Genetic diversity and population divergences of an indigenous tree (Coffea mauritiana) in Reunion Island: role of climatic and geographical factors

Edith Garot, Thierry Joët, Marie-Christine Combes, Philippe Lashermes

the **genetics**society

ARTICLE

Genetic diversity and population divergences of an indigenous tree (*Coffea mauritiana*) in Reunion Island: role of climatic and geographical factors

Edith Garot [1] · Thierry Joët [1] · Marie-Christine Combes[1] · Philippe Lashermes [1]

Abstract

Oceanic islands are commonly considered as natural laboratories for studies on evolution and speciation. The evolutionary specificities of islands associated with species biology provide unique scenarios to study the role of geography and climate in driving population divergence. However, few studies have addressed this subject in small oceanic islands with heterogeneous climates. Being widely distributed in Reunion Island forest, *Coffea mauritiana* represents an interesting model case for investigating patterns of within-island differentiation at small spatial scale. In this study, we examined the genetic diversity and population divergences of *C. mauritiana* using SNP markers obtained from 323 individuals across 34 locations in Reunion Island. Using redundancy analysis, we further evaluated the contribution of geographic and climatic factors to shaping genetic divergence among populations. Genetic diversity analyses revealed that accessions clustered according to the source population, with further grouping in regional clusters. Genetic relationships among the regional clusters underlined a recent process of expansion in the form of step-by-step colonization on both sides of the island. Divergence among source populations was mostly driven by the joint effect of geographic distance and climatic heterogeneity. The pattern of isolation-by-geography was in accordance with the dispersal characteristics of the species, while isolation-by-environment was mostly explained by the heterogeneous rainfall patterns, probably associated with an asynchronous flowering among populations. These findings advance our knowledge on the patterns of genetic diversity and factors of population differentiation of species native to Reunion Island, and will also usefully guide forest management for conservation.

Introduction

Oceanic islands are de novo islands that are usually result of volcanic activity and have never been connected to landmasses (García-Verdugo and Fay 2014). Since Darwin's observations in the Galapagos Islands, oceanic islands have been considered as ideal systems for understanding processes of evolution and speciation. Their particularities are based on several characteristics (Emerson 2002; Losos and Ricklefs 2009). First, patterns of evolution are easier to

observe and interpret in oceanic islands thanks to the restricted distribution areas of the species and the simplified biota. Second, despite their small physical dimension, oceanic islands may offer a diversity of habitats with as many possible directions for diversification and speciation. Third, provided that contacts with the neighboring lands have been limited, species have evolved in relative genetic isolation and have diverged along their own evolutionary trajectories. When volcanic activity is ongoing, it also provides a dynamic evolutionary context with the creation/suppression of biological niches. All the above characteristics have favored high levels of endemism in oceanic islands that contribute to their frequent biodiversity hotspot status (Myers et al. 2000).

Reunion Island (2512 km²) is a young island belonging to the Mascarene archipelago (~−2 Ma, Chevallier and Vatin-Perignon 1982), located in the southern Indian Ocean (55°30′E; 21°10′S), around 700 km off the coast of Madagascar. Volcanic activity and erosion have shaped a

✉ Philippe Lashermes
  philippe.lashermes@ird.fr

[1] IRD, University of Montpellier, DIADE, Montpellier, France

high relief (Piton des Neiges, 3071 m a.s.l.) with dramatically rugged topography, leading to notable climatic heterogeneity at small spatial scale and high habitat diversity (Strasberg et al. 2005). Compared to the surrounding islands, human impact on the native forest have been limited: around one third of the original vegetation has been preserved and shows substantial endemism, with 28% of endemism in flowering plants (CBNM 2018). Human disturbances were concentrated in lowlands, almost preserving the native forest of intermediary and high elevation. Although the insular context of Reunion Island offers an excellent framework for studying patterns and processes of evolution, in fact, few studies have focused on the subject.

Gene flows by dispersal strongly influence the genetic diversity and spatial genetic structure of plant populations. Two main patterns of gene flow are reported to be associated with population divergence: isolation-by-distance (IBD) and isolation-by-environment (IBE) (Wang et al. 2013; Sexton et al. 2014). Under IBD, genetic differentiation increases with geographic distance irrespective of environmental differences among populations. It results from reduced gene flow between distant populations and drift (Wright 1943). Hence, in a case of strict IBD, the dispersal abilities of the species and the size of the population are the main factors to take into consideration to explain differentiation (Slatkin 1993). Alternatively, the pattern of IBE is depicted by increasing genetic differentiation with environmental dissimilarity independently of geographic distance. Lower gene flow occurs among populations with different environmental features (Wang et al. 2013; Sexton et al. 2014). Several mechanisms may drive the pattern of IBE including selection and local adaptation (Nosil et al. 2009; Edelaar et al. 2012), or non-random mating due to adaptation or phenotypic plasticity (Sexton et al. 2014). In practice, geographical and environmental factors are not mutually exclusive and can jointly contribute to population divergences. Especially on oceanic islands, IBD and IBE can coexist provided that the island is not too small and present substantial habitat diversity. Reunion Island offers an ideal context to assess the role of environmental factors when geographic distances are reduced. Previous studies on endemic plants showed diverse patterns for explaining population divergence according to the species (Litrico et al. 2005; Mallet et al. 2014; Jaros et al. 2016; Blambert et al. 2016; Bourgeois et al. 2018).

The genus *Coffea* (Rubiaceae) comprises 124 species most of which are distributed in tropical Africa, Madagascar and islands in the western Indian Ocean, and when *Psilanthus* species are included, even in tropical Asia and Australia (Davis et al. 2006, 2011). *Coffea* species are perennial woody bushes or trees. With the exception of *C. arabica* ($2n = 4 \times = 44$), all *Coffea* species are diploid ($2n = 2 \times = 22$). Among the latter, coffee trees from the Mascarene Islands (Mauritius and Reunion Island) are of monophyletic origin (Nowak et al. 2012; Hamon et al. 2017; Kainulainen et al. 2017) and form the Mascarocoffea group. The ancestral lineage of *Coffea* is estimated to have colonized the island of Mauritius at ~1.8–2.6 Ma BP, likely from either Madagascar or Africa, through long-distance dispersal events (Nowak et al. 2014). In Mauritius, four species were characterized (*C. myrtifolia*, *C. bernardiniana*, *C. macrocarpa* and *C. mauritiana*). Several cases of natural hybridization between them were described, especially considering the crossing *C. macrocarpa* × *C. mauritiana* (Dulloo and Couturon 2009). In Reunion Island, a unique species was described as a botanical form of *C. mauritiana* (Davis et al. 2006). However, the origin of *C. mauritiana* from Reunion Island and its genetic relationships with its Mauritian counterparts remain to be elucidated. Although medicinal properties of *C. mauritiana* are well known (Giraud-Techer et al. 2016), the human use of the species has remained very marginal. During the XVIII century, cultivated *C. arabica* was introduced in Mauritius and Reunion Island. However, the possibility of gene flows from the introduced *C. arabica* species to the endemic *C. mauritiana* species can be discarded due to strong reproductive barriers such as non-overlapping flowering time (Noirot et al. 2016) and differences in ploidy levels (Mahé et al. 2007).

This study focusses on genetic diversity and population divergence of an indigenous tree (*C. mauritiana*) in Reunion Island forest. *C. mauritiana* is present as small patches of individuals widely distributed throughout the rainforest (270–1500 m a.s.l.) (Dulloo et al. 1999). Its extant habitat encompasses an altitudinal gradient of temperature and contrasted rainfall regimes. Hence, *C. mauritiana* has a wide distribution in a heterogeneous landscape. Despite the lake of genomic studies on *C. mauritiana*, the species represents an interesting model case to gain insights into the patterns and factors of differentiation within a small oceanic island. Based on previous studies on endemic species (Litrico et al. 2005; Mallet et al. 2014; Jaros et al. 2016; Blambert et al. 2016; Bourgeois et al. 2018), genetic divergences between local populations were expected in spite of the small area of the island. Among factors explaining for population divergences, the importance of environmental factors has already been highlighted in Reunion Island species (Mallet et al. 2014). Hence, an influence of the environment in population divergences may also be considered here. Using genome-wide DArTseq markers, we more specifically aim to (1) examine patterns of genetic diversity of *C. mauritiana* in Reunion Island, (2) explore the genetic connectivity between *C. mauritiana* from Reunion Island and the Mauritian coffee species, (3) investigate the role of climate and geographic distance in

