Universidad de Valladolid

Escuela de Doctorado Universidad de Valladolid

PROGRAMA DE DOCTORADO EN CONSERVACIÓN Y USO SOSTENIBLE DE SISTEMAS FORESTALES

DOCTORAL THESIS / TESIS DOCTORAL:

Genetic structure, phylogeography, adaptive variation and speciation in the tropical tree genus *Symphonia*

Estructura genética, filogeografía, variación adaptativa y especiación en árboles tropicales del género *Symphonia*

Presentada por Mª de la Paloma Torroba Balmori para optar al grado de Doctora por la Universidad de Valladolid

Dirigida por:

Doctora Myriam Heuertz

Doctora Sanna Olsson

Palencia, 2022

Gobierno de España · Ministerio de Ciencia e Innovación · CSIC · INIA Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria · iuFOR Instituto Universitario de Investigación Gestión Forestal Sostenible

–... Bien, el caso es que así fue como empezó todo. Verán, pensé que las cosas irían mucho mejor si creaba criaturas que pudieran alterar sus propias instrucciones cuando tuvieran necesidad de hacerlo, y entonces…

– Oh, se refiere a la evolución – dijo Ponder Stibbons.

– ¿Usted cree? – El dios puso expresión pensativa – Oh, sí. Cambiar con el paso del tiempo… Sí, ese término nos proporciona una forma muy acertada de describirlo, ¿verdad? Evolución. Sí, supongo que eso es lo que hago. Por desgracia, no parece estar funcionando correctamente.

El país del Fin del Mundo, Terry Pratchett.

Genetic structure, phylogeography, adaptive variation and speciation in the tropical tree genus *Symphonia*

Estructura genética, filogeografía, variación adaptativa y especiación en árboles tropicales del género *Symphonia*

PhD Student: Mª de la Paloma Torroba Balmori

Supervisors: Dr. Myriam Heuertz
UMR BIOGECO
INRAE, University of Bordeaux
Cestas (France)

Dr. Sanna Olsson
Department of Forest Ecology and Genetics
INIA Forest Research Centre
Madrid (Spain)

Supervisor of the International Scientific Visit:

Dr. Thomas L. Parchman
Department of Biology
University of Nevada
Reno (EEUU)

External Reviewers:

Dr. Dario I. Ojeda Alayon
Departmentof Forest Genetics and Biodiversity
Norwegian Institute of Bioeconomy Research (NIBIO)
Norway

Dr. Armel Donkpegan
SYSAAF – Centre INRAE Val de Loire
INRAE
France

Doctorate programme / Programa de doctorado:

Conservación y Uso Sostenible de Sistemas Forestales
Escuela Técnica Superior de Ingenierías Agrarias
Universidad de Valladolid, Palencia (Spain)

Place of Publication: Palencia, Spain

Year of Publication: 2022

Online Publication: http://biblioteca.uva.es/export/sites/biblioteca/

Agradecimientos

Como la tesis es un trabajo muy personal, pero que a la vez tiene influencia de mucha gente, es necesario destacar las siguientes aportaciones:

Agradezco la dirección de mi tesis a Myriam Heuertz, Sanna Olsson, Ricardo Alía y Santiago González Martínez. Gracias a ellos me embarqué en un proceso de desafío intelectual y personal que me ha formado a muchos niveles.

Y agradezco a Dario I. Ojeda Alayon y a Armel Donkpegan su interés y amabilidad en la revisión de la tesis.

Y puesto que para este proceso ha sido muy importante una buena base, agradezco a Pilar Zaldivar y Carolina Martínez Ruiz todo lo que me enseñaron durante la carrera, a Carmen García Barriga y a Zaida Lorenzo las bases que me dieron para empezar la tesis, y a mis compañeros de doctorado y otros investigadores por lo que me transmitieron.

Agradezco a Tom Parchman su acogida en Reno, en la que aprendí conceptos importantes de programación y secuenciación y algo sobre los grandes desiertos.

Agradezco a mi familia su apoyo, muy necesario, y quiero destacar el momento en que mis familiares de Burdeos me dejaron invadir su casa dos meses sin apenas conocerme.

A mis amigos les agradezco su existencia, especialmente a los que siguieron siendo amigos aún cuando les hablaba de mi tesis.

A mis compañeros de trabajo les quiero agradecer su apoyo, y su infinita paciencia con el recurrente tema de mi tesis.

También fue muy importante la gente con la que afortunadamente me crucé en ámbitos académicos o "civiles", a veces solo durante un rato, y que me transmitieron buenas ideas sobre ciencia o sobre la vida en general que me ayudaron a avanzar con la tesis.

Como bonus, quisiera agradecer a Terry Pratchett y a Cher su efecto positivo en mi estado de ánimo.

Y quiero dedicar esta tesis a Sofía, la nueva incorporación a la familia.

También hay que destacar que la tesis no se habría podido llevar a cabo sin los siguientes organismos y convocatorias, a los cuales agradezco su apoyo:

Ayudas para becas y contratos de Formación de Profesorado Universitario del Programa Nacional de Formación de Recursos Humanos de Investigación del ejercicio 2012, del Ministerio de Educación, Cultura y Deporte, para la realización de la tesis doctoral.

Ayudas para la formación de profesorado universitario, de los subprogramas de Formación y de Movilidad incluidos en el Programa Estatal de Promoción del Talento y su Empleabilidad, en el marco del Plan Estatal de Investigación Científica y Técnica y de

Innovación 2013-2016, del Ministerio de Educación, Cultura y Deporte, para una estancia científica en la Universidad de Nevada, Reno, EEUU.

Proyecto CGL2012-40129-C02-02/AFFLORA: DEMOGRAPHIC HISTORY AND ADAPTATION IN TROPICAL TREES (2013-2015) financiado por el Ministerio de Economía y Competitividad en la Convocatoria 2012 de Proyectos de Investigación Fundamental No Orientada

Programa COST Action FP1202, en la modalidad Short Term Scientific Missions, para la financiación de una estancia científica en la Universidad de Friburgo, Suiza.

Ayuda en el marco del programa Trees4Future, un proyecto co-financiado por el séptimo programa marco de la Unión Europea FP7, para el acceso a las instalaciones de genómica y transcriptoma del INRA en Burdeos y la financiación para el desarrollo de los marcadores funcionales basados en SNPs del género *Symphonia*.

Agradezco también al Centro de Supercomputación y Bioinnovación (SCBI) de la Universidad de Málaga (www.scbi.uma.es/site) por el acceso a sus recursos computacionales y el apoyo técnico, indispensables para los análisis del Estudio II. Al laboratorio de Dr. João Loureiro, en el Centro para la Ecología Funcional, del Departamento de Ciencias de la Vida de la Universidad de Coimbra, por su ayuda imprescindible en los análisis de citometría de flujo del Estudio III. Y al Departamento de Biología Molecular y Bioquímica de la Universidad de Málaga, por su ayuda fundamental en el ensamblaje de las secuencias de transcriptoma de *Symphonia* en el Estudio III.

INDEX

List of Tables

classes; 1st DC, maximum distance of the first class (m); $F_{ij(1),}$ mean kinship coefficient of the first distance class; *Sp*, intensity of FSGS and *P*-value of one-sided test of the regression slope *b* of F_{ij} on the logarithm of spatial distance; *b* (SE), jackknife mean of *b* and its standard error; *eig.sPCA*: eigenvalue of the first sPCA axis and significance of G-test. ns, not significant. ***, $P \leq 0.001$; **, $P \leq 0.01$; nc, not calculated (no coordinates available).

Table 11. Strength of genetic differentiation between nuclear gene pools (GPs) within *Symphonia globulifera* populations. *K*, number of STRUCTURE clusters; $F_{ST(Q \geq 0.5)}$, F_{ST} among GPs with individual assignment based on $Q \geq 0.5$; $F_{ST(Q \geq 0.875)}$, F_{ST} among GPs with $Q \geq 0.875$; PI50 (%), proportion of individuals assigned to a GP based on $Q \geq 0.5$; PI87 (%), proportion of individuals assigned to a GP based on $Q \geq 0.875$. nd, not defined; ***, $P \leq 0.001$. **70**

Table 12. Spatial genetic heterogeneity in SSR data and its association with plastid DNA haplotypes (i.e., cytonuclear disequilibria) and altitude. The mean sPCA score for the first sPCA axis is given for individuals carrying the same haplotype, sPCA (hap), or belonging to the same *ad hoc* altitudinal class sPCA (alt); n, sample size range per altitudinal class. *P* values represent the significance of one-way ANOVA analyses testing differences in the mean sPCA score for haplotypes, *P*(hap) or altitudinal classes, *P*(alt). nc[1], not computed because coordinates were unavailable or populations were monomorphic; nc[2], not computed because SSR and plastid DNA data were collected from different individuals; ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$. **71**

Table 13. Genetic diversity estimates in sampling locations or gene pools of *Symphonia globulifera* and the alternative morphotype in Africa and the Neotropics, based on 4921- SNPs genotyped with high certainty. GP or sampling location (ID), gene pools from Entropy considered in the analysis, or sampling location in case of strong admixture, and their ID; **All individuals**: N, number of individuals assigned to GP or location; $MD_{m\text{-}M}$ (ind), minimum and maximum levels of missing data per individual within GP or location (%) and individuals with more than 20% of missing alleles within GP or location (%); NPS, number of polymorphic sites; **Sampled subset**: n, sample size for gene diversity estimates; nps, number of polymorphic sites; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS} *(95%IC)*, fixation index and 95% confidence interval from bootstrapping. **73**

Table 14. Estimates of genetic distance based on Nei's D among gene pools of *Symphonia globulifera* and the alternative morphotype in Africa and the Neotropics. **74**

Table 15. Loci identified as outliers (q-value < 0.05) in BayeScEnv analysis for the five continental locations in Africa. For each locus and covariate, the table displays q-value on the *g* parameter in the model. SNP ID: names identifying the loci on the original 4921 SNPs dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. Covariates: uncorrelated environmental variables. **79**

(min – max); extreme Ne: highest and lowest values from 95% confidence intervals obtained from simulations in the studies (min – max).

Table S9.3.1.1. Climatic and soil variables for analysis of loci under selection in continental populations in Africa. Type: type of data for each variable (continuous, categorical). **179**

Table S9.3.2.1. Loci identified as outliers (q-value < 0.01) in BayeScan analysis for the five continental locations in Africa. SNP ID: names identifying the loci on the original 4921 SNP dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. F_{ST}: the F_{ST} coefficient averaged over populations. *For the model including selection*: Prob: posterior probability. log10 (PO): the logarithm of Posterior Odds to base 10. q-value: a false-discovery rate analogue to the p-value. α: the estimated alpha coefficient representing the direction and strength of selection. **181**

Table S9.3.2.2. Loci identified as outliers using the core model and the XtX statistics in BayPass for the five continental locations in Africa (99% quantile of XtX values from the simulated pseudo-observed dataset: 9.671). SNP ID: names identifying the loci on the original 4921 SNP dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. **182**

Table S9.3.2.3. Matrices used for the Mantel tests to detect if there are coincidence of extreme values of environmental variables and of extreme neutral allele frequencies in the same locations: the correlation matrix derived from **Ω** from BayPass and matrices for pairwise environmental distances for BIO17 and PH_T. Colours represent the magnitude of values within each matrix from higher (green) to lower (red) values. **185**

Table S9.4.1.1. Species, locations, accession numbers (ID) and assignment to SSR gene pools in Malagasy *Symphonia* (rGP) for *Symphonia* samples used for transcriptome sequencing and candidate SNP identification. *Spp*.: Putative species identified based on the plant specimens (we could not collect branches with leaves for all samples, neither all plant specimens collected could be identified). **187**

Table S9.4.3.1. Estimates of genetic distance based on Nei's D among gene pools of Malagasy *Symphonia* and populations of *S. globulifera* (including the alternative morphotype) in Africa and the Neotropics: A) based on 53 diploid snps, B) based on 124 diploid SNPs (nGP1 not included). **189**

Table S9.4.4.1. Details on the 20 polymorphic SSR markers used for Malagasy *Symphonia* SSR genotyping, including primer sequences and GenBank accession numbers. Repeat: Number of repeats found in the sequence that corresponds to the accession number. Fluorochrome in 50 was 6-FAM for Q2 (TAGGAGTGCAGCAAGCAT), VIC for Q3 (CACTGCTTAGAGCGATGC) and NED for Q4 (CTAGTTATTGCTCAGCGGT). GenBank Acc.: The accession number in GenBank. **192**

List of Figures

based on model log likelihood (L(K)) and Delta K (Δ*K)*. In TESS the deviance information criterion (DIC) and the model log likelihood indicated increasing support for increasing *K* values, and the Δ*K* criterion was undefined for *K*=2, because TESS does not run with *K*=1.

Structure of the Thesis

In this thesis we present three different approaches to thoroughly investigate the genetic structure of the genus *Symphonia*, a genus of tropical trees with a wide distribution in Africa, Madagascar and the Neotropics, and the mechanisms involved. The document is written in English and includes an Abstract, an Introduction, the Objectives of the thesis, a Materials and methods section describing study sites, laboratory methods to obtain the genetic data and the analyses performed, a Results section, a General Discussion and Conclusions section. The abstract and the conclusions are written in English and Spanish. The supplementary information with additional figures, tables and datasets of molecular markers, is located at the end of this document.

Abstract

The genetic structure within a species is the result of the levels of the genetic diversity and its spatial distribution. Also, it depends significantly on the specific evolutionary history experienced by the species. Thus, to disentangle the overlapping evolutionary processes acting at different levels in a species or a taxon, it will be necessary to work at different spatial scales and at different taxonomic levels as complementary approaches.

The study of the fine-scale spatial genetic structure in plants (the micro scale approach) will imply to work at the shortest spatial scales and to capture detailed information on the spatial distribution of genotypes at within-population scale. The analysis at this scale will help to detect mainly evolutionary and ecological processes more related to short-term periods of time and/or smaller spatial scales such as habitat fragmentation and other disturbances, efficiency of dispersal mechanisms or gene dispersal distances. On the other side, the study of the genetic structure at wider scales (the macro scale approach), including both geographical and taxonomic (i.e., speciation) points of view, will usually imply to detect larger spatio-temporal processes and to work with deeper evolutionary timescales. In this sense, the spatial genetic structure within a species at this macro scale will be the result of different historical and contemporary influences such as connectivity across the range of the species and landscape barriers, environmental adaptation, demographic history or climatic events, among others. Finally, if we include the taxonomic perspective in the analysis of genetic structure in a group of closely related species, we will be able to analyse the processes leading to speciation, which also may involve those previously mentioned.

Tree species are good study systems to analyse their genetic structure and underlying evolutionary processes at different levels. Trees are sessile life-forms that often present large populations connected by gene flow, frequently occurring in different environments, with high levels of usually undomesticated genetic diversity compared to other plant forms: all of these features boost the efficiency of selection leading to local adaptation, facilitate traceability of gene movements and set an appropriate scenario to disentangle selective from stochastic evolutionary forces. However, working with tropical tree species includes several challenges, such as logistic challenges, difficult field conditions, a variety of complications related to morphological identification, and other problems which have slowed down the pace in evolutionary research in tropical tree species. Nonetheless, work with tropical trees is fundamental. Many different questions related to the raise and maintenance of the high floristic diversity in tropical rainforests, and the specific evolutionary processes which contribute to the generation of such hyper-rich ecosystems, have been raised. Moreover, increasing the knowledge is of utmost importance to perform an effective sustainable management and conservation of this ecosystem, especially regarding the ongoing effects of global climate change and the high rates of global deforestation.

This thesis is focused on the tropical tree genus *Symphonia*, which belongs to the Clusiaceae family, as a study group to analyse its genetic structure at different spatial

scales and at different taxonomic levels. The main objective was to unveil the evolutionary dynamics at micro and macro scales that have led the genus *Symphonia* to its current situation in terms of ecology, geographical distribution, local adaptation and genetic diversity. This species-rich genus, with its widespread distribution in many habitats across continents and its high diversity in Madagascar, is a particularly suitable model to investigate the evolutionary processes related to spatial and taxonomic genetic structure that produce the patterns of genetic diversity in tropical trees.

Specifically, we analysed the fine-scale genetic structure of *S. globulifera* in several African and Neotropical populations in relation to the different genetic dispersal strategies and examined the relationships among similarities in the strength and patterns of fine-scale genetic structure in groups of populations and their concurring drivers, such as similar disperser communities, habitat features or biogeographic history. We also investigated the colonization and diversification history of African and Neotropical populations of *S. globulifera,* considering the influence of past climatic events as well as of other environmental and geographical factors at those large geographic scales. Also, we addressed the genomic signatures of local adaptation of *S. globulifera* in Africa, identifying some of the drivers involved in this regional adaptation. Finally, we developed new functional single nucleotide polymorphism (SNP) markers to infer phylogenetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species) for the first time, to test the current species delimitation in Malagasy *Symphonia* using integrative taxonomy and to investigate the evolutionary history underlying the radiation of *Symphonia* species in Madagascar.

Throughout the different approaches involved in this thesis, we found that the genus *Symphonia* presents an outstanding amount of genetic structure at different levels, coupled with a wide distribution range covering a variety of habitats. On one hand, we detected a wide diversity of FSGS patterns within *S. globulifera* populations, from non-significant or weak FSGS in Neotropical populations to pronounced structure in African ones. The strength of FSGS correlated with both disperser communities and altitudinal sampling range, while our data also contained evidence for co-occurrence of differentiated lineages and gene pool aggregation following habitat features.

On the other hand, we revealed that the most probable origin of colonization events from Africa to America was performed through São Tomé Island, based on the African populations analysed, reinforcing the evidence that *S. globulifera* is able to perform marine dispersal. We also evidenced that both morphotypes in French Guiana and the two nearby populations of the same common morphotype in French Guiana were evolutionarily distinct lineages. Moreover, our work also presents the first insight into the basis of local adaptation of *S. globulifera* at large scale, where the ability of our species to cope with water stress and acidic soils seems to underlie its extensive African distribution range.

Finally, our study presents phylogenetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species) and congruent results about species delimitation on Malagasy *Symphonia* based on genetic and nongenetic sources of

data and on an unprecedented sampling effort in three major regions in Madagascar, both analyses performed for the first time on this group of species.

Overall, this thesis evidences that analysing the influence of the same factors at different spatial and taxonomic scales can give complementary insights and a more comprehensive view of the evolutionary processes affecting a taxon, as well as helping to explain its contemporary evolutionary situation.

Resumen

La estructura genética que presenta una especie es el resultado de sus niveles de diversidad genética, así como de su distribución espacial. Además, depende significativamente de la historia evolutiva específica que ha sufrido la especie. Así, para desentrañar los procesos evolutivos que actúan simultáneamente a diferentes niveles en una especie o taxon, será necesario trabajar mediante enfoques complementarios, orientados a diferentes escalas espaciales y a diferentes niveles taxonómicos.

El estudio de la estructura genética espacial a pequeña escala en plantas (un enfoque en escala micro) implicará trabajar en las escalas espaciales más pequeñas, así como recoger información detallada de la distribución espacial de genotipos dentro de las poblaciones. El análisis a esta escala ayudará a detectar principalmente procesos evolutivos y ecológicos relacionados principalmente con periodos de tiempo cortos y/o a escala espacial pequeña, tales como fragmentación de hábitats y otras perturbaciones, eficiencia de los mecanismos de dispersión o distancias de dispersión genética. Por otro lado, el estudio de la estructura genética a escala más amplia (un enfoque en escala macro), incluyendo los puntos de vista geográfico y taxonómico (es decir, de especiación), normalmente implicará detectar procesos espacio-temporales más amplios y trabajar con escalas evolutivas de tiempo más largas. En este sentido, la estructura genética espacial de una especie a escala macro será el resultado de diferentes influencias históricas y contemporáneas, tales como la conectividad a lo largo de la distribución de la especie, las barreras del paisaje, la adaptación al medio, la historia demográfica o los eventos climáticos, entre otros. Finalmente, si incluímos la perspectiva taxonómica en el análisis de la estructura genética de un grupo de especies muy relacionadas, podremos ser capaces de analizar los procesos que conducen a la especiación, entre los cuales se pueden encontrar también los ya previamente mencionados.

Las especies arbóreas son buenos sistemas de estudio para analizar su estructura genética y los procesos evolutivos subyacentes a diferentes niveles. Los árboles son formas de vida sésiles que, a menudo, presentan grandes poblaciones conectadas mediante flujos genéticos. Además, aparecen frecuentemente en diferentes hábitats y suelen acumular altos niveles de diversidad genética no domesticada en comparación con otras formas de vida en plantas. Todas estas características impulsan la eficiencia de los procesos de selección que conducen a la adaptación local, facilitan la trazabilidad de los movimientos de los genes y establecen un escenario apropiado para desenmarañar las fuerzas evolutivas selectivas de aquellas que son estocásticas. Sin embargo, trabajar con especies arbóreas tropicales implica diferentes desafíos, como dificultades logísticas y de muestreo de campo, dificultades relacionadas con la identificación morfológica cierta, así como otros problemas que, en conjunto, han ralentizado el ritmo de la investigación evolutiva en especies arbóreas tropicales. Sin embargo, trabajar con árboles tropicales es fundamental y hay planteadas muchas cuestiones relacionadas con el establecimiento y el mantenimiento de la alta diversidad florística en los bosques tropicales, así como sobre los procesos evolutivos específicos que han contribuido a la generación de estos ecosistemas hiper-ricos en especies. Además, aumentar el conocimiento en este bioma es

de gran importancia para su gestión forestal sostenible y su conservación, especialmente teniendo en cuenta los efectos ya perceptibles del cambio climático y las altas tasas de deforestación global.

Esta tesis está centrada en el género de árboles tropicales *Symphonia*, que pertenece a la familia Clusiaceae, como grupo de estudio para analizar su estructura genética a diferentes escalas espaciales y a diferentes niveles taxonómicos. El objetivo principal ha sido descubrir las dinámicas evolutivas a escalas micro y macro que han conducido al género *Symphonia* a su situación actual en cuanto a ecología, distribución geográfica, adaptación local y diversidad genética. Este género rico en especies, con su amplia distribución en numerosos hábitats a lo largo de dos continenetes y su alta diversidad en Madagascar, es un modelo particularmente adecuado para investigar los procesos evolutivos relacionados con la estructura genética espacial y taxonómica que producen los patrones de diversidad genética en árboles tropicales.

Específicamente, hemos analizado la estructura genética de *S. globulifera* a pequeña escala en varias poblaciones de África y de la región Neotropical en relación con las diferentes estrategias de dispersión genética. Además, hemos examinado las relaciones entre las similitudes en los patrones y en la intensidad de la estructura genética a escala pequeña en grupos de poblaciones respecto de los factores concurrentes que influyen en dicha estructura genética, tales como comunidades similares de dispersantes, características de los hábitats o la historia biogeográfica. También hemos investigado la colonización y la historia de diversificación de las poblaciones africanas y neotropicales, teniendo en cuenta la influencia de eventos climáticos del pasado, así como otros factores ambientales y geográficos a gran escala. Asimismo, hemos estudiado las huellas genómicas de la adaptación local de *S. globulifera* en África, identificando algunos de los factores involucrados en esta adaptación a nivel regional. Finalmente, hemos desarrollado nuevos marcadores funcionales basados en polimorfismos de un solo nucleótido (SNP) con el objetivo de inferir por primera vez las relaciones filogenéticas dentro del género *Symphonia* (que incluye *S. globulifera* y las especies de *Symphonia* de Madagascar), comprobar la actual delimitación de especies de *Symphonia* de Madagascar mediante taxonomía integrativa, e investigar la historia evolutiva que subyace a la radiación de las especies de *Symphonia* en Madagascar.

A lo largo de los distintos enfoques que componen la tesis, hemos encontrado que el género *Symphonia* presenta una acumulación muy llamativa de estructuras genéticas a diferentes niveles, junto con un amplio rango de distribución que cubre una gran variedad de hábitats. Por un lado, hemos detectado una amplia diversidad de patrones de estructura genética a pequeña escala dentro de las poblaciones de *S. globulifera*, desde estructuras genéticas débiles o no significativas en poblaciones neotropicales a estructuras pronunciadas en las poblaciones africanas. La intensidad de la estructura genética a pequeña escala correlacionó con las comunidades de dispersantes y el rango espacial de muestreo en su dimensión altitudinal. Además, los datos también mostraron señales de ocurrencia simultánea de diferentes linajes, así como de agrupación de grupos genéticos en relación con las características del hábitat.

Por otro lado, hemos desvelado que el origen más probable de los eventos de colonización de África a América ha sido la isla de São Tomé, en base a las poblaciones africanas analizadas, lo que refuerza la evidencia de que *S. globulifera* es capaz de realizar dispersiones a través del mar. También hemos encontrado señales de que los dos morfotipos de la Guyana Francesa, así como dos poblaciones cercanas del morfotipo común en la Guyana Francesa, son linajes evolutivos distintos. Además, nuestro trabajo también presenta una primera aproximación a las bases de la adaptación local de *S. globulifera* a gran escala, donde la habilidad de nuestra especie en resistir el estrés hídrico y los suelos ácidos parece sustentar su amplio rango de distribución en África.

Finalmente, nuestro estudio presenta las relaciones filogenéticas dentro del género *Symphonia* (incluyendo *S. globulifera* y las especies de *Symphonia* de Madagascar), y resultados congruentes en cuanto a la delimitación de especies en Madagascar en base a fuentes de datos genéticos y no genéticos y a un esfuerzo de muestreo sin precedentes en tres grandes regiones de Madagascar, ambos análisis realizados por primera vez en este grupo de especies.

En conjunto, esta tesis demuestra que analizar la influencia de los mismos factores a diferentes escalas espaciales y taxonómicas puede dar visiones complementarias y más completas respecto a los procesos evolutivos que afecta a un taxon, así como ayudar a explicar su situación evolutiva actual.

1. Introduction

The genetic structure, as an output of the levels of the genetic diversity and its spatial distribution within a species, depends significantly on the specific evolutionary history experienced by the species (Ahuja & Mohan Jain, 2017). Thus, to disentangle the overlapping evolutionary processes acting at different levels in a species or a taxon, it will be necessary to work at different scales.

The study of the fine-scale spatial genetic structure in plants (the micro scale approach) implies to work at the shortest spatial scales to properly detect it, as well as to capture detailed information on the spatial distribution of genotypes at within-population scale (Vekemans & Hardy, 2004). Considering also that it evolves on a few generations (see e.g., Troupin, Nathan, & Vendramin, 2006), its analysis helps to detect evolutionary and ecological processes more related to short-term periods of time and/or smaller spatial scales (i.e., microevolutionary and microgeographic scales; Wang, 2010; Scotti, González-Martínez, Budde, & Lalagüe, 2015) such as habitat fragmentation and other disturbances, cohort dynamics, efficiency of dispersal mechanisms or gene dispersal distances (Marquardt & Epperson, 2004; Hardy et al., 2006; Troupin, Nathan, & Vendramin, 2006; Collevatti et al., 2014; Duminil et al., 2016). Nonetheless, historical, long-term, or wide-scale processes might also be detected (e.g., see Kalisz, Nason, Hanzawa, & Tonsor, 2001; Audigeos, Brousseau, Traissac, Scotti-Saintagne, & Scotti, 2013). Despite this small scale of work, the comparison of local situations in different locations may also help to get general conclusions valid across wider scales for the species (see Torroba-Balmori et al., 2017).

Conversely, the study of the genetic structure at wider scales (the macro scale approach), including both geographical and taxonomic (i.e., speciation) points of view, will usually imply to detect larger spatio-temporal processes and to work with deeper evolutionary timescales (Wang, 2010; Marske, Rahbek, & Nogués-Bravo, 2013). In this sense, the spatial genetic structure within a species at this macro scale will be the result of different historical and contemporary influences such as connectivity across the range of the species and landscape barriers, environmental adaptation, demographic history, climatic events, or even geomorphological events, among others (Hewitt, 2000; Lee & Mitchell-Olds, 2011; Mairal et al., 2017; Nistelberger, Tapper, Coates, McArthur, & Byrne, 2021). Further, if we include the taxonomic perspective in the analysis of genetic structure in a group of closely related species, we will be able to analyse the processes leading to speciation, which also may involve those previously mentioned (Hart, 2011; Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; see e.g., Zhou et al., 2012; Ren, Mateo, Guisan, Conti, & Salamin, 2018; Zhao, Gugger, Xia, & Li, 2016).

Overall, the analyses of the genetic structure within a species (or a group of closely related species), at different spatial scales, and at different taxonomic levels are complementary approaches. Together, they provide an opportunity to get a comprehensive picture of the relevant evolutionary and ecological processes that have led a taxon to its current geographical structure of genetic variation. At the same time and in return, since such

studies can inform on mechanisms underlying demographic processes and spatial genetic heterogeneity within and among populations, they may provide guidance for sustainable forest management and conservation practises based on the genetic structure of populations (Plomion et al., 2016; Ahuja & Mohan Jain, 2017; e.g., Degen et al., 2006; Hansen et al., 2015; Duminil et al., 2016).

1.1. Fine-scale spatial genetic structure and its drivers

Fine-scale spatial genetic structure in plants (FSGS), the non-random spatial distribution of genotypes within populations, is shaped by microevolutionary processes such as dispersal, local genetic drift and selection (Vekemans & Hardy, 2004). One of the most commonly evaluated patterns in FSGS studies in plants is isolation by distance (IBD, Wright, 1943; Meirmans, 2012). The IBD model predicts that, at drift-dispersal equilibrium, genetic differentiation among individuals is an increasing function of geographic distance due to spatially limited isotropic gene dispersal and local genetic drift (Wright, 1943; Malécot, 1951; Rousset, 2000; Vekemans & Hardy, 2004). A linear relationship is predicted with distance for 1-dimensional populations or with the logarithm of distance for 2-dimensional populations (Wright, 1943; Malécot, 1951; Rousset, 2000; Vekemans & Hardy, 2004). As pollen and seed dispersal are usually spatially restricted, the strength of FSGS under IBD assumptions can provide information on the historical gene dispersal distance in the population (Rousset, 2000; Vekemans & Hardy, 2004). While useful as the basic expected pattern (null model), the IBD model does not consider other features or processes that can constrain gene flow or generate spatial heterogeneity or discontinuities in allele frequencies.

The strength of FSGS depends primarily on an organism's life history traits, of which life form and breeding system are the most relevant in plants. Indeed, stronger FSGS is found in herbaceous plants than in trees, as well as in partially or completely selfing than in outcrossing species (Vekemans & Hardy, 2004). Population density has also a very important effect, with stronger FSGS found in low-density populations (Vekemans & Hardy, 2004; Sagnard, Oddou-Muratorio, Pichot, Vendramin, & Fady, 2011). Further, dispersal vectors matter as they determine the scale and spatial pattern of dispersal. In tropical trees for example, it has been shown that animal-pollination typically results in stronger FSGS than wind-pollination, and that gravity- or rodent-mediated seed dispersal generates stronger FSGS than dispersal by birds or larger animals (Hardy et al., 2006; Dick, Hardy, Jones, & Petit, 2008).

Beyond IBD expectations, intrinsic and extrinsic factors, often associated with landscape features, determine heterogeneity in FSGS patterns. Topographic features or complex relief can directly hinder genetic connectivity and thereby lead to anisotropic and/or heterogeneous FSGS. This occurs for instance if steep slopes or mountain ridges restrict gene flow, which contributes to genetic differentiation even in species with wide-ranging gene dispersal (Robledo-Arnuncio, Collada, Alía, & Gil, 2005; Rhodes, Fant, & Skogen, 2014). Habitat features can also influence the behaviour of seed and pollen dispersers,

affecting the genetic structure of the plants they disperse (Dick, Etchelecu, & Austerlitz, 2003; Cordeiro, Ndangalasi, McEntee, & Howe, 2009; Dyer, Chan, Gardiakos, & Meadows, 2012; Côrtes & Uriarte, 2013). For example, Jordano, García, Godoy, & García-Castaño, (2007) showed that small birds tended to disperse *Prunus mahaleb* seeds into covered microhabitats but medium-sized birds and small mammals preferentially deposited seed into open habitats. Genetic heterogeneity can also result from historical processes related to range dynamics, such as secondary contact of previously differentiated gene pools (Mitton, Kreiser, & Latta, 2000; Born et al., 2008). Finally, another factor shaping FSGS is habitat-mediated selection, which can generate adaptive differentiation, i.e., isolation by adaptation also known as isolation by environment (Andrew, Ostevik, Ebert, & Rieseberg, 2012; Shafer & Wolf, 2013). Although this process first affects only loci under selection, at later stages it can lead to genome-wide differentiation due to genetic hitchhiking (Feder, Egan, & Nosil, 2012).

It is challenging to determine whether a given FSGS pattern reflects spatial autocorrelation due to IBD alone, or whether it contains an additional spatial genetic heterogeneity (SGH, i.e., allele frequency discontinuities or locally co-occurring differentiated gene pools [GPs]) signal due to historical or contemporary processes. This is because spatial autocorrelation (the expected result of IBD) affects the analysis of spatial genetic discontinuities, and *vice versa* (Meirmans, 2012). Bayesian clustering methods employed to detect SGH can fail to detect genetic clines when genetic structure is weak, but they can also overestimate the number of genetic clusters due to the influence of IBD (Frantz, Cellina, Krier, Schley, & Burke, 2009; François & Durand, 2010). Incorporating spatial information into clustering methods can improve their results (Chen, Durand, Forbes, & François, 2007; François & Durand, 2010). Conversely, methods that quantify FSGS based on the IBD model at drift-dispersal equilibrium cannot independently assess the effect of IBD when genetic discontinuities are present (Vekemans & Hardy, 2004). Moreover, different combinations of historical and contemporary processes can produce similar FSGS patterns, further complicating the inference (Yuan, Sun, Comes, Fu, & Qiu, 2014). Aware of these issues, some authors have used sequential approaches of genetic cluster detection and IBD assessment (or *vice versa*) to infer population genetic processes (Born et al., 2008; Piotti et al., 2013), or relied on non-parametric methods to assess heterogeneous spatial genetic patterns (Jombart, Devillard, Dufour, & Pontier, 2008; Petkova, Novembre, & Stephens, 2015). A complementary approach to disentangle the factors contributing to FSGS is to compare the FSGS at biparentally inherited nuclear markers to FSGS at maternally inherited markers. Maternal markers will inform about gene flow due to seed dispersal (e.g., Ndiade-Bourobou et al., 2010; Budde, González-Martínez, Hardy, & Heuertz, 2013) and, due to their lower mutation rates, they can identify signatures of processes at deeper temporal scales than nuclear markers (Wang, 2011).

1.2. Large-scale genetic structure, demographic history and adaptive evolution

Examining the large-scale genetic structure of a species in a spatial context belongs to the field of phylogeography. This discipline studies genetic variation within species or among closely related ones in relation to their geographical context, considering both historical and contemporary factors impacting the species, and it is a bridge connecting microevolutionary (i.e., within populations) and macroevolutionary patterns and processes (i.e., among species or higher taxa) because it focuses on genealogy and historical population demography at the same time (Avise, 2009; Knowles, 2009b).

As easily deduced, the questions addressed by phylogeography encompass all processes generating the patterns of spatial genetic diversity within and among related taxa, including what processes have led a species to its contemporary geographical range and observed genetic structure (see Marske, Rahbek, & Nogués-Bravo, 2013; Marske, 2016; for a review). Regarding this specific question, many studies have focused on understanding species responses to past climate events with a special interest on the effect of Pleistocene climatic history, which is critical to understand their contemporary distributions (Davis & Shaw, 2001; Marske, Rahbek, & Nogués-Bravo, 2013; Marske, 2016; e.g., Hardy et al., 2013; Dauby et al., 2014; Piñeiro, Dauby, Kaymak, & Hardy, 2017; Helmstetter, Béthune, Kamdem, Sonké, & Couvreur, 2020). Pleistocene climate oscillations had significant impacts on the distributions of many species. The range expansions and contractions undergone by species, associated to changes in environmental conditions caused by glacial (cool/dry) and interglacial (warm/humid) periods of the Pleistocene, left genetic signatures that are still detectable (Hewitt, 2000; Petit, Feng, & Dick, 2008). That information also makes it possible to compare Pleistocene range shifts in different species, which may reveal if they responded similarly to climate change events at the regional scale and/or during the postglacial colonization (e.g., refugia and dispersal routes; see Petit et al., 2002; Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009; Hardy et al., 2013). All in all, the understanding of how species responded to past climate changes, for which phylogeography has a major role, provides information to predict the response of species to contemporary climate change and allows to guide the current conservation and management policies (Petit, Feng, & Dick, 2008; Fordham, Brook, Moritz, & Nogués-Bravo, 2014; Gavin et al., 2014).

To infer the demographic history of species or populations, that is, to be able to describe events of population split and mixture and other related demographic parameters related, the use of multiple nuclear genetic markers has increased in the recent years because they provide more accurate results than organellar markers regarding past population history and phylogenetic relationships among taxa and are less affected by stochastic events that may arise from a single genealogy (Nielsen & Beaumont, 2009; Hickerson et al., 2010; Plomion et al., 2016). Classically, organelle genomes (i.e., one single locus) were used in phylogeographic assessments in animals and plants because of their potential to reflect genealogical histories due to their suitable sequence variation, mutation levels and inheritance modes (Avise, 2009). However, those inferences involved uncertainty

because they only reflected one possible genealogy per individual among multiple random possibilities for the same demographic processes in the populations (Nielsen & Beaumont, 2009). Nowadays, coalescent-based methods are one of the most common statistical approaches used in phylogeography, which allows to analyse the species trees (i.e., the species divergence history) in contrast to the gene tree (i.e., gene genealogy) when multi-locus datasets are available (Beichman, Huerta-Sanchez, & Lohmueller, 2018). Coalescence modelling consists of a probabilistic modelling approach that, working with a large number of possible genealogies of the loci considered under a specific demographic hypothesis, analyses how genes from the present merge into their common ancestor towards the past and tries to explain how the neutral genetic variation observed was generated (Knowles, 2009b; Hickerson et al., 2010). Thus, coalescent approaches are based on models that describe the probability of a particular demographic model of generating the neutral genetic variation observed, estimating different parameters such as effective population sizes, migration rates or divergence times (Nielsen & Beaumont, 2009; Knowles, 2009a). While independent loci can show discordant diversification due to stochastic processes, coalescence methods that estimate species trees aim to reveal the history of species or populations integrating information across multiple loci while accounting for those incongruences between gene trees (Knowles, 2009a). Therefore, these analyses allow to use the whole potential associated with high-throughput sequencing technologies, which can efficiently generate multilocus datasets of thousands of loci throughout the genome, including for non-model species (McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013). Since there is a wide variety of methods to perform demographic inference, each of them with different strengths and limitations and susceptible to be affected by demographic processes not included in the model, it is advised to compare the results from different methods (review in Beichman, Huerta-Sanchez, & Lohmueller, 2018).

Besides analysing genetic structure and its underlying past demographic drivers, we can also investigate the genomic signatures of natural selection, that is, local adaptation. This is the field of ecological genomics (or adaptation genomics or environmental genetics), which analyses the genetic basis of local adaptation to understand how species adapt and respond to different environmental conditions across their distribution ranges (Savolainen, Lascoux, & Merilä, 2013). Local adaptation is defined as the process of divergent natural selection that promotes the development of advantageous traits in populations with respect to their local environmental conditions, and the consequently emerged adaptive phenotypic and genetic patterns (Kawecki & Ebert, 2004). This field of research helps to achieve a more comprehensive knowledge about what processes underlie the genetic diversity of a species, since it focuses on adaptive processes and their genomic signatures as a complementary view to the phylogeography, which focuses on putatively neutral loci to explain demographic processes (McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013; Gavin et al., 2014). The analysis of spatial patterns of adaptive genetic variation will allow to disentangle which of the observed genetic variation within the species has been shaped by natural selection (rather than by demography, including gene flow and other evolutionary neutral processes), what

environmental factors are driving local adaptation and what traits are affected (Kawecki & Ebert, 2004; Tiffin & Ross-Ibarra, 2014). Since a broad range of topics can be addressed in this field, some of the topics which stand out are speciation through ecological adaptation, species response to changing environments and climate change, or management and conservation of genetic resources (Hendry, Nosil, & Rieseberg, 2007; Shafer & Wolf, 2013; Flanagan, Forester, Latch, Aitken, & Hoban, 2018; Sork, 2018).

When we deal with these questions directly focusing on variation at the molecular level (population genetic approaches; see Savolainen, Lascoux, & Merilä, 2013; Hoban et al., 2016; Sork, 2018 for a review on methodologies, including those based on phenotype data) it is possible to use genome scan methods, which screen large pools of loci to find markers with extreme levels of differentiation (putatively affected by natural selection) among previously defined groups (populations, phenotypes, etc.). Since genome scan methods can use exclusively genomic (genome-wide markers, usually SNPs) data (e.g., without measuring phenotypes, which are not always easy to measure; Stapley et al., 2010) and high-throughput sequencing technologies makes the acquisition of large genomic datasets feasible also in non-model species without reference genomes, these methods greatly facilitate the research on local adaptation. Among them, we can find two different approaches to look for putative adaptive loci from a large pool of neutral loci experiencing other evolutionary forces: i) differentiation outlier methods, which identify outlier loci based on high genetic differentiation between populations and ii) gene-environment association methods, which are based on associations between locus frequencies and environmental gradients (Hoban et al., 2016). Both types of methodologies include different statistical models to detect the outliers, but all have in common the incorporation of different approaches to control for the confounding population structure created by demographic processes. Yet, those methods may be affected by false positives despite the correction for the genetic structure and, additionally, other processes such as unknown environmental variables or background selection can produce outliers as well (De Mita et al., 2013; de Villemereuil & Gaggiotti, 2015; Hoban et al., 2016). For that reason, the outlier loci detected must be considered as candidate loci under selection that will need to be further investigated to validate their influence on adaptive species traits (i.e., gene annotation and functional characterisation, Tiffin & Ross-Ibarra, 2014; Pardo-Diaz, Salazar, & Jiggins, 2015).

A third aspect of examining the large genetic structure is considering the taxonomic level when analysing the genetic structure in taxa above the species level. With this taxonomic approach to genetic structure, it is possible to detect the macroevolutionary patterns and processes acting at the species level or above, gain insight into the drivers and mechanisms of diversification and speciation, and connect them to patterns of biodiversity, which are major questions in evolutionary biology (Butlin et al., 2012; Myers, McKelvy, & Burbrink, 2020). We may distinguish ecological drivers of diversification and speciation, such as adaptation to different environments, and non-ecological drivers, such as geographic barriers or polyploidy that lead to reproductive barriers (although usually interacting with ecological characteristics; Schluter, 2009; Sobel, Chen, Watt, & Schemske, 2010). Distinguishing which speciation mechanisms

operate in a given study system is an important question, as their outcomes would differ in terms of patterns of divergence (e.g., considering the spatial scale: sympatry vs. allopatry, or phylogenetic relationships, Weir & Schluter, 2007; Fitzpatrick, Fordyce, & Gavrilets, 2009; Pyron & Burbrink, 2010) or speed (e.g., faster if mediated by hybridization or sexual selection; Hoskin & Higgie, 2010; Sobel, Chen, Watt, & Schemske, 2010; Kraaijeveld, Kraaijeveld-Smit, & Maan, 2011).

The analysis of genetic structure at species level or above may also help in the delimitation of species, especially when combined with other information such as morphological and ecological traits (Duminil & Di Michele, 2009; Edwards & Knowles, 2014; Noguerales, Cordero, & Ortego, 2018). This approach is called integrative taxonomy (Schlick-Steiner et al., 2010) and has facilitated the identification of errors in morphological identifications (Dexter, Pennington, & Cunningham, 2010), the discovery of new cryptic taxa or the characterization of taxonomic groups (Sim-Sim et al., 2017; Johnson et al., 2018; Younger et al., 2018). In this sense, the botanical identification of tree species is fundamental for many fields, including documenting biodiversity in ecosystems, analysing the ecology and the evolutionary processes of species, as well as the correct implementation of management practices and conservation measures in forest areas (Lacerda & Nimmo, 2010; Baker et al., 2017). It is, for example, stated as key information to ensure the legality in timber trade in the international context (e.g., see the EU Timber Regulation 2010, the EU FLEGT Regulation 2005, the US Lacey Act 2008, and the Australian Illegal Logging Prohibition Act 2012[1]).

The use of multiple unlinked nuclear markers is suitable to perform species delimitation in plants, particularly in angiosperms, because these markers are less affected by introgression (Petit & Excoffier, 2009). As the chloroplast (or mitochondrial) genome is a single molecule, it easily passes from a species A to a species B (i.e., introgresses). When an individual of species A is pollinated by species B, a hybrid offspring AB is formed. This offspring can backcross with species B, leading to introgression of the full maternally inherited plastome from A to B. Alternatively, since unlinked nuclear markers sort randomly into gametes, only a fraction of them introgresses. Therefore, species delimitation based on nuclear markers is particularly relevant when the distribution ranges of species overlap and there is no information about whether they hybridize.

It has also been shown that the genomic location of the nuclear markers (genic or non-genic) influences their utility in population genetic analyses, especially in the case of microsatellites (SSR, simple sequence repeats) where genic location improves their

[1]Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market Text with EEA relevance.

Council Regulation (EC) No 2173/2005 of 20 December 2005 on the establishment of a FLEGT licensing scheme for imports of timber into the European Community.

Lacey Act (18 USC 42-43; 16 USC 3371-3378).

Illegal Logging Prohibition Act 2012, No. 166, 2012

usefulness (Defaveri, Viitaniemi, Leder, & Merilä, 2013). One of the advantages on adaptive markers is that they often reflect better the genetic structure of genetic demes as they tend to contain less variability and an increased population divergence at the same time as they reflect response to selection pressures (one of the major drivers of speciation) better than neutral markers (Kostamo, Korpelainen, & Olsson, 2012; Defaveri, Viitaniemi, Leder, & Merilä, 2013). That would be especially important for species delimitation in sympatric species, (i.e., without physical barriers to gene flow, like mountains or distance) in which no information about their hybridization or introgression was available.

Among the currently available types of markers, nuclear SSR and SNP (single nucleotide polymorphism) markers are efficient in the detection of genetic structure. Among the advantages of SSR markers we find that they are highly polymorphic, present a high mutation rate and thus, can be very informative on genetic structure (Guichoux et al., 2011; Haasl & Payseur, 2011; Putman & Carbone, 2014). However, they also present some drawbacks such as homoplasy (i.e., the same allelic state in two individuals due to independent mutation and not because of a common ancestor) or difficulties in getting consistent genotyping across laboratories (Morin, Luikart, & Wayne, 2004; Morin, Manaster, Mesnick, & Holland, 2009; Guichoux et al., 2011). Instead, SNP markers are comparatively more numerous in the genome, with low levels of homoplasy, and present accurate information about the genetic structure when their low variability (i.e., they are usually biallelic and evolve slower, which results in lower allelic diversity) is compensated by larger number of markers (i.e., they are suitable for high-throughput sequencing technologies and new bioinformatic tools; Ljungqvist, Åkesson, & Hansson, 2010; Guichoux et al., 2011). In fact, although it has been shown that SSRs better detect the genetic structure at fine scale compared with SNPs when used in low numbers, the same power can be achieved using a sufficiently large number of SNPs (Defaveri, Viitaniemi, Leder, & Merilä, 2013; Haasl & Payseur, 2011; Putman & Carbone, 2014; e.g., Zimmerman, Aldridge, & Oyler-Mccance, 2020).

Overall, both types of markers can be of similar efficiency to detect signatures of deep/ancient population genetic structure (Defaveri, Viitaniemi, Leder, & Merilä, 2013). Therefore, both SSR and SNP markers can be used for species delimitation and hybridization analysis (Duminil & Di Michele, 2009; Viscosi, Lepais, Gerber, & Fortini, 2009; Väli et al., 2010; Daïnou et al., 2016) and they are also very suitable for fine-scale genetic structure within populations (Leslie et al., 2015; Torroba-Balmori et al., 2017).

1.3. Challenges to evolutionary research in tropical rainforest trees

Tree species are good study systems to analyse their genetic structure and underlying evolutionary processes at different levels. Trees are sessile life-forms that often present large populations connected by gene flow, frequently occurring in different environments, with high levels of usually undomesticated genetic diversity compared to other plant forms: all of these features boost the efficiency of selection leading to local adaptation,

facilitate traceability of gene movements and set an appropriate scenario to disentangle selective from stochastic evolutionary forces (Petit & Hampe, 2006; Neale & Kremer, 2011; Sork et al., 2013; Fetter, Gugger, & Keller, 2017).

Traditionally, evolutionary research on trees has been more focused on boreal and temperate trees (Neale & Kremer, 2011), since working with tropical tree species includes several challenges. On one hand, working with tropical species poses logistic challenges such as transport, forest access or permission access, and difficult field conditions (Dick, 2010). As a result, when performing research on a tropical species, it is not untypical to gather samples from different sources, sometimes obtained through collaborations, which may also imply a variable number of samples per geographical location of the species. This thesis reflects all those situations.

Besides the sampling obstacles, morphological identification on tropical tree species has to overcome absence or scarcity of reproductive characters on individuals during the sampling, large species diversity, morphological variants within species or the existence of undescribed species, among other difficulties (Duminil & Di Michele, 2009; Dexter, Pennington, & Cunningham, 2010, e.g., Duminil, Kenfack, Viscosi, Grumiau, & Hardy, 2011). Moreover, ensuring consistent and accurate identifications across their distribution ranges can be a difficult task, especially in species-rich clades, due to the current high numbers of specimens and herbaria that need to be revised and their species confirmed by a low number of existing taxonomists, and the scarce taxonomic revisions across their geographical distribution (Goodwin, Harris, Filer, Wood, & Scotland, 2015; Baker et al., 2017).

These difficulties have slowed down the pace in evolutionary research in tropical tree species, an understudied group of trees yet the most diverse, compared to the more accessible boreal and temperate species (Neale & Kremer, 2011; Fetter, Gugger, & Keller, 2017). For example, research on local adaptation, especially at large scales, has advanced faster in boreal and temperate species (e.g., Eckert & Dyer, 2012; Parchman et al., 2012; Alberto et al., 2013), although research on tropical species is increasingly making progress (e.g., Collevatti et al., 2019; Melo, Vieira, Novaes, Bacon, & Collevatti, 2020; Brousseau, Fine, Dreyer, Vendramin, & Scotti, 2021).

Nonetheless, work with tropical trees is fundamental. It is well-known that tropical rainforests harbour outstanding woody plant diversity compared to other ecosystems in other latitudes (Fine & Ree, 2006; Kerkhoff, Moriarty, & Weiser, 2014; Slik et al., 2015). Their existence raises many different questions related to this biome (see e.g., Antonelli et al., 2018; Dick & Pennington, 2019) among which the need to record their plant diversity and distribution (e.g., Dauby et al., 2016 ; Ter Steege et al., 2016; Sosef et al., 2017), the maintenance through time of floristic diversity (e.g., Molino & Sabatier, 2001; Wright, 2002; Hubbell, 2005; Kerkhoff, Moriarty, & Weiser, 2014; Schmitt, Tysklind, Derroire, Heuertz, & Hérault, 2021) and, noticeable, the specific evolutionary processes, particularly between closely related taxa, which contribute to the generation of such hyper-rich ecosystems (e.g., Fine, Daly, Muñoz, Mesones, & Cameron, 2005;

Pennington, Richardson, & Lavin, 2006; Kursar et al., 2009; Couvreur, Forest, & Baker, 2011; Schmitt, Tysklind, Hérault, & Heuertz, 2021) stand out.

It needs to be verified if evolutionary findings based on the more abundant research on boreal and temperate trees are equally valid in tropical trees (Fetter, Gugger, & Keller, 2017) considering the evident differences in life-history traits. For example, in contrast with temperate and boreal species, tropical tree species present relatively higher levels of genetic differentiation among populations, noticeably lower population densities, pollen and seed dispersal generally depend on animal vectors and dioecy syndrome is more frequent, all these features leading to idiosyncratic patterns of gene flow in tropical tree species (Dick, Hardy, Jones, & Petit, 2008). Increasing the knowledge on this biome is of utmost importance to perform effective sustainable management and conservation, especially regarding the ongoing effects of global climate change (Dick, Lewis, Maslin, & Bermingham, 2013; Valladares et al., 2014) and the high rates of global deforestation (FAO & UNEP, 2020).

1.4. Study group: the tropical tree genus *Symphonia* L. f.

We selected the tropical tree genus *Symphonia*, which belongs to the Clusiaceae family, as a study group to analyse the genetic structure across multiple spatial and taxonomic scales and infer the evolutionary underpinnings of the observed genetic structure. The genus evolved as early as the Eocene, with fossil pollen dated to ~45 Ma in the Niger delta (Jan-du-Chêne, Onyike, & Sowunmi, 1978; Dick, Abdul-Salim, & Bermingham, 2003), and harbours 16-21 endemic tree species in Madagascar. However, *Symphonia globulifera* is currently the only recognized *Symphonia* species distributed throughout tropical Africa and America but not Madagascar (Perrier de la Bâthie, 1951; Abdul-Salim, 2002; Oyen, 2005; The Plant List, 2013). This species-rich genus, with its widespread distribution in many habitats across continents and its high diversity in Madagascar, is a particularly suitable model to investigate the evolutionary processes related to spatial and taxonomic genetic structure that produce the patterns of genetic diversity in tropical trees.

1.4.1. *Symphonia globulifera*

Symphonia globulifera L. f. is represented by generally tall, hermaphroditic late-successional rainforest trees widespread throughout tropical Africa, Central and South America (see Fig. 1; Oyen, 2005). The species occurs today in tropical forests, in a wide range of precipitation and temperature of 650-2,800 mm and 23-27 ºC, respectively, and from sea level to 2,600 m altitude (in East Africa; Oyen, 2005). The species is mostly outcrossing (Degen, Bandou, & Caron, 2004; da Silva Carneiro, Sebbenn, Kanashiro, & Degen, 2007; da Silva Carneiro, Degen, Kanashiro, de Lacerda, & Sebbenn, 2009) although selfing can occur, especially in fragmented areas (Aldrich, Hamrick, Chavarriaga, & Kochert, 1998; Aldrich & Hamrick, 1998). Pollinators and seed dispersers vary in different parts of its range, with notable differences between African

and Neotropical populations. The red, showy, hermaphroditic flowers are pollinated by sunbirds in Africa, and by bees, Lepidoptera, hummingbirds or perching birds in different locations in the Neotropics, while the seeds of its drupaceous fruits are dispersed by various animals, including monkeys, ruminants and hornbills in Africa, and bats, scatter-hoarding rodents, monkeys and tapirs in the Neotropics (see Table 1 and Fig. 2).

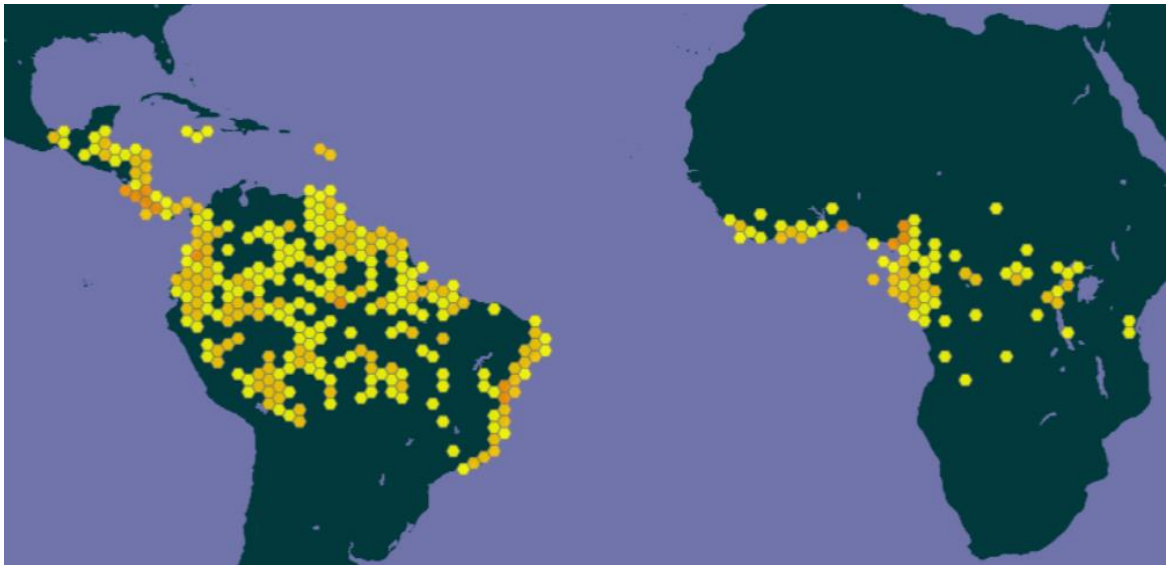


Figure 1. Distribution range of *Symphonia globulifera*. Extracted from https://www.gbif.org/es/ (Date: 9th November 2021; query: *S. globulifera* L. f.).

The species is believed to have colonized the Neotropics from Africa through sweepstakes trans-Atlantic dispersal (possibly via whole trunks or roots, since seeds are recalcitrant and cannot survive desiccation), with fossil pollen evidence from the Neotropics dating to the Early Miocene (ca. 15 and 18 Ma). The species has been widespread in both Africa and the Neotropics for millions of years (Dick, Abdul-Salim, & Bermingham, 2003). *Symphonia globulifera* has persisted locally in many sites throughout the Quaternary glaciations (Dick & Heuertz, 2008; Budde, González-Martínez, Hardy, & Heuertz, 2013; Jones, Cerón-Souza, Hardesty, & Dick, 2013; Barthe et al., 2017) and previous research identified the effect of biogeographic barriers on the continental scale geographic structure, with a pronounced effect of the Andes in the Neotropics (Dick, Abdul-Salim, & Bermingham, 2003; Dick & Heuertz, 2008; Dick, Lewis, Maslin, & Bermingham, 2013) and of the Cameroon Volcanic Line and the Gulf of Guinea in Africa (Budde, González-Martínez, Hardy, & Heuertz, 2013). Different morphotypes or suspected ecotypes of *S. globulifera* occur in several regions, including a small tree form in Costa Rica (Dick & Heuertz, 2008; Sanfiorenzo, 2018) and a suspected swamp ecotype in West Africa (Budde, González-Martínez, Hardy, & Heuertz, 2013). In French Guiana, the common flood-tolerant *S. globulifera* morphotype preferentially grows in seasonally flooded bottomlands and co-occurrs with a *terra firme* morphotype (*Symphonia sp.1*), which occurs in drier habitats at higher topographic

positions and presents smaller flowers, smooth bark and adventitious roots but no pneumatophores (Sabatier et al., 1997; Molino & Sabatier, 2001, Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007, Allié et al., 2015; Schmitt et al., 2020). It has been shown that both morphotypes present genetic and functional differences related to the fine-scale topographic position, which is a proxy for different microhabitats with different distribution of water and nutrients (Schmitt, Tysklind, Hérault, & Heuertz, 2021). They also present different performance in survival and growth in the presence of the same environmental constraints, where *Symphonia sp.1* is more generalist (Tysklind et al., 2020). Further, Schmitt, Tysklind, Hérault, & Heuertz, (2021) also recognized a third morphotype in Paracou, splitting *S. globulifera sensu stricto* into two morphotypes, associated to their own genetic and functional characteristics.

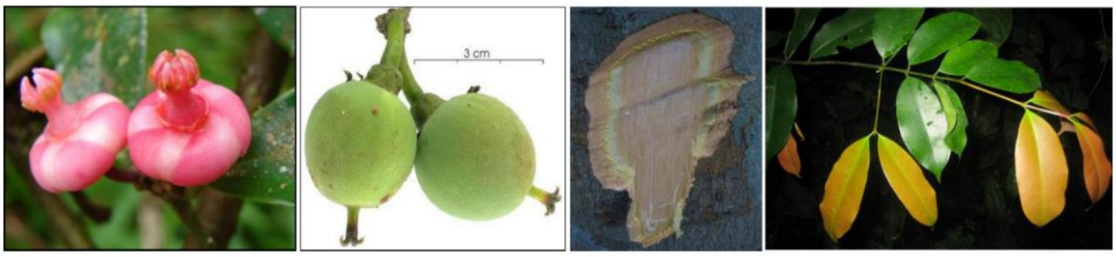


Figure 2. Morphological traits in *Symphonia globulifera*: red flowers (left, ©Tobias Sandner, University of Marburg), immature fruits (center left, © Smithsonian Tropical Research Institute), cut trunk revealing yellow latex (center right, © Myriam Heuertz), and opposite leaves with young leaves in yellow brownish color (right, © Myriam Heuertz). Reproduced with permission from Budde, (2014).

All in all, our species represents excellent characteristics to investigate the patterns and processes associated to the spatial genetic structure at both micro and macro scales:

First, *S. globulifera* presents an appropriate situation to discriminate the influence of drift-dispersal vs. landscape/ecological features and historical processes on FSGS at the population level. Because *S. globulifera* populations persisted through multiple geological time periods (Dick & Heuertz, 2008; Budde, González-Martínez, Hardy, & Heuertz, 2013; Barthe et al., 2017), we hypothesized that populations should be close to demographic equilibrium and display IBD due to drift-dispersal processes (e.g., Hardy et al., 2006). Given the life history traits of our species, i.e., an essentially outcrossed, animal-pollinated and animal-dispersed tropical tree, we expected the FSGS quantified with the *Sp* statistic to be approximately 0.01-0.02 (Vekemans & Hardy, 2004; Hardy et al., 2006), probably with substantial variation among populations with different dispersers. We also hypothesized that FSGS should vary among populations because of idiosyncratic ecological characteristics of populations, e.g., a stronger FSGS should *a priori* be expected in populations with marked topography (Robledo-Arnuncio, Collada, Alía, & Gil, 2005; Rhodes, Fant, & Skogen, 2014) or narrow-ranging dispersal (Hardy et al., 2006; Dick, Hardy, Jones, & Petit, 2008). Complex interactions between ecological

features and each population's specific history could lead to SGH. Hence, their effects on FSGS are difficult to predict.

Second, the analysis of spatial genetic structure at macro scale in such an ancient widespread tree species as *S. globulifera* will allow us to reveal environmental processes (e.g., climate or topography) that have led to its current spatial patterns of genetic diversity. Among those, the effect of past climate oscillations during the Pleistocene will be presumably among them, as it has been a major influential force shaping the genetic structure of many temperate and tropical species, including plants (Hewitt, 2000). We will also gain insight into the role of long-distance marine dispersal in the colonization of our species across continents, an extreme mode of dispersal that has also affected the range establishment in other tropical species across the same continents (Pennington & Dick, 2004; Renner, 2004).

