JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12858

Research Article

Integrative taxonomy and phylogeny of leafless *Vanilla* orchids from the South-West Indian Ocean region reveal two new Malagasy species

Cathucia F. Andriamihaja[1,2], Alemao Botomanga[2], Chantal Misandeau[3], Aro V. Ramarosandratana[2], Michel Grisoni[4], Denis Da Silva[1], Thierry Pailler[1], Vololoniaina H. Jeannoda[2], and Pascale Besse[1]*

[1]Université de la Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France
[2]Department of Plant Biology and Ecology, University of Antananarivo, 101 Antananarivo, Madagascar
[3]ADAFAM, 3 rue des cocotiers, F-97436 Saint-Leu, La Réunion, France
[4]CIRAD, UMR PVBMT, 501 Toamasina, Madagascar
*Author for correspondence. E-mail: pascale.besse@univ-reunion.fr
Received 28 May 2021; Accepted 6 May 2022; Article first published online 14 May 2022

Abstract The leafless *Vanilla* species complex from the South-West Indian Ocean (SWIO) region has long been a taxonomic challenge, due to limited patterns of morphological differentiation and an absence of variation within chloroplast sequences. This complex includes seven known morphospecies: *V. madagascariensis*, *V. bosseri*, *V. decaryana*, and *V. perrieri* endemic to Madagascar, *V. humblotii* presumed as endemic to the Comoros Archipelago, but also present in Madagascar, *V. roscheri* from the East African coast, and *V. phalaenopsis* endemic to Seychelles. A previous population genetic study using microsatellite markers allowed us to distinguish, in addition to the five recognized Malagasy taxa, two other genetic clusters present in the East of the island. An integrative taxonomy approach was therefore conducted by combining microsatellite and morphological data used in the previous study with new data sets, and by adding ITS sequencing data, to validate the taxonomic level of these Malagasy genetic clusters and unravel phylogenetic relationships between SWIO species. As a result, based on phylogenetic, genotypic and morphological evidence, nine species were discriminated in the SWIO region, including seven in Madagascar, with two new eastern species. The leafless *Vanilla* group originated and diversified in Madagascar, from an ancestor of African descent, with three subsequent independent colonization events from Madagascar to the other territories of SWIO within the two main lineages (white versus yellow flower species). The new Malagasy species, *V. allorgeae* Andriamihaja & Pailler sp. nov., and *V. atsinananensis* Andriamihaja & Pailler sp. nov., are described and a new identification key is proposed.

Key words: ITS, Madagascar, microsatellites, morphology, phylogeography, species delimitation, systematics.

1 Introduction

All biological scientists and conservationists strongly need to properly define the taxonomic units they work on. For many disciplines, such as botany, species are the fundamental units on which studies will be based. A species is denoted on the basis of many criteria of which resemblance between members is the most widely accepted (Aldhebiani, 2018). These criteria used to define species boundaries have changed over time due essentially to the rapid development of new technologies and the birth of evolutionary concepts (Rouhan & Gaudeul, 2021). Linnaeus was one of the first systematists to propose a species definition. According to him, members of a species have relatively constant characters and tend to breed (Larson, 1968). Years later, de Candolle proposed another definition which integrated, in addition to the criteria of resemblance and reproduction, a condition of fertile offspring of species members (de Candolle, 1813). In the 19th century, Darwin's theory of evolution revolutionized species concepts. According to the book "On the origin of species," Darwin's species concept can be formulated as follows: a fundamental unit of evolution that is constituted by populations of organisms united by common descent and that have been differentiated from the other species after natural selection by adaptation to environmental conditions (Darwin, 1859; Stamos, 2013). On the basis of these earlier theories, many species concepts have been

proposed from the 20th century (de Queiroz, 2005, 2007; Aldhebiani, 2018; Rouhan & Gaudeul, 2021), among which the most widely held are the following: biological (BSC) (Mayr, 1942), evolutionary (EvSC) (Simpson, 1951; Wiley, 1978), ecological (EcSC) (Van Valen, 1976), phenetic (PheSC) (Sokal, 1986), genotypic (GSC) (Mallet, 1995), and phylogenetic species concept (PhySC) (de Queiroz, 2005, 2007; Aldhebiani, 2018).

