



OPEN

Phenotypic variation, functional traits repeatability and core collection inference in *Synsepalum dulcificum* (Schumach & Thonn.) Daniell reveals the Dahomey Gap as a centre of diversity

Dédéou A. Tchokponhoué^{1,2✉}, Enoch G. Achigan-Dako², Sognigbé N'Danikou^{2,5}, Daniel Nyadanu³, Rémi Kahane⁴, Jacob Houéto², Nicodème V. Fassinou Hotegni², Alfred O. Odindo¹ & Julia Sibiyá¹

The miracle plant *Synsepalum dulcificum* is a multipurpose natural sweetener and a promising West African orphan fruit shrub candidate for genetic improvement. Unfortunately, basic knowledge such as phenotypic variation and inheritance estimates required for implementing a breeding program are still lacking. A set of 203 accessions were sampled in two habitats from seven populations spread across the Dahomey Gap (DG) and the Upper Guinea forest (UG) in West Africa. The phenotypic diversity and allometric relationships among functional traits were analysed; the broad-sense heritability was estimated for fruit-traits, and a mini-core collection was developed in the species. Quantitative variation in tree- and fruit-traits was recorded, and multivariate analyses were performed to assess relationships among accessions, whereas heritability was estimated using the coefficient of repeatability. Tree-traits observed in *S. dulcificum* were more variable than fruit-traits. While habitat-type only affected tree-traits, the provenance population significantly affected both fruit- and tree-traits, with the UG populations outperforming the DG ones. Significant correlations were observed among fruit-traits on one hand, and among tree-traits on the other hand, whereas poor correlations were observed between tree- and fruit-traits. The multivariate analysis grouped accessions in three clusters. Promising individuals for high fruit mass and pulp-dense genotypes' selection were identified within clusters. Repeatability estimates for fruit-traits ranged from 0.015 (edible ratio) to 0.88 (fruit mass). The Core Hunter algorithm enabled the extraction of 41 individuals as robust representatives of the initial set of 203 accessions, and the mapping of this core collection suggested Dahomey Gap as a centre of diversity of the species. These original findings offer opportunities, not only for the genetic improvement of *S. dulcificum*, but also for targeted ex-situ conservation in the species.

The miracle plant *Synsepalum dulcificum* (Schumach & Thonn.) Daniell [Syn. *Richardella dulcifica* (Schumach & Thonn.) Baehni] (Fig. 1a,b)—Sapotaceae—is a West African native shrub that produces a red fruit known as “miracle berry” (Fig. 1c,d). The berry is a unique natural source of miraculin¹, a glycoprotein contained in the pulp that induces sweetness. *Synsepalum dulcificum* has numerous other applications ranging from traditional to modern uses². Traditionally, all the non-edible parts of the species are involved in the treatment of many ailments including malaria, enuresis, coughing, and tooth decay among others^{3,4}. Currently, the miracle berry is considered as a reliable alternative to synthetic sugar^{5,6} and helps to control obesity⁷ and diabetes^{8–10} in modern

¹School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Scottsville, Private Bag X01, Pietermaritzburg 3209, South Africa. ²Laboratory of Genetics, Horticulture, and Seed Science (GBioS), School of Plant Sciences, University of Abomey-Calavi, 01 BP 526, Abomey-Calavi, Benin. ³Cocoa Research Institute of Ghana (CRIG), P. O. Box 8, Akim Tafo, Ghana. ⁴Research Unit HortSys, Department Persyst, CIRAD, Campus de Baillarguet, 34398 Montpellier Cedex 5, France. ⁵World Vegetable Center, East and Southern Africa, PoBox 10, Duluti, Arusha, Tanzania. ✉email: dedeoutchokponhoue@gmail.com

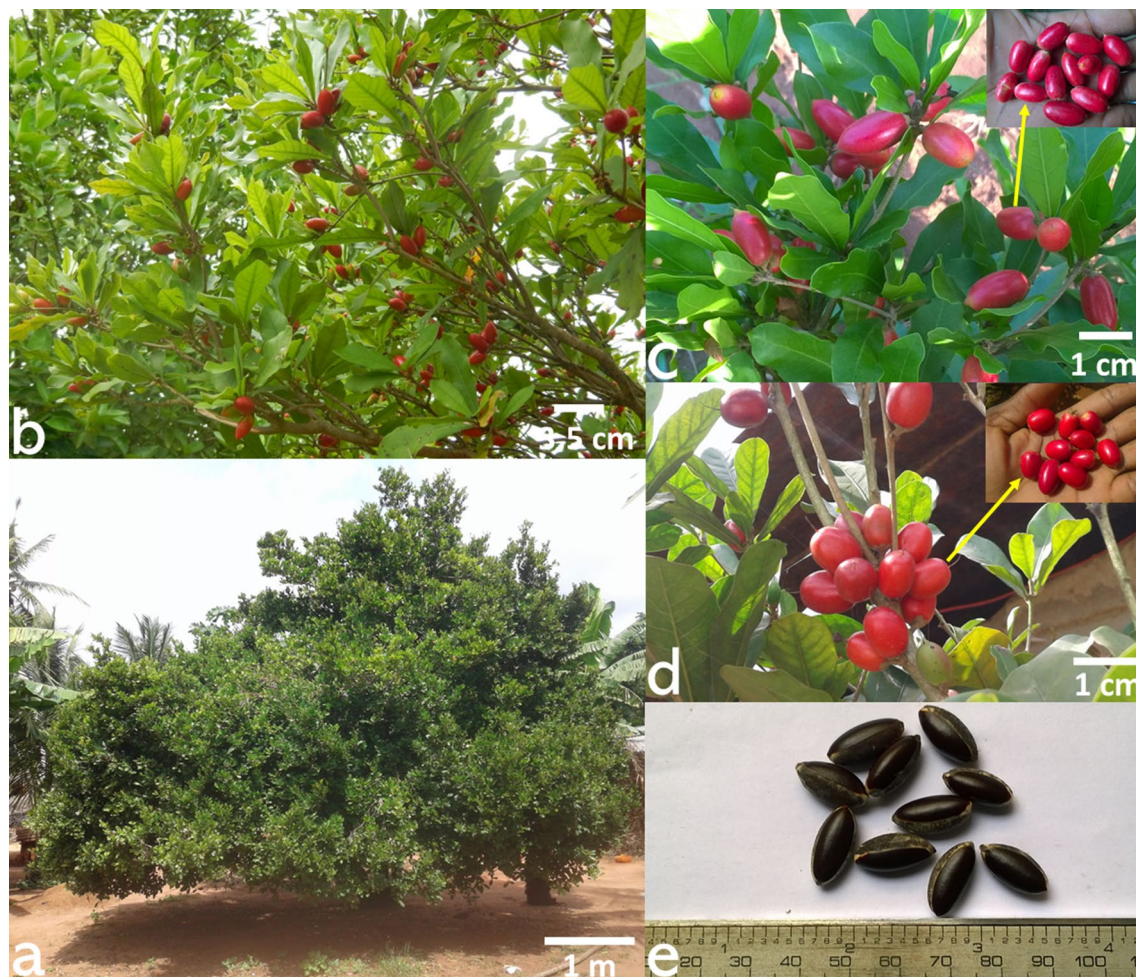


Figure 1. *Synsepalum dulcificum*, the studied species and some of its key organs. (a) A 110-year-old tree in home garden in Mission-Tove, Togo; (b) *S. dulcificum* branches bearing ripe fruits (miracle berries); (c) Oblong-shaped miracle berries; (d) Ovoid-shaped miracle berries; and (e) *S. dulcificum* seeds.

pharmaceuticals. It also serves as a food additive with its red skin and whitish pulp being, respectively, used for colouring and sweetening beverages and foods¹¹. Applications of the species in cosmetics relate to the seed (Fig. 1e) oil used in hair breakage control and the improvement of hand and finger motor skills^{12,13}. In West Africa, particularly in Benin, the miracle fruit is also sold on the open market and thus contributes to improving household livelihoods through income generation³.

Despite its growing market value (link1, link2), *S. dulcificum* is on the red list of the International Union for Conservation of Nature and classified as an endangered species in Benin¹⁴, with a risk of total depletion if no action is taken. Unfortunately, no conservation effort at either the local, regional and global scale has been so far reported in the species. Significant advances have been made in the development of farmer-friendly regeneration protocols¹⁵ and cultivation practices^{16,17} for production of *S. dulcificum* as a horticultural crop. However, foundation research for the genetic improvement and gene pool conservation of the species is still scanty. With an increasing interest in *S. dulcificum*^{18,19} worldwide, the economic potential of the shrub can be exploited if elite cultivars with improved traits are developed. However, the lack of basic knowledge on the genetics of the species is perceived as a bottleneck to the development and implementation of a relevant breeding program. For instance, the extent of genetic variation in *S. dulcificum* is elusive, and until now, there has been no study on the phenotypic diversity of the species. In their report on the molecular diversity of the species, Chibuzor et al.²⁰ did not include any quantitative phenotypic variation. While ex-situ evaluation is a common approach for the phenotypic characterization of annual and biannual plant genetic resources²¹, in-situ evaluation is important in any preliminary study of the morphological variation in perennial species^{22,23}. This has been shown in many fruit tree species including black raspberry (*Rubus occidentalis* L.)²⁴, sumac (*Rhus coriaria* (L.) Kuntze)²⁵, ber (*Ziziphus* spp.)²⁶, wild almond (*Prunus scoparia* L.)²⁷ and the African bush mangoes (*Irvingia* spp.)²⁸, among others. This approach also provides primary information to shape a gene pool conservation strategy²⁹, and sounds relevant when it comes to an orphan and threatened species such as the miracle plant.

Given that the fruit is the most valued and economically important organ in *S. dulcificum*, fruit and seed traits thus represent good candidates for genetic diversity studies and any selection. Predicting genetic gain for a target trait requires prior knowledge of its heritability³⁰, and such information is currently unavailable in *S. dulcificum*.

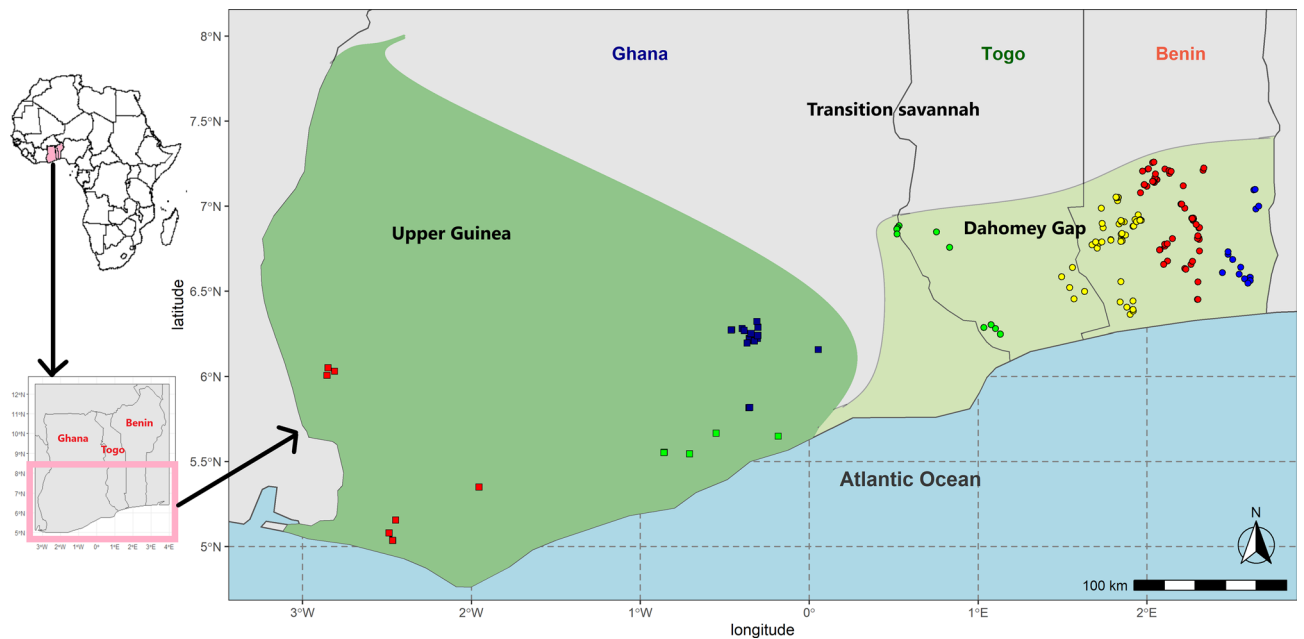


Figure 2. Study area showing the spatial distribution of the 203 accessions of *Synsepalum dulcificum* sampled in West Africa. Accessions are coloured according to the seven population sites under study. Square shapes: Upper Guinea forest block populations (filled red square: Western, filled green square: Central, filled blue square: Eastern); and Round shapes: Dahomey Gap populations (filled green circle: Volta, filled yellow circle: Mono, filled red circle: Zou, filled blue circle: Oueme). Map generated using the R environment (<https://www.r-project.org/>) (Version 3.6.2).