Third, deeper research focused on processes leading to species diversification and the formation of cryptic species (i.e., those classified as the same species as the main species but presenting subtle genetic, morphological, or ecological differences, among other traits, Bickford et al., 2007) is also a perspective in S. *globulifera*, as the presence of morphotypes or ecotypes may allow discover distinct lineages within the species.

And finally, understanding the mechanisms that influence a species' range margins are a major topic in plant ecology (Sutherland et al., 2013). The wide geographic ranges of species would imply that they have had excellent adaptive potential to adapt to different local conditions and persist in diverse communities over time, as expected in large populations occurring in habitats with contrasting environments (see Leimu & Fischer, 2008 and Hereford, 2009, for reviews on local adaptation). In our study case, the diverse ecological conditions in which *S. globulifera* occurs across its wide distribution range allow us to investigate the genomic signatures of local adaptation. Also, the investigation will facilitate gaining insight into what evolutionary strategies underlie the wide ecological ranges and tolerances in our species, so we might be able to identify putative drivers of adaptation associated with gene pools.

Table 1. Review of animals reported as seed dispersers or pollinators of *Symphonia globulifera* in Africa or the Neotropics and characteristics of their dispersal range**.** P, pollinator; sd, seed disperser.

Visitors	Cited genus/species	Function	Country or region	Source	References	Dispersal range	References
Africa							
sunbirds	*Cyanomitra, Nectarinia, Cinnyris, Chalcomitra, Hedydipna*	P	Central and South Africa	bibliographical compilation	(Mann & Cheke, 2001; Oyen, 2005)	max: 50-100 m (*Chalcomitra amethystina*)	(Van der Niet, Cozien, & Johnson, 2015)
monkeys	*Cercopithecus lhoesti*	sd (defecation, spitting out, transportation in cheek pouches)	Uganda, Cameroon, Gabon	direct observation, seed traps	(Gautier-Hion, Emmons, & Dubost, 1980; Gautier-Hion et al., 1985; Clark, Poulsen, & Parker, 2001; Poulsen, Clark, & Smith, 2001; Ukizintambara, 2009)	range: a few meters – 100m (*Cercopithecus* monkeys, seeds >1cm)	(Kaplin & Lambert, 2002)
small ruminants	*Cephalophus monticola, Hyemoschus aquaticus*	sd (defecation, regurgitation, predation?)	Gabon	direct observation, stomach content	(Gautier-Hion et al., 1980; Dubost, 1984; Gautier-Hion et al., 1985; Forget et al., 2007)	no information	
hornbills	Putatively: *Tockus fasciatus, Bycanistes fistulator, B. albotibialis, Ceratogymna atrata*	sd (defecation?, regurgitation)	Gabon	direct observation, stomach content	(Gautier-Hion et al., 1985; Forget et al., 2007)	max: >500 m (*Ceratogyma atrata, C. cylindricus*)	(Holbrook & Smith, 2000)
Neotropics							
hummingbirds	*Chlorestes notatus, Thalurania furcata*	P	Costa Rica, Brazil, French Guiana	direct observation	(Bawa, Bullock, Perry, Coville, & Grayum, 1985; Bittrich & Amaral, 1996)	max: 1-100 m (depending on the species)	(Linhart & Mendenhall, 1977; Webb & Bawa, 1983)
perching birds	*Cacicus cela, Dacnis lineata, Dacnis cayana, Chlorophanes spiza, Cyanerpes caeruleus, Cyanerpes cyaneus*	P	Brazil	direct observation	(Bittrich & Amaral, 1996 ; Gill et al., 1998)	no information	
lepidoptera	unidentified	P	Brazil, Costa Rica	direct observation	(Bittrich & Amaral, 1996 ; Pascarella, 1992)	max: 8-10 m (species pollinating *Cnidoscolus urens* in Costa Rica and *Lindenia rivalis* in Belize)	(Linhart & Mendenhall, 1977; Webb & Bawa, 1983)
bees	*Trigona cf. branneri*	P	Brazil	direct observation	(Bittrich & Amaral, 1996)	mean: 260-590 m in buzz-pollinating bees (*Scaptotrigona, Trigona, Xylocopa*)	(Jha & Dick, 2010)
bats	*Artibeus lituratus, Artibeus jamaicensis, Artibeus watsoni*	sd (exozoochorous)	Costa Rica, French Guiana	direct observation, seed rain under feeding roosts	(Charles-Dominique, 1986; Aldrich, Hamrick, Chavarriaga, & Kochert, 1998; Melo, Rodriguez-Herrera, Chazdon, Medellin, & Ceballos, 2009)	max: 100 m (*Artibeus lituratus*), max: 25-400 m (*Artibeus jamaicensis*)	(Morrison, 1978; Morrison, 1980)
scatter-hoarding rodents	unidentified	sd (exozoochorous)	French Guiana	cited	(Abdul-Salim, 2002; Hardy et al., 2006)	mean: 5-15 m (rodents)	(Forget, 1990; Brewer & Rejmánek, 1999)
nocturnal arboreal mammals	unidentified	sd (unknown)	French Guiana	cited	(Hardy et al., 2006)	no information	
monkeys	*Leontopithecus rosalia*	sd (unknown, defecation is possible)	Brazil	direct observation	(Lapenta, Procópio de Oliveira, Kierluff, & Motta-Junior, 2003; Miller & Dietz, 2006)	mean: 105 m	(Lapenta & Procópio-de-Oliveira, 2008)
tapirs	*Tapirus terrestris*, *Tapirus bairdii*	sd (defecation)	French Guiana, Central America	stomach content, bibliographical compilation	(Brooks, Bodmer, & Matola, 1997 ; Henry, Feer, & Sabatier, 2000; Forget et al., 2007)	max: 2 km	(Fragoso, 1997; Giombini, Bravo, & Tosto, 2016)

1.4.2. Malagasy *Symphonia*

Like many other groups of tropical species, Malagasy *Symphonia* species have experienced almost no revisions on their taxonomy and nomenclature in the last 50 years (Goodwin, Harris, Filer, Wood, & Scotland, 2015; for last revisions see Perrier de la Bâthie, 1951; Abdul-Salim, 2002). As a result, the *Symphonia* genus in Madagascar encompasses between 16 and 21 endemic species in sympatry (see Table 2), with high levels of uncertainty in their delimitation, as the most recent taxonomic revision did not identify enough segregating morphological characters to clearly discriminate among species due to high variability within species (Abdul-Salim, 2002). Therefore, Malagasy *Symphonia* is an understudied system with poor criteria for morphological species discrimination.

Table 2. Taxonomic and nomenclature revision of the genus *Symphonia* in Madagascar

Taxonomic and nomenclature revision		
Perrier de la Bâthie (1951)	**Abdul-Salim (2002)**	**Currently accepted names (from http://powo.science.kew.org/, consulted: 11th October 2021)**
S. ambrensis H. Perr.		
	S. andriantianii Abdul-Salim	
S. clusioides Bak.	*S. clusioides* Baker	*S. clusioides* Baker
S. eugenioides Bak.	*S. eugenioides* Baker	*S. eugenioides* Baker
S. fasciculata Benth. et Hook. f. ex Vesque	*S. fasciculata* (Norontha ex Thou.) Vesque	*S. fasciculata* (Thouars) Baill.
		S. gymnoclada (Planch. & Triana) Benth. & Hook.f. ex Vesque
	S. humbertii Abdul-Salim	
S. lepidocarpa Bak.	*S. lepidocarpa* Baker	*S. lepidocarpa* Baker
S. linearis H. Perr.	*S. linearis* H. Perrier	*S. linearis* H. Perrier
S. louveli Jum. et Perr.	*S. louveli* Jumelle	*S. louvelii* Jum.
S. macrocarpa Jum. et Perr.		
S. microphylla Benth. et Hook. f. ex Vesque	*S. microphylla* (Hilsenberg & Bojer ex Cambess) Benth. & Hook. f. ex Vesque	*S. microphylla* (Hils. & Bojer ex Cambess.) Benth. & Hook.f. ex Vesque
S. nectarifera Jum et. Perr.	*S. nectarifera* Jumelle & Perrier	*S. nectarifera* Jum. & H. Perrier
S. oligantha Bak.	*S. oligantha* E G. Baker	*S. oligantha* Baker f. (synonym: *S. ambrensis* H. Perrier)
	S. paraeugenioides Abdul-Salim	
	S. parviflora Abdul-Salim	
S. pauciflora Bak.	*S. pauciflora* Baker	*S. pauciflora* Baker
	S. perrieri Abdul-Salim	
	S. sambiranensis Abdul-Salim	
S. sessiliflora H. Perr.	*S. sessiliflora* H. Perrier	*S. sessiliflora* H. Perrier
	S. spatulata Abdul-Salim	
S. tanalensis Jum. et Perr.	*S. tanalensis* Jumelle	*S. tanalensis* Jum.
S. urophylla (Dene.) Benth. et Hook. f. ex Vesque	*S. urophylla* (Decne. Ex Planch. & Triana) Vesque	*S. urophylla* (Decne. ex Planch. & Triana) Vesque
S. verrucosa Benth. et Hook. f.	*S. verrucosa* (Hilsenberg & Bojer ex Planch. & Triana) Vesque	*S. verrucosa* (Hils. & Bojer ex Planch. & Triana) Vesque

Ambiguous delimitation in morphology allows us to perform a hypothesis-driven approach of integrative taxonomy (Schlick-Steiner et al., 2010) where we can examine the morphology using genetic information and other sorts of data (e.g., Edwards & Knowles, 2014). Thus, Malagasy *Symphonia* constitutes an interesting study system to test if the combination of genetic structure and other genetic information, morphology and environmental aspects such as geography can facilitate and clarify the delimitation of Malagasy *Symphonia* species. At the same time, we will be able to analyse if non-morphological information is congruent with the current taxonomic classification. The results will improve the characterization of this poorly defined group and increase the botanical knowledge of the system.

Moreover, such analysis may help to identify the mechanisms underpinning species diversification and accumulation of biodiversity in species-rich environments such as the hotspot of Madagascar (Sites & Marshall, 2003; Myers, Mittermeler, Mittermeler, Da Fonseca, & Kent, 2000). Closely related species of *Symphonia* in Madagascar, an oceanic island, present a suitable study case to analyse the processes of their radiation and may help us to increase the knowledge about speciation in tropical taxa. That question is even more relevant when we compare the surprising species diversity of *Symphonia* in the island of Madagascar with the existence of a unique *Symphonia* species outside Madagascar, with a broad distribution range which comprises two continents. The results will allow further research into the ecology of the constituent sister species and their process of radiation.

2. Objectives of the thesis

The main objective of this thesis is to analyse the genetic structure of the genus *Symphonia* at different spatial scales and at different taxonomic levels to unveil the evolutionary dynamics at micro and macro scales that have led the genus *Symphonia* to its current situation in terms of ecology, geographical distribution, local adaptation and genetic diversity. Therefore, with this aim, I performed a three-fold approach on the genetic structure analysis of the genus (from micro to macro scale), considering (i) within-population scale and (ii) continental scale within one species (*S. globulifera*), and (iii) genus scale involving several sympatric *Symphonia* species in Madagascar. The specific objectives of this thesis were (see also summary in Table 3):

- To analyse the fine-scale genetic structure of *S. globulifera* in several African and Neotropical populations in relation to the different genetic dispersal strategies (micro-scale approach) and address the following specific questions: (i) Is within-population FSGS in *S. globulifera* in agreement with expectations based on the species' life history traits, and to what extent does its strength vary among populations? (ii) Is FSGS in agreement with drift-dispersal equilibrium as predicted by IBD theory or are there within-population discontinuities in allele frequencies (SGH)? (iii) Are there any similarities in the strength and patterns of FSGS in groups of populations, and do they concur with, e.g., similar disperser communities, habitat features or biogeographic history? (Study I).
- To investigate the colonization and diversification history of African and Neotropical populations of *S. globulifera,* considering the spatial genetic structure at continental scale (macro-scale approach) and the influence of past climatic events as well as of other environmental and geographical factors at those large geographic scales (Study II, part 1).
- To address the genomic signatures of local adaptation of *S. globulifera* in Africa and identify some of the drivers involved in this regional adaptation that may underlie the ability of our species to occur throughout a wide range of habitats and a large distribution range (macro-scale approach; Study II, part 2).
- To develop new single nucleotide polymorphism (SNP) markers with potential relevance for adaptation in the genus *Symphonia* (i.e., S. *globulifera* and Malagasy *Symphonia spp*.) and use them to get evolutionary insights into the *Symphonia* species group in Madagascar. The specific aims are to (i) infer phylogenetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species) for the first time, (ii) to test the current species delimitation in Malagasy *Symphonia* using integrative taxonomy, (iii) to investigate the evolutionary history underlying the radiation of *Symphonia* species in Madagascar (macro scale approach at the genus level; Study III).

Table 3. Overview of the structure of this thesis, including objectives and materials and methods

	Objectives	**Materials and Methods**			
		Species	**Locations***	**Molecular Markers**	**Methods**
Study I see also: Torroba-Balmori et al., (2017) PLoS ONE 12(8): e0182515. https://doi.org/10.1371/journal.pone.0182515 (see Supplementary Information S9.1.6.)	Fine-scale spatial genetic structure (FSGS) vs. genetic dispersal strategies; drivers of FSGS	*S. globulifera*	*Africa* (ST, NM, MB) *The Neotropics* (BCI, YS, PR, IT)	3-5 nSSR (nuclear microsatellites) psbA-trnH plastid DNA (cpDNA) sequences	Genetic diversity analyses SPAGeDi MicroChecker sPCA STRUCTURE TESS Fisher tests Partial Mantel Tests One-way ANOVA
Study II, part 1	Colonization & diversification history	Two morphotypes: *S. globulifera* rangewide and *S. sp.1* in French Guiana	*Africa* (BN, KR, MB, NK, ST, GB) *The Neotropics* (IT, PR, RG)	4921 SNPs from GBS (genotyping-by-sequencing)	Genetic diversity analyses PCA (principal component analysis) Entropy NJ tree based on Nei's D TreeMix
Study II, part 2	Genomic signatures of local adaptation Drivers of local adaptation	*S. globulifera*	*Mainland Africa* (BN, KR, MB, NK, GB)	3399 SNPs from GBS (genotyping-by-sequencing)	BayeScan BayeScEnv BayPass Gene Annotation
Study III	Development of SNPs with potential adaptive relevance To test the current species delimitation Evolutionary history	Genus *Symphonia*	*Africa* (BN, KR, MB, NK, ST, GB) *Madagascar* (10 locations) *The Neotropics* (IT, PR, RG)	144 SNPs with potential adaptive relevance using Sequenom technology	Flow cytometry NJ tree based on Nei's D STRUCTURE

*BCI-Barro Colorado Island, Panama; BN-Porto Novo, Benin; GB-Ngounié, Gabon; IT-Ituberá, Brazil; KR-Korup, Cameroon W; MB-Mbikiliki, Cameroon SW; NK-Nkong Mekak, Cameroon SW; PR-Paracou, French Guiana; RG-Regina, French Guiana; ST-São Tomé, São Tomé and Principe; YS-Yasuní, Ecuador.

3. Materials and Methods

3.1. Study sites, molecular markers and other related information

3.1.1. Fine-scale spatial genetic structure in *S. globulifera* (Study I)

Study sites and plant material

In our Study I, we examined seven populations from Africa and the Neotropics, all located in natural ancient forests (Table 4). Between 2007 and 2010, we collected leaf or cambium samples on 34-148 georeferenced trees per population and dried samples on silica gel. Sampled trees generally had ≥ 10 cm diameter at breast height (dbh), except in Barro Colorado Island (BCI, Panama) and Yasuní (Ecuador) where density was lower and sampled individuals had ≥ 1.0 cm dbh (Table 4). Sampling ranges spanned ca. 1-4 km, in transect-like design following topographic features for ease of orientation, except in the forest monitoring sites of Paracou (French Guiana), BCI and Yasuní, where random sampling was conducted in established plots.

Our research complied with national and international legislation: research and sampling permits were obtained from the Ministry of Scientific Research and Innovation of Cameroon (59/MINRESI/B00/C00/C10/C13), from the responsible of the Paracou station (i.e., the French Agricultural Research Institute for Development, CIRAD), and from the managers of the BCI and Yasuní plots (i.e., the Smithsonian Tropical Research Institute). For Ituberá (Brazil), we obtained a sampling permit from the Chico Mendes Institute for Biodiversity (SISBIO 19053-1) and an export permit from the Ministério do Meio Ambiente, Brazil (Requerimento Nº 107231).

Table 4. Physical and ecological characteristics of sampled *Symphonia globulifera* populations for Study I. $H_{m\text{-}M}$, minimum and maximum sampling altitude (m); T, annual mean temperature (ºC); P, annual precipitation (mm); and D (d), density of *S. globulifera* stems ≥10 cm dbh (≥1.0 cm dbh) with d only available for BCI and Yasuni (stems/ha); for populations not corresponding to monitoring sites, the values are approximate estimates (~); nnuc, sample size for SSR data; ncp, sample size for plastid DNA.

Population	Latitude	Longitude	$H_{m\text{-}M}$	T	P	D (d)	n_{nuc}	n_{cp}
Neotropics								
Barro Colorado Island, Panama (BCI)	9.15	-79.85	149-196	25.9	2632	0.48 (3.12)	147	10
Yasuní, Ecuador (YS)	-0.68	-76.39	231-273	23.8	2380	0.68 (1.76)	34	10
Paracou, French Guiana (PR)	5.27	-52.93	38-67	22.0	2496	10.5	148	96
Ituberá, Brazil (IT)	-13.80	-39.18	92-164	25.8	2817	~ 6.55	85	50
Africa								
São Tomé, São Tomé and Principe (ST)	0.27	6.56	671-1896	23.8	2058	~ 1.43	42	38
Nkong Mekak, Cameroon (NK)	2.77	10.53	473-838	25.8	2837	~ 9.58	70	49
Mbikiliki, Cameroon (MB)	3.19	10.53	467-911	25.6	2806	~ 9.10	94	50

The selected populations spanned a wide range of climatic (WorldClim 1.4 dataset; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and topographic (ASTER Global Digital Elevation Model, http://reverb.echo.nasa.gov/) conditions (Table 4). The altitudinal range of sampled populations was larger in Africa (365-1225 m) than in the Neotropics (25-72m). There was also a marked variation in dispersal vectors between continents (Table 1). In Paracou, samples included swamp and *terra firme* morphotypes (Sabatier et al., 1997). For this population, samples were collected for plastid DNA analysis, whereas SSR data (for different trees) was reanalysed from Degen, Bandou, & Caron, (2004).

Molecular markers

DNA was extracted using the Qiagen DNeasy plant kit (Qiagen Corporation, Valencia, CA) or the Invisorb DNA Plant HTS 96 Kit (Invitek, Berlin, Germany). SSR data (five loci) were generated at the University of Michigan, Ann Arbor, USA, for populations BCI and Yasuní, as described in Dick & Heuertz, (2008), and at INIA-CIFOR, Madrid, Spain, for populations Ituberá, São Tomé, Nkong Mekak and Mbikiliki, following the protocols of Budde, González-Martínez, Hardy, & Heuertz, (2013). SSR data for Paracou was obtained from Degen, Bandou, & Caron, (2004); and contained three loci. All SSR data were resolved on capillary sequencers (Applied Biosystems, Carlsbad, USA; see Supplementary Information S9.1.1. for experiment details and genotype matrices). The SSR loci of all populations belonged to a total set of six loci (Sg03 and Sg18, Degen, Bandou, & Caron, (2004); SgC4 and Sg19, Aldrich, Hamrick, Chavarriaga, & Kochert, (1998); Sg06 and Sg10, Vinson, Amaral, Sampaio, & Ciampi, (2005)), but they were not exactly the same ones in each population because of variable amplification and genotyping success caused by large genetic distances among some of the populations (Dick & Heuertz, 2008; Olsson et al., 2017).

Sequences of the *psbA-trnH* plastid DNA (cpDNA) intergenic spacer were generated for random subsamples from Paracou, Ituberá, São Tomé, Nkong Mekak and Mbikiliki, completing the data sets of Dick & Heuertz (2008) and Budde, González-Martínez, Hardy, & Heuertz, (2013) (see sample sizes in Table 4). Amplification with *psbAF* and *trnHR* primers (Sang, Crawford, & Stuessy, 1997) was performed at INIA-CIFOR as in Budde, González-Martínez, Hardy, & Heuertz, (2013) with a modified PCR profile: 30 s at 98ºC, 35 cycles of 5 s at 98ºC, 10 s at 50ºC and 35 s at 72º, and a final elongation of 3 min at 72ºC. PCR products were purified using Exonuclease I and Calf Intestinal Alkaline Phosphatase (New England Biolabs) and sequenced using the services of Macrogen Europe (The Netherlands). Sequences were assembled, edited and aligned in CodonCode Aligner 4.2.5 (CodonCode Corporation, Dedham, MA, USA).

3.1.2. Large-scale genetic structure in *S. globulifera*, demographic history and adaptive evolution (Study II, parts 1 & 2)

3.1.2.1. Spatial genetic structure of S. globulifera across continents

Study sites and plant material

In our Study II, part 1, we obtained samples from natural ancient forests in nine locations from tropical Africa and the Neotropics (Fig. 3, Table 5), believed to represent eight diverged gene pools according to previous knowledge of the species' biogeographic history (Dick & Heuertz 2008; Budde, González-Martínez, Hardy, & Heuertz, 2013; Torroba-Balmori et al., 2017). Individuals from all locations but Regina (French Guiana) belonged to previously studied populations (see Budde, González-Martínez, Hardy, & Heuertz, 2013; Torroba-Balmori et al., 2017 and Section 3.1.1.).

Samples in Africa were collected in three locations in Cameroon representing distinct west and southwest Cameroonian gene pools (southwest: Mbikiliki in the Ngovayang massif; Nkong Mekak in the buffer zone of the Campo Ma'an National Park, 48 km apart; west: Korup National Park), in one location in the south west of Gabon (Ngounié), in São Tomé island and in southern Benin (Porto Novo, in the Dahomey Gap). In the Neotropics, one location was sampled in the Brazilian Atlantic Forest (Ituberá) and two in French Guiana (Regina and Paracou). French Guiana locations included two previously recognized morphotypes with contrasted habitat preference, occurring in valley bottoms vs. *terra firme* environments, respectively (Sabatier et al., 1997).

For Ituberá, Regina and the African locations São Tomé, Mbikiliki and Nkong Mekak, a transect-like sampling design was used for ease of fieldwork. In Benin, Korup and Gabon, samples were collected randomly in several sampling missions. In Paracou, random sampling was conducted in established plots. In all sites, leaf or cambium samples were collected from trees (≥10 cm dbh) randomly sampled across the microenvironmental conditions in which they occurred. Trees were georeferenced and plant tissue dried on silica gel. Morphotype information was recorded for all samples from French Guiana.

Table 5. Sample size (N) and coordinates of sampling locations for *Symphonia globulifera* for Study II.

Sampling location	N	Latitude	Longitude
Africa			
Porto Novo, Benin (BN)	14	6.389	2.623
Korup, Cameroon W (KR)	19	5.072	8.836
Mbikiliki, Cameroon SW (MB)	65	3.197	10.524
Nkong Mekak, Cameroon SW (NK)	64	2.762	10.531
Ngounié, Gabon (GB)	20	-1.432	10.289
São Tomé, São Tomé and Principe (ST)	40	0.280	6.591
Neotropics			
Ituberá, Brazil (IT)	65	-13.795	-39.181
Paracou, French Guiana (PR)	60*	5.260	-52.924
Regina, French Guiana (RG)	20*	4.308	-52.235

*Those locations include *S. globulifera* and *S. sp.1* morphotypes

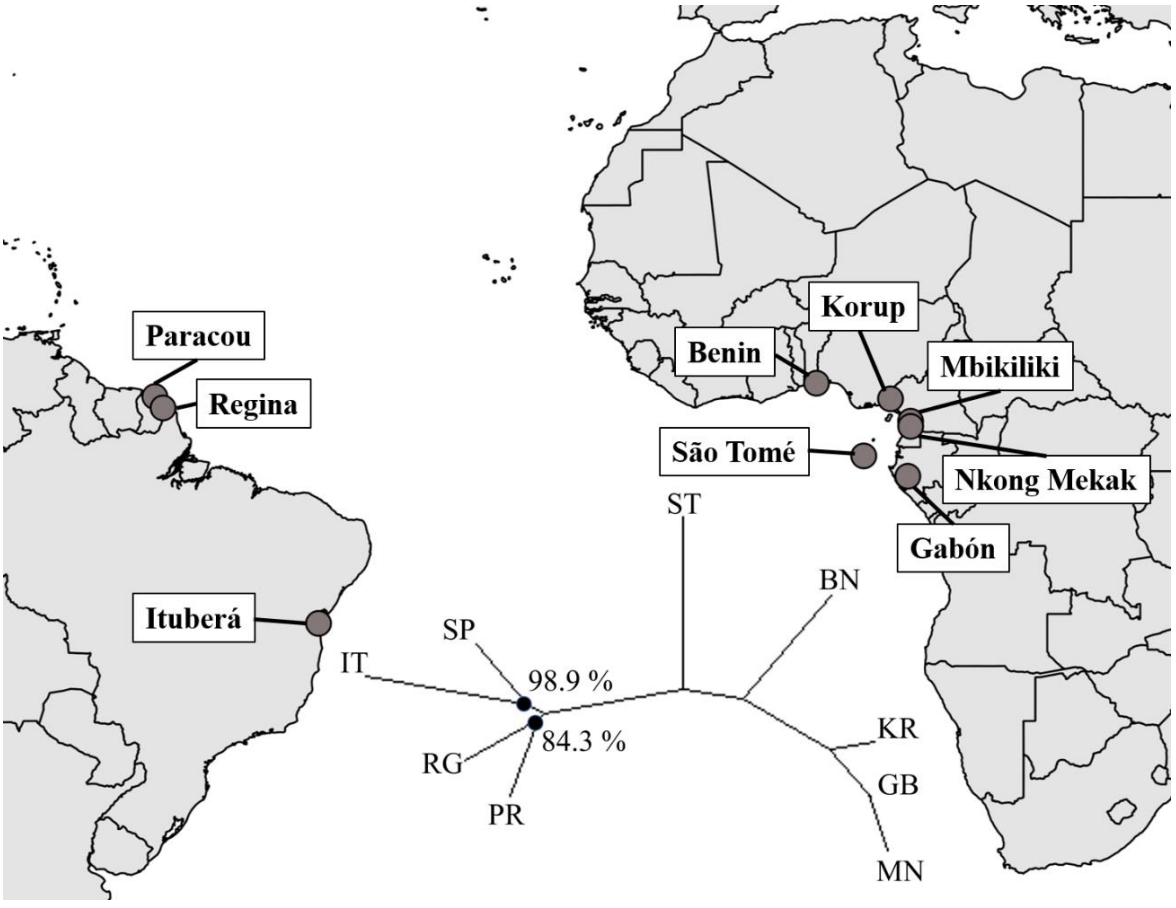


Figure 3. Location of the *Symphonia globulifera* populations in Study II and unrooted neighbour-joining tree of Nei's genetic distance with scaled branch lengths showing the genetic distance among gene pools of *Symphonia globulifera* (the values indicate the frequency in which bifurcating nodes occurs out of 1000 bootstrap replicates only when lower than 100 %).

Molecular markers

The methodology used to generate the molecular markers was Genotyping-by-Sequencing, which is a restriction-site-associated DNA sequencing methodology that uses high-throughput sequencing technologies (GBS; Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). DNA was extracted using the Invisorb DNA Plant HTS 96 Kit (Invitek, Berlin, Germany) or the Qiagen DNeasy Plant Kit (Qiagen Corporation, Valencia, CA). The quantity of DNA extract was increased following a whole genome amplification protocol (REPLI-g Mini kit, Qiagen) to ensure sufficient DNA quantity for library construction. We then prepared highly-multiplexed libraries for sequencing on the Illumina platform following the laboratory protocol for genomic enrichment of reduced complexity libraries described in Parchman et al. (2012). First, genomic DNA (20 - 150 ng/μL) was digested with two restriction enzymes, EcoRI and MseI. This produced sticky-ended fragments to which we ligated adaptor sequences containing the Illumina adaptor sequences and a unique 8–10 bp barcode for each sample to allow unambiguous identification of sample origin after pooled sequencing. Fragments from all individuals were then amplified through a PCR step using Illumina PCR primers and an extra PCR step to convert remaining single stranded templates to double stranded. Finally, all individual libraries (384 in total, all with unique barcodes) were pooled and divided into three library pools for size selection. To increase the number of reads sequenced for a reduced portion of the genome, we sized selected fragments in the range of 350-400 bp using a Pippin (Sage Science, Beverly, MA) quantitative electrophoresis system. Sequencing was carried out on three lanes of an Illumina HiSeq 2500 instrument at University of Texas producing single-end 100 base reads.

Prior to assembly, we cleaned contaminant or aberrant sequences (*Escherichia coli*, PhiX libraries, Illumina oligos) from the raw data using custom Perl scripts. Then, we parsed barcodes, so that barcodes were identified, trimmed and the corresponding sample name was added to the identification line of each sequence in the resulting fastq file. We also removed individuals with a deficit or excess of reads (under 65 Mb or above 2Gb, 17 individuals in total) to avoid problems with low quality data and paralogous loci, and to regulate coverage per individual. Next, we created an artificial reference based on a subset of 40 million reads (including reads of all sampling locations) using Seqman Ngen 2.0 (DNAstar, Inc.) to conduct a *de novo* assembly. The *de novo* assembly was formed by the consensus sequences of contigs assuming: i) a minimum coverage depth of 9x, ii) a minimum match of 93% among sequences into the contig and iii) a consensus length range of 84-90 bases in order to avoid paralogs (which would generate long consensus sequences due to sequence variability) and low-quality sequences. Then, we performed a template-guided assembly of all sequences from all individuals using the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009), allowing mismatches of up to 5 bases. We identified and filtered variant sites using custom Perl scripts along with samtools and bcftools (Li et al., 2009). BAM files obtained from BWA were processed through samtools, which computes the probability of each genotype found in the data, and the information was stored in BCF format. The calling of variant sites was carried out by bcftools by means of a Bayesian model, which takes into account the probabilities of genotypes. We considered only SNPs that were present in at least 70% of the individuals and presented a probability <0.01 under the null hypothesis that all the samples were homozygous

for the allele in the consensus sequence. After the identification of SNPs, the variant call format where genotypes probabilities were stored was transformed in a continuous range of values representing genotype uncertainty for following analyses: from 0 (homozygous for the first allele) to 2 (homozygous for the second allele), where 1 represented the heterozygous state and values closer to 0.5 or 1.5 indicated high uncertainty of genotypes. We excluded low frequency variants (minor allele frequency <1%) as they can reflect different demographic and evolutionary histories (Gompert et al., 2014), and chose randomly one SNP per contig in contigs with multiple SNPs to reduce linkage disequilibrium among SNPs. We obtained genotype likelihood data for 4921 SNPs in a final set of 367 individuals, which were improved while estimating ancestry proportions of individuals using a hierarchical Bayesian model (Entropy, Gompert et al., 2014) in a second step (see analysis in Section 3.2.2.1.). These analyses were performed thanks to the collaboration of Dr. Thomas Parchman, in the Department of Biology, University of Nevada, Reno.

3.1.2.2. Local adaptation of S. globulifera at continental scale in Africa

Study sites and molecular markers

In our Study II, part 2, to detect genomic signatures of local adaptation and identify drivers of adaptation in *Symphonia globulifera,* which could shed light on its ability to thrive in many different habitats throughout its large range, we performed several outlier and environmental association tests. Analyses were restricted to the five continental locations from Africa (MB, NK, GB, KR, BN, see Table 5), which included a variety of climatic and soil conditions to facilitate discovering which environmental conditions are leading to local adaptation in our species (see Nadeau, Meirmans, Aitken, Ritland, & Isabel, 2016).

Neotropical locations and São Tomé were not included because the spatial genetic differentiation in strongly differentiated populations may increase false positive rates (Hoban et al., 2016). Moreover, near-absent gene flow between separated regions can lead to a lack of correlation between loci and environmental pressures (Coop, Witonsky, Di Rienzo, & Pritchard, 2010), particularly if different mutations can promote the adaptation to the same environmental conditions (Hoban et al., 2016). No separate analysis was conducted in the Neotropics because of the low number of Neotropical populations sampled (see Foll & Gaggiotti 2008).

To enhance the power of outlier detection we curated the 4921 SNP dataset obtained for the Study II, part1, to reduce the frequency of missing data, reassigning discrete genotype values as above based on a minimum of 75% probability of genotype certainty. In this way, missing SNP calls were ≤15% in all except 2 individuals (17% and 19%) from Africa. We also removed SNPs that were monomorphic in Africa, obtaining a data set with 3399 SNPs in 182 individuals.

Environmental Data Selection

We extracted the 19 climatic variables plus the altitude from WorldClim 1.4 (resolution of 30 arc seconds, dataset for the period 1960–1990, Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), the aridity index (resolution of 30 arc seconds, The CGIAR-CSI Global Aridity Index, Trabucco & Zomer, 2009) and soil data (resolution of 5 arc minutes, GeoNetwork opensource

software, http://www.fao.org/geonetwork/srv/en/main.home, Land and Water Development Division, FAO, Rome) for each individual 's position from continental populations in Africa. Altitude was added as a potential driver of selection since it has been shown to have some influence on the evolutionary processes of *S. globulifera* at local scale (Torroba-Balmori et al., 2017) and it is related to many geophysical drivers (Körner, 2007). Soil data included 17 categorical variables (each categorical level representing ranges of continuous values) and 2 continuous variables (see Supplementary Information S9.3.1.). Soil categorical variables were stored as raster datasets, where each pixel presented two values for each variable: a value related to the 75% (or 60%) of the surface within the grid cell (i.e., the environmental value that occupies the largest area of the pixel), and another value related to the remaining 25% (or 40%, see Jones et al., 2013 for more details on format development). For each pixel selected, we took its two values and performed a weighted average based on the surface occupied by each value to get a unique value per pixel. Finally, to obtain the values for each environmental variable (climatic and soil) in each population, we calculated the average value over individuals in each location (see Table S9.3.1.1. for more details, 37 variables in total).

To assess the influence of our environmental variables on *S. globulifera*, we selected independent variables. Thus, we computed the correlation matrix (Pearson correlations) among climatic and soil variables and visualized it as a hierarchical clustering tree (hclust function, complete linkage clustering, R Core Team, 2014). Each group of correlated variables corresponded to correlations higher than 0.75 (see Serra-Varela et al., 2015). Then, we selected one variable per group prioritizing i) climate variables (better estimations at large scale) over soil variables, ii) annual or seasonal variables over monthly climate variables to obtain more comprehensive variables of local climate and iii) variables of easier interpretation or of easier measurement (and probably more accurate, e.g., pH). See Table 6 and Fig. S9.3.1.1. for final selection.

Table 6. Independent climatic and soil variables selected for the analysis of loci under selection in continental locations in Africa.

Environmental variable	Population				
	BN	**KR**	**MB**	**NK**	**GB**
BIO7	9.73	10.93	10.47	10.44	13.87
BIO11	25.75	24.52	21.59	21.56	23.50
BIO15	77.85	57.40	57.00	56.86	70.80
BIO17	67.50	154.30	174.71	209.20	17.55
BIO18	227.25	477.00	566.27	524.40	710.95
BIO19	314	1218.8	573.76	492.55	17.55
aridity index	0.89	1.84	1.55	1.52	1.22
PH_T	2.55	2.75	2.00	2.00	2.49

3.1.3. Genetic structure within the genus *Symphonia* in Madagascar (Study III)

Study sites and plant material

For our Study III, we sampled 10 locations from Madagascar where *Symphonia* species occur, corresponding to humid tropical rainforest in the east of the island (see Fig. 16, Table 7). Research permits Nº 272/14/MEEF/SG/DGF/DCB.SAP/ SCB, and Nº 67/13/MEF/SG/DGF/DCB.SAP/ SCB were issued by the "Ministère de l'environnement, de l'écologie et des forêts de la République de Madagascar". Sampling involved individuals of all sizes but, preferentially, sub-adult and adult. A balanced representation of gene pools in targeted areas was intended to be achieved through the sampling, and a transect-like sampling design was used for ease of fieldwork. We collected leaf or cambium samples and georeferenced individuals at the same time (sample sizes and locations are given in Table 7). Sampled plant material was dried with silica gel and DNA was extracted using the Invisorb DNA Plant HTS 96 Kit (Stratec Molecular, Germany) or the DNeasy Plant mini Kit (Qiagen, The Netherlands).

Some samples were also preserved in RNAlater (Qiagen) solution for transcriptome sequencing (see Supplementary Information S9.4.1). Additionally, for each putative species, samples including branches with leaves (and flowers and fruits when available) were taken to be deposited as herbarium vouchers at the Royal Botanical Garden Madrid (RJB-CSIC, MAD herbarium) and Antananarivo University herbarium after tentative botanical identification based on Abdul-Salim (2002), Perrier de la Bâthie (1951) and consultation of previously identified herbarium specimens from Missouri Botanical Garden Herbarium.

For the study III, we also included a subset of individuals from the nine locations of *Symphonia globulifera* in Africa and the Neotropics used in Section 3.1.2.1, all located in natural ancient forests (see Table 5). As explained above, three populations were located in Cameroon (Mbikiliki in the Ngovayang massif, Nkong-Mekak in the buffer zone of the Campo Ma'an National park, 48 km apart, and Korup National Park); one in southern Benin (Porto Novo, in the Dahomey Gap), one in south-west of Gabon (Ngounié), one on the island of São Tomé (São Tomé and Principe), one in Ituberá (Brazil), and two in French Guiana (Regina and Paracou, both populations including two morphotypes). For Ituberá, Regina, the African populations São Tomé, Mbikiliki and Nkong Mekak, a transect-like sampling design was used for ease of fieldwork. In Benin, Korup and Gabon, samples were collected randomly in different sampling missions. In Paracou, random sampling was conducted in established plots. Sampling generally involved adult or subadult trees (≥10 cm dbh). As the sampling in Paracou and Regina included individuals from swamp and terra firme morphotypes, the morphotype information was recorded for all samples from French Guiana.

Table 7. Characteristics of sampling locations in *Symphonia* individuals for Study III. Loc. ID: abbreviation for the location name, n: sample size for SNP data, *Spp*.: putative species identified in each location (not all plant specimens collected could be identified).

Continent	Location	Loc. ID	Latitude	Longitude	n	*Spp.*
America	Paracou, French Guiana	PR SP	5.260	-52.924	23	*S. globulifera* (*S. sp.1*)
		PR			23	*S. globulifera*
	Regina, French Guiana	RG SP	4.308	-52.235	11	*S. globulifera* (*S. sp.1*)
		RG			9	*S. globulifera*
	Ituberá, Brazil	IT	-13.795	-39.181	22	*S. globulifera*
Continental Africa	Porto Novo, Benin	BN	6.389	2.623	20	*S. globulifera*
	São Tomé, São Tomé and Principe (ST)	ST	0.280	6.591	22	*S. globulifera*
	Nkong Mekak, Cameroon SW	NM	2.762	10.531	12	*S. globulifera*
	Mbikiliki, Cameroon SW	MB	3.197	10.524	12	*S. globulifera*
	Korup, Cameroon W	KR	5.072	8.836	20	*S. globulifera*
	Ngounié, Gabon	GB	-1.432	10.289	20	*S. globulifera*
Madagascar (Africa)	Amboatoaranana village, Mangoro region	AB	-18.832	48.270	22	*S. eugenioides, S. louvelii*
	Anboasarinala village, Anjozorobe region	AN	-18.460	47.952	14	*S. louvelii, S. microphylla*
	Andasibe-Mantadia National Park	AM	-18.938	48.421	145	*S. clusioides, S. eugenioides, S. fasciculata, S. louvelii, S. nectarifera, S. sessiliflora, S. urophylla*
	Ankazomivady	AK	-20.775	47.183	39	*S. clusioides*
	Ambatondrazaka	AT	-18.067	48.251	1	
	Farankaraina	FR	-15.436	49.843	20	*S. sp.1 (Farankaraina), S. sp.1 (Nosy Mangabe), S. sessiliflora*
	Ialatsara_1	IA1	-21.219	47.383	34	*S. microphylla*
	Ialatsara_2	IA2	-21.066	47.208	9	*S. clusioides, S. nectarifera*
	Nosy Mangabe Island	NMI	-15.494	49.768	84	*S. sp.1 (Farankaraina), S. sp.1 (Nosy Mangabe), S. eugenioides, S. pauciflora, S. sessiliflora, S. urophylla*
	Ranomafana National Park	RN	-21.279	47.426	63	*S. eugenioides, S. nectarifera, S. tanalensis*
	unknown	-	-	-	3	

Molecular markers

In our pipeline, we first identified candidate SNPs from 20 *de novo* transcriptome assemblies (18 Malagasy *Symphonia* accessions and two African *Symphonia globulifera* accessions, see Table S9.4.1.1.). Briefly, a *de novo* assembly of transcriptomes was performed following Seoane et al. (2016; see details in Supplementary Information S9.4.1.). Based on the predicted open reading frames (ORF[2]) from this step, the transcriptome sequences that we called unigenes (i.e., full sequences including ORF and UTR) were aligned with Clustal Omega (Sievers & Higgins, 2014) and then, clustered into groups of putative orthologs (sequence identity ≥ 90%).

[2] ORF is a sequence of nucleotide triplets. Those triplets are read as codons which specify amino acids. The ORF does not contain stop codons.

Based on those unigene clusters, candidate SNPs were selected following two procedures. The first method consisted of a visual inspection of sequence alignments within clusters in order to detect SNPs across individuals. This method was time-consuming but allowed us to discard clusters with likely paralogs. The second method was an automated selection of candidate SNPs, applying filtering criteria on the SNPs previously detected by an automated SNP calling step performed on the groups of putative orthologs (see details in Supplementary Information S9.4.1). This was a time-efficient method but with a higher risk of conserving paralogs. Both methods allowed us to select SNPs randomly from the genome, including loci with common or rare alleles. We proceeded as described below:

a) Visual identification of candidate SNPs based on the inspection of alignments of unigene clusters. For this procedure, we selected clusters which contained from 6 to 20 *Symphonia* accessions and, then, we aligned sequences across individuals within each cluster using MAFFT (Katoh, 2002). By visual inspection in PhyDE (Müller, Quandt, Müller, & Neinhuis, 2006), we detected that alignments with more than 18 accessions often showed two groups of differentiated sequences (putative paralogs), whereas SNPs were very scarce in alignments up to seven individuals. Thus, we decided to keep only those alignments which included sequences from ten accessions and then, to select SNPs with the same alternative allele present in at least two accessions. To avoid linked SNPs and to comply with experimental design requirements of primers for the SNP typing using the Sequenom iPLEX^{TM} MassARRAY® technology, we selected a maximum of two visually non-redundant SNPs per alignment. Then, we extracted those sequences including the SNP and its flanking regions (flanking regions with a minimum length of 60 bp each, only one extra SNP represented by IUPAC ambiguous DNA codes on those flanking regions was allowed). Based on this method, we chose 233 candidate SNPs for the screening step.

b) Automated identification of candidate SNPs. After the *de novo* assembly of transcriptomes and the clustering of sequences into groups of putative orthologs, an automated SNP calling was performed (see details in Supplementary Information S9.4.1). Briefly, i) clusters with ten different accessions were selected, ii) a reference accession, MH2809, was chosen as the one with the highest representation across clusters, and iii) unigenes of each accession were mapped against the reference. Then, the SNPs calling was performed with VarScan2 (Koboldt et al., 2012) using default parameters. Subsequently, we applied automated filters on the vcf file (i.e., the output from the previous SNP calling step), based on the following criteria: i) minimum distance between SNPs: 24 bp, ii) depth $\geq$ 14 reads, iii) genotypes had to present the alternative allele in three accessions minimum, iv) first and last SNPs of unigenes excluded (to avoid short flanking regions and low-quality sequences), and iv) we allowed a maximum of two SNPs per alignment. As a result, 617 candidate SNPs fulfilled the criteria, and we extracted the SNPs with flanking regions (a length of 100 bp each) based on the reference transcriptome sequences of MH2809. For the screening step, we selected 252 candidate SNPs based on this method, allowing a maximum of 2 SNPs per cluster.

For the SNP screening and genotyping steps, the Sequenom iPLEX^{TM} MassARRAY® technology at the INRA Genomics-Transcriptomics Facility (PGTB) in Bordeaux (France) was used (see Oeth et al. 2007 and Bradic, Costa, & Chelo, 2011 for further details). The screening step was designed to identify the SNPs with the best performance to be included in the final

SNP set to genotype. Eight independent multiplexes of SNPs were designed with the software MassARRAY® Assay Design 3.1 Software (available for the Sequenom platform) using as input the selected candidate SNPs from the visual and the automated identification steps. Each multiplex contained 40 SNPs (assay design settings: minimum mass separation among extension primers in Daltons (Da): 20 Da; lower and upper limits for the set of extension primers: 3000-10000 Da; maximum multiplex level: 40). Thus, we tested 160 SNPs from each SNP selection strategy (320 in total) on a testing set of 95 individuals (16 individuals of *S. globulifera* and 79 individuals from Malagasy *Symphonia spp.*). The selection of those Malagasy individuals for the screening step was based on *a priori* genotyping of the complete individual Malagasy dataset (i.e., our 434 Malagasy individuals) using 20 polymorphic SSR markers developed in Olsson et al. (2017), genetic structure analysis using STRUCTURE, and gene pool assignment based on ancestry proportions (Q) ≥ 0.5 (see Supplementary Information S9.4.4. for more details). As we found five differentiated gene pools based on SSR markers (hereafter rGP; Fig. S9.4.4.1.), we included Malagasy individuals from the five rGPs detected in our testing set to be able to test and *a priori* select those SNPs amplifying in all genetic groups present in our Malagasy individuals.

Based on the result of the multiplexes, we selected the best set of SNPs using the MassARRAY® Analyzer 4 software from Sequenom, which provided plots of SNPs based on magnitude (a value based on the signal intensity detected by the mass spectrometer) and angle (a quantity derived from mass signals that is used for genotype identification) values. The criteria for a successful SNP were: > 70% of genotyped individuals with a magnitude > 4.5 and 2-3 visually differentiated clusters of genotypes (based on different values of angle) which included three patterns of allelic frequencies in the set of screened individuals: i) presence of the three expected genotypes, ii) presence of only both homozygous genotypes, or iii) presence of rare alleles (i.e., most individuals presenting the same homozygous genotype and few individuals as heterozygotes as well as alternative homozygotes, see Fig. 4). The screening resulted in 97 successful SNPs based on the visual SNP identification and 80 successful SNPs based on the automated selection method. Those 177 SNPs were used to design four multiplexes of 36-40 SNPs for the genotyping step (assay design settings: minimum mass separation among extension primers: 25 Da, lower and upper limits for the set of extension primers: 3000-10000 Da, maximum multiplex level: 40). As a result, we used a final set of 156 SNPs (81 and 75 SNPs from visual and automated selection respectively) to genotype 630 individuals (196 individuals of *S. globulifera* and 434 individuals from different gene pools of Malagasy *Symphonia* based on SSR markers, see Table 7 and Table 8), and genotypes were inferred from the masses of products obtained by the MassARRAY® mass spectrometer (MALDI-TOF).

Table 8. Number of successful selected SNPs regarding each method used for selection of candidate SNPs and each step of the workflow.

Selection method	candidate SNPs	screening design	screening success	genotyping design	genotyping success
Visual	233	160	97	81	75
Automated	252	160	80	75	69
Total	485	320	177	156	144

The Sequenom raw output data was processed using VIClust 1.1 (Garnier-Géré, Harmand, Laizet, & Mariette, 2014) to obtain the values for magnitude and angle for each SNP and individual. Those values were plotted using custom R scripts, and visual calling of genotypes was performed based on patterns of visually differentiated genotype clusters (i.e., different values of angle, genotypes were considered null when magnitude < 3). Such a pipeline was needed because the MassARRAY® Analyzer 4 software and VIClust 1.1 only perform automated genotype clustering and calling on diploid SNPs. However, during the SNP calling, we detected a subset of individuals which presented signals for more than three clusters for a subset of SNPs (i.e., certain individuals were putative tetraploids, with more than three genotypes). To strengthen the confidence in the ploidy inference, we verified that this subset of individuals corresponded exactly to a certain gene pool based on SSR markers: rGP1 (Fig. S9.4.4.1.**)**. Thus, we genotyped this subset of individuals for those SNPs as putative tetraploids and excluded putative polyploidy from the rest of individuals (see Fig. 4).

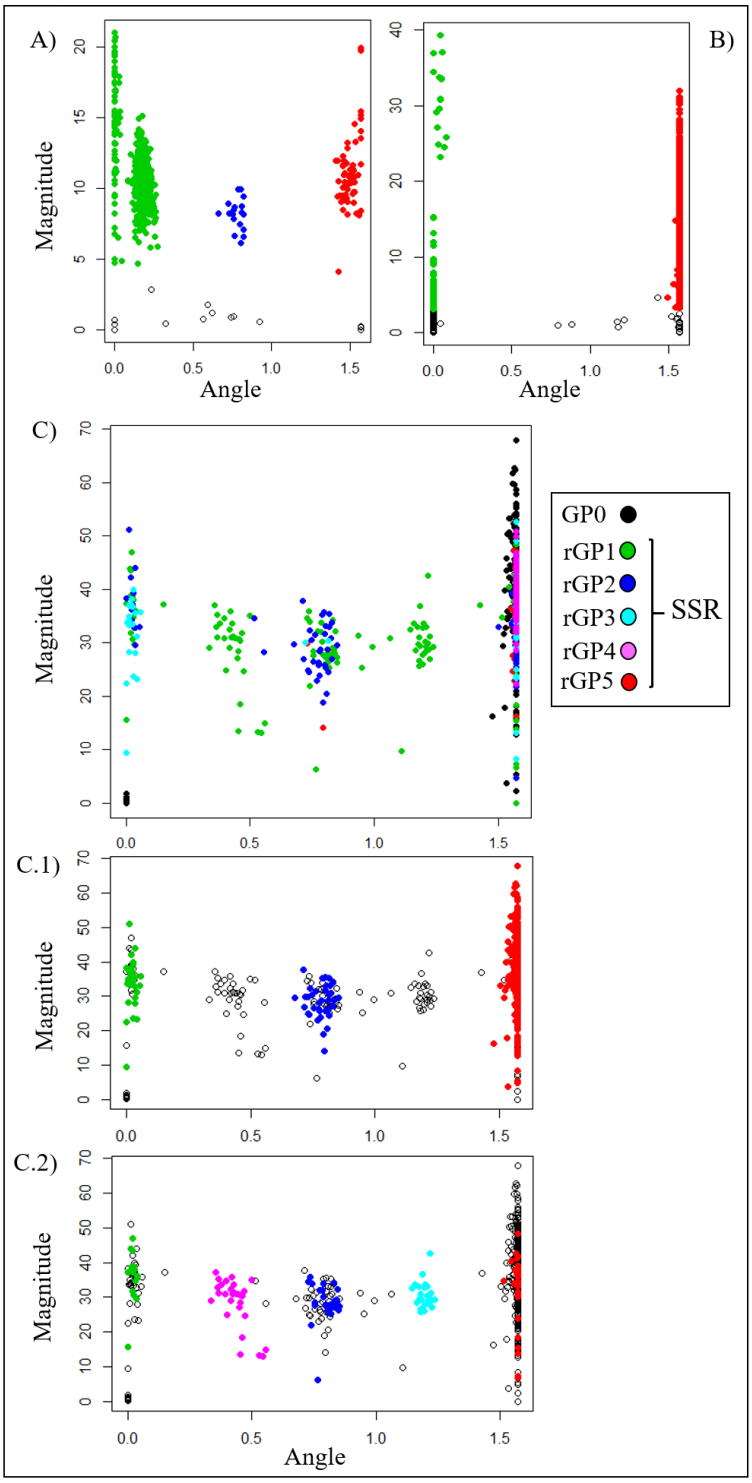


Figure 4. Plots of magnitude and angle values for each SNP and individual after Sequenom genotyping. A) Example of SNP showing the the three expected genotypes. B) Example of SNP showing only both homozygous genotypes. C) Example of SNP showing signals for five clusters (i.e., putative tetraploids) where colours represent the gene pool assignment based on SSRs (i.e., rGP) in Malagasy *Symphonia spp*. GP0: S. *globulifera* individuals (in black), rGP1: putative tetraploids (green). C.1) Diploid individuals (rGP0, rGP2, rGP3, rGP4, rGP5), showing the three expected genotypes, each one in a different colour. C.2) Putative tetraploids (rGP1) showing a pattern of five genotypes, each one in a different colour.

3.2. Data analysis

3.2.1. Fine-scale spatial genetic structure in *S. globulifera* (Study I)

Genetic diversity

The rarefied allelic richness (A_R) and the expected heterozygosity corrected for sample size (H_E) were computed for nuclear SSRs using SPAGeDi 1.4c (Hardy & Vekemans, 2002); the standard error of H_E was estimated from jackknife replicates using the PopGenKit package (Paquette, 2015) in *R* v. 3.1.1 (R Core Team, 2014). To assess deviations from Hardy-Weinberg genotypic proportions, e.g., caused by non-random mating or null alleles, we computed the fixation index (F_{IS}) and tested deviation from zero using 10,000 permutations of alleles within populations in SPAGeDi. We estimated the frequencies of possible null alleles using the Brookfield2 estimator (Brookfield, 1996; Girard & Angers, 2008) and estimated a fixation index corrected for null alleles in MicroChecker v. 2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004). Occasionally, multilocus genotypes were found more than once within populations. The probability for these copies to be derived from distinct sexual reproductive events, *psex*, was computed in GenClone (Arnaud-Haond & Belkhir, 2007). Plastid haplotypes were defined combining nucleotide polymorphisms, indels and inversions in the sequence. Rarefied plastid haplotype richness, A_{Rp}, and haplotypic diversity, *h*, were obtained in SPAGeDi 1.4c (Hardy & Vekemans, 2002).

Test and quantification of FSGS

To test for the presence of overall FSGS in each population and quantify its strength, we followed the approach of Vekemans and Hardy (Vekemans & Hardy, 2004) for nuclear and plastid DNA markers separately. Pairwise kinship coefficients F_{ij} (Loiselle, Sork, Nason, & Graham, 1995) were calculated in SPAGeDi 1.4c in all populations and were regressed on the logarithm of pairwise spatial distances between individuals. The significance of the regression slope *b* was tested using 10,000 permutations of the spatial position of the individuals. The strength of FSGS was estimated as $Sp = -b/(1-F_{ij(1)})$ (Vekemans & Hardy, 2004) where $F_{ij(1)}$ is the average kinship coefficient of individuals in the first distance class. The number and size of distance classes was defined for each population according to recommendations from the SPAGeDi user manual (Hardy & Vekemans, 2013): similar numbers of pairwise comparisons across classes, > 50% of individuals present in each class and a coefficient of variation < 1 of the number of times each individual was represented in each class.

As an alternative to *Sp*, we also performed a spatial principal component analysis (sPCA) in the *adegenet* package in *R* (Jombart, 2008). This allowed us to test for overall FSGS (e.g., patches of related individuals, allele frequency gradients) using a G-test, and to estimate the strength of spatial structure as the eigenvalue of the first sPCA axis, *eig.sPCA*. Since we observed *a priori* a stronger FSGS in African than Neotropical populations, we assessed differences in FSGS between continents using T-tests based on either *Sp* values or *eig.sPCA*. To specifically address the relationship between FSGS and altitudinal sampling range within populations, we performed a Spearman rank correlation test between *Sp* or *eig.sPCA* and the standard deviation of sampling altitude using *R* (see also next section).

Spatial genetic heterogeneity and its causes

Besides FSGS due to drift-dispersal equilibrium, non-equilibrium processes such as selection or barriers to reproduction can lead to spatial genetic heterogeneity (SGH). We tested for SGH using two types of approaches based on nuclear SSRs, and then investigated its potential causes through examining its strength, congruence with SGH at maternally inherited plastid markers and specific spatial arrangement. For SSRs, we used: 1) a G-test to detect global structure (see above) and an L-test to detect local structure, the latter corresponding to an increased differentiation between spatially close individuals, and estimated sPCA scores [-1,1] for each individual on the first global or local sPCA axes, respectively (Jombart, Devillard, Dufour, & Pontier, 2008); 2) the Bayesian clustering analysis implemented in STRUCTURE v. 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to detect sympatric gene pools (GPs) and estimate ancestry proportions (*Q*, [0,1]) for each individual in each GP. STRUCTURE was chosen because it is particularly efficient at detecting GPs that co-occur in the same geographical site when spatial structure is weak (Chen, Durand, Forbes, & François, 2007) and because it allows to test whether a model with differentiated GPs ($K>1$) fits the data better than a model with a single GP ($K=1$). To detect shared GPs among populations, STRUCTURE was run separately for the groups of populations that were genotyped together (to avoid mixing SSRs datasets for which allele identities were not cross-standardised: Ituberá and African populations; Yasuní and BCI). Then the analysis was run within each population. We used an admixture model with correlated allele frequencies for codominant markers and 10 repetitions for each number of clusters, *K*, from 1 to 7, using a burn-in length of 20,000 and a run length of 80,000 iterations. Chain convergence was checked visually. The *K* that best described the data was determined as the one with the highest logarithm probability of data, ln Pr(X|*K*) (also referred to as L(*K*)), following Pritchard, Wen, & Falush (2010), and using the Delta *K* (ΔK) method described by Evanno, Regnaut, & Goudet (2005) (see Supplementary Information S9.1.2.). Since the codominant markers model is not necessarily robust to the effect of null alleles, we repeated all analyses using the recessive alleles model as explained in the STRUCTURE documentation to assess the effect of null alleles on the clustering solution. Also, for sites with $K > 1$, analyses were repeated using an admixture model in the spatially explicit clustering program TESS v. 2.3.1. (Chen, Durand, Forbes, & François, 2007) using either a constant or a linear trend, and for the latter, a spatial interaction parameter of 0.6 or 1 (Supplementary Information S9.1.2.). Maps of the geographical distribution of GPs and haplotypes per sampling location were built in QGIS v. 2.4 (QGIS Development Team, 2014).

If within-population SGH occurs, its strength, its association with plastid DNA and its spatial arrangement, e.g., in the context of habitat variation, can provide further information on its ecological and evolutionary determinants. As measures of SGH, we used the individual sPCA scores and GP ancestry proportions (*Q*) computed above. First, to examine the strength of SGH, we assessed *Q* values and genetic differentiation (F_{ST}) among co-occurring GPs, considering higher *Q* and higher F_{ST} as indicators of stronger divergence. F_{ST} among GPs within populations was estimated and tested with permutation tests using SPAGeDi, assigning individuals to GPs based on $Q\geq50\%$ or $Q\geq87.5\%$, the latter category susceptible to include genetically pure, first and later-generation backcrosses (Guichoux et al., 2013). Within GPs based on $Q\geq87.5\%$, e.g., representing putative distinct reproductive demes, we assessed the

presence of null alleles, deviation from Hardy-Weinberg proportions and strength of FSGS as above. Second, to test for historically diverged lineages, we assessed cyto-nuclear disequilibria (Fields, McCauley, McAssey, & Taylor, 2014) between plastid haplotypes and nuclear SGH. We performed one-way ANOVA in *R* to detect if groups of individuals carrying the same haplotype differed in their mean individual sPCA scores. We also tested for association of haplotypes and GPs ($Q \geq 50\%$) using Fisher tests that ignored spatial autocorrelation in the data and we performed Partial Mantel Tests based on similarity matrices ("1" for pairs of individuals sharing the same haplotype or GP, "0" for pairs with different haplotypes or GPs) in which we controlled for spatial autocorrelation through a spatial distance matrix using the Ecodist package (Goslee & Urban, 2007) in *R*. In Paracou, we further tested the association between GPs or haplotypes and morphotypes, similarly using Fisher and Partial Mantel Tests. Third, we examined altitudinal stratification of genetic variation in each population as would be expected, for example, in the case of restricted mobility of dispersers due to slopes (Giordano, Ridenhour, & Storfer, 2007; Pérez-Espona et al., 2008). We performed one-way ANOVA to test if three *ad hoc* defined altitudinal classes differed in their mean individual sPCA scores, and in their mean *Q* value for each GP. Association with altitudinal classes was preferred over a regression analysis because the relationship to test is not necessarily linear.

3.2.2. Large-scale genetic structure in *S. globulifera*, demographic history and adaptive evolution (Study II)

3.2.2.1. Spatial genetic structure of S. globulifera across continents (Study II, part I)

Inference of improved genotypes and gene pool delimitation

We first used a hierarchical Bayesian model (Entropy, Gompert et al. 2014) to infer improved genotypes as well as ancestry proportions of individuals in putative ancestral clusters and cluster allele frequencies in a manner similar to STRUCTURE (Pritchard, Stephens, & Donnelly, 2000; Falush, Stephens, & Pritchard, 2003). Similarly to STRUCTURE, the program only requires individual genotypes and the assumed number of ancestral clusters (*K*) to consider in the model. Entropy uses genotype likelihoods from bcftools as a basis to estimate the probability of each genotype at each locus (Li, 2011). Thus, the model provides updated genotype probabilities along with other model parameters such as proportion of ancestry (Mandeville, Parchman, McDonald, & Buerkle, 2015). In addition, Entropy calculates the deviance information criterion (DIC) as a metric for comparing models with different numbers of ancestral population clusters (see Gompert et al. 2014; Mandeville, Parchman, McDonald, & Buerkle, 2015, for further explanation).

To facilitate the convergence of MCMC chains in Entropy, we initialized admixture proportions using probabilities of cluster membership. These probabilities were calculated first performing a principal component analysis (PCA) in R (prcomp, R Core Team 2014) with the raw genotype likelihoods from bcftools. Then, we used *K*-means clustering of the principal component scores to define *a priori* groups and, finally, we performed a DAPC (Discriminant Analysis of PC) of PC scores based on the best supported number of clusters from the previous step, with the lda package in R following Jombart, Devillard, & Balloux, (2010).

We ran 4 independent 100,000 MCMC step chains of Entropy and discarded the first 40,000 values as burn-in, with a thinning interval of 10, with *K* values (number of source populations) from 2 to 17. Testing for different *K*-values was performed in spite of having samples from nine locations, in order to check possible within-population clustering as detected previously using microsatellites and chloroplast sequences (Torroba-Balmori et al., 2017). DIC values of models with different *K* were compared, the model with the lowest DIC was chosen as the best (*K*=9) and individuals were assigned to the nine GPs based on a minimum of 54% assignment probability.

