

Evidence of high genetic differentiation driven by limited gene flow in a lower canopy African tropical rainforest tree species, *Coula edulis* Baill. (Coulaceae)

Narcisse Guy Kamdem

guynarcissekamdem@yahoo.fr

Evolutionary Biology and Ecology Unit, CP 160/12, Faculté des Sciences

Bonaventure Sonké

University of Yaoundé I

Saskia Sergeant

UMR AGAP - SEAPAG team Bâtiment C

Tariq Stevart

Herbarium et Bibliothèque de Botanique africaine

Vincent Deblauwe

International Institute of Tropical Agriculture

Olivier J. Hardy

Evolutionary Biology and Ecology Unit, CP 160/12, Faculté des Sciences

Research Article

Keywords: *Coula edulis*, isolation by distance, limited dispersal, phylogeographic signal, relict population

Posted Date: September 6th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7285744/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Ecology and Evolution on December 3rd, 2025. See the published version at <https://doi.org/10.1186/s12862-025-02483-8>.

Abstract

Background

The distribution of intra-specific genetic diversity in tropical African forest tree species often reveals discontinuities in the form of genetic clusters distributed in parapatry or allopatry. To date, most population genetic studies have focused on canopy trees with potentially high gene dispersal capacities. In contrast, few studies have examined lower canopy tree species, whose more limited gene dispersal may exacerbate genetic discontinuities. In this study, we used nuclear microsatellites (SSRs) to characterize the genetic structure of populations of *Coula edulis* Baill., a lower canopy species commonly known as "African walnut", exploited for its edible seeds.

Results

Across its distribution range, we detected one genetic cluster in Upper Guinea (UG) and six in Lower Guinea (LG). High levels of genetic differentiation ($F_{ST} = 0.39$ to 0.59) were found between UG and LG, while differentiation within LG varied widely ($F_{ST} = 0.08$ to 0.50). Genetic discontinuities associated to high differentiation and a phylogeographic signal ($R_{ST} > F_{ST}$) suggest ancient divergence, possibly driven by population fragmentation during historical climatic fluctuations, while those associated with limited differentiation may reflect more recent divergence and/or genetic homogenization effect due to ongoing gene flow. Genetic diversity within LG clusters was highly variable ($H_e = 0.40$ – 0.71), with evidence of a founder or bottleneck effect observed in the southeastern Cameroon cluster, resulting in reduced diversity of a relict population. The UG population also showed low diversity ($H_e = 0.38$), likely attributable to a historical bottleneck. Morphometric analyses of herbarium specimens revealed some differentiation between LG and UG populations, questioning the taxonomic status of the taxon.

Conclusions

Our results suggest that, given the limited gene dispersal observed in *C. edulis*, the observed genetic discontinuities are expected to persist over extended timescales and provide baseline information for the conservation and potential domestication of the species' genetic resources.

1. Background

Tropical rainforests are among the most complex and biodiverse ecosystems on earth. For this reason, Leigh and collaborators [1] referred to them as "museums of diversity". In Africa, our understanding of the evolutionary history of rainforest tree species, which are the main structural components of these ecosystems, has greatly advanced over the past two decades, thanks to several phylogeographic studies (e.g., [2–14]).

In Africa, tropical rainforests are predominantly found in the Guineo-Congolian region [15], which forms the second largest block of tropical forest globally after the Amazon Basin. Although less diverse in plant species than the rainforests of Amazonia and Southeast Asia [16, 17], African rainforests are nevertheless considered biodiversity refugia due to their high levels of species richness and endemism despite their level of disturbances [18, 19]. Today, these ecosystems are facing increasing threats from deforestation and degradation, with forest loss affecting 83.3%, 9.0%, and 93.0% of forest areas in West, Central, and East Africa, respectively [20, 21].

Like tropical forests on other continents, African tropical rainforests have not been ecologically stable throughout their history [22–24]. Paleobotanical and palynological reconstructions indicate that major climatic fluctuations during the Quaternary caused rainforest cover in Africa to alternate between phases of contraction and expansion [25–33]. Other disciplines, such as phylogeography, have interpreted the spatial genetic structure of certain tree species as a legacy of climatic fluctuations, suggesting the existence of major historical discontinuities in population connectivity and distribution, due to the presence of gene flow barriers [5, 8, 9, 12, 34–37]. The alternation of glacial and interglacial periods during the Pleistocene, particularly since 1.05 million years ago [38, 39], had profound impacts on the composition of African forests and likely contributed to shaping the phylogeographic patterns of their tree species [40–44].

A consistent finding across these phylogeographic studies is the presence of geographic discontinuities in the genetic variation among tree populations, whereby differentiated genetic clusters are distributed in parapatry or allopatry. These discontinuities point to historical population isolation, likely driven by past climatic oscillations and/or the presence of inhospitable habitat barriers [5, 9, 12, 14, 34–37]. These barriers play a key role in genetic differentiation between populations of a species and can be the starting point for a process of speciation, the driver of biological diversification in plants [45, 46]. In addition, several population genetic studies have revealed cryptic species among African trees (e.g., [47–50]). It is therefore useful to check whether well-differentiated genetic groups (genetic clusters) might in fact be hiding distinct species.

Population genetics studies have also shown that individuals in closer geographic proximity tend to be more genetically similar, indicating the presence of a spatial genetic structure (SGS) at either fine or large scales. This structure results from restricted or disrupted gene flow within or between populations [51–53]. Investigating SGS within each genetic cluster provides insights into gene dispersal capacity and potential barriers to gene flow.

Most phylogeographic studies conducted on African rainforest trees have focused on timber species, which are typically emergent or upper-canopy species that are heliophilous, with high pollen and/or seed dispersal potential [36, 54–57]. This may be explained by the fact that flower production at the top of the canopy favours longer-distance dispersal by wind (exposure to strong winds) or by highly mobile animals such as birds or bats [58–62], for both seeds and pollen. In contrast, the phylogeography of lower-canopy tree species with potentially more limited gene flow remains understudied, while it might reveal

patterns resulting from older events that have not yet been erased by subsequent gene flow. Recently, tree species providing non-timber forest products like edible fruits or nuts, especially those that are domesticated, have received increased attention [63–66], while wild, undomesticated fruit tree species remain understudied. In addition, archaeological records from Central Africa [67–69] have identified several forest species that have witnessed ancestral human activities, but their evolutionary history remains largely understudied. For these important genetic resources, knowledge of the geographic patterns of genetic diversity would be useful to orient domestication or conservation efforts [70–72].

To fill this gap, we focus on *Coula edulis* Baill., a widespread lower canopy tree species in African rainforests, commonly known as the "African walnut". The seeds of *C. edulis* have been used by humans since ancient times [68, 69, 73], and it continues to serve as an important food and economic resource for local communities today [74, 75]. A previous study based on limited sample size ($n = 18$ individuals in 9 populations) and chloroplastic markers [10] suggested that its populations may have undergone a large expansion over a short period of time, possibly mediated by human dispersal, to explain the very low genetic diversity detected. However, the recent development of nuclear microsatellite markers (SSRs) has revealed high levels of genetic polymorphism [76] and limited pollen and seed dispersal distances [77].

Here, we use SSRs to characterize the genetic structure of *C. edulis* populations throughout its natural range. Specifically, we address the following questions: (i) Are there genetic discontinuities within *C. edulis* populations in the form of differentiated genetic clusters? (ii) Do patterns of intra- and inter-cluster genetic diversity provide clues about the origin of the genetic structure in relation to past climate fluctuations or barriers to gene flow? (iii) Are patterns of genetic divergence reflected in morphological variation?

2. Methods

2.1. Study taxon and sampling

Coula edulis Baill., commonly known as "African walnut", is a tree species endemic to the Guineo-Congolian tropical rainforests. It is highly valued by rural communities for its edible seeds, which are frequently sold along roadsides and in urban markets [75, 78]. *Coula edulis* belongs to the Coulaceae family, a small pantropical family, previously included in the Olacaceae, consisting of three monotypic genera distributed on distinct continents [79]. It is found in tropical evergreen forests of Africa, with occasional occurrences found in semi-deciduous forests [15, 80]. Its distribution extends from Sierra Leone to the Democratic Republic of the Congo [80]. *Coula edulis* is a medium-sized tree, reaching up to 25 m in height, and is typically found in the lower canopy of mature forests [75]. Seeds are dispersed by small rodents such as *Atherurus africanus* "African brush-tailed porcupine", *Cricetomys emini* "Emin's rat", and squirrels, while larger mammals like *Loxodonta cyclotis* "forest elephants" and some primates predate the seeds [77, 81, 82].

To ensure broad geographic representation of the species across its natural range, we assembled a collection of 405 georeferenced samples consisting of leaf or cambium tissue dried in silicagel. These samples were collected during field missions conducted by the authors and collaborators in most countries where the species occurs. The reference specimen of *Coula edulis* Baill. used in this study was identified by Dr. Gilles Dauby. It was collected in Gabon, specifically in the Ogooué-Lolo province (coordinates: -0.23325° N, 12.726614° E), and deposited at the herbarium of the Université Libre de Bruxelles (BRLU) under the voucher number GD2428.

To reduce sampling bias due to unequal collection intensity across locations, we subsampled the database by retaining no more than two individuals per km² cell. This strategy also helped to reduce the risk of biased clustering due to sample overrepresentation in certain areas [83]. In addition to silicagel dried tissues, we incorporated herbarium material comprising 14 specimens from the herbarium of the Université Libre de Bruxelles (BRLU) and 40 specimens from the Meise Botanic Garden (BR). The final dataset comprised 400 individuals. Of these, 352 originated from Lower Guinea (Cameroon, Republic of the Congo, Gabon, Equatorial Guinea, Nigeria, and southwestern Democratic Republic of the Congo) and 48 from Upper Guinea (Ivory Coast, Ghana, and Liberia). To our knowledge, this is the first phylogeographic study of *C. edulis* that covers nearly its entire distribution range, except in Sierra Leone and Nigeria.

All these samples analyzed were collected from state-owned lands, with the necessary collection permits granted by national authorities. In Cameroon, authorization was obtained from the Ministry of Scientific Research and Innovation (MINRESI; Permit No. 000102/MINRESI/B00/C00/C10/C13); in Liberia, from the Forestry Development Authority (FDA; Ref: MD/04/2016/-2); and in Gabon, from the National Center for Scientific and Technological Research (CENAREST; Permit No. AR0034/19/MESRSTT/CENAREST/CG/CST/CSAR).

Additional samples from countries such as Ivory Coast, Ghana, Nigeria, Equatorial Guinea, Republic of the Congo, and Democratic Republic of the Congo were obtained from herbarium specimens at BRLU and BR.