In perennial species, the coefficient of repeatability, which is the correlation between repeated measurements of the same individual over time and/or space is considered a reliable estimator of broad-sense heritability³¹. It has been used for fruit-traits heritability estimation in several tropical, subtropical and temperate tree species including mangabeira (*Hancornia speciosa* Gomes)³², akebia [*Akebia trifoliata* (Thumb.) Koidz.]³³, shea tree (*Vitellaria paradoxa* C.F. Gaertn.)³⁴ and peach tree [*Prunus persica* (L.) Batch]³⁵, among others. This metric appears relevant in the case of *S. dulcificum* to inform on the potential inheritance of fruit traits.

Core collection development emerged as an approach to optimize plant genetic resources conservation and utilization, but is still in its infancy for perennial and fruit tree species^{36–38}. More importantly, no evidence exists for such resources in African indigenous underutilized fruit tree species, despite its potential to rationalize orphan fruit tree breeding process by helping to skirt the necessity to evaluate multi-traits on a large sample of individuals. Because *Synsepalum dulcificum* exhibits recalcitrant seeds³⁹, cryopreservation appears to be the most adequate way for the maintenance of its genetic resources^{40,41}. Unfortunately, cryopreservation is currently inexistent in West Africa. The only cost-effective alternative ways to conserve the species ex-situ is to establish it as living collections, whose size will be tremendously reduced, and the management much eased with the development of a core collection.

The current natural distribution area of *S. dulcificum* ranges from Ivory Coast to the Democratic Republic of Congo (link3). This area covers four major ecological regions of the African lowland rain forest including the Upper Guinea forest block (UG), the Dahomey Gap (DG), the Lower Guinea forest block (LG) and the Congolian forest block (CB)⁴². However, the species has been reported to originate from a complex of countries including Ghana, Togo and Benin^{43–45}, which encompasses UG and DG.

Therefore, the objectives of the present study were to: (1) assess the phenotypic diversity of *S. dulcificum* among genetic resources from Benin, Togo and Ghana in West Africa; (2) estimate the heritability of some major fruit traits, and; (3) develop a core collection in the species. The overall goal of this study is to design and implement a relevant breeding program for the miracle plant in West Africa in order to sustain the utilization of the species. The study was guided by the following research questions: (1) what is the extent of quantitative variations in the miracle berry tree- and fruit-traits in its centre of origin and how are these traits affected by the ecological conditions?, (2) what is the clustering pattern among the miracle plant accessions?, (3) how heritable are the miracle berry traits? and (iv) what is the structure of the phenotype-based core collection in the species in the distribution area of West Africa including Benin, Togo and Ghana?

Materials and methods

Study area for sampling. The data collection focused on the Upper Guinea forest block (UG) and the Dahomey Gap (DG) (Fig. 2) and was expected to capture maximum diversity. The DG is a savannah corridor disrupting the zonal West African rain forest. It was created during the late Holocene following an abrupt climatic change between 4500 and 3400 cal. years BP⁴⁶. It is considered as a geographical barrier between UG demarcated at the West through the Volta river, and LG demarcated at the East through the Weme river⁴⁷. The

Traits	Traits code	N	Min	Max	Mean	SD	CV (%)
Tree-traits							
Tree diameter at ground level (cm)	DiamGr	203	4.28	68.78	19.12	10.20	53.34
Tree height (m)	PIHgt	203	0.86	7.50	3.47	1.27	36.59
Tree crown diameter (m)	CrDiam	203	0.75	8.80	4.24	1.62	38.20
Fruit-traits							
Fruit length (mm)	AvFL	1015	13.04	26.41	18.86	1.82	9.65
Fruit width (mm)	AvFW	1015	7.39	18.30	10.59	1.28	12.08
Individual fruit mass (g)	AvFM	8875	0.43	2.03	1.09	0.28	25.68
Individual seed mass (g)	AvSM	8875	0.16	0.60	0.35	0.07	20.00
Edible ratio (%)	EdRt	8875	45.25	83.01	66.99	7.17	10.70
Fruit shape index	FrSh	1015	1.06	2.39	1.79	0.18	10.05

Table 1. Quantitative variation in *Synsepalum dulcificum* tree- and fruit-traits (n = sample size for the measurement, Min: Minimum, Max: Maximum, SD: Standard deviation; CV: Coefficient of variation).

DG zone experiences 1000–1200 mm annual rainfall. The vegetation is a mixture of savannahs, gallery forests, and fragments of swamp forests, as compared to that of UG, dominated by rain forest with >2000 mm annual rainfall⁴⁸. The UG block corresponds to the rain forest belt extending from Sierra Leone to Ghana⁴². A total of seven collecting sites were investigated including four (Oueme, Zou, Mono and Volta) in DG and three (Eastern, Central and Western) in UG, giving rise to seven populations (Fig. 2). A population of *S. dulcificum* was defined as a set of accessions possibly interbreeding and randomly distributed in an agroforestry or tree-based production system within the same area or environment. Specifically, two distant populations of *S. dulcificum* (e.g. Mono and Volta, Eastern and Western) were separated by the minimum geographical distance of 50 km whereas two adjacent populations of *S. dulcificum* were separated by a maximum distance of 15 km coupled with the existence of a natural barrier such as a mountain (as observed between the adjacent populations Central and Eastern, in Ghana) or a river (as observed between the adjacent populations Zou and Oueme, and between the adjacent populations Mono and Zou, in Benin).

From September 2016 to August 2019, a total of 203 accessions (individual trees) of *S. dulcificum* were sampled from the seven populations, with the number of accessions sampled per population varying from 4 to 66 (Oueme: 19, Zou: 67, Mono: 67, Volta: 11, Eastern: 25 Central: 4, and Western: 10). The sampled accessions were all of reproductive age (i.e., having borne fruits at least once), and bearing fruits during the field sampling. The minimum distance of 100 m was observed between two sampled shrubs in order to reduce occurrence of closely related individuals.

Data collection. For each accession, data on the habitat type, tree-traits, fruit-traits, and the tree age (where possible) were recorded. The habitat was recorded as either a home garden, which is here defined according to Niñez⁴⁹ as a production system characterized by its proximity to dwellings, and over which the owner has a certain control on the target products; or as a farm, which is defined as an open agricultural production space, and usually farther away from the dwellings than a home garden. The tree age was estimated by the owner (the planter) where applicable; otherwise one of his closest relatives (e.g. brother, child or grandchild) gave the information.

Three tree-traits and six fruit-traits were measured per accession (Table 1). The tree-traits included: 1. Plant height (PIHgt), 2. Tree diameter at ground level (DiamGr), and 3. Tree crown diameter (CrDiam). The fruit traits included: 4. Fruit length (AvFL), 5. Fruit width (AvFW), 6. Fresh fruit mass (AvFM), 7. Seed mass (AvSM), 8. Edible ratio (EdRt) which is the pulp-to-fruit ratio, and 9. Fruit shape index (FrSh).

The plant height was measured from the soil surface to the tallest point of the shrub using a tape measure. The crown diameter was obtained by averaging the measures taken in the south-north and east–west directions. The diameter at ground level was obtained using the Eq. (1):

$$DiamGr = C_{Gr} / \pi \quad (1)$$

where DiamGr is the tree diameter at ground level, and C_{Gr} is the circumference at ground level measured with a tape measure.

The fruit-traits were recorded in-situ on 20–50 mature, fully ripe (red) and pests-free fruits that were randomly collected from the four sides (South, East, West and North) of each tree and bulked in replicates of 10 fruits each. A total of 8875 single fruits were assessed for this study.

The fresh mass of each replicate (FM_{10}) was measured (± 0.01 g) using a portable electronic scale (OHAUS, model: PA512, Port Melbourne, Australia). The mean value of the replicates was then divided by 10 to obtain an average fruit mass (AvFM) of the accession. The same procedure was used to obtain the average seed mass (AvSM) of each accession after the fruits were depulped, while being kept in their replicate lot. The Edible Ratio (EdRt) was determined on each replicate, and the mean EdRt of an accession was the average value of all replicates of the accession. The EdRt was computed following the Eq. (2):

$$EdRt = (FM_{10} - SM_{10}) / FM_{10} \quad (2)$$

where FM_{10} is the fresh mass of 10 fruits, and SM_{10} is the fresh mass of the seeds of the same 10 fruits.

The fruit length and width were measured on five fruits randomly sampled from the different replicates per accession, using a digital Vernier calliper (± 0.01 mm). The mean values of all replicates per accession were used as the average fruit length (AvFL) and width (AvFW) of the accession. The fruit shape index (FrSh) of each accession was determined from five fruits following the Eq. (3):

$$FrSh = AvFL / AvFW \quad (3)$$

Data analysis. All the analyses were performed using the R language Version 3.6.2⁵⁰.

Tree morphology and fruit-trait variation. Descriptive statistics (minimum, maximum, mean, standard deviation, coefficient of variation) were computed to understand the overall variation in the species. The effect of the habitat-type on the tree height, diameter at ground level, crown diameter, fruit mass, fruit length, fruit width, fruit shape index and seed mass were assessed using a t-test, or a Mann Whitney's test, where relevant. Likewise, the effect of the provenance population of accessions on the same variables was analysed using a type-II analysis of variance (ANOVA) to account for sample size imbalance among populations, or a Kruskal–Wallis test, where relevant. The edible ratio was assessed using a generalized linear model with a quasi-binomial error structure to account for overdispersion. Where the effect of the population was significant, a contrast analysis was employed to depict the influence of the ecological region (Upper Guinea vs Dahomey Gap) on the different parameters tested.

Relationship among tree morphology and fruit-traits. Relationships among tree- and fruit-traits were sought, and their significance were tested using Pearson's or Spearman's correlation tests, as appropriate. The correlation strength and significance were illustrated using the *chart.Correlation()* function of the "PerformanceAnalytics" package⁵¹. To depict the effect of the ecological region on the relationship among functional fruit-traits in the species, allometric regressions were used with sequential tests for slope equality and elevation shift between DG and UG accessions for various regression lines. This was done using the *sma()* function of the "smatr" package⁵², whereas regression lines were drawn using the *ggPredict()* function of the "ggiraphExtra" package⁵³.

To assess the relationships among and between accessions and tree- and fruit-traits, a series of multivariate analyses were conducted. First, a Principal Component Analysis (PCA) was performed using the *PCA()* function implemented in the "FactoMineR" package⁵⁴ to retain the most meaningful components. Second, the clustering tendency in the dataset was tested using the Hopkins statistics "H" through the *hopkins()* function implemented in the "clustertend" package⁵⁵. A value of H close to 0.5 indicated that the dataset was not clusterable, whereas a value of H close to 0 indicated a clusterable dataset. In the latter case, a hierarchical clustering on the retained principal components (HCPC) was performed using the *HCPC()* function of the "FactoMineR" package. The graphical outputs were visualized using the function *fviz_cluster()* of the "factoextra" package⁵⁶.

