla (*Liocichla steerii*). *BMC Evol Biol* **10**, 71.
- Météo France (2007) *Atlas climatique de la Nouvelle-Calédonie* Météo France, Nouméa.
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci U S A* **104**, 19363-19368.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

- Neall VE, Trewick SA (2008) The age and origin of the Pacific islands: a geological overview. *Philos Trans R Soc Lond B Biol Sci* **363**, 3293-3308.
- Oginuma K, Jaffre T, Tobe H (2000) The karyotype analysis of somatic chromosomes in *Amborella trichopoda* (Amborellaceae). *Journal of Plant Research* **113**, 281-283.
- Pelletier B (2006) Geology of the New Caledonia region and its implications for the study of the New Caledonian biodiversity. In: *Compendium of marine species of New Caledonia, Doc. Sci. Tech. II7, seconde édition* (eds. Payri CE, Richer de Forges B), pp. 19-32. IRD, Nouméa.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.
- Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161-175.
- Pillon Y (2012) Time and tempo of diversification in the flora of New Caledonia. *Botanical Journal of the Linnean Society* **170**, 288-298.
- Pillon Y, Munzinger J (2005) Amborella fever and its (little) implication in conservation. *Trends Plant Sci* **10**, 519-520.
- Pillon Y, Munzinger J, Amir H, Hopkins HC, Chase MW (2009) Reticulate evolution on a mosaic of soils: diversification of the New Caledonian endemic genus *Codia* (Cunoniaceae). *Mol Ecol* **18**, 2263-2275.
- Pintaud J-C, Jaffre T, Puig H (2001) Chorology of New Caledonian palms and possible evidence of Pleistocene rain forest refugia. *C R Acad Sci III* **324**, 453-463.
- Podwojewski P, Beaudou A (1987) Carte morpho-pédologique de la Nouvelle Calédonie à 1/200000 (ed. ORSTOM), pp. 4 cartes, 3 fascicules.
- Poncet V, Couderc M, Tranchant-Dubreuil C, et al. (2012a) Microsatellite markers for *Amborella* (Amborellaceae), a monotypic genus endemic to New Caledonia. *Am J Bot* **99**, e411-414.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol Evol* **23**, 564-571.
- Qi XS, Chen C, Comes HP, et al. (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytol* **196**, 617-630.
- Qiu YL, Lee J, Bernasconi-Quadroni F, et al. (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404-407.
- Qiu YX, Sun Y, Zhang XP, et al. (2009) Molecular phylogeography of East Asian Kirengeshoma (Hydrangeaceae) in relation to quaternary climate change and landbridge configurations. *New Phytol* **183**, 480-495.
- Rebernick CA, Schneeweiss GM, Bardy KE, et al. (2010) Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (*Melampodium leucanthum*, Asteraceae). *Molecular Ecology* **19**, 3421-3443.
- Robbrecht E (1996) Geography of African Rubiaceae with reference to glacial rain forest refugia. In: *The biodiversity of African plants: Proceedings XIVth AETFAT congress* (eds. van der Maesen LJG, van der Burgt XM, van Medenbach de Rooy JM). Kluwer Academic, Netherlands.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
- Stevenson J (2004) A late-Holocene record of human impact from the southwest coast of New Caledonia. *Holocene* **14**, 888-898.
- Stevenson J, Dodson JR, Prosser IP (2001) A late Quaternary record of environmental change and human impact from New Caledonia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **168**, 97-123.
- Stevenson J, Hope GS (2005) A comparison of late Quaternary forest changes in New Caledonia and northeastern Australia. *Quaternary Research* **64**, 372-383.
- Svenning JC, Flojgaard C, Marske KA, Nogues-Bravo D, Normand S (2011) Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews* **30**, 2930-2947.

- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**, 453-464.
- Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739.
- Thien LB, Sage TL, Jaffré T, *et al.* (2003) The population structure and floral biology of *Amborella trichopoda* (Amborellaceae). *Annals of the Missouri Botanical Garden* **90**, 466-490.
- Vergeer P, Rengeling R, Copal A, Ouborg NJ (2003) The interacting effects of genetic variation, habitat quality and population size on performance of *Succisa pratensis*. *Journal of Ecology* **91**, 18-26.
- Verhaegen D, Assoumane A, Serret J, *et al.* (2013) Structure and genetic diversity of *Ixora margaretae* an endangered species. *Tree Genetics & Genomes* **9**, 511-524.
- Wellenreuther M, Sanchez-Guillen RA, Cordero-Rivera A, Svensson EI, Hansson B (2011) Environmental and climatic determinants of molecular diversity and genetic population structure in a coenagrionid damselfly. *PLoS One* **6**, e20440.

### **Data Accessibility**

Microsatellite data: doi:10.5061/dryad.8nb7v. *Amborella trichopoda* occurrence records used for the Maxent models and sampling locations of the populations are provided in Table S1 (Supporting Information)

### **Supporting Information**

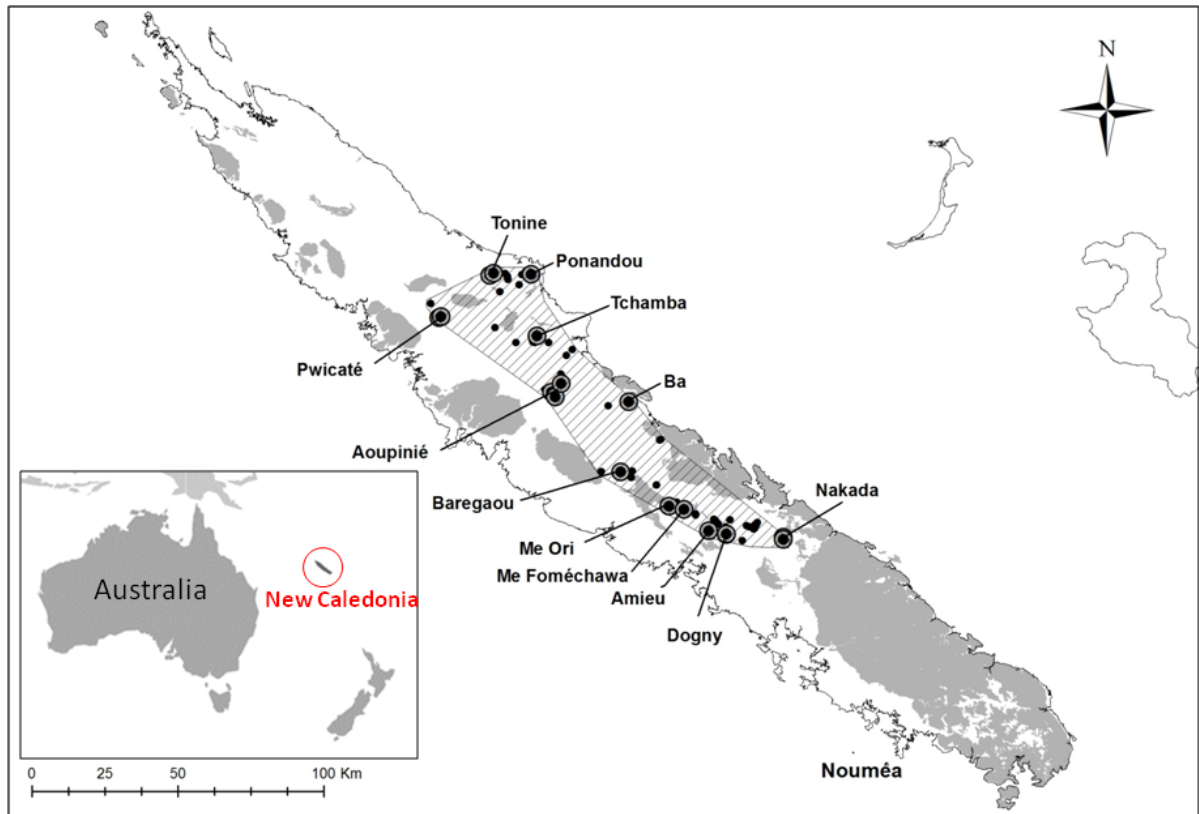
Additional supporting information may be found in the online version of this article: supplementary Tables and Figures.