Fig. 1 Sampling localities and climates on Reunion Island. **a** Provenance of source populations of *C. mauritiana* on Reunion Island. PdN Piton des Neiges, PdF Piton de la Fournaise. **b** Spatial rainfall distribution over Reunion Island
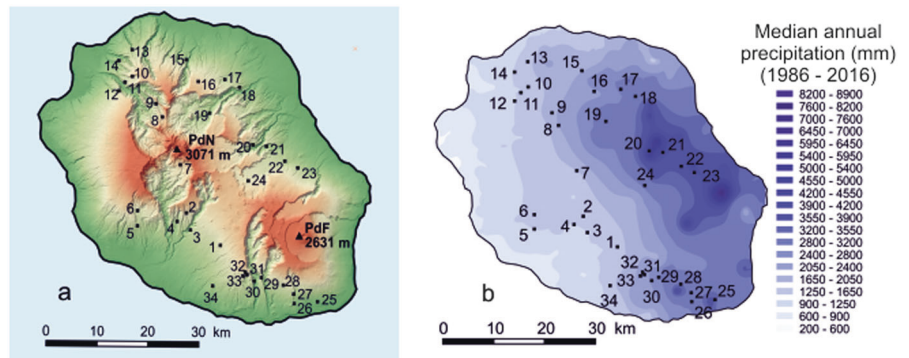
shaping population divergences in Reunion Island, and (4) highlight historical migration and mixture events.

Material & methods

Plant material and DNA extraction

C. mauritiana is a sub-canopy tree widely distributed in the forest of low and intermediary altitude from Reunion Island. In its natural range, the species is present in diverse conditions of temperature (13–24 °C in mean annual temperature) and rainfall (from 1100 to up to 5000 mm in median annual precipitation). Apart from lowlands, the original habitat of *C. mauritiana* in Reunion Island was almost well preserved from human disturbances. Collecting missions were organized mainly based on previous records available in the "Mascarine" database (Mascarine, *système d'information sur la flore et les habitats de La Réunion*) developed by the *Conservatoire Botanique National de Mascarin* (CBNM, Saint-Leu, Reunion Island). Sampling was designed to cover the environmental range of habitats and the geographic distribution of the species in Reunion Island as well as possible. The number of samples per source population (~ 10) differed depending on accessibility and on the number of trees identified. Special care was taken to collect distantly located trees to minimize relatedness between the trees sampled. Nineteen additional *C. mauritiana* trees were sampled at the international *Coffea* collection (IRD, Saint-Pierre, Reunion Island), not only providing more individuals per source population but also individuals from two additional source populations (Salazie and Piton Fougères). In all, a total of 326 *C. mauritiana* individuals from 34 source populations were collected (Fig. 1a, Table 1). To develop phylogenetic approaches, eight representatives of the Mauritian endemic coffee species were also sampled at the IRD *Coffea* collection, as well as representatives of the main other biogeographic groups identified for *Coffea* species (West and Central Africa, East Africa and Madagascar) (Table S1).

Leaves were freeze-dried and stored at −80 °C until DNA extraction. Approximately 70 mg of lyophilized leaf material were used for genomic DNA extraction using an in-house MATAB based protocol. The DNA was treated with RNase A (Promega, USA). DNA concentration was estimated by 1% (*w/v*) agarose gel electrophoresis, in comparison with a standard series of lambda *DNA/EcoRI + HindIII* markers (Promega, USA).

Climatic data

The climate in Reunion Island is mainly characterized by a thermal gradient associated with elevation (annual mean temperature 0–25 °C) and contrasted annual rainfall regimes (500–11,000 mm; Météo-France 2018). High rainfall is recorded on the windward side of the island (east coast), whereas the climate is substantially drier on the leeward side (west coast; Fig. 1b).

For each sampled source population, daily climatic data (precipitation, temperature, potential evapotranspiration and incident radiation) were obtained from the METEOR meteorological data repository (CIRAD, UR AÏDA 2018) for the period 2002–2016 and using an optimized interpolation of records for any geographical point in Reunion Island (https://smartis.re/METEOR, Chopart et al. 2002). The number of meteorological stations used for interpolation varies from 57 (temperature records) to 124 (rainfall records) according to the climatic record being considered. Twelve derived climate variables were calculated as mean values over the period: mean of minimum, average, and maximum daily temperature (Tmin, Tmean, Tmax, respectively, °C), average daily temperature amplitude (Trange, °C), mean of minimum, average, and maximum monthly precipitation (Pmin, Pmean, Pmax, mm), average monthly potential evapotranspiration (PET, mm), mean number of consecutive dry days (CDD, counts of days with less than 5 mm of precipitation), rainfall concentration index (RCI, mm) (average intra-annual variation in monthly precipitation), drought severity index (DSI, dry period weighted intensity), average water balance of driest months (WBmin,

Table 1 Source population characteristics and summary statistics of genetic diversity estimates obtained from the analysis of 3953 SNP markers

Population number	Sampling sites	Symbol	Elevation (m)	N	He	Ho
	West				*0.13*	*0.13*
1	**Pointe de Bras-Creux**	▲	1077	10	0.14	0.14
2	**Dimitile**	▼	1138	10	0.13	0.13
3	**Bras de Pontho**	●	689	10	0.13	0.13
4	**Bras-Long Entre-Deux**	✳	670	10	0.13	0.13
5	**Pièce Jeanne**	✳	805	9	0.14	0.13
6	**Les Makes**	■	965	10	0.13	0.13
7	**Cilaos Roche Merveilleuse**	▲	1429	10	0.11	0.11
	North-West				*0.13*	*0.13*
8	**Mafate – Sentier Augustave**	◇	1271	10	0.16	0.15
9	**Source oiseaux/Aurère**	✿	897	10	0.12	0.12
10	**Dos d'Âne**	■	1251	9	0.13	0.12
11	**Forêt de Mafate – Dos d'Âne**	▲	845	9	0.12	0.12
12	**Forêt de Sans-Soucis**	✳	1071	10	0.13	0.13
13	**La Grande Montagne**	▼	613	10	0.12	0.12
14	**La Grande Chaloupe**	●	636	10	0.13	0.13
15	**Ravine Montauban**	○	569	5	0.14	0.14
	North-East				*0.16*	*0.16*
16	**Piton Fougères**	●	1364	3	0.16	0.15
17	**Hauts de la Perrière**	▲	792	10	0.17	0.17
18	**Forêt de Dioré**	✳	760	10	0.17	0.17
19	**Salazie**	■	600	3	0.12	0.14
	East				*0.17*	*0.16*
20	**Takamaka**	✛	803	12	0.16	0.16
21	**Grand-Etang**	✿	526	13	0.17	0.17
22	**Forêt de Sainte-Marguerite**	▼	702	10	0.17	0.16
23	**Piton Armand les Hauts**	○	723	10	0.17	0.16
24	**Piton des Songes**	◇	1250	11	0.16	0.15
	South				*0.16*	*0.16*
25	**Mare Longue**	○	223	13	0.17	0.17
26	**Piton Bernard**	▼	466	10	0.16	0.16
27	**Basse Vallée**	◆	718	10	0.16	0.17
28	**Forêt Jacques Payet**	✛	969	9	0.17	0.16
29	**Grand Coude**	▲	893	12	0.16	0.16
30	**Trois sources/Rivière des remparts**	●	410	9	0.15	0.15
31	**Plaine des Grègues**	✿	864	9	0.15	0.15
32	**Piton du Rond**	✳	1245	5	0.15	0.15
33	**Petite Ile**	■	1077	12	0.16	0.16
34	**Piton Mont-Vert**	◇	556	10	0.15	0.15
	All sites		-	323	0.14	0.18

Arithmetic means for each region are in italics.

N number of individual plants used for subsequent analyses, *He* expected heterozygosity, *Ho* observed heterozygosity.

mm). Pearson correlation coefficients and their *p*-value adjusted by Holm's method were calculated between all complete pairs of variables using the Rcmdr R package (Fox J 2005) and a principal component analysis (PCA) was

performed in R 3.3.2 to determine the main climatic factors differentiating the natural habitats.

DArTseq analysis

Genomic DNA from each accession was standardized to a concentration of 50–100 ng μl^{-1} and sent to Diversity Array Technology Pty. Ltd. Australia (http://www.diversityarrays.com/). One DNA sample was duplicated to be used as technical replicate. Samples were genotyped by Diversity Arrays Technology Pty Ltd using the DArTseq platform (Sansaloni et al. 2011). The enzyme combination *PstI/HapII* was selected for library construction and fragments were sequenced using Illumina technology (Illumina, San Diego, CA, USA). Genotype data were generated using proprietary analytical pipelines (DArT Pty Ltd). Identical sequences were collapsed into "fastqcall files". These files were then processed to produce "SNP" tables (codominant markers) and "silicoDArT" tables (dominant markers). Finally, the different sequences were blasted (E-value = 5×10^{-7}; Minimum sequence identity = 80%) against the publicly available *C. canephora* genome (http://coffee-genome.org/) to determine the corresponding genomic positions.