Estimation of repeatability. For each of the six fruit-traits, the repeatability estimation was conducted using the linear mixed effect models framework implemented in the "rptR" package⁵⁷. To that, the provenance population was set as a fixed effect and the tree/accession as a random effect. The repeatability of the edible ratio was estimated using a generalized linear mixed-effect model with the *rptProportion()* function whereas repeatability estimates of other fruit traits (fruit mass, seed mass, fruit length, fruit width and fruit shape index) were determined using a generalized linear mixed-effects model fitted by restricted maximum likelihood (REML) with the *rptGaussian()* function. The interest in using the "rptR" package over the classical approach was mainly its ability to deal with non-gaussian data (e.g. the edible ratio in our case that is a proportion data). The significance of each repeatability was tested using the Likelihood Ratio Test (LRT), and the uncertainties (standard error: S.E. and Confidence Interval: C.I.) associated to the repeatability estimates were determined using a parametric bootstrapping (n = 1000 data samples).

Development and evaluation of the core collection. A Core Hunter phenotype data was developed using the *phenotype()* function implemented in Core Hunter version 3 of the "Core Hunter" package⁵⁸. Thereafter the *samplecore()* function of the same package was applied to the Core Hunter data previously generated to develop the core set of accessions using the average-entry-to-nearest-entry distance optimization objective supported by the Gower distance measure. The quality of the core collection developed was evaluated by first computing the Coincidence Rate of range (CR%) (Eq. 4), the Variable Rate (VR%) (Eq. 5), the Variance Difference percentage (VD%) (Eq. 6), and the Mean Difference percentage (MD%) (Eq. 7) between the whole collection and the core collection, following Kim et al.⁵⁹ and Hu et al.⁶⁰:

$$CR\% = \frac{1}{m} \sum_{i=1}^m \frac{R_{C_i}}{R_{W_i}} \times 100 \quad (4)$$

where R_{C_i} is the range of the core collection for the trait i , and R_{W_i} is the range of the whole collection for the trait i ;

$$VR\% = \frac{1}{m} \sum_{i=1}^m \frac{CV_{C_i}}{CV_{W_i}} \times 100 \quad (5)$$

where CV_{C_i} is the coefficient of variation of the core collection for the trait i , and CV_{W_i} is the coefficient of variation of the whole collection for the trait i ;

$$VD\% = \frac{1}{m} \sum_{i=1}^m \frac{|V_{W_i} - V_{C_i}|}{V_{C_i}} \times 100 \quad (6)$$

where V_{C_i} is the variance of the core collection for the trait i , and V_{W_i} is the variance of the whole collection for the trait i ;

$$MD\% = \frac{1}{m} \sum_{i=1}^m \frac{|M_{W_i} - M_{C_i}|}{M_{C_i}} \times 100 \quad (7)$$

where M_{C_i} is the mean of the core collection for the trait i , and M_{W_i} is the mean of the whole collection for the trait i .

Then, the existence of significant difference of phenotypic traits was tested between the whole collection and the core collection, using a t-test (for seed mass), a Wilcoxon's test (for fruit mass, fruit length, fruit width, tree stem diameter, tree height, tree crown diameter and fruit shape index), or a generalized linear model with a quasi-binomial error distribution (for edible ratio).

Additional packages such as "ggplot2"⁶¹ (for basic graphics management), "sf"⁶², "rworldxtra"⁶³ and "rnatu-ralearth"⁶⁴ (for mapping and geo-referencing) were also used.

Results

Quantitative variation in tree- and fruit-traits. Tree-traits generally exhibited larger variation than fruit-traits (Table 1). The phenotypic coefficients of variation ranged from 36.59% (tree height) to 53.34% (tree diameter at ground level) for tree-traits, and from 9.65% (fruit length) to 25.68% (fruit mass) for fruit-traits. The effect of the population provenance was significant ($p < 0.05$) on all the assessed traits except on tree diameter ($p = 0.211$) and tree crown diameter ($p = 0.337$) (Fig. 3a–i). In particular, trees were taller in the Upper Guinea forest (UG) than in the Dahomey Gap (DG) ($p = 0.001$, Fig. 3b). Likewise, accessions from the UG produced heavier, longer and larger fruits with higher edible ratio than those from the DG ($p < 0.0001$, Fig. 3d–h), whereas fruit shape index ($p < 0.0001$, Fig. 3i) was greater in accessions from DG (oblong fruit) than from UG (round fruit). Among the screened populations, the Volta population produced bigger fruits and heavier seeds ($p < 0.0001$).

Contrary to the provenance population, habitat-type had lower effect on most studied traits (Fig. 4a–i), only significantly affecting tree-traits (tree height, $p = 0.004$, and crown diameter, $p = 0.02$), with relatively bigger shrubs in home gardens than in farms. Tree diameter at ground level ($p = 0.07$) was relatively greater in shrubs of farms than in those of home gardens. The age estimate of the studied trees ranged from 9 to 150 years and was on average 57.69 ± 2.04 years. No significant age difference was found between individuals on farms (57.75 ± 3.44 years) and those in home gardens (57.66 ± 2.56 years) ($p = 0.56$) (see Supplementary Fig. S1).

Relationships among tree- and fruit-traits. Coefficients of correlation ranged from -0.00286 to 0.83 (Fig. 5). All tree-traits were positively and significantly correlated, and the strongest relationship was observed between tree height and crown diameter ($r = 0.62$, $p < 0.001$). The strongest relationship among fruit-traits was found between fruit mass and fruit width ($r = 0.83$, $p < 0.001$). However, fruit-traits were not uniformly correlated. Fruit mass was significantly correlated with all other fruit-traits ($r \geq 0.49$, $p < 0.01$) except with fruit shape index ($r = -0.4$), and fruit shape index was only correlated with fruit length ($r = 0.34$, $p < 0.001$). Edible ratio was positively correlated with fruit mass ($r = 0.63$, $p < 0.001$), but negatively with the seed mass ($r = -0.26$, $p < 0.001$), meaning that larger fruits had shorter seed and denser pulp. Overall, there was a poor relationship between tree-traits and fruit-traits in *S. dulcificum*, and the highest correlation was observed between tree height and fruit width ($r = 0.31$, $p < 0.01$) (Fig. 5). Similarly, out of all the traits assessed, the tree age estimate was only correlated, and moderately to the tree diameter ($r = 0.53$, $p < 0.001$) and weakly to the tree crown diameter ($r = 0.3$, $p < 0.001$) (Supplementary Table S2). The ecological conditions seemed to influence the relationships observed between fruit-traits (Fig. 6a–i). For example, stronger correlations were observed between fruit mass and seed mass (Fig. 6a) on one hand and between fruit mass and fruit length (Fig. 6b) on the other hand for accessions from the UG compared with those from the DG. However, the correlation between fruit mass and fruit width was not influenced by ecological conditions (Fig. 6c). Similarly, correlations between edible ratio and fruit mass (Fig. 6d), edible ratio and fruit width (Fig. 6g), and between seed mass and fruit length (Fig. 6h) were not affected by ecological conditions. Nonetheless, a highly significant effect of the ecological region was noted on correlations between edible ratio and seed mass (Fig. 6e), edible ratio and fruit length (Fig. 6f) and between seed mass and fruit width (Fig. 6i). In this case, an increase in fruit width resulted in a greater seed mass gain in accessions from DG than those from UG, whereas a decrease in seed mass yielded a greater gain in edible ratio in accessions from UG than those from the DG. A greater correlation between edible ratio and fruit length was observed under UG ecological conditions than under DG conditions.

Relationships among accessions of *S. dulcificum*. From the principal component analysis, it emerged that the first four components explained a significant proportion (83.3%) of the total variance (Fig. 7a–c) and these were retained for further analyses. The most correlated traits to the first component (35.1% of the total variability) were fruit mass ($r = 0.93$, $p < 0.0001$), fruit width ($r = 0.91$, $p < 0.0001$) and edible ratio ($r = 0.61$, $p < 0.0001$). This first component is that of the "fruit size". Similarly, the crown diameter ($r = 0.83$, $p < 0.0001$), tree

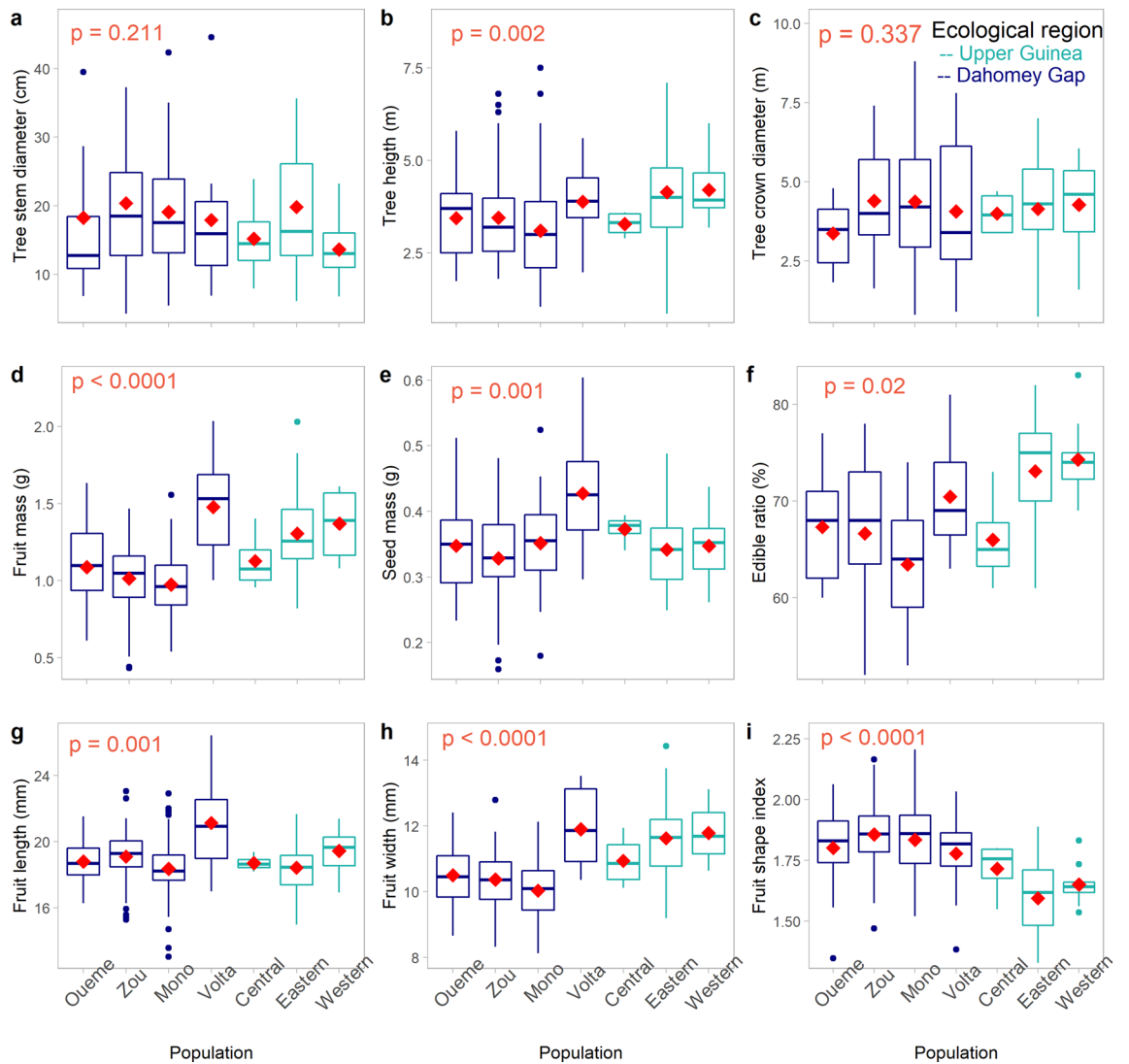


Figure 3. Quantitative variation of *Synsepalum dulcificum* tree- and fruit-traits across the seven sampled populations. Median values are in bold, red diamond shapes represent mean values, dots above and below boxplots are outliers, and lower and upper tails represent minimum and maximum values, respectively.

height ($r = 0.69$, $p < 0.0001$) and tree diameter ($r = 0.61$, $p < 0.0001$) were the most correlated to the second component, which accounted for 19.6% of the total variability. This component is that of “tree-shape”. The fruit shape index ($r = 0.756$, $p < 0.001$) and fruit length ($r = 0.71$, $p < 0.001$) were the most represented variables on the third component “the fruit shape component” that accounted for 16.8% of the total variability. Though the component 4 accounted for 11.8% of the total variance, none of the variables were found to significantly correlate to it.