Thus, for the subsequent analyses we considered all individuals sampled in IT, ST, KR, BN as corresponding to a specific GP (and sampling location). Moreover, three GPs were found in French Guiana, two corresponded to *S. globulifera* in PR and RG and one corresponded to *S. spp.1* (hereafter SP). All individuals from Mbikiliki and Nkong Mekak, except three, fell into the same gene pool and were thus pooled (GP named "MN") for further analysis. Gabon was an admixed location, but it was also considered a distinct gene pool in order to disentangle its demographic history. And finally, we excluded a gene pool with only three individuals (origin: MN) as that low sample size was not enough for accurate analyses (see Results in Section 4.2.1.).

In addition, the improved genotypes obtained with Entropy were employed for subsequent analysis. As genotype probabilities from Entropy were coded in a continuous range of values representing genotype uncertainty, and all subsequent analyses did not allow for such genotype uncertainty, we reassigned the continuous values obtained to discrete genotypes (i.e., homozygous: 0 or 2; heterozygous: 1) if they lay within 0.1 of the closest discrete value. Otherwise, SNPs were considered missing data. Therefore, we obtained genotypes with a minimum of 90% probability of being correct. That way, the follow-up analyses were based on genotype data with high certainty for 4921 SNPs in a final set of 367 individuals (see results in Section 4.2.1.)

Genetic diversity

Using Arlequin 3.5.2. and the R package hierfstat (Excoffier & Lischer 2010; Goudet & Jombart, 2015), we calculated the observed and expected heterozygosity, the number of polymorphic loci and the inbreeding coefficient (10,000 bootstraps) in a random subset of 20 individuals in each gene pool (GP) from Entropy (exceptionally, we took all individuals available in GPs with less than 20 samples) to correct the estimates for uneven sample size. We also calculated the number of polymorphic loci per GP (all individuals considered).

To characterize patterns of genetic distances among gene pools we calculated an unrooted neighbour-joining tree (Saitou & Nei, 1987) based on Nei's standard genetic distance D (Nei's D; Nei, 1972) using the Seqboot, gendist, neighbor and consense options (1,000 bootstrapping, majority rule consensus tree) in PHYLIP (Felstein, 1991). Missing data in population allele frequencies was not allowed in the analysis, so the analysis was based on 4814 out of 4921 SNPs. The final tree was visualized using the APE package (Paradis, Claude, & Strimmer, 2004) in R using the genetic distance matrix among the gene pools.

Phylogeographic history

We inferred the phylogeny of our gene pools using two approaches. First, we used SNAPP 2.4.0. (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012) to infer the "species tree" of our gene pools. This coalescence-based method can integrate all possible gene trees in a single species tree directly from unlinked biallelic markers, using a Markov chain Monte Carlo sampler. It also estimates relative divergence times and population sizes from the root and branches of the tree. Analyses were done under the assumption of no gene flow between lineages. As SNAPP is computationally intensive and the analysis needs at least two sequences per gene pool (Drummond & Bouckaert, 2015), we randomly selected three individuals per gene pool (six sequences) to get representative allele frequencies from populations (21 individuals in total, missing data: <14% in 19 ind., ~19% in 2 ind. from different GPs).

Following Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, (2012) and Drummond & Bouckaert (2015), we tested four groups of priors under alternative hypotheses to detect if estimates were informed by the data and not just sampled from the priors. The priors for alternative models (M) were defined based on *Symphonia globulifera* studies: For all models, we assumed a generation time of 100 years (Budde, González-Martínez, Hardy, & Heuertz, 2013; Jones, Cerón-Souza, Hardesty, & Dick, 2013), mutation rate (μ) of 10^{-9} per site per generation (Ossowski et al., 2010) and a divergence time for *S. globulifera* populations between Africa and America of 17.36 Ma (Dick, Abdul-Salim, & Bermingham, 2003) as expectation for the root height of the tree to calculate the birth rate lambda (λ=10536) in the Yule prior (see further information about the Yule model in Drummond & Bouckaert 2015).

For the first three models we set weakly informative priors. The forth model was intended to assess the effect of parameter specifications under which SNAPP is usually run (e.g., Harvey & Brumfield 2014, Bell, Drewes, & Zamudio, 2015) and also included a larger population size prior (mean population mutation rate, Θ). Briefly, the model settings were: **M1)** gamma distribution with alpha (α) = 2 and beta (β) = 25,000 (uniform β hyperprior with limits: 0.04774963 – 56100922; mean Θ = 0.00008), lambda hyperprior = 1/X; **M2)** uniform distribution for Θ (limits: 3.56 x 10^{-8} – 41.88), lambda hyperprior = 1/X; **M3)** uniform distribution for Θ (limits: 3.56 x 10^{-8} – 41.88), lambda uniform hyperprior (limits: 4064 – 12193); **M4)** gamma distribution with α = 2, β = 200 (no hyperpriors, mean Θ = 0.01), lambda hyperprior = 1/X. Besides, we used the birth rate lambda (λ = 10536) for the Yule prior in models M1, M2 and M4. For extended information about selection of parameters for the different models see Supplementary Information S9.2.2.

For each model, we conducted 15 runs. For each run, we retained 1 iteration every 1,000 to generate chains of 2 million iterations. Then, we examined the effective sample size (it should be above 200 to indicate convergence) and traced plots to determine the burn-in, and we assessed MCMC convergence using Tracer v.1.6 (Rambaut, Suchard, Xie, & Drummond, 2014). Then, we combined runs using LogCombiner 2.4.0. in BEAST (Bouckaert et al., 2014) generating one chain of 19,000 iterations minimum for each model. Finally, we summarized the distribution of topologies from the different models with TreeSetAnalyser package in SNAPP, visualized the distribution of species tree topologies and branch lengths using DensiTree (Bouckaert, 2010) and obtained the posterior probability of consensus tree nodes

using TreeAnnotator (Target tree type: Maximum Clade Credibility Tree; Rambaut & Drummond, 2016).

As in SNAPP we only included 3 individuals per GP to infer the phylogeny, we also performed the maximum likelihood approach implemented in TreeMix v. 1.13 (Pickrell & Pritchard, 2012), which is also suitable for data exploration (Beichman et al., 2018). TreeMix models the genetic drift between groups of individuals, represents the topology of relationships using genome-wide allele frequency data and links the groups to their most common ancestor. Optionally, it can also model migration events among populations when populations do not fit well with the bifurcating tree model, for example if a sampled population has origins in more than one source population. All analyses were performed under the assumption that all sites were independent. Standard errors were calculated with the -se option and we rooted the tree with the MN gene pool. This choice was motivated by the fact that African populations are older than those from the Neotropics (Dick, Abdul-Salim, & Bermingham, 2003), Nei's D indicated that MN was the African GP with the highest genetic distance from Neotropical GPs and high genetic diversity (*He* and number of polymorphic sites) suggested that the MN gene pool could be more ancient than other African GPs. To assess the stability of the tree topology, we generated 100 bootstrap replicates by resampling blocks of 100 SNPs. Trees and residuals were visualized using scripts in R provided by the software.

As we will explained more detailed in Section 3.2.2.2., we also obtained a hierarchical cluster tree of mainland populations in Africa based on the correlation matrix obtained using a covariate-free approach on BayPass (Gautier, 2015), performed to detect differentiation-based outlier loci putatively under selection among populations.

3.2.2.2. Local adaptation of S. globulifera at continental scale in Africa (Study II, part 2)

Outlier tests

We used three Bayesian approaches (BayeScan, BayeScEnv and BayPass; Foll & Gaggiotti, 2008; de Villemereuil & Gaggiotti, 2015; Gautier, 2015), to detect differentiation-based outlier loci putatively under selection and to test hypotheses about specific drivers of local adaptation. These statistical methods correct for population structure based on different working hypotheses and allow to detect loci with high differentiation in the genome (BayeScan, BayPass), to test the association between allele frequencies and environmental variables along populations (BayPass), or both (BayeScEnv).

BayeScan v. 2.1 (Foll & Gaggiotti, 2008) was used to identify candidate loci putatively under natural selection, as it presents lower type I and type II error rates compared with other methods and it is robust for different demographic scenarios, including the isolation-by distance (IBD) model (Foll & Gaggiotti, 2008; Narum & Hess, 2011). The model, based on a multinomial-Dirichlet distribution, considers a number of populations which have split from an ancestral population and their differences with the common gene pool are measured by subpopulation F_{ST} values. The method uses a logistic regression, which decomposes F_{ST} values into a population-specific component shared by all loci (β) and a locus specific component shared by all populations (α). When the model needs the locus-specific component at a given locus to

explain the observed pattern of diversity, the locus is considered to depart from neutrality. Positive α values suggest diversifying selection, whereas negative values suggest balancing selection. The method also uses short successive pilot runs to choose the proposal distribution to initiate the analysis. The analysis was performed with default settings: i) analysis preceded by 20 pilot runs before starting the calculation to choose the proposal distribution for the analysis, length of each pilot run: 5000; ii) Resulting total number of iterations in the main analysis (pilot runs not included): 100,000, burn-in applied on the total of iterations: 50,000, thinning interval: 10, final sample size: 5000. Prior odds were set to 100 to account for the relatively high number of loci. All loci showing q-values (a false-discovery rate analogue to the p-value) lower than 0.01 on the parameter α were considered outliers, which implied that 1 % of the outliers were expected to be false positives. For each outlier locus, the probability of being putatively under selection was inferred using the Posterior Odds (PO), which indicates the likelihood of the model with selection in comparison to the neutral model, and its significance was assessed using the Jeffreys' scale of evidence (Jeffreys, 1961).

BayeScEnv (de Villemereuil & Gaggiotti, 2015) is a genome-scan software based on the *F* model behind BayeScan to account for population structure and it is also robust to different demographic scenarios. Additionally, it allows us to test for association between allele frequencies and continuous environmental variables and to reduce the false positive rate (de Villemereuil & Gaggiotti, 2015). The model considers three types of effects on loci: demography, locus-specific effects due to local adaptation caused by particular environmental variables and locus-specific effects due to other causes (i.e., unknown environmental variables, allele surfing, background selection, etc.). The model decomposes F_{ST} values into the α and β components (previously explained) and the *g* component, which considers the impact of an environmental differentiation on a given locus (de Villemereuil & Gaggiotti, 2015). For our analysis, we standardized our environmental variables and computed the environmental differentiation of each population regarding the mean for the eight environmental variables (i.e., covariates) selected (Table 6). Then, we ran separate BayeScEnv analysis for each environmental variable. The analysis was performed with default settings for most parameters: i) analysis preceded by 20 pilot runs before starting the calculation before starting the calculation to choose the proposal distribution for the analysis, length of each pilot run: 5000; ii) Resulting total number of iterations in the main analysis (pilot runs not included): 150,000; burn-in applied on the total of iterations: 100,000; thinning interval: 10, sample size: 5000, prior preference for the locus-specific model (*P*) = 0.5. We set the prior probability for non-neutral models (*Pi*) = 0.01, which was equivalent to prior odds = 100 in BayeScan. In this analysis, all loci showing q-values lower than 0.05 on the parameter *g* were considered candidate loci for being under the influence of selection, which implied that 5 % of these outliers were expected to be false positives.

BayPass accounts for the hierarchical population structure and the sampling noise based on the population allele frequencies and removes that covariance among populations before performing the genome scan for signals of positive selection (Gautier, 2015). This kind of method may present similar or higher power compared to genome scans based on F_{ST} (Lotterhos & Whitlock, 2014). We performed two complementary approaches with this software: a covariate-free approach and a covariate approach.

For the covariate-free approach, we ran the core model (default parameters) which computed the covariance matrix of population allele frequencies (Ω) and estimated the XtX statistics (an estimate analogous to F_{ST} but accounting for the variance-covariance structure of populations involved, Günther & Coop, 2013). We compared the observed XtX values to the 99% quantile of XtX values from a simulated pseudo-observed dataset (based on 3399 simulated SNPs, i.e., with the same sample size as the original dataset) which is the default threshold used to discriminate between outlier and neutral SNPs in BayPass. The correlation matrix based on the Ω matrix was visualized as a hierarchical cluster tree using R (Hclust function, R Core Team, 2014; see Fig. S9.3.2.1.) to be compared to the Inferred Maximum likelihood tree of populations from TreeMix, also based on an interpretation of the Ω matrix (Pickrell & Pritchard, 2012), and to the unrooted neighbour-joining tree of Nei's genetic distance, both obtained through analyses detailed in the previous section.

For the covariate approach, aimed to detect significant correlations between environmental variables and allele frequencies, we ran BayPass using the standard covariate model with default options. The default options used an importance sampling algorithm (IS, Coop, Witonsky, Di Rienzo, & Pritchard, 2010) which allows the computation of Bayes Factors to compare the model with association against the null model (Gautier, 2015). This model captures only linear relationships between allele frequency differences and the covariable (Gautier, 2015). Significance of the association of each SNP with covariates was evaluated by means of Bayes Factors (BFs, expressed in deciban (dB) units via the Transformation $10 \times \log_{10}(BF)$) using the Jeffreys' scale of evidence (Jeffreys, 1961) only when the empirical Bayesian p-value (eBPis) was above three (support in favour of a non-null regression coefficient).

Finally, when the same significant locus was related to the same covariate in both covariate-approach analyses, we selected those environmental variables detected as significant putative drivers of adaptive evolution. To discard the possibility of spurious outliers due to the coincidence of extreme values of environmental variables and extreme neutral allele frequencies in the same population (Günther & Coop, 2013), we performed Mantel tests between the correlation matrix derived from Ω (BayPass) and each of the matrices of pairwise environmental distances between populations using the Ecodist package (Goslee & Urban, 2007). Also, we visually compared the values of those pairs of matrices (see Table S9.3.2.3.).

Gene Annotation

The loci identified as outliers with at least two different methods were subjected to an attempted annotation through BlastX searches (i.e., searching protein databases using a translated nucleotide query; Camacho et al., 2009) against NCBI´s (National Center for Biotechnology Information) non-redundant nucleotide (nr) database as well as Swiss-Prot and TrEMBL, the UniProt Knowledgebase (UniProtKB) protein databases. In addition, local Blast searches of the rather short flanking sequences (up to 90 bp) were performed against the *Symphonia globulifera* genome (Olsson et al. 2017). The local Blast hit results were converted into bed coordinates with blast2bed (available from https://github.com/nterhoeven/blast2bed) and a maximum of an additional 500 bp before and after the locus of interest were extracted with BEDTools 2.29.2 (Quinlan & Hall, 2010). BlastX searches with these extended sequences were performed against the three databases as described above.

3.2.3. Genetic structure within the genus *Symphonia* in Madagascar (Study III)

Genome size estimation and ploidy level inference using flow cytometry

To confirm that Malagasy individuals genotyped as putative tetraploids were true tetraploids, the nuclear DNA content of 37 individuals (40 samples) from the five differentiated gene pools based on SSR markers (hereafter rGPs; see Supplementary Information S9.4.4.) in Malagasy *Symphonia* was estimated using flow cytometry (FCM). Five individuals from each rGP (ancestry proportions (Q) $\geq$ 0.78) were chosen including both desiccated leaf and cambium tissues within each rGP. We also tested the suitability of tissues preserved in RNAlater for FCM analyses and, therefore, we selected samples from each rGP collected during the field sampling with tissues preserved in RNAlater solution.

Nuclear suspensions were obtained following Galbraith et al., (1983) by chopping tissue of the studied species and fresh leaf tissue of *Pisum sativum* 'Ctirad' (internal reference standard) in 1 ml of WPB buffer (Loureiro, Rodriguez, Dolezel, & Santos, 2007). The nuclear suspension was then filtered using a 50 µm nylon mesh and 50 $\mu g.ml^{-1}$ of propidium iodide (PI, Fluka, Buchs, Switzerland) and 50 $\mu g.ml^{-1}$ of RNAse (Fluka, Buchs, Switzerland) were added. Samples were analyzed in a Partec CyFlow Space flow cytometer (Partec GmbH., Görlitz, Germany; 532 nm green solid-state laser, operating at 30 mW) and results were acquired using Partec FloMax software v2.4d (Partec GmbH, Münster, Germany). The 1C genome size in Mbp was obtained by the ratio between G_1 mean peaks of *Symphonia* and *P. sativum*, multiplied by the genome size of the reference standard (2C = 8908 Mbp; Doležel et al., 1998), and further divided by two. All analyses were performed in the laboratory of Dr. João Loureiro, in the Centre for Functional Ecology, Department of Life Sciences, University of Coimbra. Samples whose nuclear DNA content was approximately double that of most samples were flagged as tetraploid.

Gene pool delimitation and phylogenetic relationships

Once the polyploidy of the individuals assigned to rGP1 was confirmed by FCM (see results in Section 4.3.), the software STRUCTURE v. 2.3.4 (Pritchard et al., 2000) was used to analyze the full SNP dataset (144 SNPs) and delimit gene pools (hereafter nGP) within the genus *Symphonia*, with the aim to contrast the genetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species) with the Malagasy *Symphonia* genetic structure discovered through the SSR analysis (see Supplementary Information S9.4.4.). This software, as explained in the Section 3.2.1., is a Bayesian clustering analysis that infers GPs andestimates ancestry proportions (Q, [0,1]) for each individual in each GP. It performs efficient analysis even if the genetic structure is weak (Chen, Durand, Forbes, & François, 2007) and also allows to compare the performance of models with differentiated GPs including the model with a single GP (K=1). STRUCTURE analysis was carried out using an admixture model with correlated allele frequencies for codominant markers and for clusters, K, from 1 to 20 (burn-in length: 50,000, run length: 100,000 iterations, 10 repetitions for each number of clusters, chain convergence visually checked). The best K was inferred based on the highest logarithm posterior probability of data (L(K)) and Delta K (ΔK), following Pritchard, Wen, & Falush (2010) and Evanno, Regnaut, & Goudet (2005). Then, to test for hierarchical genetic structure, i.e., sub-structure within the inferred GPs, the same analysis was run within each gene

pool discovered (for clusters, *K*, from 1 to 10; Evanno, Regnaut, & Goudet, 2005). As these analyses included both diploid and polyploid SNPs, we scored all SNPs as tetraploid genotypes (i.e., using four characters) indicating both states: diploid (e.g., AA99) or tetraploid patterns (e.g., AGGG), and missing data (9999, see Supplementary Information S9.4.2.). The only exception for this format was the STRUCTURE analysis performed only on the individuals of *S. globulifera* to analyze its genetic structure based on these new developed SNPs. As all individuals of *S. globulifera* showed a diploid state in all 144 SNPs, we scored all SNP genotypes as diploids (i.e., using two characteres).

To estimate the degree of relatedness among genetic groups based on SNP data, we calculated a genetic distance matrix based on Nei's D (Nei's standard genetic distance measure, Nei, 1972). We defined the genetic groups based on the following criteria: i) The tetraploid gene pool (nGP1) was considered a single genetic group (n = 135 ind.); ii) *S. globulifera* individuals were grouped based on the gene pools defined for GBS analysis in Study II, part 1, where most of the GPs matched with their populations (although we merged MB and NK, and the alternative morphotype in PR and RG, as two gene pools respectively, see Table 7 for number of individuals in each population); iii) pure individuals ($Q > 0.875$) from diploid nGPs detected using STRUCTURE were grouped to represent the four diploid gene pools from Madagascar (number of genetically pure individuals for nGP2, n = 74; nGP3, n = 56; nGP4, n = 74; nGP5, n = 72). The $Q > 0.875$ threshold was chosen because it is expected to discriminate between genetically pure individuals (and second- or later-generation backcrosses) from first generation backcrosses, in case that hybridization occurs (Guichoux et al., 2013).

Then, for the construction of the phylogenetic relationships among the defined genetic groups, we proceeded as in Section 3.2.2.1., using the Seqboot, gendist, neighbor and consense options in PHYLIP (Felstein, 1991) on the matrix with Nei's D values. Next, we calculated an unrooted neighbour-joining tree (Saitou & Nei, 1987) which works well for different evolutionary histories (Kalinowski, 2009) based on the Nei´s distance (1000 bootstrap, majority rule consensus tree). Considering that this type of analysis did not allow missing data for any locus within population and diploid SNPs were required, we worked with two sets of data: a) *S. globulifera* populations and the five nGPs from Madagascar (53 diploid snps in total), and b) *S. globulifera* populations and the four diploid nGPs from Madagascar (tetraploid nGP1 not included, 124 diploid snps in total). The final trees were built in R (APE package, Paradis, Claude, & Strimmer, 2004) using the genetic distance matrix among the genetic groups previously defined (Table S9.4.3.1.).

Congruence of genetic and morphological species delimitation

For those Malagasy individuals which could be identified as a botanical species based on morphological delimitation, we compared their putative identification against their assignment to gene pools for both sets of markers (SSR and SNPs). The threshold of STRUCTURE ancestry proportion considered for gene pools from both sets of markers was $Q > 0.5$. We also compared the previous information with nuclear genome sizes and the geographical distribution of individuals.

4. Results

4.1. Fine-scale spatial genetic structure in *S. globulifera* (Study I)

Genetic diversity

Nuclear microsatellite data and individual coordinates are reported in Supplementary Information S9.1.1. The number of SSR alleles per locus and population ranged from three to 35. In Ituberá, four multilocus genotypes (genets) occurred in more than one (2-3) trees (ramets) sampled in close proximity (4-25m), and in São Tomé, two trees carried the same genotype, but spatial coordinates were unknown for one copy. Plastid DNA haplotype data was only available for both genotype copies in São Tomé, which bore identical haplotypes. *Psex* for the genotype copies was low, from 1.13×10^{-8} to 2.17×10^{-4}, suggesting that trees with identical multilocus genotypes represented clonal copies. Heterozygosity and allelic richness estimates were high and similar in all populations with the exception of Ituberá, where both statistics were slightly lower, although not significantly different from other populations (Table 9). Significant inbreeding was detected in all populations but Yasuní (Table 9). Microchecker detected null alleles in all inbred populations, but F_{IS} corrected for null alleles remained significant in several populations, especially in Africa, suggesting non-random mating (Table 9).

Twenty-five plastid DNA haplotypes were detected across populations. The *psbA-trnH* alignment varied at 28 positions, for sequence lengths of 289-515 bp. Polymorphism varied strongly between populations, from one haplotype in Yasuní to 6 and 7 haplotypes in the Cameroonian populations (Table 9, see Supplementary Information S9.1.3. for the complete list of Genbank accession numbers, including the newly generated sequences KX572421 - KX572686), a result that was mirrored in the estimates of rarefied haplotype richness (Table 9).

Fine scale spatial genetic structure (FSGS)

Significant FSGS was observed in all populations except Yasuní, with an estimated strength of FSGS from *Sp*=0.0003 in Yasuní to *Sp*=0.0341 in São Tomé (Table 10, Fig. 5). *Sp* values and their significance remained similar when the analysis was restricted to three loci in all populations (the loci assessed by Degen, Bandou, & Caron, (2004); in Paracou, Supplementary Information S9.1.1.). These results suggested that the analysed SSRs had sufficient power to detect FSGS and estimate its strength. The sPCA analysis identified a significant global FSGS by means of a G-test within all populations, with eigenvalues of the first sPCA axis, *eig.sPCA*, ranging from 0.029 in BCI to 0.272 in São Tomé (Table 10). FSGS was stronger in African than in Neotropical populations with mean *Sp*=0.025 in Africa *vs.* 0.008 in America and mean *eig.sPCA*=0.180 in Africa *vs.* 0.049 in America (*P*=0.029 for *Sp* and *P*=0.014 for *eig.sPCA* using one-tailed T-tests). *Sp* and *eig.sPCA* were both positively correlated with altitudinal sampling range (for *Sp*: Spearman rho=0.76, *P*=0.033; for *eig.sPCA*: rho=0.89, *P*=0.006).

Most populations displayed significant FSGS for maternally inherited plastid DNA sequences with *Sp* ranging from -0.0032 to 0.4883. The signal was an order of magnitude greater than at

nuclear markers, with stronger structure in African ($Sp \geq 0.277$) than in Neotropical populations ($Sp \leq 0.102$, Table 10).

Table 9. Genetic diversity estimates of *Symphonia globulifera* populations**.** n_{nuc}, sample size for SSR data; SSR, number of SSR loci genotyped; *A*, mean number of alleles per locus; A_R (SD), allelic richness or number of alleles expected in a sample of 34 individuals and its standard deviation; H_E (SE), expected heterozygosity and its standard error based on jackknife resampling; F_{IS}, fixation index; F_{IS}*, fixation index after null allele correction; n_{cp}, sample size for plastid DNA; hap, number of plastid haplotypes; A_{Rp}, plastid haplotype richness or number of haplotypes expected in a sample of 10 individuals; *h*, gene diversity for plastid haplotypes corrected for sample size.

Population	n_{nuc}	SSR	*A*	A_R (SD)	H_E (SE)	F_{IS}	F_{IS} *	n_{cp}	hap	A_{Rp}	*h*
Neotropics											
BCI	147	5	13.8	8.68 (2.07)	0.831 (0.016)	0.148^{***}	-0.049^{**}	10	2	2	0.356
Yasuní	34	5	11	9.21 (2.46)	0.783 (0.038)	0.057^{ns}	nc	10	1	1	0
Paracou	148	3	23.6	12.57 (3.56)	0.880 (0.031)	0.172^{***}	-0.003^{ns}	96	5	3.21	0.494
Ituberá	85	5	10.8	7.26 (5.10)	0.632 (0.061)	0.107^{***}	0.072^{*}	50	4	3.03	0.594
Africa											
São Tomé	42	5	12.6	9.81 (2.76)	0.813 (0.022)	0.183^{***}	0.111^{***}	38	4	2.48	0.450
Nkong Mekak	70	5	14.2	9.95 (5.59)	0.801 (0.044)	0.148^{***}	0.081^{***}	49	6	4.04	0.729
Mbikiliki,	94	5	16.2	9.81 (4.82)	0.748 (0.043)	0.154^{***}	0.086^{***}	50	7	3.34	0.571

Table 10. Estimates of FSGS parameters in *Symphonia globulifera* populations. n_{nuc} , sample size for SSR data; n_{cp}, sample size for plastid DNA; DC, number of distance classes; 1st DC, maximum distance of the first class (m); $F_{ij(1),}$ mean kinship coefficient of the first distance class; *Sp*, intensity of FSGS and *P*-value of one-sided test of the regression slope *b* of F_{ij} on the logarithm of spatial distance; *b* (SE), jackknife mean of *b* and its standard error; *eig.sPCA*: eigenvalue of the first sPCA axis and significance of G-test. ns, not significant. ***, $P \leq 0.001$; **, $P \leq 0.01$; nc, not calculated (no coordinates available).

	SSRs							*Plastid DNA*		
Population	n_{nuc}	**DC**	**1st DC**	$F_{ij(1)}$	***Sp***	***b* (SE)**	***eig.sPCA***	n_{cp}	***Sp***	***b***
Neotropics										
BCI	147	7	113	0.030	0.0166^{***}	-0.0161 (0.0060)	0.029^{***}	10	nc	nc
Yasuní	34	6	131	0.003	0.0003^{ns}	-0.0003 (0.0057)	0.057^{***}	10	nc	nc
Paracou	148	4	203	0.015	0.0090^{***}	-0.0088 (0.0029)	0.038^{***}	96	0.1021^{***}	-0.0925
Ituberá	85	5	152	0.009	0.0074^{**}	-0.0074 (0.0023)	0.073^{***}	50	-0.0032^{ns}	0.0032
Africa										
São Tomé	42	6	856	0.084	0.0341^{***}	-0.0312 (0.0096)	0.272^{***}	38	0.4951^{***}	-0.2802
Nkong Mekak	70	5	312	0.020	0.0124^{***}	-0.0122 (0.0034)	0.111^{***}	49	0.2769^{***}	-0.1787
Mbikiliki	94	7	240	0.072	0.0273^{***}	-0.0253 (0.0105)	0.154^{***}	50	0.4883^{***}	-0.2069

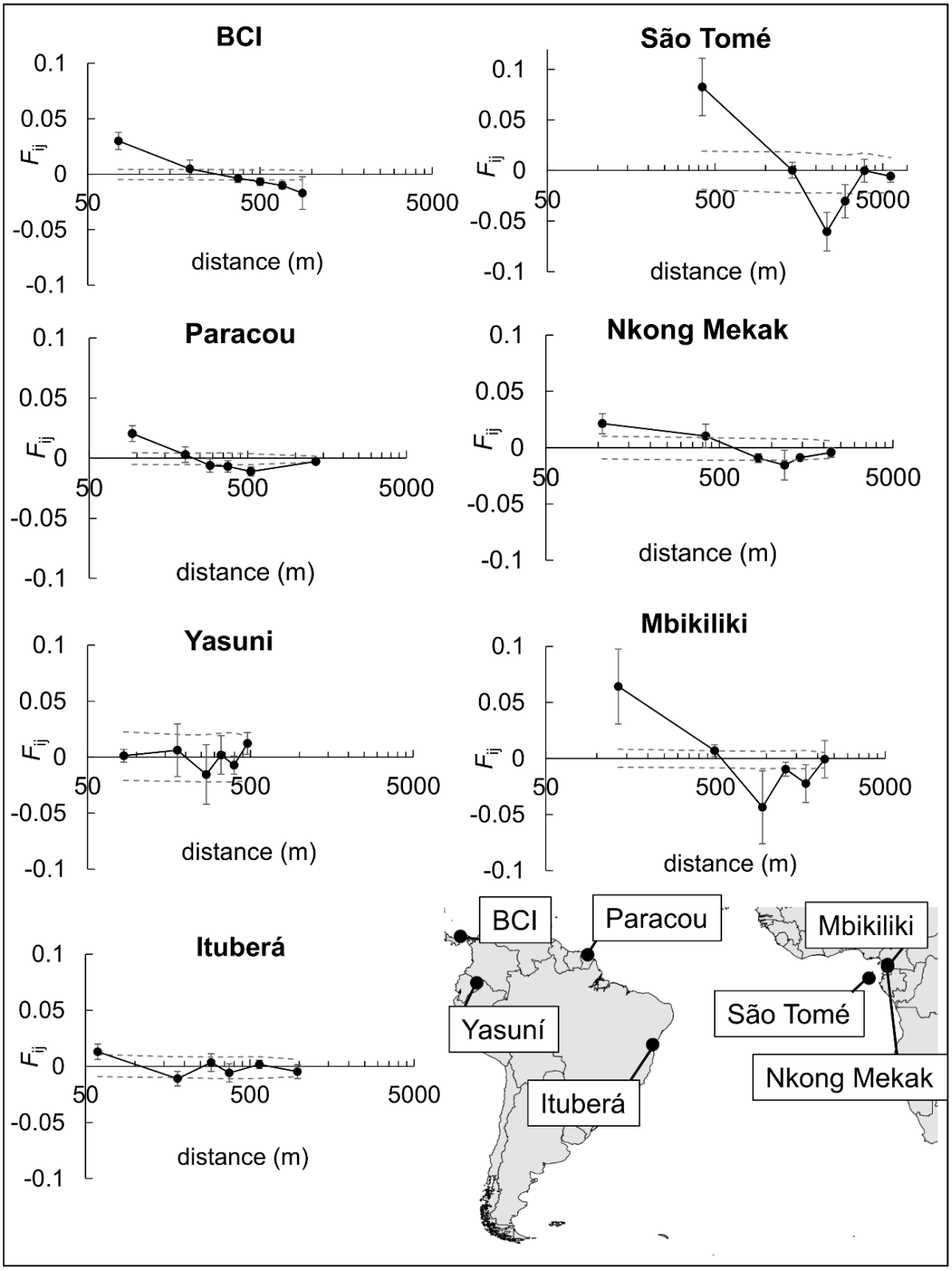


Figure 5. Location of *Symphonia globulifera* populations examined in this study and kinship-distance relationships within populations. The mean jackknife estimate of the kinship coefficient F_{ij} (± standard error) is plotted per distance class, as well as the permutation-based 95% CI for absence of FSGS (dashed grey lines).

Spatial genetic heterogeneity and its causes

Based on sPCA, we detected global structure in all populations (G-test, see above), but the L-test for local structure was not significant in any population, suggesting that neighbouring individuals were not strongly differentiated. STRUCTURE analysis for codominant markers across populations revealed that each population segregated into its own GP except the two Cameroonian populations, which shared the same two GPs. Within populations, the number of GPs that best explained the data was K=1 in the American populations Yasuní and Ituberá, and K=2 or K=3 in the African populations (Fig. 6 and 7, Table 11, Supplementary Information S9.1.2.). For the American populations Paracou and BCI, the selection of the best K was not trivial: STRUCTURE barplots reflected subtle substructure with uneven ancestry proportions Q across individuals in up to K=3 clusters, but L(K) was highest for K=3 in BCI and for K=1 in Paracou, the solutions we eventually retained (Supplementary Information S9.1.2.). The recessive alleles model in STRUCTURE gave similar results, with Pearson correlation coefficients $r \geq 0.94$ for individual ancestry proportions between the codominant and recessive alleles models (Supplementary Information S9.1.2.). We thus considered that null alleles had a negligible effect on the STRUCTURE analysis and retained only results from the codominant marker model for further analyses. In populations with multiple GPs, the proportion of individuals assigned at Q>0.875 was high (57-70%) in the African populations reflecting putative coexisting demes, while individuals in American populations were more homogeneous or admixed on average (Table 11; compare bar plots representing Q in American (Fig. 6) vs. African populations (Fig. 7), Supplementary Information S9.1.2.). STRUCTURE results were broadly congruent with those obtained in TESS (Supplementary Information S9.1.2).

In the African populations, F_{IS} within GPs ($Q \geq 0.875$) was generally non-significant (Table S9.1.5.1.), suggesting that deviation from Hardy-Weinberg equilibrium at the population level (Table 9) was largely due to population substructure. F_{IS} was however significant in GP2 in both Mbikiliki and Nkong Mekak and remained significant after correction for null alleles in Nkong Mekak, suggesting deviation from random mating within this GP, e.g., due to selfing or biparental inbreeding.

Cyto-nuclear disequilibria based on one-way ANOVA were detected in the three African populations only: individuals carrying different haplotypes differed in their mean sPCA score (Table 12). Haplotype-GP association tests were only significant in the two Cameroonian populations, where GP1 was associated with haplotype H19 and GP2 with H24 in both populations (Fisher test, P<0.001; Fig. 7). The associations were still significant after controlling for geographical distance (Partial Mantel Tests, P<0.001 in both populations). In Paracou, individuals from the same morphotype were genetically more related at plastid DNA than expected at random (Fisher test: P<0.01; partial Mantel test: P<0.05).

Finally, we detected a clear altitudinal stratification in African but not in Neotropical populations: in Africa, individuals from different *ad hoc* altitude classes differed in their mean sPCA scores (Table 12). Further, in all populations with multiple GPs, at least one GP was associated with a specific altitudinal class (Table S9.1.7.1.). These results also supported stronger altitudinal stratification of GPs in Africa.

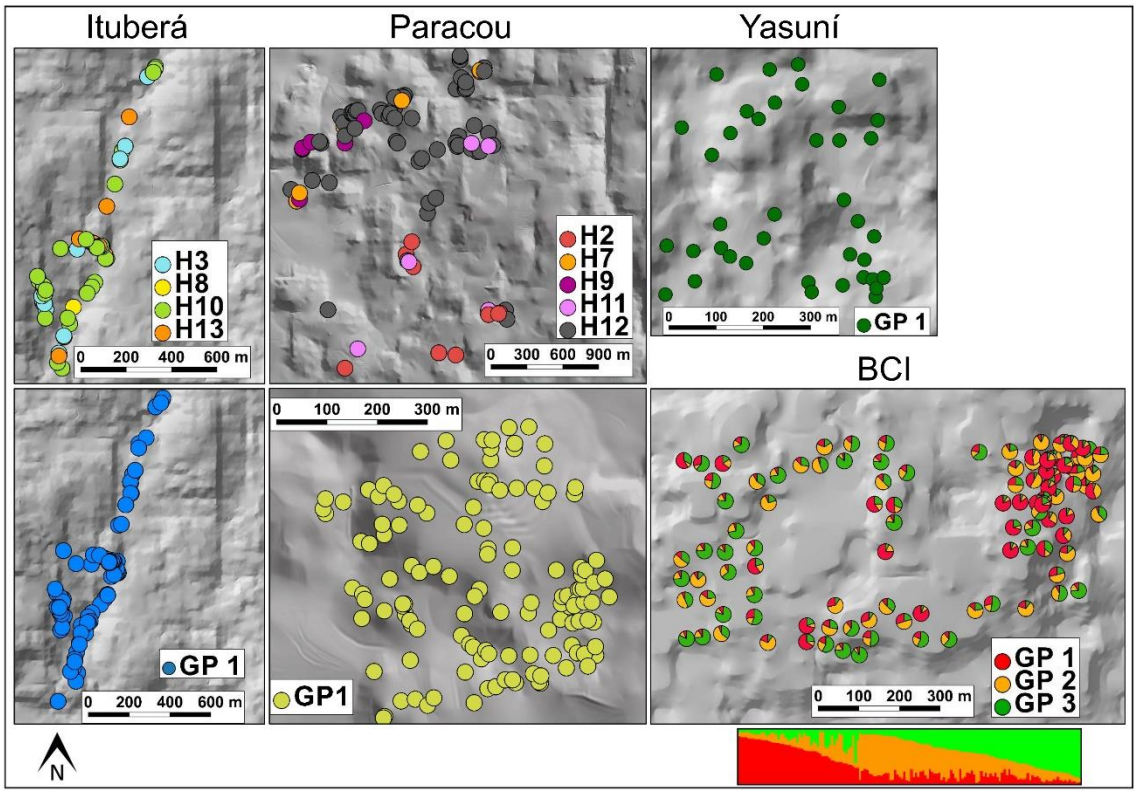


Figure 6. Fine-scale spatial genetic structure in Neotropical populations of *Symphonia globulifera*. Each individual is plotted on the map as a disc representing the colour of its specific plastid DNA haplotype ("H") or as a pie chart indicating the ancestry proportions, Q, in different genetic clusters ("GP"), as defined in the STRUCTURE analysis for the number of clusters *K* best describing the data. Individual STRUCTURE barplots below each population map illustrate the distribution of ancestry proportion for each of the *K* gene pools.

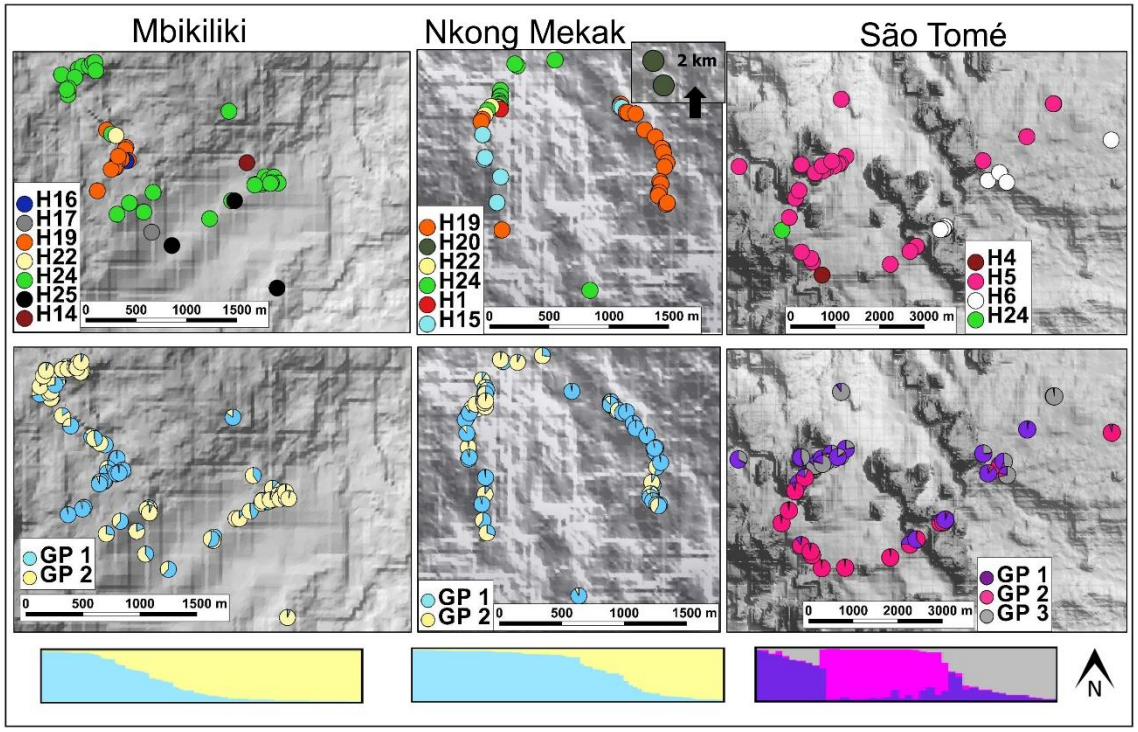


Figure 7. Fine-scale spatial genetic structure in African populations of *Symphonia globulifera***.** Each individual is plotted on the map as a disc representing the colour of its specific plastid DNA haplotype ("H") or as a pie chart indicating the ancestry proportions, Q, in different genetic clusters ("GP"), as defined in the STRUCTURE analysis for the number of clusters *K* best describing the data. Individual STRUCTURE barplots below each population map illustrate the distribution of ancestry proportion for each of the *K* gene pools.

Table 11. Strength of genetic differentiation between nuclear gene pools (GPs) within *Symphonia globulifera* populations. *K*, number of STRUCTURE clusters; $F_{ST(Q\geq0.5)}$, F_{ST} among GPs with individual assignment based on $Q\geq 0.5$; $F_{ST(Q\geq0.875)}$, F_{ST} among GPs with $Q\geq0.875$; PI50 (%), proportion of individuals assigned to a GP based on $Q\geq 0.5$; PI87 (%), proportion of individuals assigned to a GP based on $Q\geq0.875$. nd, not defined; ***, $P\leq0.001$.

Population	*K*	$F_{ST(Q\geq0.5)}$	$F_{ST(Q\geq0.875)}$	PI50 (%)	PI87 (%)
Neotropics					
BCI	3	0.082***	nd	76.9	0
Yasuní	1	nd	nd	nd	nd
Paracou	1	nd	nd	nd	nd
Ituberá	1	nd	nd	nd	nd
Africa					
São Tomé	3	0.168***	0.234***	97.6	57.1
Nkong Mekak	2	0.082***	0.141***	100.0	70.0
Mbikiliki	2	0.102***	0.188***	100.0	67.0

Table 12. Spatial genetic heterogeneity in SSR data and its association with plastid DNA haplotypes (i.e., cytonuclear disequilibria) and altitude. The mean sPCA score for the first sPCA axis is given for individuals carrying the same haplotype, sPCA (hap), or belonging to the same *ad hoc* altitudinal class sPCA (alt); n, sample size range per altitudinal class. *P* values represent the significance of one-way ANOVA analyses testing differences in the mean sPCA score for haplotypes, *P*(hap) or altitudinal classes, *P*(alt). nc[1], not computed because coordinates were unavailable or populations were monomorphic; nc[2], not computed because SSR and plastid DNA data were collected from different individuals; ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$.

Population	*P*(hap)	sPCA (hap1)	sPCA (hap2)	sPCA (hap3)	*P*(alt)	n	sPCA (alt1)	sPCA (alt2)	sPCA (alt3)
Neotropics									
BCI	nc[1]	nc[1]	nc[1]	nc[1]	ns	46-51	-0.022	-0.011	0.031
Yasuní	nc[1]	nc[1]	nc[1]	nc[1]	ns	11-12	0.053	0.003	-0.057
Paracou	nc[2]	nc[2]	nc[2]	nc[2]	ns	46-55	0.036	-0.044	0.001
Ituberá	ns	-0.086 (H10)	-0.104 (H13)	0.091 (H3)	ns	28-29	0.127	-0.071	-0.055
Africa									
São Tomé	*	-0.209 (H5)	0.431 (H6)	-	***	12-16	-0.568	0.036	0.716
Nkong Mekak	***	-0.412 (H15)	-0.309 (H19)	0.323 (H24)	**	23-24	-0.075	0.300	-0.216
Mbikiliki	***	-0.688 (H19)	0.182 (H25)	0.258 (H24)	***	31-32	0.230	-0.335	0.102

4.2. Large-scale genetic structure in *S. globulifera*, demographic history and adaptive evolution (Study II)

4.2.1. Spatial genetic structure of *S. globulifera* across continents (Study II, part 1)

Inference of genotypes and gene pool delimitation

DNA sequencing resulted in 634,404,152 reads (84-90 base pairs long) from identified barcoded individuals. Using the *de novo assembly* (162.186 contigs from a subset of 40 million reads) and targeting SNPs with common variants and the independence of loci, we obtained genotype likelihood data for 4921 SNPs in a final set of 367 individuals with a good coverage (mean coverage of 16.9 reads per locus per individual, 30,657,415 reads in total, see Table 13 for amount of missing data).

At *K*=2, Entropy identified two distinct gene pools (GPs) representing mainly Neotropical vs. African samples, with admixture for some African locations (GB, KR and BN showed a progressive increase in the admixture proportions with the Neotropical GP), and São Tomé surprisingly assigned to the Neotropical GP. For *K*=9, the best supported model as by DIC, individuals were assigned into nine differentiated GPs, (Fig. 8, Fig. S9.2.1.1.; see Table 13 for details on GPs). Four GPs had a straightforward correspondence with four locations (BN, KR, ST, IT) with almost no signs of admixture with other GPs. MB and NK were merged into the same homogeneous GP (the MN gene pool), except for three individuals (MB: 1 ind., NK: 2

ind.) that formed an independent GP. Individuals from GB showed strong admixture between MN and KR gene pools. Entropy revealed three GPs in French Guiana despite the proximity of populations and the sympatry of morphotypes within those populations. The widespread *Symphonia globulifera* morphotype split into two separate GPs, typical of locations PR and RG, respectively, whereas the third GP was present in both populations and matched with the information recorded for the alternative morphotype, abbreviated as SP for *S. sp.1* (see Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007). Additionally, in each location, individuals from both morphotypes showed evidence of admixture between their GPs (see Table 13 for numbers of individuals assigned to each GPs).

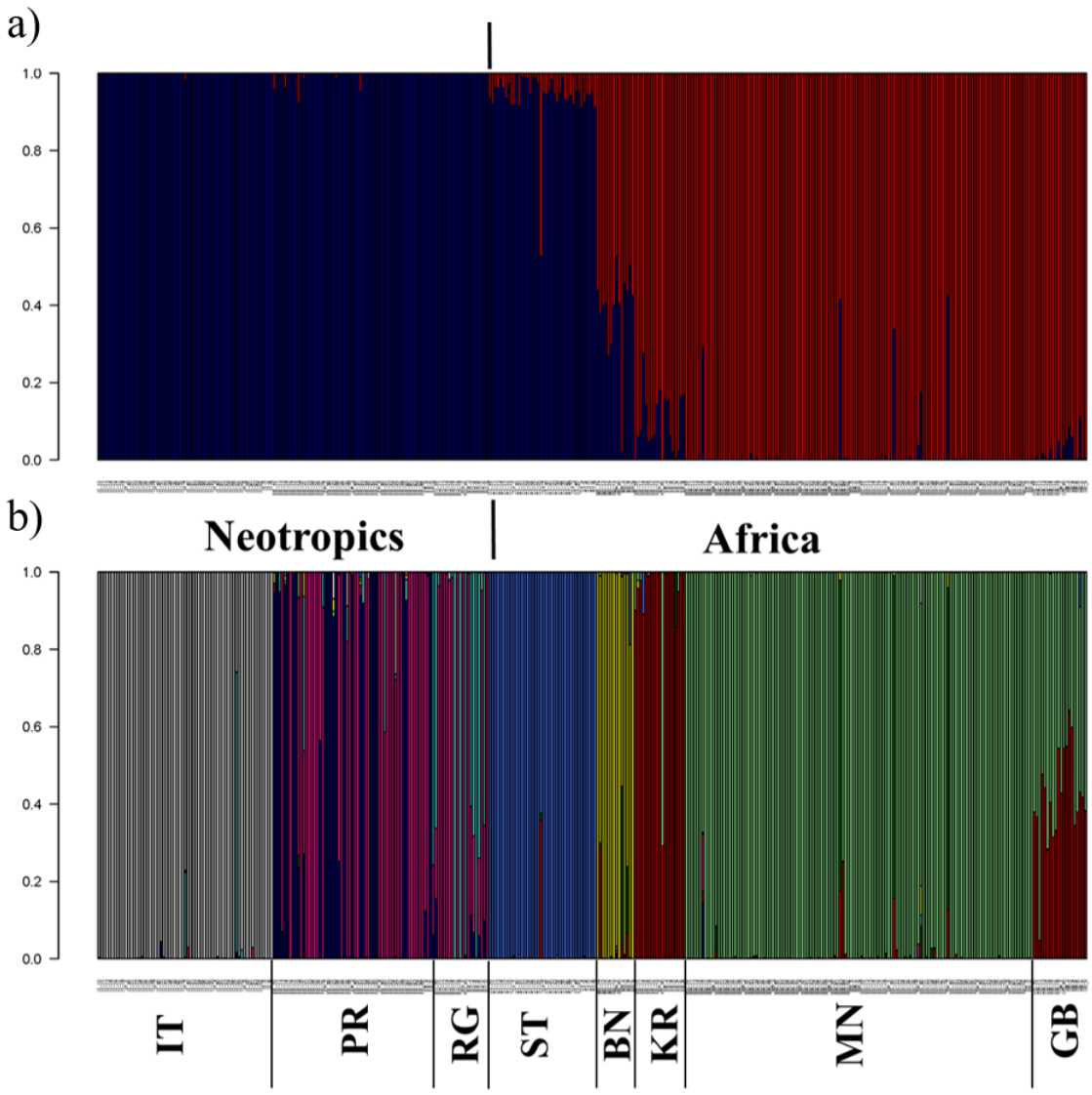


Figure 8. Entropy barplots illustrating the distribution of ancestry proportion of individuals in each of the *K* gene pools from our populations: a) *K*=2, b) *K*=9 (pink GP: alternative morphotype in PR and RG).

Genetic diversity

We computed genetic diversity statistics for eight gene pools as well as for the location GB that was admixed (see Materials and Methods). Based on estimates corrected for uneven sample size, GB presented the highest number of polymorphic sites, whereas BN and RG were the GPs with more fixed alleles. When all individuals were considered, no GP presented both alleles for all SNPs: only MN showed polymorphisms in more than half of the final set of SNPs (Table 13). Estimates of heterozygosity (particularly *He*) together with the proportion of polymorphic sites indicated that genetic diversity was greater within MN and GB (the latter probably because of the admixture between two different gene pools) and minimum in BN. The inbreeding coefficient (F_{IS}) ranged from -0.354 to -0.014, with only one F_{IS} value not significantly different from zero in KR.

Table 13. Genetic diversity estimates in sampling locations or gene pools of *Symphonia globulifera* and the alternative morphotype in Africa and the Neotropics, based on 4921- SNPs genotyped with high certainty. GP or sampling location (ID), gene pools from Entropy considered in the analysis, or sampling location in case of strong admixture, and their ID; **All individuals**: N, number of individuals assigned to GP or location; $MD_{m\text{-}M}$ (ind), minimum and maximum levels of missing data per individual within GP or location (%) and individuals with more than 20% of missing alleles within GP or location (%); NPS, number of polymorphic sites; **Sampled subset**: n, sample size for gene diversity estimates; nps, number of polymorphic sites; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS} *(95%IC)*, fixation index and 95% confidence interval from bootstrapping.

	All ind.			**Subset**				
GP or sampling location (ID)	**N**	**$MD_{m\text{-}M}$ (ind.) %**	**NPS**	**n**	**nps**	***Ho***	***H_E***	***F_{IS} (95%IC)***
Benin (BN)	14	6.8-29.8 (21.4)	832	14	832	0.066	0.058	-0.264 (-0.344; -0.221)
Korup (KR)	19	4.2-28.5 (21.1)	1435	19	1435	0.074	0.073	-0.014 (-0.049; 0.019)
Mbikiliki & Nkong Mekak (MN)	129	4.7-40.7 (28.7)	2518	20	1484	0.103	0.080	-0.354 (-0.396; -0.330)
Gabon (GB)	20	5.4-33.8 (25.0)	1709	20	1709	0.091	0.081	-0.132 (-0.162; -0.104)
São Tomé (ST)	40	3.5-33.4 (2.5)	1400	20	1235	0.073	0.064	-0.169 (-0.207; -0.133)
Itubera (IT)	65	4.1-26.3 (4.6)	1422	20	1178	0.084	0.069	-0.241 (-0.279; -0.207)
Paracou (PR)	25	4.4-24.2 (24.0)	1431	20	1375	0.078	0.070	-0.135 (-0.170; -0.102)
Regina (RG)	9	5.0-18.2 (0.0)	881	9	881	0.078	0.064	-0.300 (-0.363 -0.261)
Alternative morphotype in French Guiana (SP) (ind. from PR/RG)	46 (35/11)	3.7-29.9 (15.2)	1805	20	1457	0.087	0.074	-0.1756 (-0.207; -0.143)

Genetic distance

Nei's genetic distance among GPs or locations ranged from 0.0112 to 0.1059 and the tree topology was very consistent for all GPs but two (main topology for SP and PR: 98.9% and 84.3% of bootstrap replicates, Fig. 3 and Table 14). Gene pools from the Neotropics were more closely related to each other than to those from Africa and showed a closer relationship between IT and SP than previously expected based on the sampling location of samples. African gene pools presented genetic distances very correlated with geography (based on location of gene pools). In broad terms, their genetic distances from American GPs decreased following a south-to-north direction, with the GP from São Tomé Island in between the continental gene pools.

Table 14. Estimates of genetic distance based on Nei's D among gene pools of *Symphonia globulifera* and the alternative morphotype in Africa and the Neotropics.

	BN	**KR**	**MN**	**GB**	**ST**	**IT**	**PR**	**RG**	**SP**
BN	-								
KR	0.0510	-							
MN	0.0664	0.0325	-						
GB	0.0569	0.0194	0.0112	-					
ST	0.0682	0.0692	0.0849	0.0743	-				
IT	0.0938	0.0962	0.1059	0.0970	0.0923	-			
PR	0.0806	0.0820	0.0936	0.0837	0.0757	0.0521	-		
RG	0.0795	0.0814	0.0950	0.0844	0.0736	0.0505	0.0308	-	
SP	0.0819	0.0842	0.0975	0.0869	0.0760	0.0445	0.0328	0.0362	-

Phylogeographic history

Topologies in SNAPP (Fig. 9) presented similarities among all models (M1: three topologies, M2 and M4: one main topology), showing a tree with two main branches (one for each continent), each one presenting hierarchical tree topologies. M3, with the least informative priors, failed to converge. All except three nodes in the topologies were very well supported (posterior probability = 1, see Fig. 9D).

In the three successful analyses, the African branch presented a constant topology among continental GPs, with more ancient divergence events than in the Neotropics. MN and GB were the most closely related GPs, whereas KR and BN presented an increasing time of divergence with respect to MN and GB with ST as the most ancient gene pools in Africa. Neotropical lineages were more recent: RG and PR were always closely related, together with SP in most topologies, whereas IT was the most divergent. Interestingly, M1 showed three alternative topologies for ST, which indicated that there are important admixture proportions with GPs from Africa and the Neotropics, but not enough data information to clearly define the topology in this model.

Overall, topologies were steady and well delimited in all models. However, divergence time varied for some populations among models and some theta values did not reach convergence, despite the ESS values (effective sample size) being > 200 in most of the model parameters.

The analysis in TreeMix showed a consistent topology among African GPs, identical to the results in SNAPP, as well as between them and Neotropical GPs (100 % of bootstrap replicates). The analysis only showed a slightly variable topology among American populations (main topology: 70% of bootstrap replicates, two main alternative topologies with 10-13% of bootstrap replicates, Fig. 10), consistent with the variability found in M1 in SNAPP. The other 7% of bootstrap replicates corresponded with four other models for the Neotropical locations (not shown).

The TreeMix topology showed an increase of genetic drift among gene pools from Africa, from GB north- and west-ward to São Tomé Island and the American block sequentially. IT was consistently the most diverged population from the root in all bootstrapped models, in agreement with its higher time of divergence in SNAPP with respect to the other Neotropical GPs. Also, IT was closely related to SP in the three main alternative models for the Neotropical locations, although this result differed from the Neotropical topology in SNAPP, where RG, PR and SP were closely related for almost all topologies.

The model without migration already explained most of the variance in ancestry between populations (99.8%, see Pickrell & Pritchard, 2012). Also, differences among scaled residuals were small (≤ 2 SE, see Fig. S9.2.3.1.). Large positive residuals would indicate populations where the fit might be improved by adding migration edges, if they were as they are more closely related in the data than expected under the maximum likelihood tree (Pickrell & Pritchard, 2012). As that was not our case (see scaled residuals in Pickrell & Pritchard, 2012 for comparison), it was not necessary to include migration events to improve the model.

Using BayPass (see Section 3.2.2.2.), confidence in the correlation matrix derived from Ω was assessed by visually comparing its hierarchical clustering tree (see Fig. S9.3.2.1.) against the Inferred Maximum likelihood tree implemented in TreeMix and the unrooted neighbour-joining tree of Nei's genetic distance. The relationships of the five African locations were similar in the three topologies. The only exception was BN in the hierarchical clustering tree, probably because of slight variations in methods since the tree showed almost equal distance between KR and BN with respect to the other populations.

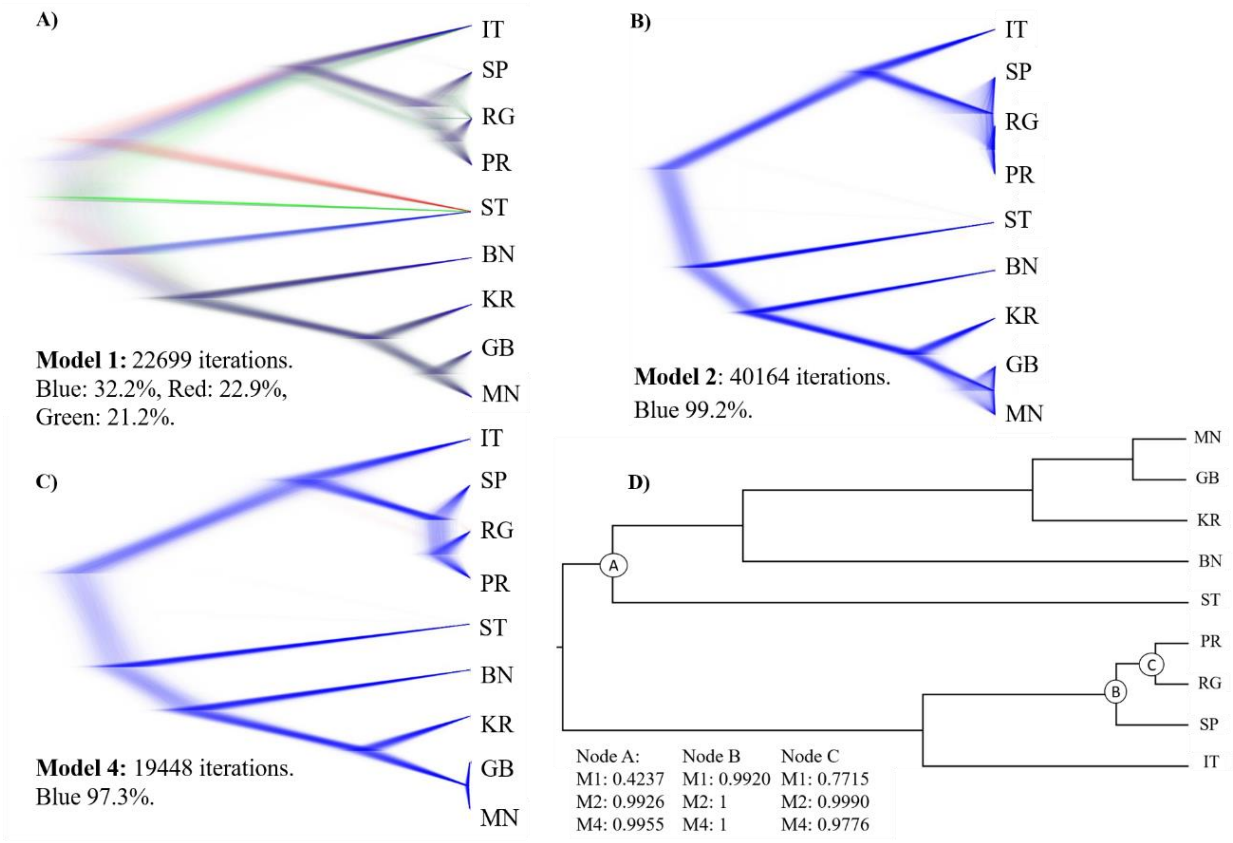


Figure 9. A-C) SNAPP species trees for *Symphonia globulifera* inferred from 4921 nuclear bi-allelic SNPs in African and Neotropical GPs using four alternative hypotheses for the model (model 3 failed). For each model, the total of iterations and the support for topologies are indicated as % of the iterations corresponding to each color. D) Tree representing the main topology for the three models successfully performed in SNAPP and their posterior probabilities of nodes (M1: model 1, M2: model 2, M3: model 3).