Data filtering for clustering analysis

The DArTseq-derived SNP dataset was subset for phylogenetic analysis purposes. Regarding *C. mauritiana* from Reunion Island, seven accessions were selected to cover the natural distribution of the species. Concerning the other DArTseq genotyped species, all accessions were included. The resulting dataset included 24 accessions. SNP showing missing values (not available, NAs) higher than 8% in all *Coffea* accessions were discarded. SNP with minor allele frequency (MAF) lower than 0.02 (just one alternative allele) among the seven *C. mauritiana* from Reunion Island were filtered out. Regarding the other *Coffea* accessions, no MAF filtering was performed.

The genetic distance (simple matching) between accessions was calculated from the filtered SNPs using DARwin 6.0.14 software (Perrier and Jacquemoud-Collet 2006), and a neighbor-joining tree was computed. Robustness of the nodes was evaluated by bootstrapping (1000 bootstraps).

Data filtering for population genetic studies and assessment of genetic diversity

The DArTseq-derived SNP data set corresponding to the *C. mauritiana* samples from Reunion Island was filtered to only retain individuals with NAs < 25%. SNPs with NAs > 5% and MAF < 0.002 were discarded. The expected heterozygosity (He), and the observed heterozygosity (Ho) were calculated for the 34 source populations, using

ARLEQUIN version 3.5.2 (Excoffier et al. 2005). Next, average He and Ho were calculated per geographic regions of origin (Table 1).

Otherwise, the SilicoDArT data set was filtered to avoid markers with NAs >5%; markers for which a restriction site was detected only once; markers whose location on *C. canephora* genome could not be determined. Individuals were also filtered to only retain those with NAs <5%.

Assessment of genetic diversity structure in *C. mauritiana*

A tree-based method was first performed using DARwin. Genetic distance was computed between each pair of individuals (at least 80% of pairwise completion; 1000 bootstraps) using a simple matching distance and the corresponding neighbor-joining tree was computed.

Next, the sNMF program (sparse non-negative matrix factorization algorithms) implemented in the LEA R package v. 2.0.0. (Frichot and François 2015) was used to infer individual ancestry coefficients. The program relies on sparse non-negative matrix factorization (sNMF) algorithms to estimate population clustering, and to compute least squares estimates of the ancestry coefficients (Frichot et al. 2014). sNMF algorithms were run 100 times for each putative number of ancestral populations (K), ranging from K = 1–30 (α = 1000). The cross-entropy criteria (CEC) helped to choose the optimal K value, a smaller value of the CEC generally indicating a better run (Frichot et al. 2014). The K values minimizing the CEC were first retained. The Q matrices and the bar plots associated with each selected K value were also screened to help in decision. The K value leading to the lowest number of admixture events and greatest genetic proximity of individuals originating from the same source population was selected. Finally, the best run (e.g. lowest CEC) out of the 100 runs corresponding to the selected K was retained for the graphical representation of ancestry estimates.

Population differentiation analysis

Subsequent analyses were performed on 32 source populations, excluding source populations comprising fewer than 5 individuals. Analyses of molecular variance (AMOVA) were performed using ARLEQUIN. The tested model was the partitioning of total variance among the inferred genetic clusters from the sNMF analysis, among source populations within clusters, among individuals within source populations and within all individuals. When individuals from the same source population were distributed in several clusters, or showed more than 30% admixture, these individuals were discarded from the analysis. Weir and Cockerham's F_{ST} was also computed among pairs of the 32 source

populations retained, using ARLEQUIN. The statistical significance of the Weir and Cockerham's F_{ST} was assessed by 1000 random permutations of individuals among populations.

SNP filtering and Treemix analysis

The topology of relationships and migrations between the K populations of *C. mauritiana* were assessed by Treemix v. 1. 12 (Pickrell and Pritchard 2012) using population allelic frequencies estimated from the above mentioned filtered SNP dataset. As *C. myrtifolia* and *C. bernardiniana* were genotyped for three accessions respectively, these species were set as outgroups.

The DArTseq-derived SNP dataset was subset for including *C. mauritiana* from Reunion Island, *C. myrtifolia* and *C. bernardiniana* samples. The resulting dataset was filtered to discard SNPs with missing data and markers with MAF <0.002. Two different filtering methods were used to account for linkage disequilibrium, depending on whether information was available on the position on the coffee reference genome or not. When information on the position of the SNP was available, SNPs were filtered for $r^2 < 0.2$ using the bigsnpr and bigstatsr R package (function *snp_clumping*; Privé et al. 2018). When no information on position was available, only one randomly chosen SNP per sequence was retained.

The dataset was divided into windows of 20 SNPs to account for potential remaining linkage disequilibrium by jackknifing. Up to three migration edges were tested (*m* ranging from 1–3). Each Treemix analysis was computed three times to ensure consistency of the results. As all three runs were consistent, only one run is presented in the results.

The three-population test *f3*(X;Y,W) included in the TreeMix package was performed to test for mixture events for all combinations of three points out of the *K* clusters of *C. mauritiana* samples. A window of 20 SNPs was used to account for linkage disequilibrium by jackknifing. A negative *f3* statistic provided evidence that X originated from a mixture between Y and W. According to Reich et al. (2009), Z-scores below −2 were considered as significant.

Effects of environmental and geographical factors on reproductive isolation

A distance-based redundancy analysis (dbRDA) was used to estimate the relative contribution of IBD and IBE to the differentiation of *C. mauritiana* source populations. The analysis consisted in multivariate linear regression between a first matrix of dependent variables (pairwise genetic distances) and a second matrix of explanatory variables (either

geographic or environmental factors or both) where collinearity among variables must be minimized (Legendre and Legendre 1998; Orsini et al. 2013). The pairwise F_{ST} were set as dependent variable. The explanatory variables were set as follows: (i) Geographic variables. Principal coordinates of neighborhood matrix (PCNM) was used to transform the original geographical matrix of Euclidean distances into uncorrelated vectors. As positive eigenvectors better describe the geographical distance (Dray et al. 2006), we kept the eigenvalues of first four positive vectors (geo1, geo2, geo3, geo4) to account for patterns of IBD. (ii) Environmental variables. Climatic variables that minimized collinearity were selected previous to dbRDA. Taking into consideration both the results of Pearson's correlations and PCA analyses, five climatic variables were retained: Tmean, Trange, CDD, PET, RCI (Fig. S1, Table S2).

dbRDA was performed using the function *capscale* in the vegan R package v. 2.4-2 (Oksanen et al. 2017). To identify the best model, the dbRDA was run on nested models, which optimizes the model by taking the Akaike information criterion (AIC) into consideration. Both forward and backward processes were used to select direction. The significance of the joint effect of predictors as well as their marginal effects were tested using multivariate F-statistics with 10,000 permutations. Multi-collinearity among the explanatory variables was also evaluated by computing the variance inflation factor (VIF), a VIF <10 being indicative of low multi-collinearity in the variables. The R^2_{adj} was computed to evaluate the quality of the model. This procedure was implemented for the 32 source populations which pairwise F_{ST} was available, and was next implemented for 10 source populations from the southern part of the island. The contribution of each variable from the optimized model, kept independently from the remaining explanatory variables, was estimated by partial dbRDA. It also enabled estimation of IBD (sum of contributions of variables of geography) and IBE (sum of contributions of environmental variables). Finally, the relative contribution of both IBD and IBE and their intersection was estimated by variance partitioning.