The Hopkins statistics computed on the data was $H = 0.23$ (< 0.5), indicating the existence of significant clusters in the dataset. The hierarchical clustering on the first four components produced three clusters (Fig. 8). Clusters 1, 2 and 3 were made up of 46.8%, 26.5% and 26.7% of the total number of accessions, respectively and a χ^2 test on the clustering structure indicated a significant representation of accessions from the Upper Guinea in cluster 3 ($p = 0.001$) (see Supplementary Fig. S3). The most determinant quantitative characteristics of each cluster are presented in Table 2.

Cluster 1 was significantly characterized by all the nine traits studied, and mainly encompassed the least performing accessions observed in the study (Table 2). These accessions were of weak architecture, producing poor standard fruits that are markedly oblong in shape. Cluster 2 was significantly described by fruit length and seed mass in addition to tree traits (Table 2). Accessions in this cluster had a strong tree architecture and a high seed mass. Cluster 3 was characterized by individuals with a weak tree architecture, but with the highest performing fruit-traits. In part, accessions of cluster 3 had the highest fruit mass, fruit width, edible ratio, and seed mass (Table 2) and are the elite trees in this study.

Fruit traits repeatability in *S. dulcificum*. Repeatability coefficients for the six fruit traits under study are presented in Table 3. All the repeatability estimates were highly significant ($p < 0.0001$) with most traits

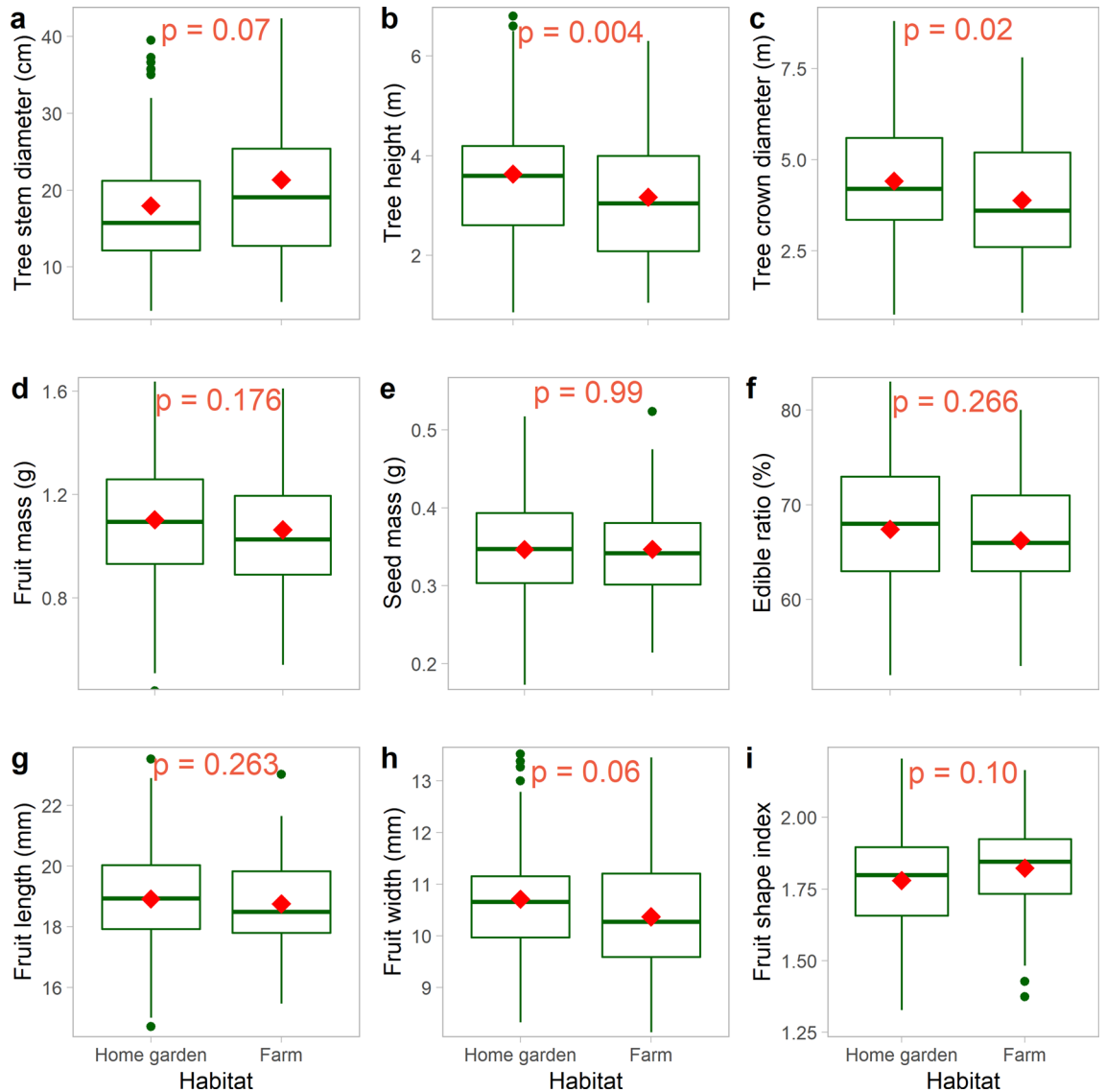


Figure 4. Quantitative variation of *Synsepalum dulcificum* tree- and fruit-traits across the habitat-types under study. Median values are in bold, red diamond shapes represent mean values, dots above and below boxplots are outliers, and lower and upper tails represent minimum and maximum values, respectively.

exhibiting high repeatability. Thus, the trait with the highest potential broad-sense heritability based on the repeatability estimates was fruit mass, whereas the one with the least broad-sense heritability was the edible ratio.

Mapping the core collection of *S. dulcificum*. Based on the phenotypic data from the 203 accessions evaluated, the Core Hunter algorithm returned as core collection a total of 41 accessions. The spatial distribution of these 41 accessions is presented in Fig. 9. This core set is made up of eight accessions from the Upper Guinea forest and 33 from the Dahomey Gap. The repartition of the core set accessions in the three clusters obtained above indicated that Clusters 1, 2 and 3 were represented by 15, 13 and 13 accessions, respectively, with the Upper Guinea ecological region having mainly accessions of Cluster 3 against the Dahomey Gap region encompassing accessions from all the three clusters.

The coincidence rate of range (CR) per trait varied from 87.57 to 100% for an overall CR of 97.6% (Table 4). Similarly, all the variable rates (VR) per trait were higher than 100%, whereas the variance difference percentage between the whole collection and the core collection was 46.95%. More captivatingly, the difference between the whole collection and the core collection was not significant for any of the nine phenotypic traits evaluated ($0.26 < p < 0.99$). This observation was supported by the low mean difference percentage between the core and the whole collection, which was 3.23% (Table 4).

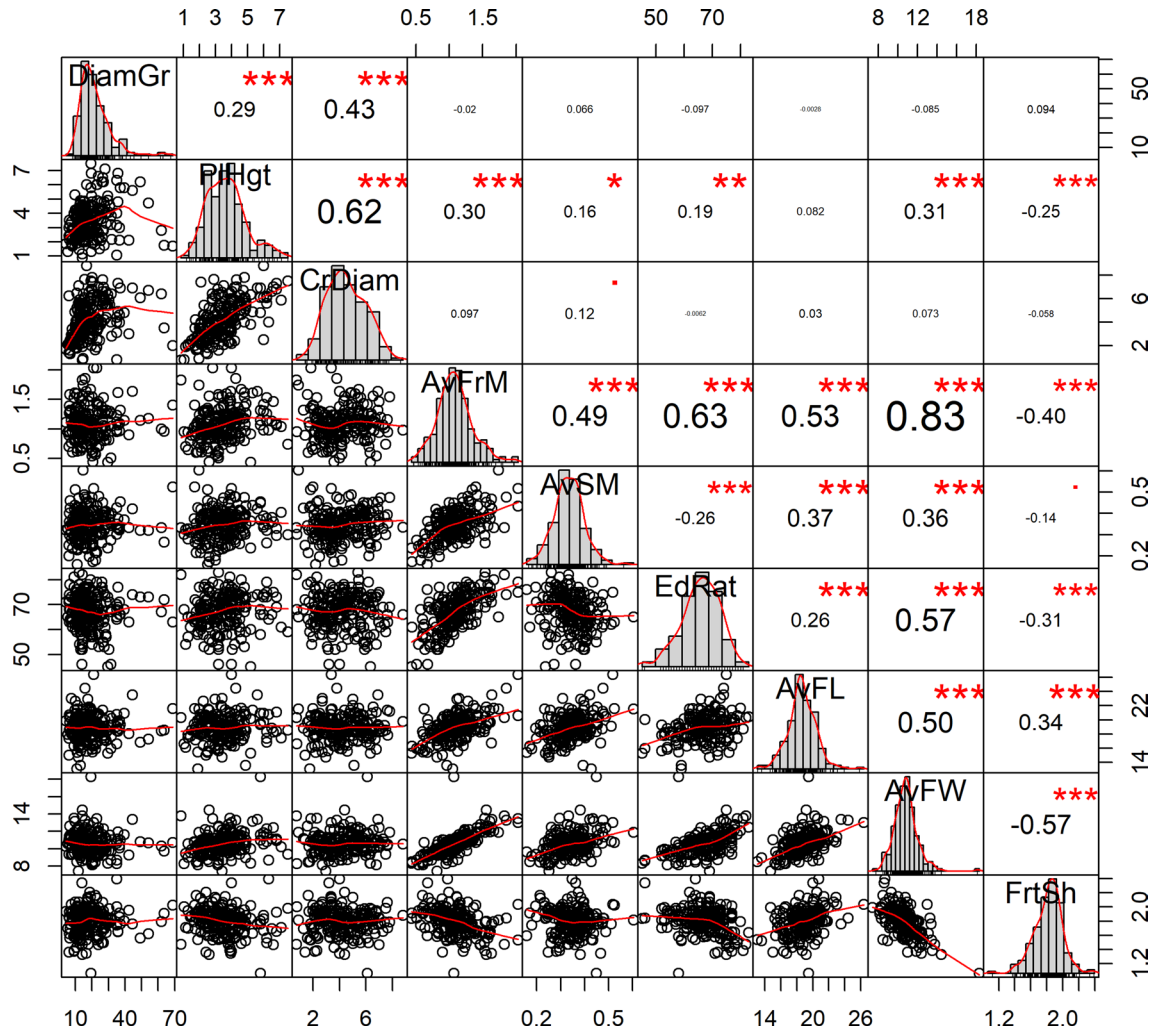


Figure 5. Mixed correlogram based on nine *Synsepalum dulcificum* tree- and fruit-traits (Traits list detailed in Table 1). On the main diagonal histograms are presented for each trait. Figures at the lower triangle present bivariate scatterplots, with a fitted line. Upper triangle contains correlation coefficients (r) with significance value from a correlation test. Magnification of r values is proportional to the correlation strength and magnification of the p-value value symbol is proportional to the significance level: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$, ■ for $p < 0.1$.