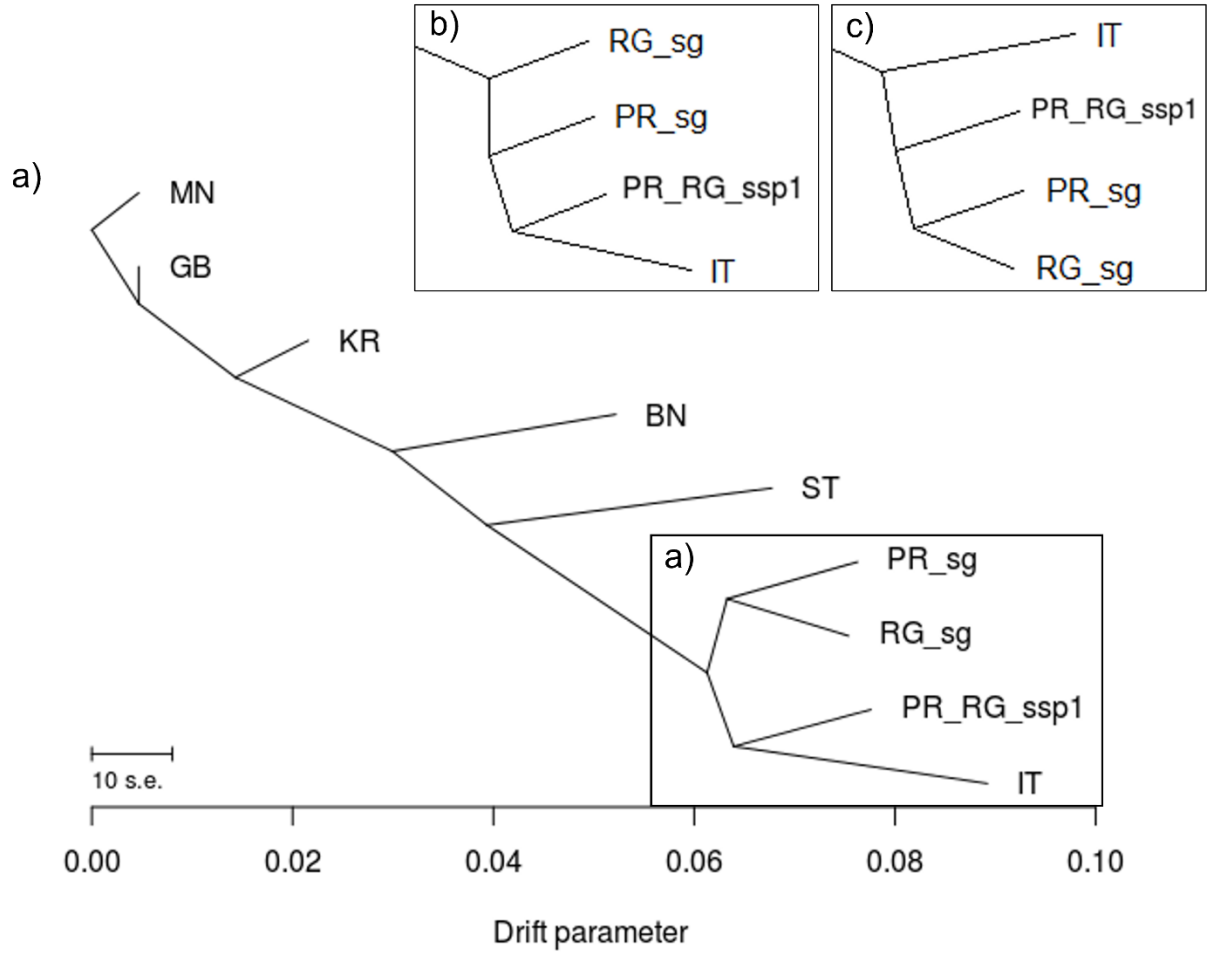


Figure 10. Inferred Maximum likelihood tree implemented in TreeMix from African and Neotropical GPs when no migration edges were fit. The amount of genetic drift in each branch is proportional to the horizontal branch lengths. The scale bar represents 10 times the average standard error of population relatedness in the sample covariance matrix of the model. Support of bootstrap replicates for topologies: A) 100%, B) 70%, C) 10%, D) 13%.

4.2.2. Local adaptation of *S. globulifera* at continental scale in Africa (Study II, part 2)

We performed the following analyses on a data set with 3399 SNPs in 182 individuals, where missing SNP calls were ≤15% in all except 2 individuals (17% and 19%) from Africa.

BayeScan identified 24 loci as outliers, with a minimum posterior probability of 0.93. All alpha values were positive, suggesting diversifying selection. Based on the values for log10 (PO) and using the Jeffreys' scale, the evidence for selection affecting those loci was "decisive" (18 loci, log10 (PO) > 2), "very strong" (5 loci, log10 (PO): 1.5-2) and "strong" (1 locus, log10 (PO): 1-1.5; see Fig. 11 and Table S9.3.2.1.).

BayeScEnv identified 8 loci as outliers associated with the Aridity Index (5 loci), BIO17 (1 locus) and PH_T (6 loci). Only four of these loci were correlated with unique covariates (Table 15).

For the covariate-free approach in BayPass, 23 outlier SNPs were observed (Fig. 12, Table S9.3.2.2.). In the covariate approach in Baypass, 21 outlier SNPs were detected with different levels of evidence on the association, and some of these SNPs were correlated with several covariates. The association of SNPs with one or more covariates showed "decisive" evidence for 3 SNPs, "very strong" evidence for 6 SNPs and "strong" evidence for 19 SNPs (Table 16, Fig. S9.3.2.2.).

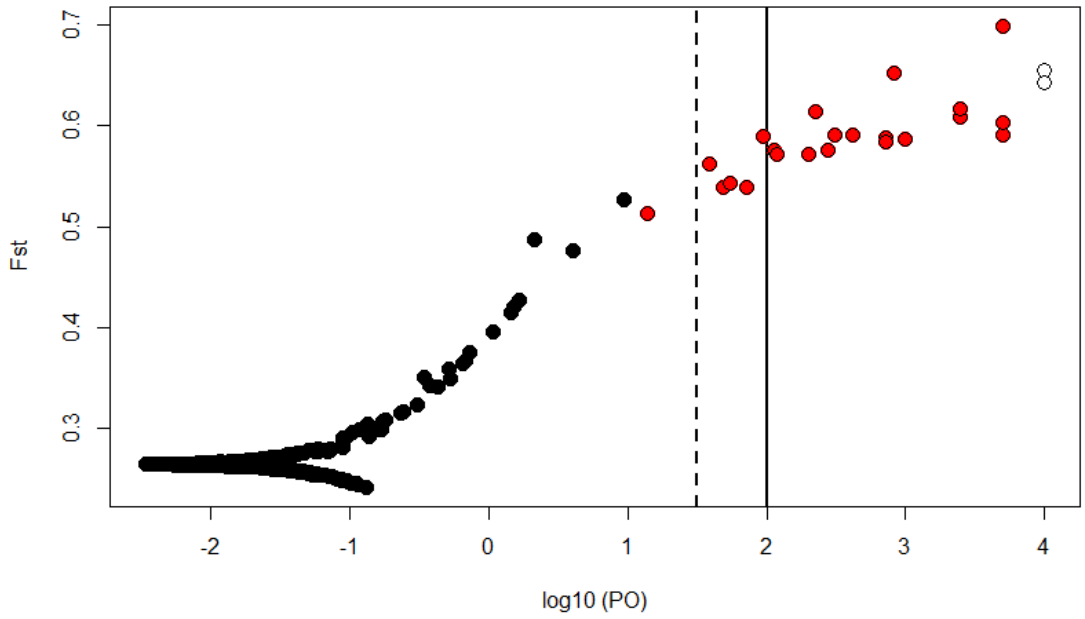


Figure 11. BayeScan results: distribution of log-transformed posterior probabilities and locus specific F_{ST}. Loci identified as outliers are shown in red and white (the posterior probabilities for loci in white were 1 so their log10 (PO) values would be infinity). The dashed and solid lines indicate log10 (PO) of 1.5 and 2, which correspond to posterior probabilities of locus effects of 0.97 and 0.99, respectively.

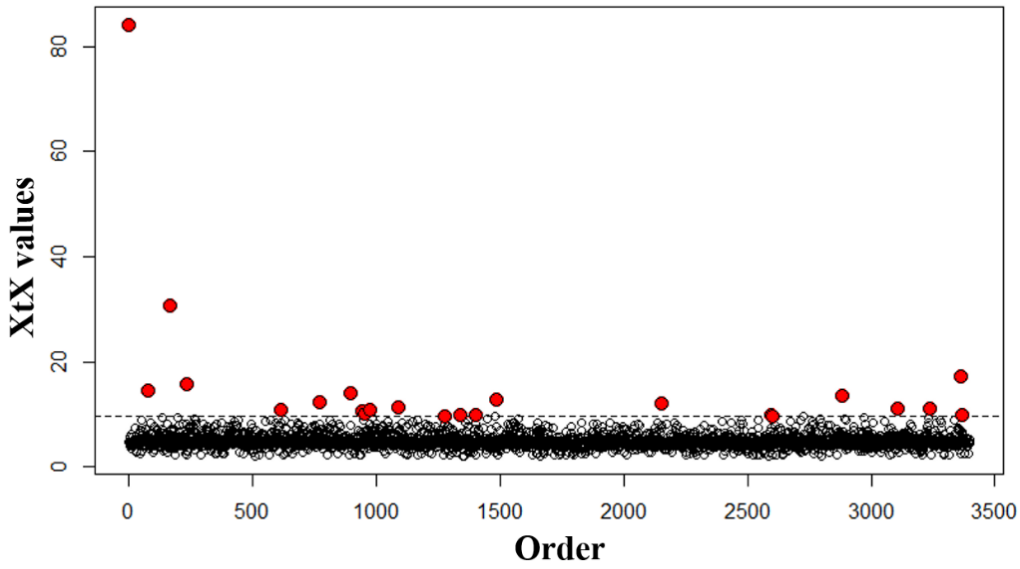


Figure 12. Loci identified as outliers (in red) using the core model and the XtX statistics in BayPass for the five continental locations in Africa.

Table 15. Loci identified as outliers (q-value < 0.05) in BayeScEnv analysis for the five continental locations in Africa. For each locus and covariate, the table displays q-value on the *g* parameter in the model. SNP ID: names identifying the loci on the original 4921 SNPs dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. Covariates: uncorrelated environmental variables.

		COVARIATES		
Locus ID	**Order**	**aridity index**	**BIO17**	**PH_T**
Locus3	1			0.0224
Locus98	78	0.0404		
Locus219	164		0.0268	
Locus1571	1087	0.0295		0.0428
Locus2151	1485			0.0038
Locus4685	3236	0.0090		0.0004
Locus4722	3261	0.0484		0.0337
Locus4871	3361	0.0103		0.0108

Synthesis of Outlier Tests

Although we have detected 54 different outlier loci using four types of analyses, only 12 SNPs were identified as outliers in at least two methods (Table 17). Both covariate approaches detected several SNPs showing significant associations with one or more variables (Table 15 and 16). All outliers identified by BayeScEnv were also detected using BayeScan, whereas only 7 loci were common outliers when comparing the results of both complementary approaches in BayPass. Comparing results from the covariate-free analysis approaches in BayeScan and

BayPass, 9 loci were common outliers. However, the covariate approaches carried out in BayeScEnv and BayPass only found 4 loci in common (Locus 3, Locus 98, Locus 219, Locus 4871) although, remarkably, they were the only significant outliers for all four analyses indicating loci under strong selection. These four loci were outliers for more than one environmental variable considering both covariate-approach analyses but, interestingly, two of those loci were associated to one specific environmental variable in both analyses, different for each locus: Locus 219 (BIO17: Precipitation of Driest Quarter) and Locus 4871 (PH_T: pH in topsoil).

Table 16. Loci identified as outliers (eBPis > 3) using the standard covariate model (IS estimator) in BayPass for the five continental locations in Africa. SNP ID: names identifying the loci on the original 4921 SNPs dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. Covariates: uncorrelated environmental variables. Jeffreys' scale of evidence (BFis in DB): Decisive evidence (D: BFis > 20), Very Strong evidence (VS: 15 < BFis < 20), Strong evidence (S: 10 < BFis < 15).

		COVARIATES							
SNP ID	**Order**	**BIO7**	**BIO18**	**BIO11**	**aridity index**	**BIO17**	**BIO15**	**BIO19**	**PH_T**
Locus3	1		S			D		S	
Locus98	78	S							S
Locus219	164	D	D	S	D	D	D	D	VS
Locus309	235				S			S	
Locus346	256				S			S	
Locus365	269				S			S	
Locus652	462			D					S
Locus844	593				S			VS	
Locus880	614	VS							
Locus1371	949							S	
Locus1410	975		S						
Locus1446	996				S				
Locus1466	1011	S				S	VS		
Locus1812	1257		S						
Locus2645	1830								S
Locus2887	1994					S			
Locus3913	2672								S
Locus4050	2772			VS					
Locus4196	2880								S
Locus4551	3137				S				
Locus4871	3361	S		S		S	S		VS

The genotype frequency distribution of these 12 outlier loci revealed that they were monomorphic for two or more populations, except for Locus 880 (see Fig. 13). However, the strongest genetic structure (four monomorphic populations out of five) was observed only for three SNPs (Locus 309, Locus 2151, Locus 4685). As those three loci did not appear as significant for all four analyses, it seems that the strong genetic structure did not have a decisive influence on the analysis performed. Additionally, we discarded an association between trends of values in both environmental variables detected through the covariate-approach analysis and the neutral allele frequencies along the populations (Mantel test, $P>0.05$ for both covariates, no similar patterns detected between matrices through visual comparison, see Table S9.3.2.3.).

Finally, in spite of performing Blast searches against multiple databases (including local searches against the published *Symphonia globulifera* genome), gene annotation of the 12 outliers was unsuccessful (data not shown), probably due to the short GBS sequence tags (84-90 bases) on which they were located. We only obtained one locus annotation for one outlier (Locus 4685, *E*-value: 2.87e-10, Description: ABC transporter G family member 35).

Table 17. Loci identified as outliers in two or more methods for the five continental locations in Africa. SNP ID: names identifying the loci on the original 4921 SNPs dataset.

SNP ID	BayeScan	BayeScenv	BayPass XtX	BayPass Env.
Locus3	x	x	x	x
Locus98	x	x	x	x
Locus219	x	x	x	x
Locus309			x	x
Locus880			x	x
Locus1571	x	x	x	
Locus2151	x	x	x	
Locus4196	x		x	x
Locus4506	x		x	
Locus4685	x	x	x	
Locus4722	x	x		
Locus4871	x	x	x	x

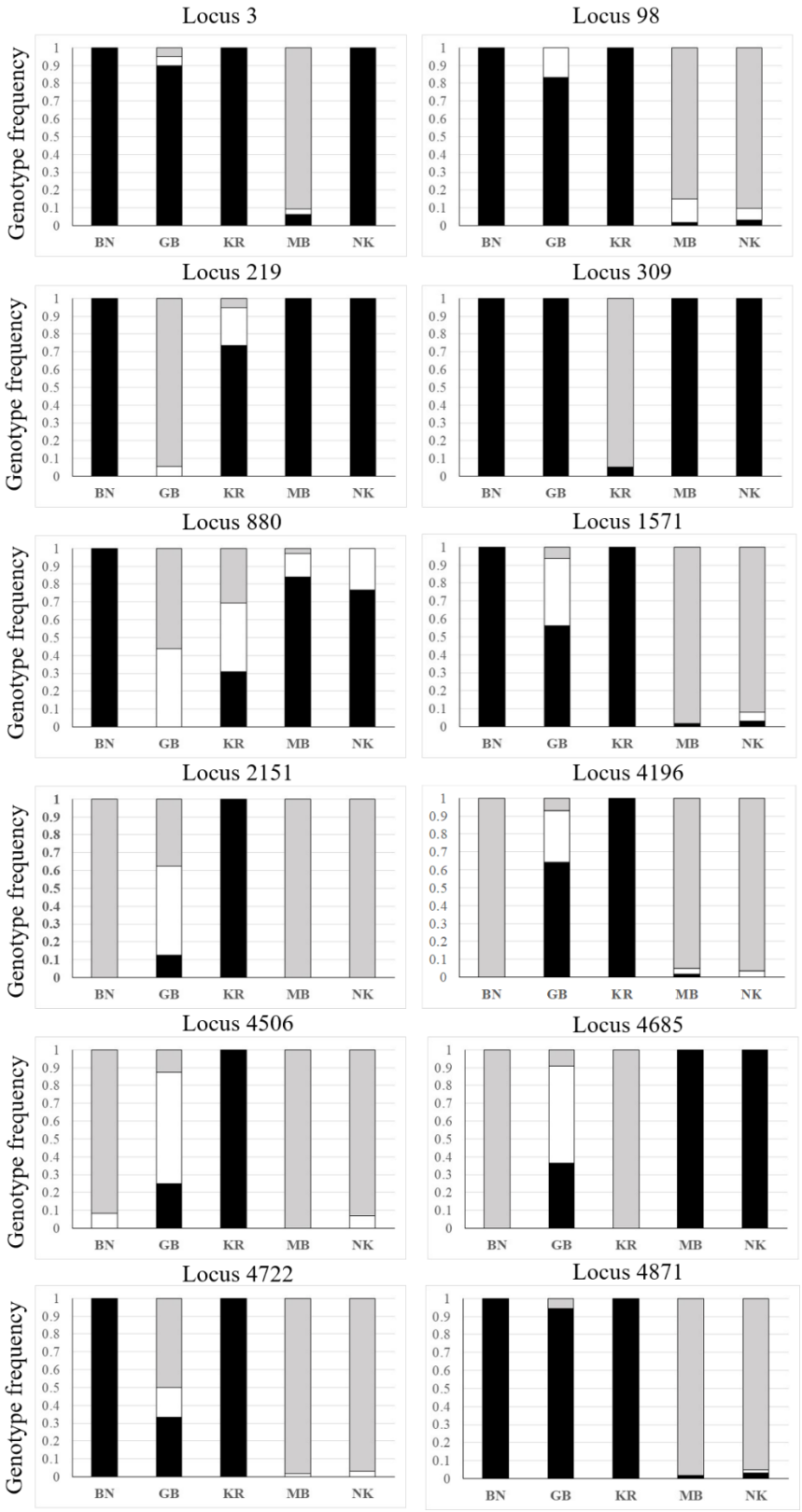


Figure 13. Distribution of the frequency of genotypes per population for the 12 SNPs detected under selection by at least two methods of analysis (missing data in genotypes have not been included). Black: frequency for genotype 0 (homozygous for one allele), White: frequency for genotype 1 (heterozygous), Grey: frequency for genotype 2 (homozygous for the alternative allele).

4.3. Genetic structure within the genus *Symphonia* in Madagascar (Study III)

SNP genotyping and inference of ploidy

Based on the automatic and manual SNP selection methods (see Section 3.1.3.) and using the Sequenom iPLEX^{TM} MassARRAY® technology, we obtained a total of 144 successfully - genotyped, variable SNPs shared between *S. globulifera* and Malagasy *Symphonia*. Selection of candidate SNPs through the visual method worked slightly better than the automated method in the screening step (Table 8). Once SNPs were selected for the genotyping step, the failure rate of both methods was low and very similar. Although missing data in genotyped individuals ranged from 0% to 90.3%, 68.7% of individuals presented less than 10% missing data and 97.1% of individuals presented less than 40% missing data.

During the SNP calling, we observed SNP raw data with four or five genotype groups in 84 SNPs in agreement with polyploidy (tetraploids) in individuals belonging to a specific gene pool inferred in Malagasy samples based on SSR markers (hereafter, rGP): rGP1. It is important to remark that 13 SNPs belonging to the group of those 84 SNPs also presented patterns which could be attributed to polyploidy in individuals from other Malagasy rGPs and we proceeded to scored them as such. However, there was not a consistent group of individuals (i.e., a consistent rGP) that always displayed genotypes consistent with "tetraploid" status based on these 13 SNPs (i.e., the putative polyploid status for a given rGP varied depending on each of these 13 SNP selected). On the other side, 60 SNPs always presented diploid patterns for all individuals.

Ploidy levels inferred from nuclear genome size data and SNP genotypes

Our genome size estimates (1C values) ranged from 2945 to 3589 Mbp in rGP1 while values decreased approximately to a half for the other four rGPs (values from 1317 to 1799 Mbp) (Table 18). These results were consistent with the suspicion of polyploidy inferred during the genotyping of *Symphonia* individuals and suggested a correlation between genome size and ploidy levels, with the higher genome estimates corresponding to tetraploid individuals and the lower ones to diploids. Values for the diploid rGPs were consistent with the genome size estimated by FCM on *S. globulifera* (1522 Mb) by Ewédjè (2012). The use of cambium tissue combined or not with preservation in RNAlater solution was successful for obtaining estimations of genome size using FCM in *Symphonia*. The use of desiccated leaves, usually not preserved in RNAlater, in some cases did not enable to isolate nuclei of sufficient quality for FCM analyses.

Table 18. Genome size estimations (in 1C, Mbp) and inferred ploidy level in Malagasy *Symphonia* individuals (ID) antheir membership in gene pools based on SSRs (rGP). For some of the individuals, different tissues were analysed or the analysis on the same tissue was repeated (two values for 1C). Tissue: type of tissue used on the analysis. RNAlater: tissues preserved in RNAlater solution. 1C (Mbp): genome size of the individuals expressed for an equivalent unreplicated haploid nucleus (thousands of megabase pairs).

ID	rGP	Tissue	RNAlater	1C (Mbp)	Inferred ploidy level using FCM	Inferred ploidy level using SNPs
MH2774	rGP1	cambium		3186	4*x*	4*x*
MH2776	rGP1	cambium		3079	4*x*	4*x*
MH2778	rGP1	cambium		3004	4*x*	4*x*
MH2779	rGP1	leaves		Failed	-	4*x*
MH2836	rGP1	leaves		Failed	-	4*x*
MH3020	rGP1	cambium	RNAlater	3348	4*x*	4*x*
MH3105	rGP1	leaves	RNAlater	2945	4*x*	4*x*
MH3115	rGP1	cambium	RNAlater	3589	4*x*	4*x*
MH3140	rGP1	leaves	RNAlater	3279	4*x*	4*x*
MH2812	rGP2	cambium		1542	2*x*	2*x*
MH2816	rGP2	leaves (A)		1634	2*x*	2*x*
MH2816	rGP2	leaves (B)	RNAlater	1664	2*x*	2*x*
MH2855	rGP2	cambium		1590	2*x*	2*x*
MH2906	rGP2	leaves		1447	2*x*	2*x*
MH2948	rGP2	cambium	RNAlater	1454	2*x*	2*x*
MH2956	rGP2	leaves		Failed	-	2*x*
MH3090	rGP2	cambium (A)	RNAlater	1662	2*x*	2*x*
MH3090	rGP2	leaves (B)	RNAlater	1406	2*x*	2*x*
MH2728	rGP3	leaves		Failed	-	2*x*
MH2747	rGP3	cambium		1388	2*x*	2*x*
MH2825	rGP3	leaves		1600	2*x*	2*x*
MH2826	rGP3	cambium	RNAlater	1317	2*x*	2*x*
MH2832	rGP3	leaves	RNAlater	1527	2*x*	2*x*
MH2834	rGP3	cambium		1418	2*x*	2*x*
MH2845	rGP3	cambium		1384	2*x*	2*x*
MH2890	rGP4	leaves		1529 / 1614	2*x*	2*x*
MH2908	rGP4	cambium		1682	2*x*	2*x*
MH2933	rGP4	cambium		1541 / 1628	2*x*	2*x*
MH2946	rGP4	leaves		1602	2*x*	2*x*
MH2947	rGP4	leaves		1560	2*x*	2*x*
MH2964	rGP4	leaves	RNAlater	1373	2*x*	2*x*
MH2723	rGP5	leaves		1745	2*x*	2*x*
MH2771	rGP5	cambium		1663	2*x*	2*x*
MH2772	rGP5	cambium		1660	2*x*	2*x*
MH2809	rGP5	leaves	RNAlater	1512	2*x*	2*x*
MH2841	rGP5	cambium (A)	RNAlater	1641	2*x*	2*x*
MH2841	rGP5	cambium (B)	RNAlater	1423	2*x*	2*x*
MH2954	rGP5	leaves		1799	2*x*	2*x*
MH2957	rGP5	leaves		Failed	-	2*x*
MH3010	rGP5	cambium	RNAlater	1576	2*x*	2*x*

Genetic structure of Symphonia in Madagascar

The STRUCTURE analysis of the SNP dataset (144 loci in 630 individuals, all SNPs scored as tetraploid genotypes using four characters and indicating both states: diploid - e.g., AA99 - or tetraploid patterns - e.g., AGGG - and missing data: 9999) including all populations from Africa and America, assigned the individuals into three clearly differentiated gene pools for the best model (*K*=3): *S. globulifera*, the tetraploid Malagasy *Symphonia* (coincident with rGP1) and the diploid Malagasy *Symphonia* (Fig. 14). When Malagasy individuals were sorted according the five rGPs (based on SSR data), we could observe that SNP genotypes of a group of individuals belonging to rGP2 and rGP3 (which also corresponded to GP3 based on SNP markers, see below) presented some ancestry in the tetraploid gene pool (rGP1) and in the *S. globulifera* gene pool (blue and green, respectively, in Fig. 14).

However, since models with higher *K* suggested a more complex genetic structure (for example, see *K*=6 in Fig. 14), we performed independent STRUCTURE analysis on each of those three main GPs. For these analyses, the 144 SNPs were scored as tetraploid genotypes in all individuals from Malagasy gene pools as previously, due to the tetraploid gene pool but also because diploid Malagasy individuals still presented 13 SNPs with patterns which could be attributed to polyploidy (see Section 4.3.). Exceptionally, we used diploid scoring in the analysis including only *S. globulifera* individuals. These analyses revealed strong genetic substructure (Fig. 15), with some unexpected results (hereafter, gene pools based on SNP data: nGP). The analysis of the tetraploid gene pool (defined as individuals from rGP1, identified as nGP1 based on SNP markers) revealed two gene pools (best model *K*=2) showing uneven ancestry proportions (*Q*) across individuals and a low proportion of individuals (31%) assigned at *Q*>0.875, probably reflecting within gene pool structure instead of differentiated gene pools. The number of nGPs that best explained the genetic structure of diploid Malagasy *Symphonia* individuals (defined as rGP2, rGP3, rGP4 and rGP5) was *K*=4, reflecting almost completely the structure detected using SSR with the following correspondence: nGP2-rGP2, nGP3-rGP3, nGP4-rGP4 and nGP5-rGP5. The only exception was the fact that more than a third (36.7%) of the individuals from rGP2 were assigned to the gene pool nGP3 and all of them but three occurred in the two Northern Malagasy locations (see Fig. 15 and Fig. 16). Finally, STRUCTURE assigned four populations of Africa (GB, MB, NK, KR) to one gene pool, whereas American populations, BN and ST were assigned to another, with only BN and KR showing evident signs of admixture of both gene pools.

The spatial genetic structure (SGS) in Madagascar revealed by the two types of molecular markers was very similar (see Fig. 16). The only GP which corresponded exclusively to a single population (Nosy Mangabe Island) was nGP4-rGP4, while nGP1-rGP1 was mostly present in central-south and central populations (although very scarcely in Andasibe-Mantadia) and nGP5-rGP5 was mainly present in Andasibe-Mantadia and Farankaraina. The main differences in the SGS revealed by SNPs and SSR markers were related to nGP3-rGP3 and nGP2-rGP2. While rGP2 occurred in the three sampled regions (north, central and central-south) with different levels of presence and rGP3 was present in central-south and central populations (but absent from Andasibe-Mantadia and Anboasarinala), nGP2 and nGP3 showed almost the same distribution except that nGP2 was absent from northern populations and was replaced by nGP3 (see Fig. 16).

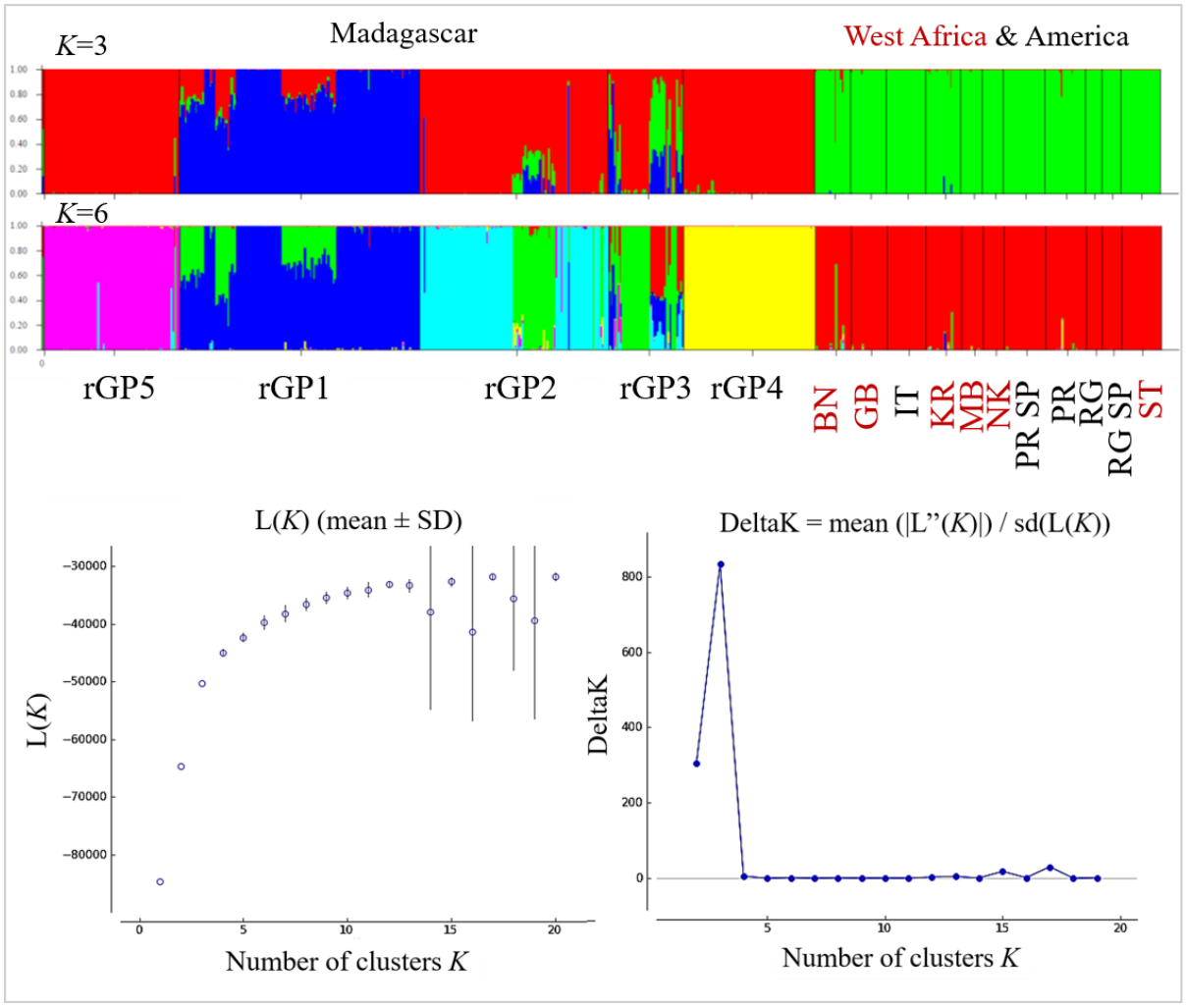


Figure 14. Illustration of *K*=3 (the best number of genetic clusters) and *K*=6 in the STRUCTURE analyses for SNP markers in Malagasy individuals and West Africa & America *S. globulifera* populations, based on tetraploid scoring of 144 SNPs. The best *K* was supported based on the logarithm probability of data (L(*K*)) and Delta *K* (Δ*K*) (plots modified after outputs from Structure Harvester software, Earl & VonHoldt, 2012). Barplots were based on the best run for *K*=3 and *K*=6. Malagasy individuals were sorted according to their membership in five rGPs (based on SSR data), while the plot colours illustrate the ancestry proportions of individuals in each of the *K* gene pools based on SNP data (nGP). Thus, the correspondence between Malagasy GPs detected, based on SSR (rGP) and on SNPs (nGPs) is shown.

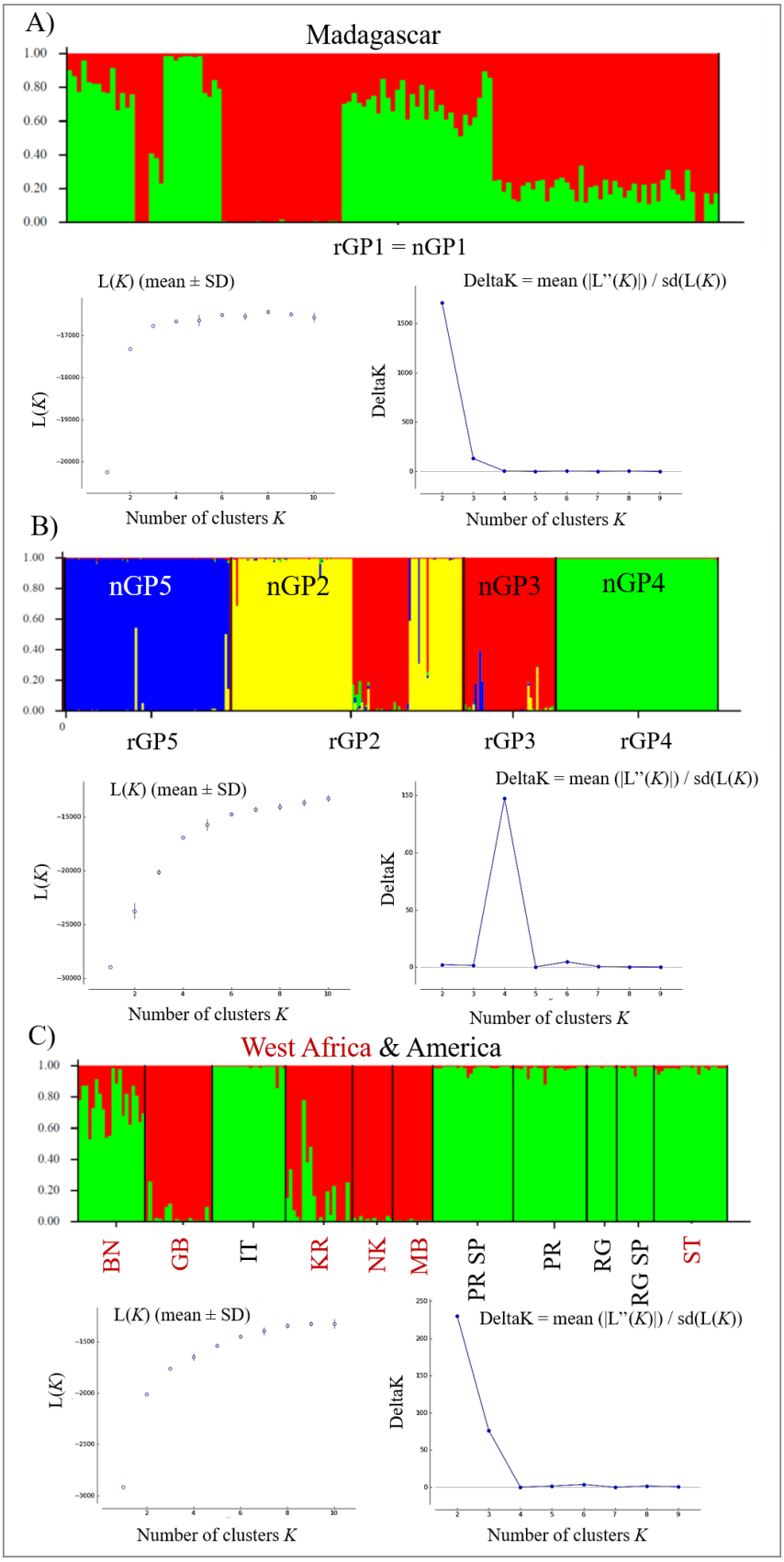


Figure 15. Illustration of the best number of genetic clusters (*K*) from independent STRUCTURE analysis on the three main gene pools discovered based on 144 SNP markers (tetraploid scoring of Malagasy individuals, diploid scoring of *S. globulifera* individuals): A) *K*=2, B) *K*=4, C) *K*=2. The best *K* was supported based on the logarithm probability of data (L(*K*)) and Delta *K* (Δ*K*).

Barplots were based on the best run for each analysis (plots modified after outputs from STRUCTURE Harvester software; Earl & VonHoldt, 2012). Malagasy individuals were sorted according to their membership in five rGPs (based on SSR data), while the plot colours illustrate the ancestry proportions of individuals in each of the *K* gene pools based on SNP data (nGP). Thus, the correspondence between Malagasy GPs detected, based on SSR (rGP) and on SNPs (nGPs) is shown.

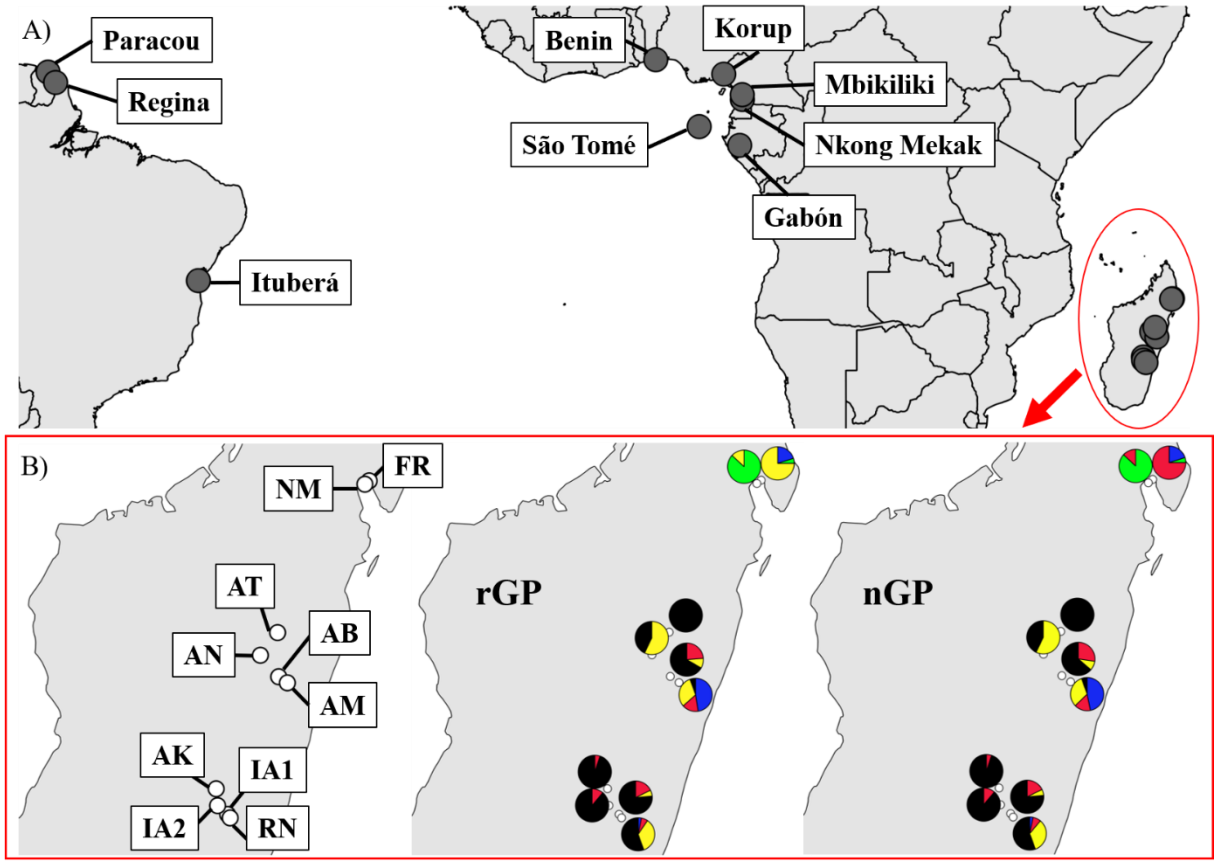


Figure 16. A) Location of sampling sites for *Symphonia globulifera* in the Neotropics and continental Africa and *Symphonia spp.* in Madagascar for Study III. B) Short names for Malagasy sites (see Table 7 for the corresponding non-abbreviated location names) and proportion of the different rGPs and nGPs found in each location (individuals were assigned to a gene pool when (*Q)*>0.5 in STRUCTURE analyses). The pattern of colours for gene pools follows Figure 15 (Black: rGP1/nGP1, yellow: rGP2/nGP2, red: rGP3/nGP3, green: rGP4/nGP4, blue: rGP5/nGP5).

Phylogenetic relationships

Nei's genetic distance among gene pools of *S. globulifera* and the five Malagasy nGPs (based on 53 SNPs due to the requirements of the analysis, see Section 3.2.2.) ranged from 0.0012 to 0.4513 but the inferred neighbour joining tree topology was not very robust. When the tetraploid nGP1 was excluded, the resulting dataset of 124 SNPs (See section 3.2.3.) displayed a similar Nei's genetic distance values (min. - max.: 0.0018 - 0.3397) and NJ tree topology, although its topology was more robust (see Fig. 17 and Table S9.4.3.1.). The main differences between both trees were related to the topology of *S. globulifera* populations, since the 124-SNP tree showed a topology more similar to the one discovered using ca. 5000 SNPs from GBS (compare Figs. 10 and 17).

In both cases, the NJ distance trees indicated two broad genetic clusters: the identified Malagasy gene pools vs. *S. globulifera* populations. The bootstrap support was in general poor and the more reliable nodes were related to Malagasy gene pools (especially to nGP2 and nGP5, the

group of the most differentiated Malagasy gene pools, which surprisingly did not include the tetraploid nGP1) and to the split of *S. globulifera* gene pools from the most southern populations sampled in central Africa (GB and MN, but also KR for the 124-SNP tree). As expected, the gene pools from Madagascar were more closely related to each other than to those from *S. globulifera*. However, despite the occurrence of populations across two distant continents, the genetic distances among *S. globulifera* gene pools did not reach the genetic distance values found among Malagasy gene pools.

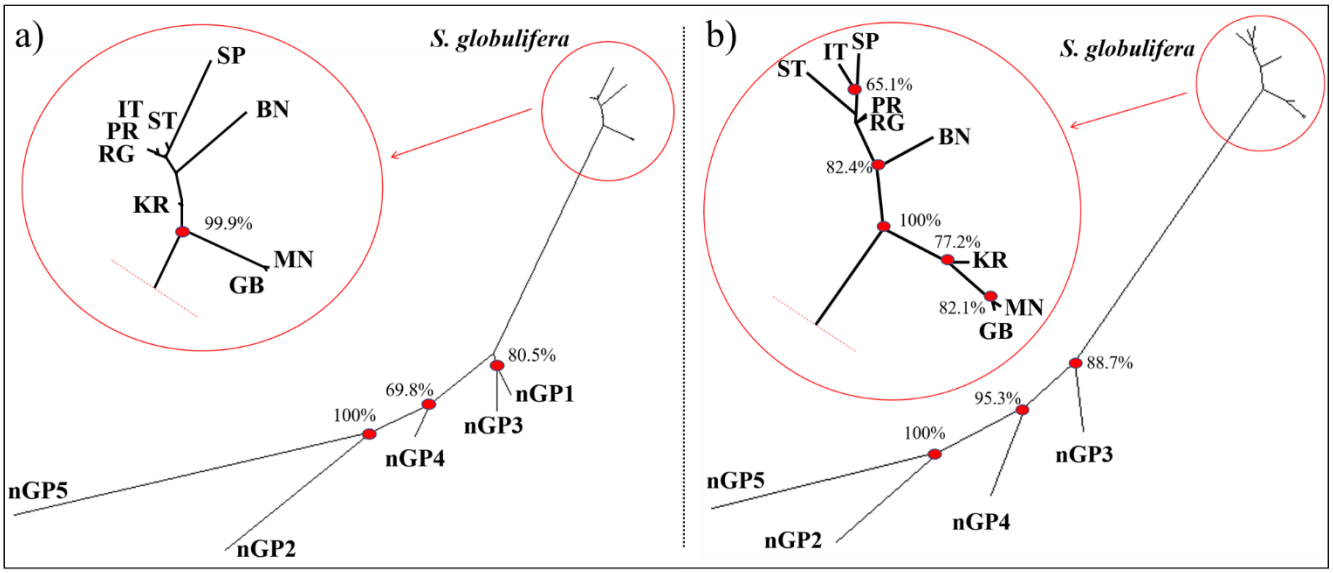


Figure 17. Unrooted neighbour-joining tree based on Nei's genetic distance with scaled branch lengths showing the genetic distance among gene pools of Malagasy *Symphonia* and populations of *Symphonia globulifera*. Values indicate the frequency in which bifurcating nodes occur out of 1000 bootstrap replicates (highest values of nodes displayed). a) Tree including *S. globulifera* populations and the five nGPs from Madagascar based on 53 diploid SNPs. b) Tree including *S. globulifera* populations and the four diploid nGPs from Madagascar (nGP1 not included) based on 124 diploid SNPs.

Congruence of genetic and morphological species delimitation

From a total of 199 plant specimens collected during the sampling in Madagascar for which information on morphological traits of branches and leaves (and flowers or fruits if available, see Supplementary Information S9.4.5.) was collected, putative botanical species determination was obtained for 132 individuals (see Table 7 and Table 19). Additionally, another 51 individuals were assigned to two unidentified morphospecies named after their populations: *S. sp.1* (Farankaraina) and *S. sp.1* (Nosy Mangabe). Considering these groups of individuals with putative botanical species determination (183 ind. in total), 57 individuals belonged to the tetraploid group nGP1-rGP1 and most of them (50 individuals) were assigned to three species: *S. clusioides*, *S. eugenioides* and *S. microphylla*. Only five individuals putatively determined as *S. clusioides* or *S. eugenioides* belonged to other gene pools. Almost all individuals of *S. sp.1* (Nosy Mangabe) were contained in nGP4-rGP4 and represented the main putative species of these gene pools, whereas *S. fasciculata* was the most represented putative species in nGP5-rGP5, gathering all individuals from this species. A mix of species was included in nGP2-rGP2 and nGP3-rGP3, and some of them had different GP assignments based on rGP2 and nGP3 (the most relevant case being *S. sp.*1 (Farankaraina). Most individuals of *S. louvelii* and *S. nectarifera* were gathered in nGP2-rGP2, whereas nGP3-rGP3 contained most individuals from *S. urophylla*.

Table 19. Putative botanical identification of plant specimens of Malagasy *Symphonia* collected in Madagascar, and their assignment to SSR (rGP) and SNP (nGP) gene pools determined by STRUCTURE analyses. Only samples with an ancestry proportion above 0.5 were assigned to a gene pool.

	SSR GP					**SNP GP**				
Putative species	**rGP1**	**rGP2**	**rGP3**	**rGP4**	**rGP5**	**nGP1**	**nGP2**	**nGP3**	**nGP4**	**nGP5**
S. sp.1 (Farankaraina)		10						10		
S. sp.1 (Nosy Mangabe)		2		39				2	39	
S. clusioides	6	2				6	2			
S. eugenioides	26	2	1			26	1	2		
S. fasciculata					13					13
S. louvelii	2	15			4	2	15			4
S. microphylla	18					18				
S. nectarifera	1	6				1	6			
S. pauciflora		1						1		
S. sessiliflora		5			5			5		5
S. tanalensis	2					2				
S. urophylla	2		18	3		2		18	3	
Total	57	43	19	42	22	57	24	38	42	22

5. Discussion

5.1. Fine-scale spatial genetic structure in *S. globulifera* (Study I)

Six out of seven *Symphonia globulifera* populations from Africa and America displayed fine-scale spatial genetic structure based on *Sp*, and all seven had a significant FSGS based on sPCA. The magnitude of FSGS was overall in agreement with expectations for outcrossed tropical trees but varied strongly among populations, from *Sp*=0.000 to *Sp*=0.034 for SSRs. African populations had a much stronger FSGS signal than Neotropical populations, based on both nuclear and plastid markers, and the signal was associated with larger altitudinal gradients in Africa than in America. These results suggested on average a more restricted gene flow, and especially a more restricted seed-based gene flow, in African than in American populations, reflecting a more restricted movement of dispersers in rugged African populations. There was limited evidence for selfing in *S. globulifera*, but null alleles and population substructure (SGH) contributed to deviations from Hardy-Weinberg genotypic proportions within populations. There was evidence for cyto-nuclear disequilibria and historical gene pool differentiation in the two Cameroonian populations, while the population from French Guiana displayed an association of plastid haplotypes with two morphotypes characterized by differential habitat preferences.

Methodological considerations

Some methodological issues are worth discussing with regard to our results. First, samples were collected either randomly or following approximate transects in different populations (Fig. 6, Fig. 7). This should not have affected meaningfully the estimation of FSGS using the *Sp* statistic, which is robust towards differences in sampling scheme (Vekemans & Hardy, 2004; Heuertz, Vekemans, Hausman, Palada, & Hardy, 2003). Second, our sampling covered large distances (maximum distance of one to several km) within populations, likely covering the suitable distance range where kinship decays linearly with the logarithm of distance (Vekemans & Hardy, 2004), hence minimizing the risk of overestimating FSGS due to too short sampling distances (De-Lucas, González-Martínez, Vendramin, Hidalgo, & Heuertz, 2009). Third, our sampling scheme had probably a low power to estimate the decay of kinship at short distance because only a low proportion of sample pairs corresponded to true nearest neighbours in the populations. Fourth, our populations featured different densities of *S. globulifera*. This can affect FSGS, which is expected to increase in low-density populations because of a reduced overlap of seed shadows (Vekemans & Hardy, 2004; Sagnard, Oddou-Muratorio, Pichot, Vendramin, & Fady, 2011). Against *a priori* expectations, however, weak FSGS was observed in the low-density Neotropical populations BCI and Yasuni, where sampling included also younger individuals (>1cm dbh) potentially representing cohorts of related individuals. The weaker than expected FSGS in these populations could have been caused by confounding factors, e.g., increased animal-mediated dispersal distance in low-density populations (Dick, Etchelecu, & Austerlitz, 2003). In any case, considering the stronger FSGS in African than in Neotropical populations observed in our study, variation in population density and age of

sampled individuals did not appear to magnify the pattern of FSGS differences among populations.

Although the number of SSRs used in our study was low (3-5), these highly polymorphic markers were able to detect significant FSGS in all and in six out of seven populations, by means of the G-test and *Sp*, respectively. This indicates a sufficient power for the purpose of the study. In fact, a dataset of 18 genic SSRs on ca. 30 individuals (Olsson et al., 2017) had a lower power than our FSGS analyses (FSGS analyses on data from Olsson et al., 2017 are reported in Table S9.1.4.2.). Another risk of using a low marker number is that it can lead to erroneous GP inference (e.g., using STRUCTURE) because few markers do not capture well the diversity of stochastic lineage sorting processes due to random genetic drift (Orozco-terWengel, Corander, & Schlötterer, 2011). To mitigate this potential problem, we used two types of cyto-nuclear and habitat association analyses, i.e., based on GPs and based on sPCA scores, which gave congruent results. Several studies also reported that IBD can lead to overestimation of the number of GPs inferred by STRUCTURE (Frantz, Cellina, Krier, Schley, & Burke, 2009; Schwartz & McKelvey, 2009). Explicitly adjusting for IBD in our populations by using a spatial prior in the TESS analysis did however not reduce the number of inferred GPs (Supplementary Information S9.1.2.). Our results suggest that ancestry proportions *Q* should complement the interpretation of *K* because *K* alone does not characterize population substructure well. Examining both statistics in the populations where *K*=3, we can interpret GPs in São Tomé as putative distinct demes with some degree of reproductive isolation, whereas in BCI, we conclude that GPs are mostly a result of allele frequency gradients (Table 11).

Biotic and abiotic determinants of within-population spatial genetic structure

The observed FSGS patterns in our study can be explained through a series of factors, including topographic complexity, seed and pollen dispersal features, biogeographic history and, potentially, microenvironmental adaptation.

At SSRs, FSGS was in the range expected for species with outcrossing or mixed mating systems (*Sp* from 0.0126 to 0.0372) and animal- or gravity-mediated seed dispersal (*Sp* from 0.0088 to 0.0281, (Vekemans & Hardy, 2004; Hardy et al., 2006; Debout, Doucet, & Hardy, 2011), in agreement with *S. globulifera*'s life history traits (Aldrich, Hamrick, Chavarriaga, & Kochert, 1998; da Silva Carneiro, Sebbenn, Kanashiro, & Degen, 2007). *Sp* at maternally inherited plastid DNA was generally an order of magnitude greater than at biparentally inherited SSRs. This pattern suggests that, among other factors, restricted seed dispersal shapes FSGS in *S. globulifera* whereas pollen is the long-distance component of gene flow (see Budde, González-Martínez, Hardy, & Heuertz, 2013), a typical pattern in tropical trees (e.g., Hardy et al., 2006; Ndiade-Bourobou et al., 2010).

The factor that most clearly co-varied with FSGS was altitudinal sampling range: stronger FSGS was observed in populations sampled in more prominent altitudinal gradients, specifically, in African populations with large altitudinal gradients (>350 m in Cameroon, >1200 m in São Tomé). Steep topography is known to restrict the mobility of animal species (Giordano, Ridenhour, & Storfer, 2007; Pérez-Espona et al., 2008; Storfer, Murphy, Spear, Holderegger, & Waits, 2010), thus reducing gene flow and increasing the genetic structure of the plants that these animals disperse (Dyer, Chan, Gardiakos, & Meadows, 2012; Côrtes &

Uriarte, 2013). This mechanism could partially explain the strong FSGS in African populations of *S. globulifera*. In addition, the complexity of habitats and the vegetation associated with such gradients could have favoured microenvironmental adaptation of *S. globulifera* (see below) and/or specialization of its dispersers, in terms of behaviour or community composition (Canova, 1993; McCain, 2005), restricting seed dispersal. Unfortunately, no data on the precise composition of disperser communities or the behaviour of *S. globulifera* dispersers are available for our study populations. On the other hand, large-scale differences in disperser communities between continents can contribute to explaining the observed pattern. Bats (*Artibeus spp.*) or tapirs (*Tapirus terrestris* and *Tapirus bairdii*), endemic to the Neotropics, can disperse propagules from hundreds of meters to several kilometres (Morrison, 1978; Morrison, 1980; Miller & Dietz, 2006; Ortega & Castro-Arellano, 2001; see Table 1). Both bats and tapirs (the latter not present in BCI) can concentrate a wide spectrum of seed genotypes at their feeding roosts or latrines, respectively, promoting seed shadow overlap and thereby, decreasing FSGS (Aldrich, Hamrick, Chavarriaga, & Kochert, 1998; Giombini, Bravo, & Tosto, 2016). Bat-mediated seed dispersal of *S. globulifera* has only been reported in American populations, although frugivorous bats occur also in Africa. Hornbills could constitute an equivalent long-seed disperser in Africa but they putatively regurgitate the seeds of *S. globulifera* (Forget et al., 2007), which could reduce dispersal distances compared to endozoochory (Schupp, 1993; Whitney et al., 1998; Holbrook & Smith, 2000).

The two Cameroonian populations Mbikiliki and Nkong Mekak provided an interesting example of biogeographic history shaping within-population structure. In these populations, we observed the same associations between GPs and plastid haplotypes, an evidence of preferential reproduction within GPs (Wahlund effect). Such cyto-nuclear associations reflect the sympatric occurrence of differentiated lineages. An allopatric differentiation of such lineages is most commonly proposed, for instance in distinct refugia where rainforest species persisted during the dry and cold periods of the Pleistocene (Maley, 1996). Cameroonian *S. globulifera* lineages now co-occur in the Ngovayang massif, a region that corresponds to a proposed Pleistocene refuge area (Maley, 1996; Tchouto, de Wilde, de Boer, van der Maesen, & Cleef, 2009; Gonmadje, Doumenge, Sunderland, Balinga, & Sonké, 2012). A comparison with plastid haplotypes widely sampled across Lower Guinea (Gabon and Cameroon) suggested a restricted distribution of the concerned lineages, in agreement with previous suggestions of local population persistence and absence of evidence for pronounced range shifts in *S. globulifera* (Budde, González-Martínez, Hardy, & Heuertz, 2013). The cyto-nuclear disequilibria are thus unlikely to reflect insufficient time for genetic homogenization after colonization (e.g., Kremer et al., 2002; Goslee & Urban, 2007). Rather, we believe they reveal a persistent historical or adaptive pattern maintained by partial reproductive isolation or assortative mating (Asmussen, Arnold, & Avise, 1989; Fields, McCauley, McAssey, & Taylor, 2014).

Adaptation to locally heterogeneous habitats, e.g., to specific soil properties or associated vegetation, could also explain the genetic clustering and altitudinal stratification of GPs in our study (Andrew, Ostevik, Ebert, & Rieseberg, 2012; Misiewicz & Fine, 2014; Brousseau, Foll, Scotti-Saintagne, & Scotti, 2015; see Scotti, González-Martínez, Budde, & Lalagüe (2015) for an overview). In Paracou, where soil moisture content decreases with relative elevation, Allié and collaborators (2015) showed that the common *S. globulifera* morphotype is associated with

moist valley bottoms whereas the alternative morphotype preferentially grows in the upper part of slopes. The morphotype – haplotype association in our data and the morphotype – GP association based on genic SSRs (Supplementary Information S9.1.4.) indicates that differential habitat preferences are paralleled by genetic differentiation in *S. globulifera* in Paracou. Similarly, local-scale genetic differentiation in the Neotropical tree *Eperua falcata* has been attributed to edaphic specialization (Audigeos, Brousseau, Traissac, Scotti-Saintagne, & Scotti, 2013; Brousseau, Foll, Scotti-Saintagne, & Scotti, 2015). Signals of microenvironmental selection can be detected in neutral markers (Shafer & Wolf, 2013; Misiewicz & Fine, 2014; Budde et al., 2017) when they are linked to makers under selection, or when emerging reproductive barriers foster linkage among physically unlinked markers (Nosil, Egan, & Funk, 2008; Feder, Egan, & Nosil, 2012). Adaptive divergence can thus potentially lead to cyto-nuclear disequilibria resulting in patterns like those observed in the Cameroonian populations.

Assortative mating can interact with other forces to enhance genetic structure, potentially resulting also in significant inbreeding. Mass flowering events and asynchronous flowering promote pollinator movements between flowers of the same tree (see Augspurger, 1980; Loveless & Hamrick, 1984; Aldrich & Hamrick, 1998) leading to temporal assortative mating (Hendry & Day, 2005). This is likely in *S. globulifera* which may produce up to 200 open flowers per tree each day and for which unsynchronized flowering is suspected (Bittrich & Amaral, 1996; Degen, Bandou, & Caron, 2004; da Silva Carneiro, Degen, Kanashiro, de Lacerda, & Sebbenn, 2009). Further, agamospermy (seed development without fertilization) has been observed in other Clusiaceae (e.g., in the genera *Garcinia* and *Clusia* (Maguire, 1976*;* Richards, 1990*;* Sweeney, 2008) and leads to groups of genetically identical individuals as observed in two of the studied *S. globulifera* populations. However, additional data is needed to determine whether the observed clonality is due to agamospermy or to root suckers.

5.2. Large-scale genetic structure in *S. globulifera*, demographic history and adaptive evolution (Study II)

In the present section we have addressed the spatial patterns of genetic divergence of *S. globulifera* across continents to unveil the route of its range expansion and the signals of adaptation on loci from African gene pools. We performed a gene pool delimitation based on SNPs obtained through GBS, analysed their phylogenetic relationships and presented the most probable origin of colonization events from Africa to America and, finally, unveiled loci bearing signatures of selection in African locations and gained insights into the drivers of adaptation in *S. globulifera*.

5.2.1. Spatial genetic structure of *S. globulifera* across continents (Study II, part 1)

Geographical patterns of genetic structure

Sequencing of nine widely distant locations of *S. globulifera* from two continents for 4921 SNPs revealed nine differentiated gene pools, of which seven perfectly matched geographical locations. Entropy also discriminated the *terra firme* morphotype in French Guiana and

delimited two gene pools within the common morphotype. Only one GP was represented by too few individuals (three individuals from MB and NK) to consider it in further analyses.

Overall, we found a strong genetic structure as expected for a widespread species with a large distribution range involving many biogeographic barriers and great species age (Dick, Abdul-Salim, & Bermingham, 2003, Dick & Heuertz, 2008, Budde, González-Martínez, Hardy, & Heuertz, 2013) and possibly highly ecologically adapted to local environments as species strategy, as evidenced by the high number of ecotypes attributed to this species (see Sabatier et al., 1997, Sabatier & Molino, 2001, Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007, Dick & Heuertz, 2008, Budde, González-Martínez, Hardy, & Heuertz, 2013, Allié et al., 2015, Schmitt et al., 2020).

Setting the gene pool from Cameroon (MN) as a starting point, the geographic pattern of genetic divergence indicated an increase of genetic differentiation among locations following a geographical line starting from the southern populations of the Gulf of Guinea, in continental Africa, towards the North- and the West-, then to São Tomé island and, finally, the American gene pools as a group. Genetic similarities in African GPs decreased when geographic distances increased from south to north locations (with the island as the end of the path) in agreement with a model of isolation by distance (IBD), which is expected when limitation in dispersal exists (Wright, 1943; Vekemans & Hardy, 2004). However, based on the NJ tree based on Nei´s D, a stepping-stone pattern (i.e., individuals which are distributed in groups and may migrate to nearby groups instead of being continuously spread as in IBD, Kimura & Weiss, 1964) may also underlie those genetic differences, as this metric of genetic distance can be linearly related to geographical distance in the stepping-stone model (Nei, 1972) which roughly corresponded to the trend found between geographical and genetic distance among our African gene pools. Nei´s D can also be linearly related to evolutionary time if the rate of gene substitutions per year is constant (Nei, 1972), which also seems to occur as evidenced through the similarities in the four tree topologies analysed in this study.

The four African gene pools recovered by Entropy, including São Tomé, mirrored spatially those found by Budde, González-Martínez, Hardy, & Heuertz, (2013), who demonstrated that biogeographic features (the ocean, mountain chains like the Cameroon Volcanic line and dryer habitats like the Dahomey Gap) influenced the genetic structure in African populations. Particularly, a clearly differentiated gene pool was expected for São Tomé due to its isolated position, as this location corresponds to an oceanic island in the Gulf of Guinea, emerged $\geq$ 13Ma and ca. 242 km away from the African continent at its closest point (seas ca. 1800 m deep), and never connected to the mainland (Jones, 1994, Lee, Halliday, Fitton, & Poli, 1994, Meyers, Rosendahl, Harrison, & Ding, 1998). Such a situation reinforces the evidence that *S. globulifera* is able to perform marine dispersal, already stated by Dick & Heuertz (2008). Strikingly, the allocation by Entropy of this isolated gene pool in the same group as Neotropical locations for K=2 in comparison with the other African gene pools, closer in distance, points out that São Tomé was probably a middle point in the *S. globulifera* overseas dispersal between continents.