---

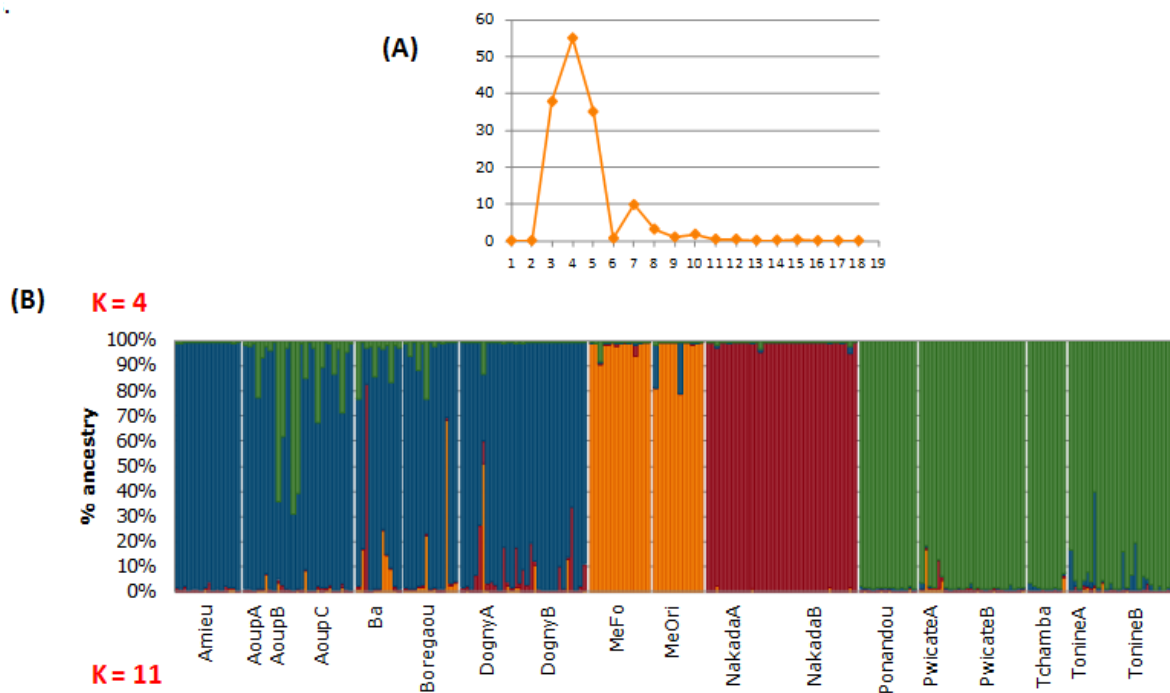
V.P. and A.d.K. obtained funding, designed and coordinated the study; Y.P. collected the samples; F.M. and J.M. performed the ecological research; V.P., C.G., Y.P., C.T.-D. and M.C. performed the genetic research; V.P., F.M. and J.M. analysed the results; SH supported the work. VP and FM wrote the manuscript.