Discussion

The study reports for the first-time, the phenotypic variation, heritability estimates for functional fruit traits and the development of a core collection for *S. dulcificum* in West Africa.

The miracle plant *S. dulcificum* is known to be an endangered species in Benin where it has been mostly reported in swamp forests¹⁴. However, from this current study, although not a purely ecological one, it is worth noting that no accession was found in swamp forest in Benin, but rather in farms and mostly in home gardens. In Togo, another portion of the Dahomey Gap, the species was notably absent from forests, especially from sacred forests where it had a rarity index of 0.98. This trend suggests that the predominant habitat of the species in the Dahomey Gap might have shifted from the natural environment to human-close environment such as home gardens, probably due to anthropogenic pressures such as land clearing for agricultural expansion. This seemed to align with the idea of home gardens as a biodiversity hotspot for endangered species⁶⁵. In the Upper Guinea forest, the current absence of the species from forests might not only illustrate an in-situ anthropogenic pressure on the resource, but could also reflect a land conversion scenario in which forests were degraded in favour of tree plantation systems. This hypothesis is supported by previous reports indicating that the Upper Guinea forest has been experiencing serious degradations, which have already resulted in the loss of nearly 80% of the original forest cover⁶⁶. Almost all the farms in which the species was recorded in the Upper Guinea forest zone were characterized by a tree plantation system where *S. dulcificum* was associated with cocoa (*Theobroma cacao* L.), rubber tree [*Hevea brasiliensis* (Willd. ex A.Juss.) Müll.Arg.], orange tree (*Citrus* spp) or *Musa* spp.

To date, no information existed on the variation of quantitative traits in *S. dulcificum*. In this study, using the populations sampled from the species' centre of origin, trends of traits variations were similar to those observed in sister Sapotaceae species such as *V. paradoxa*^{34,67} from tropical Africa, and also *Argania spinosa* (L.) Skeels⁶⁸

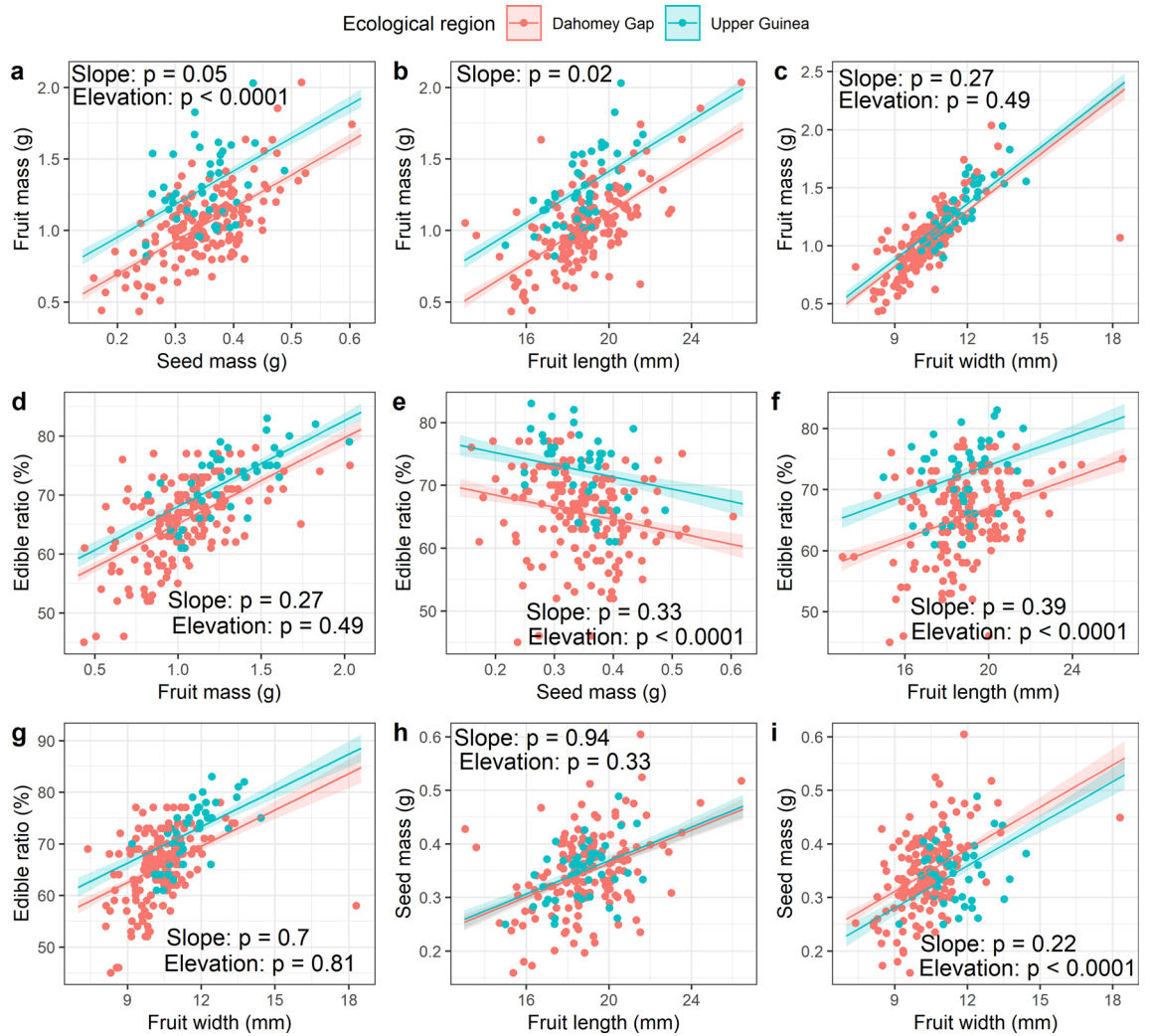


Figure 6. Allometric relationships among selected functional fruit-traits in *Synsepalum dulcificum* across the Dahomey Gap (in red) and the Upper Guinea forest (in blue) ecological regions. Slope: Test on difference between slopes; Elevation: Test on elevation shift in regression lines.

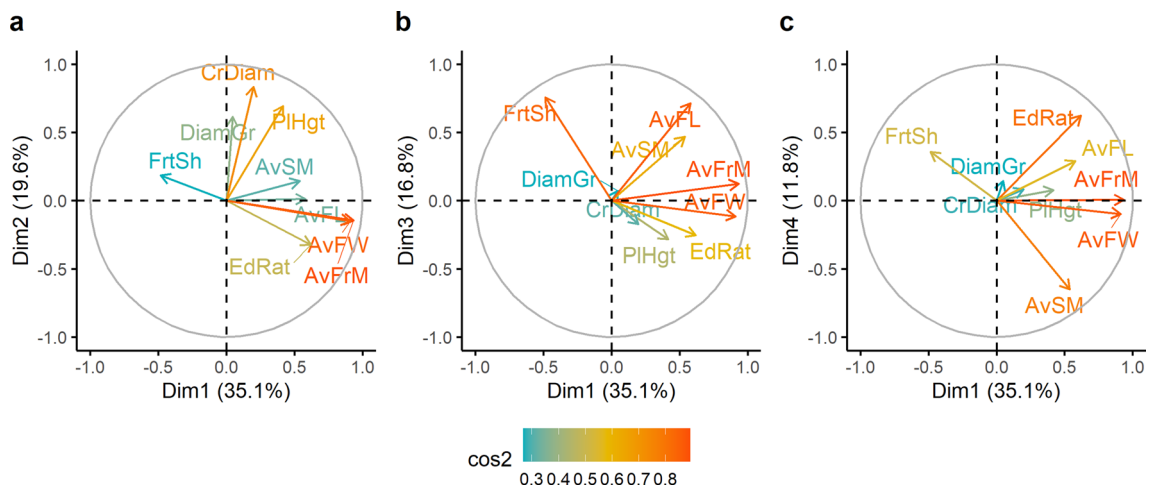


Figure 7. Correlation circle indicating *Synsepalum dulcificum* tree- and fruit-traits projection on the first two components: (a); first and third components: (b); and first and fourth components: (c) (traits list is detailed in Table 1).

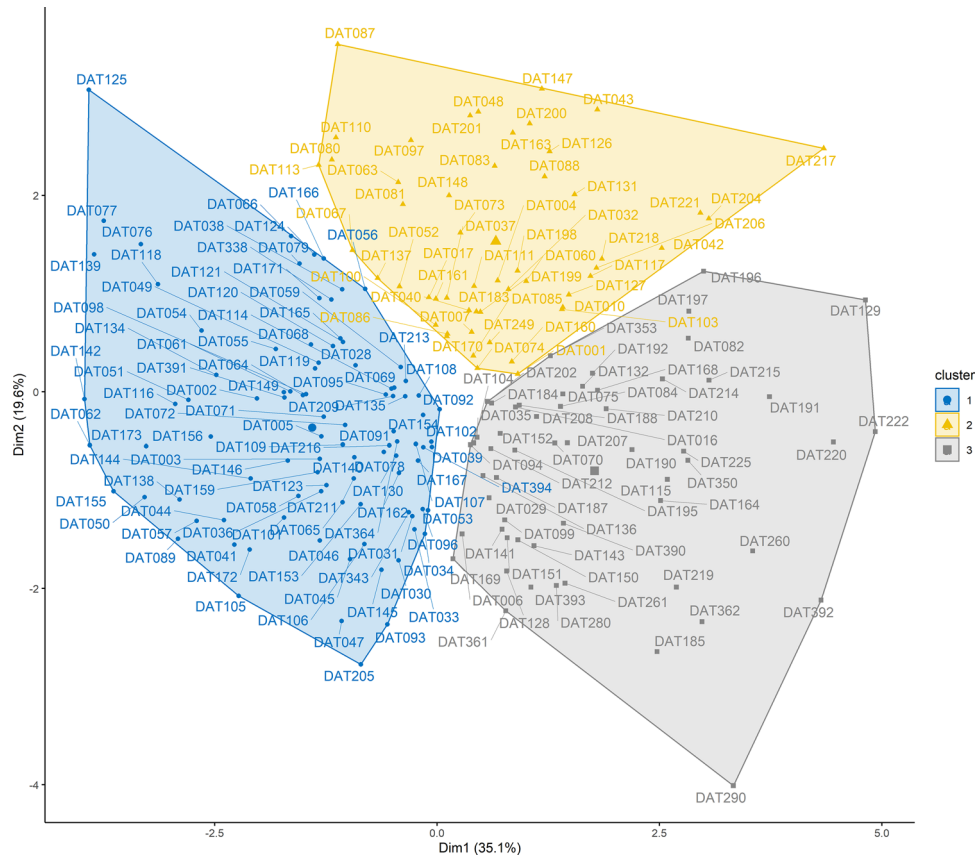


Figure 8. Factor map showing *Synsepalum dulcificum* accessions grouping in clusters.