2.2. DNA extraction and nuclear microsatellite genotyping

For the samples dried in silicagel, DNA was extracted from 20–25 mg dry weight of leaves and 30–35 mg dry weight of cambium per sample using the NucleoSpin 96 kit for plants (Macherey-Nagel) according to the manufacturer's instructions. For herbarium samples, DNA was extracted from 25 to 35 mg dry weight of leaves per sample following the protocol described in [84].

Using 21 nuclear microsatellites (SSRs), we genotyped 400 samples according to the protocol described in [76] using an ABI3730 sequencer (Applied Biosystems, Lennik, Netherlands). Only samples for which at least 8 of the 21 loci had amplified were used in subsequent analyses, resulting in the suppression of 15 samples. Initially, we maintained all 21 loci for detecting different genetic clusters; in a second phase,

we eliminated four loci (Ced18, Ced22, Ced23, and Ced25) from subsequent analyses as they showed a relatively high frequency of null alleles.

2.3. Identification of genetic discontinuities

For the identification of different genetic clusters, we used the Bayesian clustering algorithm implemented in STRUCTURE 2.3.4 [85]. We used the model with admixture and uncorrelated allele frequencies without a priori grouping of individuals. We modeled genetic clusters for a number of clusters ranging from $K = 1$ to 10, with 10 iterations for each K , setting a burn-in period of 10,000 MCMC steps followed by 100,000 MCMC replicates. For the different runs, an alternative ancestry prior was chosen following [86] with $\alpha = 0.2$ to improve inferences of K and individuals assignment to populations. We assessed the optimal value of K using STRUCTURE HARVESTER [87] based on both the mean log-likelihood of the data [$\ln P(K)$] and the ΔK method of [88].

To validate the clustering results, we first conducted additional STRUCTURE analyses within each identified genetic cluster to assess whether any further subdivision was justified. Secondly, we applied on our genotypes a maximum-likelihood clustering algorithm implemented in the function 'snapclust' of the 'adegenet' package [89] implemented in R v. 4.3.1 [90].

For each individual, the proportion of the genome assigned to a given genetic cluster (q-value) was used to define cluster membership. Individuals with $q \geq 0.8$ for a cluster were assigned to a single genetic cluster. Those with $q \geq 0.2$ for at least two clusters were considered admixed. Then, we estimated the introgression rate between pairs of clusters, as Introgression Rate (%) = $(n/N) \times 100$, where n represents the number of individuals assigned to both clusters with $q \geq 0.2$, and N represents total number of individuals assigned to at least one of these clusters with $q \geq 0.2$.

The spatial limits of each genetic cluster were visually assessed based on the map of individuals with $q \geq 0.8$ and admixed individuals. We have added on the map an approximate distribution limit of the species, drawn manually according to the distribution of *C. edulis* samples found in GBIF (Global Biodiversity Information Facility), CJBG (Conservatory and Botanical Garden of the City of Geneva), and World Flora online databases, after excluding specimens mentioned in botanical gardens.

2.4. Diversity and differentiation parameters: detecting a phylogeographic signal

For each genetic cluster, the following multilocus genetic diversity parameters were calculated using SPAGeDi v. 1.5 [91]: total number of alleles (A_o), allelic richness (R_s), and expected heterozygosity (H_e). Analyses of variance were performed using R software [90] across the different genetic clusters to determine if there were significant differences in terms of genetic diversity (H_e), considering the loci as a random variable. We used INEst 1.0 [92] to estimate in each genetic cluster a corrected inbreeding coefficient (F_i) taking into account the effect of null alleles.

F_{ST} [93] and R_{ST} [94] were calculated to evaluate pairwise genetic differentiation between the inferred genetic clusters. F_{ST} is based on allele identity, whereas R_{ST} is based on allele size and can be used to estimate the contribution of stepwise SSR mutations to genetic differentiation [95]. Therefore, the observed R_{ST} values were compared to the F_{ST} values, and using 10,000 permutations of allele sizes, we tested whether $R_{ST} > F_{ST}$ ($P < 0.05$), as expected in the presence of a phylogeographic signal at the level of SSRs [95].

2.5. Characterization of the spatial genetic structure within and between genetic clusters

Gene flow patterns influence the spatial genetic structure (SGS) in plant populations. Given the observed genetic discontinuities, we assessed whether genetic discontinuities resulted from isolation by distance (IBD) or from true gene flow barriers. SGS was characterized (*i*) within each genetic cluster where we expect to detect an isolation by distance pattern, and (*ii*) between pairs of geographically contiguous genetic clusters to assess whether their genetic structures are spatially independent (no correlation between genetic similarity and spatial distance), or still connected through gene flow (decay of genetic similarity with spatial distance; [95]). To this end, following the approach of [51], we analyzed the kinship coefficient (F_{ij}) between pairs of individuals i and j and assessed how it evolves with spatial distance when i and j belong to the same genetic cluster or different ones. If clusters result from IBD or a secondary contact with long-standing gene flow, we expect that the F_{ij} curve between genetic clusters decreases with distance in parallel to the intra-cluster F_{ij} curves. This indicates that despite the two clusters have differentiated, they continue to be in contact through gene flow. Alternatively, if clusters result from an old-standing complete barrier to gene flow, the inter-cluster F_{ij} curve should be independent of distance (horizontal) with values much lower than those of the intra-cluster F_{ij} curves, indicating that clusters are evolving independently.

The kinship coefficient (F_{ij}) was estimated as a correlation coefficient between allelic states, following J. Nason's estimator [96], and calculated with SPAGeDi v.1–5 [91]. To quantify and evaluate the importance of SGS, we estimated the Sp statistic [51], defined as $Sp = -b_{LD}/(1 - F_N)$, where b_{LD} is the observed regression slope of F_{ij} over the logarithmic distance d_{ij} , and F_N is the mean F_{ij} between neighboring individuals (in practice, the F_{ij} between samples separated by < 2 km). The statistical significance of b_{LD} was quantified in SPAGeDi from 10,000 permutations of the rows and columns of the inter-individual distance matrix (Mantel test).

2.6. Morphometric traits and analyses

To assess whether the observed genetic divergence among genetic clusters was also reflected in morphology, we measured and analyzed vegetative and reproductive morphological traits for 43 herbarium specimens from Meise Botanic Garden (BR) (Table S1). These morphological traits are those used in species determination keys [97, 98]. Vegetative traits included (*i*) four qualitative variables: leaf apex shape (obtuse or acuminate), petiole indumentum (glabrous or pubescent), leaf blade shape (oval,

elliptical, or lanceolate), and leaf base shape (cordate, obtuse, or rounded); and (ii) five quantitative variables: petiole length, acumen length, number of secondary vein pairs, length and width of the lamina. Reproductive traits included eight qualitative variables: inflorescence position (axillary or terminal), type of inflorescence (spike, head, or raceme), pedicel indumentum (glabrous or pubescent), inflorescence and pedicel length, fruit stipe length, and fruit length and width. A principal component analysis (PCA) was performed on the quantitative variables to explore patterns of morphological variation. Kruskal-Wallis tests were used to assess differences between genetic clusters for both vegetative and reproductive traits.

3. Results

3.1. Genetic discontinuities within *C. edulis* populations

Bayesian clustering analyses implemented in STRUCTURE showed that the mean likelihood ($L(K)$) of the data increased gradually from $K = 1$ to $K = 4$, before decreasing at $K = 5$ and $K = 6$ with a high variance among runs, and increasing again at $K = 7$, above which a plateau was reached (Fig. 1a). For $K \leq 7$, every inferred cluster had a membership coefficient $q > 0.8$ for at least one individual; for $K > 7$, no additional clusters met this criterion, indicating model saturation. The number of admixed individuals reached 29 at $K = 4$, and 36 at $K = 7$.

At $K = 4$, a first genetic cluster (UG) was confined to Upper Guinean forests in West Africa. The three other genetic clusters were restricted to Lower Guinea and will be called LG_N, LG_W, and LG_S, being distributed in parapatry from North to South in this order (Fig. 1b). LG_N ranged from Nigeria to the coastal area of Gabon, covering Cameroon and Equatorial Guinea. LG_W was restricted to northern and central Gabon. LG_S was present in the southwest of Gabon, Republic of the Congo, and Democratic Republic of the Congo, between 2°S and 6°S latitude, covering the Mayumbe forests (Fig. 2). While UG and LG_S clusters were rarely admixed with other clusters (1 and 3 cases, respectively), about 10.3% of the individuals assigned to the LG_N and/or LG_W clusters were admixed.

At $K = 7$, we recovered the UG and LG_S clusters as described above, and the LG_W cluster was also largely recovered but slightly more restricted and less admixed (hereafter, LG_W will be used to denote the cluster as defined under $K = 7$). By contrast, the LG_N cluster was now subdivided into four clusters, three of them called LG_Nw1, LG_Nw2, and LG_Nw3 were situated westwards in parapatry from South to North, and one cluster called LG_Ne was situated eastwards, in allopatry with respect to all the other clusters (Fig. 2). LG_Nw1 appeared already at $K = 5$ (Fig. 1b) and ranges from the southwestern corner of Cameroon (Campo, Ma'an, and surrounding localities) to Equatorial Guinea and northwestern Gabon, with one off-center individual reaching the limit of the Dja Faunal Reserve in Cameroon. LG_Ne first appeared at $K = 6$ (Fig. 1b) with a narrow and isolated range in southeastern Cameroon, near the border point between Cameroon, Gabon, and the Republic of the Congo. Finally, LG_Nw2 and LG_Nw3 clusters appeared only at $K = 7$ (Fig. 1b). LG_Nw2 occupies a large range in western Cameroon (localities of Douala, Edéa, Fifinda, Bipindi, Ngovayang, Mbalmayo, and Ebolowa) and includes two off-center

individuals in northeastern Gabon (Fig. 2). LG_Nw3 occurs on either side of the Cameroonian volcanic line and extends into southern Nigeria, represented by a single individual (Fig. 2).

To validate the seven clusters solution, we re-applied separately the same Bayesian clustering approach to the genotypes assigned to UG and then to LG_S. In both cases, no further subdivision was detected, with the optimal number of clusters being $K = 1$. In contrast, when applied to genotypes assigned to LG_N and/or LG_W under $K = 4$, the analysis revealed five clusters which corresponded precisely to the clusters described above at $K = 7$. Moreover, when applying the maximum likelihood algorithm implemented in 'snapclust', the optimal number of clusters was also seven and corresponded very closely to the ones described above.

The 36 admixed genotypes under $K = 7$ (9.3% of all 385 samples) were mostly found in western Cameroon, Equatorial Guinea, and western Gabon. The percentage of admixed samples was highest between LG_Nw1, LG_W, and LG_Nw2 (2–8.3% per pair of clusters; Table 2). Only three samples (2.4%) were admixed between LG_S and LG_W, one (0.9%) between LG_Nw1 and LG_Ne, two (1.5%) between LG_Nw2 and LG_Ne, and one (0.8%) between LG_Nw3 and LG_Nw2. Most of these samples occurred in the contact areas of the respective genetic clusters. No admixed sample was found between UG and any other cluster.