---

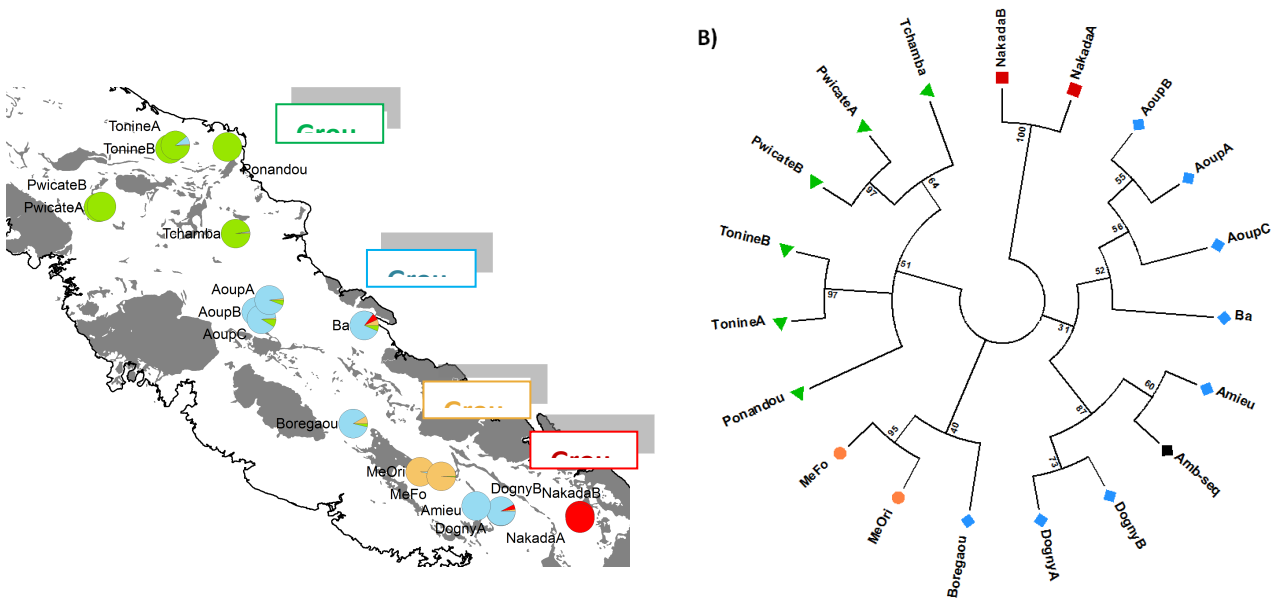
**Figure 1.** Geographical distribution of the study sites. The hatched area corresponds to the geographical range of *Amborella trichopoda* as defined by georeferenced vouchers or sites (black dots). The circled dots represent populations sampled for the genetic analysis. The ultramafic substrates are represented by grey areas. More information about the locations is available in Table S1 (Supporting information).



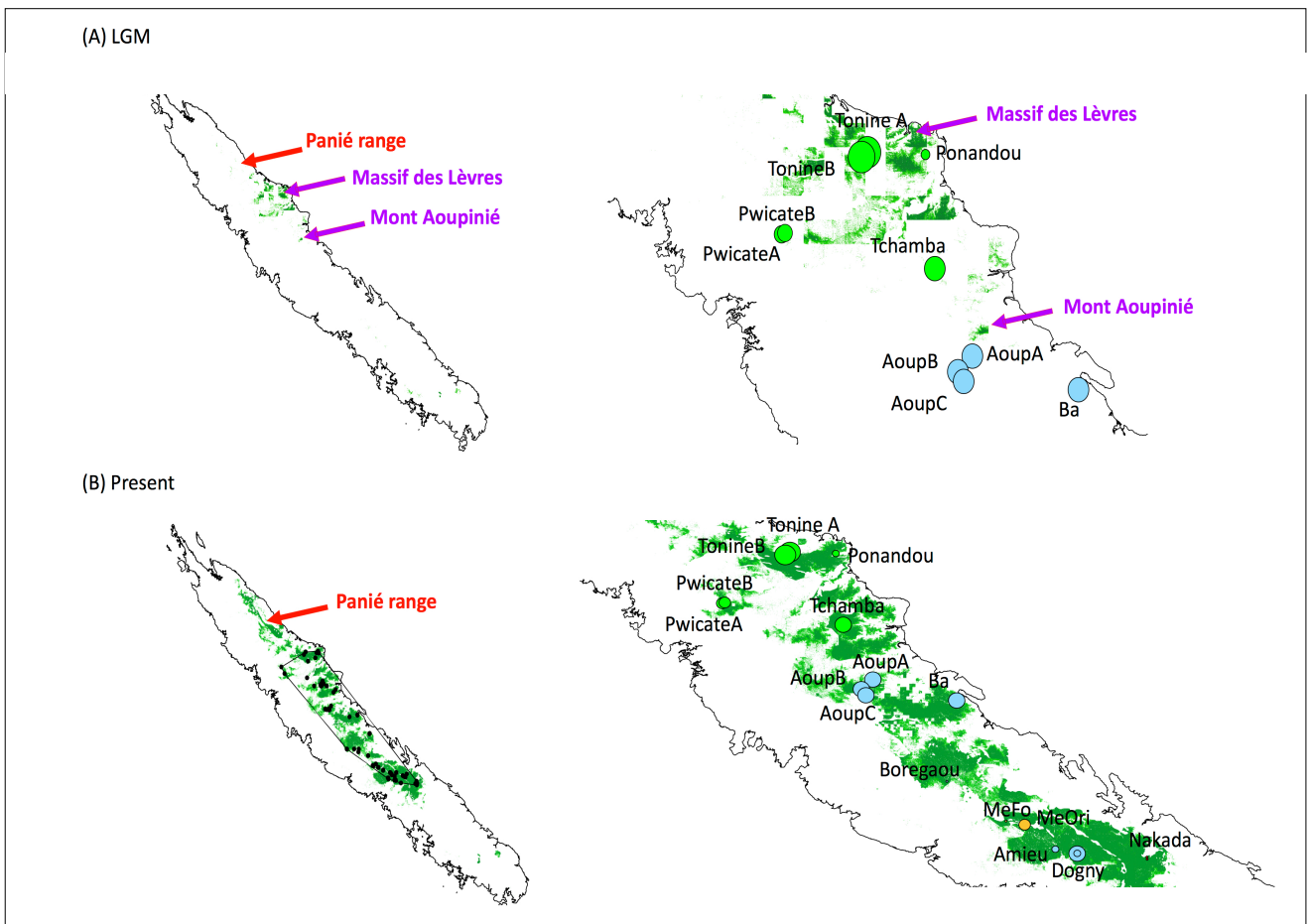
**Figure 2.** Estimated population structure of the 18 *A. trichopoda* populations from Bayesian structure analysis using the STRUCTURE 2.3.3. program. (A) Delta K values for different Ks, suggesting that K=4 is the most likely structure according to Evanno et al. (2005); (B) Individual Bayesian assignment probabilities for K=4 for 18 *A. trichopoda* populations and runs with the highest posterior probability (grouped for visualization into geographically close groups). Individuals are presented by thin vertical lines, which are partitioned into segments representing each individual's estimated membership fraction.