Results

DArTseq analysis and SNP detection

A total of 31,398 codominant SNPs and 16,775 Silico-DArTs were obtained from the DArTseq analysis of the overall coffee accessions. In the analysis, special attention was paid to the SNP data. The average genotype call rate was of 83.2%, with SNP scoring reproducibility of 99.7%. The rate of SNP genotyping error was 1.7%.
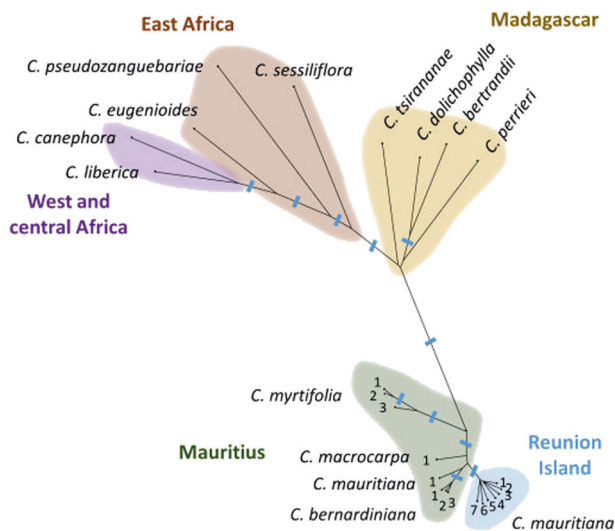
Fig. 2 Unweighted NJ tree based on simple matching distances estimated from 2251 SNP markers, among 24 coffee accessions. Clades supported by a bootstrap value of 100% are indicated by a blue dash

Origin of *C. mauritiana* from Reunion Island and genetic connectivity between *Mascarocoffea* species

Genetic distances among the 24 selected coffee accessions were calculated from the 2251 filtered SNPs. The resulting neighbor-joining tree showed that accessions were grouped according to their geographic origin (Fig. 2). Among the Mauritian representatives, *C. myrtifolia* appeared to be the most divergent, while *C. macrocarpa*, *C. mauritiana*, and *C. bernardiniana* were more closely related. Accessions of *C. mauritiana* from Reunion Island formed a distinct cluster that is divergent from the Mauritian accessions. This cluster was supported by high bootstrap values. Finally, *C. mauritiana* from Reunion Island was as distant from *C. bernardiana* representatives, *C. macrocarpa* and *C. mauritiana* from Mauritius.

SNP filtering and assessment of genetic diversity

The complete DArTseq-derived SNP dataset was filtered for quality criteria leading to 3953 SNPs from 323 *C. mauritiana* samples from Reunion Island. This dataset was considered for analysis of genetic diversity and population structure. Total expected heterozygosity (He) and observed heterozygosity (Ho) among the 34 *C. mauritiana* source populations collected in Reunion Island were 0.14 and 0.18, respectively (Table 1). He values were low among source populations from the north-western and western parts of the island (He values <0.15), except for source population 8 (He = 0.16). Source populations from the east (mean He = 0.17), north–east (mean He = 0.16) and south (mean He = 0.16) showed the highest He values. The same pattern was observed regarding Ho values.

Detailed assessment of population genetic structure

The overall structure in the panel was visualized in a neighbor-joining (NJ) tree (Fig. 3a). Individuals belonging to the same source population preferentially clustered together, leading to a clear separation between most source populations. The majority of bootstrap values associated with clusters of individuals from the same source population were higher than 80%. Geographically close source populations also tended to cluster together, indicating genetic proximity according to geographic distribution. This mostly concerned north-eastern and eastern (17, 18, 20, 21, 22, 23, 24), north-western (9, 10, 11, 12, 13, 13, 14), western (2, 3, 4, 5, 6), and southern regions (south-eastern: 27, 28; south-western: 29, 31, 32, 33). It should be noted that within the eastern, north-eastern, western, and southern regions, some individuals were not grouped according to the source populations. Furthermore, the shortest orthodromic distances were recorded between source populations in the southern region (32–33: 870 m; 31–33: 840 m; 31–32: 615 m) (Table S3).

The overall observation of individual clustering according to source population affiliation was confirmed by the NJ tree using filtered SilicoDArT markers (Fig. S2). The tendency of geographically close populations to be grouped together was also observed. As reported previously, individuals from the western and south-western regions displayed imperfect grouping according to the source population. In the western region, both NJ tree based on SilicoDArT markers and SNP markers showed that one individual sampled in location 3 clustered with samples from location 5 while the distance between the two locations was 10.5 km (Table S3). In contrast, eastern individuals were better grouped than previously.

The sNMF program was used to investigate population clustering of the 323 *C. mauritiana* individuals in K clusters and to infer individual ancestry coefficients. The K values ranging from 5 to 8 minimized the cross-entropy criteria (CEC) (Fig. S3a). An overview of the Q matrices (data not shown) and their graphical representations (Fig. S3b,c,d) led us to select a structure in seven main clusters as the most relevant since K = 7 minimized admixture events while mostly preserving the source population affiliation (Fig. 3b). The bar plot revealed that clusters of individuals matched their geographic region of origin on the island (north–east, east, west, north–west). Source populations from the south were distributed in two clusters corresponding to south–eastern and south–western regions, respectively. Moreover, one cluster (hereafter called central cluster) was formed by two source populations (8 and 19) located respectively in two neighboring cirques in the center of the island (Fig. 3c). Individuals from location 8 showed little admixture (<10%) with the north–western cluster.
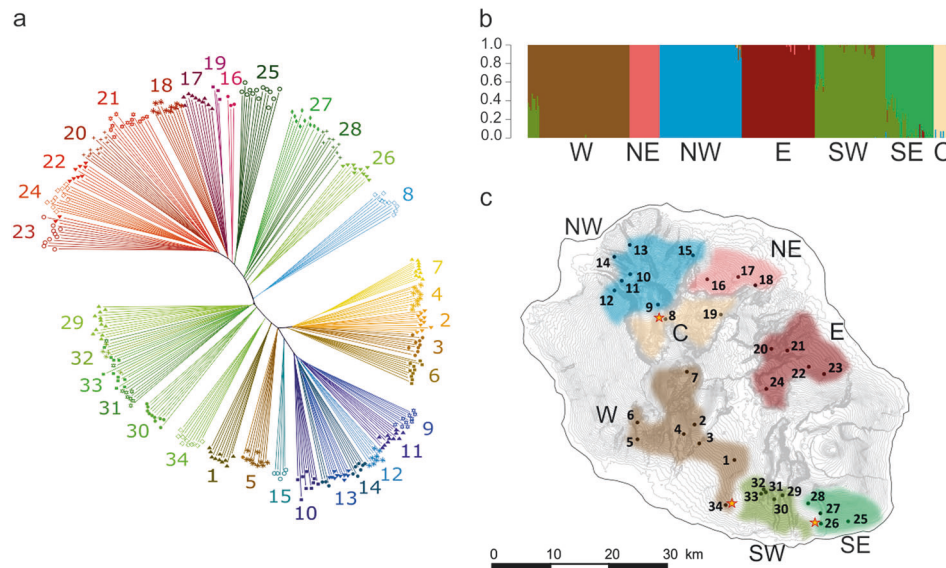
Fig. 3 Spatial genetic structure of 323 *C. mauritiana* trees across 34 sampling sites in Reunion Island, estimated from 3953 SNP markers. **a** Unweighted NJ tree based on simple matching distances. Correspondence between the population number and the name of the population are listed in Table 1. Colors distinguish the main geographical sampling regions. Red: east and north–east; Blue: north-west; Yellow and brown: west; Green: south. **b** Group assignment from sNMF at $K = 7$. Biogeographic regions associated with the clusters are represented by seven colors. Each individual tree is represented by a bar and coefficient ancestry relative to each cluster is indicated by colored segments. W western genetic group, NE north-eastern genetic group, NW north-western genetic group, E eastern genetic group, SW south-western genetic group, SE south-eastern genetic group, C central genetic group. **c** Spatial representation of the genetic structure of *C. mauritiana* populations obtained from sNMF analysis ($K = 7$). Each sampling site is represented by a dot. Colors distinguish the genetic groups inferred from sNMF. Yellow stars indicate source populations showing admixed individuals

Individuals from the east and north–east were divided into two separate clusters. The limits of the regional clusters almost preserved the population source origin except for location 26: six individuals belonged to the south–western cluster while the other four individuals were placed in the south–eastern cluster, revealing very high admixture levels (up to 48%). Individuals from location 34 (western cluster) also showed high admixture levels (from 29 to 47%) indicating that admixture occurred between the western and south–western clusters. The results of the NJ trees, and sNMF analyses were largely consistent with an overall genetic structure at regional scale and finer substructure in source populations, except for accessions of the south–west regional cluster for which less grouping by source population was observed.