Figure 1 Results of the Bayesian clustering analysis performed with STRUCTURE [99] on 385 *Coula edulis* samples genotyped at 21 microsatellite loci. a: Posterior log-likelihood of the data according to the number K of genetic clusters: mean (circle) $\pm$ standard deviation (vertical bars) over 10 repetitions for each K ; b: Histograms of the individual admixture proportions (q values) to genetic clusters for $K = 4$ to 7, considering the respective repetitions with highest log-likelihood, with indication of their geographic distributions: Upper Guinea (UG) or Lower Guinea (LG)(see Fig. 2). Each cluster is represented by a different color.

Table 1

Parameters of genetic diversity of seven inferred genetic clusters in populations of *C. edulis*.

Genetic clusters	<i>N</i>	<i>A_o</i>	<i>R_s</i> (24)	<i>H_e</i>	<i>F_i</i>	<i>Sp</i> ($\pm$ SE)
UG	43	4.59	3.34	0.38 ^c	0.16 (0.09 – .0.24)	0.057 $\pm$ 0.007
LG_Nw3	16	3.35	3.24	0.46 ^{bc}	0.18 (0.08–0.28)	0.029 $\pm$ 0.006
LG_Nw2	98	11.06	6.38	0.71 ^a	0.06 (0.00–0.13)	0.024 $\pm$ 0.002
LG_Nw1	68	10.24	6.61	0.70 ^a	0.02 (0.01–0.04)	0.008 $\pm$ 0.001
LG_Ne	23	3.18	2.77	0.40 ^c	0.04 (0.00–0.11)	0.004 $\pm$ 0.002
LG_W	68	9.71	5.64	0.66 ^{ab}	0.12 (0.04–0.18)	0.036 $\pm$ 0.004
LG_S	33	6.53	4.93	0.50 ^{abc}	0.13 (0.09–0.17)	0.061 $\pm$ 0.008
All genetic clusters	349	16.76	8.42	0.83	0.23 (0.21–0.26)	0.088 $\pm$ 0.005

N: number of individuals analyzed, *A_o*: number of alleles, *R_s* (*k* = 24) allelic richness or number of alleles among 24 gene copies, *H_e*: expected heterozygosity, *F_i*: inbreeding coefficient estimated by INEst accounting for null alleles (95% posterior range), *Sp*: statistic used to quantify the decay of kinship-distance curves within genetic clusters, *SE*: standard error. Values sharing the same letter are not significantly different (*P* > 0.05).

Figure 2 Distribution of 385 *Coula edulis* samples and their corresponding genetic cluster assignments as inferred with STRUCTURE under *K* = 7. Country codes are as follows: SIL: Sierra Leone; LBR: Liberia; CIV: Ivory Coast; GHA: Ghana; NGA: Nigeria; CMR: Cameroon; GNQ: Equatorial Guinea; GAB: Gabon; COG: Republic of the Congo; COD: Democratic Republic of the Congo. The dotted line delimits approximately the previously known distribution of *Coula edulis* according to GBIF, CJBG, and World Flora online databases, after excluding individuals observed in botanical gardens. The potential rainforest distribution (grey area) is based on [100].

3.2. Genetic diversity and differentiation across genetic clusters

Genetic diversity indices varied significantly across genetic clusters (Table 1), being highest for LG_Nw1, LG_Nw2 and LG_W (*R_s* = 5.64 to 6.61; *H_e* = 0.66 to 0.71), lowest for LG_Ne, LG_Nw3 and UG (*R_s* = 2.77 to 3.34; *H_e* = 0.38 to 0.46), and intermediate for LG_S (*R_s* = 4.93; *H_e* = 0.50). The inbreeding coefficients were close to zero (*F_i* = 0.02 to 0.06) for LG_Ne, LG_Nw1, and LG_Nw2 clusters, but significantly higher than zero for the other ones (*F_i* = 0.12 to 0.18; Table 1).

In terms of genetic differentiation, pairwise F_{ST} values between genetic clusters ranged widely, from 0.08 to 0.59. The clusters most differentiated from all the other ones were UG ($F_{ST} \geq 0.41$), LG_S ($F_{ST} \geq 0.34$), LG_Nw3 ($F_{ST} \geq 0.21$), LG_Ne ($F_{ST} \geq 0.20$), and LG_W ($F_{ST} \geq 0.17$). The least differentiated pair of clusters were LG_Nw1 and LG_Nw2 ($F_{ST} = 0.08$). In general, geographically adjacent genetic clusters had an $F_{ST} \leq 0.21$, except LG_W and LG_S, which are parapatric at about 2°S in Gabon but remain well differentiated ($F_{ST} = 0.35$; Table 2).

Pairwise R_{ST} values ranged from 0.10 to 0.86 and followed the same trends as pairwise F_{ST} (Table 2). Permutation tests revealed phylogeographic signals (i.e., R_{ST} significantly higher than F_{ST}) for pairs of clusters involving UG, LG_S (except with LG_Ne), and LG_W (except with LG_Ne), while there was no phylogeographic signal between the four clusters from northern LG (Table 2).

Table 2
Genetic differentiation estimates (F_{ST} below diagonal, R_{ST} above diagonal) and percentages of admixed individuals between the seven inferred genetic clusters in *Coula edulis*

Genetic clusters	UG	LG_Nw3	LG_Nw2	LG_Nw1	LG_Ne	LG_W	LG_S
Genetic differentiation F_{ST}/R_{ST}							
UG		0.86***	0.63***	0.66**	0.76***	0.69***	0.76***
LG_Nw3	0.58		0.14 ^{ns}	0.35 ^{ns}	0.48 ^{ns}	0.57**	0.64**
LG_Nw2	0.39	0.21		0.10 ^{ns}	0.26 ^{ns}	0.29**	0.44*
LG_Nw1	0.41	0.24	0.08		0.38 ^{ns}	0.22*	0.54**
LG_Ne	0.59	0.50	0.20	0.29		0.47 ^{ns}	0.59 ^{ns}
LG_W	0.41	0.35	0.22	0.17	0.39		0.55*
LG_S	0.56	0.48	0.34	0.35	0.53	0.35	
Percentage of admixed individuals (%)							
UG		0.00	0.00	0.00	0.00	0.00	0.00
LG_Nw3			0.80	0.00	0.00	0.00	0.00
LG_Nw2				4.00	1.50	2.00	0.00
LG_Nw1					0.90	8.30	0.00
LG_Ne						0.00	0.00
LG_W							2.40
LG_S							

* indicates that allele size permutation tests detected a significant shift in allele size between genetic clusters ($R_{ST} > F_{ST}$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; *ns*: not significant.

3.3. Characterization of spatial genetic structure

The spatial genetic structure (SGS) of *C. edulis* throughout its distribution range, showed an isolation-by-distance (IBD) pattern: the kinship coefficient (F_{ij}) between pairs of individuals decreases nearly linearly with the logarithm of the distance, from a mean value of 0.25 between individuals separated by less than 2 km, reaching negative values between individuals separated by more than about 440 km (Fig. 3a). The Sp statistic was high and significant ($Sp = 0.088 \pm 0.005$), indicating strong SGS at a large scale.

At the level of each genetic cluster, the kinship-distance curves also showed a rapid decline (Fig. 3a, where results are given only for clusters represented by ≥ 30 samples). These results were associated with statistically significant Sp values (Table 1). The SGS was strong within the UG and LG_S clusters ($Sp = 0.057$ to 0.061), intermediate within LG_W, LG_Nw2, and LG_Nw3 ($Sp = 0.024$ to 0.036), and weak within LG_Nw1 and LG_Ne ($Sp = 0.004$ to 0.008), paralleling approximately the trends observed for the inbreeding coefficients (F_i).

The kinship-distance curves for pairs of samples belonging to distinct and parapatric clusters (Figs. 3b, c, d) showed clear IBD patterns between LG_Nw1 and LG_Nw2 ($b_{LD} = -0.022 \pm 0.008$; Fig. 3b), and between LG_Nw1 and LG_W ($b_{LD} = -0.039 \pm 0.018$; Fig. 3c). By contrast, no significant IBD pattern was detected between LG_W and LG_S ($b_{LD} = -0.016 \pm 0.025$; Fig. 3d), suggesting that these clusters have independent genetic structures.

Figure 3 Kinship-distance curves for pairs of individuals assigned to a cluster ($q \geq 0.8$) according to $\log(\text{distance})$. a: within well-sampled genetic clusters (UG, LG_Nw2, LG_Nw1, LG_W, and LG_S) and across the whole distribution ranges ('All genetic clusters'); b: between LG_Nw1 and LG_Nw2; c: between LG_Nw1 and LG_W; d: between LG_W and LG_S. For pairs of clusters (panels b, c, and d), two curves represent the SGS within each cluster, and one curve corresponds to the kinship between samples from distinct clusters, representing the inter-cluster SGS dependency.

3.4. Divergence in morphological traits

We analyzed vegetative traits on 43 herbarium specimens, which we grouped according to their geographic origins into UG (21 samples from Upper Guinea), LG_N + W (17 samples from Lower Guinea north of 2°S latitude), and LG_S (5 samples from Lower Guinea south of 2°S latitude). We did not observe diagnostic qualitative traits, but among the five quantitative traits measured, we observed a high variability in the number of pairs of secondary veins, which was higher in LG_N + W samples than in UG or LG_S (Kruskal-Wallis test: $P = 0.016$; Fig. 4a).

A principal component analysis (PCA) using both reproductive (inflorescence and pedicel lengths) and vegetative (number of pairs of secondary veins) traits on 22 specimens (but where LG_S was represented by only two samples), separated most of the UG and the LG samples along the first principal

component, accounting for 60% of the total variance (Fig. 4b). Kruskal-Wallis tests conducted between the LG_N + W and UG groups revealed significant differences for the number of pairs of secondary veins ($P = 0.003$), the length of the inflorescence ($P = 0.0003$), and the length of the pedicel ($P = 0.001$).

As only nine herbarium samples bore fruits, we did not include fruit-related traits in the morphometric analysis.

Figure 4 Morphological variation observed in herbarium specimens from three different regions corresponding to distinct genetic clusters. a: Boxplots of the number of pairs of secondary veins in 43 specimens (17 from LG_N + W, 5 from LG_S, 21 from UG). Values sharing the same letter are not significantly different ($P > 0.05$); b: Ordination according to the two first components of a PCA, explaining 59.7% and 20.9% of the variation expressed in the number of pairs of secondary veins, inflorescence length, and pedicel length, measured in 22 specimens. LG_N + W: include genetic clusters LG_Nw1, LG_Nw2, LG_Nw3, and LG_W.