Quantitative traits	Clusters		
	1	2	3
Tree-traits			
DiamGr	17.14 ± 0.88	26.70 ± 12.21	15.42 ± 6.21
PlHgt	2.81 ± 0.09	4.59 ± 1.30	NA
CrDiam	3.60 ± 0.12	5.97 ± 1.20	3.68 ± 1.36
Fruit-traits			
AvFL	17.99 ± 0.16	19.45 ± 1.42	19.79 ± 1.78
AvFW	9.76 ± 0.07	NA	11.89 ± 1.21
AvFM	0.88 ± 0.02	NA	1.37 ± 0.22
AvSM	0.31 ± 0.01	0.37 ± 0.05	0.37 ± 0.06
EdRt	64.01 ± 0.69	NA	72.32 ± 5.34
FrSh	1.84 ± 0.02	NA	1.67 ± 0.19

Table 2. Phenotypic differentiation among *S. dulcificum* accession clusters. Figures in bold indicate values that are significantly greater than the overall means, while the rest of the figures indicate values that are significantly lower than the overall means for all accessions. NA: quantitative trait not significantly describing the cluster.

from Northern Africa. However, the magnitude of the variation observed in *S. dulcificum* is greater than in *V. paradoxa* and *A. spinosa* for tree-traits and lower in *S. dulcificum* than in *V. paradoxa* and *A. spinosa* for fruit-traits. Considering tree-traits variation for instance, the difference in magnitude between *V. paradoxa* and *S. dulcificum* might be attributed to the degree of heterogeneity of the habitats of each species. Fundamentally, *V. paradoxa* is found in similar habitat-types (i.e. wild and parkland), whereas *S. dulcificum* is found in more diversified habitats including farms and home gardens where it undergoes different management intensities. In home gardens, *S. dulcificum* benefits from management practices such as watering and occasional fertilization compared to farms where there is no special management provided to the species. This is likely to increase the magnitude of the differences observed in tree-traits between these two habitats. Information on tree age is rarely reported in studies tackling phenotypic diversity in perennial fruit trees^{69–71}. In this study, the age estimates for *S.*

Fruit-traits	Repeatability (r_{adj})	Uncertainty ^a		Likelihood ratio test
		S.E	Confidence interval	
AvFL	0.745	0.023	[0.700, 0.786]	<0.0001
AvFW	0.666	0.028	[0.608, 0.717]	<0.0001
AvFrM	0.883	0.013	[0.855, 0.906]	<0.0001
AvSM	0.821	0.019	[0.784, 0.857]	<0.0001
EdRt	0.015	0.002	[0.010, 0.018]	<0.0001
FrSh	0.580	0.030	[0.527, 0.647]	<0.0001

Table 3. Estimation of repeatability (adjusted) coefficient for six fruit-traits in *S. dulcificum* (S.E. standard error). ^aDetermined using 1000 parametric bootstraps.

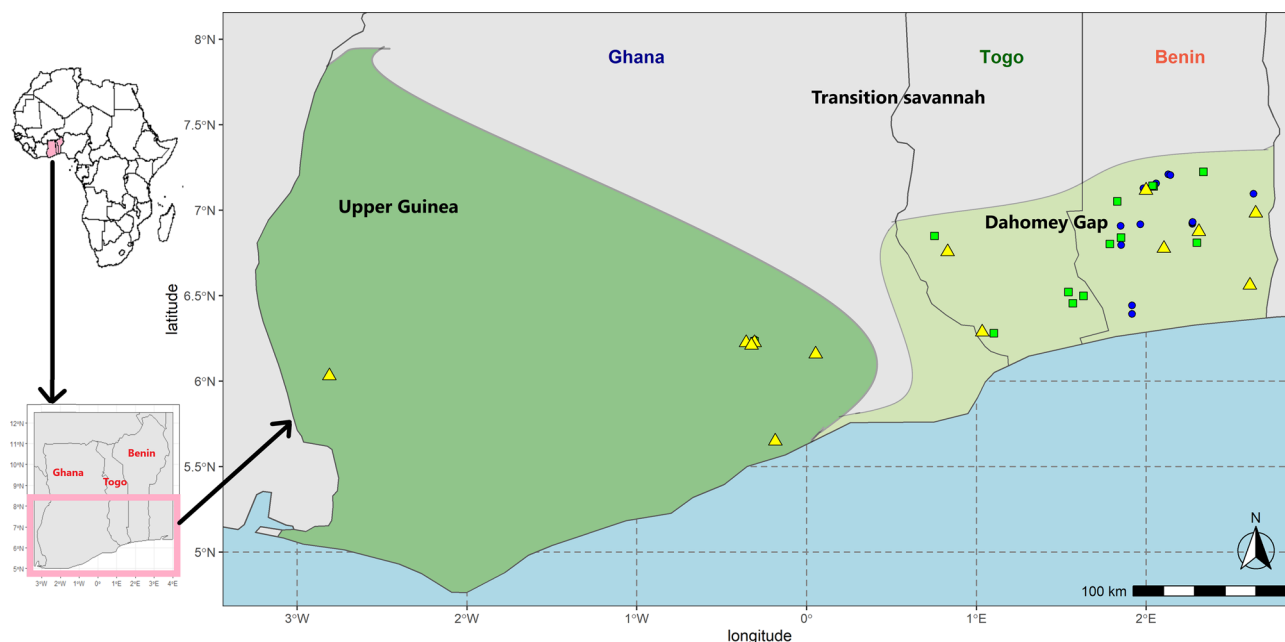


Figure 9. Spatial distribution of the core set (41 accessions: 8 accessions for UG and 33 for DG) extracted from 203 accessions of *Synsepalum dulcificum* in West Africa. Accessions are coloured following their belonging to the three clusters identified in this study: blue filled circle: Cluster 1 (least performant accessions with regards to the evaluated traits); green filled square: Cluster 2 (elite accessions for tree-traits); yellow filled triangle: Cluster 3 (potential elite accessions for fruit-traits). Map generated using the R environment (<https://www.r-project.org/>) (Version 3.6.2).

Traits	Range		CRI (%)	CV		VRi (%)	Mean		MDi (%)	Variance		VDi (%)	Means comparison (whole vs core)
	Whole	Core		Whole	Core		Whole	Core		Whole	Core		
DiamGr	64.50	62.68	97.17	53.34	64.05	120.07	19.12	22.33	14.37	104.17	204.63	49.08	$p=0.26$
PIHgt	6.64	6.64	100.00	36.59	45.88	125.38	3.47	3.70	6.21	1.613	2.87	43.90	$p=0.61$
CrDiam	8.05	7.05	87.57	38.20	47.92	125.44	4.23	4.23	0.00	2.61	4.11	36.37	$p=0.99$
AvFL	13.37	12.83	95.96	9.65	14.06	145.69	18.86	18.95	0.47	3.31	7.10	53.35	$p=0.89$
AvFW	10.91	10.91	100.00	12.08	18.58	153.80	10.59	10.82	2.12	1.64	4.04	59.46	$p=0.52$
AvFM	1.60	1.60	100.00	25.68	32.06	124.84	1.09	1.14	4.38	0.078	0.13	41.66	$p=0.43$
AvSM	0.44	0.43	97.72	20.00	27.07	135.35	0.35	0.35	0.00	0.004	0.008	46.70	$p=0.82$
EdRt	37.76	37.76	100.00	10.70	13.53	126.44	66.99	67.66	0.99	0.59	0.83	38.67	$p=0.44$
FrSh	1.33	1.33	100.00	10.05	15.07	149.95	1.79	1.78	0.56	0.03	0.07	53.42	$p=0.97$
			97.60			134.10			3.23			46.95	-

Table 4. Quality attributes of the core collection (Core) extracted form the whole collection (Whole). CV coefficient of variation, CRI coincidence range for the trait I, VRi variable range for the trait I, VDi variance difference for the trait I, MDi mean difference percentage for the trait i.

dulcificum confirmed the prominence of long-living species in the Sapotaceae family⁷². The absence of difference of age between home garden trees and those found on farms ruled out the hypothesis that tree age is a driver of tree-traits variation between these two habitats in *S. dulcificum*. This is furthermore supported by the overall poor correlation between tree age estimate and tree-traits.

This study provided for the first time information on *S. dulcificum* tree and crown diameter variation and expanded knowledge on the height growth potential of the species. *Synsepalum dulcificum* is a long-living microphanerophyte¹⁴ whose maximum height was reported to be around four meters². However, in favourable conditions this height may reach up to seven meters and the crown diameter up to eleven meters. Such phenotypes are found mainly in home gardens since individuals on farms are prone to high human pressures including harvesting of leaves and branches as well as debarking of stem for medicinal purposes³, thus impeding their normal growth.

Ayensu⁷³ successfully described the anatomy and morphology of *S. dulcificum* but made no mention of the fruit and seed morphometry. In an attempt to fill this gap, Lim⁷⁴ in his botanical description of the species indicated the miracle fruit being roughly 1.0 cm wide for a length ranging from 2.0 to 2.5 cm. Nevertheless, this current study revealed greater variations for both fruit length (1.30–2.64 cm) and fruit width (0.73–1.83 cm). Given the significant effect the provenance population exhibited on fruit-traits in this study, differences with Lim's⁷⁴ findings could be attributed to environmental variation. On the other hand, a difference in sampling effort coupled with potential genetic variation among the individuals measured might also explain the differences observed in the reported fruit width and length. While Lim⁷⁴ did not indicate the number of fruits used to compute the statistics, in this study nearly 8875 fruits were measured.

Due to the increasing interest in *S. dulcificum*, this study purposively targeted the fruit and its components. For instance, while the edible ratio is of great interest in medicinal, pharmaceutical, food and beverages industries¹⁹, the seed mass is of interest in pharmaceutical and cosmetics^{12,13,75}. Consequently, knowledge of the tree-to tree or population-to population variation for these specific traits would be useful for elite tree selection by the breeders and decision making for investment by stakeholders. This study reported for the first time the extent of the edible ratio and seed mass variation in the miracle berry and showed the influence of the ecological conditions on the fruit traits in general. An average edible ratio of roughly 67% was observed in *S. dulcificum*, which is one of the highest when compared to other tropical Africa indigenous fruit species such as *V. paradoxa* (61.33–62.0% of edible ratio)^{34,67}, *Balanites aegyptica* (L.) Delile. (44.9–50.65%)⁷⁶, and subtropical species such as *A. trifoliata* (23.52–27.63%)⁷⁷. More importantly, under favourable conditions, the edible ratio in *S. dulcificum* can increase up to 72% as observed for the set of individuals in the Upper Guinea forest block, a region characterized by higher rainfall and less sunlight exposure. This region was also favourable for heavier fruit production than the Dahomey Gap, which was characterized by drier conditions with savannah as main vegetation type. This strengthened arguments in favour of the previously highlighted beneficial effect of watering as well as the detrimental effect of sunlight exposure on *S. dulcificum* growth, reproductive and fruiting performances^{16,17}. The influence of climatic conditions (e.g. rainfall) on fruit traits in perennial species was also previously reported in *Azelia Africana* Smith⁷⁸ in which individuals in wet conditions performed better than those in dry conditions.

The poor correlation observed in this study between *S. dulcificum* tree-traits and fruit-traits is in line with findings in other Africa indigenous fruit trees species (e.g. *V. paradoxa*^{34,67}, *Sclerocaria birrea* (A. Rich) Hochst.⁷⁹), and suggests that tree architecture cannot be used to predict fruiting performance in adult individuals of *S. dulcificum*. Depending on the habitat, different pressures may interfere with the standard species growth, thus changing the trajectory of the relationship between tree morphology traits and any other group of traits. On the contrary, the significant and positive correlation between fruit mass and fruit width, and between fruit mass and the edible ratio, combined with the negative correlation between the edible ratio and seed mass are of significant interest for the development of a selection strategy in breeding of improved varieties of *S. dulcificum*. Such associations make it possible to simultaneously improve the fruit size and the edible ratio while improving the fruit mass. This could translate into an increased benefit and profitability for the food and beverages industries, which are more interested in the edible part of the fruit. On the other hand, it might be difficult to effectively select for seed mass using the fruit mass to benefit the cosmetics industry, thus a direct selection for high seed mass yielding genotypes would be required. Selection of dual-purpose miracle berry genotypes (e.g. for simultaneous use in both cosmetics and food and beverages industries) could be explored by developing a selection index that assigns weights of economic value for each trait (seed mass and edible ratio)⁸⁰. More importantly, the influence of the ecological conditions on the relationships between most of the functional fruit traits in *S. dulcificum* suggests the possibility to conduct environment-specific traits selection and production. It further suggests the need to involve multi-environments or locations in evaluation and development of improved varieties of *S. dulcificum*.