Population differentiation analysis

Pairwise F_{ST} values were computed among the source populations comprising more than five individuals (e.g. 32 source populations). The F_{ST} values ranged from 0.014–0.430, with an average genetic differentiation of 0.168 (Table S4). All F_{ST} values were statistically significant ($P < 0.001$). The F_{ST} values for source populations belonging to the same regional cluster were low whereas higher F_{ST} values were observed between source

populations from different regional clusters. Pairwise F_{ST} was also computed between the seven regional clusters (Table 2). Low F_{ST} values were observed between regional clusters in the east (north–east, east, and south–east) and west (north–west, west, and south–west), respectively. The highest differentiation was observed between clusters from the east and west coast, the maximum F_{ST} value was recorded for the north–western and eastern pair ($F_{ST} = 0.26$), while the minimum F_{ST} value was recorded for the north–eastern and eastern pair ($F_{ST} = 0.04$). A low F_{ST} value was also observed between southern clusters ($F_{ST} = 0.08$)

The results from the AMOVA were consistent with the differentiation depicted by the pairwise F_{ST}, with 15.2% of the global variation between regional clusters, and 5.0% between source populations belonging to the same genetic cluster.

Population genetic relationships

After filtering for quality criteria and linkage disequilibrium, a total of 1280 SNPs for the 323 *C. mauritiana*, the three *C. myrtifolia* and the three *C. bernardiniana* samples were retained. A Treemix analysis was conducted to examine the topology of relationships and migration history among the seven regional clusters. With *C.*

Table 2 Weir and Cockerham's pairwise F_{ST} calculated between the seven clusters inferred by sNMF analysis.

	W	NE	NW	E	SW	SE	C
W	0	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
NE	0.22	0	<0.001	<0.001	<0.001	<0.001	<0.001
NW	0.11	0.24	0	<0.001	<0.001	<0.001	<0.001
E	0.24	0.04	0.26	0	<0.001	<0.001	<0.001
SW	0.10	0.13	0.17	0.15	0	<0.001	<0.001
SE	0.18	0.08	0.23	0.07	0.08	0	<0.001
C	0.17	0.09	0.16	0.12	0.13	0.13	0

F_{ST} values are below the diagonal and *p*-values above. Yellow: $F_{ST} \leq 0.10$; Green: $0.10 < F_{ST} \leq 0.20$; Red: $F_{ST} > 0.20$.

W western genetic group, *NE* north-eastern genetic group, *NW* north-western genetic group, *E* eastern genetic group, *SW* south-western genetic group, *SE* south-eastern genetic group, *C* central genetic group.
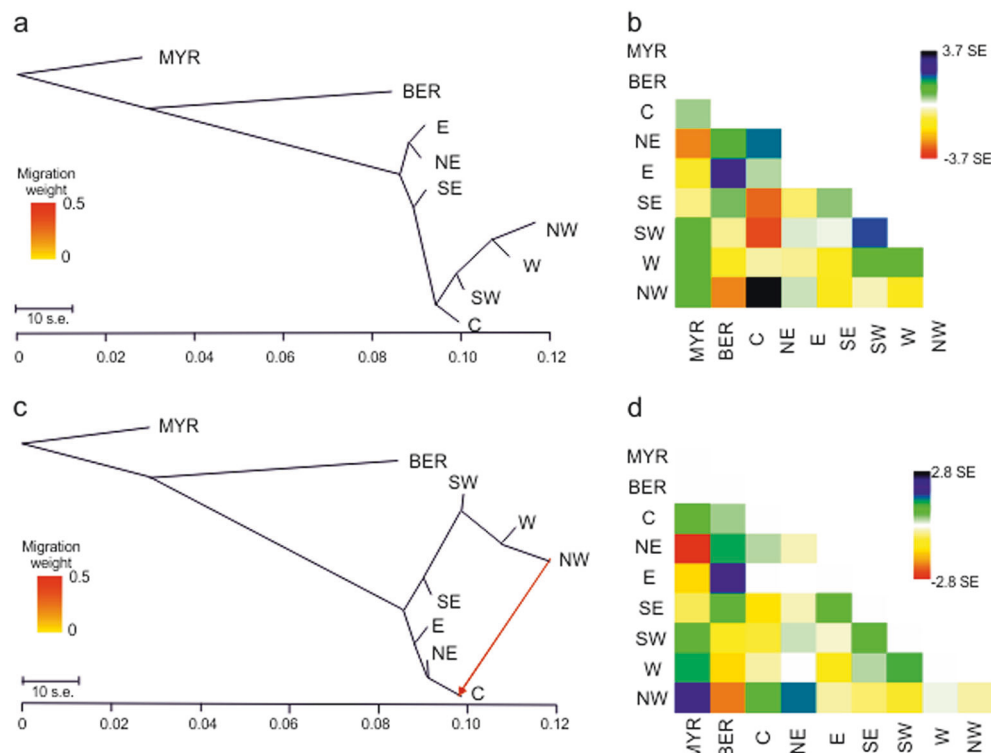

Fig. 4 Results of TreeMix analysis of the phylogenic relationships among the seven clusters defined by sNMF analysis. *C. myrtifolia* (MYR) and *C. bernardiniana* (BER) served as outgroup populations. E eastern cluster, NE north-eastern cluster, SE south-eastern cluster, C central cluster, SW south-western cluster, W western cluster, NW north-western cluster. **a** Maximum-likelihood tree of 323 *C. mauritiana* divided into the seven population groups included in the analysis. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred in each branch. **b** Residual heatmap from the tree in **a**. Pairs with high residual values are candidates for migration events. **c** Maximum- likelihood tree that best fits the data and its associated residual heatmap **d**. The migration weight indicates the proportion of ancestry deriving from the migration edge

myrtifolia and *C. bernardiniana* individuals used as outgroups, the ML tree showed a split between regional clusters from the east coast (north–eastern and eastern clusters) and the other clusters from the south, the center and the west coast (Fig. 4a). The drift parameter ranged from 0.09 to 0.12, indicating low genetic divergence among the *C. mauritiana* regional clusters. The north–western cluster showed the highest drift parameter (~0.12), indicating the highest divergence from a set of ancestral allele frequencies.

The tree representation captured the total variation (98.9%) well. The residual heatmap revealed that the north–western and the central clusters were the worst represented, with a standard error of 3.7 (Fig. 4b). Migration edges (parameter m) were then computed to see whether gene flow events would improve the percentage of total variance and reduce the global standard error. An optimized model was obtained by inferring one migration edge from the north–western cluster to the center (Fig. 4c). This historical migration was in accordance with the few admixture rates found in the

sNMF analysis (<10%) between individuals from source population 8 (central cluster) and the north–western cluster. The ML tree supported an additional 0.54% of total variance while decreasing the global standard error and improving the representation of both the north–western and central clusters (Fig. 4d). Adding more migration edges did not significantly improve the ML tree. So an ML tree with only one migration edge was retained (weight: 41.6%; se: 0.021; P-value $<2.23E^{-6}$).

An exhaustive *f3-statistics* test (three-population test) was applied to combinations of three regional groups to further investigate their mixture history. The results presented in Table S5 suggested that the south–eastern cluster originated from a mixture between the western and eastern clusters (Z-score: −2.12). In addition, a complex history was obtained for the south-western cluster: significant evidence was found for a mixture between western and eastern clusters (Z-score: −2.98), and between western and south-eastern clusters (Z-score: −3.29). Weak evidence was also found for a mixed origin of the central cluster from the north-eastern and the north-western clusters (Z-score: −1.95).

Importance of environmental and geographical factors in explaining genetic differentiation

A series of full and partial redundancy analyses (RDA) was used to assess the amount of genetic divergence among source populations (pairwise F_{ST}) explained by uncorrelated geographic vectors and climatic factors. The optimized model (AIC: −72.96; P-value < 0.001; R^2_{adj}: 85.0%) constrained 88.4% of total inertia, and retained seven variables: the four vectors of geographical distance, CDD (mean number of consecutive dry days), PET (average monthly potential evapotranspiration) and Tmean (mean of average daily temperature) (Table 3, Fig. S4). The VIF values were low, revealing relatively low collinearity in predictors, except for CDD (10.25). This relatively high value could be due to the strong collinearity between the geographic position and the rainfall regime in Reunion Island. The pure contribution of each independent variable was obtained after removing the confounding effects caused by the other significant factors (partial dbRDA). The contribution of geographic variables (IBD) to genetic divergence was higher than the contribution of environmental variables (IBE) (21.37 and 7.20%, respectively).