It was already clear that the *terra firme* morphotype of *S. globulifera* (*S. sp*1.) is morphologically and ecologically different from the common flood-tolerant *S. globulifera*

morphotype (Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007, Allié et al., 2015, Schmitt et al., 2020, Tysklind et al., 2020; Schmitt, Tysklind, Hérault, & Heuertz, 2021). Also, genetic differentiation between morphotypes had already been detected by Olsson et al., (2017), Torroba-Balmori et al., (2017) and Schmitt, Tysklind, Hérault, & Heuertz, (2021). However, our study evidenced that strong genetic differentiation exists not only between morphotypes, but also within the common morphotype (*S. globulifera*) between locations from the same region (French Guiana). Schmitt, (2020) and Schmitt, Tysklind, Hérault, & Heuertz, (2021) identified three gene pools in Paracou based on 454,262 biallelic SNPs, apparently coincident with the GPs presented here. In their study, they use the pool of sequences from French Guiana developed through GBS within the present study, among other sources, as genome reference to design a novel gene capture experiment to obtain SNPs in their individuals. As we have demonstrated, the analysis based on our 4921 genomic SNPs provided results with high enough resolution to clearly discriminate among the three gene pools. Congruently, Schmitt, (2020) and Schmitt, Tysklind, Hérault, & Heuertz, (2021) reached a similar resolution with a larger data set.

Thanks to the wide range of sampling in *S. globulifera*, we have been able to show high levels of genetic differentiation present in French Guiana, not only for the alternative morphotype *S. sp.1*, but for the three gene pools in this region compared to other widely distant locations across both continents. Specifically, the high levels of genetic differentiation in *S. sp.1* regarding *S. globulifera* in French Guiana and Ituberá, as conspicuous as the genetic differentiation found among the other gene pools in America, support considering *S. globulifera* as a putative species complex in the same way as other tropical genera (Schmitt et al., 2020, Schmitt, Tysklind, Hérault, & Heuertz, 2021; see Pinheiro, Dantas-Queiroz, & Palma-Silva, 2018).

Phylogeographic history: Africa

All our species trees in SNAPP pointed to an ancient genetic divergence between continents in the history of *S. globulifera* in line with Dick, Abdul-Salim, & Bermingham, (2003), who suggested that *S. globulifera* achieved its contemporary distribution (i.e., its presence in both continents) early in its history of expansion (estimated divergence time 17.36 ±1.53 Ma). Thus, the current genetic structure of *S. globulifera* in Africa originated subsequently to the colonization of America (see also Budde, González-Martínez, Hardy, & Heuertz, 2013).

The order of divergence events leading to the recognized GPs in all models presented São Tomé as the most ancient African GP from all our locations. Particularly, M1 in SNAPP evidenced closer phylogenetic relationships between São Tomé and American gene pools, suggesting that colonization of the island and its GP differentiation could even date from the period where colonization of America occurred (first American fossils date from 18-15 My ago in Mesoamerica and ≈15 My ago in South America, São Tomé emerged ≥ 13My ago, Lee, Halliday, Fitton, & Poli, 1994; Dick, Abdul-Salim, & Bermingham, 2003; see Fig. 9.A in green). The third possibility inferred based on M1 would be that the genetic differentiation in São Tomé had been more related to the genetic differentiation processes within the American continent, which would have involved higher levels of gene flow events with the American continent compared to Africa (Fig. 9.A in red).

Additionally, considering the African origin of our species, drift levels of São Tomé in TreeMix situated this location in the middle of both continents, suggesting that sweepstakes dispersal to São Tomé originated from Benin (which is the previous African location with more similar levels of drift with respect to South-West Cameroon GPs) and discarded connections between Lower Guinea populations and America (see Budde, González-Martínez, Hardy, & Heuertz, 2013 for this alternative hypothesis). Thus, our study supports this island as the most probable origin of the colonization events from Africa to America and suggests that colonization of São Tomé could originate from Benin, or from other nearby populations more genetically related to Beninese populations in the West African rainforest (Upper Guinea), the Dahomey Gap (a ca. 200 km area covered by savanna vegetation from eastern Ghana to western Nigeria, separating West African from Atlantic Central African rainforest in Lower Guinea) or Nigeria (see Daïnou et al., 2010 and Demenou et al., 2020 for colonization sources of the Dahomey Gap region).

The branch lengths reflecting divergence time in our SNAPP trees indicated that the genetic divergence between Benin (at the Dahomey Gap) and the other locations in Lower Guinea was the earliest among our mainland locations in Africa, while divergence from Korup was clearly posterior. Signatures of divergence between Southwest Cameroon and Gabon were more imprecise, indicating either a relatively recent divergence or no divergence at all.

Most probably, the ancient divergence events experienced by Benin and Korup were related to the impact of the Pleistocene climatic fluctuations on our species, since numerous studies have shown that intraspecific genetic structure of many rainforest species in the Guineo-Congolian rainforest mainly emerged during the Pleistocene (2.6-0.01 My ago) because of rainforest degradation and re-expansion caused by glacial and interglacial periods (see for example Hardy et al., 2013; Dauby et al., 2014; Piñeiro, Dauby, Kaymak, & Hardy, 2017; Helmstetter, Béthune, Kamdem, Sonké, & Couvreur, 2020). Such an impact has also been revealed as a driver for species divergence (Duminil et al., 2015), although diversification of tree species in Africa has mainly been detected before the Pleistocene (Faye et al., 2016a; Migliore et al., 2019).

Particularly, the divergence we detected in Korup would reasonably correspond to that period, as that gene pool was situated on the west side of the Cameroon volcanic line (CVL, a geographic feature formed by a long chain of volcanoes and seamounts that extends 1600 km from the archipelago in the Gulf of Guinea to mainland Africa along the Cameroon-Nigeria border formed 30 My ago; Meyers, Rosendahl, Harrison, & Ding, 1998) near a postulated Pleistocene forest refuge (Maley, 1996; Maley & Brenac, 1998; see Budde, González-Martínez, Hardy, & Heuertz, 2013). Glacial refugia, under the refuge theory developed with reference to the Pleistocene climatic fluctuations, refers to regions where species could have persisted despite the climate changes of the Plio-Pleistocene glaciations (Bennett & Provan, 2008). Following this theory, an accumulation of endemic variants is expected in the refugia and patterns of lower diversity outside refugia because of loss of alleles caused during the re-colonisation processes (founder effects; Hewitt, 2000; Petit et al., 2003). Following this theory, the existence of refugia for rainforest trees has been supported by the discovery of endemic genetic variants in many tropical tree species in the CVL, including *S. globulifera* (Budde, González-Martínez, Hardy, & Heuertz, 2013; Hardy et al., 2013; Demenou et al., 2020).

Remarkably, although bottleneck signals attributed to forest reduction during the last glacial maximum (LGM, approximately 21 Kya, Gibbard, Ehlers, & Hughes, 2017) have been found in Benin and Korup (Budde, González-Martínez, Hardy, & Heuertz, 2013), our estimated divergence times indicated that both populations diverged in different moments (i.e., one of them before the other), possibly related to different glacial cycles during the Pleistocene, although both gene pools were then affected by LGM. In the same line, although LGM has been analysed as an important influence on the genetic structure within species (Duminil et al., 2015; Faye et al., 2016b), other studies have also evidenced pre-LGM divergence times shaping the genetic structure in different species (Piñeiro, Dauby, Kaymak, & Hardy, 2017; Demenou et al., 2020; Ndiade-Bourobou et al., 2020).

During successive glacial periods in the Pleistocene, the Dahomey Gap has been affected by important breaks in continuity between adjacent forest blocks (Upper and Lower Guinea), which caused range shifts, fragmentation and demographic changes in species (Salzmann & Hoelzmann, 2005; Anhuf et al., 2006; Demenou et al., 2020). Those oscillations are believed to have shaped conspicuous biogeographic patterns (i.e., sub-centres of endemism and high levels of species diversity in Upper and Lower Guinea for a large number of species; White, 1979; Linder, 2001) and left similar patterns of genetic differentiation between both blocks in many tree species, including *S. globulifera* (Budde, González-Martínez, Hardy, & Heuertz, 2013; Hardy et al., 2013; Demenou et al., 2020). In our species, it is possible that the glacial periods of Pleistocene may have isolated Upper Guinea lineages of *S. globulifera* earlier in its history. Similarly, in Upper Guinea early divergent lineages (pre-LGM: ca. 0.5 My ago) for vertebrates (chimpanzees and woodpeckers; Bjork, Liu, Wertheim, Hahn, & Worobey, 2011; Fuchs & Bowie, 2015) have been detected at both sides of the Dahomey Gap.

In contrast, we suggest that the divergence between Gabon and South-West Cameroon is posterior to Pleistocene climatic events. Unlike Benin and Korup, more severely affected by past climatic oscillations, Atlantic Central African forests in southern parts of Lower Guinea could have offered a relatively more stable habitat for our species during Pleistocene climatic oscillations, similarly observed in *Anthonotha macrophylla* and *Distemonanthus benthamianus* (Demenou et al., 2020). Budde, González-Martínez, Hardy, & Heuertz, (2013) evidenced that S. *globulifera* was possibly less affected by the LGM in this region and recovered signs of its long-term persistence in a wide geographic range within the region, outside the postulated refugia. Dauby et al., (2014) also suggested western Cameroon and possibly a large region in Gabon as refugium areas for several tree species including *S. globulifera*, whereas Eastern areas of their distribution ranges (eastern Cameroon, north-eastern Gabon, Central African Republic and northern Congo) are believed to have shown generalized recolonization patterns. Moreover, results on relationships among African GPs suggested a decreasing south-north gradient of genetic diversity across the Gulf of Guinea, in mainland Africa, in line with the pattern found by Heuertz, Duminil, Dauby, Savolainen, & Hardy (2014), despite one of our northern locations occurred near the proposed refuge in CLV.

Overall, our findings suggest that our southern locations in Lower Guinea could be an ancestral population for our species, since long-term populations in climatically stable areas usually tend to harbour more diversity than others more severely impacted by past climate change (Svenning, Fløjgaard, Marske, Nógues-Bravo, & Normand, 2011; see for example Carnaval,

Hickerson, Haddad, Rodrigues, & Moritz, 2009; Faye et al., 2016b; Demenou et al., 2020). In this sense, Cowling et al., (2008) pointed out a larger presence of tropical broadleaf forests in Central Africa during the Pleistocene than previously expected based on the Refugia Hypothesis (Maley, 1996; Anhuf et al., 2006). The topology of relationships among mainland African populations, with a pattern of increasing divergence from southern to northern populations in the Gulf of Guinea is also congruent with such an hypothesis (see for example Garot, Joët, Combes, & Lashermes, 2019). Therefore, it would be interesting for future studies to identify the geographic distributions of differentiated *S. globulifera* clusters accross its continuous range in Lower Guinea, date their divergence events, and test the hypothesis of the southern region of Lower Guinea as an ancestral location for our species.

Phylogeographic history: The Neotropics

The sequence of divergence events estimated in our species trees in SNAPP for Neotropical gene pools supported a scenario where Ituberá presented an early divergence with respect to the other GPs with a significant time gap. The second split, relatively recent (roughly at the same level that southern populations in Lower Guinea), separated *S. sp.1* from the *S. globulifera* gene pools in French Guiana in most topologies, indicating that it is an evolutionarily distinct lineage. Both results were concordant with the main topology in TreeMix (Fig. 10B), which showed that the highest and second highest depth of divergence in Neotropical populations corresponded to Ituberá and *S. sp.1*, respectively (although the latter conspicuously less divergent). The last and most recent divergence event occurred between gene pools from Regina and Paracou, and it was surprising considering the short distance (ca. 140 km) and the absence of steep topography between locations (for example, distance between locations in South-West Cameroon and Gabon: ca. 500 km; see Guitet et al., 2013).

The ancient split of Ituberá within the Neotropics, largely divergent from Africa gene pools, indicated a long-term presence of *S. globulifera* in this area. Our result in this location, situated in the Atlantic Forest of Brazil, is congruent with the identification of this area as climatically stable during the climate oscillations of the Pleistocene (Carnaval & Moritz, 2008; Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009; Leite et al., 2016) although, because of that, the genetic diversity was not so high as we expected compared with other Neotropical GPs.

The inferred splitting time between PR and RG suggested a recent process of differentiation between gene pools. On one hand, RG was located within a putative refuge, climatically stable during the Pleistocene, while PR occurred within an area strongly disturbed during the Pleistocene and Holocene periods (see Dutech, Maggia, Tardy, Joly, & Jarne, 2003). Evidence of recolonization processed from putative refugia in French Guiana has been found in *Vouacapoua americana*, a shade-tolerant rainforest tree species (Dutech, Maggia, & Joly, 2000; Dutech, Maggia, Tardy, Joly, & Jarne, 2003). Also, Barthe et al., (2017) found signatures of stability or expansion in several rainforest species in French Guiana, probably occurring during or after the LGM, although in favor of stability in the case of both *Symphonia* morphotypes. However, other studies have revealed population expansion of our species in this region (Dick & Heuertz, 2008; Schmitt, 2020). Therefore, the genetic differentiation we have found could have been promoted by recent recolonization processes, keeping in mind that other factors might also be involved (e.g., landscape barriers, limitation in dispersal). On the other hand,

apparently, both *S. globulifera* gene pools in Paracou and Regina would correspond to two sub-morphotypes with differences in trunk morphology and topographic niches (Schmitt, 2020). In addition, Paracou and Regina occur in two regions within French Guiana ("east" and "west") with contrasted rainfall regimes, and local adaptation to environmental constraints has been found in those regions for S. *globulifera*, particularly related to water availability (Tysklind et al., 2020; Schmitt, Tysklind, Hérault, & Heuertz, 2021). Thus, the genetic differentiation could be also reflecting ecologically driven isolation processes (Feder, Egan, & Nosil, 2012). Moreover, both scenarios are not mutually exclusive. Therefore, further research would be needed to clarify the mechanisms behind this recent and strong divergence between these nearby locations within a relatively soft landscape (see Torroba-Balmori et al., 2017 for influence of altitudinal gradients on the SGS of *S. globulifera*) in a species capable to occur across large genetically homogeneous regions.

In contrast, we propose that the geographic origin of *S. sp.1* does not correspond to French Guiana, as its presence in this region seems more consistent with divergence in allopatry and secondary contact with GPs in French Guiana (see e.g., Torroba-Balmori et al., 2017) rather than with sympatric genetic differentiation due to segregation on different ecological niches (i.e., with gene flow between morphotypes, Feder, Egan, & Nosil, 2012). The reasoning behind this hypothesis is threefold: i) The genetic relationship between IT and SP is closer than previously expected based on the distance between locations: the *S. sp.1* gene pool is more similar to Ituberá than to gene pools from French Guiana and both gene pools seem to share a common ancestor based on TreeMix results; ii) PR and RG have differentiated more recently than SP and seem to share their own common ancestor, and iii) it is reasonable to think that the SP gene pool would present similar patterns to PR and RG regarding the genetic differentiation if SP would have experienced the same regional influences driving their divergence processes (especially if the divergence was driven by contraction-expansion dynamics) since both morphotypes present similar life-history traits. In this sense, Tysklind et al., (2020) showed that *S. sp.1* presented a habitat generalist behaviour related to habitats with contrasted levels of water availability.

Thus, we have found evidence that both morphotypes do not represent emerging divergent lineages (Feder, Egan, & Nosil, 2012) but sympatric occurrence of previously diverged lineages. Schmitt et al., (2020) showed that there was a niche overlap between both morphotypes related to decreasing water availability along topographic positions (i.e., from bottomland to plateau) and, in this sense, they hypothesised that such convergence could be derived from genetic exchange between both morphotypes. Based on that, in the absence of reproductive isolating mechanisms, it is still unclear if i) the demographic tendency between morphotypes will lead to a homogenization of gene pools and niches in time, or if ii) local adaptation to specific microenvironments will maintain the genetic divergence while helping to increase their respective adaptive potential to a wider range of microenvironments (Feder, Egan, & Nosil, 2012; Richardson, Urban, Bolnick, & Skelly, 2014; Schmitt et al., 2020).

Methodological considerations

Among the methodological considerations for our analysis, we want to highlight the differences reached in accuracy between tree topologies and branch lengths in SNAPP, where our three

models gave a good visual example of such variations. Accuracy problems in the three output parameters of SNAPP due to difficulties in convergence was already suggested in Drummond & Bouckaert (2015), who indicated their decrease in this order: tree topology, branch length and effective population size, as is also our case. Additionally, thanks to comparing different model settings, we have demonstrated that SNAPP allows us to test a variety of parameter specifications (including those usually used by default) and still gives robust and accurate estimates, providing the data contains enough information for the model.

Regarding TreeMix analyses, neither did a literature review provide information about the most ancestral location in Africa to root the analysis, nor did we have SNP data from a sister species of *S. globulifera* to use as an outgroup. Such a situation forced us to interpret the African topology as unrooted. However, as it is known that the presence of *S. globulifera* in Africa preceded the American colonization based on fossil data, Neotropical topology of *S. globulifera* showed the real population dynamics since it was rooted to the more ancient African locations.

Finally, not to account for gene flow in the estimation of relationships among gene pools, which were assumed to be isolated but could be affected by gene flow, may affect topology in coalescent-based methods (i.e., species tree methods; Solís-Lemus, Yang, & Ané, 2016). Although such inconsistency has not been tested in SNAPP, apparently the model may be affected and would recover different topologies than other methods that consider gene flow such as TreeMix (Thom et al., 2018). In our case, the main topologies in SNAPP and TreeMix were not mutually consistent regarding the branching order of the three gene pools from French Guiana, which presented evident signatures of admixture. However, our topology in TreeMix fitted well without adding any migration event among populations, which would have indicated more than one parental population for the gene pool affected. Overall, both models seemed to be affected by the levels of admixture because their topologies for those branches were not fully supported.

Generation of SNPs through GBS

Genotyping-by-sequencing (GBS) is a very useful genotyping tool which can provide many thousands of SNP in non-model species because marker discovery and direct genotyping can be performed without first sequencing the species genome (e.g., Pais, Whetten, & Xiang, 2017, Fernández-Mazuecos et al., 2018). Therefore, it gives the opportunity to study the population genetic structure, the potentially adaptive genetic variation, as well as the drivers for population divergence of non-model species (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013) such as *S. globulifera*.

Based on our experience, when the genetic differences among populations are too high, such as in our case, the efficiency of the method to generate markers decreases (see Parchman et al., 2012 as an example of the power of the GBS using a reference-based assembly for almost a hundred thousand markers). Thus, the number of SNPs we obtained was similar to those found in multispecies studies (see Mandeville, Parchman, McDonald, & Buerkle, 2015, Fernández-Mazuecos et al., 2018).

Our results (a few thousands of markers, many alleles fixed for the different locations) were probably affected by the difficulty to find enough homogeneous regions with a low number of

mismatches (the SNPs) in most individuals, due to large divergence, the absence of the homogenising effects of gene flow among so distant locations (among others, the ocean clearly is a strong geographical barrier to gene flow among our locations) and, possibly, local adaptation. Thus, our results evidence one of the potential drawbacks of restriction-digest based methods (as GBS) to generate markers using high-throughput sequencing technologies: a lower suitability of those methods to generate appropriate data sets for phylogenetic studies when groups included are highly divergent (McCormack et al., 2013). The reason behind is that long phylogenetic distances among groups are associated to higher frequency of mutations in the restriction sites, which results in failure to cut the DNA because the restriction enzyme does not recognise that site and this reduces the number of homologous fragments found for all individuals (i.e., allele dropout; McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Consequently, the marker yields are reduced.

Additionally, our process seems to have selected conservative regions within the genome and, consequently, many SNPs showed low variability within locations due to fixed alleles (see number of polymorphic sites in Table 13). Additionally, we observed extensive within-population heterozygosity in some markers that could be caused by paralogous copies that are cotransmitted, as negative values in the fixation index (F_{IS}) and *Ho* > *He* values (Table 13) would indicate a tendency to "clonal transmission" for these markers (i.e., without recombination, similar to asexual reproduction) that is usually related to higher levels of heterozygosity than expected (Stoeckel et al., 2006). In this sense, it is worth remembering that *S. globulifera* is an outcrossing species with tendency to non-random mating (i.e., leading to positive F_{IS} values, Torroba-Balmori et al., 2017).

5.2.2. Local adaptation of *S. globulifera* at continental scale in Africa (Study II, part 2)

Our objective was to gain insight into the genomic signatures of local adaptation in *S. globulifera* and identify some of the drivers that underlie the ability of our species to occur in a wide range of habitats. The combination of four methods to detect outlier loci in *S. globulifera* led us to identify 12 loci putatively under selection and significant SNP associations with influential environmental variables related to water stress and pH in the topsoil, which conspicuously differ across its African distribution range.

Environmental drivers of local adaptation in S. globulifera

We have found signatures of local adaptation in *S. globulifera* related to water availability, specifically to the ability of the species to cope with water stress, as the environmental variable related to the Locus 219 was BIO17 ("Precipitation of driest quarter"). Our work also suggests that adaptation to pH levels may play a role, since Locus 4871 was related to pH in the topsoil. Additionally, both environmental variables were involved in the greatest number of genetic-environmental associations regarding the 4 outlier loci found in all analyses. When considering the set of 12 loci significant in two or more tests, the most frequent environmental variables related were pH in topsoil and the "Aridity index", also related to water stress (it represents

moisture availability for potential growth of vegetation). Thus, the environmental variation covered in our sampling was wide enough to detect the adaptive differences that underlie the establishment of our species in differentiated climate regimes and soil conditions (see Wang & Bradburd, 2014; Nadeau, Meirmans, Aitken, Ritland, & Isabel, 2016).

Thanks to the geographical distribution of our sampled locations in mainland Africa, we could contrast the effect of different rainfall regimes across different parts of the range of our widely distributed species. According to the climate classification of West and Central Africa using the Köppen-Geiger system revised by Peel, Finlayson, & McMahon (2007), all our locations occurred under tropical climate, with two seasons. However, Benin and Gabon presented a tropical climate with a dry winter, while Korup, Mbikiliki and Nkong Mekak occurred under tropical monsoon climate (higher precipitation of the driest month compared to the other climate). In addition, our sampling crossed the Central African climatic inversion (i.e., the climate hinge, a latitude of N-S seasonal inversion, ca. 0-2º N), which separates boreal and austral climatic regimes with inverted seasons (Suchel, 1990). Thus, the climate presents a sunny main dry season (December to February) towards the North (i.e., Korup and Benin) while the cloudy dry season (June to August) helps to preserve high levels of humidity towards the South (i.e., Gabon; Suchel, 1990; Tsalefac et al., 2015). Mbikiliki and Nkong Mekak were situated also in the Northern Hemisphere, although near the transition area. Therefore, our sampling covered locations with different conditions for water stress in plants, not only due to different climates but also to cloud regimes and their effect on the evapotranspiration, with dry seasons more intense towards the North (for example, Benin was the most arid of our locations despite Gabon presenting the lowest precipitation of the driest quarter).

The seasonal inversion might be also a relevant driver to the local adaptation of *S. globulifera* to water availability. It has been reported that this seasonal inversion has an impact in the floristic composition on both sides of the climate hinge, stronger than the barrier effect that the CVL may produce (Gonmadje, Doumenge, Sunderland, Balinga, & Sonké, 2012). Also, this inversion has been pointed out as an important process shaping genetic structure and differentiation within plant species (e.g., Helmstetter et al., 2020; Ndiade-Bourobou et al., 2020) and, based on chloroplast markers, Heuertz, Duminil, Dauby, Savolainen, & Hardy (2014) evidenced that our species displayed a strong north-south differentiation pattern. Several non-mutually exclusive hypotheses have been proposed to explain the existence of this phylogeographic barrier within plant species (Hardy et al., 2013): (i) repeated isolation and recolonization of both sides during favourable periods along past climatic fluctuations during the Pleistocene (e.g., Ndiade-Bourobou et al., 2020); (ii) existence of a barrier to reproduction because of difference in flowering phenology due to inverted seasons (e.g., Ndiade-Bourobou et al., 2020) and (iii) limited colonization by migrants from opposite sides because of natural selection against them, since the environmental differences would lead to local adaptation (e.g., Faye et al., 2016b; Helmstetter et al., 2020). This last hypothesis remarks that the differences in the climatic conditions on both sides of the climatic hinge are particularly important with regard to the differences in water stress that plants may suffer (Hardy et al., 2013).

The relevance of limitation in water availability for *S. globulifera* related to ecological differentiation has been also highlighted in several studies. For example, Tysklind et al., (2020) showed that *S. globulifera* presents physiological adjustments to habitats with contrasted

conditions in soil water availability in French Guiana, with individuals from wetter provenances performing worse in harsher conditions. Similarly, Schmitt et al., (2020), Tysklind et al., (2020) and Schmitt, Tysklind, Hérault, & Heuertz, (2021) evidenced that both *Symphonia* morphotypes in French Guiana present differentiated patterns of microenvironmental distribution as well as morphological and functional responses related to water availability. Finally, Budde, González-Martínez, Hardy, & Heuertz, (2013) pointed out a possible differentiation of putative swamp and *terra firme* morphotypes in Africa. Overall, drought tolerance seems a relevant driver for local adaptation in our species at small and large geographical scales.

Regarding the influence of pH in the topsoil in our species, our locations in mainland Africa belonged to two of the seven soil regions of Africa: soils from Sahel and Savannah (i.e., Benin), characterized by a good drainage and thin layer of organic matter, and soils from forests (i.e., the rest of locations), poor in nutrients, acidic, with high rates of organic matter decomposition and heavy leaching (Jones et al., 2013). Although all locations occurred within a region with a generalised pattern of very acidic soils, common in areas of high rainfall (Jones et al., 2013), both locations in South-Western Cameroon presented very acidic soils (pH 4.5 - 5.5) while in the other locations there was a mixed presence of both very acidic and acidic to neutral (pH 5.5 - 7.2) soils (see Fig. 18 and Table S9.3.1.1.).

Thus, the genetic-environmental association found with pH may be explained because of the important effect of acidic soils ($pH \geq 5.5$) in plants related to deficiency in nutrients (especially phosphorous) and toxicity by metals such as aluminum (main limiting growth factor in acidic soils, it inhibits root growth and its activity, causing drought and mineral deficits to the plant), manganese (it produces chlorosis, reduced growth and damaged leaves), and iron (it affects growth and development and can cause cell death; Kochian, Hoekenga, & Piñeros, 2004; Gupta, Gaurav, & Kumar, 2013; Anjum et al., 2015). In such harsh environments, the plants need to develop tolerance mechanisms to survive. In the case of aluminum toxicity, many tolerance mechanisms are focused on the root, among which ABC transporters also play a role (Kochian, Piñeros, Liu, & Magalhaes, 2015; Bojórquez-Quintal, Escalante-Magaña, Echevarría-Machado, & Martínez-Estévez, 2017). The locus annotation that was obtained for Locus 4685 (related to climate aridity and the pH in the topsoil) indicated location within the region of an ABC transporter protein (ABC transporter G family member 35 or ABCG35). This family of proteins works as active membrane transport proteins in the cells, acts in many physiological functions related to plant development, fitness and survival, and presents highly conserved sections (Gräfe & Schmitt, 2020). In particular, ABCG35 is expressed in the roots and is involved in root exudation of phytochemicals (which have a role in the rhizosphere interactions and helps the root to penetrate the soil) and in cadmium response related to heavy metal resistance (Kim, Bovet, Maeshima, Martinoia, & Lee, 2007; Gräfe & Schmitt, 2020; Badri et al., 2008). Altogether, these results indicate that pH in the topsoil is a potential predominant driver of selection in our species in Africa and point to Locus 4685 as a good candidate for being under selection, even when its strong genetic structure could have led to classify it as a false positive. As acidic soils are very frequent in tropical rainforest around the world (von Uexküll & Mutert, 1995), it is probable that tolerance to acidic soils also plays a selective role in Neotropical populations of *S. globulifera*.

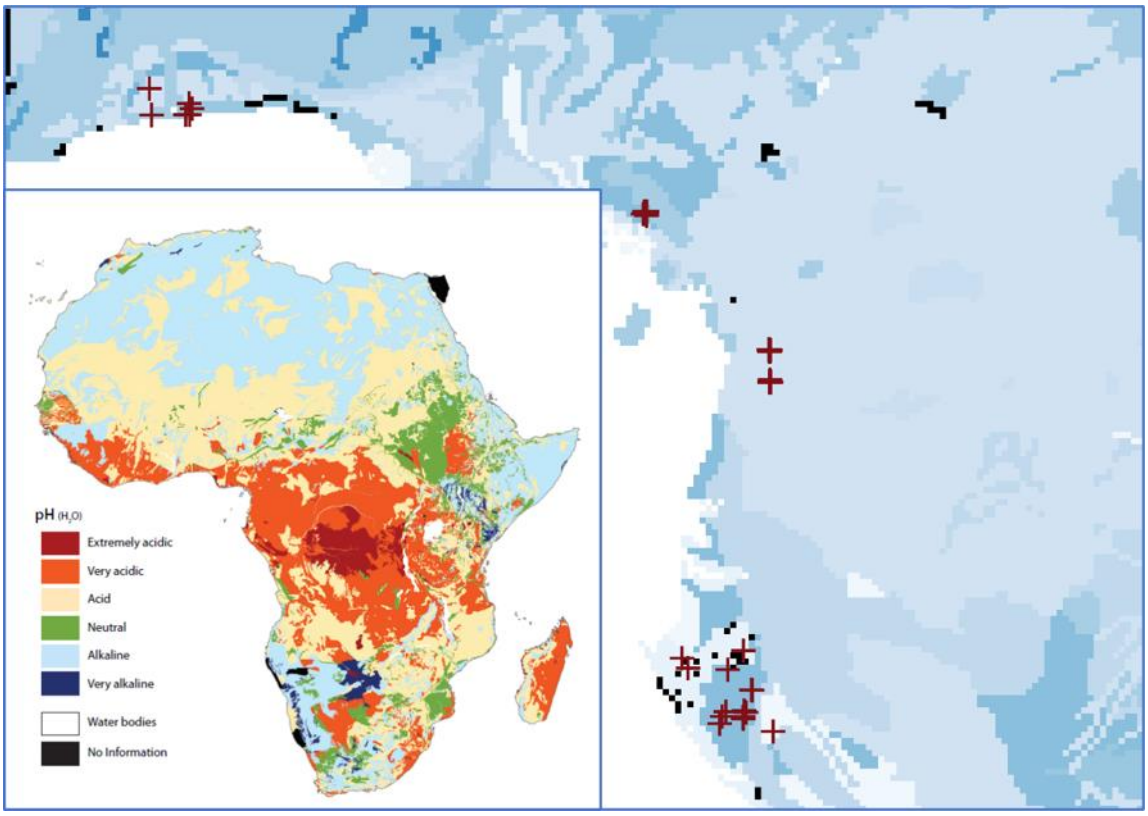


Figure 18. General map (reproduced from Jones et al., 2013) and detailed map of pH values in the topsoil. Darker blue shading indicates higher pH values. Rock, water or missing data in black. Individuals indicated as crosses.

Methodological considerations

The demographic history of populations (e.g., range contractions and expansions, barriers to gene flow, etc) affects the genetic structure of populations and may confuse the signals of local adaptation generating confounding effects (Nadeau, Meirmans, Aitken, Ritland, & Isabel, 2016). The demographic history and geographical barriers to gene flow have contributed to shape a strong genetic structure in our species. Moreover, limited gene flow between regions because of isolation by distance probably has helped to reinforce the genetic divergences creating a scenario where genetic differences increased with distance (see Study II, part 1). In this situation, a false signature of local adaptation could arise if environmental variables also follow a gradient with distance because those circumstances could generate correlations between the variables and neutral alleles (i.e., spatial correlation; Meirmans, 2012; Wang & Bradburd, 2014). Thus, following Hoban et al., (2016), we used both most common approaches to deal with the confounding effects of demography: 1) demographic null models: BayeScan, BayeScEnv, and 2) models with null model that takes into account the neutral population structure in the data: BayPass. We also tested the correlation between the outlier loci and their associated environmental variables to discard such spatial correlations. Additionally, given that the overlap detected through significant loci from different statistical frameworks is never complete, that Bayesian models are sensitive to priors (especially in presence of not enough

informative data), and considering the high levels of allele fixation in our SNP dataset, we took a conservative approach to get a reduced but more robust set of outlier SNPs and only considered those loci significant in two or more tests as robust candidates for being under selection (Nadeau, Meirmans, Aitken, Ritland, & Isabel, 2016; Ruiz Daniels et al., 2019). Therefore, the use of our four different methods helped to maximize the detection of outlier loci, to reduce false positive signatures of selection in a complicated dataset with strong genetic differentiation, and to disentangle potential drivers of adaptive evolution in *S. globulifera.*

By taking the intersection of SNPs in all four tests we discovered two top-candidate SNPs, either under selection or linked to regions under selection, and two top environmental drivers putatively influencing the local adaptation of *S. globulifera* in Africa. Locus annotation allowed the identification of a third top-candidate loci, putatively involved in physiological functions related to soil pH. Moreover, the high frequency of associations between SNPs and both environmental pressures might be related to their influence on loci relevant for local adaptation but under weaker selection due to polygenic adaptation. Polygenic adaptation involves traits influenced by many loci presenting small allele frequency changes and may not be correctly detected by F_{ST}-based methods, especially under strong genetic structure among populations (Le Corre & Kremer, 2012) and its relevance in local adaptation processed has been highlighted, including its possible influence in forest trees and their climate-related traits (Pritchard & Di Rienzo, 2010; Le Corre & Kremer, 2012).

5.3. Genetic structure within the genus *Symphonia* in Madagascar (Study III)

Polyploidy in the genus Symphonia

The genotypes based on our newly developed SNPs, together with the FCM results, revealed that one of the drivers of the genetic differentiation within the Malagasy genus *Symphonia* was the existence of different ploidy levels. The analysed individuals from nGP1-rGP1 (i.e., the tetraploid gene pool detected based on both SSR and SNP data), which presented signals of five genotype categories in more than a half of the loci analysed (84 SNPs out of 144 SNPs), exhibited consistently a nuclear content approximately doubled the genome size of individuals from the other four gene pools detected based on SSR data (see Table 18).

It is important to keep in mind that ploidy is not always a property of the whole genome. Instead, it is possible to find different ploidy levels related to genes, scaffolds or chromosomes (Margarido & Heckerman, 2015, McKinney, Waples, Seeb, & Seeb, 2017). In this sense, Leitch & Bennett, (2004) indicated that the polyploid formation may lead to genome downsizing, and one of the mechanisms of such is eliminating specific DNA sequences. This could explain why we found some SNPs with diploid patterns for individuals from nGP1-rGP1, despite more than half of the SNPs and the FCM being congruent with a whole genome duplication with respect to the other gene pools detected using SSR markers (hereafter, rGP) and SNP markers (hereafter, nGP).

Overall, evidence pointed out a positive correlation between DNA content and a genome duplication in nGP1-rGP1, without any clear signal of downsizing, confirming that Malagasy

Symphonia is a polyploid group of species which includes *a priori* diploid and tetraploid species. Still, the 1C estimates should be always considered with caution in case genome downsizing has occurred after polyploid formation (Leitch & Bennett, 2004; Loureiro, Kopecký, Castro, Santos, & Silveira, 2007).

It is also worth mentioning that the FCM analysis was successful in spite of two possible drawbacks in the conditions of the tissues: i) the use of desiccated plant tissues (collected during two missions in 2013 and 2014, in difficult conditions for a proper dehydration, and then analysed in 2017), and ii) the secondary metabolites within tissues (particularly latex in the genus *Symphonia*), which might have interfered and compromised the FCM assays (Suda & Trávnícek, 2006, Bainard et al., 2011).

Genetic structure and phylogenetic relationships within the genus Symphonia

Based on our comparison between the similar genetic structures detected by both 20 SSR markers and 144 SNPs, we showed that our putative functional SNPs are good markers to characterize the genetic structure of the *Symphonia* genus in Madagascar, despite that SSR markers are frequently considered best for this kind of analysis because of their higher resolution power (Guichoux et al., 2011; Haasl & Payseur, 2011), especially when the SNPs numbers are moderate (<300 SNPs) (Putman & Carbone, 2014). The obtained resolution is congruent with Guichoux et al., (2011) and Haasl & Payseur (2011) who indicated that between 4(5)-12(15) times more SNPs than SSR markers would be needed to result in an equivalent power for the analysis of genetic structure. However, it is worth remembering that both types of markers can reach comparable resolution, or even a better resolution for SNPs, when divergence time among populations increases or targeted SNPs are selected based on informativeness (Rosenberg, Li, Ward, & Pritchard, 2003, Haasl & Payseur, 2011). In this sense, our SNP markers were highly informative because they were specifically selected to reflect variability among putative *Symphonia* species.

The genetic clustering results obtained with different markers increased the confidence on the identified genetic structure of Malagasy *Symphonia*, as both sets of markers with different advantages and shortcomings (e.g., SSR: homoplasy; SNPs: biallelic nature) yielded rather similar results (see Zimmerman, Aldridge, & Oyler-Mccance, 2020). Also, even though their number was appreciably lower compared with the thousands of SNPs developed through the genotyping-by-sequencing (GBS) technique (see Study II, part 1), the set of SNPs added useful information for the analysis of genetic structure in *S. globulifera*.

The STRUCTURE analysis based on SNP markers revealed the existence of three main gene pools supporting the expected genetic differences among *S. globulifera*, tetraploid Malagasy *Symphonia* and diploid Malagasy *Symphonia*. However, the levels of mixed ancestry among those three gene pools displayed by individuals from nGP3 were unexpected. Also, considering the location of nGP3 in the middle of *S. globulifera* and both the tetraploid and diploid Malagasy GPs in the NJ trees, nGP3 might represent a middle point of connection between Malagasy *Symphonia* species (both polyploids and diploids) and *S. globulifera* along their speciation history and could be a starting point to disentangle the phylogenetic relationships within the genus *Symphonia*.

We also observed a significant genetic substructure within each of these three gene pools, but apparently driven by different evolutionary forces. Within the tetraploid gene pool nGP1-rGP1, the ancestry distribution of the two detected clusters among the individuals was in general relatively even, although there were individuals with high proportions of one of the gene pools. Such spatial genetic structure may be caused by the sympatric occurrence of differentiated lineages (e.g., Torroba-Balmori et al., 2017), or alternatively, it could indicate hybridization among different species (e.g., Burgarella et al., 2009). Diploid Malagasy *Symphonia* revealed four gene pools with almost no signs of admixture using both sets of markers, even when those gene pools co-occurred in sympatry, broadly pointing out to strong barriers to gene flow among different species. Finally, the analysis on *S. globulifera* resulted in the same genetic differentiation between continents, where BN and ST were strongly genetically related to the Neotropical populations, a phylogeographic conclusion already discussed in Study II, part 1.

A strong evolutionary divergence among genetic groups can be inferred from our two independent sets of markers. Both resulted in similar clustering patterns, which also evidenced absence or scarcity of gene flow among clusters. Such a high divergence among nGPs was corroborated by their high genetic distances, especially considering the genetic distances among Malagasy clusters with regard to those presented among *S. globulifera* populations. Moreover, the relationships among gene pools displayed by the NJ tree are an important step for achieving comprehensive insight into the evolutionary divergence among groups (Kalinowski, 2011; Carstens, Pelletier, Reid, & Satler, 2013). The NJ tree showed two broad genetic clusters, Madagascar and S. *globulifera*, congruent with the STRUCTURE clustering. However, the genetic differentiation of polyploid nGP1 with respect to the other Malagasy gene pools was surprisingly low, especially compared to the unexpected high genetic distances shown by nGP2 and nGP5. The case of nGP1 could be influenced by the use of the small subset of 53 diploid SNPs, which does not sufficiently represent the enormous genetic difference of the gene pool due to polyploidy. Another explanation could be a 'fairly recent' event of polyploid formation in nGP1. Polyploidization can result in rapid isolation and major structural changes in the new species in a few generations (Soltis & Soltis, 2009) and, for that reason, might not involve enough time in isolation to detect high levels of genetic differentiation among groups. On the other side, both nGP2 and nGP5 were almost as distant from other Malagasy gene pools as nGP1, nGP3 and nGP4 were from *S. globulifera*, indicating the extreme genetic dissimilarities of those former gene pools with respect to the others.

Nei´s D is a metric of genetic distance roughly linear to evolutionary time, assuming a drift-mutation equilibrium (Nei, 1972). As we were working with putative functional markers, it is likely that natural selection has counteracted the effects of genetic drift on GPs and, thus, has also influenced the genetic dissimilarities found. For this reason, the high genetic distances discovered among GPs cannot be directly connected to long divergence times. Nevertheless, considering the rather linear shape of the unrooted NJ tree, for future studies it could be interesting to rely on outgroups to test the hypothesis of nGP2 and nGP5 as the most recent gene pools in the genus *Symphonia* in Madagascar because of their divergence with respect to *S. globulifera* or, alternatively, as the oldest gene pools within the genus *Symphonia*. Moreover, it would be interesting to test if the divergence of nGP1 was more or less ancient than the

aforementioned gene pools, based on its higher genetic similarities with nGP3 and nGP4 and considering the levels of admixture between nGP3 and nGP1.

Regarding *S. globulifera*, STRUCTURE analysis and genetic distances revealed a clear differentiation with respect to Malagasy *Symphonia*. The phylogenetic relationships were congruent and very much expected based on the GBS analysis in Study II, part 1. Also, differences between GBS- and functional SNPs-based analyses were most probably due to the much lower number of markers used in the present study, an already known effect (Kalinowski, 2002) also evidenced when we compared the better supported 124-SNPs tree against the 53-SNPs tree. It is worth remembering that in our analysis of *S. globulifera* we included populations from two distant continents, which diverged ca. 17 Ma ago (Dick, Abdul-Salim, & Bermingham, 2003). The dataset of individuals also encompassed the presence of contrasted morphotypes derived from genetic differences within the species (see Study II, part1, Abdul-Salim, 2002; Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007; Torroba-Balmori et al., 2017) which could be a signal of cryptic species (Duminil & Di Michele, 2009; Schmitt, Tysklind, Hérault, & Heuertz, 2021). However, based on our SNP markers, the very low genetic distances among *S. globulifera* populations when compared to the genetic distances among Malagasy gene pools (which indeed encompass different species) seems to point out to the integrity of *S. globulifera* as a single species in agreement with its current taxonomic classification. Although these results are apparently contradictory regarding those from Study II, part 1, they are probably due to the lower amount of SNPs used in this study, which seems not to be enough to show the genetic structure in the same degree of detail.

Testing species delimitation and insights into drivers of radiation in Malagasy Symphonia

A starting point to delimit species at an exploratory stage, when there is no *a priori* partitioning of samples to lineages, is to carry out a population genetic structure analysis to assign individuals to putative groups (i.e., species discovery methods, see Ence & Carstens, 2011; Carstens, Pelletier, Reid, & Satler, 2013; Leavitt, Divakar, Crespo, & Lumbsch, 2016). STRUCTURE has been suggested as one of the tools to perform such an analysis (Carstens, Pelletier, Reid, & Satler, 2013); Choi, 2016; Leavitt, Divakar, Crespo, & Lumbsch, 2016), since this software allows to carry out a "blind" analysis to delimit gene pools. Therefore, this approach can reveal species or groups of closely related species without a priori classification of individuals, minimizing Hardy–Weinberg disequilibrium within clusters for a given partitioning level and representing the ancestry of individuals and the levels of admixture for that clustering solution (Pritchard et al., 2000). Minimizing admixture is a major consideration when delimiting species following the biological species concept (Mayr, 1963).

However, there were two shortcomings in the use of this general approach. On one hand, the temporal divergence among groups is not estimated and, consequently, there is not any measure of the evolutionary relationships among clusters (Kalinowski, 2011; Carstens, Pelletier, Reid, & Satler, 2013). In our case, that situation was overcome with the NJ analysis. On the other hand, the taxonomic relevance of the population genetic structure inferred with this kind of analysis is unknown, as it could reveal intraspecific population structure below the species level due to the efficiency of the software for a given marker set (Chen, Durand, Forbes, & François, 2007; Torroba-Balmori et al., 2017) or, alternatively, lineages above to the species level.

However, we solved that question by contrasting the delimitation of species/population boundaries based on two sets of molecular markers against genome size estimates and other non-genetic characters such as morphology and geographical distribution. Such an approach is methodologically advised to achieve accurate classifications of individuals to their correct species (see Dexter, Pennington, & Cunningham, 2010; Duminil & Di Michele, 2009; Carstens, Pelletier, Reid, & Satler, 2013).

First, the genetic structure of Malagasy individuals already gave some hints on the taxonomic relevance and the evolutionary forces shaping the genetic differentiation. In the first place, the discovery of a tetraploid gene pool clearly delimited a species or group of species within the genus and, remarkably, revealed one of the drivers of the radiation of the *Symphonia* genus in Madagascar: polyploidy. This is an important driving force of speciation in angiosperms which may arise from hybridization between different species (allopolyploidy) or from the fusion of unreduced gametes within a species (autopolyploidy), multiplying complete sets of chromosomes in the new species and leading to radiation events (Soltis et al., 2009; Soltis, 2015; Wendel, 2015; Gillespie et al., 2020). Importantly, polyploidy reveals that *Symphonia* speciation in Madagascar could partially respond to a sympatric speciation pattern. Speciation via polyploidy is one of the modes of sympatric speciation in plants since polyploidization can result in genetic and genomic changes that may produce ecological divergences, changes in morphology and physiology, and reproductive barriers with their parental species (Otto & Whitton, 2000; Köhler, Mittelsten Scheid, & Erilova, 2010; Ainouche & Wendel, 2014).

Second, the presence of several gene pools per location (particularly in Andasibe, Ranomafana and Farankaraina) led us to discard the hypothesis that clusters were the result of the influence of geographic distance or landscape barriers to gene flow (reproduction), which would have resulted in isolation-by-distance patterns or geographically isolated gene pools. Thus, the discovered pattern together with the scarce signals of admixture among gene pools pointed to reproductive barriers (intrinsic reproductive isolation), providing evidence of the existence of different species (De Queiroz, 2007). As additional support, the patterns of the NJ trees did not reflect the geographic locations of the Malagasy gene pools, which would have been an indication of an isolation-by distance structure (for example see results for *S. globulifera* populations in Study II, part 1). The large genetic differentiation among Malagasy gene pools compared with the lower genetic differentiation among *S. globulifera* populations was also coherent with a genetic structure at species or species-cluster level, although within-species genetic structure signals can also be involved.

Finally, our tentative identification of Malagasy *Symphonia* species based on morphological characters gave some interesting hints regarding their genetic differences, pointing out that some gene pools could gather several species which differed in their morphological characters (e.g., the polyploid species cluster) whereas others could be nearly equivalent to a single species (e.g., rGP3 for *S. urophylla*, nGP4-rGP4 for *S. sp.1* in Nosy Mangabe, individuals in the intersection of rGP2-nGP3 for *S. sp.1* (Farankaraina)). It is worth remembering that the species within the Malagasy *Symphonia* group are morphologically variable and many of the morphological traits overlap between species (Abdul-Salim, 2002). Therefore, the partial correspondence found between genetic and morphological information of species only

reinforces the evidence that Malagasy *Symphonia* species are complex to delimit based on morphology alone.

Overall, the results presented in the Study III represent a first attempt of a species delimitation including genetic data in the species delimitation. Our combined results of genetic (SSR and SNP markers, genome size estimates) and non-genetic (morphology, geographical location) data pointed out that both functional marker sets (SSR and SNP) were informative on species differentiation over intraspecific population structure. Altogether, the different sources of data were rather congruent on the general situation (a necessary step to increase the confidence on species delimitation), in which Malagasy gene pools presented reproductive barriers (i.e., they did not depend on geographic barriers or distance) and gathered one or more Malagasy species (although they did not perfectly match due to their complex morphology; Carstens, Pelletier, Reid, & Satler, 2013; Pante, Schoelinck, & Puillandre, 2015). Nevertheless, through our analysis we could also detect intraspecific population structure putatively driven by other types of barriers to gene flow, such as geographic isolation in the case of Nosy Mangabe island (see Torroba-Balmori et al., 2017 for a revision of biotic and abiotic drivers of within-population spatial genetic structure). For that reason, we need to remember that our analysis might be merging signals of both between-species and within-species genetic structure and their corresponding drivers, and also that drivers for both levels of genetic structure are not mutually exclusive.

Development of non-neutral SNPs in a non-model genus

We have presented results based on novel SNP markers, developed from transcriptome sequences in a non-model genus and sequenced using the Sequenom technology. By using two methods of selection of candidate SNPs from transcriptome alignments, we encountered some of the problems of SNP detection which may arise in development workflows. In our study, based on the genotyping results of the screening step, we identified two causes which could yield non-successful SNPs. The first was a difficulty to define homogeneous flanking regions when designing primers, which caused the non-amplification of alleles during the PCR step (i.e., null alleles in all or most individuals). The second cause was the selection of non-polymorphic sites as SNP markers, which led to monomorphic SNPs during the screening step. Both types of errors occurred using both methods of selection.

A probable cause for erroneous detection of homogeneous flanking regions was the presence of polymorphisms at primer binding sites in the DNA sequence (such as insertion/deletion or point mutations), a well-known cause of allele dropout (e.g., Guichoux et al., 2011, Gautier et al., 2013), especially considering that we were working with a group of species which could gather a large number of polymorphisms. Due to its nature, this cause is likely to have equally affected both methods of SNP selection. In contrast, reasons for the selection of non-polymorphic sites seemed to be more linked to the selection method. Sequence misalignments arising from putative paralogy or repetitive regions are known to increase the degree of polymorphism observed and lead to detecting spurious alleles in individuals (Bryc, Patterson, & Reich, 2013, Verdu et al., 2016). We believe that such a problem could have affected the detection of SNPs using the automated method, as such a situation was frequently detected and avoided through the visual method. The automated method was based on the alignment of each

accession against the reference accession, by pairs, which possibly hindered the detection of paralogs. For the visual method we could not control for the depth of coverage, which is a parameter useful to control for sequencing errors (Margarido & Heckerman, 2015, Gompert & Mock, 2017). In conclusion, the visual SNP selection worked slightly better than the automated method, since it seems that our visual method could more successfully distinguish the presence of real diploid SNPs (i.e., the presence of two alternative alleles) than the automated method.

5.4. Concluding remarks

The goal of this thesis was to explore the evolutionary dynamics at micro and macro-geographic and evolutionary scales that have led the genus *Symphonia* to its current situation in terms of ecology, geographical distribution, local adaptation, and genetic diversity. My research emphasizes complementary approaches obtained by performing genetic structure studies within a genus, from micro to macro scales, considering different geographical and taxonomic levels.

At both micro and macro- geographic scales, *S. globulifera* presented a surprisingly strong and extended within-population structure and clustering in natural populations (see Studies I and II).

On one hand, we detected a wide diversity of FSGS patterns within *S. globulifera* populations, from non-significant or weak FSGS in Neotropical populations to pronounced structure in African ones. The strength of FSGS correlated with both disperser communities and altitudinal sampling range, while our data also contained evidence for co-occurrence of differentiated lineages and GP aggregation following habitat features (Study I). Therefore, the microenvironmental scale seems crucial for evolutionary processes in persistent populations of tree species, as has recently been shown in reports on microenvironmental adaptation in forest trees (Allié et al., 2015; Brousseau, Foll, Scotti-Saintagne, & Scotti, 2015; Scotti, González-Martínez, Budde, & Lalagüe, 2015; Budde et al., 2017).

Additionally, the complex spatial genetic structure of *S. globulifera* at macro-geographic scale allowed insights into the complex biogeographic history that affected an ancient rainforest tree species on two continents (Study II, part 1). Here, we revealed that the current genetic structure of *S. globulifera* in Africa originated after the colonization of America. We also presented the most probable origin of colonization events from Africa to America through São Tomé, among the African populations analysed, reinforcing the evidence that *S. globulifera* is able to perform marine dispersal. Finally, we analysed the impact of the Pleistocene climatic fluctuations on the genetic discontinuities within the species in both continents and found signatures of a possible ancestral location for our species in our southern populations in mainland Africa.

On the other hand, based on the complex spatial genetic structure revealed within the species at micro and macro scales, deeper research focused on cryptic species and species complexes is also a perspective in *S. globulifera*. This species encompasses different morphotypes or ecotypes occurring in different regions, including putative swamp and *terra firme* morphotypes in Africa (Budde, González-Martínez, Hardy, & Heuertz, 2013), the common valley bottom (*S. globulifera*) and the *terra firme* (*Symphonia sp.1*) ecotypes in French Guiana (Sabatier et al.

1997, Molino and Sabatier 2001, Baraloto, Morneau, Bonal, Blanc, & Ferry, 2007, Allié et al., 2015) as well as an understorey ecotype in Costa Rica (Dick & Heuertz, 2008; Sanfiorenzo, 2018). In our study (Study II, part 1), we evidenced that both morphotypes in French Guiana and the two nearby populations of the same common morphotype in French Guiana were evolutionarily distinct lineages, at the same level as geographically distant populations and despite the presence of gene flow among them. Since in both cases, they presented ecologically differentiated habitats or microhabitats and differentiated functional responses (see comments on Study II, part 2), mechanisms such as isolation by environment might be maintaining the genetic differentiation among genetic clusters, which would be a step forward to the properties related to the ecological species concept within our species (De Queiroz, 2007).

Our work also presents the first insight into the basis of local adaptation of *S. globulifera* at large scale, where the ability of our species to cope with water stress and acidic soils seems to underlie its extensive African distribution range and most likely, the Neotropical distribution range as well (Study II, part 2). However, it is worth keeping in mind that our environmental variables detected as influential were closely related to others (see Fig. S9.3.1.1.) that might also have had an influence in the adaptive evolution detected in *S. globulifera*. Also, further analysis with different sets of markers and other locations in both continents would be needed to consistently test if the species presents convergent evolution across continents and to clarify if the importance of different selective drivers varies across its wide distribution range.

Based on GBS and the Sequenom technologies in our studies, we presented two examples of how the current novel marker development workflows and the associated bioinformatic treatment of data, although very useful for the development of single nucleotide polymorphism (SNP) markers for non-model species, may come with some drawbacks (Studies II and III). Specifically, a high-throughput sequencing-based approach such as genotyping-by-sequencing, applied to strongly genetically differentiated populations, may lead to lower efficiencies than expected regarding the yield of genetic marker discovery and to low genetic variability within groups due to fixed alleles. The second example comes from our Sequenom workflow, in which we showed that "false SNPs" may arise from the SNP selection step and, importantly, they could bias further biological conclusions if not validated by genotyping as we did (see for example Verdu et al., 2016)

As significant outcomes from both workflows, we developed two sets of markers: (i) 4921 putatively neutral SNPs shared across populations of *S. globulifera* in two continents and (ii) 144 validated transcriptome SNPs for the genus *Symphonia*. Both resulting SNP sets could be used in the future to discriminate gene pools or ecotypes for conservation management in *Symphonia globulifera*, since this widespread species is ecologically important in tropical Africa and America and presents many local uses but, yet, its genetic constitution is poorly known (Oyen, 2005; Budde, González-Martínez, Hardy, & Heuertz, 2013). The transcriptome SNPs would, in addition, be most suitable to study the genomic basis of adaptation and speciation in the non-model genus *Symphonia*, and particularly useful to look into the drivers of speciation in this tropical group of tree species. Another interesting application would be their use for forest management and conservation practises, as Sequenom technology has been proposed for many tropical tree species for timber tracking purposes (e.g., in Carapa and Jacaranda, Sebbenn et al., 2019; Tysklind et al., 2019).

Finally, our study presents phylogenetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species) and congruent results about species delimitation on Malagasy *Symphonia* based on genetic and nongenetic sources of data (genetic data based on two sets of independent markers, nuclear genome sizes, morphology and geographical distribution) and on an unprecedented sampling effort in three major regions in Madagascar, both performed for the first time on this group of species (Study III). The concordance among different data strengthens the confidence in the taxonomic level revealed by the genetic analysis, although our approach was conservative as we could not differentiate genetically every species included in the analysis. Therefore, our findings are the first step towards a revision of the current species delimitation in the group of *Symphonia* species in Madagascar, in which polyploidy seems to play an important role in speciation.

6. Conclusions

1. The microenvironmental scale is crucial for evolutionary processes in persistent populations of tree species. Our study showed that altitudinal gradients, disperser communities, biogeographic history and microhabitat adaptation affect fine-scale spatial genetic structure in African and Neotropical populations of *S. globulifera,* an ancient tropical tree species (Study I).

    - **Microevolutionary processes have a major relevance in shaping fine-scale spatial genetic structure in African and Neotropical populations of *S. globulifera*.**

2. To understand how the current spatial genetic structure and the distribution of ancient tree species such as *S. globulifera* has arisen, it is necessary to take into account not only the current barriers to gene flow that maintain and reinforce its contemporary genetic differentiation, but also the historical drivers that affected the demographic history of the species through time, among which climate oscillations and ancient landscape barriers appear as most influential in our species (Study II, part 1).

    - **Climate oscillations and ancient landscape barriers appear as highly influential historical drivers shaping the spatial genetic structure in our species.**

3. Among the African populations analysed, the ones most closely related to Neotropical populations of *S. globulifera* were those from Benin and São Tomé Island, suggesting them as a possible source of colonization towards the American continent and reinforcing the evidence that *S. globulifera* is able to perform marine dispersal (Study II, part 1).

    - **Benin and São Tomé Island are probable sources of colonization of *S. globulifera* from Africa to America.**

4. Our results allow for a redefinition of genetically delimited gene pools within *S. globulifera*. Genetic clustering methods clearly delimited *S. sp.1* from the common morphotype in French Guiana which, the same time, presented two different gene pools. Also, the occurrence of *S. sp.1* in that region pointed to a different origin source. In addition, the sympatric occurrence of *S. sp.1* and the common morphotype in French Guiana suggests their genetic differentiation is maintained by isolation by environment, in the absence of a priori barriers to gene flow. The mechanism seems also relevant for the genetic differentiation of close locations in French Guiana. Thus, our study suggests that the ability of our tropical widespread species to occupy a vast distribution range characterized by a wide range of environments might be supported by these mechanisms, at least partially (Study II, part 1).

    - **The alternative morphotype *S. sp.1* is clearly genetically differentiated from the other two gene pools found related to the common morphotype in French Guiana and, probably, from a different origin source. The genetic**

differentiation of all gene pools found in French Guiana is possibly maintained by isolation by environment.

5. Water stress and acidic soils have been detected as two harsh environmental conditions that have a strong influence on the local adaptation of *S. globulifera* at the regional scale in Africa. However, given the high frequency of acidic soils in tropical locations in the world, and the evidence that *S. globulifera* is capable of physiological adjustments to cope with water stress at micro and macro scales in French Guiana, it is most probable that both environmental drivers play an important selective role for the species throughout its distribution range (Study II, part 2).

   - **Local adaptation of *S. globulifera* at the regional scale in Africa seems to be influenced by harsh environmental conditions related to water stress and acidic soils.**

6. One of the drivers underlying the radiation of the genus *Symphonia* in Madagascar is polyploidy, pointing out a probable sympatric speciation pattern related to this mechanism of speciation. Moreover, our study points out that Malagasy *Symphonia* includes *a priori* diploid and tetraploid species (Study III).

   - **Based on the *a priori* existence of diploid and tetraploid species, polyploidy is one of the drivers underlying the radiation of the genus *Symphonia* in Madagascar.**

7. The inferred phylogenetic relationships within the genus *Symphonia* (including *S. globulifera* and Malagasy *Symphonia* species), inferred with unprecedented resolution in this thesis, revealed very high genetic dissimilarities among Malagasy *Symphonia* gene pools in comparison with populations within *S. globulifera* and suggested a possible sequence of speciation events within the *Symphonia* genus to be further investigated (Study III).

   - **This thesis presents the phylogenetic relationships within the genus *Symphonia* with unprecedented resolution and suggest a possible sequence of speciation events.**

8. Our newly developed functional SNP markers were able to successfully delimit Malagasy *Symphonia* gene pools at species or supra-species level. However, based on the different resolution in genetic structure that markers can detect, it is necessary to use also other genetic and nongenetic information to cross-check taxonomic delimitations (Study III).

   - **The newly developed functional SNP markers provide essential information for cross-checking taxonomic delimitations in Malagasy *Symphonia*.**

9. Comparing approaches to analyse spatial genetic structure at different scales, the same drivers that affect genetic structure within and among populations of a single species can also affect spatial genetic structure in locations where closely related species occur in sympatry. Therefore, our study evidences that analysing the influence of the same factors at different spatial and taxonomic scales can give complementary insights and a

more comprehensive view of the evolutionary processes affecting a taxon, and may help to explain its contemporary evolutionary situation (Study I, II and III).

- **We can get a more comprehensive view of the evolutionary processes affecting the contemporary evolutionary situation in a species or genus by analysing the influence of the same factors at different spatial and taxonomic scales.**

10. Our study reveals *Symphonia* as an interesting genus with an outstanding spatial (at both micro and macro scale) and taxonomic genetic structure. Such characteristics, together with its wide distribution range covering a variety of habitats and its conspicuous diversification in the island of Madagascar, represents a biological model to study the evolutionary drivers and forces that influenced the genus throughout its history. Therefore, the genetic structure found within the genus *Symphonia* reinforces the suitability of this taxon as a model to study the evolutionary process in tropical plants (Study I, II and III).

    - ***Symphonia* is a suitable genus to study the evolutionary processes in tropical trees, as the taxon shows an outstanding amount of genetic structure at different levels, coupled with a wide distribution range covering a variety of habitats.**

7. Conclusiones

1. La escala microambiental es crucial en los procesos evolutivos de las poblaciones estables de especies arbóreas. Nuestro estudio mostró que los gradientes altitudinales, las comunidades de dispersantes, la historia biogeográfica y la adaptación a los microhábitats afecta a la estructura genética espacial a pequeña escala en las poblaciones africanas y neotropicales de *S. globulifera*, una especie arbórea tropical de gran antigüedad (Estudio I).

    - **Los procesos evolutivos a escalas pequeña tienen una gran importancia a la hora de moldear la estructura genética espacial a pequeña escala de las poblaciones africanas y neotropicales de *S. globulifera*.**

2. Para entender cómo han surgido la estructura genética espacial y la distribución actuales de una especie arbórea de gran antigüedad como *S. globulifera*, es necesario tener en cuenta no solo las actuales barreras al flujo genético que mantienen y refuerzan la actual diferenciación genética, sino también las causas históricas que afectaron la historia demográfica de la especie a lo largo del tiempo. Entre estas últimas hemos encontrado que las oscilaciones climáticas y las antiguas barreras geográficas han tenido una influencia importante en nuestra especie (Estudio II, parte 1).