4. Discussion

In this study, we used microsatellites (SSRs) to characterize the phylogeographic pattern of *Coula edulis*, allowing comparison with patterns observed in other African rainforest tree species. We expected similar responses that would lead to congruent phylogeographic patterns across species, considering that past climatic changes have influenced vegetation cover and may have left common imprints on the distribution of genetic diversity. *Coula edulis* being an outcrossing species [77], we expected a low level of inbreeding within population. However, its limited seed and pollen dispersal capacities may lead to biparental inbreeding. Consistent with this, we observe significant inbreeding ($F_i > 0$) within several genetic clusters, especially in those covering broad areas and displaying substantial SGS (high Sp), such as in UG, LG_W, LG_S, and LG_Nw3.

4.1. Patterns of genetic discontinuity and intra/inter-population variability in *Coula edulis*

Our results revealed strong and weaker genetic discontinuities in *C. edulis* across its distribution range. We identified two well differentiated and rarely admixed genetic clusters at each extremity of the distribution range (UG and LG_S clusters), and five genetic clusters with various levels of genetic differentiation and admixture ranging from Nigeria to Gabon, distributed in parapatry (LG_Nw3, LG_Nw2, LG_Nw1, and LG_W) or allopatry (LG_Ne).

The observed genetic clusters indicate genetic discontinuities, as reported in many studies of Guineo-Congolian tree species [6, 8, 9, 14, 35, 36, 42, 55, 101, 102]. Genetic clusters at putatively neutral SSR markers can reflect (i) current barriers to gene flow, (ii) the presence of ancient barriers to gene flow due to past population fragmentation, or (iii) founder effects associated with recent colonization events in regions. The UG cluster can be explained by the Dahomey Gap, a savanna corridor isolating rainforests

from UG and LG (Fig. 2), but within LG, there are currently no obvious barriers to gene flow for *C. edulis*, whose distribution seems spatially continuous, except for the isolated LG_Ne population. However, genetic clusters can also arise in the absence of actual genetic discontinuities under isolation by distance, and appear well differentiated and little admixed under a spatially discontinuous sampling [13]. This explanation must be considered in *C. edulis* because its seed and pollen dispersal occurs over short distances [77], leading to substantial isolation by distance at all scales (Fig. 3). Our sampling covered well the whole distribution range of the species, except in Nigeria (one sample) and Sierra Leone (no sample), and we kept no more than two individuals per km² (by subsampling the initial collection) to avoid sampling closely related individuals that could form a cluster. Moreover, we controlled the effect of isolation by distance using kinship-distance curves, and when applied on parapatric pairs of genetic clusters, the much higher kinship within clusters than between clusters at similar spatial distances confirmed the presence of actual genetic discontinuities (Figs. 3b,c,d). Therefore, genetic discontinuities in LG can be interpreted as evidence of past population fragmentation, perhaps due to Pleistocene climatic fluctuations or even earlier events [103], but also as recent fragmentation or founder events [35, 36].

4.1.1. Genetic homogeneity and low diversity within UG, and strong differentiation between UG and LG populations

In West Africa, our results revealed the presence of a single genetic cluster, which we called UG, suggesting a relatively high genetic homogeneity over a large area, which is nevertheless subject to substantial isolation by distance (Fig. 3a; $S_p = 0.057$). This cluster was characterized by low genetic diversity ($H_e = 0.38$) compared to clusters in Lower Guinea (Table 1). Similar results have been reported in other African forest tree species, such as *Pentadesma butyracea* [104], *Milicia excelsa* [105], and *Distemonanthus benthamianus* [36]. These findings support the hypothesis that the West African rainforest experienced severe contraction during Pleistocene climatic fluctuations. This may have led to geographic isolation and demographic bottleneck of tree species populations, resulting in increased genetic drift, thus explaining the low genetic diversity [4, 36, 54, 104, 105].

However, our results contrast with those of [35], who reported a much higher value of $H_e = 0.70$ in *Terminalia superba*. This difference suggests that *T. superba* maintained a relatively large effective population size in UG during past climatic changes, possibly because it is a light-demanding species well adapted to disturbed forests, while *C. edulis* is a shade-tolerant species found in old-growth forests.

The high levels of genetic differentiation ($F_{ST} = 0.39–0.59$) between UG and LG populations, which are currently isolated by a savanna corridor called the Dahomey gap, are comparable to those found in *Guibourtia ehie* (F_{ST} values ranging from 0.47 to 0.55; [101]), but which probably correspond to populations of distinct *Guibourtia* species (Tosso and Wieringa, pers. comm.). In contrast, lower F_{ST} values between UG and LG populations have been reported in other African tree species such as *T. superba* ($F_{ST} = 0.07$; [35]), *Erythrophleum ivorense* ($F_{ST} = 0.09$; [11]), *M. excelsa* ($F_{ST} = 0.20$; [106]), *P.*

butyracea ($F_{ST} = 0.26$; [105]), *D. benthamianus* ($F_{ST} = 0.35$; [36]), and *Entandrophragma cylindricum* ($F_{ST} = 0.07$ to 0.12 ; [57]).

The high F_{ST} values associated to even higher R_{ST} values between UG and LG clusters (Table 2) indicate ancient divergence, allowing for the accumulation of mutations and the emergence of a phylogeographic signal. At SSR markers, $R_{ST} > F_{ST}$ is expected between isolated populations when the number of generations since divergence is higher than the reverse of the mutation rate [95]. Similar phylogeographic signals at SSR markers have been reported in other tree species such as *E. cylindricum* and *E. candollei* [56], *G. ehie* [101], and *E. ivorenses* [9]. Hence, assuming a mutation rate in the $\mu = 10^{-3}$ – 10^{-4} range [106] and a generation time of 100–200 years for rainforest trees, UG and LG populations could have diverged for at least 0.1 to 2 million years. This biogeographic separation between UG and LG is also reflected in patterns of plant and animal diversity [107–109]. Whole DNA sequence (plastome) data could help estimate the divergence time between *C. edulis* UG and LG populations. By analyzing substitution rates and molecular clocks from plastome sequences, researchers can reconstruct evolutionary histories and infer the timing of population splits, offering deeper insight into historical biogeographic processes [11].

4.1.2. Genetic heterogeneity and high variability in genetic diversity and differentiation within LG

In Lower Guinea (LG), we identified six geographically coherent genetic clusters in *Coula edulis*. In addition, our analyses revealed a high level of variability in both genetic diversity and population differentiation (Tables 1 and 2), a pattern rarely documented in other studies. This substantial variation highlights the complexity of the evolutionary dynamics affecting *C. edulis* populations in this region. These results also support the hypothesis that the Lower Guinea region is the most genetically diverse Guineo-Congolian sub-center, both floristically [107, 110, 111] and at the intraspecific level in terms of genetic diversity.

Deep divergence between LG_N, LG_W and LG_S

In this study, two north-south genetic discontinuities associated to significant phylogeographic signals ($R_{ST} > F_{ST}$) were detected within LG (Fig. 2): (i) one across 1° N latitude, (ii) another across 2° S latitude.

At the level of 1° N, genetic discontinuities between northern (Cameroon, Nigeria, Equatorial Guinea) and southern (Gabon, Republic of the Congo, and Democratic Republic of the Congo) populations have been frequently reported in other African tree species [12–14, 35, 36, 55]. Our results are consistent with those reviewed by [9], who identified similar discontinuities in eight out of nine Guineo-Congolian tree species, typically located between 0.5° and 2.5° North latitude.

The high F_{ST} values (ranging from 0.17 to 0.35) observed across the 1° N discontinuity between LG_W and the LG_N clusters tend to exceed those reported in other African tree species, such as *B. toxisperma* ($F_{ST} = 0.14$ to 0.18 ; [56]), *Scorodophloeus zenkeri* ($F_{ST} = 0.12$ to 0.24 ; [5]), and *Staudtia kamerunensis*

($F_{ST} = 0.13$; [6]). These high levels of F_{ST} values were associated with significantly higher R_{ST} values, indicating that the divergence time has been sufficiently long for stepwise mutations to accumulate and generate a phylogeographic signal [95]. This pattern suggests that ancient fragmentation events isolated surviving *C. edulis* populations for a long time, before being reconnected nowadays.

The spatial genetic structure (SGS) between the parapatric LG_Nw1 and LG_W clusters showed a statistically significant isolation-by-distance (IBD) effect for the inter-cluster kinship-distance curve (Fig. 3c). This suggests that although these populations were historically separated, their current distribution results from secondary contact and ongoing gene flow for sufficient time. Indeed, we identified 21 admixed individuals between LG_Nw1 and LG_W (8.3% admixture, Table 2), which are gradually eroding the historical signal of fragmentation over time. This erosion is probably a very slow process given the very limited seed and pollen dispersal capacities and the putatively long generation time of *C. edulis*, characterized by a slow annual growth rate [74]. A rough estimate of generation time could be obtained by dividing the mean diameter of adults by the mean annual dbh growth rate.

At the level of 2°S, the genetic discontinuity was detected between LG_W, which includes only individuals from central and northern Gabon, and LG_S, which includes individuals from southwestern Gabon, the Republic of the Congo, and the Democratic Republic of the Congo (Fig. 2). A north-south discontinuity around this latitude was less commonly reported in other African tree species, perhaps because fewer phylogeographic studies have had sufficient sampling south of 2°S [111], but it was also found in *Terminalia superba* [35], *Scorodophloeus zenkeri* [13], and *Staudtia kamerunensis* [112].

The high F_{ST} values associated with even higher R_{ST} values between LG_W and LG_S clusters (Table 2) indicate ancient divergence, allowing for the accumulation of mutations and the emergence of a phylogeographic signal [95]. In addition, spatial genetic structure (SGS) analyses between these two genetic clusters show no signal of isolation by distance (IBD) (Fig. 3d), indicating that these populations have evolved independently. This pattern suggests a relatively recent secondary contact between LG_W and LG_S, as evidenced by our finding of three admixed individuals between them (Table 2; Fig. 2), or the development of a reproductive barrier between these clusters. The latter hypothesis of an incipient speciation process would be worth testing by conducting population genetics investigations across the contact zone.

The wide range of genetic diversity ($H_e = 0.40\text{--}0.71$) observed in LG genetic clusters may be attributed to the presence of multiple forest refugia in the LG region, which likely helped maintain genetic diversity within stable populations during periods of climatic disturbance [14, 37]. Additionally, LG is characterized by a heterogeneous landscape (variation in elevation and savanna–forest mosaics), which reinforces differentiation among populations isolated in distinct forest refugia and contributes to the observed genetic diversity [37, 113, 114].