All these species concepts are complementary but also have some limitations. The problem of species delimitation is amplified when species have recently diverged or when they are cryptic or closely related (de Queiroz, 2005; Barley et al., 2013; Pinheiro et al., 2018). Indeed, characters resulting from completed speciation may not all be observable in a recently diverged species (de Queiroz, 2005). For cryptic and closely related species, phenetic (morphology, cytology, phytochemistry, anatomy) traits may not be able to distinguish them even if they formed distinct genetic groups (Bickford et al., 2007). Thus, in such cases, the concept of morphological or phenetic species is difficult to apply, especially when the number of phenetic markers is limited. There are also some cases (e.g., adaptive radiation) where genetic divergence between species is low although morphological differentiation is high (Barley et al., 2013). Given the existence of all these definitions of species, which are sometimes incompatible, de Queiroz (2007) proposed the Unified Species Concept (USC). The USC retains only the common elements of all existing species concepts as delimitation criteria. According to the USC, a species is a lineage that evolves separately from other such lineages (de Queiroz, 2007). Today, it is recognized that the best way to circumscribe species is to use several methods (e.g., morphological analyses, phylogenetics, genetic structuring) and consider different species concepts (Rouhan & Gaudeul, 2021). This principle of taxonomy, called integrative taxonomy (Dayrat, 2005), is based on the idea that species delimitation using different data sources and methods is more rigorous than using a single type of data (Dayrat, 2005; Schlick-Steiner et al., 2010). By combining different species concepts, the limitations of each species concept are offset by the advantages of the others, making integrative taxonomy a now widely used approach (Dayrat, 2005). Hence, integrative taxonomy has been shown as being most efficient for the delimitation of recently diverged, cryptic or closely related species (e.g., Dayrat, 2005; Duminil et al., 2012; Pessoa et al., 2012; Pansarin & Ferreira, 2019).

In this paper, we applied this strategy to resolve the taxonomic puzzle associated with an evolutionarily young group of leafless *Vanilla* Plum. ex Mill. species (Orchidaceae) from the South-West Indian Ocean region (SWIO) (Madagascar, Comoros archipelago, Seychelles, Eastern coast of Africa) (Andriamihaja et al., 2020). Seven leafless species are described in this region, but their floral morphology is very similar. Our recent population genetic work suggested that there might be more species than the five initially described for Madagascar (Andriamihaja et al., 2020, 2021), and nothing is known about the phylogenetic relationships between the Malagasy species and sister species from the SWIO region. Some species are allopatric (species found only in Madagascar, Seychelles, Comoros or East of Africa, and species found only in the West or the East of Madagascar), but *Vanilla* species from western Madagascar often occur in sympatry (Andriamihaja et al., 2020, 2021). Hybrid individuals were detected in areas where at least two species were found to occur sympatrically (Andriamihaja et al., 2021). As interspecific hybridization has also already been observed between several *Vanilla* species (either naturally or through breeding), even between the most genetically distant (Nielsen, 2000; Divakaran et al., 2006; Minoo et al., 2008), adopting the BSC would call into question many species so far described in the genus. Indeed, strict reproductive isolation, on which BSC is based, is rarely achieved in the case of closely related species living in sympatry (Rieseberg & Carney, 1998; Nielsen & Siegismund, 1999; Marchelli & Gallo, 2001). In addition, the fertility of interspecific hybrids was demonstrated, as in the case of the species *V. x tahitensis* J.W.Moore resulting from the hybridization of *V. planifolia* Andrews and *V. odorata* C.Presl (Lubinsky et al., 2008). The EcSC, defining species based on their occurrence in distinct adaptive zones, called niches (Van Valen, 1976), is also not applicable to *Vanilla* species because they are often sympatric (Nielsen, 2000; Arenas & Dressler, 2010; Andriamihaja et al., 2020), and most likely occupy the same ecological niches, using the same resources. Since it is challenging to test how Malagasy leafless *Vanilla* species are differing in their evolutionary trajectories, given that they are sympatric and can hybridize, the EvSC concept, though theoretically being the best-suited concept, is practically difficult to apply.