    - **Las oscilaciones climáticas y las antiguas barreras geográficas aparecen como causas de gran influencia a la hora de dar forma a la estructura genética espacial de nuestra especie.**

3. Entre las poblaciones africanas analizadas, aquellas más relacionadas con las poblaciones neotropicales de *S. globulifera* fueron las de Benin y la isla de São Tomé, lo que sugiere que son una posible fuente de colonización del continente americano y refuerzan la evidencia de que *S. globulifera* es capaz de dispersarse a través del mar (Estudio II, parte 1).

    - **Benin y la isla de São Tomé son orígenes probables de la colonización de *S. globulifera* desde África hacia América.**

4. Nuestros resultados han permitido una redefinición de los grupos genéticos dentro de *S. globulifera*. En la Guyana Francesa, los métodos de agrupación genética separaron claramente a *S. sp.1* del morfotipo común, el cual presentó a su vez dos grupos genéticos diferentes. Además, la presencia de *S. sp.1* en la región apuntaba a que provenía de un origen diferente. Por otro lado, la presencia en simpatría de *S. sp.1* y el morfotipo común en la Guyana Francesa sugiere que su diferenciación genética se mantiene en principio gracias al aislamiento mediado por el ambiente, debido a la ausencia de barreras evidentes al flujo genético. Este mecanismo también parece relevante en la diferenciación genética de localizaciones cercanas dentro de la Guyana Francesa. Así, nuestro estudio sugiere que la habilidad para ocupar amplios rangos de distribución caracterizados por un amplio rango de hábitats que presenta especie tropical puede basarse en estos mecanismos, al menos parcialmente (Estudio II, parte 1).

- **El morfotipo alternativo *S. sp.1* es genéticamente diferente de los otros dos grupos genéticos encontrados para el morfotipo común en la Guyana Francesa y probablemente procede de una región distinta. La diferenciación de todos los grupos genéticos encontrados en la Guyana Francesa probablemente se mantiene gracias al aislamiento mediado por el ambiente.**

5. Se ha detectado que el estrés hídrico y la acidez del suelo son dos factores ambientales limitantes que tienen una gran influencia en la adaptación local de *S. globulifera* a escala regional en África. Sin embargo, dada la alta frecuencia de los suelos ácidos en regiones tropicales del mundo, y la evidencia de que *S. globulifera* es capaz de hacer ajustes fisiológicos para hacer frente al estrés hídrico a escalas micro y macro en la Guyana Francesa, es muy probable que ambos factores ambientales jueguen un papel selectivo importante para la especie a lo largo de toda su área de distribución (Estudio II, parte 2).

   - **La adaptación local de *S. globulifera* a escala regional en África parece estar influenciada por factores limitantes ambientales relacionados con el estrés hídirco y los suelos ácidos.**

6. Uno de los factores que ha impulsado la radiación del género *Symphonia* en Madagascar es la poliploidía, lo que posiblemente apunta a un patrón de especiación en simpatría relacionado con este mecanismo de especiación. Además, nuestro estudio señala que las especies de *Symphonia* en Madagascar incluyen en principio especies diploides y tetraploides (Estudio III).

   - **En base a la existencia de especies diploides y tetraploides, la poliploidía parece uno de los factores causantes de la radiación del género *Symphonia* en Madagascar.**

7. Las relaciones filogenéticas inferidas dentro del género *Symphonia* (que incluye *S. globulifera* y especies de *Symphonia* de Madagascar), obtenidas en esta tesis con una resolución sin precedentes hasta la fecha, han revelado fuertes diferencias entre los grupos genéticos de *Symphonia* en Madagascar en comparación con las poblaciones de *S. globulifera*, y además sugieren una posible secuencia de eventos de especiación dentro del género *Symphonia* para futuras investigaciones (Estudio III).

   - **Esta tesis presenta las relaciones filogenéticas dentro del género *Symphonia* con una resolución sin precedentes y sugiere una posible secuencia de eventos de especiación.**

8. Nuestros nuevos marcadores funcionales, basados en SNPs, fueron capaces de delimitar con éxito los grupos genéticos a nivel de especie o superior dentro del género *Symphonia* en Madagascar. Sin embargo, puesto que los marcadores pueden detectar estructuras genéticas a diferentes resoluciones, es necesario emplear también otras fuentes de información genética y no genética para cotejarlas con la delimitación taxonómica (Estudio III).

- **Los nuevos marcadores funcionales, basados en SNPs, han aportado información esencial para cotejar la delimitación taxonómica de especies de *Symphonia* en Madagascar.**

9. Comparando las estrategias de análisis de la estructura genética espacial a diferentes escalas, los mismos factores que afectan a la estructura genética dentro y entre poblaciones de una misma especie pueden también afectar a la estructura genética espacial en lugares donde especies emparentadas viven en simpatría. Así, nuestro estudio muestra que analizar la influencia de los mismos factores a diferentes escalas espaciales y taxonómicas puede dar visiones complementarias y más completas sobre los procesos evolutivos que afectan a las especies, ayudando a explicar su situación evolutiva actual (Estudio I, II y III)

    - **Es posible obtener una vision más completa de los procesos evolutivos que afectan a la situación evolutiva actual de una especie o género analizando la influencia de los mismos factores a diferentes escalas espaciales y taxonómicas.**

10. Nuestro estudio señala a *Symphonia* como un género interesante con una estructura espacial (a escala micro y macro) y taxonómica muy destacable. Esta característica, junto con su amplia distribución geográfica que cubre una gran variedad de hábitats y su llamativa diversificación en la isla de Madagascar, señala un modelo biológico para estudiar los factores evolutivos y fuerzas que han influido en el género a través de la historia. Así, la estructura genética encontrada en el género *Symphonia* refuerza la idoneidad de este taxón como modelo para estudiar los procesos evolutivos en plantas tropicales (Estudio I, II y III).

    - ***Symphonia* es un genéro muy adecuado para estudiar los procesos evolutivos en árboles tropicales, ya que es un taxón que muestra una acumulación muy llamativa de estructuras genéticas a diferentes niveles, junto con un amplio rango de distribución que cubre una gran variedad de hábitats.**

8. References

Abdul-Salim, K. (2002). *Systematics and biology of Symphonia L. f. (Clusiaceae)*. (Doctoral dissertation). Harvard University, Cambridge, Massachusetts (US).

Ahuja, M. R., & Mohan Jain, S. (Eds.). (2017). *Biodiversity and Conservation of Woody Plants*. Springer. https://doi.org/10.1007/978-3-319-66426-2

Ainouche, M. L., & Wendel, J. F. (2014). Polyploid Speciation and Genome Evolution: Lessons from Recent Allopolyploids. In P. Pontarotti (Ed.), *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life* (pp. 87–113). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07623-2

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., … Savolainen, O. (2013). Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology*, *19*(6), 1645–1661. https://doi.org/10.1111/gcb.12181

Aldrich, P. R., & Hamrick, J. L. (1998). Reproductive Dominance of Pasture Trees in a Fragmented Tropical Forest Mosaic. *Science*, *281*(5373), 103–105. https://doi.org/10.1126/science.281.5373.103

Aldrich, P. R., Hamrick, J. L., Chavarriaga, P., & Kochert, G. (1998). Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, *7*(8), 933–944. https://doi.org/10.1046/j.1365-294x.1998.00396.x

Allié, E., Pélissier, R., Engel, J., Petronelli, P., Freycon, V., Deblauwe, V., … Baraloto, C. (2015). Pervasive Local-Scale Tree-Soil Habitat Association in a Tropical Forest Community. *PLOS ONE*, *10*(11), e0141488. https://doi.org/10.1371/journal.pone.0141488

Andrew, R. L., Ostevik, K. L., Ebert, D. P., & Rieseberg, L. H. (2012). Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, *21*(9), 2078–2091. https://doi.org/10.1111/j.1365-294X.2012.05454.x

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Anhuf, D., Ledru, M. P., Behling, H., Da Cruz, F. W., Cordeiro, R. C., Van der Hammen, T., … Da Silva Dias, P. L. (2006). Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *239*(3–4), 510–527. https://doi.org/10.1016/j.palaeo.2006.01.017

Anjum, N. A., Singh, H. P., Khan, M. I. R., Masood, A., Per, T. S., Negi, A., … Ahmad, I. (2015). Too much is bad—an appraisal of phytotoxicity of elevated plant-beneficial heavy metal ions. *Environmental Science and Pollution Research*, *22*(5), 3361–3382. https://doi.org/10.1007/s11356-014-3849-9

Antonelli, A., Ariza, M., Albert, J., Andermann, T., Azevedo, J., Bacon, C., … Edwards, S. V. (2018). Conceptual and empirical advances in Neotropical biodiversity research. *PeerJ*, *6*(10), e5644. https://doi.org/10.7717/peerj.5644

Arnaud-Haond, S., & Belkhir, K. (2007). GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular*

Ecology Notes, *7*(1), 15–17. https://doi.org/10.1111/j.1471-8286.2006.01522.x

Asmussen, M. A., Arnold, J., & Avise, J. C. (1989). The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics*, *122*(4), 923–934.

Audigeos, D., Brousseau, L., Traissac, S., Scotti-Saintagne, C., & Scotti, I. (2013). Molecular divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary Biology*, *26*(3), 529–544. https://doi.org/10.1111/jeb.12069

Augspurger, C. K. (1980). Mass-Flowering of a Tropical Shrub (*Hybanthus prunifolius*): Influence on Pollinator Attraction and Movement. *Evolution*, *34*(3), 475–488. https://doi.org/10.2307/2408217

Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15. https://doi.org/10.1111/j.1365-2699.2008.02032.x

Badri, D. V., Loyola-Vargas, V. M., Broeckling, C. D., De-la-Peña, C., Jasinski, M., Santelia, D., … Vivanco, J. M. (2008). Altered profile of secondary metabolites in the root exudates of arabidopsis ATP-binding cassette transporter mutants. *Plant Physiology*, *146*(2), 762–771. https://doi.org/10.1104/pp.107.109587

Baer, C. F., Miyamoto, M. M., & Denver, D. R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, *8*(8), 619–631. https://doi.org/10.1038/nrg2158

Bainard, J. D., Husband, B. C., Baldwin, S. J., Fazekas, A. J., Gregory, T. R., Newmaster, S. G., & Kron, P. (2011). The effects of rapid desiccation on estimates of plant genome size. *Chromosome Research*, *19*(6), 825–842. https://doi.org/10.1007/s10577-011-9232-5

Baker, T. R., Pennington, R. T., Dexter, K. G., Fine, P. V. A., Fortune-Hopkins, H., Honorio, E. N., … Vasquez, R. (2017). Maximising Synergy among Tropical Plant Systematists, Ecologists, and Evolutionary Biologists. *Trends in Ecology and Evolution*, *32*(4), 258–267. https://doi.org/10.1016/j.tree.2017.01.007

Baraloto, C., Morneau, F., Bonal, D., Blanc, L., & Ferry, B. (2007). Seasonal Water Stress Tolerance and Habitat Associations within Four Neotropical Tree Genera. *Ecology*, *88*(2), 478–489. https://doi.org/http://www.jstor.org/stable/27651120

Barthe, S., Binelli, G., Hérault, B., Scotti-Saintagne, C., Sabatier, D., & Scotti, I. (2017). Tropical rainforests that persisted: inferences from the Quaternary demographic history of eight tree species in the Guiana shield. *Molecular Ecology*, *26*(4), 1161–1174. https://doi.org/10.1111/mec.13949

Bawa, K., Bullock, S., Perry, D., Coville, R., & Grayum, M. (1985). Reproductive biology of tropical lowland rain forest trees. II. Pollination Systems. *American Journal of Botany*, *72*(3), 346–356.

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Bell, R. C., Drewes, R. C., & Zamudio, K. R. (2015). Reed frog diversification in the Gulf of Guinea: Overseas dispersal, the progression rule, and in situ speciation. *Evolution*, *69*(4), 904–915. https://doi.org/10.1111/evo.12623

Bennett, K. D., & Provan, J. (2008). What do we mean by "refugia"? *Quaternary Science Reviews*, *27*(27–28), 2449–2455. https://doi.org/10.1016/j.quascirev.2008.08.019

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., … Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Bittrich, V., & Amaral, M. C. E. (1996). Pollination biology of *Symphonia globulifera* (Clusiaceae). *Plant Systematics and Evolution*, *200*, 101–110.

Bjork, A., Liu, W., Wertheim, J. O., Hahn, B. H., & Worobey, M. (2011). Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Molecular Biology and Evolution*, *28*(1), 615–623. https://doi.org/10.1093/molbev/msq227

Bojórquez-Quintal, E., Escalante-Magaña, C., Echevarría-Machado, I., & Martínez-Estévez, M. (2017). Aluminum, a Friend or Foe of Higher Plants in Acid Soils. *Frontiers in Plant Science*, *8*, 1767. https://doi.org/10.3389/fpls.2017.01767

Born, C., Hardy, O. J., Chevallier, M. H., Ossari, S., Attéké, C., Wickings, E. J., … Bruxelles, L. De. (2008). Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: A stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology*, *17*(8), 2041–2050. https://doi.org/10.1111/j.1365-294X.2007.03685.x

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bouckaert, R. R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Bradic, M., Costa, J., & Chelo, I. M. (2011). Genotyping with Sequenom. In V. Orgogozo & M. V. Rockman (Eds.), *Molecular Methods for Evolutionary Genetics* (Vol. 772, pp. 193–210). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-61779-228-1

Brewer, S., & Rejmánek, M. (1999). Small rodents as significant dispersers of tree seeds in a Neotropical Forest. *Journal of Vegetation Science*, *10*, 165–174. https://doi.org/10.2307/3237138

Brookfield, J. F. Y. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, *5*(3), 453–455. https://doi.org/10.1046/j.1365-294X.1996.00098.x

Brooks, D., Bodmer, R., & Matola, S. (1997). *Tapirs–Status survey and conservation action plan.* Gland, Switzerland: IUCN/SSC Tapir Specialist Group, IUCN.

Brousseau, L., Fine, P. V. A., Dreyer, E., Vendramin, G. G., & Scotti, I. (2020). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. (Fabaceae). *Molecular Ecology*, *30*(5), 1136–1154. https://doi.org/10.1111/mec.15595

Brousseau, L., Foll, M., Scotti-Saintagne, C., & Scotti, I. (2015). Neutral and Adaptive Drivers of Microgeographic Genetic Divergence within Continuous Populations: The Case of the Neotropical Tree Eperua falcata (Aubl.). *Plos One*, *10*(3), e0121394. https://doi.org/10.1371/journal.pone.0121394

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. a., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Bryc, K., Patterson, N., & Reich, D. (2013). A novel approach to estimating heterozygosity from low-coverage genome sequence. *Genetics*, *195*(2), 553–561. https://doi.org/10.1534/genetics.113.154500

Budde, K. B. (2014). *Genetic structure of forest trees in biodiversity hotspots at different spatial scales*. (Doctoral dissertation).Universidad Complutense de Madrid, Madrid (Spain).

Budde, K. B., González-Martínez, S. C., Hardy, O. J., & Heuertz, M. (2013). The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. *Heredity*, *111*(1), 66–76. https://doi.org/10.1038/hdy.2013.21

Budde, K. B., González-Martínez, S. C., Navascués, M., Burgarella, C., Mosca, E., Lorenzo, Z., … Heuertz, M. (2017). Increased fire frequency promotes stronger spatial genetic structure and natural selection at regional and local scales in *Pinus halepensis* Mill. *Annals of Botany*, *119*(6), 1061–1072. https://doi.org/10.1093/aob/mcw286

Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R. J., … Gil, L. (2009). Detection of hybrids in nature: Application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*, *102*(5), 442–452. https://doi.org/10.1038/hdy.2009.8

Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., Castillo Cajas, R. F., … Schilthuizen, M. (2012). What do we need to know about speciation? *Trends in Ecology and Evolution*, *27*(1), 27–39. https://doi.org/10.1016/j.tree.2011.09.002

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Canova, L. (1993). Resource partitioning between the bank vole *Clethrionomys glareolus* and the wood mouse *Apodemus sylvaticus* in woodland habitats. *Bolletino Di Zoologia*, *60*(2), 193–198. https://doi.org/10.1080/11250009309355809

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, *323*(5915), 785–789. https://doi.org/10.1126/science.1166955

Carnaval, A. C., & Moritz, C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, *35*(7), 1187–1201. https://doi.org/10.1111/j.1365-2699.2007.01870.x

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, *22*, 4369–4383. https://doi.org/10.1111/mec.12413

Charles-Dominique, P. (1986). Inter-relations between frugivorous vertebrates and pioneer plants: *Cecropia*, birds and bats in French Guyana. In A. Estrada & T. H. Fleming (Eds.), *Frugivores and seed dispersal* (pp. 119–135). Dordrecht: Dr W. Junk Publishers. https://doi.org/10.2307/2260583

Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, *7*(5), 747–756. https://doi.org/10.1111/j.1471-8286.2007.01769.x

Choi, S. C. (2016). The role of arboreal seed dispersal groups on the seed rain of a lowland tropical forest. *Genes and Genomics*, *38*(10), 905–915. https://doi.org/10.1007/s13258-016-0458-7

Clark, C. C. J., Poulsen, J. J. R., & Parker, V. T. (2001). The role of arboreal seed dispersal groups on the seed rain of a lowland tropical forest. *Biotropica*, *33*(4), 606–620. https://doi.org/10.1111/j.1744-7429.2001.tb00219.x

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*(10), 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Collevatti, R. G., Estolano, R., Ribeiro, M. L., Rabelo, S. G., Lima, E. J., & Munhoz, C. B. R. (2014). High genetic diversity and contrasting fine-scale spatial genetic structure in four seasonally dry tropical forest tree species. *Plant Systematics and Evolution*, *300*(7), 1671–1681. https://doi.org/10.1007/s00606-014-0993-0

Collevatti, R. G., Novaes, E., Silva-Junior, O. B., Vieira, L. D., Lima-Ribeiro, M. S., & Grattapaglia, D. (2019). A genome-wide scan shows evidence for local adaptation in a widespread keystone Neotropical forest tree. *Heredity*, *123*(2), 117–137. https://doi.org/10.1038/s41437-019-0188-0

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423. https://doi.org/10.1534/genetics.110.114819

Cordeiro, N. J., Ndangalasi, H. J., McEntee, J. P., & Howe, H. F. (2009). Disperser limitation and recruitment of an endemic African tree in a fragmented landscape. *Ecology*, *90*(4), 1030–1041. https://doi.org/10.1890/07-1208.1

Côrtes, M. C., & Uriarte, M. (2013). Integrating frugivory and animal movement: a review of the evidence and implications for scaling seed dispersal. *Biological Reviews*, *88*(2), 255–272. https://doi.org/10.1111/j.1469-185X.2012.00250.x

Couvreur, T. L. P., Forest, F., & Baker, W. J. (2011). Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology*, *9*, 44. https://doi.org/10.1186/1741-7007-9-44

Cowling, S. A., Cox, P. M., Jones, C. D., Maslin, M. A., Peros, M., & Spall, S. A. (2008). Simulated glacial and interglacial vegetation across Africa: Implications for species phylogenies and trans-African migration of plants and animals. *Global Change Biology*, *14*(4), 827–840. https://doi.org/10.1111/j.1365-2486.2007.01524.x

da Silva Carneiro, F., Degen, B., Kanashiro, M., de Lacerda, A. E. B., & Sebbenn, A. M. (2009). High levels of pollen dispersal detected through paternity analysis from a continuous *Symphonia globulifera* population in the Brazilian Amazon. *Forest Ecology and Management*, *258*(7), 1260–1266. https://doi.org/10.1016/j.foreco.2009.06.019

da Silva Carneiro, F., Sebbenn, A. M., Kanashiro, M., & Degen, B. (2007). Low Interannual Variation of Mating System and Gene Flow of Symphonia globulifera in the Brazilian Amazon. *Biotropica*, *39*(5), 628–636. https://doi.org/10.1111/j.1744-7429.2007.00314.x

Daïnou, K., Bizoux, J. P., Doucet, J. L., Mahy, G., Hardy, O. J., & Heuertz, M. (2010). Forest refugia revisited: NSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). *Molecular Ecology*, *19*(20), 4462–4477. https://doi.org/10.1111/j.1365-294X.2010.04831.x

Daïnou, K., Blanc-Jolivet, C., Degen, B., Kimani, P., Ndiade-Bourobou, D., Donkpegan, A. S. L., … Hardy, O. J. (2016). Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences - a case study in the tree genus *Milicia*.

BMC Evolutionary Biology, *16*, 259. https://doi.org/10.1186/s12862-016-0831-9

Dauby, G., Duminil, J., Heuertz, M., Koffi, G. K., Stévart, T., & Hardy, O. J. (2014). Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Molecular Ecology*, *23*(9), 2299–2312. https://doi.org/10.1111/mec.12724

Dauby, G., Zaiss, R., Blach-Overgaard, A., Catarino, L., Damen, T., Deblauwe, V., … Couvreur, T. L. P. (2016). RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys*, *74*, 1–18. https://doi.org/10.3897/phytokeys.74.9723

Davis, M. B., & Shaw, R. G. (2001). Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, *292*(5517), 673–679. https://doi.org/10.1126/science.292.5517.673

De-Lucas, A. I., González-Martínez, S. C., Vendramin, G. G., Hidalgo, E., & Heuertz, M. (2009). Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* Aiton. *Molecular Ecology*, *18*(22), 4564–4576. https://doi.org/10.1111/j.1365-294X.2009.04372.x

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399. https://doi.org/10.1111/mec.12182

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*(11), 1248–1258. https://doi.org/10.1111/2041-210X.12418

Debout, G. D. G., Doucet, J.-L., & Hardy, O. J. (2011). Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity*, *106*, 88–99. https://doi.org/10.1038/hdy.2010.35

Defaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources*, *13*(3), 377–392. https://doi.org/10.1111/1755-0998.12071

Degen, B., Bandou, E., & Caron, H. (2004). Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity*, *93*, 585–591. https://doi.org/10.1038/sj.hdy.6800560

Degen, B., Blanc, L., Caron, H., Maggia, L., Kremer, A., & Gourlet-Fleury, S. (2006). Impact of selective logging on genetic composition and demographic structure of four tropical tree species. *Biological Conservation*, *131*, 386–401. https://doi.org/10.1016/j.biocon.2006.02.014

Demenou, B. B., Migliore, J., Heuertz, M., Monthe, F. K., Ojeda, D. I., Wieringa, J. J., … Hardy, O. J. (2020). Plastome phylogeography in two African rain forest legume trees reveals that Dahomey Gap populations originate from the Cameroon volcanic line. *Molecular Phylogenetics and Evolution*, *150*, 106854. https://doi.org/10.1016/j.ympev.2020.106854

Dexter, K. G., Pennington, T. D., & Cunningham, C. W. (2010). Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter?

Ecological Monographs, *80*(2), 267–286. https://doi.org/10.1890/09-0267.1

Dick, C. W. (2010). Phylogeography and population structure of tropical trees. *Tropical Plant Biology*, *3*, 1–3. https://doi.org/10.1007/s12042-009-9039-0

Dick, C. W., Abdul-Salim, K., & Bermingham, E. (2003). Molecular Systematic Analysis Reveals Cryptic Tertiary Diversification Diversification of a Widespread Tropical Rain Forest Tree. *The American Naturalist*, *162*(6), 691–703. https://doi.org/10.1086/379795

Dick, C. W., Etchelecu, G., & Austerlitz, F. (2003). Pollen dispersal of tropical trees (Dinizia excelsa: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology*, *12*(3), 753–764. https://doi.org/10.1046/j.1365-294X.2003.01760.x

Dick, C. W., Hardy, O. J., Jones, F. A., & Petit, R. J. (2008). Spatial Scales of Pollen and Seed-Mediated Gene Flow in Tropical Rain Forest Trees. *Tropical Plant Biology*, *1*(1), 20–33. https://doi.org/10.1007/s12042-007-9006-6

Dick, C. W., & Heuertz, M. (2008). The complex biogeographic history of a widespread tropical tree species. *Evolution*, *62*(11), 2760–2774. https://doi.org/10.1111/j.1558-5646.2008.00506.x

Dick, C. W., Lewis, S. L., Maslin, M., & Bermingham, E. (2013). Neogene origins and implied warmth tolerance of Amazon tree species. *Ecology and Evolution*, *3*(1), 162–169. https://doi.org/10.1002/ece3.441

Dick, C. W., & Pennington, R. T. (2019). History and Geography of Neotropical Tree Diversity. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 279–301. https://doi.org/10.1146/annurev-ecolsys-110617-062314

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M. A., Nardi, L., & Obermayer, R. (1998). Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Annals of Botany*, *82*(SUPPL. A), 17–26. https://doi.org/10.1006/anbo.1998.0730

Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST 2*. Cambridge University Press.

Dubost, G. (1984). Comparison of the diets of frugivorous forest ruminants of Gabon. *Journal of Mammalogy*, *65*(2), 298–316. https://doi.org/http://dx.doi.org/10.2307/1381169

Duminil, J., Daïnou, K., Kaviriri, D. K., Gillet, P., Loo, J., Doucet, J.-L., & Hardy, O. J. (2016). Relationships between population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. *Heredity*, *116*(3), 295–303. https://doi.org/10.1038/hdy.2015.101

Duminil, J., & Di Michele, M. (2009). Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosystems*, *143*(3), 528–542. https://doi.org/10.1080/11263500902722964

Duminil, J., Kenfack, D., Viscosi, V., Grumiau, L., & Hardy, O. J. (2011). Testing species delimitation in sympatric species complexes: The case of an African tropical tree, Carapa spp. (Meliaceae). *Molecular Phylogenetics and Evolution*, *62*(1), 275–285. https://doi.org/10.1016/j.ympev.2011.09.020

Duminil, J., Mona, S., Mardulyn, P., Doumenge, C., Walmacq, F., Doucet, J. L., & Hardy, O. J. (2015). Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. *Journal of Biogeography*, *42*(8), 1443–1454.

https://doi.org/10.1111/jbi.12510

Durand, E., Chen, C., & François, O. (2009). Tess version 2.3 - Reference Manual.

Durand, E., Jay, F., Gaggiotti, O. E., & François, O. (2009). Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, *26*(9), 1963–1973. https://doi.org/10.1093/molbev/msp106

Dutech, C., Maggia, L., & Joly, H. I. (2000). Chloroplast diversity in Vouacapoua americana (Caesalpiniaceae), a neotropical forest tree. *Molecular Ecology*, *9*(9), 1427–1432. https://doi.org/10.1046/j.1365-294X.2000.01027.x

Dutech, C., Maggia, L., Tardy, C., Joly, H. I., & Jarne, P. (2003). Tracking a genetic signal of extinction-recolonization events in a neotropical tree species: *Vouacapoua americana* aublet in French Guiana. *Evolution*, *57*(12), 2753–2764. https://doi.org/10.1111/j.0014-3820.2003.tb01517.x

Dyer, R. J., Chan, D. M., Gardiakos, V. A., & Meadows, C. A. (2012). Pollination graphs: Quantifying pollen pool covariance networks and the influence of intervening landscape on genetic connectivity in the North American understory tree, *Cornus florida* L. *Landscape Ecology*, *27*(2), 239–251. https://doi.org/10.1007/s10980-011-9696-x

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eckert, A. J., & Dyer, R. J. (2012). Defining the landscape of adaptive genetic diversity. *Molecular Ecology*, *21*(12), 2836–2838. https://doi.org/10.1111/j.1365-294X.2012.05615.x

Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1777), 20132765. https://doi.org/10.1098/rspb.2013.2765

Ence, D. D., & Carstens, B. C. (2011). SpedeSTEM: A rapid and accurate method for species delimitation. *Molecular Ecology Resources*, *11*(3), 473–480. https://doi.org/10.1111/j.1755-0998.2010.02947.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewédjè, E. (2012). *Biologie de la reproduction, phylogéographie et diversité de l'arbre à beurre, Pentadesma butyracea Sabine (Clusiaceae) - implications pour sa conservation au Bénin.* (Doctoral dissertation). Université Libre de Bruxelles, Brussels (Belgium).

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Falush, D., Stephens, M., & Pritchard, J. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

FAO, & UNEP. (2020). The State of the World's Forests 2020. Forests, biodiversity and people. Rome. https://doi.org/10.4060/ca8642en.

Faye, A., Deblauwe, V., Mariac, C., Richard, D., Sonké, B., Vigouroux, Y., & Couvreur, T. L.

P. (2016b). Phylogeography of the genus Podococcus (Palmae/Arecaceae) in Central African rain forests: Climate stability predicts unique genetic diversity. *Molecular Phylogenetics and Evolution*, *105*, 126–138. https://doi.org/10.1016/j.ympev.2016.08.005

Faye, A., Pintaud, J. C., Baker, W. J., Vigouroux, Y., Sonke, B., & Couvreur, T. L. P. (2016a). (n.d.). Phylogenetics and diversification history of African rattans (Calamoideae, Ancistrophyllinae). *Botanical Journal of the Linnean Society*, *182*(2), 256–271. https://doi.org/10.1111/boj.12454

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Felstein, J. (1991). PHYLIP v. 3.696. University of Washington, Seattle.

Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., & Glover, B. J. (2018). Resolving Recent Plant Radiations: Power and Robustness of Genotyping-by-Sequencing. *Systematic Biology*, *67*(2), 250–268. https://doi.org/10.1093/sysbio/syx062

Fetter, K. C., Gugger, P. F., & Keller, S. R. (2017). Landscape Genomics of Angiosperm Trees: From Historic Roots to Discovering New Branches of Adaptive Evolution. In A. T. Groover & Q. C. B. Cronk (Eds.), *Comparative and Evolutionary Genomics of Angiosperm Trees* (pp. 303–333). Springer, Cham. https://doi.org/10.1007/7397_2016_19

Fields, P. D., McCauley, D. E., McAssey, E. V, & Taylor, D. R. (2014). Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability. *Heredity*, *112*(2), 99–104. https://doi.org/10.1038/hdy.2013.79

Fine, P. V. A., Daly, D. C., Muñoz, G. V., Mesones, I., & Cameron, K. M. (2005). The Contribution of Edaphic Heterogeneity To the Evolution and Diversity of Burseraceae Trees in the Western Amazon. *Evolution*, *59*(7), 1464–1478. https://doi.org/10.1554/04-745

Fine, P. V. A., & Ree, R. H. (2006). Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *American Naturalist*, *168*(6), 796–804. https://doi.org/10.1086/508635

Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, *22*(11), 2342–2347. https://doi.org/10.1111/j.1420-9101.2009.01833.x

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, *11*(7), 1035–1052. https://doi.org/10.1111/eva.12569

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Fordham, D. A., Brook, B. W., Moritz, C., & Nogués-Bravo, D. (2014). Better forecasts of range dynamics using genetic data. *Trends in Ecology and Evolution*, *29*(8), 436–443. https://doi.org/10.1016/j.tree.2014.05.007

Forget, P.-M. (1990). Seed-dispersal of *Vouacapoua americana* (Caesalpiniaceae) by caviomorph rodents in French Guiana. *Journal of Tropical Ecology*, *6*, 459–468.

Forget, P., Dennis, A., Mazer, S., Jansen, P., Kitamura, S., Lambert, J., & Westcott, D. (2007). Seed allometry and disperser assemblages in tropical rainforests: A comparison of four

floras on different continents. In A. Dennis, E. W. Schupp, R. Green, & D. Westcott (Eds.), *Seed dispersal. Theory and its Application in a Changing World* (pp. 5–37). Oxfordshire: CAB International.

Fragoso, J. M. V. (1997). Tapir-generated seed shadows: scale-dependent patchiness in the Amazon rain forest. *Journal of Ecology*, *85*(4), 519–529. https://doi.org/10.2307/2960574

François, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, *10*(5), 773–784. https://doi.org/10.1111/j.1755-0998.2010.02868.x

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Fuchs, J., & Bowie, R. C. K. (2015). Concordant genetic structure in two species of woodpecker distributed across the primary West African biogeographic barriers. *Molecular Phylogenetics and Evolution*, *88*, 64–74. https://doi.org/10.1016/j.ympev.2015.03.011

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, *27*(9), 480–488. https://doi.org/10.1016/j.tree.2012.04.012

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., & Firoozabady, E. (1983). Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues. *Science*, *220*(4601), 1049–1051. https://doi.org/10.1126/science.220.4601.1049

Garnier-Géré, P., Harmand, N., Laizet, Y., & Mariette, S. (2014). A R program implemented in Galaxy for Sequenom SNP genotypes batch visualization and alternative clustering (unpublished).

Garot, E., Joët, T., Combes, M. C., & Lashermes, P. (2019). Genetic diversity and population divergences of an indigenous tree (Coffea mauritiana) in Reunion Island: role of climatic and geographical factors. *Heredity*, *122*(6), 833–847. https://doi.org/10.1038/s41437-018-0168-9

Gautier-Hion, A., Duplantier, J., Quris, R., Feer, F., Sourd, C., Decoux, G., … Thiollay, J. (1985). Fruit characters as a basis of fruit choice and seed dispersal in a tropical forest vertebrate community. *Oecologia*, *65*, 324–337.

Gautier-Hion, A., Emmons, L. H., & Dubost, G. (1980). A Comparison of the Diets of Three Major Groups of Primary Consumers of Gabon (Primates, Squirrels and Ruminants). *Oecologia*, *45*, 182–189.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., … Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, *22*(11), 3165–3178. https://doi.org/10.1111/mec.12089

Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodríguez-Sánchez, F., Dobrowski, S. Z., … Williams, J. W. (2014). Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, *204*(1), 37–54. https://doi.org/10.1111/nph.12929

Gibbard, P. L., Ehlers, J., & Hughes, P. D. (2017). Quaternary Glaciations. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston. (Eds.), *The International Encyclopedia of Geography: People, the Earth, Environment and Technology* (pp. 1–10). Oxford, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118786352.wbieg0562

Gill, G. E., Fowler, R. T., & Mori, S. A. (1998). Pollination Biology of *Symphonia globulifera* (Clusiaceae) in Central French Guiana. *Biotropica*, *30*(1), 139–144.

Gillespie, R. G., Bennett, G. M., De Meester, L., Feder, J. L., Fleischer, R. C., Harmon, L. J., … Wogan, G. O. U. (2020). Comparing Adaptive Radiations Across Space, Time, and Taxa. *Journal of Heredity*, *111*(1), 1–20. https://doi.org/10.1093/jhered/esz064

Giombini, M. I., Bravo, S. P., & Tosto, D. S. (2016). The key role of the largest extant Neotropical frugivore (*Tapirus terrestris*) in promoting admixture of plant genotypes across the landscape. *Biotropica*, *0*(0), 1–10. https://doi.org/10.1111/btp.12328

Giordano, A. R., Ridenhour, B. J., & Storfer, A. (2007). The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactulym*). *Molecular Ecology*, *16*(8), 1625–1637. https://doi.org/10.1111/j.1365-294X.2006.03223.x

Girard, P., & Angers, B. (2008). Assessment of power and accuracy of methods for detection and frequency-estimation of null alleles. *Genetica*, *134*(2), 187–197. https://doi.org/10.1007/s10709-007-9224-8

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. a., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, 4555–4573. https://doi.org/10.1111/mec.12811

Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, *17*(6), 1156–1167. https://doi.org/10.1111/1755-0998.12657

Gonmadje, C. F., Doumenge, C., Sunderland, T. C. H., Balinga, M. P. B., & Sonké, B. (2012). Analyse phytogéographique des forêts d'Afrique Centrale: le cas du massif de Ngovayang (Cameroun). *Plant Ecology and Evolution*, *145*(2), 152–164. https://doi.org/10.5091/plecevo.2012.573

Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. I., & Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, *25*(22), R1066–R1067. https://doi.org/10.1016/j.cub.2015.10.002

Goslee, S. C., & Urban, D. L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Goudet, J., & Jombart, T. (2015). Estimation and Tests of Hierarchical F-Statistics. Package 'hierfstat'. Retrieved from http://github.com/jgx65/hierfstat

Gräfe, K., & Schmitt, L. (2020). The ABC transporter G subfamily in *Arabidopsis thaliana*. *Journal of Experimental Botany*, *72*(1), 92–106. https://doi.org/10.1093/jxb/eraa260

Guichoux, E., Garnier-Géré, P., Lagache, L., Lang, T., Boury, C., & Petit, R. J. (2013). Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology*, *22*(2), 450–462. https://doi.org/10.1111/mec.12125

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., … Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, *11*(4), 591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

Guitet, S., Cornu, J. F., Brunaux, O., Betbeder, J., Carozza, J. M., & Richard-Hansen, C. (2013). Landform and landscape mapping, French Guiana (South America). *Journal of Maps*, *9*(3), 325–335. https://doi.org/10.1080/17445647.2013.785371

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Gupta, N., Gaurav, S. S., & Kumar, A. (2013). Molecular Basis of Aluminium Toxicity in Plants: A Review. *American Journal of Plant Sciences*, *4*(12), 21–37. https://doi.org/10.4236/ajps.2013.412a3004

Haasl, R. J., & Payseur, B. A. (2011). Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, *106*(1), 158–171. https://doi.org/10.1038/hdy.2010.21

Hansen, O. K., Changtragoon, S., Ponoy, B., Kjær, E. D., Minn, Y., Finkeldey, R., … Graudal, L. (2015). Genetic resources of teak (*Tectona grandis* Linn. f.)—strong genetic structure among natural populations. *Tree Genetics and Genomes*, *11*(1), 802. https://doi.org/10.1007/s11295-014-0802-5

Hardy, O. J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J. Ô., … Poncet, V. (2013). Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus - Geoscience*, *345*(7–8), 284–296. https://doi.org/10.1016/j.crte.2013.05.001

Hardy, O. J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M. H., … Degen, B. (2006). Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, *15*(2), 559–571. https://doi.org/10.1111/j.1365-294X.2005.02785.x

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi : a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*, 618–620. https://doi.org/10.1046/j.1471-8278 .2002.00305.x

Hardy, O. J., & Vekemans, X. (2013). SPAGeDi 1.4 a program for Spatial Pattern Analysis of Genetic Diversity. User´s manual.

Hart, M. W. (2011). The species concept as an emergent property of population biology. *Evolution*, *65*(3), 613–616. https://doi.org/10.1111/j.1558-5646.2010.01202.x

Harvey, M. G., & Brumfield, R. T. (2014). Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *ArXiv Preprint ArXiv:1405.6571*, *83*, 305–316. https://doi.org/10.1016/j.ympev.2014.10.023

Helmstetter, A. J., Amoussou, B. E. N., Bethune, K., Kamdem, N. G., Glèlè Kakaï, R., Sonké, B., & Couvreur, T. L. P. (2020). Phylogenomic approaches reveal how climate shapes patterns of genetic diversity in an African rain forest tree species. *Molecular Ecology*, *29*(18), 3560–3573. https://doi.org/10.1111/mec.15572

Helmstetter, A. J., Béthune, K., Kamdem, N. G., Sonké, B., & Couvreur, T. L. P. (2020). Individualistic evolutionary responses of Central African rain forest plants to Pleistocene climatic fluctuations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(51), 32509–32518. https://doi.org/10.1073/pnas.2001018117

Hendry, A. P., & Day, T. (2005). Population structure attributable to reproductive time: Isolation by time and adaptation by time. *Molecular Ecology*, *14*(4), 901–916. https://doi.org/10.1111/j.1365-294X.2005.02480.x

Hendry, A. P., Nosil, P., & Rieseberg, L. H. (2007). The speed of ecological speciation. *Functional Ecology*, *21*(3), 455–464. https://doi.org/10.1111/j.1365-2435.2007.01240.x

Henry, O., Feer, F., & Sabatier, D. (2000). Diet of the Lowland Tapir (*Tapirus terrestris* L.) in French Guiana. *Biotropica*, *32*(2), 364–368. https://doi.org/10.1111/j.1744-7429.2000.tb00480.x

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, *173*(5), 579–588. https://doi.org/10.1086/597611

Heuertz, M., Duminil, J., Dauby, G., Savolainen, V., & Hardy, O. J. (2014). Comparative phylogeography in rainforest trees from lower Guinea, Africa. *PLoS ONE*, *9*(1), e84307. https://doi.org/10.1371/journal.pone.0084307

Heuertz, M., Vekemans, X., Hausman, J. F., Palada, M., & Hardy, O. J. (2003). Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, *12*(9), 2483–2495. https://doi.org/10.1046/j.1365-294X.2003.01923.x

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., … Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, *54*(1), 291–301. https://doi.org/10.1016/j.ympev.2009.09.016

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Holbrook, K. M., & Smith, T. B. (2000). Seed dispersal and movement patterns in two species of *Ceratogymna hornbills* in a West African tropical lowland forest. *Oecologia*, *125*, 249–257. https://doi.org/10.1007/s004420000445

Hoskin, C. J., & Higgie, M. (2010). Speciation via species interactions: The divergence of mating traits within species. *Ecology Letters*, *13*(4), 409–420. https://doi.org/10.1111/j.1461-0248.2010.01448.x

Hubbell, S. P. (2005). Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, *19*(1), 166–172. https://doi.org/10.1111/j.0269-8463.2005.00965.x

Jan-du-Chêne, R., Onyike, M., & Sowunmi, M. (1978). Some new Eocene pollen of the Ogwashi-Asaba Formation, southeastern Nigeria. *Revista Española de Micropaleontologia*, *10*, 285–322.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Clarendon Press.

Jha, S., & Dick, C. W. (2010). Native bees mediate long-distance pollen dispersal in a shade coffee landscape mosaic. *Proceedings of the National Academy of Sciences of the United*

States of America, *107*(31), 13760–13764. https://doi.org/10.1073/pnas.1002490107

Johnson, N. A., Smith, C. H., Pfeiffer, J. M., Randklev, C. R., Williams, J. D., & Austin, J. D. (2018). Integrative taxonomy resolves taxonomic uncertainty for freshwater mussels being considered for protection under the U.S. Endangered Species Act. *Scientific Reports*, *8*(1), 15892. https://doi.org/10.1038/s41598-018-33806-z

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/doi: 10.1186/1471-2156-11-94

Jombart, T., Devillard, S., Dufour, A.-B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, *101*(1), 92–103. https://doi.org/10.1038/hdy.2008.34

Jones, A., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Dewitte, O., … Zougmoré, R. (Eds.). (2013). *Soil Atlas of Africa*. Luxembourg: European Commission, Publications Office of the European Union.

Jones, F. A., Cerón-Souza, I., Hardesty, B. D., & Dick, C. W. (2013). Genetic evidence of Quaternary demographic changes in four rain forest tree species sampled across the Isthmus of Panama. *Journal of Biogeography*, *40*(4), 720–731. https://doi.org/10.1111/jbi.12037

Jones, P. J. (1994). Biodiversity in the Gulf of Guinea: an overview. *Biodiversity and Conservation*, *3*(9), 772–784. https://doi.org/10.1007/BF00129657

Jordano, P., García, C., Godoy, J. A., & García-Castaño, J. L. (2007). Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3278–3282. https://doi.org/10.1073/pnas.0606793104

Kalinowski, S. T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity*, *88*(1), 62–65. https://doi.org/10.1038/sj.hdy.6800009

Kalinowski, S. T. (2009). How well do evolutionary trees describe genetic relationships among populations? *Heredity*, *102*(5), 506–513. https://doi.org/10.1038/hdy.2008.136

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, *106*(4), 625–632. https://doi.org/10.1038/hdy.2010.95

Kalisz, S., Nason, J. D., Hanzawa, F. M., & Tonsor, S. J. (2001). Spatial population genetic structure in *Trillium grandiflorum*: The roles of dispersal, mating, history, and selection. *Evolution*, *55*(8), 1560–1568. https://doi.org/10.1111/j.0014-3820.2001.tb00675.x

Kaplin, B. A., & Lambert, J. E. (2002). Effectiveness of seed dispersal by *Cercopithecus* monkeys: implications for seed input into degraded areas. In D. J. Levey, W. R. Silva, & M. Galetti (Eds.), *Seed dispersal and frugivory: ecology, evolution and conservation* (pp. 351–364). Wallinforf, UK: CABI Publishing. Retrieved from http://www.cabdirect.org/abstracts/20023028597.html

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kerkhoff, A. J., Moriarty, P. E., & Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8125–8130. https://doi.org/10.1073/pnas.1308932111

Kim, D. Y., Bovet, L., Maeshima, M., Martinoia, E., & Lee, Y. (2007). The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant Journal*, *50*(2), 207–218. https://doi.org/10.1111/j.1365-313X.2007.03044.x

Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, *49*(4), 561–576.

Knowles, L. L. (2009a). Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology*, *58*(5), 463–467. https://doi.org/10.1093/sysbio/syp061

Knowles, L. L. (2009b). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 593–612. https://doi.org/10.1146/annurev.ecolsys.38.091206.095702

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., … Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576. https://doi.org/10.1101/gr.129684.111

Kochian, L. V., Hoekenga, O. A., & Piñeros, M. A. (2004). How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. *Annual Review of Plant Biology*, *55*, 459–493. https://doi.org/10.1146/annurev.arplant.55.031903.141655

Kochian, L. V., Piñeros, M. A., Liu, J., & Magalhaes, J. V. (2015). Plant adaptation to acid soils: The molecular basis for crop aluminum resistance. *Annual Review of Plant Biology*, *66*, 571–598. https://doi.org/10.1146/annurev-arplant-043014-114822

Köhler, C., Mittelsten Scheid, O., & Erilova, A. (2010). The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics*, *26*(3), 142–148. https://doi.org/10.1016/j.tig.2009.12.006

Körner, C. (2007). The use of "altitude" in ecological research. *Trends in Ecology and Evolution*, *22*(11), 569–574. https://doi.org/10.1016/j.tree.2007.09.006

Kostamo, K., Korpelainen, H., & Olsson, S. (2012). Comparative study on the population genetics of the red algae *Furcellaria lumbricalis* occupying different salinity conditions. *Marine Biology*, *159*(3), 561–571. https://doi.org/10.1007/s00227-011-1835-z

Kraaijeveld, K., Kraaijeveld-Smit, F. J. L., & Maan, M. E. (2011). Sexual selection and speciation: The comparative evidence revisited. *Biological Reviews*, *86*(2), 367–377. https://doi.org/10.1111/j.1469-185X.2010.00150.x

Kremer, A., Kleinschmit, J., Cottrell, J., Cundall, E. P., Deans, J. D., Ducousso, A., … Stephan, B. R. (2002). Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management*, *156*, 75–87. https://doi.org/10.1016/S0378-1127(01)00635-1

Kursar, T. A., Dexter, K. G., Lokvam, J., Pennington, R. T., Richardson, J. E., Weber, M. G., … Coley, P. D. (2009). The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy*

of Sciences of the United States of America, *106*(43), 18073–18078. https://doi.org/10.1073/pnas.0904786106

Lacerda, A. E. B. de, & Nimmo, E. R. (2010). Can we really manage tropical forests without knowing the species within? Getting back to the basics of forest management through taxonomy. *Forest Ecology and Management*, *259*(5), 995–1002. https://doi.org/10.1016/j.foreco.2009.12.005

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lapenta, M. J., & Procópio-de-Oliveira, P. (2008). Some aspects of seed dispersal effectiveness of golden lion tamarins (Leontopithecus rosalia) in a Brazilian Atlantic Forest. *Tropical Conservation Science*, *1*, 122–139.

Lapenta, M. J., Procópio de Oliveira, P., Kierluff, M. C. M., & Motta-Junior, J. C. (2003). Fruit exploitation by Golden Lion Tamarins (*Leontopithecus rosalia*) in the União Biological Reserve, Rio das Ostras, RJ - Brazil. *Mammalia*, *67*(1), 41–46. https://doi.org/10.1515/mamm.2003.67.1.41

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*(7), 1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP Data. *Systematic Biology*, *63*(4), 534–542. https://doi.org/10.1093/sysbio/syu018

Leavitt, S. D., Divakar, P. K., Crespo, A., & Lumbsch, H. T. (2016). A Matter of Time — Understanding the Limits of the Power of Molecular Data for Delimiting Species Boundaries. *Herzogia*, *29*(2), 479–492. https://doi.org/10.13158/heia.29.2.2016.479

Lee, C. R., & Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, *20*(22), 4631–4642. https://doi.org/10.1111/j.1365-294X.2011.05310.x

Lee, D. C., Halliday, A. N., Fitton, J. G., & Poli, G. (1994). Isotopic variations with distance and time in the volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth and Planetary Science Letters*, *123*(1–3), 119–138. https://doi.org/10.1016/0012-821X(94)90262-3

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*(12), e4010. https://doi.org/10.1371/journal.pone.0004010

Leitch, I. J., & Bennett, M. D. (2004). Genome downsizing in polyploid plants I. *Biological Journal of the Linnean Society*, *82*(4), 651–663. Retrieved from http://www.mendeley.com/research/biological-relevance-polyploidy-ecology-genomics-polyploidy-arctic-plants/

Leite, Y. L. R., Costa, L. P., Loss, A. C., Rocha, R. G., Batalha-Filho, H., Bastos, A. C., … Pardini, R. (2016). Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), 1008–1013. https://doi.org/10.1073/pnas.1513062113

Leslie, S., Hellenthal, G., Winney, B., Davisona, D., Boumertit, A., Day, T., … Robinsonll, M. (2015). Fine scale genetic structure of the British population. *Nature*, *519*, 309–314. https://doi.org/10.1038/nature14230

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linder, H. P. (2001). Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography*, *28*(2), 169–182. https://doi.org/10.1046/j.1365-2699.2001.00527.x

Linhart, Y. B., & Mendenhall, J. A. (1977). Pollen Dispersal by Hawkmoths in a *Lindenia rivalis* Benth. Population in Belize. *Biotropica*, *9*(2), 143.

Ljungqvist, M., Åkesson, M., & Hansson, B. (2010). Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Molecular Ecology*, *19*(5), 851–855. https://doi.org/10.1111/j.1365-294X.2010.04522.x

Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial Genetic Structure of a Tropical Understory Shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, *82*(11), 1420–1425.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*(9), 2178–2192. https://doi.org/10.1111/mec.12725

Loureiro, J., Kopecký, D., Castro, S., Santos, C., & Silveira, P. (2007). Flow cytometric and cytogenetic analyses of Iberian Peninsula Festuca spp. *Plant Systematics and Evolution*, *269*, 89–105. https://doi.org/10.1007/s00606-007-0564-8

Loureiro, J., Rodriguez, E., Dolezel, J., & Santos, C. (2007). Two New Nuclear Isolation Buffers for Plant DNA Flow Cytometry: A Test with 37 Species, *100*, 875–888. https://doi.org/10.1093/annbot/mcm152

Loveless, J. L., & Hamrick, M. D. (1984). Ecological Determinants of Genetic Structure in plant populations. *Annual Review of Ecology and Systematics*, *15*, 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433

Maguire, B. (1976). Apomixis in the genus *Clusia* (Clusiaceae). A preliminary report. *Taxon*, *25*(2/3), 241–244. https://doi.org/10.2307/1219446

Mairal, M., Sanmartín, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J. J., & Alarcón, M. (2017). Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports*, *7*, 45749. https://doi.org/10.1038/srep45749

Malécot, G. (1951). Quelques schémas probabilistes sur la variabilité des populations naturelles. *Annales de L´Université de Lyon*, *13*, 339–340.

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. Section B. Biological Sciences*, *104*, 31–73. https://doi.org/10.1017/S0269727000006114

Maley, J., & Brenac, P. (1998). Vegetation dynamics, palaeoenvironments and climatic changes

in the forests of western Cameroon during the last 28,000 years B.P. *Review of Palaeobotany and Palynology*, *99*(2), 157–187. https://doi.org/10.1016/S0034-6667(97)00047-X

Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular Ecology*, *24*(8), 1856–1872. https://doi.org/10.1111/mec.13118

Mann, C. F., Cheke, R. A., & Allen, R. (2001). *A Guide to the Sunbirds, Flowerpeckers, Spiderhunters and Sugarbirds of the world*. London: Christopher Helm Publishers.

Margarido, G. R. A., & Heckerman, D. (2015). ConPADE: genome assembly ploidy estimation from next-generation sequencing data. *PLoS Computational Biology*, *11*(4), e1004229. https://doi.org/10.1371/journal.pcbi.1004229

Marquardt, P. E., & Epperson, B. K. (2004). Spatial and population genetic structure of microsatellites in white pine. *Molecular Ecology*, *13*(11), 3305–3315. https://doi.org/10.1111/j.1365-294X.2004.02341.x

Marske, K. (2016). Phylogeography. In R. M. Kliman (Ed.), *Encyclopedia of Evolutionary Biology* (Vol. 3, pp. 291–296). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-800049-6.00109-8

Marske, K. A., Rahbek, C., & Nogués-Bravo, D. (2013). Phylogeography: Spanning the ecology-evolution continuum. *Ecography*, *36*(11), 1169–1181. https://doi.org/10.1111/j.1600-0587.2013.00244.x

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

McCain, C. M. (2005). Elevational gradients in diversity of small mammals. *Ecology*, *86*(2), 366–372. https://doi.org/10.1890/03-3147

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526–538. https://doi.org/10.1016/j.ympev.2011.12.007

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. https://doi.org/10.1111/1755-0998.12613

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Melo, F. P. L., Rodriguez-Herrera, B., Chazdon, R. L., Medellin, R. A., & Ceballos, G. G. (2009). Small Tent-Roosting Bats Promote Dispersal of Large-Seeded Plants in a Neotropical Forest. *Biotropica*, *41*(6), 737–743. https://doi.org/10.1111/j.1744-7429.2009.00528.x

Melo, W. A., Vieira, L. D., Novaes, E., Bacon, C. D., & Collevatti, R. G. (2020). Selective Sweeps Lead to Evolutionary Success in an Amazonian Hyperdominant Palm. *Frontiers in Genetics*, *11*, 596662. https://doi.org/10.3389/fgene.2020.596662

Meyers, J. B., Rosendahl, B. R., Harrison, C. G. A., & Ding, Z. D. (1998). Deep-imaging seismic and gravity results from the offshore Cameroon Volcanic Line, and speculation of African hotlines. *Tectonophysics*, *284*(1–2), 31–63. https://doi.org/10.1016/S0040-1951(97)00173-X

Migliore, J., Kaymak, E., Mariac, C., Couvreur, T. L. P., Lissambou, B., Piñeiro, R., & Hardy, O. J. (2019). Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from Greenwayodendron (Annonaceae) phylogenomics. *Journal of Biogeography*, *46*(1), 212–223. https://doi.org/10.1111/jbi.13476

Miller, K. E., & Dietz, J. M. (2004). Effects of Individual and Group Characteristics on Feeding Behaviors in Wild *Leontopithecus rosalia*. *International Journal of Primatology*, *25*, 27–39. https://doi.org/10.1007/s10764-005-8854-7

Misiewicz, T. M., & Fine, P. V. A. (2014). Evidence for ecological divergence across a mosaic of soil types in an Amazonian tropical tree: *Protium subserratum* (Burseraceae). *Molecular Ecology*, *23*(10), 2543–2558. https://doi.org/10.1111/mec.12746

Mitton, J. B., Kreiser, B. R., & Latta, R. G. (2000). Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. *Molecular Ecology*, *9*(1), 91–97. https://doi.org/10.1046/j.1365-294X.2000.00840.x

Molino, J., & Sabatier, D. (2001). Tree Diversity in Tropical Rain Forests: A Validation of the Intermediate Disturbance Hypothesis. *Science*, *294*(5547), 1702–1704.

Morin, P. A., Luikart, G., Wayne, R. K., & the SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, *19*(4), 208–216. https://doi.org/10.1016/j.tree.2004.01.009

Morin, P. A., Manaster, C., Mesnick, S. L., & Holland, R. (2009). Normalization and binning of historical and multi-source microsatellite data: Overcoming the problems of allele size shift with allelogram. *Molecular Ecology Resources*, *9*(6), 1451–1455. https://doi.org/10.1111/j.1755-0998.2009.02672.x

Morrison, D. W. (1978). Foraging Ecology and Energetics of the Frugivorous Bat Artibeus jamaicensis. *Ecology*, *59*, 716–723. https://doi.org/10.2307/1938775

Morrison, D. W. (1980). Foraging and Day-Roosting Dynamics of Canopy Fruit Bats in Panama. *Journal of Mammalogy*, *61*(1), 20–29. https://doi.org/10.2307/1379953

Müller, K., Quandt, D., Müller, J., & Neinhuis, C. (2006). PhyDE®: Phylogenetic Data Editor, version 0.995. Retrieved from http://www.phyde.de.

Myers, E. A., McKelvy, A. D., & Burbrink, F. T. (2020). Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Molecular Ecology*, *29*(4), 797–811. https://doi.org/10.1111/mec.15358

Myers, N., Mittermeler, R. A., Mittermeler, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858. https://doi.org/10.1038/35002501

Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., & Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecology and Evolution*, *6*(24), 8649–8664. https://doi.org/10.1002/ece3.2550

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. a. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847. https://doi.org/10.1111/mec.12350

Narum, S. R., & Hess, J. E. (2011). Comparison of F ST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*(SUPPL. 1), 184–194.

https://doi.org/10.1111/j.1755-0998.2011.02987.x

Ndiade-Bourobou, D., Daïnou, K., Hardy, O. J., Doumenge, C., Tosso, F., & Bouvet, J. M. (2020). Revisiting the North-South genetic discontinuity in Central African tree populations: the case of the low-density tree species *Baillonella toxisperma*. *Tree Genetics and Genomes*, *16*(1), 1–11. https://doi.org/10.1007/s11295-019-1408-8

Ndiade-Bourobou, D., Hardy, O. J., Favreau, B., Moussavou, H., Nzengue, E., Mignot, A., & Bouvet, J. M. (2010). Long-distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. *Molecular Ecology*, *19*(22), 4949–4962. https://doi.org/10.1111/j.1365-294X.2010.04864.x

Neale, D. B., & Kremer, A. (2011). Forest tree genomics: Growing resources and applications. *Nature Reviews Genetics*, *12*(2), 111–122. https://doi.org/10.1038/nrg2931

Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist*, *106*(949), 283–292. https://doi.org/http://www.jstor.org/stable/2459777?origin=JSTOR-pdf

Nielsen, R., & Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Molecular Ecology*, *18*(6), 1034–1047. https://doi.org/10.1111/j.1365-294X.2008.04059.x

Nistelberger, H. M., Tapper, S. L., Coates, D. J., McArthur, S. L., & Byrne, M. (2021). As old as the hills: Pliocene palaeogeographical processes influence patterns of genetic structure in the widespread, common shrub *Banksia sessilis*. *Ecology and Evolution*, *11*(2), 1069–1082. https://doi.org/10.1002/ece3.7127

Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Integrating genomic and phenotypic data to evaluate alternative phylogenetic and species delimitation hypotheses in a recent evolutionary radiation of grasshoppers. *Molecular Ecology*, *27*(5), 1229–1244. https://doi.org/10.1111/mec.14504

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, *62*(2), 316–336. https://doi.org/10.1111/j.1558-5646.2007.00299.x

Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van Den Boom, D., & Jurinke, C. (2007). *iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassARRAY ® System Through Single Base Primer Extension with Mass-Modified Terminators*.

Olsson, S., Seoane-Zonjic, P., Bautista, R., Claros, M. G., González-Martínez, S. C., Scotti, I., … Heuertz, M. (2017). Development of genomic tools in a widespread tropical tree, *Symphonia globulifera* L.f.: a new low-coverage draft genome, SNP and SSR markers. *Molecular Ecology Resources*, *17*(4), 614–630. https://doi.org/10.1111/1755-0998.12605

Orozco-terWengel, P., Corander, J., & Schlötterer, C. (2011). Genealogical lineage sorting leads to significant, but incorrect Bayesian multilocus inference of population structure. *Molecular Ecology*, *20*(6), 1108–1121. https://doi.org/10.1111/j.1365-294X.2010.04990.x

Ortega, J., & Castro-Arellano, I. (2001). *Artibeus jamicensis*. *Mammalian Species*, *662*, 1–9.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., … Lynch, M. (2010). The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science*, *327*(5961), 92–94. https://doi.org/10.1126/science.1180677

Otto, S. P., & Whitton, J. (2000). Polyploid Incidence and Evolution. *Annual Review of*

Genetics, *34*(1), 401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Oyen, L. (2005). *Symphonia globulifera* L.f. (Internet) Record from Protabase. In: Louppe D, Oteng-Amoako AA, Brink M (eds.) PROTA (Plant Resources of Tropical Africa/ Ressources vegetales de l'Afrique tropicale).Wageningen, Netherlands.

Pais, A. L., Whetten, R. W., & Xiang, Q. Y. J. (2017). Ecological genomics of local adaptation in *Cornus florida* L. by genotyping by sequencing. *Ecology and Evolution*, *7*(1), 441–465. https://doi.org/10.1002/ece3.2623

Pante, E., Schoelinck, C., & Puillandre, N. (2015). From integrative taxonomy to species description: One step beyond. *Systematic Biology*, *64*(1), 152–160. https://doi.org/10.1093/sysbio/syu083

Paquette, S. R. (2015). Package PopGenKit: Useful functions for (batch) file conversion and data resampling in microsatellite datasets. R package version 1.0. Retrieved from https://cran.r-project.org/web/packages/PopGenKit/

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*(4), 445–464. https://doi.org/10.1111/2041-210X.12324

Pascarella, J. B. (1992). Notes on flowering phenology, nectar robbing and pollination of *Symphonia globulifera* L.f. (Clusiaceae) in a lowland rain forest in Costa Rica. *Brenesia*, *38*, 83–86.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Pennington, R. T., & Dick, C. W. (2004). The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1450), 1611–1622. https://doi.org/10.1098/rstb.2004.1532

Pennington, R. T., Richardson, J. E., & Lavin, M. (2006). Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytologist*, *172*(4), 605–616. https://doi.org/10.1111/j.1469-8137.2006.01902.x

Pérez-Espona, S., Pérez-Barbería, F. J., Mcleod, J. E., Jiggins, C. D., Gordon, I. J., & Pemberton, J. M. (2008). Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, *17*(4), 981–996. https://doi.org/10.1111/j.1365-294X.2007.03629.x

Perrier de la Bâthie, H. (1951). Guttiferes. In H. Humbert & J.-F. Leroy (Eds.), *Flore de Madagascar et des Comores (Plantes vasculaires)*. Tananarive, Paris: Imprimerie officielle. Museum national d'histoire naturelle.