A first variance partitioning test was then conducted to assess the relative contribution of IBD, IBE and their intersection in explaining the genetic differentiation. The largest contributor to the genetic divergence was the intersection IBD ∩ IBE explaining 56.4% (Fig. 5a). A second variance partitioning test was then performed to determine which environmental variable was mainly involved in

Table 3 Detailed results of dbRDA significance tests for the global dbRDA analysis (after removing non-significant terms) and the marginal test, considering 32 source populations of *C. mauritiana*

	dbRDA			Partial dbRDA		
	% of variance explained	d.f.	P-value	% of variance explained	d.f.	P-value
Global analysis	85.03	7	<0.001			
Residuals	14.97	24	–		–	–
Marginal test						
Geography	47.56			21.37		
geo1	22.30	1	<0.001	10.54	1	<0.001
geo2	15.24	1	<0.001	7.01	1	<0.001
geo3	6.21	1	0.006	2.51	1	0.002
geo4	3.81	1	0.048	1.31	1	0.013
Environment	23.66			7.20[a]		
CDD	9.75	1	<0.001	4.27	1	<0.001
Tmean	6.06	1	0.008	2.43	1	0.014
PET	7.85	1	0.002	3.32	1	0.004
Residual	28.76	24	–	28.76	24	–

CDD mean number of consecutive dry days, *Tmean* mean average daily temperature, *PET* average monthly potential evapotranspiration.

In the dbRDA analysis, the marginal effects of each constraining factor were tested through 10,000 permutation tests, by removing each term one by one from the model containing all other terms. In partial dbRDA, the pure contribution of each factor was obtained after removing the confounding effects from the other constraining factors. geo1, geo2, geo3, geo4 are uncorrelated geographic factors

[a]Due to remnant correlations between environmental variables, the estimated contribution of environmental factors in partial dbRDA (7.20%) does not correspond to the sum of individual contributions of each environmental factor (10.02%).

IBD ∩ IBE (Fig. 5b). The intersection IBD ∩ CDD was responsible for 47% of variation, indicating that the main environmental factor involved in the intersection with geographic distance was CDD.

Even if we attempted to control for collinearity in data processing, this result highlighted the difficulty of disentangling the effect of geography and climate in Reunion Island, since geographically distant sites also tend to differ environmentally. To test for climatic effects while controlling for geographic distance, a dbRDA analysis was again performed on a smaller region of the island. The southern region was chosen as it contained sufficient source populations (10) to perform the analysis, with different cluster affiliation and a context of climatic heterogeneity at small spatial scale. Among the five climatic variables and the five geographic vectors, the optimized model only retained CDD and geo2. As the marginal effect of geo2 was not significant in the ANOVA table, it was removed from the explanatory variables, leading to a model with a single explanatory variable (CDD) that explained 49.4% of overall genetic
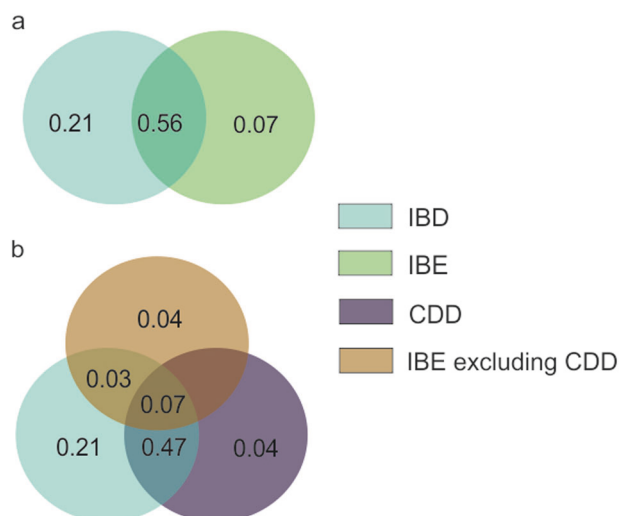
Fig. 5 Variance partitioning results of dbRDA analyses. IBD comprises four geographic vectors (geo1, geo2, geo3, geo4). IBE comprises three environmental variables: Tmean mean of average daily temperature, CDD mean number of consecutive dry days, PET average monthly potential evapotranspiration. **a** Variance partitioning between IBD and IBE. The overlapping zone represents the intersection between IBD and IBE (IBD ∩ IBE). **b** Variance partitioning between IBD, IBE excluding CDD and CDD. Values below zero are not shown

differentiation (AIC: -37.92, P-value $= 0.002$, R^2adj: 49.43%). These results confirmed the role of climate in explaining genetic divergence at reduced spatial scale and more precisely, the influence of the rainfall pattern.

Discussion

Colonization of Reunion Island and genetic connectivity with the Mauritian coffee species

The origin of *C. mauritiana* in Reunion Island and its genetic connectivity with the Mauritian counterparts are important to consider for a right understanding of the patterns of genetic differentiation of the species. The SNP-based clustering of *Coffea* species accessions was consistent with the known phylogeny among the *Coffea* genus (Nowak et al. 2012; Hamon et al. 2017; Kainulainen et al. 2017), supporting an independent grouping of the Mascarocoffea species. Since Mauritius is older (~-8 Ma, Moore et al. 2011) than Reunion Island (~-2 Ma, Chevallier and Vatin-Perignon 1982), an arrival of *C. mauritiana* in provenance of Mauritius was proposed as very likely (Nowak et al. 2014). Moreover considering our results, *C. mauritiana* from Reunion Island was as distant from *C. mauritiana* (Mauritius), than from *C. bernardiniana* representatives (Mauritius) and *C. macrocarpa* (Mauritius). Thus, the ancestors of Reunionnese *C. mauritiana* most likely diverged from the ancestral lineage at the origin of

Mauritian *C. mauritiana*, *C. bernardiniana* and *C. macrocarpa* species. Otherwise, the cluster formed by *C. mauritiana* accessions from Reunion Island was clearly divergent from the Mauritian species, arguing for genetic isolation after the ancestral events of colonization. So, Reunion Island colonization would result from few ancient events of long distance dispersal from Mauritius.

Besides their genetic divergences, the two forms of *C. mauritiana* in Reunion Island and Mauritius were also reported to be distinct for fruits and leaves phenotypes (Leroy 1963) making doubtful the unity of the species among islands. It pointed that the biological systematics of the Mascarocoffea species would require further clarification, especially regarding the two forms of *C. mauritiana*.

Genetic diversity

For many years, genetic variation has been considered to be reduced in plant species endemic to oceanic islands compared to their continental counterparts. This pattern was frequently attributed to the initial founder effect of colonization, limited immigration from mainland source areas and the consequence of genetic drift within small populations (Barrett et al. 1996; Frankham 1996, 1997). To better understand the level of genetic diversity in *C. mauritiana* (Ho $= 0.18$ and He $= 0.14$), indices can be compared to those obtained when considering a continental coffee species. Since *C. canephora* is an allogamous species and also presents the widest geographic distribution among *Coffea* species in Africa, it was chosen for comparison. Bearing *C. canephora* evolutionary history in mind (Cubry et al. 2013), one would thus expect much higher levels of genetic diversity in *C. canephora* than in *C. mauritiana*. Surprisingly, compared to *C. canephora*, genetic diversity indices in *C. mauritiana* were not so restricted as its evolutionary context would lead one to suppose (Ho: 0.1381 and He: 0.1763 in *C. mauritiana* versus Ho: 0.1405 and He: 0.1933 in *C. canephora*, Garavito et al. 2016). In fact, recent studies argue that after many generations on the island, genetic evidence of the initial founder effect is often no longer detected (García-Verdugo et al. 2015) and has no lasting impact on molecular evolution in endemic plants (James et al. 2016). For an appropriate interpretation of the level of genetic diversity in *C. mauritiana*, the biological characteristics of the species, such as their breeding system, aptitudes for dispersion and perenniality must be taken into account (García-Verdugo et al. 2014; Stuessy et al. 2014), as well as the island characteristics (Stuessy et al. 2014).

First, the breeding system has been shown to be a major determinant of plant genetic diversity (Glemin et al. 2006). In the *Coffea* genus, an S-RNase system was reported to be responsible for gametophytic self-incompatibility (Lashermes et al. 1996; Nowak et al. 2011; Asquini et al.

2011). Evidence for self-incompatibility in Mauritian coffee species has also been provided (Nowak et al. 2014), so self-incompatibility in *C. mauritiana* from Reunion Island is highly reliable, implying a strictly allogamous reproductive regime. Allogamy must have played a role in the persistence of significant levels of genetic diversity in *C. mauritiana*, despite a major contraction in the size of the population during its establishment on the island.

The life cycle characteristics of *C. mauritiana* could also have facilitated spatial expansion on Reunion Island. *C. mauritiana* is a woody species with a juvenile phase that can persist for several years before reproduction. This lengthy juvenile phase must have facilitated expansion and limited founder effects: delayed reproduction allowed the accumulation of several founders from the source population before reproduction began (Austerlitz et al. 2000; Born et al. 2008). Being long-lived, one individual can also contribute to the next generation several times.