**Figure 3.** (A) Geographic distributions of four *Amborella trichopoda* gene pools estimated by STRUCTURE analysis. The four groups are illustrated by different colors. (B) Consensus bootstrapped Neighbor-joining tree of the 18 populations based on shared allele distances. The symbol colors and shapes indicate the four clusters. The sample used for the whole genome sequencing is indicated in black (Amb-seq). When nearby populations were sampled at the same site but at different elevations, they are noted with the site name followed by A, B or C.



**Figure 4.** Potential distribution of *Amborella trichopoda* and population structure. (A) Last glacial maximum (LGM, ~ 21 000 years BP) model (B) Current distribution and actual range (georeferenced vouchers and sites in black dots) model. Distributions are presented both on the whole mainland and on an enlarged view. The population gene pools estimated by STRUCTURE analysis are superimposed, colored according to genetic group and with circle sizes proportional to the level of current genetic diversity (or expected heterozygosity,  $H_E$ ). The area of occurrence predicted by the Maxent SDM is shown in dark green for  $p \geq 0.379$  and light green for  $0.379 > p \geq 0.293$ . The Tonine populations are located in the Massif des Lèvres putative refugium. The Mt Aoupinié (Aoup) populations are located close to another putative refugium.



**Table 1.** Information on the 18 *Amborella trichopoda* populations.

<b>Population</b>	<b>Site</b>	<b>Elevation (m)</b>	<b>Latitude</b>	<b>Longitude</b>	<b>No. of individuals</b>
<b>Amieu</b>	La Foa, Col d'Amieu	490	-21,61063889	165,8104722	16
<b>AoupA</b>	Ponerihouen, Mont Aoupinié, versant Goa	500	-21,15602778	165,3195833	6
<b>AoupB</b>	Ponerihouen, Mont Aoupinié, versant Goa	860	-21,18222222	165,2889167	8
<b>AoupC</b>	Poya, Mont Aoupinié, versant Goapin	610	-21,19855556	165,3011667	13
<b>Ba</b>	Houailou, Vallée de Ba	260	-21,21177778	165,544	11
<b>Boregaou</b>	Bourail, Crête au nord de Boregaou	560	-21,42927778	165,5185833	13
<b>DognyA</b>	Sarraméa, Plateau de Dogny	560	-21,62213889	165,8688889	15
<b>DognyB</b>	Sarraméa, Plateau de Dogny	850	-21,62158333	165,86925	16
<b>MeFo</b>	Moindou, Mé Foméchwawa	540	-21,54508333	165,7274722	15
<b>MeOri</b>	Moindou, Mé Ori	500	-21,53547222	165,6775833	12
<b>NakadaA</b>	Thio, Mont Nakada	680	-21,63613889	166,0565556	18
<b>NakadaB</b>	Thio, Mont Nakada	860	-21,63025	166,05625	19
<b>Ponandou</b>	Touho, Ponandou	110	-20,81719444	165,2199722	14
<b>PwicateA</b>	Koné, Mont Pwicate	700	-20,95141667	164,9165556	8
<b>PwicateB</b>	Koné, Mont Pwicate	920	-20,94936111	164,9236389	18
<b>Tchamba</b>	Ponerihouen, Vallée de la Tchamba	540	-21,00902778	165,2398611	9
<b>TonineA</b>	Touho, Mont Tonine	450	-20,81363889	165,0975	7
<b>TonineB</b>	Touho, Mont Tonine	670	-20,82147222	165,0851667	19

**Table 2.** Summary of genetic diversity estimates obtained for the 12 *Amborella trichopoda* sites as defined by Geneland. Allelic richness correction for sample size ( $A_E$ , based on a minimum sample size of 9 individuals per site), number of private alleles (PA), genetic diversity (expected heterozygosity  $H_E$ ), observed heterozygosity ( $H_o$ ), inbreeding coefficient ( $F_{IS}$ ), and Hardy-Weinberg equilibrium (HW).

Site	$A_E$	PA	$H_E$	$H_o$	$F_{IS}^a$	HW <sup>b</sup>
Amieu	2.7	1	0.41	0.39	0.06	Amt-43
Aoup	4.3	4	0.55	0.50	0.10	Amt-43
Ba	3.7	0	0.55	0.52	0.06	
Boregaou	2.9	0	0.35	0.33	0.04	
Dogny	3.2	0	0.47	0.46	0.03	Amt-16
MeFo	2.9	2	0.45	0.49	-0.08	
MeOri	2.5	0	0.32	0.37	-0.16	
Nakada	2.2	3	0.34	0.34	-0.01	Amt-43
Ponandou	3.1	0	0.41	0.40	0.00	
Pwicaté	3.4	5	0.47	0.44	0.07	
Tchamba	3.8	1	0.51	0.46	0.10	
Tonine	4.7	6	0.59	0.56	0.06 <sup>a</sup>	Amt-43
<b>Mean</b>	<b>3.28</b>	<b>3.38</b>	<b>0.45</b>	<b>0.44</b>	<b>0.02</b>	

<sup>a</sup> indicates that none of the  $F_{IS}$  values were significantly different from zero  $P < 0.05$  (adjusted value), except Tonine for Amt-43

<sup>b</sup> names of markers for which a deviation from H-W equilibrium was observed.

**Table 3.** Percentage of variation of molecular variance (AMOVA) attributed to various hierarchical population structure levels for *Amborella trichopoda*, according to: (a) sites; (b) groups as defined by Structure analysis

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
<b>(a)</b>				
Among sites	11	378.2	0.826 Va	33.3***
Among populations within sites	6	17.827	0.057 Vb	2.3***
Within populations	456	728.367	1.597 Vc	64.4***
Total	473	1133.395	2.480	
FST: 0.356; FSC: 0.034; FCT: 0.333				
<b>(b)</b>				
Among groups	3	237.991	0.504 Va	19.55***
Among sites within groups	8	147.351	0.474 Vb	18.39***
Within sites	462	739.477	1.600 Vc	62.06***
Total	473	1124.819	2.579	
FST: 0.379; FSC: 0.229 ; FCT: 0.195				

\*\*\*:  $P < 0.0001$