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., … Vendramin, G. G. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, *300*(5625), 1563–1565. https://doi.org/10.1126/science.1083264

Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., … Kremer, A. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, *156*(1–3), 49–74. https://doi.org/10.1016/S0378-1127(01)00634-X

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution*, *24*(7), 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Petit, R. J., Feng, S. H., & Dick, C. W. (2008). Forests of the past: A window to future changes. *Science*, *320*(5882), 1450–1452. https://doi.org/10.1126/science.1155457

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Petkova, D., Novembre, J., & Stephens, M. (2015). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Piñeiro, R., Dauby, G., Kaymak, E., & Hardy, O. J. (2017). Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the ice ages. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1866), 20171800. https://doi.org/10.1098/rspb.2017.1800

Pinheiro, F., Dantas-Queiroz, M. V., & Palma-Silva, C. (2018). Plant Species Complexes as Models to Understand Speciation and Evolution: A Review of South American Studies. *Critical Reviews in Plant Sciences*, *37*(1), 54–80. https://doi.org/10.1080/07352689.2018.1471565

Piotti, A., Leonardi, S., Heuertz, M., Buiteveld, J., Geburek, T., Gerber, S., … Vendramin, G. G. (2013). Within-Population Genetic Structure in Beech (Fagus sylvatica L.) Stands Characterized by Different Disturbance Histories: Does Forest Management Simplify Population Substructure? *PLoS ONE*, *8*(9), e73391. https://doi.org/10.1371/journal.pone.0073391

Plomion, C., Bastien, C., Bogeat-Triboulot, M. B., Bouffier, L., Déjardin, A., Duplessis, S., … Vacher, C. (2016). Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science*, *73*(1), 77–103. https://doi.org/10.1007/s13595-015-0488-3

Poulsen, J. R., Clark, C. J., & Smith, T. B. (2001). Seed dispersal by a diurnal primate community in the Dja Reserve, Cameroon. *Journal of Tropical Ecology*, *17*(6), 787–808. https://doi.org/10.1017/S0266467401001602

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation - Not by sweeps alone. *Nature Reviews Genetics*, *11*(10), 665–667. https://doi.org/10.1038/nrg2880

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Pritchard, J. K., Wen, X., & Falush, D. (2010). *Documentation for structure software: Version 2.3*.

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, *4*(22), 4399–4428. https://doi.org/10.1002/ece3.1305

Pyron, R. A., & Burbrink, F. T. (2010). Hard and soft allopatry: Physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography*, *37*(10), 2005–2015. https://doi.org/10.1111/j.1365-2699.2010.02336.x

QGIS Development Team. (2014). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from http://qgis.osgeo.org

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/.

Rambaut, A., & Drummond, A. (2016). TreeAnnotator v2.4.1: MCMC Output Analysis. Retrieved from http://www.beast2.org/

Rambaut, A., Suchard, M., Xie, D., & Drummond, A. (2014). Tracer v1.6. Retrieved from http://beast.bio.ed.ac.uk/Tracer

Rannala, B., & Yang, Z. (2003). Using DNA Sequences From Multiple Loci. *Genetics*, *164*, 1645–1656.

Ren, G., Mateo, R. G., Guisan, A., Conti, E., & Salamin, N. (2018). Species divergence and maintenance of species cohesion of three closely related *Primula* species in the Qinghai–Tibet Plateau. *Journal of Biogeography*, *45*(11), 2495–2507. https://doi.org/10.1111/jbi.13415

Renner, S. (2004). Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences*, *165*(4 SUPPL.), S23–S33. https://doi.org/10.1086/383334

Rhodes, M. K., Fant, J. B., & Skogen, K. A. (2014). Local Topography Shapes Fine-Scale Spatial Genetic Structure in the Arkansas Valley Evening Primrose, *Oenothera harringtonii* (Onagraceae). *Journal of Heredity*, *105*(6), 900–909. https://doi.org/10.1093/jhered/esu051

Richards, A. J. (1990). Studies in *Garcinia*, dioecious tropical forest trees: agamospermy. *Botanical Journal of the Linnean Society*, *103*, 233–250. https://doi.org/10.1111/j.1095-8339.1990.tb00186.x

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, *29*(3), 165–176. https://doi.org/10.1016/j.tree.2014.01.002

Robledo-Arnuncio, J. J., Collada, C., Alía, R., & Gil, L. (2005). Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*, *32*(4), 595–605. https://doi.org/10.1111/j.1365-2699.2004.01196.x

Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of Genetic Markers for Inference of Ancestry. *American Journal of Human Genetics*, *73*(6), 1402–1422. https://doi.org/10.1086/380416

Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, *13*(1), 58–62. https://doi.org/10.1046/j.1420-9101.2000.00137.x

Ruiz Daniels, R., Taylor, R. S., González-Martínez, S. C., Vendramin, G. G., Fady, B., Oddou-Muratorio, S., … Beaumont, M. A. (2019). Looking for Local Adaptation: Convergent Microevolution in Aleppo Pine (*Pinus halepensis*). *Genes*, *10*(9), 673. https://doi.org/10.3390/genes10090673

Sabatier, D., Grimaldi, M., Prévost, M. F., Guillaume, J., Godron, M., Dosso, M., & Curmi, P. (1997). The influence of soil cover organization on the floristic and structural heterogeneity of a Guianan rain forest. *Plant Ecology*, *131*(1), 81–108. https://doi.org/10.1023/A:1009775025850

Sagnard, F., Oddou-Muratorio, S., Pichot, C., Vendramin, G. G., & Fady, B. (2011). Effects of seed dispersal, adult tree and seedling density on the spatial genetic structure of regeneration at fine temporal and spatial scales. *Tree Genetics & Genomes*, *7*(1), 37–48. https://doi.org/10.1007/s11295-010-0313-y

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Salzmann, U., & Hoelzmann, P. (2005). The Dahomey Gap: An abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene*, *15*(2), 190–199. https://doi.org/10.1191/0959683605hl799rp

Sanfiorenzo, A. (2018). Potential Pollinators of understory populations of *Symphonia globulifera* in the Neotropics. *Journal of Pollination Ecology*, *22*(1), 1–10. https://doi.org/10.26786/1920-7603(2018)one

Sang, T., Crawford, D. J., & Stuessy, T. F. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, *84*(9), 1120–1136.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. https://doi.org/10.1038/nrg3522

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology*, *55*, 421–438. https://doi.org/10.1146/annurev-ento-112408-085432

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, *323*(5915), 737–741. https://doi.org/10.1126/science.1160006

Schmitt, S. (2020). *Génomique écologique de l'exploitation de niche et de la performance individuelle chez les arbres forestiers tropicaux.* (Doctoral dissertation). Université de Bordeaux, France.

Schmitt, S., Hérault, B., Ducouret, É., Baranger, A., Tysklind, N., Heuertz, M., … Derroire, G. (2020). Topography consistently drives intra- and inter-specific leaf trait variation within tree species complexes in a Neotropical forest. *Oikos*, *129*(10), 1521–1530. https://doi.org/10.1111/oik.07488

Schmitt, S., Tysklind, N., Derroire, G., Heuertz, M., & Hérault, B. (2021). Topography shapes the local coexistence of tree species within species complexes of Neotropical forests. *Oecologia*, *196*, 389–398. https://doi.org/10.1007/s00442-021-04939-2

Schmitt, S., Tysklind, N., Hérault, B., & Heuertz, M. (2021). Topography drives microgeographic adaptations of closely related species in two tropical tree species complexes. *Molecular Ecology*, *30*(20), 5080–5093. https://doi.org/10.1111/mec.16116

Schupp, E. W. (1993). Quantity, quality and the effectiveness of seed dispersal by animals. *Vegetatio*, *107/108*, 15–29.

Schwartz, M. K., & McKelvey, K. S. (2009). Why sampling scheme matters: The effect of sampling scheme on landscape genetic results. *Conservation Genetics*, *10*(2), 441–452. https://doi.org/10.1007/s10592-008-9622-1

Scotti, I., González-Martínez, S. C., Budde, K. B., & Lalagüe, H. (2015). Fifty years of genetic studies: what to make of the large amounts of variation found within populations? *Annals of Forest Science*, *73*(1), 69–75. https://doi.org/10.1007/s13595-015-0471-z

Sebbenn, A. M., Blanc-Jolivet, C., Mader, M., Meyer-Sand, B. R. V., Paredes-Villanueva, K., Honorio Coronado, E. N., … Degen, B. (2019). Nuclear and plastidial SNP and INDEL markers for genetic tracking studies of *Jacaranda copaia*. *Conservation Genetics Resources*, *11*(3), 341–343. https://doi.org/10.1007/s12686-019-01097-9

Seoane, P., Ocaña, S., Carmona, R., Bautista, R., Madrid, E., M. Torres, A., & Gonzalo Claros, M. (2016). AutoFlow, a Versatile Workflow Engine Illustrated by Assembling an Optimised de novo Transcriptome for a Non-Model Species, such as Faba Bean (*Vicia faba*). *Current Bioinformatics*, *11*(4), 440–450. https://doi.org/10.2174/1574893611666160212235117

Serra-Varela, M. J., Grivet, D., Vincenot, L., Broennimann, O., Gonzalo-Jiménez, J., & Zimmermann, N. E. (2015). Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* Ait.). *Global Ecology and Biogeography*, *24*(11), 1302–1313. https://doi.org/10.1111/geb.12369

Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, *16*(7), 940–950. https://doi.org/10.1111/ele.12120

Sievers, F., & Higgins, D. G. (2014). Clustal Omega. *Current Protocols in Bioinformatics*, *48*, 3.13.1-3.13.16. https://doi.org/10.1002/0471250953.bi0313s48

Sim-Sim, M., Afonina, O. M., Almeida, T., Désamoré, A., Laenen, B., Garcia, C. A., … Stech, M. (2017). Integrative taxonomy reveals too extensive lumping and a new species in the moss genus *Amphidium* (Bryophyta). *Systematics and Biodiversity*, *15*(5), 451–463. https://doi.org/10.1080/14772000.2016.1271059

Sites, J. W., & Marshall, J. C. (2003). Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, *18*(9), 462–470. https://doi.org/10.1016/S0169-5347(03)00184-8

Slik, J. W. F., Arroyo-Rodríguez, V., Aiba, S. I., Alvarez-Loayza, P., Alves, L. F., Ashton, P., … Venticinque, E. M. (2015). Correction for "An estimate of the number of tropical tree species." *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), E4628–E4629. https://doi.org/10.1073/pnas.1512611112

Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*(2), 295–315. https://doi.org/10.1111/j.1558-5646.2009.00877.x

Solís-Lemus, C., Yang, M., & Ané, C. (2016). Inconsistency of Species Tree Methods under Gene Flow. *Systematic Biology*, *65*(5), 843–851. https://doi.org/10.1093/sysbio/syw030

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., … Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96*(1), 336–348. https://doi.org/10.3732/ajb.0800079

Soltis, P. S. (2015). Polyploidy and genome evolution in plants. *Molecular Ecology Resources*, *18*(6), 135–141. https://doi.org/10.1038/nbt0493-508

Soltis, P. S., & Soltis, D. E. (2009). The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology*, *60*(1), 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039

Sork, V. L. (2018). Genomic Studies of Local Adaptation in Natural Plant Populations. *Journal of Heredity*, *109*(1), 3–15. https://doi.org/10.1093/jhered/esx091

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., & Neale, D. B. (2013). Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes*, *9*(4), 901–911. https://doi.org/10.1007/s11295-013-0596-x

Sosef, M. S. M., Dauby, G., Blach-Overgaard, A., van der Burgt, X., Catarino, L., Damen, T., … Couvreur, T. L. P. (2017). Exploring the floristic diversity of tropical Africa. *BMC Biology*, *15*(1), 15. https://doi.org/10.1186/s12915-017-0356-8

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., … Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, *25*(12), 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. *Molecular Ecology*, *15*(8), 2109–2118. https://doi.org/10.1111/j.1365-294X.2006.02926.x

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: Where are we now? *Molecular Ecology*, *19*(17), 3496–3514. https://doi.org/10.1111/j.1365-294X.2010.04691.x

Suchel, J. B. (1990). Les modalités du passage du régime climatique boréal au régime climatique austral dans le Sud-Ouest camerounais. *Cahier Du Centre de Recherche de Climatologie.Université de Bourgogne, Dijon*, *13*, 63–76.

Suda, J., & Trávnícek, P. (2006). Reliable DNA Ploidy Determination in Dehydrated Tissues of Vascular Plants by DAPI Flow Cytometry—New Prospects for Plant Research. *Cytometry. Part A*, *69A*(4), 273–280. https://doi.org/10.1002/cyto.a

Sutherland, W. J., Freckleton, R. P., Godfray, H. C. J., Beissinger, S. R., Benton, T., Cameron, D. D., … Wiegand, T. (2013). Identification of 100 fundamental ecological questions. *Journal of Ecology*, *101*(1), 58–67. https://doi.org/10.1111/1365-2745.12025

Svenning, J. C., Fløjgaard, C., Marske, K. A., Nógues-Bravo, D., & Normand, S. (2011). Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, *30*(21–22), 2930–2947. https://doi.org/10.1016/j.quascirev.2011.06.012

Sweeney, P. W. (2008). Phylogeny and Floral Diversity in the Genus *Garcinia* (Clusiaceae) and Relatives. *International Journal of Plant Sciences*, *169*(9), 1288–1303. https://doi.org/10.1086/591990

Tchouto, M. G. P., de Wilde, J. J. F. E., de Boer, W. F., van der Maesen, L. J. G., & Cleef, a. M. (2009). Bio-indicator species and Central African rain forest refuges in the Campo-

Ma'an area, Cameroon. *Systematics and Biodiversity*, *7*(1), 21–31. https://doi.org/10.1017/S1477200008002892

Ter Steege, H., Vaessen, R. W., Cárdenas-López, D., Sabatier, D., Antonelli, A., De Oliveira, S. M., … Salomão, R. P. (2016). The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Scientific Reports*, *6*, 29549. https://doi.org/10.1038/srep29549

The Plant List. Version 1.1. (2013). Retrieved June 27, 2016, from http://www.theplantlist.org/

Thom, G., Amaral, F. R. Do, Hickerson, M. J., Aleixo, A., Araujo-Silva, L. E., Ribas, C. C., … Miyaki, C. Y. (2018). Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian floodplain endemic species. *Systematic Biology*, *67*(4), 700–718. https://doi.org/10.1093/sysbio/syy004

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*, *29*(12), 673–680. https://doi.org/10.1016/j.tree.2014.10.004

Torroba-Balmori, P., Budde, K. B., Heer, K., González-Martínez, S. C., Olsson, S., Scotti-Saintagne, C., … Heuertz, M. (2017). Altitudinal gradients, biogeographic history and microhabitat adaptation affect fine-scale spatial genetic structure in African and Neotropical populations of an ancient tropical tree species. *PLOS ONE*, *12*(8), e0182515. https://doi.org/10.1371/journal.pone.0182515

Trabucco, A., & Zomer, R. J. (2009). Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. *CGIAR Consortium for Spatial Information. Published Online, Available from the CGIAR-CSI GeoPortal at: Http://Www.Csi.Cgiar.Org.*

Troupin, D., Nathan, R., & Vendramin, G. G. (2006). Analysis of spatial genetic structure in an expanding Pinus halepensis population reveals development of fine-scale genetic clustering over time. *Molecular Ecology*, *15*(12), 3617–3630. https://doi.org/10.1111/j.1365-294X.2006.03047.x

Tsalefac, M., Laraque, A., Sonwa, D., Scholte, P., Pokam, W., Beyene, T., … Ndjatsana, M. (2015). Climate of Central Africa : past, present and future. In C. de Wasseige, M. Tadoum, R. Ebaa-Atyi, & C. Doumenge (Eds.), *The Forests of the Congo Basin : Forests and climate change* (pp. 37–52). Weyrich.

Tysklind, N., Blanc-Jolivet, C., Mader, M., Meyer-Sand, B. R. V., Paredes-Villanueva, K., Honorio Coronado, E. N., … Degen, B. (2019). Development of nuclear and plastid SNP and INDEL markers for population genetic studies and timber traceability of *Carapa* species. *Conservation Genetics Resources*, *11*(3), 337–339. https://doi.org/10.1007/s12686-019-01090-2

Tysklind, N., Etienne, M. P., Scotti-Saintagne, C., Tinaut, A., Casalis, M., Troispoux, V., … Scotti, I. (2020). Microgeographic local adaptation and ecotype distributions: The role of selective processes on early life-history traits in sympatric, ecologically divergent *Symphonia* populations. *Ecology and Evolution*, *10*(19), 10735–10753. https://doi.org/10.1002/ece3.6731

Ukizintambara, T. (2009). *Forest edge effects on the behavioral ecology of L´Hoest´S monkey (Cercopithecus lhoesti) in Biwindi impenetrable national park, Uganda.* (Doctoral Dissertation). Antioch University, New England (US).

Väli, Ü., Saag, P., Dombrovski, V., Meyburg, B. U., Maciorowski, G., Mizera, T., …

Fagerberg, S. (2010). Microsatellites and single nucleotide polymorphisms in avian hybrid identification: A comparative case study. *Journal of Avian Biology*, *41*(1), 34–49. https://doi.org/10.1111/j.1600-048X.2009.04730.x

Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M. B., Balaguer, L., Benito-Garzón, M., … Zavala, M. A. (2014). The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, *17*(11), 1351–1364. https://doi.org/10.1111/ele.12348

Van der Niet, T., Cozien, R. J., & Johnson, S. D. (2015). Experimental evidence for specialized bird pollination in the endangered South African orchid *Satyrium rhodanthum* and analysis of associated floral traits. *Botanical Journal of the Linnean Society*, *177*(1), 141–150. https://doi.org/10.1111/boj.12229

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, *4*(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935. https://doi.org/10.1046/j.1365-294X.2004.02076.x

Verdu, C. F., Guichoux, E., Quevauvillers, S., De Thier, O., Laizet, Y., Delcamp, A., … Mariette, S. (2016). Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecology and Evolution*, *6*(20), 7323–7333. https://doi.org/10.1002/ece3.2466

Vinson, C. C., Amaral, A. C., Sampaio, I., & Ciampi, A. Y. (2005). Characterization and isolation of DNA microsatellite primers for the tropical tree, *Symphonia globulifera* Linn. f. *Molecular Ecology Notes*, *5*(2), 202–204. https://doi.org/10.1111/j.1471-8286.2005.00876.x

Viscosi, V., Lepais, O., Gerber, S., & Fortini, P. (2009). Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: A comparison of traditional and geometric morphometric methods. *Plant Biosystems*, *143*(3), 564–574. https://doi.org/10.1080/11263500902723129

von Uexküll, H. R., & Mutert, E. (1995). Global extent, development and economic impact of acid soils. *Plant and Soil*, *171*(1), 1–15. https://doi.org/10.1007/BF00009558

Wang, I. J. (2010). Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, *19*(13), 2605–2608. https://doi.org/10.1111/j.1365-294X.2010.04715.x

Wang, I. J. (2011). Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses. *Molecular Ecology*, *20*(12), 2480–2482. https://doi.org/10.1111/j.1365-294X.2011.05123.x

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Webb, C., & Bawa, K. (1983). Pollen Dispersal by Hummingbirds and Butterflies : A Comparative Study of Two Lowland Tropical Plants. *Evolution*, *37*(6), 1258–1270.

Weir, J. T., & Schluter, D. (2007). The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, *315*(5818), 1574–1576. https://doi.org/10.1126/science.1135590

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *American Journal of Botany*, *102*(11), 1753–1756. https://doi.org/10.3732/ajb.1500320

White, F. (1979). The Guineo-Congolian Region and Its Relationships to Other Phytochoria, *49*(1), 11–55.

Whitney, K. D., Fogiel, M. K., Lamperti, A. M., Holbrook, K. M., Stauffer, D. J., Hardesty, B. D., … Thomas, B. (1998). Seed dispersal by *Ceratogymna hornbills* in the Dja Reserve, Cameroon. *Journal of Tropical Ecology*, *14*, 351–371.

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138.

Wright, S. J. (2002). Plant diversity in tropical forests: A review of mechanisms of species coexistence. *Oecologia*, *130*(1), 1–14. https://doi.org/10.1007/s004420100809

Younger, J. L., Strozier, L., Maddox, J. D., Nyári, Á. S., Bonfitto, M. T., Raherilalao, M. J., … Reddy, S. (2018). Hidden diversity of forest birds in Madagascar revealed using integrative taxonomy. *Molecular Phylogenetics and Evolution*, *124*, 16–26. https://doi.org/10.1016/j.ympev.2018.02.017

Yuan, N., Sun, Y., Comes, H. P., Fu, C. X., & Qiu, Y. X. (2014). Understanding population structure and historica demography in a conservation context: Population genetics of the endangered *Kirengeshoma palmata* (Hydrangeaceae). *American Journal of Botany*, *101*(3), 521–529. https://doi.org/10.3732/ajb.1400043

Zhao, J. L., Gugger, P. F., Xia, Y. M., & Li, Q. J. (2016). Ecological divergence of two closely related *Roscoea* species associated with late Quaternary climate change. *Journal of Biogeography*, *43*(10), 1990–2001. https://doi.org/10.1111/jbi.12809

Zhou, W. W., Wen, Y., Fu, J., Xu, Y. B., Jin, J. Q., Ding, L., … Zhang, Y. P. (2012). Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai-Tibetan Plateau. *Molecular Ecology*, *21*(4), 960–973. https://doi.org/10.1111/j.1365-294X.2011.05411.x

Zimmerman, S. J., Aldridge, C. L., & Oyler-Mccance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, *21*(1), 382. https://doi.org/10.1186/s12864-020-06783-9

9. Supplementary information

9.1. Fine-scale spatial genetic structure in *S. globulifera*

S9.1.1. Geographic coordinates and microsatellite genotypes of Symphonia globulifera samples used in this study.

This section contains individual IDs, population IDs, geographic coordinates (UTM), altitude (m above sea level), altitude class used in analyses, and SSR allele sizes coded in two columns (one for each observed allele).

Each tab corresponds to a different population.

The data from Paracou was obtained from Degen, Bandou, & Caron, 2004, Heredity 93: 585–591. They contained three loci (Sg03, Sg18 and SgC4) genotyped on an ABI 310 genetic analyzer (Applied Biosystems, Carlsbad, USA).

The data of Yasuni and BCI contained five SSRs (Sg03, Sg18, SgC4, Sg19 and Sg06) and were genotyped on an ABI 3700 Genetic Analyzer (Applied Biosystems). They are cross-standardized for allele identities.

The data of Ituberá, Mbikiliki, Nkong Mekak and São Tomé contained five loci (Sg03 – PET, Sg18 – FAM, SgC4 – FAM, Sg19 – VIC, Sg10 – PET [locus – fluorescent label]) and were separated on an ABI 3730 Genetic Analyzer (Applied Biosystems). They are cross-standardized for allele identities.

In São Tomé, individual FO0001 did not have original coordinates available. We placed it next to FO004 for SGS analysis, as both were probably sampled close-by (nearly sequential naming).

Original references of loci used in this study:

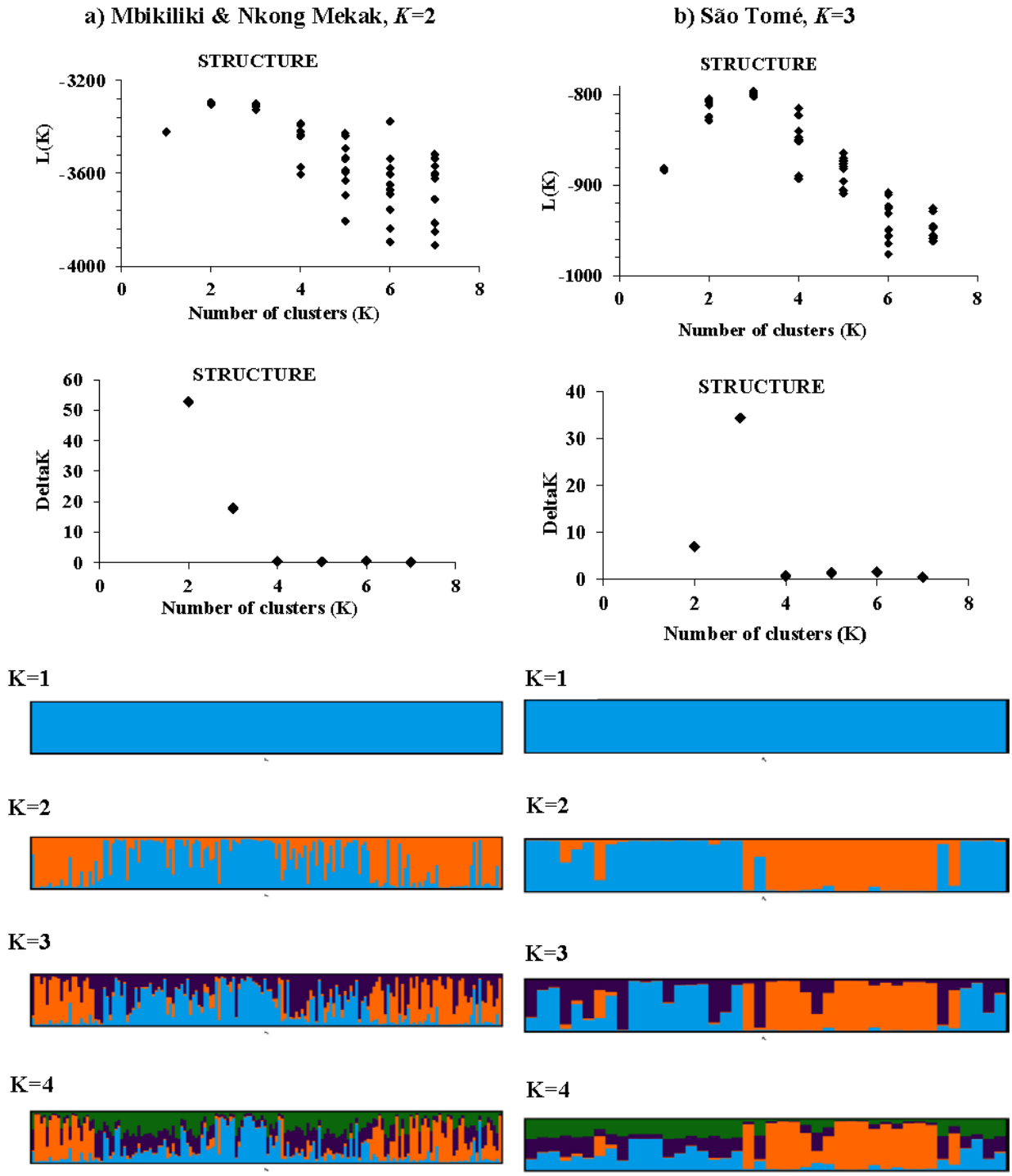
Locus	Reference	Comment
Sg03	Degen, Bandou, & Caron, (2004)	
Sg06	Vinson, Amaral, Sampaio & Ciampi (2005)	
Sg10	Vinson, Amaral, Sampaio & Ciampi (2005)	
Sg18	Degen, Bandou, & Caron, (2004)	
Sg19	Aldrich, Hamrick, Chavarriaga, & Kochert, (1998)	
SgC4	Aldrich, Hamrick, Chavarriaga, & Kochert, (1998)	
Sg92	Dick & Heuertz (2008)	same locus as Sg18

The file can be download from: https://doi.org/10.1371/journal.pone.0182515.s001

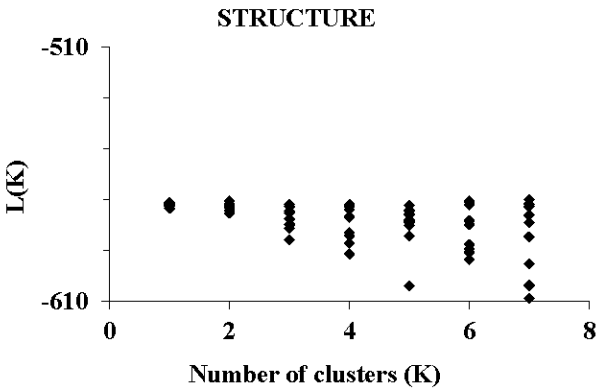
S9.1.2. Genetic clustering based on STRUCTURE and TESS

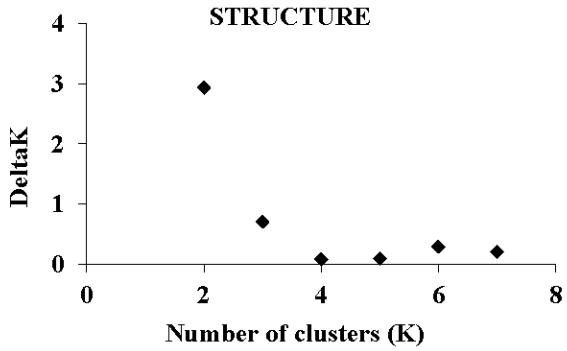
Codominant marker model in STRUCTURE

Figure S9.1.2.1. Illustration of the best number of genetic clusters *K* in STRUCTURE analyses (codominant marker model) for the African and American populations; results for *K*=1 to *K*=4. The best *K* was supported based on model log likelihood (L(*K*)) and Delta *K* (Δ*K*). The consensus barplot for each *K* from 10 independent runs confirmed the selection visually.

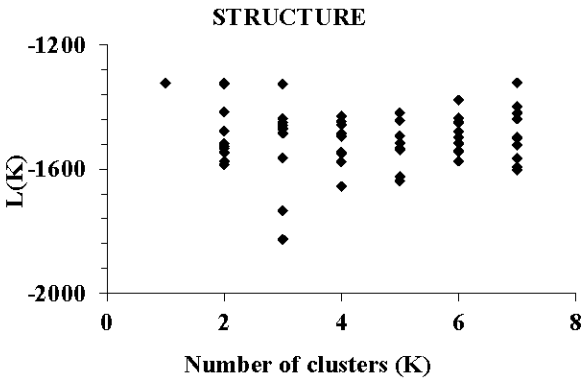


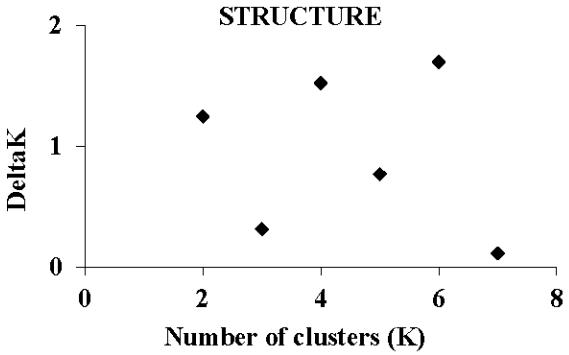
c) Yasuní, *K*=1

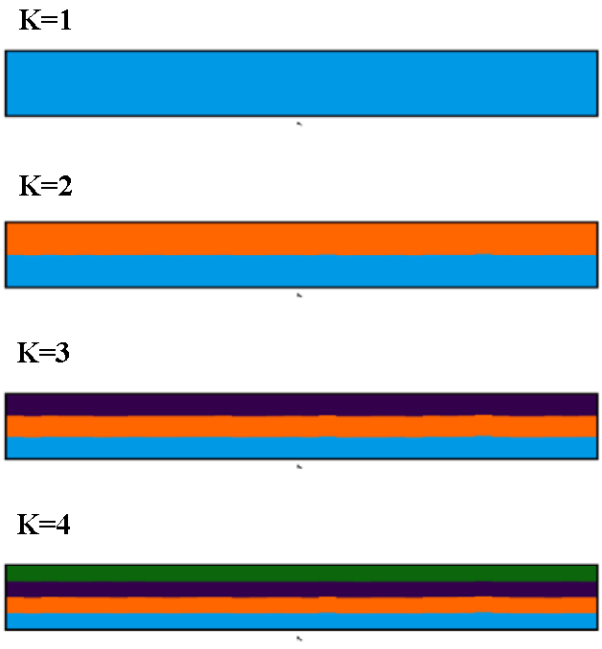


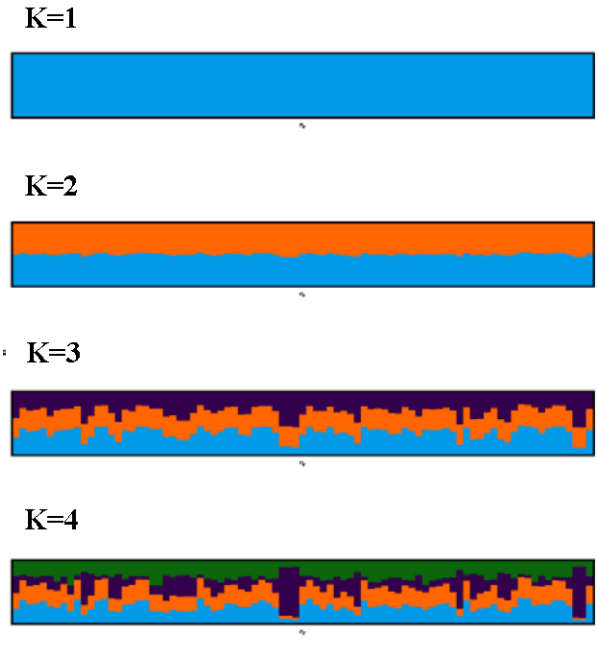


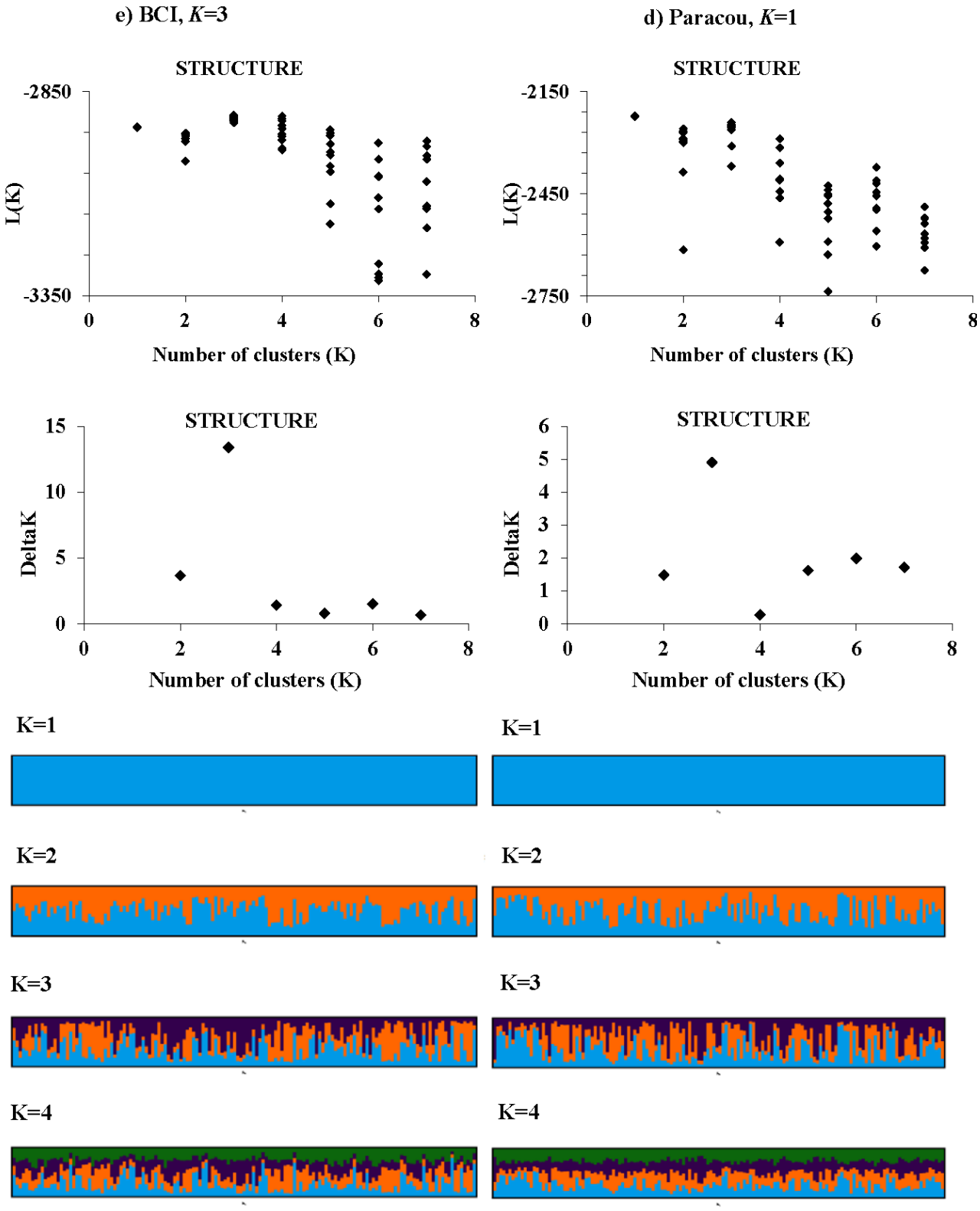
d) Ituberá, *K=1*










e) BCI, K=3
d) Paracou, K=1
STRUCTURE
L(K)
-2850
-3350
-2150
-2450
-2750
Number of clusters (K)
DeltaK
15
10
5
0
6
4
3
2
1
0
2
4
6
8
K=1
K=2
K=3
K=4

Comparison of the codominant and recessive marker models in STRUCTURE

Table S9.1.2.1. Comparison of the codominant and recessive marker models in STRUCTURE analyses. *K*, number of clusters under the codominant marker model, *K(null)*, number of clusters under the recessive marker model which accounts for null alleles. $r\ Q_{GP1}$ - $r\ Q_{GP3}$, Pearson's *r* correlation coefficients between ancestry proportions for each gene pool between the codominant and recessive allele models.

Population	***K***	***K(null)***	$r\ Q_{GP1}$	$r\ Q_{GP2}$	$r\ Q_{GP3}$
Neotropics					
BCI	3	3	0.988	0.968	0.989
Yasuní	1	1	-	-	-
Paracou	1	1	-	-	-
Ituberá	1	1	-	-	-
Africa					
São Tomé	3	3	0.938	0.992	0.973
Nkong Mekak	2	3	0.993	0.993	-
Mbikiliki	2	3	0.988	0.988	-

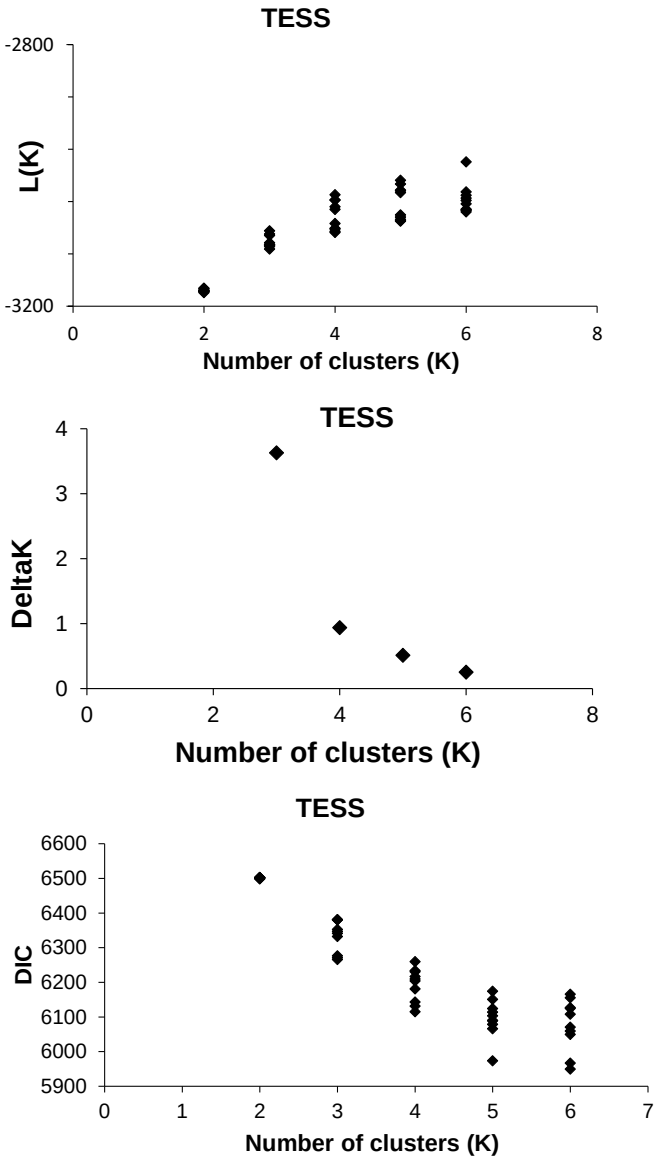
TESS analysis and comparison with the codominant marker model in STRUCTURE

A Bayesian clustering analysis of individual multilocus genotypes was performed with the program TESS v. 2.3.1 (Chen, Durand, Forbes, & François, 2007), which can include the individuals' spatial position as prior information. We used an admixture model based on trend surfaces and on a conditional autoregressive model (CAR) where each individual's multilocus genotype is composed of fractions that originate in up to *K* different potentially unobserved source populations or clusters. The degree of trend surface in the model determines whether spatial information is included (0, no spatial information; ≤1, spatial information) and the spatial interaction parameter Ψ determines the intensity of spatial autocorrelation. We tested three different models: model 1 with trend degree = 0 and $\Psi = 0$; model 2 with trend degree = 1 and $\Psi = 0.6$; and model 3 with trend degree = 1 and $\Psi = 1$, and used default options for the other parameters (see Eric Durand, Jay, Gaggiotti, & François, 2009). Model 1 without spatial prior is equivalent to STRUCTURE, as used in our work (see Section 3.2.1.). Each model was run for 100,000 MCMC steps, including a burn-in of 20,000 steps, and we repeated the analysis ten times from *K*=2 to *K*=7. The number of clusters *K* that best described the data was based on the model's log-likelihood, the deviance information criterion (DIC), and ΔK (Durand, Chen, & François, 2009; Evanno, Regnaut, & Goudet, 2005; Pritchard, Wen, & Falush 2010). The choice of the *K* that best explained the data was not trivial in TESS, and we systematically obtained a larger *K* than in STRUCTURE (Figures S9.1.2.1. and S9.1.2.2.). In Table S9.1.2.2., we illustrate the correlation of cluster membership in STRUCTURE and TESS, choosing *K* based on the STRUCTURE analysis.

Table S9.1.2.2. Pearson correlations between ancestry coefficients inferred by STRUCTURE for the best run and *K* (K > 1) and ancestry coefficients of the three models tested in TESS for the best run (choice of *K* based on STRUCTURE analysis). M1: model 1 in TESS, M2: model 2 in TESS; M3: model 3 in TESS. GP1, GP2, GP3: the different GPs identified by the Bayesian clustering. Significance values refer to significance of Pearson correlation tests after Bonferroni correction: ns, not significant; ***, P≤0.001; **, P≤0.01; *, P≤0.05.

	GP1			**GP2**			**GP3**		
Population	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**
Mbikiliki and Nkong Mekak	0.97^{***}	0.96^{***}	0.95^{***}	0.97^{***}	0.96^{***}	0.95^{***}	-	-	-
São Tomé	0.68^{***}	0.63^{***}	0.35^{ns}	0.99^{***}	0.92^{***}	0.88^{***}	0.85^{***}	0.81^{***}	0.64^{***}
BCI	-0.25^{ns}	-0.19^{ns}	-0.21^{ns}	-0.53^{***}	0.03^{ns}	-0.01^{ns}	-0.21^{ns}	0.74^{***}	0.71^{***}

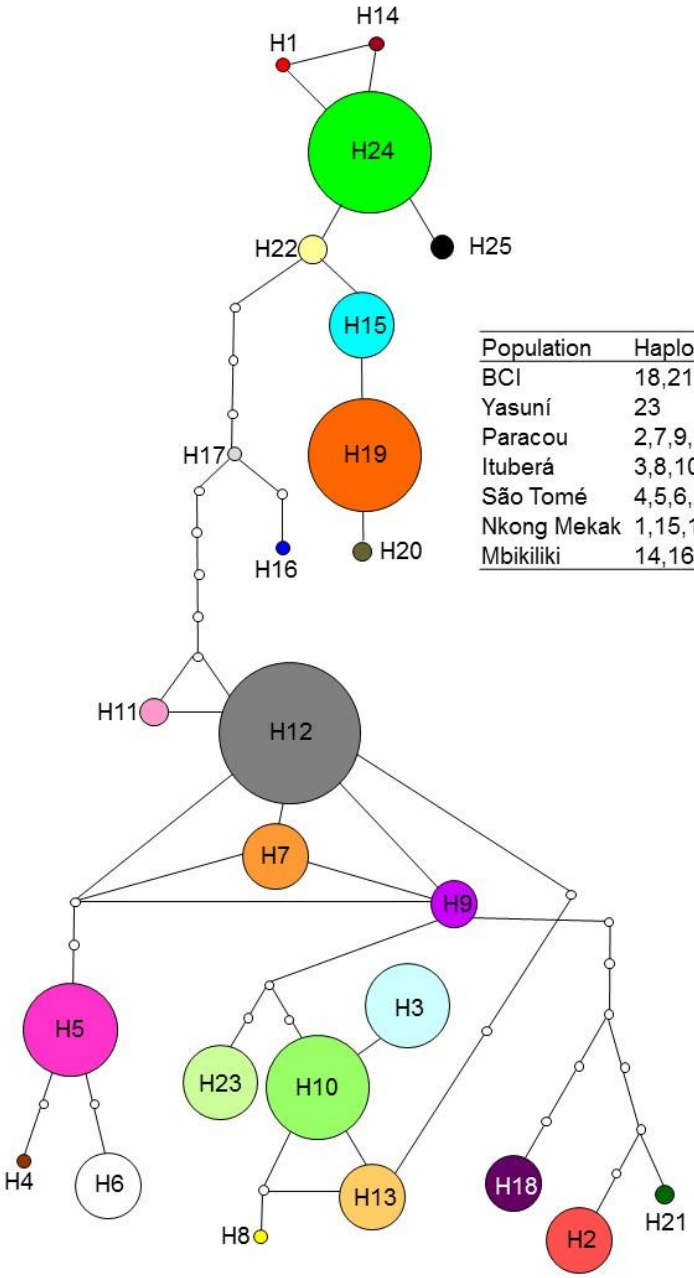
Figure S9.1.2.2. Illustration of different criteria for the choice of the best number of genetic clusters *K* in TESS analyses for the Cameroonian Mbikiliki and Nkong Mekak populations. In STRUCTURE (see Fig. S9.1.2.1.), *K*=2 was supported based on model log likelihood (L(K)) and Delta K (Δ*K*). In TESS the deviance information criterion (DIC) and the model log likelihood indicated increasing support for increasing *K* values, and the Δ*K* criterion was undefined for *K*=2, because TESS does not run with *K*=1.



S9.1.3. Evolutionary relationships among plastid DNA haplotypes

A haplotype network for *psbA- trnH* sequences was created in TCS (Clement, Posada, & Crandall, 2000) using statistical parsimony based on a haplotype distance matrix where distance between two haplotypes was defined as the number of genetic differences (nucleotide differences or insertion/deletion polymorphisms).

Figure S9.1.3.1. Haplotype network in *Symphonia globulifera* and geographic distribution of haplotypes. Chart size increases with sample size (from 1 to 67 individuals). Haplotype numbers and colors correspond to those of Fig. 6 and 7 in the Study I. Each line corresponds to one mutation and small white circles indicate non-observed haplotypes.



Population	Haplotypes
BCI	18,21
Yasuní	23
Paracou	2,7,9,11,12
Ituberá	3,8,10,13
São Tomé	4,5,6,24
Nkong Mekak	1,15,19,20,22,24
Mbikiliki	14,16,17,19,22,24,25

Table S9.1.3.1. Plastid DNA haplotype definition in *Symphonia globulifera* based on sequences of the *psbA-trnH* intergenic spacer region. *n*, sample size; column headings indicate positions of single nucleotide polymorphisms, insertion-deletion polymorphisms, microsatellites and inversions.

	Polymorphic positions in *psbA-trnH*																														
Hap	**n**	**70**	**170**	**226**	**248-253**	**265**	**313-314**	**320**	**328-330**	**352**	**355-361**	**411**	**413-425**	**431-444**	**445**	**446-451**	**453**	**483**	**505**	**509**	**511-519**	**520**	**534**	**538-561**	**593**	**614**	**616**	**620-649**	**653**
H1	1	?	T	T	TAAGAA	A	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTTT---	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	-	G	A----------	C
H2	9	T	T	-	-	A	AT	G	AAA	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-	T	G	C	T	-	A	C	ATAA-------AAAATAAA-----G	A	T	G	A----------	G
H3	11	T	G	T	TAAGAA	A	AT	G	AAA	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-	T	T	A	T	TTAAATTTG	A	C	ATAA-------AGAATAAA-----G	A	T	A	A----------	G
H4	1	T	T	T	TAAGAA	A	AT	G	TTT	T	-	-	-	TTTTTTTTTTT--	T	-----A	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H5	27	T	T	T	TAAGAA	A	AT	G	TTT	T	-	-	-	TTTTTTTTTTTT-	T	-	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H6	9	T	T	T	TAAGAA	A	AT	G	TTT	T	-	-	-	TTTTTTTTTTTT	T	-	G	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H7	9	T	T	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTT-------	T	-	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H8	1	T	G	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTT------	T	-	T	T	A	T	TTAAATTTG	A	C	ATAA-------AGAATAAA-----G	C	T	A	A----------	G
H9	6	T	T	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H10	29	T	G	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-	T	T	A	T	TTAAATTTG	A	C	ATAA-------AGAATAAA-----G	A	T	A	A----------	G
H11	5	T	T	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTTT----	T	-	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H12	67	T	T	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTTT----	T	-	T	T	A	T	-	A	C	ATAA-------AGAATAAA-----G	A	T	G	A----------	G
H13	9	T	G	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTTT----	T	-	T	T	A	T	TTAAATTTG	A	C	ATAA-------AGAATAAA-----G	A	T	A	A----------	G
H14	1	A	T	T	TAAGAA	A	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTTT---	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H15	9	T	T	T	TAAGAA	C	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTT----	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H16	1	?	?	?	?	A	TA	T	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	G	TTTTTA	T	A	A	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	G
H17	1	T	T	T	TAAGAA	A	TA	T	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTTT----	G	TTTTTA	T	T	A	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	G
H18	8	T	T	-	TAAGAA	A	AT	G	TTT	G	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-	T	T	C	T	-	T	C	A------- A-------AAAAAAATAAA-----G	A	T	G	A------- TTATTCTTTTATTTTAGTGAA	G
H19	32	T	T	T	TAAGAA	C	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTT-----	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H20	2	T	T	T	TAAGAA	C	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTT-----	G	TTTTT-	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H21	2	T	T	-	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTTTT---	T	-	T	G	C	T	-	A	C	ATAA-------AAAATAAA-----G	A	T	G	A----------	G
H22	5	T	T	T	TAAGAA	A	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTT----	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H23	10	T	T	T	TAAGAA	A	AT	G	TTT	T	-	A	CTCATTTTTCTTT	TTTTTTTTT-----	T	-----A	T	T	A	A	-	A	C	ATAA-------AGAATAAA-----G	A	T	A	A----------	G
H24	44	T	T	T	TAAGAA	A	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTTT---	G	TTTTTA	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C
H25	4	T	T	T	TAAGAA	A	TA	T	TTT	T	TTTT TAC	-	CTCATTTTTCTTT	TTTTTTTTTTT---	G	TTTTT-	T	T	T	T	-	A	A	ATAA-------A----------G	A	T	G	A----------	C

Table S9.1.3.2. *Psba-trnH* sequence data set in *Symphonia globulifera* with Genbank accession numbers. In Paracou, the *S. globulifera* morphotype is given in the Population field: *S.glo* for the common morphotype, and *S.sp1* for the alternative morphotype.

The file can be download from: https://doi.org/10.1371/journal.pone.0182515.s003

S9.1.4. Genetic diversity and fine-scale spatial genetic structure statistics in Symphonia globulifera based on different groups of SSRs.

Table S9.1.4.1. Genetic diversity and fine-scale spatial genetic structure statistics in *Symphonia globulifera* in a subset of loci corresponding to the three nuclear SSRs used in Paracou (Degen, Bandou, & Caron, 2004). n, sample size for SSR data; *A*, mean number of alleles per locus; A_R, allelic richness or number of alleles expected in a sample of 34 individuals and standard deviation; H_E, expected heterozygosity; F_{IS}, fixation index; DC, number of distance classes; 1st DC, maximum distance of the first class (m); F_{ij}-intra, intra individual kinship coefficient; $F_{ij(1)}$, average kinship coefficient of the first distance class; *Sp*, intensity of FSGS and *P*-value of one-sided test of the regression slope *b*; *b* mean jacknife ± SE, jackknife mean of *b* and standard error. ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$.

Population	n	SSR	*A*	A_R(SD)	H_E	F_{IS}	DC	1st DC (m)	F_{ij}-*intra*	$F_{ij(1)}$	*Sp*	*b* mean jackknife±SE
BCI	147	3	5.84	8.09 (2.60)	0.808	0.157^{***}	7	113	0.1538	0.0373	0.0223^{***}	-0.0215±0.0104
Yasuní	34	3	4.81	8.10 (2.69)	0.727	0.054^{ns}	10	94	0.1042	-0.0634	0.0035^{ns}	-0.0038±0.0094
Paracou	148	3	9.99	12.57 (3.56)	0.880	0.172^{***}	4	203	0.1709	0.0154	0.0090^{***}	-0.0088±0.0029
Ituberá	85	3	3.53	5.26 (3.54)	0.588	0.159^{***}	5	151	0.1564	0.0082	0.0072^{*}	-0.0072±0.0045
São Tomé	42	3	4.69	8.98 (2.96)	0.769	0.178^{***}	6	856	0.1783	0.1135	0.0468^{***}	-0.0415±0.0166
Nkong Mekak	70	3	6.20	7.22 (5.94)	0.729	0.205^{***}	5	312	0.2051	0.0272	0.0169^{***}	-0.0165±0.0052
Mbikiliki	94	3	4.71	7.13 (3.96)	0.686	0.265^{***}	7	240	0.2650	0.0989	0.0372^{***}	-0.0335±0.0196

Table S9.1.4.2. Genetic diversity and fine-scale spatial genetic structure statistics in *Symphonia globulifera* based a) on 18 genic nuclear SSRs, data from. Olsson et al. (2017), b). on 3-5 genic nuclear SSRs and subsets of ca. 30 individuals. n, sample size for SSR data; *A*, mean number of alleles per locus; A_R, allelic richness or number of alleles expected in a sample of 30 individuals and standard deviation; H_E, expected heterozygosity; F_{IS}, fixation index; DC, number of distance classes; 1st DC, maximum distance of the first class (m); F_{ij}-intra, intra individual kinship coefficient; $F_{ij(1)}$, average kinship coefficient of the first distance class; *Sp*, intensity of FSGS and *P*-value of one-sided test of the regression slope *b*; *b* mean jacknife ± SE, jackknife mean of *b* and standard error. ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$; ., $P \leq 0.1$.

a Population	**n**	**SSR**	***A***	***A_R*(SD)**	***H_E***	***F_{IS}***	**DC**	**1st DC (m)**	***F_{ij}-intra***	***$F_{ij(1)}$***	***Sp***	***b* mean jackknife±SE**
Paracou	32	18	4.89	2.77(2.06)	0.491	0.102^{*}	4	378	0.1015	-0.0072	0.0018.	-0.0018±0.0012
Ituberá	31	18	2.94	2.59(0.96)	0.370	-0.291^{***}	4	178	-0.2828	-0.0025	-0.0021^{ns}	0.0021±0.0042
São Tomé	30	18	3.24	2.99(2.35)	0.341	0.123^{**}	4	1371	0.0841	0.0470	0.0339^{***}	-0.0323±0.0081
Nkong Mekak	31	18	3.83	3.29(1.47)	0.457	-0.167^{***}	7	204	-0.1721	0.0106	0.0072.	-0.0071±0.0057

b Population	**n**	**SSR**	***A***	***A_R*(SD)**	***H_E***	***F_{IS}***	**DC**	**1st DC (m)**	***F_{ij}-intra***	***$F_{ij(1)}$***	***Sp***	***b* mean jackknife±SE**
Paracou	32 (subset different from Olsson et al. 2017)	3	14.67	11.28(4.42)	0.865	0.169^{***}	4	203	0.1676	-0.0022	0.0067.	-0.0064±0.0065
Ituberá	31 (same as Olsson et al. 2017)	5	8.8	6.98(5.23)	0.622	0.108^{**}	4	178	0.1051	0.0070	0.0055^{ns}	-0.0049±0.0031
São Tomé	30 (same as Olsson et al. 2017)	5	11.6	9.42(3.06)	0.824	0.211^{***}	4	1371	0.2137	0.0707	0.0475^{***}	-0.0438±0.0139
Nkong Mekak	31 (same as Olsson et al. 2017)	5	11.4	9.24(5.46)	0.807	0.163^{***}	7	204	0.1631	0.0136	0.0088	-0.0088±0.0042

S9.1.5. Estimates of mating system and FSGS parameters in genetic clusters of Symphonia globulifera

Table S9.1.5.1. Estimates of mating system and FSGS parameters in genetic clusters of *Symphonia globulifera*. GP, gene pool (GPs include individuals with ancestry proportions *Q* of 0.875-1); n, sample size; F_{IS}, fixation index; F_{IS}*, fixation index after null allele correction; DC, number of distance classes; 1st DC, maximum distance of the first class (m); *Sp*, intensity of SGS and *P*-value of one-sided test of the regression slope *b*. ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$; nc, not calculated (no null alleles or small sample size).

Population	GP	n	F_{IS}	F_{IS}*	DC	1st DC (m)	*Sp*
São Tomé	GP 1	4	nc	nc	-	-	-
	GP 2	12	0.054^{ns}	nc	3	1434	-0.0008^{ns}
	GP 3	8	-0.025^{ns}	nc	3	724	-0.0054^{ns}
Nkong Mekak	GP 1	35	0.055^{ns}	nc	7	353	0.0181^{***}
	GP 2	14	0.119^{*}	nc	3	219	0.0085^{ns}
Mbikiliki	GP 1	18	-0.011^{ns}	nc	3	125	0.0136^{**}
	GP 2	45	0.125^{***}	0.035^{ns}	4	265	0.0136^{***}

S9.1.6. Publication of results of Study I: first page.

PLOS ONE

RESEARCH ARTICLE

Altitudinal gradients, biogeographic history and microhabitat adaptation affect fine-scale spatial genetic structure in African and Neotropical populations of an ancient tropical tree species

Paloma Torroba-Balmori[1,2], Katharina B. Budde[3], Katrin Heer[4,5], Santiago C. González-Martínez[1,2,3], Sanna Olsson[1], Caroline Scotti-Saintagne[6], Maxime Casalis[7†], Bonaventure Sonké[8,9], Christopher W. Dick[10,11], Myriam Heuertz[1,3,9*]

1 Department of Forest Ecology and Genetics, INIA Forest Research Centre, Madrid, Spain, **2** Sustainable Forest Management Research Institute, University of Valladolid - INIA, Palencia, Spain, **3** UMR BIOGECO, INRA, University of Bordeaux, Cestas, France, **4** Institute of Experimental Ecology, University of Ulm, Ulm, Germany, **5** Conservation Biology and Ecology, University of Marburg, Marburg, Germany, **6** UR Écologie des Forêts Méditerranéennes, INRA, Avignon, France, **7** UMR EcoFoG, INRA, Kourou, French Guiana, **8** Ecole Normale Supérieure, Université de Yaoundé I, Yaoundé, Cameroon, **9** Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium, **10** Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America, **11** Smithsonian Tropical Research Institute, Republic of Panama

† Deceased.
* myriam.heuertz@inra.fr

Check for updates

OPEN ACCESS

Citation: Torroba-Balmori P, Budde KB, Heer K, González-Martínez SC, Olsson S, Scotti-Saintagne C, et al. (2017) Altitudinal gradients, biogeographic history and microhabitat adaptation affect fine-scale spatial genetic structure in African and Neotropical populations of an ancient tropical tree species. PLoS ONE 12(8): e0182515. https://doi.org/10.1371/journal.pone.0182515

Editor: Zhengfeng Wang, Chinese Academy of Sciences, CHINA

Received: May 25, 2017

Accepted: July 8, 2017

Published: August 3, 2017

Copyright: © 2017 Torroba-Balmori et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Geographic coordinates, sampling elevation, morphotype information (where relevant) and microsatellite genotypes of Symphonia globulifera samples are given in Supporting Information S1. Plastid DNA sequences are archived in Genbank. The full list of Genbank accession numbers is given in Supporting Information S3, including accession numbers KX572421 - KX572686 which correspond to sequences newly generated for this study.

Abstract

The analysis of fine-scale spatial genetic structure (FSGS) within populations can provide insights into eco-evolutionary processes. Restricted dispersal and locally occurring genetic drift are the primary causes for FSGS at equilibrium, as described in the isolation by distance (IBD) model. Beyond IBD expectations, spatial, environmental or historical factors can affect FSGS. We examined FSGS in seven African and Neotropical populations of the late-successional rain forest tree *Symphonia globulifera* L. f. (Clusiaceae) to discriminate the influence of drift-dispersal vs. landscape/ecological features and historical processes on FSGS. We used spatial principal component analysis and Bayesian clustering to assess spatial genetic heterogeneity at SSRs and examined its association with plastid DNA and habitat features. African populations (from Cameroon and São Tomé) displayed a stronger FSGS than Neotropical populations at both marker types (mean Sp = 0.025 *vs.* Sp = 0.008 at SSRs) and had a stronger spatial genetic heterogeneity. All three African populations occurred in pronounced altitudinal gradients, possibly restricting animal-mediated seed dispersal. Cyto-nuclear disequilibria in Cameroonian populations also suggested a legacy of biogeographic history to explain these genetic patterns. Conversely, Neotropical populations exhibited a weaker FSGS, which may reflect more efficient wide-ranging seed dispersal by Neotropical bats and other dispersers. The population from French Guiana displayed an association of plastid haplotypes with two

S9.1.7. Altitudinal clustering of gene pools in Symphonia globulifera populations

Table S9.1.7.1. Altitudinal clustering of gene pools in *Symphonia globulifera* populations. n, sample size range per altitudinal class; GP, gene pool; *P*-value, *P*-value of an one-way ANOVA contrasting individual ancestry values (*q*) in three altitudinal classes for each GP within populations; *q* mean, mean *q* of individuals in altitude class H1, H2, or H3 (from lowest to highest). ns, not significant; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$.

Populations	**n**	**GP**	***P*-value**	***q* mean (H1)**	***q* mean (H2)**	***q* mean (H3)**
BCI	46-51	GP 1	**	0.310	0.245	0.422
		GP 2	*	0.389	0.341	0.282
		GP 3	*	0.301	0.414	0.296
São Tomé	12-16	GP 1	ns	0.311	0.365	0.140
		GP 2	***	0.073	0.347	0.831
		GP 3	***	0.616	0.288	0.029
Nkong Mekak	23-24	GP 1	*	0.693	0.461	0.739
		GP 2	*	0.307	0.539	0.261
Mbikiliki	31-32	GP 1	***	0.177	0.589	0.334
		GP 2	***	0.823	0.411	0.666

9.2. Spatial genetic structure of *S. globulifera* across continents

S9.2.1. SNP genotypes of Symphonia globulifera samples generated by Genotyping-by-sequencing

This file contains:

SNP ID:

Correspondence between contigs, in which SNPs were present in at least 70% of the individuals, and names identifying the loci on the 4921 SNP dataset used in this study.