Finally, geomorphological events may interfere with the interpretation of the current levels of genetic diversity. Nevertheless, its influence seems marginal here due to the age of the major geological events in Reunion Island: the main explosive period of the former volcano (Piton des Neiges) is quite ancient (around −150 ka, Gillot and Nativel 1982; Fretzdorff et al. 2000) and the current eruptions and collapses of the second volcano (Piton de la Fournaise) are restricted to a few zones in the south of the island (Michon and Saint-Ange 2008).

A genetic structure reflecting low dispersal and environmental influence

The fine scale genetic structure found among samples provides insights into the dispersal pattern of *C. mauritiana*. Individuals clustered according to their source population affiliation, with little evidence for gene flows between source populations from the same region. F_{ST} values were variable (F_{ST} between source populations ranged from 0.01 to 0.47) revealing higher differentiation between distant source populations. These elements suggest a pattern of local dispersal with isolation-by-distance (IBD). Indeed, IBD was of considerable importance in explaining divergence between source populations (21.4% of genetic differentiation was explained by geographic distance). Pollen and seed dispersal are major determinants of spatial genetic structure. Compared to pollen transport, seed dissemination appears as marginal in several species of coffee trees. Berries seem mainly dispersed by gravity to a few meters from the mother tree and possibly transported by run-off water (Berthaud 1986). Regarding dispersal by pollen transport, in *Coffea* species it is mainly ensured by both wind and bees, but the ratio "wind-pollination" to "bee-pollination" varies considerably among species (Klein et al.

2003; Ngo et al. 2011; Noirot et al. 2016). Since the pollen of self-incompatible coffee trees like *C. mauritiana* is light and dry (Ngo et al. 2011), anemophilous dispersal is probably involved. Nevertheless, in several coffee species, the efficiency of pollen transport by wind appeared to be limited to a distance of ~100 m from the mother tree (Charrier 1972; Berthaud 1986; Ngo et al. 2011). Regarding bee-pollination, the indigenous bee (*A. m. mellifera*, Techer et al. 2017) can be an historical pollinator of *C. mauritiana*.

Long distance dispersal can occasionally occur, as exemplified in the western region: one individual clustered with the accessions of another source population sampled 10.5 km away. A few candidates can be proposed, as they are known to be efficient dispersers of several small fleshy fruits in native species: the endemic bulbul (*Hypsipetes borbonicus*, Strasberg 1996; Thébaud and Strasberg 1999), the extinct starling (*Fregilupus varius*, Cheke and Hume 2008), and an endemic bat from Reunion Island and Mauritius (*Pteropus niger*, Nyhagen et al. 2005).

Additionally, the results of our study indicate that differentiation due to limited dispersal abilities was complemented by isolation due to environmental factors (IBE, 7.2% of genetic differentiation). Whether considering all source populations or focusing on the southern ones, the indices for dry period duration, CDD, appeared as the main climatic factor associated with differentiation. In coffee species, floral bud morphogenesis is induced by a dryness period and synchronized flowering occurs at the first rain showers of the rainy season (DaMatta 2018). Because of the heterogeneous rainfall pattern in Reunion Island, the duration of drought periods and the timing of the first rain showers must differ between sites. Thus, an asynchronous flowering in *C. mauritiana* should contribute to the establishment of a reproductive barrier between source populations.

History of population expansion

Results on relationships among the regional clusters provided further insights into present genetic diversity in *C. mauritiana* and population admixture. Overall, the low level of divergence between clusters suggests recent differentiation. The architecture of the TreeMix-derived ML tree led us to hypothesize distinct evolutionary pathways for regional groups from the eastern and western coasts, which is also supported by higher differentiation in pairwise F_{ST}. The central montane barrier must have been a major barrier to gene flow, separating the two sides of the island (east coast versus west coast). Moreover, results suggest the recent expansion of the species through a step-by-step colonization on both sides of the island. Along the western coast, the colonization front apparently proceeded from south to north. The north–western cluster appeared to be the

most differentiated (higher drift parameter) and the lowest levels of genetic diversity were also found in this regional group. Apart from separate evolution on the two sides of the island, a few ancestral mixtures were identified in the south–western and south–eastern clusters, attesting to a more complex history of population establishment in the southern part of the island. Due to proximity to the younger volcano (Piton de la Fournaise), the establishment of *C. mauritiana* in the south may have been locally impacted by several geologic events linked with volcanic activity (Michon and Saint-Ange 2008). Other gene flows between the north–western and the central clusters attest to the colonization of the central cirques by neighboring populations.

In conclusion, our study provides a comprehensive picture of the genetic diversity and structure of *C. mauritiana*, a widely represented forest tree in Reunion Island. We highlighted the joint contribution of geographic distance and rainfall regimes in shaping genetic differentiation in a small oceanic island. These results can be used to guide forest management for conservation while simultaneously favoring forest resilience to climate change (Fady et al. 2016). Moreover, the distribution of *C. mauritiana* in highly contrasted environments suggests local adaptation. Consequently, this work can also be used as a starting point for investigation of potential genetic variants underlying the adaptive traits in *C. mauritiana* populations. Finally, the information acquired on genetic relationships between populations provides insights into past forest dynamics. In depth studies are now needed to confirm the expansion scenario. Major forest transitions covering a recent period (<40,000 years before present) have been reported in Mauritius in association with climate change (Van Der Plas et al. 2012; de Boer et al. 2013). Similar reassortments in taxonomic composition of Reunion Island forest can be hypothesized.

Data archiving

DArTseq genotype data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.4vc0rg0

Compliance with ethical standards

References

Asquini E, Gerdol M, Gasperini D, Igic B, Graziosi G, Pallavicini A (2011) S-RNase-like sequences in styles of *Coffea* (Rubiaceae). Evidence for S-RNase based gametophytic self-incompatibility? Trop Plant Biol 4:237–249

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics 154:1309–1321

Barrett SCH, Emerson B, Mallet J (1996) The reproductive biology and genetics of island plants. Philos Trans R Soc B Biol Sci 351:725–733

Berthaud J (1986) Structure de l'espèce *C. canephora*: polymorphisme interne des populations. In: Les ressources génétiques pour l'amélioration des caféiers africains diploïdes, Travaux et documents. Institut français de recherche scientifique pour le développement en coopération. Éditions de l'ORSTOM: Paris, pp 73–96.

Blambert L, Mallet B, Humeau L, Pailler T (2016) Reproductive patterns, genetic diversity and inbreeding depression in two closely related *Jumellea* species with contrasting patterns of commonness and distribution. Ann Bot 118:93–103

de Boer EJ, Hooghiemstra H, Vincent Florens FB, Baider C, Engels S, Dakos V et al. (2013) Rapid succession of plant associations on the small ocean island of Mauritius at the onset of the Holocene. Quat Sci Rev 68:114–125

Born C, Kjellberg F, Chevallier M-H, Vignes H, Dikangadissi J-T, Sanguie J et al. (2008) Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, *Aucoumea klaineana*. Proc R Soc B Biol Sci 275:2171–2179

Bourgeois Y, Bertrand J, Delahaie B, Holota H, Thebaud C, Mila B (2018) Local adaptation and sexual selection drive intra-island diversification in a songbird lineage: differential divergence in autosomes and sex chromosomes. *bioRxiv* 353771.