More recent divergence within LG_N

West-East genetic discontinuity was identified within LG_N, separating the LG_Ne cluster from the three western LG_N clusters. LG_Ne is particularly distinctive due to its narrow and disjoint distribution. Located in a region where the species has not previously been recorded in either the African Plant Database (<https://africanplantdatabase.ch>) or GBIF (Fig. 2), this population was known from a few Cameroonian field botanists [115]. Given its location in southern Cameroon, close the border with Gabon and the Republic of the Congo, we cannot exclude that this *C. edulis* population extends into those countries, in regions yet little explored botanically [111]. Despite the high genetic differentiation values between LG_Ne and the other LG_N clusters (F_{ST} from 0.20 to 0.50), the divergence is probably relatively recent given the absence of a phylogeographic signal (R_{ST} not significantly different from F_{ST}). Therefore, the high F_{ST} values probably result from recent colonization followed by strong genetic drift affecting the LG_Ne cluster, as supported by its low gene diversity and allelic richness compared to the other LG clusters. Two scenarios could explain the isolation and small range of the LG_Ne population: (i) a founder event after exceptional long-distance dispersal, (ii) a population fragmentation if *C. edulis* had temporarily extended its range eastwards in the past before retracting, in which case LG_Ne would represent a relict population. Although *C. edulis* is known to be used by local communities, we are not aware of any evidence that it has been planted in Central Africa. However, the possibility of human-mediated dispersal cannot be entirely excluded.

West–East genetic discontinuities in LG have also been observed in other tree species, including *Terminalia superba* [35], *Distemonanthus benthamianus* [36], *Greenwayodendron suaveolens* [12], and *Staudtia kamerunensis* [14]. However, in all these examples, both eastern and western genetic clusters occupy large areas, are not disconnected, and are much less differentiated (lower F_{ST} and R_{ST}) than reported for *C. edulis*.

In western LG_N, despite the continuous distribution of the species, two main genetic discontinuities separate three genetic clusters: LG_Nw1, LG_Nw2, and LG_Nw3, with F_{ST} values ranging from 0.08 to 0.24. The lower F_{ST} values and absence of phylogeographic signals suggest more recent divergence and/or a homogenization effect due to ongoing gene flow, as supported by the relatively high percentage of admixed individuals (Figs. 2, 3b; Table 2). Populations close to the volcanic line (LVC), like LG_Nw3, are often genetically original [50, 103].

Our results contradict the hypothesis proposed by [10], which suggested that *Coula edulis* populations expanded across a broad geographic area without accumulating mutations. Such a rapid demographic expansion, potentially linked to human-mediated dispersal, might indeed leave limited time for mutations to accumulate. However, the discrepancy between our findings and theirs likely stems from differences in marker types and sampling resolution: their analysis was based solely on a chloroplast DNA markers, which probably lacked the polymorphism necessary to reveal recent or fine-scale population structure.

4.2. Divergence in morphological traits

Analysis of morphological traits from *Coula edulis* herbarium specimens revealed notable variability between individuals from the Upper Guinean (UG) and Lower Guinean (LG) genetic clusters. Specifically, differences in the number of pairs of secondary veins, inflorescence length, and pedicel length support the existence of morphologically distinct populations, consistent with molecular results. In contrast, other morphological traits did not show significant variation between clusters, and we did not find any qualitative diagnostic trait. While the observed differentiation may be interpreted as evidence of structured populations, the possibility of incipient speciation or cryptic species cannot be ruled out. Nevertheless, caution is warranted, as this study represents the first to integrate both morphological and molecular data in *C. edulis*, and some genetically distinct clusters, such as LG_Ne, LG_Nw3, and LG_S, were underrepresented in herbarium samples.

The morphological descriptions by authors, such as [97] for West Africa and [98] or Central Africa, converge in recognizing *Coula edulis* as a single species. However, there are multiple recent examples of cryptic species discovered among African trees using an integrative approach combining molecular and morphological data [47, 50] highlighting the limitation of morphology alone in species delimitation. Therefore, a more comprehensive analysis is needed, incorporating a larger number of herbarium specimens across the full distribution range of *C. edulis*. It would also be valuable to integrate field-based observations of traits, such as the color of immature and mature fruits, flowers, bark texture, and growth form, to improve taxonomic resolution. Given the economic importance of the species for its edible nuts, it would also be valuable to investigate whether differences exist among genetic clusters in key fruit and seed traits, such as size, nutritional and organoleptic characteristics, and germination capacity.

5. Conclusions and perspectives

Microsatellite data revealed genetic discontinuities along the distribution range of *Coula edulis* Baill., a tree species of high economic importance for local communities. Our results provide baseline information for the conservation of the species' genetic resources and its potential domestication. For future research, we recommend the use of whole-genome data (e.g., plastomes), which would allow for precise dating of population divergence and provide a more detailed picture of the evolutionary history of the African tropical forest.

Abbreviations

SSR
Simple Sequence Repeat
MCDL
Markov Chain Monte Carlonn.

Declarations

Acknowledgments

We would like to express our sincere gratitude to the Ministry of Forests and Wildlife of Cameroon for granting research authorization N°. 4025/L/MINFOF/SETAT/SG/DAG/SDPSP/SP/CBF0RM/, which enabled data collection with the support of the conservation staff of Campo Ma'an National Park. The fieldwork in the Campo Ma'an National Park was supported in part by the Congo Basin Institute's Ebony Project, funded by Bob and Cindy Taylor. We thank Drs. Vincent Droissart, Gilles Dauby, Pierre Couteron, Thomas Couvreur, Nicolas Barbier, and Pierre Ploton for their valuable collaboration.

Funding

The laboratory work was funded by the Fonds de la Recherche Scientifique, F.R.S-FNRS, through grants PDR-WISD X.3040.17 and PDR T.0119.20, and by a cooperation grant of the Université Libre de Bruxelles, which provided a PhD fellowship to N.G.K.

Authors contributions

N.G.K., B.S., and O.J.H. conceived the research; all authors, except S.S. and V.D., contributed to sampling. N.G.K. and S.S. performed the genotyping. N.G.K. and O.J.H. conducted data analyses. N.G.K. wrote the first draft, and all authors contributed to the final version of the manuscript.

Data availability statement

The nSSR and morphometric traits data are available in the supplementary material Table S1

Ethical approval

Samples from Cameroon (Mbalmayo, Bideem) were collected with a research permit granted by MINRESI (000102/MINRESI/B00/C00/C10/C13). In Gabon, a research permit was granted by CENAREST (n°AR0034/19/MESRSTT/CENAREST/CG/CST/ CSAR). In Liberia, authorization was granted by the FDA (ref: MD/04/2016/-2).

Conflict of interest

The authors declare that they have no conflict of interest.

Consent for publication

Not applicable

References

1. Leigh J, Giles E, Davidar P, Dick CW, Terborgh J, Steege H, et al. Why Do Some Tropical Forests Have so Many Species of Trees ? *Biotropica*. 2004;36:447–73.

2. Ahossou OD, Dainou K, Janssens SB, Triest L, Hardy OJ. Species delimitation and phylogeography of African tree populations of the genus *Parkia* (Fabaceae). *Tree Genet Genomes*. 2020;16.
3. Bouka GUD, Doumenge C, Ekué MRM, Duminil J, Florence J, Degen B et al. Genetic differentiation in *Khaya ivorensis* A. Chev., a threaten tree of evergreen African rainforests. *Tree Genet Genomes*. 2024;20.
4. Budde KB, González-Martínez SC, Hardy OJ, Heuertz M. The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. *Heredity (Edinb)*. 2013;111:66–76.
5. Dauby G, Duminil J, Heuertz M, Hardy OJ. Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Trop Plant Biol*. 2010;3:4–13.
6. Debout GDG, Doucet JL, Hardy OJ. Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity (Edinb)*. 2011;106:88–99.
7. Demenou BB, Migliore J, Heuertz M, Monthe FK, Ojeda DI, Wieringa JJ, et al. Plastome phylogeography in two African rain forest legume trees reveals that Dahomey Gap populations originate from the Cameroon volcanic line. *Mol Phylogenet Evol*. 2020;150:106854.
8. Duminil J, Brown RP, Ewédjè EEB, Mardulyn P, Doucet JL, Hardy OJ. Large-scale pattern of genetic differentiation within African rainforest trees: Insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp (Fabaceae). *BMC Evol Biol*. 2013;13:13.
9. Hardy OJ, Born C, Budde K, Dainou K, Dauby G, Duminil J, et al. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geosci*. 2013;345:284–96.
10. Heuertz M, Duminil J, Dauby G, Savolainen V, Hardy OJ. Comparative phylogeography in rainforest trees from lower Guinea, Africa. *PLoS ONE*. 2014;9.
11. Matvijev K, Dellicour S, Kaymak E, Hardy OJ. Spatially explicit phylogeographical reconstruction sheds light on the history of the forest cover in the Congo Basin. *J Biogeogr*. 2022;49:2256–68.
12. Piñeiro R, Dauby G, Kaymak E, Hardy OJ. Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the ice ages. *Proc R Soc B Biol Sci*. 2017;284:1–9.
13. Vanden Abeele S, Janssens SB, Piñeiro R, Hardy OJ. Evidence of past forest fragmentation in the Congo Basin from the phylogeography of a shade-tolerant tree with limited seed dispersal: *Scorodophloeus zenkeri* (Fabaceae, Detarioideae). *BMC Ecol Evol*. 2021;21:50.
14. Vanden Abeele S, Matvijev K, Hardy OJ, Assumani D-M, Angoboy Ilondea B, Beeckman H, et al. Genetic breaks caused by ancient forest fragmentation: phylogeography of *Staudtia kamerunensis* (Myristicaceae) reveals distinct clusters in the Congo Basin. *Tree Genet Genomes*. 2023;19:27.
15. White F. The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. *Nat Resour Res*. 1983;20:1–356.