Sequences:

Consensus sequences of contigs for the 4921 SNPs selected

Raw probabilities:

It contains individual IDs, population IDs and genotype likelihood data for 4921 SNPs in a final set of 367 individuals inferred using a hierarchical Bayesian model in Entropy for *K*=9.

Genotype likelihood is presented as continuous values from 0 (homozygote for the first allele) to 2 (homozygote for the second allele), where 1 represented the heterozygous state and values closer to 0.5 or 1.5 indicated high uncertainty of genotypes.

Round probabilities.9:

Reassignement of continuous values of genotype likelihood to discrete genotypes if they lay within 0.1 of the closest discrete value (representing a minimum of 90% probability to be such genotype), otherwise SNPs were considered missing data (NA).

It contains individual IDs and population IDs for all populations of *S. globulifera.*

Round probabilities.75:

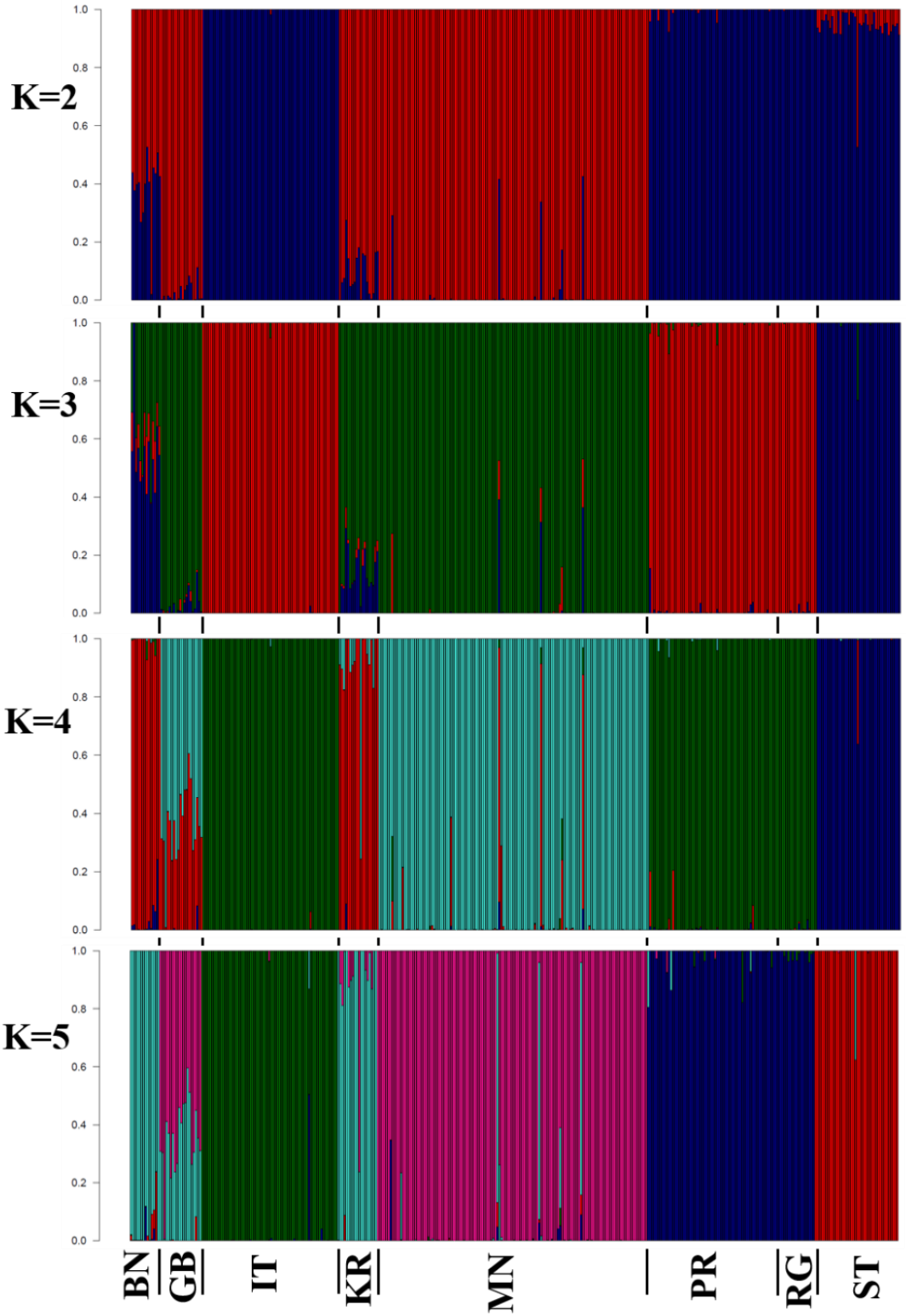
Reassignement of continuous values of genotype likelihood to discrete genotypes if they lay within 0.25 of the closest discrete value (representing a minimum of 75% probability to be such genotype), otherwise SNPs were considered missing data (NA).

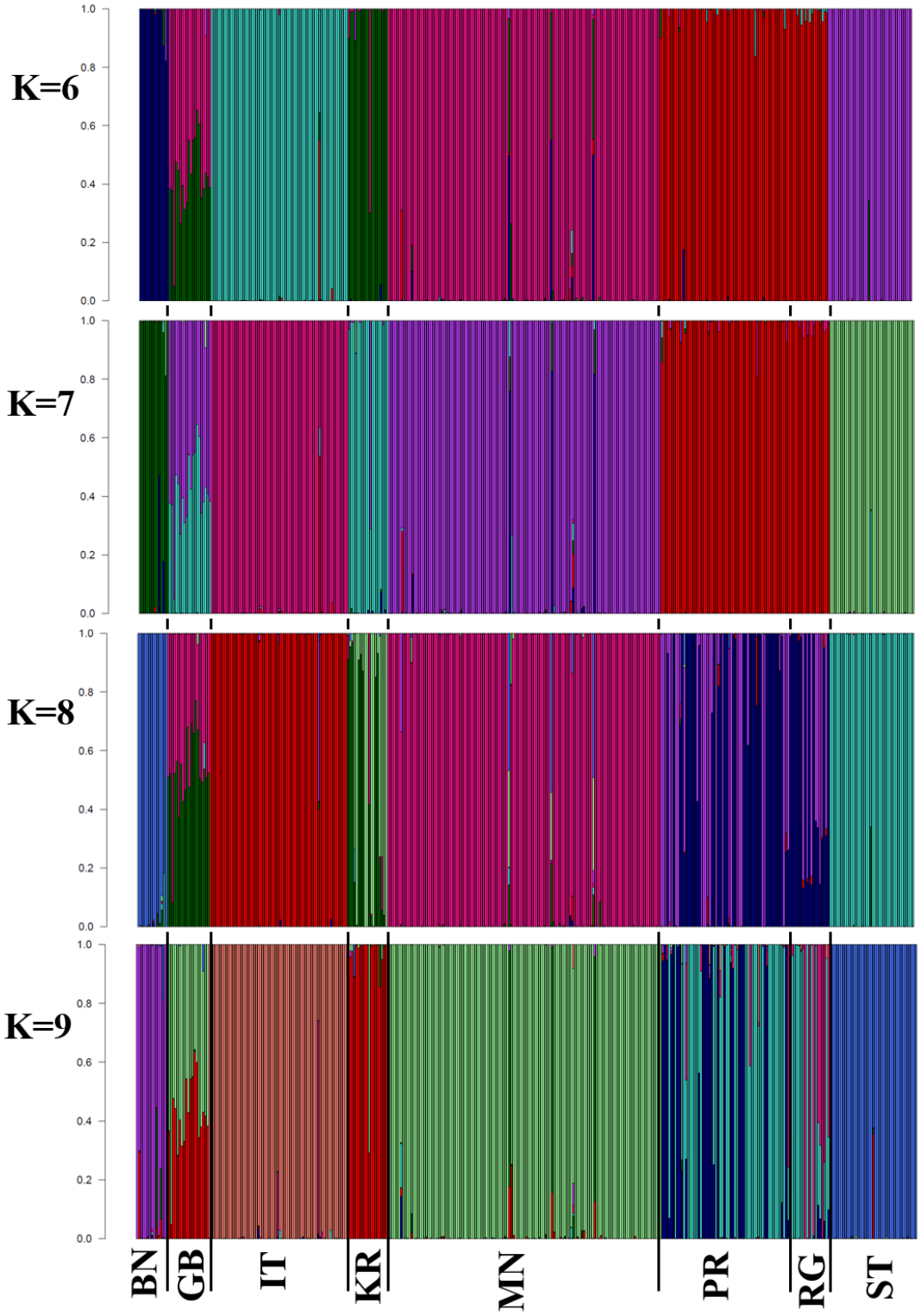
It contains individual IDs and population IDs for populations of *S. globulifera* in mainland Africa.

The dataset only contains the 3399 polimorphic SNPs for these selected populations (monomorphic SNPs were removed)

The file can be download from: https://doi.org/10.5281/zenodo.5772441

Figure S9.2.1.1. Entropy barplots illustrate the distribution of ancestry proportion of individuals in each of the *K* gene pools from our populations from *K*=2 to *K*=9 (light blue GP: alternative morphotype in PR and RG for *K*=9).




K=6
K=7
K=8
K=9
1.0
0.8
0.6
0.4
0.2
0.0
BN
GB
IT
KR
MN
PR
RG
ST

S9.2.2. Models and priors for phylogenetic analysis in SNAPP

As commented in Section 3.2.2.1, we selected several priors for the models in SNAPP based on literature. For all models, we assumed a generation time of 100 years (Budde, González-Martínez, Hardy, & Heuertz, 2013; Jones, Cerón-Souza, Hardesty, & Dick, 2013), mutation rate (μ) of 10^{-9} per site per generation (Ossowski et al., 2010) and a divergence time for *S. globulifera* populations between Africa and America of 17.36 Ma (Dick, Abdul-Salim, & Bermingham, 2003) as expectation for the root height of the tree to calculate the birth rate lambda (λ=10536) in the Yule prior (see further information about the Yule model in Drummond & Bouckaert 2015).

Uniform priors for lambda are used when no information exists on root height for the tree, whereas 1/X prior works better when a birth rate prior (λ) is available (Drummond & Bouckaert 2015). We chose 1/X for most of the models as we had a λ estimate, but we also assessed the hypothesis of no information for that parameter. In that latter case, we obtained the limits for λ uniform distribution based on pollen fossil records from *S. globulifera* in Africa (45 Ma) and South America (15 Ma; see Dick, Abdul-Salim, & Bermingham, 2003) as expected limits for the root height of the tree.

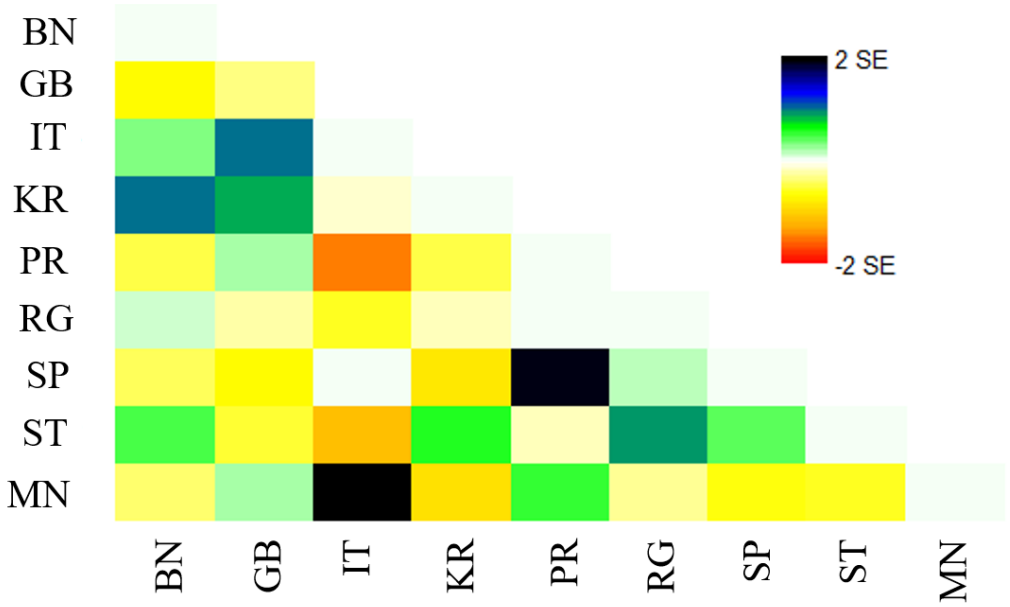
We obtained different values for effective size in current and ancient populations in *S. globulifera* in Africa and America (Jones, Cerón-Souza, Hardesty, & Dick, 2013; Budde, González-Martínez, Hardy, & Heuertz, 2013) based on the relation $\Theta = 4 \times Ne \times \mu$ for diploid populations (Drummond & Bouckaert 2015). We took the highest and lowest values from mean / median and extreme Ne estimates from the three sets of values (Table S9.2.2.1.). We used the intermediate value 20,000 from mean values (median) to calculate the mean coalescence rate for *S. globulifera* ($\Theta = 0.00008$) using $\mu = 10^{-9}$. We considered the lowest and highest values from extreme Ne values (89 - 1,047,128,548) and higher and lower mutation rate values than before ($10^{-10} – 10^{-8}$ per site per generation; Baer, Miyamoto, & Denver, 2007) to calculate the limits for uniform Θ distribution in M2 and M3. We also used two different distributions for Θ: a gamma distribution, a good non-informative prior for population size (Drummond & Bouckaert 2015), with $\alpha = 2$ reflecting the common procedure (prior mean $\Theta = \alpha/\beta$; e.g., Rannala & Yang 2003, Leaché, Fujita, Minin, & Bouckaert, 2014), or a less informative prior (uniform distribution). Values for backward (u) and forward (v) mutation rates were sampled from the data. Non-polymorphic sites option was selected due to the existence of missing data. We used default options for the other parameters.

Table S9.2.2.1. Values for effective population size (Ne) in *S. globulifera* from different studies. Ne in African populations have been deduced from the Ne x μ values and μ rates used in their simulations. m. Ne: mean or median values from simulations (min – max); extreme Ne: highest and lowest values from 95% confidence intervals obtained from simulations in the studies (min – max).

	Continent	**Population age**	**m. Ne**	**extreme Ne**
Budde, González-Martínez, Hardy, & Heuertz (2013)	Africa	recent and isolated populations	1,620 – 5,420,000	500 - 9,200,000
		the ancestral African population	21,800 - 27,200	3,250 - 48,600
Jones, Cerón-Souza, Hardesty, & Dick, (2013)	America	ancient and recent populations	661 - 281,838	89 - 1,047,128,548

S9.2.3. Inference of a maximum likelihood tree implemented in TreeMix from African and American populations.

Figure S9.2.3.1. Plot of scaled residuals from the maximum likelihood tree without migration.



9.3. Local adaptation of *S. globulifera* at continental scale in Africa

S9.3.1. Climatic and soil variables used for the analysis of loci under selection

Climatic variables

As explained in the metadata of the WorldClim 1.4 dataset (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), the bioclimatic variables derive from the monthly temperature and rainfall values. These variables represent annual trends, seasonality, and extreme or limiting environmental factors. A quarter of the year is a period of three months.

BIO1: Annual Mean Temperature (ºC)
BIO2: Mean Diurnal Range (Mean of monthly (maximun temp - minimun temp)) (ºC)
BIO3: Isothermality (BIO2/BIO7) (100) (ºC)*
*BIO4: Temperature Seasonality (standard deviation *100)*
BIO5: Max Temperature of Warmest Month (ºC)
BIO6: Min Temperature of Coldest Month (ºC)
BIO7: Temperature Annual Range (BIO5-BIO6) (ºC)
BIO8: Mean Temperature of Wettest Quarter (ºC)
BIO9: Mean Temperature of Driest Quarter (ºC)
BIO10: Mean Temperature of Warmest Quarter (ºC)
BIO11: Mean Temperature of Coldest Quarter (ºC)
BIO12: Annual Precipitation (mm)
BIO13: Precipitation of Wettest Month (mm)
BIO14: Precipitation of Driest Month (mm)
BIO15: Precipitation Seasonality (Coefficient of Variation)
BIO16: Precipitation of Wettest Quarter (mm)
BIO17: Precipitation of Driest Quarter (mm)
BIO18: Precipitation of Warmest Quarter (mm)
BIO19: Precipitation of Coldest Quarter (mm)

Aridity index

As cited in the metadata of the Aridity Index dataset (Trabucco & Zomer, 2009): "the Global-Aridity Index represents moisture availability for potential growth of reference vegetation excluding the impact of soil mediating water runoff events". This index is a function of Mean Annual Precipitation and Mean Annual Potential Evapo-Transpiration (PET). The values will increase for more humid conditions and will decrease for higher aridity.

The classes are: < 0.03 Hyper Arid; 0.03-0.2 Arid; 0.2-0.5 Semi-Arid; 0.5-0.65 Dry sub-humid; > 0.65 Humid.

Soil variables

All soil datasets (GeoNetwork opensource software, http://www.fao.org/geonetwork/srv/en/main.home, Land and Water Development Division, FAO, Rome) come from the "Derived Soil Properties" of the FAO-UNESCO Soil Map of the World. **For ease of understanding, the literal explanations ("quoted") about variables included in their metadata has been added to this annex**.

NN_S: nitrogen percentage in subsoil.

"The interpretation of this factor often involves the calculation of the C/N ratio, which is an indicator of the quality of the humus fraction".

The classes are: 1: N 0-0.02%; 2: N >0.02-0.08%; 3: N >0.08-0.2%; 4: N >0.2-0.5%; 5: N >0.5%; 97: Water; 99: Glaciers, Rock, Shifting sand, Missing data.

NN_T: nitrogen percentage in topsoil.

"The interpretation of this factor often involves the calculation of the C/N ratio, which is an indicator of the quality of the humus fraction".

The classes are: 1: N 0-0.02%; 2: N >0.02-0.08%; 3: N >0.08-0.2%; 4: N >0.2-0.5%; 5: N >0.5%; 97: Water; 99: Glaciers, Rock, Shifting sand, Missing data.

OC_S: organic carbon percentage in subsoil

OC_T: organic carbon percentage in topsoil

"Organic Carbon is, together with pH, the best indicator of the nutrient status of the soil. Moderate to high amounts of organic carbon are associated with fertile soils with a good structure".

The classes are: 1: OC <0.2%; 2: OC 0.2-0.6%; 3: OC >0.6-1.2%; 4: OC >1.2-2.0%; 5: OC >2.0%; 97: Water; 99: Glaciers, Rock, Shifting sand, Missing data.

CN_S: carbon to nitrogen ratio in subsoil

CN_T: carbon-to-nitrogen ratio in topsoil

"C/N ratio is calculated from the organic carbon and total nitrogen percentages. Good quality humus often has C/N values below 10, while humus of mediocre quality has values over 20".

The classes are: 0: water; 1: C/N <10; 2: C/N >=10-15; 3: C/N >15-20; 4: C/N >20.

BS_T: base saturation as a percentage of topsoil.

"Indication of the overall nutrient status of the soil. It is measured by the sum of exchangeable cations (nutrients) Na, Ca, Mg and K as a percentage of the overall exchange capacity of the soil (including the same cations plus H and Al)".

The classes are: 0: water; 1: BS <20%; 2: BS 20-50%; 3: BS >50-80%; 4: BS >80%.

CC_S: cation exchange between clay minerals and kaolinites in subsoil

"The type of clay mineral predominantly present in the soil is often typical for a specific set of pedogenetic factors: tropical leaching climates produce kaolinite, confined conditions rich in Ca and Mg with a pronounced dry season encourages the formation of smectite (montmorillonite). Clay minerals have typical exchange capacities with kaolinites generally having the lowest at about 16 mmol/l, while smectites have one of the highest with a CEC of 100g clay being 80mmol/l or more".

The classes are: 0: water; 1: <20 meq/100 g clay; 2: 20-50 meq/100 g clay; 3: >50-100 meq/100 g clay; 4: >100 meq/100 g clay.

CC_T: cation exchange between clay minerals and kaolinites in topsoil

"The type of clay mineral predominantly present in the soil is often typical for a specific set of pedogenetic factors: tropical leaching climates produce kaolinite, confined conditions rich in Ca and Mg with a pronounced dry season encourages the formation of smectite (montmorillionite). Clay minerals have typical exchange capacities with kaolinites generally having the lowest at about 16 mmol/l, while smectites have one of the highest with a CEC of 100g clay being 80mmol/l or more".

The classes are: 0: water; 1: <20 meq/100 g clay; 2: 20-50 meq/100 g clay; 3: >50-100 meq/100 g clay; 4: >100 meq/100 g clay.

CE_S: cation exchange in subsoil

"The total nutrient fixing capacity of a soil is well expressed by its cation exchange capacity. Soils with low CEC have little resilience and cannot build up stores of nutrients. Many sandy soils have CEC less than 4. The clay content, the clay type and the organic matter content all determine the total nutrient storage capacity. Values in excess of 10 are considered satisfactory".

The classes are: 0: water; 1: <4 meq/100 g; 2: 4-10 meq/100 g; 3: >10-20 meq/100 g; 4: >20-40 meq/100 g; 5: >40 meq/100 g.

CE_T: cation exchange in topsoil

"The total nutrient fixing capacity of a soil is well expressed by its cation exchange capacity. Soils with low CEC have little resilience and cannot build up stores of nutrients. Many sandy soils have CEC less than 4. The clay content, the clay type and the organic matter content all determine the total nutrient storage capacity. Values in excess of 10 are considered satisfactory".

The classes are: 0: water; 1: <4 meq/100 g; 2: 4-10 meq/100 g; 3: >10-20 meq/100 g; 4: >20-40 meq/100 g; 5: >40 meq/100 g.

PH_S: pH in subsoil and PH_T: pH in topsoil

"pH is a measure for the acidity of the soil. Five major pH classes are recognized each of which has a specific agronomic significance:

pH < 4.5 Extremely acid soils.

pH 4.5 - 5.5: Very Acid Soils suffering often from Al toxicity. Some crops are tolerant for these conditions (Tea, Pineapple).

pH 5.5 - 7.2 Acid to neutral soils these are the best pH conditions for nutrient availability.

pH 7.2 - 8.5 These pH values are indicative of carbonate rich soils".

The classes are: 1: pH <4.5; 2: pH >=4.5-5.5; 3: pH >5.5-7.2; 4: pH >7.2-8.5; 5: pH >8.5; 97: Water; 99: Glaciers, Rock, Shifting sand, Missing data.

SMAX: easily available water

"This is an indicator for the amount of stored soil moisture readily available to crops. The water retention at 2 bar suction is used to separate easily available water (EAV) from water which is more tightly held at higher suctions and difficult to abstract, especially from deeper subsoils, and in the use of a conceptual model of effective rooting depth".

The classes are: 1: Wetlands; 2: > 120 mm/m; 3: 100 - 120 mm/m; 4: 60 - 100 mm/m; 5: 40 - 60 mm/m; 6: 20 - 40 mm/m; 7: < 20 mm/m; 97: Water; 99: Glaciers, Rock, Shifting sand, Missing data.

DRAIN: soil drainage

"The soil drainage class indicates the possibility to evacuate excess moisture from a soil based on the soil unit's classification name, the soil phase(s) indicated for the dominant unit and the slope class. It considers the full composition of each mapping unit as given in the mapping unit composition. Soil drainage is indicated by 7 classes from very excessive to very poorly drained".

The classes are: 1: Not applicable; 2: Excessively drained; 3: Soils extremely drained; 4: Well drained; 5: Moderately well drained; 6: Imperfectly drained; 7: Poorly drained; 8: Very poorly drained; 97: Water bodies.

EAVSOILMOI: estimated easily available soil moisture in mm/m

"An average easily available soil moisture storage capacity".

HWSD_SMAX: estimated maximum available soil moisture in mm/m

"An average maximum available soil moisture storage capacity".

DEPTH: effective soil depth

"The effective soil depth is the depth to which micro-organisms are active in the soil, where roots can develop and where soil moisture can be stored. As such it is an essential indicator of soil health".

The classes are: 1: Very shallow (<10 cm); 2: Shallow (10-50 cm); 3: Moderately deep (50-100 cm); 4: Deep (100-150 cm); 5: Very deep (150-300 cm); 97: Water; 99: Missing data.

TERRSLOPE: terrain slopes derived from GTOPO30 (global digital elevation model, global coverage of 30-arc second elevation data; U.S. Geological Survey, 1996)

"The slope is the inclination of the land and depends on the horizontal distance considered. The use of the global altitude measures each kilometre has allowed to generate average slope estimates".

The classes are: 1: 0-2%; 2: 2-5%; 3: 5-8%; 4: 8-16%; 5: 16-30%; 6: 30-45%; 7:>45%; 8: Inland Water; 0: Undefined.

Table S9.3.1.1. Climatic and soil variables for analysis of loci under selection in continental populations in Africa. Type: type of data for each variable (continuous, categorical).

	Variable	Type	BN	KR	MB	NK	GB
Climatic							
	BIO1	cont.	27.28	25.59	22.92	22.76	37.87
	BIO2	cont.	6.79	8.54	7.93	8.05	9.58
	BIO3	cont.	6.91	7.74	7.53	7.65	6.85
	BIO4	cont.	114.93	74.09	84.97	77.67	127.32
	BIO5	cont.	32.515	31.71	28.358	28.189	32.105
	BIO6	cont.	22.79	20.785	17.891	17.754	18.24
	BIO7	cont.	9.73	10.93	10.47	10.44	13.87
	BIO8	cont.	26.855	24.515	22.448	22.268	25.895
	BIO9	cont.	27.98	25.74	23.524	23.334	23.5
	BIO10	cont.	28.71	26.51	23.72	23.47	26.75
	BIO11	cont.	25.75	24.52	21.59	21.56	23.50
	BIO12	cont.	1268.95	2870.4	2198.833	2163.415	2018.2
	BIO13	cont.	305.85	422.85	374.667	376.277	255.25
	BIO14	cont.	13.9	33.9	43.318	49.108	2.75
	BIO15	cont.	77.85	57.4	57.000	56.862	70.8
	BIO16	cont.	651.25	1218.8	887.394	891.492	872.35
	BIO17	cont.	67.50	154.30	174.71	209.20	17.55
	BIO18	cont.	227.25	477	566.273	524.400	710.95
	BIO19	cont.	314.00	1218.80	573.76	492.55	17.55
	ALT	cont.	13.0	183.9	717.1	681.7	132.4
	aridity index	cont.	0.89	1.84	1.55	1.52	1.22
Soil							
	NN_S	cat.	1.38	3.00	2.75	2.75	1.73
	NN_T	cat.	1.54	4.00	3.75	3.75	2.43
	OC_S	cat.	1.55	2.00	2.00	2.00	1.68
	OC_T	cat.	2.54	3.25	4.00	4.00	2.70
	CN_S	cat.	1.89	1.00	1.00	1.00	1.40
	CN_T	cat.	2.67	2.00	1.25	1.25	2.00
	BS_T	cat.	2.38	2.75	2.00	2.00	2.66
	CC_S	cat.	2.00	2.00	1.25	1.25	2.28
	CC_T	cat.	2.17	2.00	2.00	2.00	2.30
	CE_S	cat.	2.22	2.75	2.00	2.00	1.65
	CE_T	cat.	2.03	3.00	2.25	2.25	1.91
	PH_S	cat.	2.72	2.75	2.00	2.00	2.54
	PH_T	cat.	2.55	2.75	2.00	2.00	2.49
	SMAX	cat.	3.64	3.00	3.00	3.00	2.58
	DRAIN	cat.	2.38	4.75	3.75	3.75	3.05
	EAVSOILMOI	cont.	93	110	110	110	137
	HWSD_SMAX	cont.	142	162	162	162	163
	DEPTH	cat.	4.11	5.00	5.00	5.00	4.01
	TERRSLOPE	cat.	1	4	5	4	2

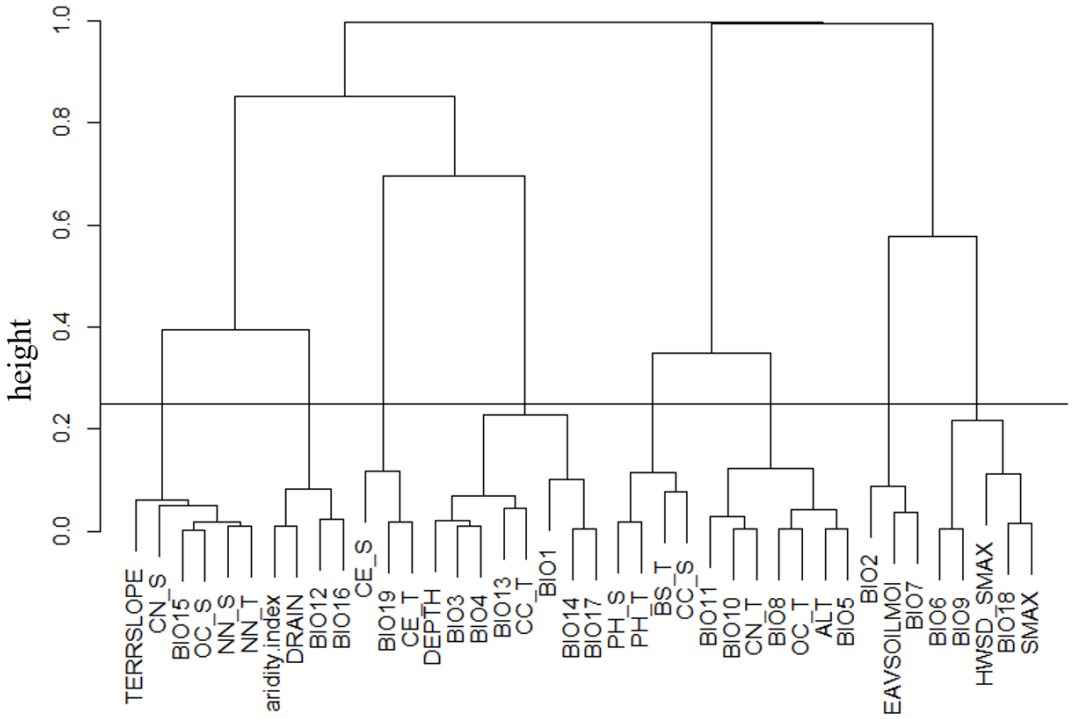


Figure S9.3.1.1. Dendrogram of environmental variables of continental populations from Africa based on Pearson correlations among variables. The horizontal line represents the limit below which we find the groups of variables correlated higher than 0.75.

S9.3.2. Analysis of loci under selection

Table S9.3.2.1. Loci identified as outliers (q-value < 0.01) in BayeScan analysis for the five continental locations in Africa. SNP ID: names identifying the loci on the original 4921 SNP dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis. F_{ST}: the F_{ST} coefficient averaged over populations. *For the model including selection*: Prob: posterior probability. log10 (PO): the logarithm of Posterior Odds to base 10. q-value: a false-discovery rate analogue to the p-value. α: the estimated alpha coefficient representing the direction and strength of selection.

SNP ID	Order	F_{ST}	prob	log10	q-value	α
Locus3	1	0.608	1.000	3.398	0.0002	1.731
Locus68	54	0.589	0.989	1.970	0.0028	1.635
Locus98	78	0.590	1.000	3.699	0.0001	1.640
Locus219	164	0.655	1.000	1,000	0.0000	1.967
Locus526	375	0.575	0.991	2.052	0.0023	1.570
Locus882	616	0.576	0.996	2.442	0.0011	1.571
Locus1083	752	0.539	0.986	1.860	0.0033	1.392
Locus1571	1087	0.603	1.000	3.699	0.0001	1.703
Locus1665	1144	0.588	0.999	2.853	0.0006	1.631
Locus2112	1451	0.614	0.996	2.355	0.0013	1.765
Locus2151	1485	0.652	0.999	2.920	0.0004	1.963
Locus2387	1652	0.572	0.992	2.072	0.0020	1.552
Locus2646	1831	0.590	0.997	2.493	0.0009	1.642
Locus2896	2001	0.539	0.980	1.690	0.0047	1.390
Locus3434	2352	0.543	0.982	1.742	0.0040	1.409
Locus3621	2474	0.572	0.995	2.299	0.0016	1.551
Locus3990	2728	0.584	0.999	2.853	0.0006	1.609
Locus4196	2880	0.587	0.999	3.000	0.0003	1.627
Locus4436	3053	0.591	0.998	2.619	0.0007	1.644
Locus4506	3108	0.562	0.975	1.587	0.0056	1.500
Locus4685	3236	0.699	1.000	3.699	0.0001	2.212
Locus4722	3261	0.616	1.000	3.398	0.0002	1.771
Locus4761	3286	0.514	0.933	1.142	0.0082	1.265
Locus4871	3361	0.643	1.000	1,000	0.0000	1.900

Table S9.3.2.2. Loci identified as outliers using the core model and the XtX statistics in BayPass for the five continental locations in Africa (99% quantile of XtX values from the simulated pseudo-observed dataset: 9.671). SNP ID: names identifying the loci on the original 4921 SNP dataset. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis.

SNP ID	Order	XtX estimates
Locus3	1	84.148
Locus98	78	14.600
Locus219	164	30.668
Locus309	235	15.719
Locus880	614	10.790
Locus1113	773	12.326
Locus1299	896	13.912
Locus1367	945	10.584
Locus1379	956	10.075
Locus1408	974	10.954
Locus1571	1087	11.322
Locus1844	1279	9.699
Locus1942	1339	9.895
Locus2037	1403	9.932
Locus2151	1485	12.765
Locus3120	2152	12.055
Locus3796	2594	9.954
Locus3814	2603	9.736
Locus4196	2880	13.497
Locus4506	3108	10.994
Locus4685	3236	11.103
Locus4871	3361	17.249
Locus4879	3368	9.804

Figure S9.3.2.1. Correlation map based on the **Ω** matrix calculated by BayPass and its visualization as a hierarchical clustering tree.

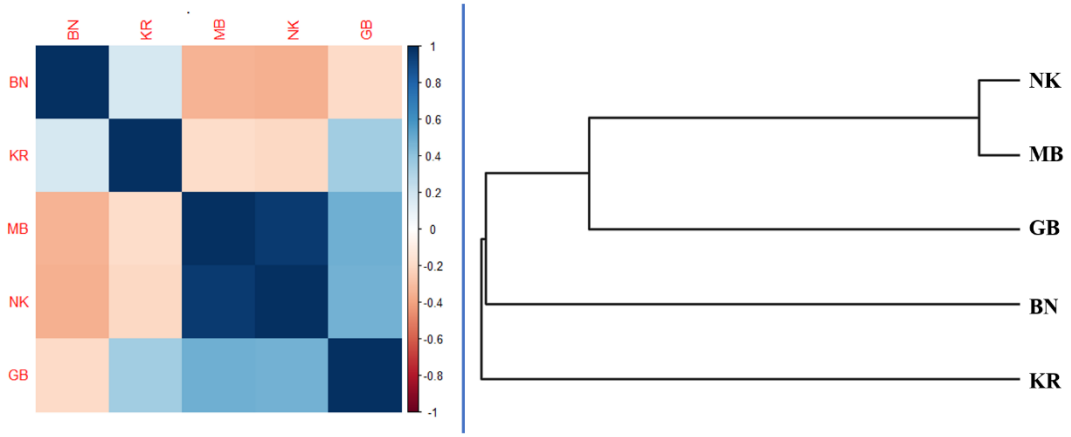


Figure S9.3.2.2. Bayes Factors for each locus and covariate in deciban (dB) units (10 x $\log_{10}$(BF)) using the standard covariate model (IS algorithm) in BayPass for the five continental locations in Africa. Order: sequential index for the 3399 SNPs selected for the outlier detection analysis and the eight covariates. The dashed line indicates BFis (in DB) > 20 (Decisive evidence in Jeffreys' scale of evidence).

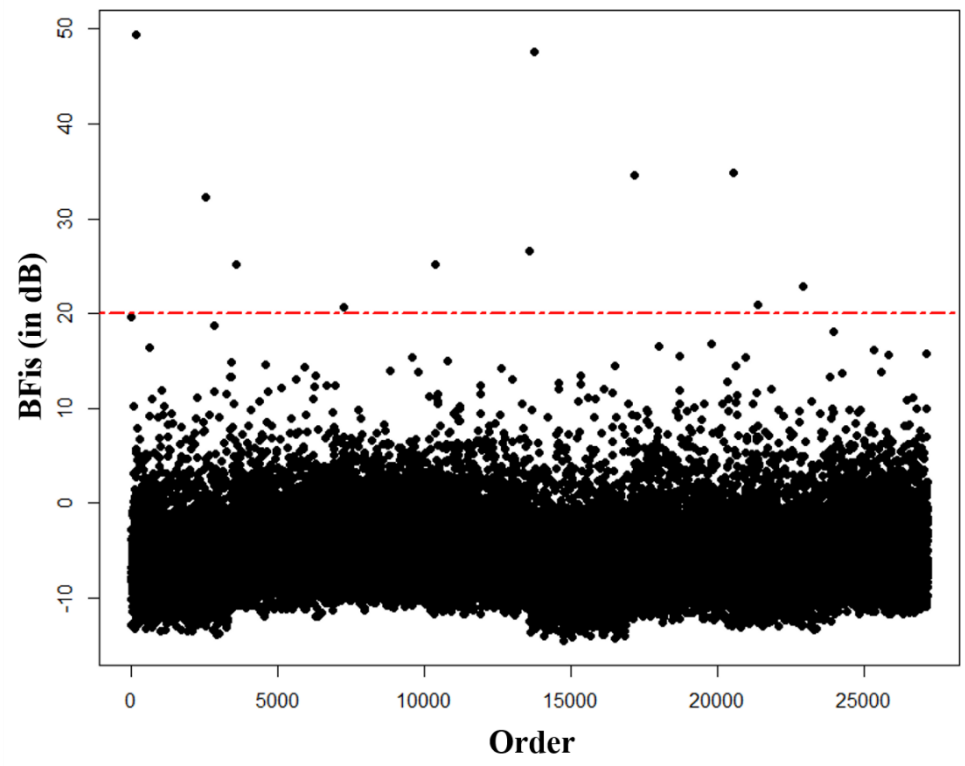


Table S9.3.2.3. Matrices used for the Mantel tests to detect if there are coincidence of extreme values of environmental variables and of extreme neutral allele frequencies in the same locations: the correlation matrix derived from Ω from BayPass and matrices for pairwise environmental distances for BIO17 and PH_T. Colours represent the magnitude of values within each matrix from higher (green) to lower (red) values.

Variables		Populations				
		BN	**GB**	**KR**	**MB**	**NK**
Correlation (Ω)	**BN**	-				
	GB	-0.193	-			
	KR	0.173	0.350	-		
	MB	-0.345	0.489	-0.180	-	
	NK	-0.352	0.472	-0.209	0.960	-
BIO17	**BN**	-				
	GB	49.950	-			
	KR	-86.800	-136.750	-		
	MB	-107.212	-157.162	-20.412	-	
	NK	-141.700	-191.650	-54.900	-34.488	-
PH_T	**BN**	-				
	GB	-0.065	-			
	KR	0.197	0.263	-		
	MB	-0.553	-0.488	-0.750	-	
	NK	-0.553	-0.488	-0.750	0.000	-

9.4. Genetic structure within the genus *Symphonia* in Madagascar

S9.4.1. Workflow to design SNP baits

This step was performed thanks to the collaboration with the Molecular Biology and Biochemistry Department, in the University of Malaga (Spain).

De novo assembly of transcriptomes

The 20 transcriptomes (18 Malagasy *Symphonia* accessions belonging to 8 putative species and two African *Symphonia globulifera* accessions, Table S9.4.1.1.) used for SNP detection in the Study III were obtained from Olsson et al. (unpublished), based on Illumina HiSeq 2000 paired end sequencing (2 x 100 bp). Transcriptomes were assembled using the workflow AutoFlow as described in Seoane et al. (2016). The full-length transcripts predicted by Full-LengtherNext (Seoane et al., 2016) were selected from each transcriptome and the ORF[3] (open reading frame) sequence of each full-length transcript was predicted as part of the workflow.

Clustering orthologues

From each transcriptome assembly, the ORFs of full-length transcripts (labelled as "Complete Sure" in the AutoFlow pipeline) were extracted. A k-mer (mbed mode) alignment was made for all complete ORF sequences (523,065 sequences) across the 20 accessions with Clustal Omega (Sievers & Higgins, 2014). Guided by the tree distance file (.dnd) given as output from the program, the transcripts were clustered in putative orthologue groups using a sequence identity threshold of 90%. In practice, each branch in the tree file was built until an identity value of 90 % minimum was reached and, then, the branch was cut to form a cluster. At this point, for visual SNP identification, we retrieved the corresponding full-length transcripts, including untranslated regions (UTRs), from the individual transcriptomes for each cluster.

Automated SNP calling

The clusters obtained in the previous stage were filtered to extract only those clusters encompassing ten different accessions. Then, for each assembly, the number of transcripts in these filtered clusters were counted. The accession with the largest number of transcripts across the clusters was selected as reference accession (i.e., MH2809). Only those clusters with transcripts from this accession were used to build the mapping reference, based on full sequences (including ORF and UTR). Each accession in each cluster was mapped against the reference (MH2809 presented a total of 935 reference transcriptome sequences) with Bowtie2 (Langmead, Trapnell, Pop, & Salzberg, 2009) using default parameters. SNP calling was done with VarScan2 (Koboldt et al., 2012) using default parameters (each accession results were recorded in a vcf file). These individual SNP calls were merged to one vcf file containing all accessions, on which we applied our criteria for the automated identification of candidate SNPs.

[3] ORF is a sequence of nucleotide triplets. Those triplets are read as codons which specify amino acids. The ORF does not contain stop codons.

Table S9.4.1.1. Species, locations, accession numbers (ID) and assignment to SSR gene pools in Malagasy *Symphonia* (rGP) for *Symphonia* samples used for transcriptome sequencing and candidate SNP identification. *Spp*.: Putative species identified based on the plant specimens (we could not collect branches with leaves for all samples, neither all plant specimens collected could be identified).

Species	Location	ID	rGP	*Spp.*
	Continental Africa			
Symphonia	Dja-et-Lobo, Cameroon	JD357	-	*S. globulifera*
globulifera	Dja-et-Lobo, Cameroon	JD413	-	*S. globulifera*
	Madagascar			
Malagasy	Andasibe-Mantadia NP	MH2809	rGP5	*S. fasciculata*
Symphonia		MH2816	rGP2	*S. louvelii*
		MH2818	rGP2	-
		MH2826	rGP3	*S. urophylla*
		MH2832	rGP3	*S. urophylla*
		MH2841	rGP5	-
		MH2853	rGP2	*S. louvelii*
	Farankaraina	MH2949	rGP2	*S. sp.1 (Farankaraina)*
		MH2964	rGP4	*S. sp.1 (Nosy Mangabe)*
		MH2966	rGP2	*S. sp.1 (Farankaraina)*
	Ranomafana NP	MH3010	rGP5	-
		MH3020	rGP1	*S. eugenioides*
		MH3064	rGP1	-
		MH3082	rGP1	-
	Ialatsara_1	MH3044	rGP3	-
		MH3049	rGP1	*S. microphylla*
	Ialatsara_2	MH3105	rGP1	*S. clusioides*
	Ankazomivady	MH3140	rGP1	*S. clusioides*

S9.4.2. SNP genotypes of Symphonia globulifera and Malagasy Symphonia samples generated by Sequenom technology

This file contains:

SNP data:

It contains individual IDs, population IDs and genotype data for 144 SNPs in a final set of 628 individuals.

Genotypes are represented with four digits or letters, to cover both situations: diploid (e.g., AA99) or tetraploid patterns (e.g., AGGG), and missing data (9999).

60 SNPs presented diploid patterns for all individuals (marked in grey).

In Paracou and Regina, the *S. globulifera* morphotype is given in the Population field: symglo for the common morphotype, and ssp1 for the alternative morphotype.

Sequences:

Consensus sequences of contigs for the 144 SNPs selected.

The file can be download from: https://doi.org/10.5281/zenodo.5772441

S9.4.3. Analysis on functional SNP markers in the genus Symphonia

Table S9.4.3.1. Estimates of genetic distance based on Nei's D among gene pools of Malagasy *Symphonia* and populations of *S. globulifera* (including the alternative morphotype) in Africa and the Neotropics: A) based on 53 diploid snps, B) based on 124 diploid SNPs (nGP1 not included).

		BN	KR	MN	GB	ST	IT	PR	RG	SP	nGP1	nGP2	nGP3	nGP4	nGP5
A)	**BN**	-													
	KR	0.0225	-												
	MN	0.0543	0.0226	-											
	GB	0.0497	0.0198	0.0031	-										
	ST	0.0246	0.0104	0.0265	0.0277	-									
	IT	0.0221	0.0113	0.0314	0.0325	0.0045	-								
	PR	0.0233	0.0129	0.0354	0.0355	0.0071	0.0048	-							
	RG	0.0208	0.0107	0.0330	0.0340	0.0033	0.0012	0.0036	-						
	SP	0.0409	0.0277	0.0480	0.0531	0.0201	0.0207	0.0233	0.0194	-					
	nGP1	0.2003	0.1593	0.1733	0.1744	0.1632	0.1754	0.1786	0.1730	0.1714	-				
	nGP2	0.3443	0.3203	0.3298	0.3294	0.3167	0.3198	0.3335	0.3251	0.3449	0.1784	-			
	nGP3	0.1967	0.1673	0.1804	0.1828	0.1706	0.1823	0.1813	0.1794	0.1878	0.0401	0.1941	-		
	nGP4	0.2306	0.1984	0.2128	0.2021	0.201	0.2135	0.2059	0.2109	0.2244	0.0899	0.1541	0.0954	-	
	nGP5	0.4477	0.4174	0.4328	0.4214	0.4197	0.4322	0.4096	0.4274	0.4513	0.2483	0.2916	0.2737	0.2391	-
B)	**BN**	-													
	KR	0.0318	-												
	MN	0.0445	0.0157	-											
	GB	0.0429	0.0125	0.0034	-										
	ST	0.0275	0.0405	0.0440	0.0452	-									
	IT	0.0260	0.0399	0.0460	0.0467	0.0180	-								
	PR	0.0181	0.0332	0.0395	0.0401	0.0113	0.0100	-							
	RG	0.0174	0.0323	0.0389	0.0399	0.0098	0.0085	0.0018	-						
	SP	0.0273	0.0405	0.0463	0.0489	0.0181	0.0116	0.0113	0.0097	-					
	nGP2	0.2779	0.2839	0.2975	0.2920	0.2836	0.2892	0.2829	0.2814	0.2912	-	-			
	nGP3	0.1833	0.1822	0.1872	0.1838	0.1930	0.2004	0.1877	0.1900	0.1947	-	0.1517	-		
	nGP4	0.2282	0.2227	0.2330	0.2245	0.2296	0.2345	0.2188	0.2237	0.2316	-	0.1359	0.0999	-	
	nGP5	0.3144	0.3225	0.3397	0.3306	0.3306	0.3352	0.3147	0.3235	0.3344	-	0.1554	0.1877	0.1831	-

S9.4.4. SSR development, genotyping and genetic structure of Malagasy Symphonia individuals

Plant material, DNA extraction and SSR genotyping

Plant material was collected during several missions to Madagascar between 2010 and 2014. For DNA extraction, leaf or cambium samples from 434 trees belonging to the genus *Symphonia* were collected in ten locations (individuals were coincident with those from SNP genotyping, see Table 7, Fig. 16), most of them in the humid tropical rainforest part in eastern Madagascar. The material was dried in paper bags surrounded by silica gel. The samples covered 12 putative *Symphonia* species. For some trees, representing the morphological diversity of the genus, plant material was also preserved in RNAlater buffer, for RNA extraction and transcriptome sequencing. Herbarium vouchers were collected representing the different morphotypes, e.g., putative species, in each site, and deposited at the herbarium of the CNARP in Antananarivo Madagascar and the MAD herbarium in Madrid, Spain.

DNA was extracted from 20 mg of dry weight tissue using the DNeasy Plant mini Kit (Qiagen, the Netherlands). A total of 21 transcriptome derived nuclear simple sequence repeat markers (SSRs) developed by Olsson et al. (2017) were amplified using the methods described by Olsson et al., (2017) with small alterations in the composition of the multiplex mixes (see Table S9.4.4.1.). Primers for locus 7972 were added in Mix1 $((AAT)^4$ F: Q4-GTTTGCCCTTCTGGCTTCG, R: TGAATGGAATACTTTACGGCACC) and locus 16672 in Mix3 $((AGGGAT)^5$ F: Q4- TGGAATGCCATCCGAATTTGAG, R: TGGACAGGAACTGTGGAGC). They were left out from the study design of Olsson et al., (2017) because they did not amplify the tested *S. globulifera* samples. Also, to avoid size conflict in the new multiplexes, locus 3131 was transferred from Mix1 to Mix3.

All SSR markers amplified samples from Madagascar except for locus 2978, which was omitted from the current study. Locus 2978 was also omitted from population genetics analyses in Olsson et al. (2017) due to low level of polymorphism in *S. globulifera.* Therefore, we genotyped 434 individuals with a final set of 20 SSR markers. During genotyping some samples showed more than two peaks per locus possibly indicating polyploidy.

Genetic structure analysis based on SSRs

We performed a genetic structure analysis using STRUCTURE (Pritchard et al., 2000), with the aim to get an overview about the genetic structure of the genus *Symphonia* in Madagascar, as a first step to unveil drivers of genetic structure in this species complex (such as species differentiation, geographic and ecological isolation, polyploidy, etc.). STRUCTURE analysis was carried out using an admixture model with correlated allele frequencies for codominant markers and for clusters, *K*, from 1 to 10 (burn-in length: 50,000, run length: 100,000 iterations, 10 repetitions for each number of clusters, chain convergence visually checked). The best *K* was supported based on the highest logarithm probability of data (L(*K*)) and Delta *K* (ΔK), following Pritchard, Wen, & Falush (2010) and Evanno, Regnaut, & Goudet (2005). Then, individuals were assigned to one of the gene pools based on SSR markers (hereafter rGP) when ancestry proportions (Q) ≥ 0.5.

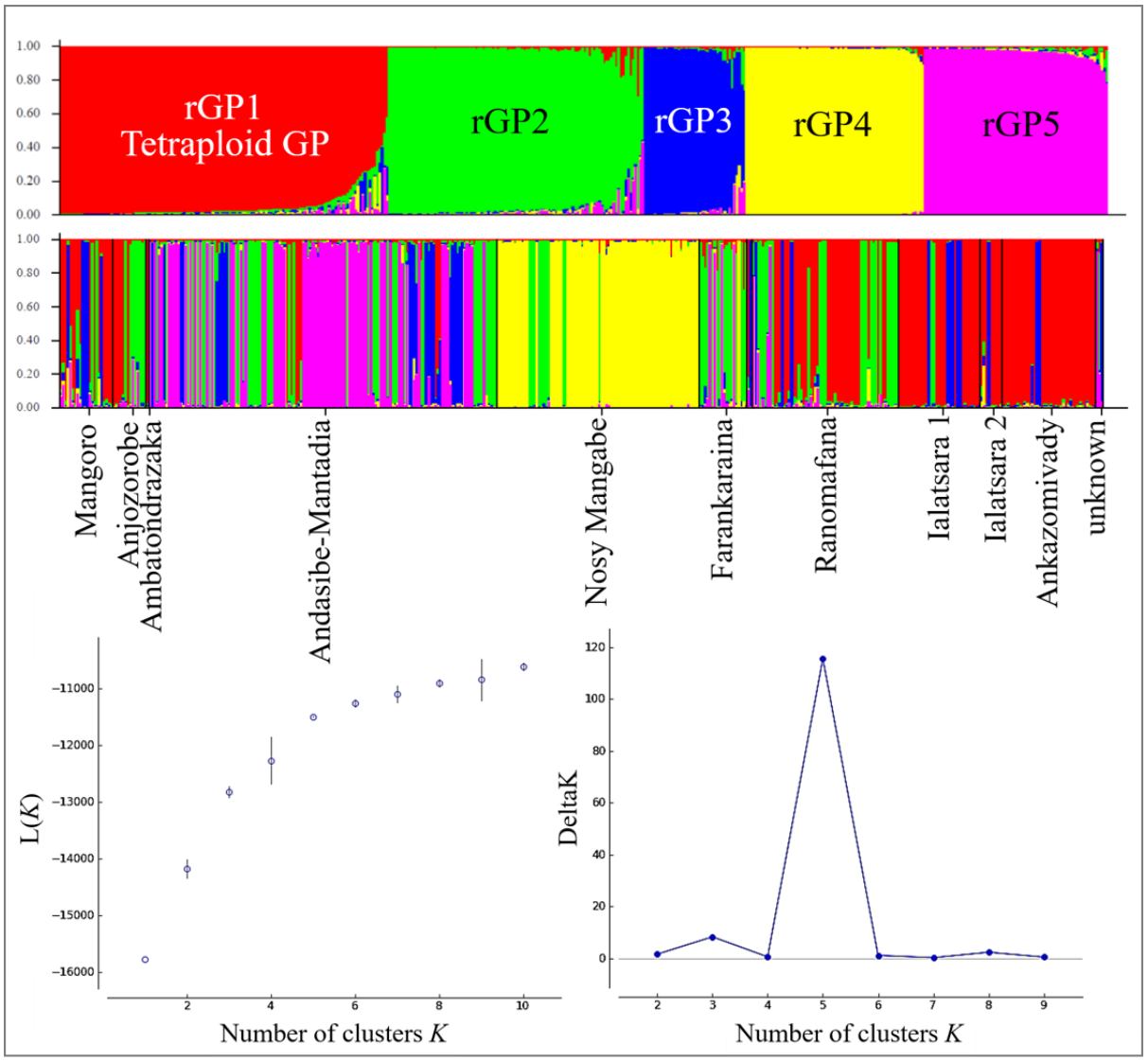
Genetic structure results based on SSRs

The number of clusters best describing the data corresponded to *K*=5. All individuals were assigned to one of the five detected rGPs, except for one individual from the Mangoro population. Only rGP4 corresponded exclusively to a single population (Nosy Mangabe Island). rGP1 and rG3 were mostly present in central-south and central populations (rGP1 and rGP3 very scarce or absent in Andasibe and Anjozorobe respectively). rGP5 was mainly present in Andasibe-Matadia and Farankaraina, and rGP2 occured in the three sampled regions (north, central and central-south) with difference rates of presence (see Fig. 16B). Transcriptome sequencing and candidate SNP identification was performed on *Symphonia* samples representing all five rGPs detected, each represented by one or more individuals (Table S9.4.1.1.).

Table S9.4.4.1. Details on the 20 polymorphic SSR markers used for Malagasy *Symphonia* SSR genotyping, including primer sequences and GenBank accession numbers. Repeat: Number of repeats found in the sequence that corresponds to the accession number. Fluorochrome in 50 was 6-FAM for Q2 (TAGGAGTGCAGCAAGCAT), VIC for Q3 (CACTGCTTAGAGCGATGC) and NED for Q4 (CTAGTTATTGCTCAGCGGT). GenBank Acc.: The accession number in GenBank.

Locus	Repeat	Q-tail	Mix	Primer sequence (5'–3 ')	GenBank Acc.
1582	$(ATC)^4$	Q3	1	F: Q3-GTGGTGGGATTGCTGCTATT	KR363116
				R: TGGCAAGGAACAAGTGAAGA	
3131	$(ACC)^5$	Q3	3	F: Q3-TCGAAGAAGAAAGCATTTACGTG	KR363118
				R: ATGAGTACGTTCCAGGGCG	
3984	$(ACC)^4$	Q2	1	F: Q2-TTACGTGCAAGAAGATTCACG	KR363119
				R: ACCACAACCCGCTCATACAC	
4464	$(CTT)^9$	Q3	3	F: Q3-CCGCTTGAATCTTCAATTTCTC	KR363120
				R: AACGAACTTGGTGGTCTTGG	
5489	$(GGATT)^4$	Q2	3	F: Q2-AGAAGGACTTGACGGTGCC	KR363109
				R: GGAGCGGAAAGTGGACTCG	
6387	$(AAT)^5$	Q2	1	F: Q2-ACGGGGATCAGATCGAGTTT	KR363121
				R: TCACACATAACAGAATTTGCAATC	
6636	$(GGTTT)^5$	Q2	1	F: Q2-CAGTGGGATGAAACCGAAAT	KR363110
				R: CCCGTAACTTTGACCCAACA	
6783	$(GCT)^4$	Q2	1	F: Q2-AATACGCAGAGATGGGCAC	KR363111
				R: GAATGCTCGGGTTCAAATGC	
7189	$(AAG)^4$	Q3	1	F: Q3-CCGACTTCACATCCCTAAACC	KR363112
				R: GACCGAGATGCTTGATTCCC	
7694	$(GTT)^7$	Q3	2	F: Q3-GGCACTAATCCGGAAACCAG	KR363122
				R: TCTCCACGAAAGCTCAGGTC	
7972	$(AAT)^4$	Q4	1	F: Q4- GTTTGCCCTTCTGGCTTCG	unpublished
				R: TGAATGGAATACTTTACGGCACC	
9610	$(ATC)^6$	Q2	2	F: Q2-GGGAGCAAGAAGCACTGTC	KR363123
				R: TGATGAGGCTTGATTGGCG	
9990	$(GCT)^7$	Q3	2	F: Q3-TCGTTGCTTTACCGAACTCC	KR363124
				R: CCATCCATATCGAAGATGACG	
10829	$(AGC)^7$	Q2	3	F: Q2-ACTATGGTTTGGGTCCCGTC	KR363113
				R: ACTCCCTGGCAAAGAACCC	
10904	$(AGC)^6$	Q2	1	F: Q2-ATCTCTCCTCCCAGTGCAG	KR363114
				R: GGCTCAAGGCAACTTGGTC	
14623	$(CTT)^5$	Q2	2	F: Q2-TAGGTGGGGAGAAGGATGC	KR363125
				R: TAAGGGAAGGAGGTGAACGA	
15834	$(AGCG)^7$	Q2	3	F: Q2-GGGTTGGTGGATCGAGTACC	KR363126
				R: AAGAGCATAGCGCTTGACG	
15979	$(GGT)^7$	Q4	1	F: Q4-GCTTTTGTCTCGGCACTTGT	KR363115
				R: CTCCAAACCGACTAGGACCA	
16615	$(AAC)^7$	Q4	2	F: Q4-GCCGAAAACCACCAAACC	KR363127
				R: CGGAAGCTATAGGAAGGGATT	
16672	$(AGGGAT)^5$	Q4	3	F: Q4- TGGAATGCCATCCGAATTTGAG	unpublished
				R: TGGACAGGAACTGTGGAGC	

Figure S9.4.4.1. Illustration for the best number of genetic clusters (*K*=5) describing the data in the STRUCTURE analyses for 20 SSR markers in 10 Malagasy locations (barplots based on the best run for *K*=5). Colours illustrate the ancestry proportions for each of the *K* gene pools. Upper and lower barplots are sorted by *Q* and population, respectively. rGP1 corresponded to the tetraploid individuals discovered during the SNP genotyping in Study III (see Section 3.1.3., 3.2.3. and 4.3.). The best *K* was supported based on the logarithm probability of data (L(*K*)) and Delta *K* (Δ*K*) (plots modified after outputs from STRUCTURE Harvester software, Earl & VonHoldt, 2012). rGP corresponds to gene pools based on SSR markers.



S9.4.5. Images of Malagasy plant specimens collected

Figure 9.4.5.1. Sample ID: MH2724 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia sessiliflora.*

Figure 9.4.5.2. Sample ID: MH2878 collected in Nosy Mangabe. Putatively identified as *Symphonia sessiliflora.*

Figure 9.4.5.3. Sample ID: MH2812 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia clusioides.*

Figure 9.4.5.4. Sample ID: MH3138 collected in Ankazomivady. Putatively identified as *Symphonia clusioides.*

Figure 9.4.5.5. Sample ID: MH2920 collected in Nosy Mangabe. Putatively identified as *Symphonia sp.1* (Nosy Mangabe).

Figure 9.4.5.6. Sample ID: MH2947 collected in Nosy Mangabe. Putatively identified as *Symphonia sp.1* (Nosy Mangabe).

Figure 9.4.5.7. Sample ID: MH2778 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia eugenioides.*

Figure 9.4.5.8. Sample ID: MH3057 collected in Ranomafana National Park. Putatively identified as *Symphonia eugenioides.*

Figure 9.4.5.9. Sample ID: MH3020 collected in Ranomafana National Park. Putatively identified as *Symphonia eugenioides.*

Figure 9.4.5.10. Sample ID: MH3026 collected in Ialatsara 1. Putatively identified as *Symphonia microphylla.*

Figure 9.4.5.11. Sample ID: MH3023 collected in Ialatsara 1. Putatively identified as *Symphonia microphylla.*

Figure 9.4.5.12. Sample ID: MH3049 collected in Ialatsara 1. Putatively identified as *Symphonia microphylla.*

Figure 9.4.5.13. Sample ID: MH2746 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia urophylla.*

Figure 9.4.5.14. Sample ID: MH2838 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia urophylla.*

Figure 9.4.5.15. Sample ID: MH2950 collected in Farankaraina. Putatively identified as *Symphonia sp.1 (Farankaraina)*.

Figure 9.4.5.16. Sample ID: MH2953 collected in Farankaraina. Putatively identified as *Symphonia sp.1 (Farankaraina)*.

Figure 9.4.5.17. Sample ID: MH2822 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia fasciculata.*

Figure 9.4.5.18. Sample ID: MH2765 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia fasciculata.*

Figure 9.4.5.19. Sample ID: MH2788 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia louvelii.*

Figure 9.4.5.20. Sample ID: MH2803 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia louvelii.*

Figure 9.4.5.21. Sample ID: MH2766 collected in Andasibe-Mantadia National Park. Putatively identified as *Symphonia nectarifera*.

Figure 9.4.5.22. Sample ID: MH3094 collected in Ranomafana National Park. Putatively identified as *Symphonia nectarifera*.

Figure 9.4.5.23. Sample ID: MH3095 collected in Ranomafana National Park. Putatively identified as *Symphonia nectarifera*.