In the absence of detailed evidence regarding evolutionary trajectories, we therefore decided to consider the application of the Phylogenetic (PhySC), Genotypic (GSC), and Phenetic (PheSC) concepts for the *Vanilla* species under survey, particularly those from Madagascar. PhySC defines a species as a set of organisms belonging to a single monophyletic group (de Queiroz, 2005, 2007; Aldhebiani, 2018). SWIO leafless species form a relatively recent (Mean age: 4.4 Myr, 95% HPD (Highest posterior density): [0.3–10.6]) monophyletic group as shown using chloroplast DNA markers (*rbcL*, *psaB*, *psbB*, *psbC*) (Bouetard et al., 2010). These cpDNA markers were however unable to resolve species under PhySC within this recent group because of the lack of sequence variation. Indeed, it is crucial to choose the adequate molecular marker in phylogenetic methods depending on the taxonomic level one wants to delimit (Besse, 2014a). As demonstrated in barcoding experiments for *Vanilla* species (Besse et al., 2021), the sequencing of the nuclear Internal Transcribed Spacer (ITS) region of the ribosomal DNA is necessary to reach a sufficient level of discrimination, and will therefore be used to apply the PhySC to this recent group of species and resolve the phylogenetic relationships within this clade. GSC considers two separate clusters as two different species based on genotypic clustering analyses (Mallet, 1995) using genotypic markers, such as allozymes, microsatellites, and single nucleotide polymorphism (SNP) (Chauhan & Rajiv, 2010). Previous thorough microsatellite markers' (SSR or Simple Sequence Repeats) study at the population level combined with morphological evidence allowed to resolve some of the Malagasy species (seven genetic groups were revealed for five described species) (Andriamihaja et al., 2021), demonstrating the power of combined GSC and PheSC analyses,

which will therefore be applied in complement to the PhySC to further resolve species delimitation. According to Mallet (1995), the GSC approach is powerful in orchids where interspecific hybridization is frequently observed.

We therefore initiated this study to answer major questions that were raised by our previous SSR study of Madagascan population differentiation (Andriamihaja et al., 2021). We studied new accessions of leafless species from the SWIO group (Seychelles, East African coast, Comoros archipelago) in addition to Madagascan accessions, using ITS sequencing and SSR genotyping (outgroups consisting of leafy African and Malagasy species were also added in the ITS analysis). We also performed new field prospections that enabled the morphological characterization of the genetic group from Manompana (Madagascar), which was not studied earlier (Andriamihaja et al., 2021). This allowed us to unravel the phylogenetic relationships and biogeography of leafless *Vanilla* species distributed in the SWIO as well as to resolve, using an integrative taxonomy approach, the number of species present in Madagascar, with the discovery and description of two new leafless *Vanilla* species in the East of Madagascar.

2 Material and Methods

2.1 Sampling design

This study included 38 leafless *Vanilla* samples collected during field surveys carried out in Madagascar (2017–2019), as well as eight samples from the SWIO region (Mayotte, Seychelles, Zanzibar and South Africa) and three leafless samples coming from botanical gardens and maintained in the Biological Resources Center (BRC) Vatel collection in Reunion island (Tables 1, S1). Samples were chosen to cover all distribution areas of the seven leafless *Vanilla* species indicated in the literature (Portères, 1954; Cribb & Hermans, 2009; Allorge-Boiteau, 2013; Andriamihaja et al., 2020) and to represent all genetic groups from Madagascar revealed in a previous study (Andriamihaja et al., 2021) (Fig. 1). Fieldwork and sampling in Madagascar were conducted according to permit numbers 201/17/MEEF/SG/DGF/DSAP/SCB.Re and 205/17/MEEF/SG/DGF/DSAP/SCB.Re delivered by the Ministry of Environment and Sustainable Development. In addition, 14 leafy specimens from Africa and Madagascar and one outgroup from America were also studied to complete the data set. All the accessions studied are registered and stored in BRC Vatel.