CBNM (2018) Statistiques sur la flore. *Mascarine Cadetiana*. Available at: https://mascarine.cbnm.org/

Charrier A (1972) Etude de la pollinisation des caféiers cultivés sur la côte Est malgache par marquage du pollen au phosphore et au soufre radioactifs. Terre Malgache Tany Malagasy 12:229–249

Cheke A, Hume J (2008) Lost Land of the Dodo. T&AD Poyser, London

Chevallier L, Vatin-Perignon N (1982) Volcano-structural evolution of Piton des Neiges, Reunion Island, Indian Ocean. Bull Volcanol 45:285–298

Chopart J-L, Mézino M, Le Mezo L (2002) Relations entre l'altitude et la température mensuelle de l'air dans l'ouest de la Réunion. Rev Agric Et Sucrière De l'Ile Maurice 80:68–72

CIRAD, UR AÏDA (2018) METEOR. Available at: https://smartis.re/METEOR

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol 60:483–501

DaMatta FM (2018) Coffee tree growth and environmental acclimation. In: Lashermes P (ed) Achieving sustainable cultivation of coffee: breeding and quality traits. Burleigh Dodds Series in Agriculture Science, Cambridge, p 21–48

Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). Bot J Linn Soc 152:465–512

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*: *Psilanthus* subsumed in *Coffea*. Bot J Linn Soc 167:357–377

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196:483–493

Dulloo ME, Couturon E (2009) Collecting wild coffee species in Mauritius. Crop Wild Relat 7:21–22

Dulloo ME, Maxted N, Guarino L, Florens D, Newbury HJ, Lloyd BVF (1999) Ecogeographic survey of the genus *Coffea* in the Mascarene Islands. Bot J Linn Soc 131:263–284

Edelaar P, Alonso D, Lagerveld S, Senar JC, BjöRklund M (2012) Population differentiation and restricted gene flow in Spanish crossbills: not isolation-by-distance but isolation-by-ecology. J Evol Biol 25:417–430

Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Mol Ecol 11:951–966

Excoffier L, Laval G, Schneider S (2005) An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Fady B, Cottrell J, Ackzell L, Alía R, Muys B, Prada A et al. (2016) Forests and global change: what can genetics contribute to the major forest management and policy challenges of the twenty-first century? Reg Environ Change 16:927–939

Fox J (2005) The R Commander : a basic statistics graphical user interface to R J Stat Softw 14:1–42

Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conserv Biol 10:1500–1508

Frankham R (1997) Do island populations have less genetic variation than mainland populations? Heredity 78(Pt 3):311–327

Fretzdorff S, Paterne M, Stoffers P, Ivanova E (2000) Explosive activity of the Reunion Island volcanoes through the past 260,000 years as recorded in deep-sea sediments. Bull Volcanol 62:266–277

Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies (B O'Meara, Ed.). Methods Ecol Evol 6:925–929

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983

Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the Coffea canephora cultivated in Vietnam and Mexico BMC Plant Biol 16:242

García-Verdugo C, Baldwin BG, Fay MF, Caujapé-Castells J (2014) Life history traits and patterns of diversification in oceanic archipelagos: a meta-analysis: plant traits and speciation on islands. Bot J Linn Soc 174:334–348

García-Verdugo C, Fay MF (2014) Ecology and evolution on oceanic islands: broadening the botanical perspective: ecology and evolution on oceanic islands. Bot J Linn Soc 174:271–275

García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J (2015) Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. Mol Ecol 24:726–741

Gillot P-Y, Nativel P (1982) K-Ar chronology of the ultimate activity of Piton des Neiges volcano, Réunion J Volcanol Geotherm Res 13:131–146

Giraud-Techer S, Amédée J, Girard-Valenciennes E, Thomas H, Brillant S, Grondin I et al. (2016) Plantes médicinales de La Réunion inscrites à la Pharmacopée française Ethnopharmacologia 56:7–33

Glemin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc B Biol Sci 273:3011–3019

Hamon P, Grover CE, Davis AP, Rakotomalala J-J, Raharimalala NE, Albert VA et al. (2017) Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species. Mol Phylogenet Evol 109:351–361

James JE, Lanfear R, Eyre-Walker A (2016) Molecular evolutionary consequences of island colonization. Genome Biol Evol 8:1876–1888

Jaros U, Fischer GA, Pailler T, Comes HP (2016) Spatial patterns of AFLP diversity in *Bulbophyllum* occultum (Orchidaceae) indicate long-term refugial isolation in Madagascar and long-distance colonization effects in La Réunion. Heredity 116:434–446

Kainulainen K, Razafimandimbison SG, Wikström N, Bremer B (2017) Island hopping, long-distance dispersal and species radiation in the Western Indian Ocean: historical biogeography of the Coffeeae alliance (Rubiaceae). J Biogeogr 44:1966–1979

Klein A-M, Steffan-Dewenter I, Tscharntke T (2003) Bee pollination and fruit set of *Coffea arabica* and *C. canephora* (Rubiaceae). Am J Bot 90:153–157

Lashermes P, Couturon E, Moreau N, Paillard M, Louarn J (1996) Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. Theor Appl Genet 93:458–462

Legendre P, Legendre L (1998) Canonical analysis. In: Numerical ecology, Developments in environmental modelling. ElsevierScience BV: Amsterdam, pp 575–635.

Leroy J-F (1963). Sur les caféiers sauvages des îles des Mascareignes. In: *Comptes rendus de l'Académie des Sciences*, Paris Vol. 256, pp 2897–2899.

Litrico I, Ronfort J, Verlaque R, Thompson JD (2005) Spatial structure of genetic variation and primary succession in the pioneer tree species *Antirhea borbonica* on La Reunion. Mol Ecol 14:1575–1584

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. Nature 457:830–836

Mahé L, Le Pierrès D, Combes M-C, Lashermes P (2007) Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. Genome 50:316–324

Mallet B, Martos F, Blambert L, Pailler T, Humeau L (2014) Evidence for isolation-by-habitat among populations of an epiphytic orchid species on a small oceanic island. PLoS ONE 9: e87469

Météo-France (2018) Climat de la Réunion. *Météo Fr La Réun.* Available at: http://www.meteofrance.re/climat/description-du-climat

Michon L, Saint-Ange F (2008) Morphology of Piton de la Fournaise basaltic shield volcano (La Réunion Island): characterization and implication in the volcano evolution J Geophys Res 113:B03203

Moore J, White WM, Paul D, Duncan RA, Abouchami W, Galer SJG (2011) Evolution of shield-building and rejuvenescent volcanism of Mauritius. J Volcanol Geotherm Res 207:47–66

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858

Ngo HT, Mojica AC, Packer L (2011) Coffee plant – pollinator interactions: a review. Can J Zool 89:647–660

Noirot M, Charrier A, Stoffelen P, Anthony F (2016) Reproductive isolation, gene flow and speciation in the former *Coffea* subgenus: a review. Trees 30:597–608

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375–402

Nowak MD, Davis AP, Anthony F, Yoder AD (2011) Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffea* (Rubiaceae). PLoS ONE 6:e21019

Nowak MD, Davis AP, Yoder AD (2012) Sequence data from new plastid and nuclear COSII regions resolves early diverging lineages in *Coffea* (Rubiaceae). Syst Bot 37:995–1005

Nowak MD, Haller BC, Yoder AD (2014) The founding of Mauritian endemic coffee trees by a synchronous long-distance dispersal event. J Evol Biol 27:1229–1239

Nyhagen DF, Turnbull SD, Olesen JM, Jones CG (2005) An investigation into the role of the Mauritian flying fox, *Pteropus niger*, in forest regeneration. Biol Conserv 122:491–497

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E and Wagner H (2017). *vegan: Community Ecology Package*. R package version 2.4-6. https://CRAN.R-project.org/package=vegan

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22:5983–5999

Perrier X, Jacquemoud-Collet J-P (2006) DARwin software. http://darwin.cirad.fr/darwin.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8: e1002967

Privé F, Aschard H, Blum M, Audrey Z (2018) Efficient management and analysis of large-scale genomewidedata with two R packages: bigstatsr and bigsnpr Bioinformatics 34:2781–2787

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461:489–494

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D et al. (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc 5:P54

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68:1–15

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Strasberg D (1996) Diversity, size composition and spatial aggregation among trees on a 1-ha rain forest plot at La Réunion. Biodivers Conserv 5:825–840

Strasberg D, Rouget M, Richardson DM, Baret S, Dupont J, Cowling RM (2005) An assessment of habitat diversity and transformation on la réunion island (mascarene islands, indian ocean) as a basis for identifying broad-scale conservation priorities. Biodivers Conserv 14:3015–3032

Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ (2014) Interpretation of patterns of genetic variation in endemic plant species of oceanic islands: genetic variation in island plants. Bot J Linn Soc 174:276–288

Techer MA, Clémencet J, Simiand C, Preeaduth S, Azali HA, Reynaud B et al. (2017) Large-scale mitochondrial DNA analysis of native honey bee Apis mellifera populations reveals a new African subgroup private to the South West Indian Ocean islands BMC Genet 18:53

Thébaud C, Strasberg D (1999) Plant dispersal in fragmented landscapes: a field study of woody colonization in rainforest remnants of the Mascarene archipelago. In: Laurence WF, Bierregaard RO (eds) Tropical Forest Remnants: Ecology, Management and Conservation of Fragmented Communities, Chicago, pp 312–332.

Van Der Plas GW, De Boer EJ, Hooghiemstra H, Vincent Florens FB, Baider C, Van Der Plicht J (2012) Mauritius since the last glacial: environmental and climatic reconstruction of the last 38 000 years from Kanaka Crater. J Quat Sci 27:159–168

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. Ecol Lett 16:175–182

Wright S (1943) Isolation by distance. Genetics 28:114–138