16. Slik JWF, Arroyo-Rodríguez V, Aiba SI, Alvarez-Loayza P, Alves LF, Ashton P, et al. An estimate of the number of tropical tree species. *Proc Natl Acad Sci U S A*. 2015;112:7472–7.
17. Couvreur TLP. Odd man out: why are there fewer plant species in African rain forests? *Plant Syst Evol*. 2015;301:1299–313.
18. Robbrecht E. Geography of African Rubiaceae with reference to glacial rain forest refuges. *The Biodiversity of African Plants*. Dordrecht: Springer Netherlands; 1996. pp. 564–81.
19. Tchouto MGP, De Wilde JJFE, De Boer WF, Van Der Maesen LJG, Cleef AM. Bio-indicator species and Central African rain forest refuges in the Campo-Ma'an area. *Cameroon Syst Biodivers*. 2009;7:21–31.
20. Aleman JC, Jarzyna MA, Staver AC. Forest extent and deforestation in tropical Africa since. *Nat Ecol Evol*. 2015;2:26–33.
21. Vancutsem C, Achard F, Pekel JF, Vieilledent G, Carboni S, Simonetti D, et al. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci Adv*. 2021;7:1–21.
22. Heaney LR. A synopsis of climatic and vegetational change in Southeast Asia. *Trop Clim*. 1991;:53–61.
23. Flenley JR. Tropical forests under the climates of the last 30,000 years. *Clim Change*. 1998;39:177–97.
24. Mayle FE, Burbridge R, Killeen TJ. Millennial-scale dynamics of southern Amazonian rain forests. *Sci (80-)*. 2000;290:2291–4.
25. Anhuf D, Ledru MP, Behling H, Da Cruz FW, Cordeiro RC, Van der Hammen T, et al. Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2006;239:510–27.
26. Maley J. Fragmentation de la Forêt Dense Humide ouest-africaine et extension d'une végétation montagnarde à basse altitude au Quaternaire récent: implications paléoclimatiques et biogéographiques. *Géodynamique*. 1987;2:127–31.
27. Maley J. A Catastrophic Destruction of African Forests about 2,500 Years Ago Still Exerts a Major Influence on Present Vegetation Formations. *IDS Bull*. 2002;33:13–30.
28. Maley J, Brenac P. Vegetation dynamics, palaeoenvironments and climatic changes in the forests of western Cameroon during the last 28,000 years B.P. *Rev Palaeobot Palynol*. 1998;99:157–87.
29. Ngomanda A, Neumann K, Schweizer A, Maley J. Seasonality change and the third millennium BP rainforest crisis in southern Cameroon (Central Africa). *Quat Res*. 2009;71:307–18.
30. Morley R. *Origin and evolution of tropical rain forest*. John Wiley Sons New York. 2000.
31. Elenga H, Schwartz D, Vincens A. Pollen evidence of late Quaternary vegetation and inferred climate changes in Congo. *Palaeogeogr Palaeoclimatol Palaeoecol*. 1994;109:345–56.
32. Elenga H, Schwartz D, Vincens A. Changements climatiques et action anthropique sur le littoral congolais au cours de l'Holocène. *Bull - Soc Geol Fr*. 1992;163:83–90.

33. Maley J. The African rainforest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proc R Soc Edinb.* 1996;104B:31–73.
34. Koffi G. Etude de la diversité génétique et de la phylogéographie de *Santiria trimera* (Burseraceae), implications pour une conservation durable des forêts humides d'Afrique. PhD Thesis. Brussels: Université Libre de Bruxelles. 2010.
35. Demenou BB, Doucet J-L, Hardy OJ. History of the fragmentation of the African rain forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*. *Heredity (Edinb).* 2018;120:547–61.
36. Demenou BB, Piñeiro R, Hardy OJ. Origin and history of the Dahomey Gap separating West and Central African rainforests: Insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. *J Biogeogr.* 2016;43:1020–31.
37. Dauby G, Duminil J, Heuertz M, Koffi GK, Stévant T, Hardy OJ. Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Mol Ecol.* 2014;23:2299–312.
38. Dupont LM, Donner B, Schneider R, Wefer G. Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. *Geology.* 2001;29:195–8.
39. Miller CS, Gosling WD. Quaternary forest associations in lowland tropical West Africa. *Quat Sci Rev.* 2014;84:7–25.
40. Couvreur TLP, Porter-Morgan H, Wieringa JJ, Chatrou LW. Little ecological divergence associated with speciation in two African rain forest tree genera. *BMC Evol Biol.* 2011;11:19.
41. Plana V. Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *R Soc.* 2004;359:1585–94.
42. Duminil J, Mona S, Mardulyn P, Doumenge C, Walmacq F, Doucet J-L, et al. Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. *J Biogeogr.* 2015;42:1443–54.
43. Ley AC, Heuertz M, Hardy OJ. The evolutionary history of central African rain forest plants: phylogeographical insights from sister species in the climber genus *Haumania* (Marantaceae). *J Biogeogr.* 2017;44:308–21.
44. Helmstetter AJ, Béthune K, Kamdem NG, Sonké B, Couvreur TLP. Individualistic evolutionary responses of Central African rain forest plants to Pleistocene climatic fluctuations. *Proc Natl Acad Sci.* 2020;117:32509–18.
45. Satler JD, Carstens BC. Do ecological communities disperse across biogeographic barriers as a unit? *Mol Ecol.* 2017;26:3533–45.
46. Harvey MG, Seeholzer GF, Smith BT, Rabosky DL, Cuervo AM, Brumfield RT. Positive association between population genetic differentiation and speciation rates in New World birds. *Proc Natl Acad Sci U S A.* 2017;114:6328–33.
47. Bouka UG, Florence J, Doumenge C, Loumeto JJ, McKey D. *Khayae* (Meliaceae) specierum Nomenclator. *Adansonia.* 2017;39:15–30.

48. Donkpegan ASL, Doucet J-L, Migliore J, Duminil J, Dainou K, Piñeiro R, et al. Evolution in African tropical trees displaying ploidy-habitat association: The genus *Afzelia* (Leguminosae). *Mol Phylogenet Evol.* 2017;107:270–81.
49. Ikabanga DU, Stévant T, Koffi KG, Kamení Monthe F, Nzigou Doubindou EC, Dauby G, et al. Combining morphology and population genetic analysis uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae). *Phytotaxa.* 2017;321:166.
50. Lissambou BJ, Couvreur TLP, Atteke C, Stévant T, Piñeiro R, Dauby G, et al. Species delimitation in the genus *Greenwayodendron* based on morphological and genetic markers reveals new species. *Taxon.* 2019;68:442–54.
51. Vekemans X, Hardy OJ. New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol.* 2004;13:921–35.
52. Barluenga M, Meyer A. Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. *BMC Evol Biol.* 2010;10:20.
53. Kyndt T, Assogbadjo AE, Hardy OJ, Kakaï RG, Sinsin B, Van Damme P, et al. Spatial genetic structuring of baobab (*Adansonia digitata*, Malvaceae) in the traditional agroforestry systems of west africa. *Am J Bot.* 2009;96:950–7.
54. Dainou K. Structuration de la diversité génétique du genre *Milicia* : Taxonomie, phylogéographie, dynamique des populations. PhD Thesis. Gembloux: Université de Liege-Gembloux Agro-Bio Tech. 2012.
55. Ndiade-Bourobou D, Dainou K, Hardy OJ, Doumenge C, Tosso F, Bouvet JM. Revisiting the North-South genetic discontinuity in Central African tree populations: the case of the low-density tree species *Baillonella toxisperma*. *Tree Genet Genomes.* 2020;16.
56. Monthe F. Origine et dynamique de la diversité génétique des arbres Guinéo-Congolais du genre *Entandrophragma* et implications pour une gestion durable. PhD Thesis. Brussels: Université Libre de Bruxelles. 2019.
57. Thomson FJ, Moles AT, Auld TD, Kingsford RT. Seed dispersal distance is more strongly correlated with plant height than with seed mass. *J Ecol.* 2011;99:1299–307.
58. Nathan R, Katul GG. Foliage shedding in deciduous forests lifts up long-distance seed dispersal by wind. *Proc Natl Acad Sci U S A.* 2005;102:8251–6.
59. Corlett RT. Seed dispersal distances and plant migration potential in tropical East Asia. *Biotropica.* 2009;41:592–8.
60. van der Pijl L. Principles of Dispersal in Higher Plants. Berlin, Heidelberg: Springer Berlin Heidelberg; 1972.
61. Ghazoul J. Pollen and seed dispersal among dispersed plants. *Biol Rev.* 2005;80:413–43.
62. Muscarella R, Fleming TH. The role of frugivorous bats in tropical forest succession. *Biol Rev.* 2007;82:573–90.

63. Rimlinger A. Histoire évolutive de l'arbre fruitier *Dacryodes edulis*: implications pour sa conservation et sa gestion durable. PhD Thesis. Montpellier : Université de Montpellier. 2021.
64. Leakey RRB, Atangana AA, Kengni E, Waruhiu AN, Usoro C, Anegbeh PO, et al. Domestication of *Dacryodes edulis* in west and central africa: Characterisation of genetic variation. *Trees Livelihoods*. 2002;12:57–71.
65. Yogom BT, Avana-Tientcheu ML, Monthé FK, Bissiengou P, Loumeto JJ, Zekraoui L et al. Genetic diversity and structure in wild and cultivated populations of an emblematic African tree species, *Garcinia kola* (Clusiaceae). *Tree Genet Genomes*. 2023;19.
66. Vihotogbe R, van den Missihoun BRG, Sinsin AA, Sosef B. Genetic diversity of bitter and sweet African bush mango trees (*Irvingia* spp., Irvingiaceae) in West and Central Africa. *Afr J Biotechnol*. 2015;14:3062–74.
67. Neumann K, Bostoen K, Höhn A, Kahlheber S, Ngomanda A, Tchiengué B. First farmers in the Central African rainforest: A view from southern Cameroon. *Quat Int*. 2012;249:53–62.
68. Oslisly R. The history of human settlement in the middle Ogooué valley (Gabon): implications for the environment. In: Weber B, White L, Vedder A, Naughton-Treves L, editors. *African rain forest ecology and conservation*. New Haven, CT: Yale University Press; 2001. pp. 101–18.
69. Oslisly R, White L. Human impact and environmental exploitation in Gabon and Cameroon during the Holocene. In: Denham T, Iriarte I, Vrydaghs L, editors. *Rethinking agriculture; archaeological and ethnoarchaeological perspectives*. Walnut Creek, CA: Left Coast Press Inc; 2007. pp. 345–58.
70. Frankel OH, Brown AHD, Burdon JJ. *The Conservation of Plant Biodiversity*. Cambridge University Press; 1995.
71. Schoen DJ, Brown AH. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc Natl Acad Sci*. 1993;90:10623–7.
72. Duminil J, Di Michele M. Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosyst*. 2009;143:528–42.
73. Assoko Ndong A. Synthèse des données archéologiques récentes sur le peuplement à l'Holocène de la réserve de faune de la Lopé. *Gabon Anthropologie*. 2002;106:135–58.
74. Moupela C. Ecologie, dynamique des populations et intérêts économiques du noisetier d'Afrique (*Coula edulis* Baill.) au Gabon. PhD Thesis. Gemboux: Université de Liège – Gembloux Agro-Bio Tech. 2013.
75. Oyog Matig O, Ndoeye O, Kengue J, Awono A. Les fruitiers forestiers comestibles du Cameroun. Cotonou, Benin. International Plant Genetic Resources Institute (IPGRI) Regional Office for West and Central Africa; 2006.
76. Kamdem NG, Sergeant S, Vercruysse C, Deblauwe V, Sonké B, Hardy OJ. Development and characterization of nuclear microsatellite markers for the African walnut *Coula edulis* Baill (Coulaceae). *Mol Biol Rep*. 2024;51:438.
77. Kamdem NG, Sonké B, Sergeant S, Deblauwe V, Hardy OJ. Short-distance seed and pollen dispersal in both hunted and intact forests in the lower canopy African rainforest tree, *Coula edulis* Baill.