2.2 Phylogenetic analyses and divergence estimation

The first analysis carried out consisted in building a phylogenetic tree. DNA was extracted using the Dneasy plant Mini kit (Qiagen, Hilden, Germany). Because preliminary assessments showed that chloroplast genes (*rbcl*, *psaB*, *psbB*, *psbC*, and *matK*) were not variable enough to discriminate this group of leafless *Vanilla* species (Bouetard et al., 2010, and data not shown), we used the nuclear ITS region. As in many plant species (China Plant BOL Group et al., 2011), the

Table 1 List of specimens used in this study, with classification and geographical origin

Classification*	Species	Geographical origin	Wild population or collection	*N*
Xanata Tethya /Aphyllae	*V.* sp1**	Madagascar	AND	4
	V. sp2**	Madagascar	MAN	5
	V. perrieri Schltr.**	Madagascar	AKL, BFD, BTM	5
	V. bosseri L.Allorge**	Madagascar	AKL, BBL, KRD, ZVB	6
	V. madagascariensis Rolfe**	Madagascar	AMB, ANK, ANM, CMK	4
	V. decaryana H.Perrier**	Madagascar	ANJ, ATD, KRM	4
	V. humblotii Rchb.f.**	Madagascar	MDF	4
	Unknown**	Madagascar	EST	1
	Unknown**	Madagascar	Diego (trad. Doctor)	2
	Unknown**	Madagascar	Tulear (Antsokay arboretum)	1
	Unknown**	Madagascar	Vohemar (Floribis collection)	2
	V. humblotii Rchb.f.**	Mayotte	Moya, Sohoa, Choungi	3
	V. phalaenopsis Rchb.f.**	Seychelles	Mahé, JB Lyon	3
	V. roscheri Rchb.f.**	East Africa	South Africa, Zanzibar	3
	Unknown**	SWOI unspecified	MNHN Cherbourg, RBG Kew	2
Xanata Tethya/Foliosae	*V. africana* Lindl.***	Africa	MNHN Cherbourg	1
	V. crenulata Rolfe***	Africa	MNHN Cherbourg	3
	V. françoisii H.Perrier***	Madagascar	MAN	7
	V. polylepis Summerh.***	Africa	RBG Kew	1
	V. imperialis Kraenzl.***	Africa	MNHN Cherbourg, Denmark	2
Xanata Xanata/Foliosae	*V. palmarum* Salzm. ex Lindl.***	America	Brazil	1
			total	64

Species names were defined according to genetic structuring analyses in Andriamihaja et al. (2021), morphological description and geographical location in the literature (Portères, 1954; Cribb & Hermans, 2009; Allorge-Boiteau, 2013). Accession voucher numbers are detailed in Table S1; *Sub genus and section according to Soto Arenas & Cribb (2010)/section according to Portères (1954); **Leafless species; ***Leafy species. *N* represents the number of individuals studied.