Genetic gain is fundamental in plant breeding program and is a function of the heritability, the phenotypic variation and the selection intensity⁸¹. A joint analysis of the variability (variance) and the heritability (coefficient of repeatability) of the six functional fruit-traits in this study indicated that under similar selection pressure, a greater genetic gain may be achieved in fruit mass, seed mass and fruit width against a lower gain in the edible ratio. In fact, by nature, breeding for traits expressed in a ratio is considered more laborious than simple traits⁷⁷, and this is well illustrated by the edible ratio of the miracle fruit that has an extremely low repeatability estimate ($R = 0.015$). On the contrary, this trait seemed to be more heritable in other fruit tree species (e.g. *A. trifoliata*³³) where the repeatability estimate for edible ratio was as high as 0.98. This suggests a potential species-specific pattern in the heritability of this important trait. Fruit mass and seed mass had high heritability estimates which suggests that these traits could be easily improved in the miracle berry once suitable testing environments are selected. While those traits were also found to be highly heritable in *A. trifoliata*³³, they were less heritable in *V. paradoxa*³⁴ and *Allanblackia floribunda* Oliver⁸² among other indigenous fruit tree species.

Whether high heritability combined with large phenotypic variability can ensure consequent genetic gain, a good baseline breeding population is desired to accelerate genetic gain⁸³. The development of a performing

breeding population in tree breeding first relies on elite tree identification⁸⁴. Findings from this study highlighted the existence of a diversity of breeding population pools in the miracle plant that can be tailored to specific breeding objectives. Fundamentally, it emerged that the Upper Guinea (UG) represented a source of promising parental lines that can be immediately exploited to initiate a breeding program in the miracle plant.

A core collection was developed, which in terms of size is approximately 20% of the size of the whole collection evaluated. This proportion is in line with the recommended size for a good core collection (10–30% of the initial collection)^{85,86}. Historically, four parameters have been used to evaluate the internal quality of a core collection: the coincidence range (CR), the mean difference percentage (MD), the variance difference (VD) and the variable rate (VR)⁶⁰. In this study, the core collection exhibited a coincidence rate CR that is by far higher than the threshold of CR > 80% required for a core collection to optimally represent an initial whole collection. While it was indicated that a well-constructed core collection should have no more than 20% of the traits deviating significantly from the whole collection, none of the traits in the core collection significantly deviated from the whole collection. The MD was also extremely low, suggesting that the core is a good representative of the whole set of accessions. Most important, the core collection encompassed individuals from all the sampled populations in this study, and all the clusters identified with a quasi-similar representativeness. Large VD and VR are preferred for a good collection³⁸ as they illustrate how much diversity of the whole collection is maintained in the core collection. An exceptionally high variable rate (> 100%) and a VD of 46.95% were observed, which together suggests that further diversity can be obtained from the core collection compared to the whole collection. Combined, these characteristics suggest that the regional core collection compiled is robust enough to represent the whole collection and can consequently be considered as a reliable working sample in implementing ex-situ conservation measures and developing a breeding population for association study and further genomic selection implementation⁸⁷ in the species. Practically, this core collection will ease and accelerate further evaluations (multi-location trials, genomic evaluation, metabolomic evaluation) and consequently shorten time to cultivar release. Besides, this core collection to be maintained as a living collection will help secure the reservoir of useful alleles the breeding process might need to tap in the future while favouring their co-evolution with environmental factors. An analysis of the core collection constitution indicated that all the three clusters identified in this study were represented in the Dahomey Gap zone whereas mainly one cluster was present in the Upper Guinea zone. This suggests the Dahomey Gap to be a centre of *S. dulcificum* diversity while providing an opportunity to properly elucidate phylogeographical relationships of the Dahomey Gap populations with the Upper Guinea, the Lower Guinea and the Congolian populations of the species.

In summary, this study evaluated the phenotypic variation in the miracle plant *S. dulcificum*, from its centre of origin, West Africa. A prominent effect of the provenance population was found on most of the phenotypic traits, with a better performance from the Upper Guinea forest accessions compared to their counterparts from the Dahomey Gap. Tree-traits were more variable than fruit-traits. Traits such as fruit mass, edible ratio and fruit width were highly correlated and exhibited high heritability estimates implying that simultaneous selection of the traits can be done. In addition, *S. dulcificum* individuals likely to serve as elite parental lines in the selection and development of high-yielding and pulp-dense cultivars were also identified. Finally, the successful development of a core collection of 41 accessions from the initial set of 203 accessions suggested the Dahomey Gap as a centre of diversity of the species. These findings pave the way for focused ex-situ conservation measures implementation as well as a cost-effective breeding program design for the species.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 13 May 2020; Accepted: 23 October 2020

Published online: 11 November 2020

References

- Kurihara, K. & Beidler, L. M. Taste-modifying protein from miracle fruit. *Science* **161**, 1241–1243 (1968).
- Achigan-Dako, E. G., Tchokponhoué, D. A., N'Danikou, S., Gebauer, J. & Vodouhè, R. S. Current knowledge and breeding perspectives for the miracle plant *Synsepalum dulcificum* (Schum. et Thonn.) Daniell. *Genet. Resour. Crop. Evol.* **62**, 465–476 (2015).
- Fandohan, A. B. *et al.* Usages traditionnels et valeur économique de *Synsepalum dulcificum* au Sud-Bénin. *Bois For. Trop.* **332**, 17–30 (2017).
- Oumorou, M., Dah-Dovonon, J., Aboh, B., Hounsoukaka, M. & Sinsin, B. Contribution à la conservation de *Synsepalum dulcificum*: régénération et importance socio-économique dans le département de l'ouémé (Bénin). *Ann. Sci. Agron.* **14**, 101–120 (2010).
- Rodrigues, J. F., da Silva Andrade, R., Bastos, S. C., Coelho, S. B. & Pinheiro, A. C. M. Miracle fruit: An alternative sugar substitute in sour beverages. *Appetite* **107**, 645–653 (2016).
- Andrade, A. C. *et al.* Effect of different quantities of miracle fruit on sour and bitter beverages. *LWT* **99**, 89–97 (2019).
- Swamy, K. B., Hadi, S. A., Sekaran, M. & Pichika, M. R. The clinical effects of *Synsepalum dulcificum*: a review. *J. Med. Food.* **17**, 1165–1169 (2014).
- Chen, C. C., Liu, I. M. & Cheng, J. T. Improvement of insulin resistance by miracle fruit (*Synsepalum dulcificum*) in fructose-rich chow-fed rats. *Phytother. Res.* **20**, 987–992 (2006).
- Han, Y. C., Wu, J. Y. & Wang, C. K. Modulatory effects of miracle fruit ethanolic extracts on glucose uptake through the insulin signaling pathway in C2C12 mouse myotubes cells. *Food Sci. Nutr.* **7**, 1035–1042 (2019).
- Obafemi, T. O., Akinmoladun, A. C., Olaleye, M. T., Agboade, S. O. & Onasanya, A. A. Antidiabetic potential of methanolic and flavonoid-rich leaf extracts of *Synsepalum dulcificum* in type 2 diabetic rats. *J. Ayurveda Integr. Med.* **8**, 238–246. <https://doi.org/10.1016/j.jaim.2017.01.008> (2017).
- Buckmire, R. & Francis, F. Pigments of miracle fruit, *Synsepalum dulcificum*, Schum, as potential food colorants. *J. Food Sci.* **43**, 908–911 (1978).