- (Coulaceae). BMC Ecol Evol. 2025;25:20.
78. Moupela C, Vermeulen C, Daïnou K, Doucet JL. Le noisetier d'Afrique (*Coula edulis* Baill.). Un produit forestier non ligneux méconnu. Biotechnol Agron Soc Environ. 2011;15:485–95.
 79. Kuijt J, Hansen B. Flowering Plants. Cham: Springer International Publishing; 2015.
 80. Vivien J, Faure JJ. Arbres des forêts denses d'Afrique Centrale. 2011.
 81. Moupela C, Doucet J, Daïnou K, Tagg N, Bourland N, Vermeulen C. Dispersal and predation of diaspores of *Coula edulis* Baill. in an evergreen forest of Gabon. Afr J Ecol. 2014;52:88–96.
 82. Boesch C, Boesch H. Mental map in wild chimpanzees: An analysis of hammer transports for nut cracking. Primates. 1984;25:160–70.
 83. Puechmaille SJ. The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. Mol Ecol Resour. 2016;16:608–27.
 84. Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, et al. A multidisciplinary study of archaeological grape seeds. Naturwissenschaften. 2010;97:205–17.
 85. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
 86. Wang J. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. Mol Ecol Resour. 2017;17:981–90.
 87. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61.
 88. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol. 2005;14:2611–20.
 89. Jombart T, Adegnet. A R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5.
 90. R Core Team. A Language and Environment for Statistical Computing. Vienna, Austria. R Found Stat Comput. 2018.
 91. Hardy OJ, Vekemans X. Spagedi: a Versatile Computer Program To Analyse Spatial. Mol Ecol Notes. 2002;:618–20.
 92. Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. J Hered. 2009;100:106–13.
 93. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution (N Y). 1984;38:1358.
 94. Slatkin M. Microsatellite Interpretation using Rst. Genetics. 1995;139:457–62.
 95. Hardy OJ, Charbonnel N, Fréville H, Heuertz M. Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. Genetics. 2003;163:1467–82.
 96. Loiselle BA, Sork VL, Nason J, Graham C. Spatial Genetic Structure of a Tropical Understory Shrub. Am J Bot. 1995;82:1420–5.

97. Hawthorne W, Jongkind C. Woody plants of Western African forests. Royal Botanic Garden, Kew; 2006.
98. Villiers JF, Icacinacées. Olacacées, Pentadiplandracées, Opiliacées, Octoknémacées. In: Aubréville, A. & Leroy, J.F., eds. In: La flore du Gabon n°20. Muséum National d'Histoire Naturel, Paris, France; 1973. pp. 144–6.
99. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
100. Parmentier I, Malhi Y, Senterre B, Whittaker RJ, Alonso NATN. The odd man out? Might climate explain the lower tree α -diversity of African rain forests relative to Amazonian rain forests? *J Ecol*. 2007;95:1058–71.
101. Tosso F. Évolution et adaptation fonctionnelle des arbres tropicaux: le cas du genre *Guibourtia* Benn. PhD Thesis. Gemboux: Université de Liège – Gembloux Agro-Bio Tech. 2018.
102. Ewédjè E, Jansen S, Koffi GK, Staquet A, Piñeiro R, Essaba RA, et al. Species delimitation in the African tree genus *Lophira* (Ochnaceae) reveals cryptic genetic variation. *Conserv Genet*. 2020;21:501–14.
103. Migliore J, Kaymak E, Mariac C, Couvreur TLP, Lissambou BJ, Piñeiro R, et al. Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from *Greenwayodendron* (Annonaceae) phylogenomics. *J Biogeogr*. 2018;46:212–23.
104. Ewédjè E. Biologie de la reproduction, phylogéographie et diversité de l'arbre à beurre *Pentadesma butyracea* Sabine (Clusiaceae)-implications pour sa conservation au Bénin. PhD Thesis. Brussels: Université Libre de Bruxelles. 2012.
105. Dainou K, Mahy G, Duminil J, Dick CW, Doucet J-L, Donkpégan ASL, et al. Speciation slowing down in widespread and long-living tree taxa: insights from the tropical timber tree genus *Milicia* (Moraceae). *Heredity (Edinb)*. 2014;113:74–85.
106. Estoup A, Angers B. Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: Mol Ecol Carvalho GR. 1998.
107. White F. The Guineo-Congolian Region and Its Relationships to Other Phytochoria Author. *Bull du Jard Bot Natl Belgique*. 1979;49:11–55.
108. Mayr E, O'Hara RJ. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evol (N Y)*. 1986;40:55–67.
109. Linder HP. Plant diversity and endemism in sub-Saharan tropical Africa. *J Biogeogr*. 2001;28:169–82.
110. Droissart V, Dauby G, Hardy OJ, Deblauwe V, Harris DJ, Janssens S, et al. Beyond trees: Biogeographical regionalization of tropical Africa. *J Biogeogr*. 2018;45:1153–67.
111. Sosef M, Dauby G, Blach-Overgaard A, van der Burgt X, Catarino L, Damen T, et al. Exploring the floristic diversity of tropical Africa. *BMC Biol*. 2017;15:1–23.

112. Vanden Abeele S, Hardy OJ, Janssens SB. Isolation of microsatellite loci in the African tree species *Staudtia kamerunensis* (Myristicaceae) using high-throughput sequencing. Mol Biol Rep. 2018;45:1539–44.
113. Ley AC, Dauby G, Köhler J, Wypior C, Röser M, Hardy OJ. Comparative phylogeography of eight herbs and lianas (Marantaceae) in central African rainforests. Front Genet. 2014;5 NOV:1–14.
114. Textier N. Biogéographie des plantes vasculaires endémiques d'Afrique centrale occidentale. PhD Thesis. Brussels: Université Libre de Bruxelles. 2022.
115. Kouob BS. Organisation de la diversité végétale dans les forêts matures de terre ferme du Sud-est Cameroun. PhD Thesis. Brussels: Université Libre de Bruxelles. 2009.

Figures

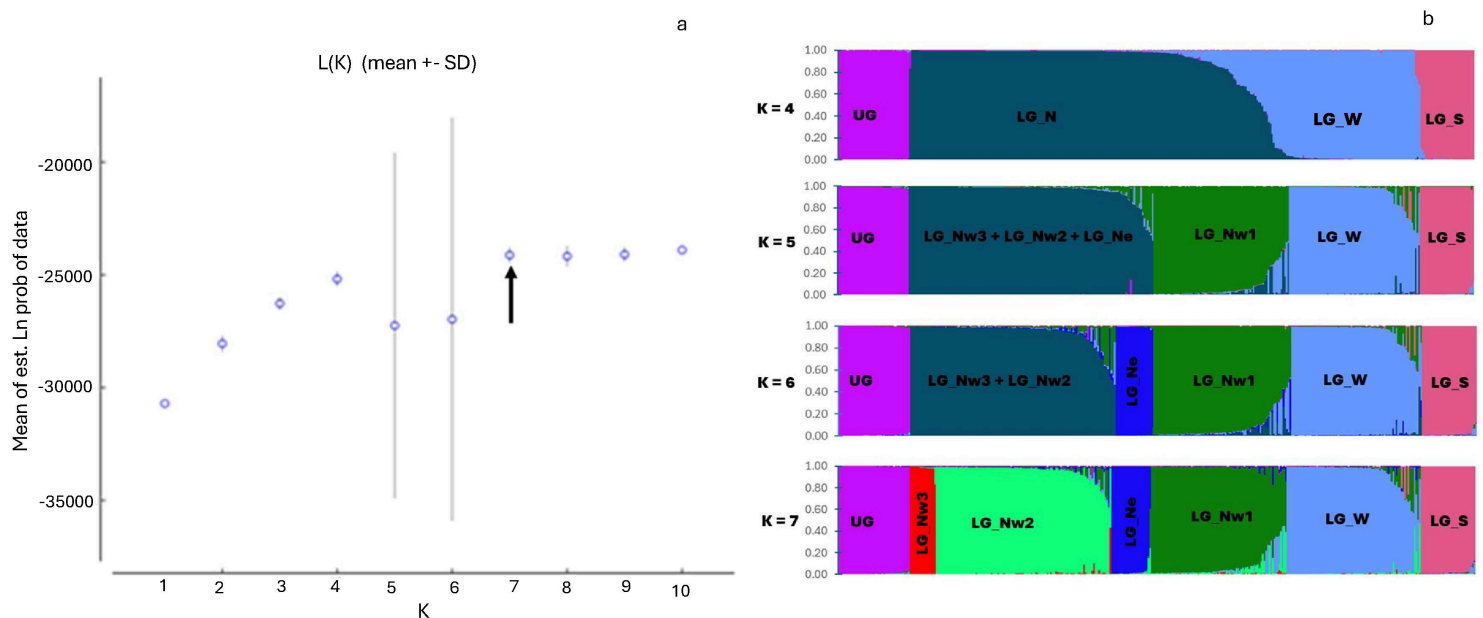


Figure 1

Results of the Bayesian clustering analysis performed with STRUCTURE [99] on 385 *Coula edulis* samples genotyped at 21 microsatellite loci. a: Posterior log-likelihood of the data according to the number K of genetic clusters: mean (circle) $\pm$ standard deviation (vertical bars) over 10 repetitions for

each K ; b: Histograms of the individual admixture proportions (q values) to genetic clusters for $K=4$ to 7, considering the respective repetitions with highest log-likelihood, with indication of their geographic distributions: Upper Guinea (UG) or Lower Guinea (LG)(see Fig. 2). Each cluster is represented by a different color.

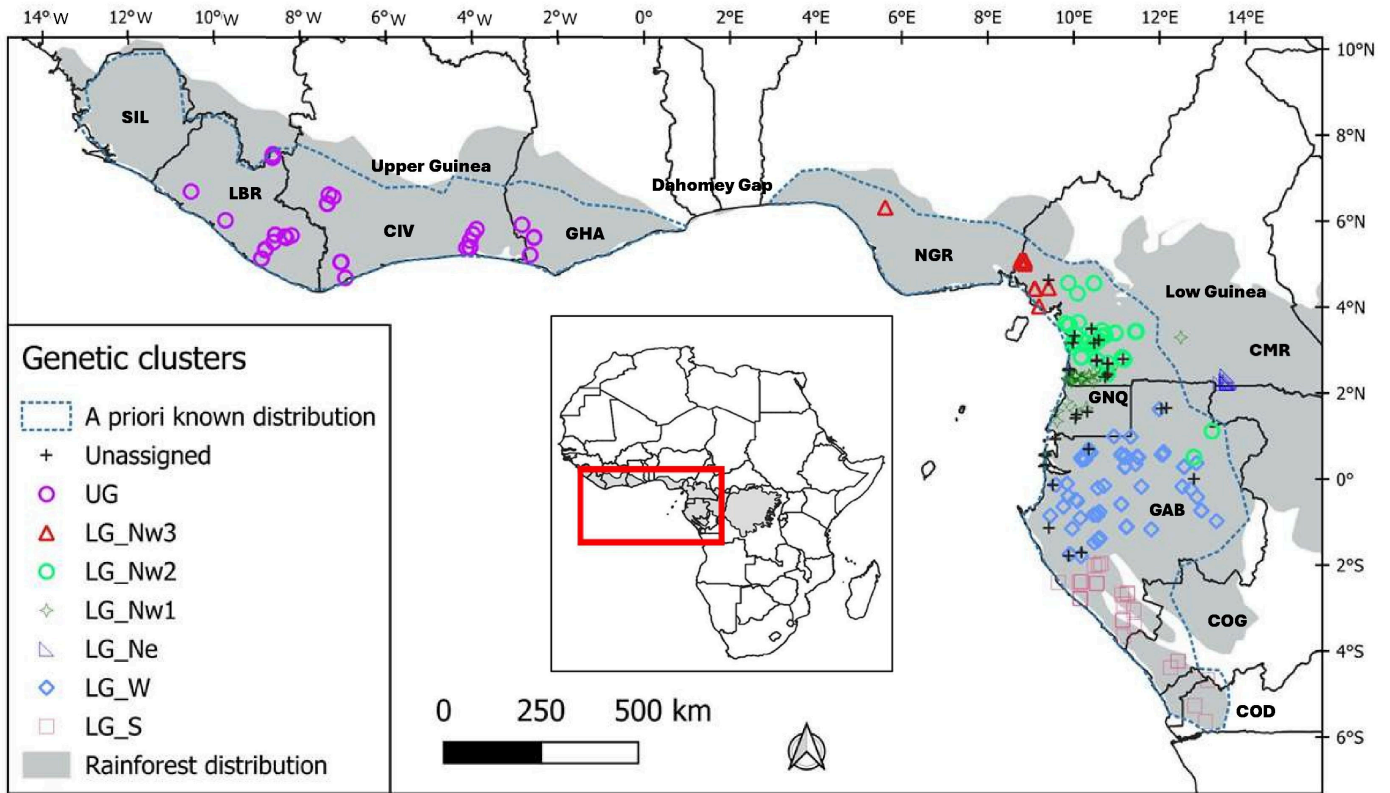


Figure 2

Distribution of 385 *Coula edulis* samples and their corresponding genetic cluster assignments as inferred with STRUCTURE under $K = 7$. Country codes are as follows: SIL: Sierra Leone; LBR: Liberia; CIV: Ivory Coast; GHA: Ghana; NGR: Nigeria; CMR: Cameroon; GNQ: Equatorial Guinea; GAB: Gabon; COG: Republic of the Congo; COD: Democratic Republic of the Congo. The dotted line delimits approximately the previously known distribution of *Coula edulis* according to GBIF, CJBG, and World Flora online databases, after excluding individuals observed in botanical gardens. The potential rainforest distribution (grey area) is based on [100].

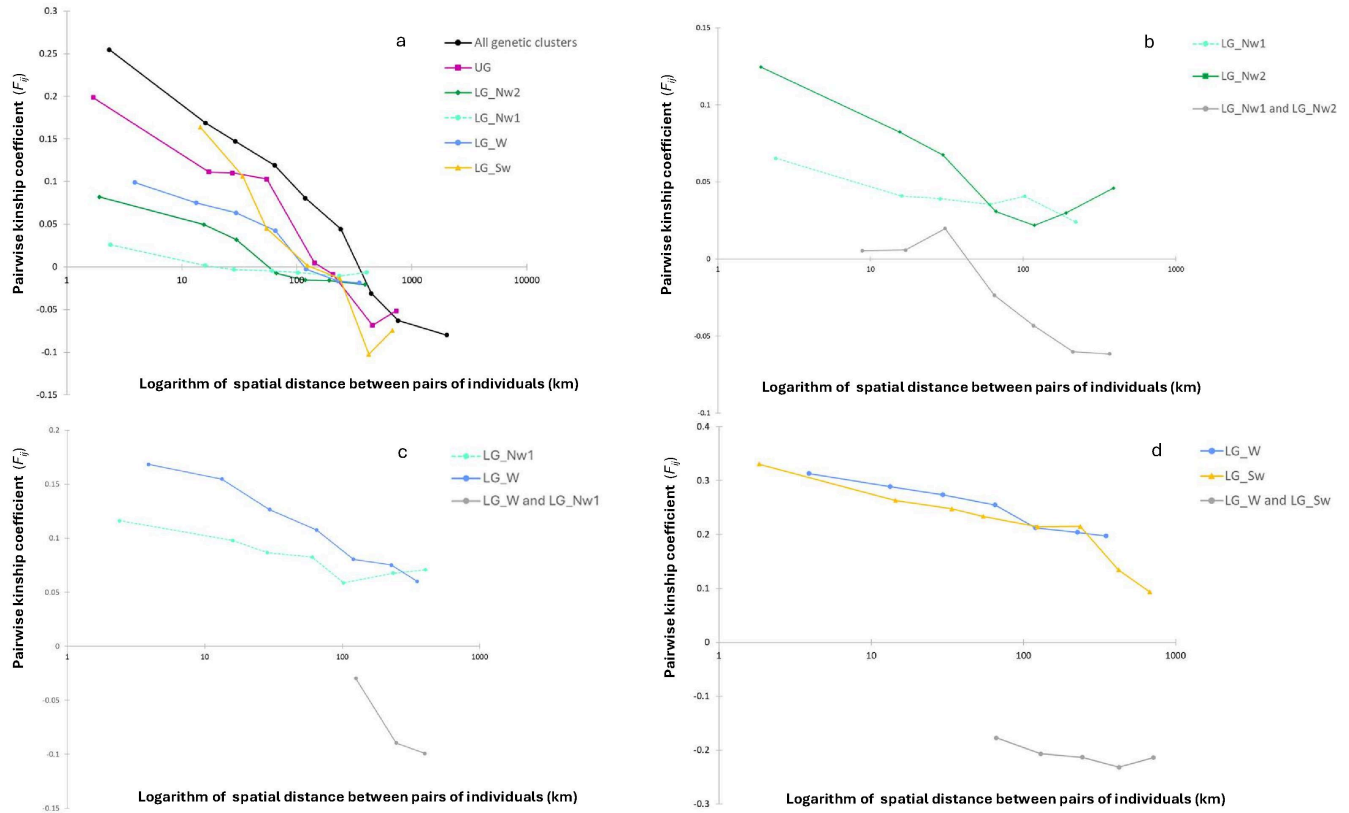


Figure 3

Kinship-distance curves for pairs of individuals assigned to a cluster ($q \geq 0.8$) according to $\log(\text{distance})$. a: within well-sampled genetic clusters (UG, LG_Nw2, LG_Nw1, LG_W, and LG_S) and across the whole distribution ranges ('All genetic clusters'); b: between LG_Nw1 and LG_Nw2; c: between LG_Nw1 and LG_W; d: between LG_W and LG_S. For pairs of clusters (panels b, c, and d), two curves represent the SGS within each cluster, and one curve corresponds to the kinship between samples from distinct clusters, representing the inter-cluster SGS dependency.

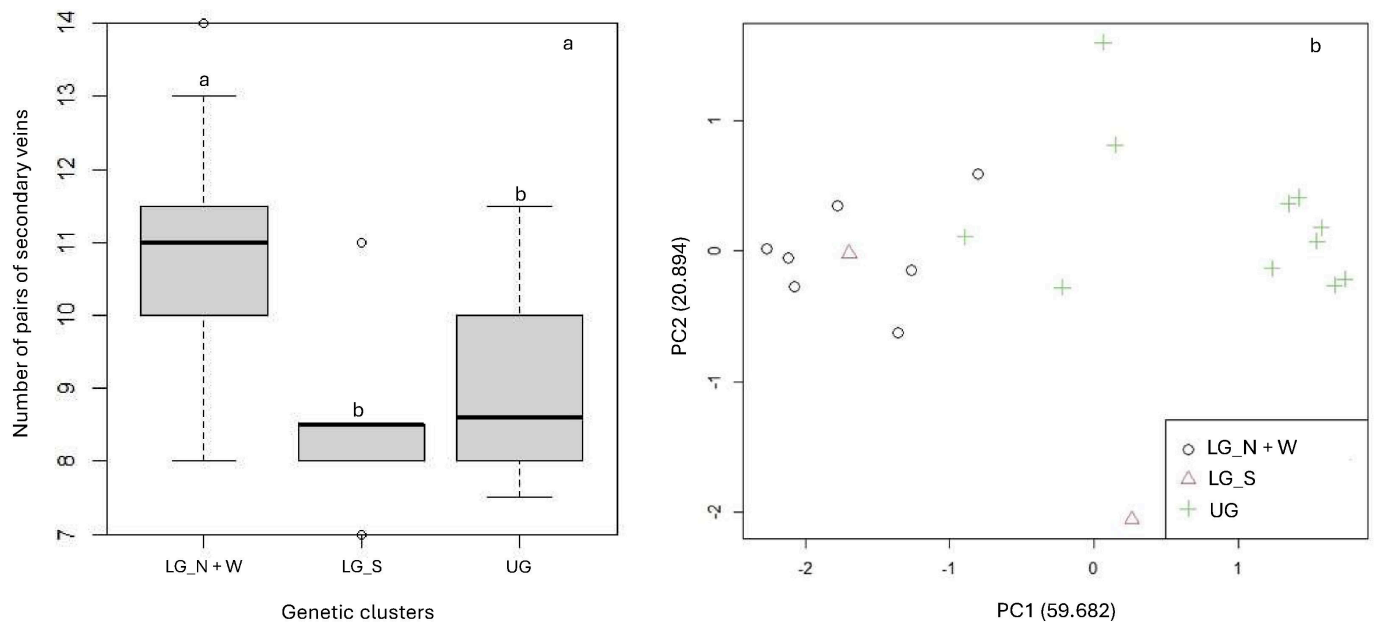


Figure 4

Morphological variation observed in herbarium specimens from three different regions corresponding to distinct genetic clusters. a: Boxplots of the number of pairs of secondary veins in 43 specimens (17 from LG_N+W, 5 from LG_S, 21 from UG). Values sharing the same letter are not significantly different ($P > 0.05$); b: Ordination according to the two first components of a PCA, explaining 59.7% and 20.9% of the variation expressed in the number of pairs of secondary veins, inflorescence length, and pedicel length, measured in 22 specimens. LG_N + W: include genetic clusters LG_Nw1, LG_Nw2, LG_Nw3, and LG_W.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)