12. Del Campo, R., Zhang, Y. & Wakeford, C. Effect of miracle fruit (*Synsepalum dulcificum*) seed oil (MFSO*) on the measurable improvement of hair breakage in women with damaged hair: a randomized, double-blind, placebo-controlled, eight-month trial. *J. Clin. Aesthet. Dermat.* **10**, 39–48 (2017).
13. Gorin, S. *et al.* Beneficial effects of an investigational wristband containing *Synsepalum dulcificum* (miracle fruit) seed oil on the performance of hand and finger motor skills in healthy subjects: a randomized controlled preliminary study. *Phytother. Res.* **32**, 321–332 (2018).
14. Adomou, A. *Vegetation patterns and environmental gradients in Benin* PhD thesis, University of Wageningen, (2005).
15. Tchokponhoué, D. *et al.* Regeneration ability and seedling growth in the miracle plant *Synsepalum dulcificum* (Schumach. & Thonn.) Daniell. *Fruits* **73**, 13–21 (2018).
16. Tchokponhoué, D. A., N'Danikou, S., Houéto, J. S. & Achigan-Dako, E. G. Shade and nutrient-mediated phenotypic plasticity in the miracle plant *Synsepalum dulcificum* (Schumach. & Thonn.) Daniell. *Sci. Rep.* **9**(5137), 1–11 (2019).
17. Tchokponhoué, D. A., N'Danikou, S., Hale, I., Van Deynze, A. & Achigan-Dako, E. G. Early fruiting in *Synsepalum dulcificum* (Schumach. & Thonn.) Daniell juveniles induced by water and inorganic nutrient management. *F1000Res.* **6**, 1–17 (2017).
18. Choi, S. E. & Garza, J. Effects of different miracle fruit products on the sensory characteristics of different types of sour foods by descriptive analysis. *J. Food Sci.* **85**, 36–49 (2020).
19. Tafazolli, S. *et al.* Safety assessment of miraculin using in silico and in vitro digestibility analyses. *Food Chem. Toxicol.* **133**(110762), 1–10 (2019).
20. Chibuzor, I. A., Bukola, O., Adejoke, A. O. & Chidozie, O. P. Genetic assessment of the shrub *Synsepalum dulcificum* (Schumach. & Thonn.) Daniell in Nigeria using the Randomly Amplified Polymorphic DNA (RAPD). *Int. J. Genet. Genom.* **4**, 45–50 (2017).
21. Sogbohossou, E. D. *et al.* Association between vitamin content, plant morphology and geographical origin in a worldwide collection of the orphan crop *Gynandropsis gynandra* (Cleomaceae). *Planta* **250**, 933–947 (2019).
22. Singh, K., Sharma, Y. P. & Gairola, S. Morphological characterization of wild Rosa L. germplasm from the Western Himalaya, India. *Euphytica* **216**(41), 1–24. <https://doi.org/https://doi.org/10.1007/s10681-020-2567-2> (2020).
23. Sun, W. *et al.* Multivariate analysis reveals phenotypic diversity of *Euscaphis japonica* population. *PLoS ONE* **14**, 1. <https://doi.org/10.1371/journal.pone.0219046> (2019).
24. Yazdanpour, F., Khadivi, A. & Etemadi-Khah, A. Phenotypic characterization of black raspberry to select the promising genotypes. *Sci. Hortic. Amsterdam* **235**, 95–105. <https://doi.org/10.1016/j.scienta.2018.02.071> (2018).
25. Fereidoonfar, H., Salehi-Arjmand, H., Khadivi, A. & Akramian, M. Morphological variability of sumac (*Rhus coriaria* L.) germplasm using multivariate analysis. *Ind. Crops Prod.* **120**, 162–170. <https://doi.org/https://doi.org/10.1016/j.indcrop.2018.04.034> (2018).
26. Norouzi, E., Erfani-Moghadam, J., Fazeli, A. & Khadivi, A. Morphological variability within and among three species of *Ziziphus* genus using multivariate analysis. *Sci. Hortic. Amsterdam* **222**, 180–186. <https://doi.org/10.1016/j.scienta.2017.05.016> (2017).
27. Khadivi-Khub, A. & Anjam, K. Morphological characterization of *Prunus scoparia* using multivariate analysis. *Plant Syst. Evol.* **300**, 1361–1372 (2014).
28. Vihotogbé, R., van den Berg, R. G. & Sosef, M. S. Morphological characterization of African bush mango trees (*Irvingia* species) in West Africa. *Genet. Resour. Crop. Evol.* **60**, 1597–1614 (2013).
29. Ouborg, N. J. Integrating population genetics and conservation biology in the era of genomics. *Biol. Lett.* **6**, 3–6 (2010).
30. Martínez-García, P. J. *et al.* Predicting breeding values and genetic components using generalized linear mixed models for categorical and continuous traits in walnut (*Juglans regia*). *Tree Genet. Genomes* **13**, 109 (2017).
31. Falconer, D. S. *Introduction to quantitative genetics*. (Oliver And Boyd; Edinburgh; London, 1960).
32. Fonseca, C. E. L. d., Morais, F. M. d., Gonçalves, H. M., Aquino, F. d. G. & Rocha, F. S. Repeatability of fruit traits from two *Hancornia speciosa* populations from the core region of the Brazilian Cerrado. *Pesqui. Agropecu Bras.* **53**, 710–716 (2018).
33. Zou, S., Yao, X., Zhong, C., Zhao, T. & Huang, H. Genetic analysis of fruit traits and selection of superior clonal lines in *Akebia trifoliata* (Lardizabalaceae). *Euphytica* **214**(111), 1–9. <https://doi.org/10.1007/s10681-018-2198-z> (2018).
34. Sanou, H. *et al.* Phenotypic variation of agromorphological traits of the shea tree, *Vitellaria paradoxa* CF Gaertn., in Mali. *Genet. Resour. Crop. Evol.* **53**, 145–161 (2006).
35. Albuquerque, A. S., Bruckner, C. H., Cruz, C. D., Salomão, L. C. C. & Neves, J. C. L. Repeatability and correlations among peach physical traits. *Crop Breed. Appl. Biot.* **4**, 441–445 (2004).
36. Belaj, A. *et al.* Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* **8**, 365–378 (2012).
37. Le Cunff, L. *et al.* Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. *subsp. sativa*. *BMC Pl. Biol.* **8**:31, 1–12 (2008).
38. Mahmoodi, R. *et al.* Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. *Sci. Hortic. Amsterdam* **249**, 439–448 (2019).
39. Tchokponhoué, D. A., N'Danikou, S. & Achigan-Dako, E. G. A combination of approaches evidenced seed storage behaviour in the miracle berry *Synsepalum dulcificum* (Schumach. et Thonn.) Daniell. *BMC Pl. Biol.* **19**:117, 1–13 (2019).
40. Edesi, J., Tolonen, J., Ruotsalainen, A. L., Aspi, J. & Häggman, H. Cryopreservation enables long-term conservation of critically endangered species *Rubus humulifolius*. *Biodivers. Conserv.* **29**, 303–314. <https://doi.org/10.1007/s10531-019-01883-9> (2020).
41. Bharuth, V., Naidoo, C., Pammenter, N. W., Lamb, J. M. & Moodley, T. Responses to chilling of recalcitrant seeds of *Ekebergia capensis* from different provenances. *S. Afr. J. Bot.* **130**, 8–24 (2020).
42. Leal, M. E. *The African rain forest during the Last Glacial Maximum an archipelago of forests in a sea of grass*. (2004).
43. Swenson, U., Richardson, J. E. & Bartish, I. V. Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): evidence of generic polyphyly and extensive morphological homoplasy. *Cladistics* **24**, 1006–1031 (2008).
44. Juhé-Beaulaton, D. “Fèves”, “pois” et “grains” dans le golfe de Guinée : problèmes d’identification des plantes dans les sources historiques. in *Plantes et paysages d’Afrique, une histoire à explorer* (ed Chastanet M) 45–68 (1998).
45. Inglett, G. E. & May, J. F. Tropical plants with unusual taste properties. *Econ. Bot.* **22**, 326–331. <https://doi.org/10.1007/BF02908127> (1968).
46. Salzmann, U. & Hoelzmann, P. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* **15**, 190–199. <https://doi.org/10.1191/0959683605hl799rp> (2005).
47. Booth, A. The Niger, the Volta and the Dahomey Gap as geographic barriers. *Evolution* **12**, 48–62 (1958).
48. White, F. The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. (1983).
49. Niñez, V. Household gardens: theoretical and policy considerations. *Agr. Syst.* **23**, 167–186. [https://doi.org/10.1016/0308-521X\(87\)90064-3](https://doi.org/10.1016/0308-521X(87)90064-3) (1987).
50. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2019).
51. Peterson, B. G. *et al.* ‘Performance analytics’: econometric tools for performance and risk analysis. R Team Cooperation. (2018).
52. Warton, D. I., Duursma, R. A., Falster, D. S. & Taskinen, S. smatr 3—an R package for estimation and inference about allometric lines. *Methods Ecol. Evol.* **3**, 257–259 (2012).
53. Moon, K.-W. *Interactive plot*. in *Learn ggplot2 Using Shiny App* (ed Keon-Woong Moon) 295–347 (Springer, 2016).
54. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

55. YiLan, L. & RuTong, Z. clustertend: Check the Clustering tendency. *R package version 1* (2015).
56. Kassambara, A. & Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. *R package version 1*, 4 (2017).
57. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644 (2017).
58. De Beukelaer, H., Davenport, G. F. & Fack, V. Core Hunter 3: flexible core subset selection. *BMC Bioinformatics* **19**(203), 1–12 (2018).
59. Kim, K.-W. *et al.* PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **23**, 2155–2162 (2007).
60. Hu, J., Zhu, J. & Xu, H. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **101**, 264–268 (2000).
61. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).
62. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439–446 (2018).
63. South, A. Rworldxtra: Country boundaries at high resolution. *R package version 1* (2012).
64. South, A. rnaturalearth: World map data from Natural Earth. *R package version 0.1.0* (2017).
65. Salako, V. K. *et al.* Home gardens: an assessment of their biodiversity and potential contribution to conservation of threatened species and crop wild relatives in Benin. *Genet. Resour. Crop. Evol.* **61**, 313 (2014).
66. van Rompaey, R. S. *Forest gradients in West Africa: a spatial gradient analysis*, Wageningen, (1993).
67. Gwali, S. *et al.* Morphological variation among shea tree (*Vitellaria paradoxa* subsp. *nilotica*) 'ethnovarieties' in Uganda. *Genet. Resour. Crop. Evol.* **59**, 1883–1898 (2012).
68. Metougui, M. L., Mokhtari, M., Maughan, P. J., Jellen, E. N. & Benlhabib, O. Morphological variability, heritability and correlation studies within an argan tree population (*Argania spinosa* (L.) Skeels) preserved in situ. *Int. J. Agr. For.* **7**, 42–51 (2017).
69. Tsobeng, A. *et al.* Tree-to-tree variation in fruit of three populations of *Trichoscypha acuminata* in Cameroon. *Sci. Afr.* **7**, 1–12 (2020).
70. Omondi, M. *et al.* Fruit morphological diversity and productivity of baobab (*Adansonia digitata* L.) in coastal and lower eastern Kenya. *For. Trees Livelihoods* **28**, 266–280 (2019).
71. Abdulai, I., Krutovsky, K. V. & Finkeldey, R. Morphological and genetic diversity of shea tree (*Vitellaria paradoxa*) in the savannah regions of Ghana. *Genet Res Crop Evol* **64**, 1253–1268 (2017).
72. Karambiri, M., Elias, M., Vinceti, B. & Grosse, A. Exploring local knowledge and preferences for shea (*Vitellaria pradoxa*) ethnovarieties in Southwest Burkina-Faso through a gender and ethnic lens. *For. Trees Livelihoods* **26**, 13–28 (2016).
73. Ayensu, E. S. Morphology and anatomy of *Synsepalum dulcificum* (Sapotaceae). *Bot. J. Linn. Soc.* **65**, 179–187 (1972).
74. Lim, T. K. in *Edible medicinal and non-medicinal plants* Vol. 6 (ed T.K. Lim) 146–150 (Springer, Dordrecht, 2013).
75. Huang, W., Chung, H. Y., Xuan, W., Wang, G. & Li, Y. The cholesterol-lowering activity of miracle fruit (*Synsepalum dulcificum*). *J. Food Biochem.* **1**, e13185. <https://doi.org/10.1111/jfbc.13185> (2020).
76. Ahmed, A. A. O. *et al.* Tree-to-tree variability in fruits and kernels of a *Balanites aegyptiaca* (L.) Del. population grown in Sudan. *Trees* **34**, 111–119 (2020).
77. Zou, S., Yao, X., Zhong, C., Zhao, T. & Huang, H. Effectiveness of recurrent selection in *Akebia trifoliata* (Lardizabalaceae) breeding. *Sci. Hortic. Amsterdam* **246**, 79–85 (2019).
78. Houehanou, T. D. *et al.* Morphological trait variation and relationships of *Azzeria africana* Sm. caused by climatic conditions and anthropogenic disturbance in Benin (West Africa). *Genet. Resour. Crop. Evol.* **66**, 1091–1105 (2019).
79. Gouwakinnou, G. N., Assogbadjo, A. E., Lykke, A. M. & Sinsin, B. Phenotypic variations in fruits and selection potential in *Sclerocarya birrea* subsp. *birrea*. *Sci. Hortic. Amsterdam* **129**, 777–783 (2011).
80. Cotterill, P. P. & Dean, C. A. *Successful tree breeding with index selection* (CSIRO, Division of Forestry and Forest Products, 1990).
81. Zobel, B. & Talbert, J. *Applied forest tree improvement*. (John Wiley & Sons, 1984).
82. Atangana, A. R. *et al.* Tree-to-tree variation in stearic and oleic acid content in seed fat from *Allanblackia floribunda* from wild stands: potential for tree breeding. *Food Chem.* **126**, 1579–1585 (2011).
83. Shelbourne, C. Genetic gains from different kinds of breeding population and seed or plant production population. *S. Afr. For. J.* **160**, 49–65 (1992).
84. Leakey, R. & Page, T. The 'ideotype concept' and its application to the selection of cultivars of trees providing agroforestry tree products. *For. Trees Livelihoods* **16**, 5–16 (2006).
85. Bhattacharjee, R., Khairwal, I., Bramel, P. J. & Reddy, K. Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica* **155**, 35–45 (2007).
86. Escribano, P., Viruel, M. & Hormaza, J. in *XII EUCARPIA Symposium on Fruit Breeding and Genetics 814*. 67–70.
87. Duan, H. *et al.* Genetic characterization of Chinese fir from six provinces in southern China and construction of a core collection. *Sci. Rep.* **7**, 1–10 (2017).

Acknowledgements

Dédéou A. Tchokponhoué (D.A.T), is a PhD candidate at the University of KwaZulu-Natal, South Africa under the project grant number 2016-2988 on “Enhancing training and research mobility for novel crops breeding in Africa (MoBreed)” funded by the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission under the “Intra-Africa Academic Mobility Scheme”. The project provided a scholarship for 48 months’ research mobility and a partial research grant to the first Author whose home institution is the University of Abomey-Calavi in Republic of Benin. We are also indebted to Rachidi Francisco, Fernand Sohindji, Felix Satognon, and Komi Emile Sogbedji for their help during our field work in Benin and Togo, and to Mathieu Ayenan, Xavier Matro, and Yakubu Abdulai Nasr for their invaluable assistance during field work in Ghana. We pay a special tribute to late Florian Ahloumessou and J. Anglow for their unforgettable help during this study.

Author contributions

D.A.T, E.A.D and J.S conceived the study. D.A.T, J.H and D.N conducted the field work. D.A.T analysed the data and wrote the manuscript. E.A.D, S.N, D.N, A.O.O, N.V.F.H, R.K and J.S reviewed the manuscript. E.A.D and J.S supervised the study. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-76103-4>.

Correspondence and requests for materials should be addressed to D.A.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020