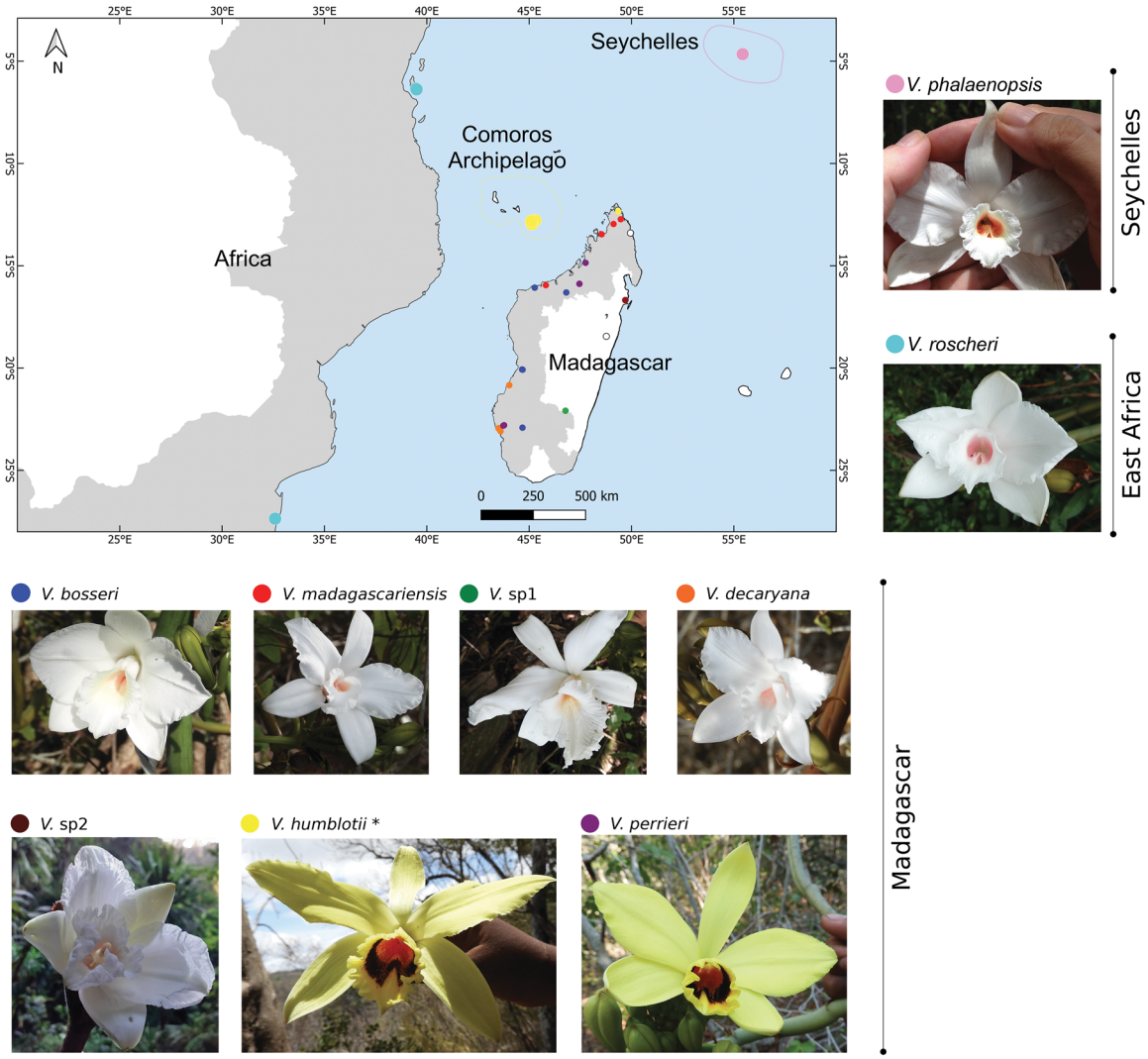

Fig. 1. Sampling map. Circles indicate sample locations. Species names were refined according to a previous study on genetic structuring (Andriamihaja et al., 2021), botanical description and geographical distribution (Portères, 1954; Cribb & Hermans, 2009; Soto Arenas & Cribb, 2010; Allorge-Boiteau, 2013). White circles surrounded by black lines correspond to unspecified samples. *V.* sp1 and *V.* sp2 represent samples that could not be matched to known species on the basis of their morphological characteristics and genetic grouping (Andriamihaja et al., 2021). Distribution of leafless *Vanilla* species in the South-West Indian Ocean region are highlighted in gray. *: *V. humblotii* is both present in Comoros Archipelago and Madagascar (Portères, 1954; Cribb & Hermans, 2009). The map was created using a world shapefile data set available online (https://public.opendatasoft.com) and QGIS software. The final figure with photos was produced using Inkscape software. (Photo *V. bosseri*; *V. madagascariensis*; *V.* sp1; *V. decaryana*: Cathucia F. Andriamihaja; *V. perrieri*: Hoby N. Nomenjanahary, *V.* sp2: Michel Grisoni, *V. humblotii*: Johnson G. Andrianantenaina, *V. roscheri*: Rodolphe Gigant, *V. phalaenopsis*: Pascale Besse).

ITS region is among the most used and efficient markers in *Vanilla* phylogenetics (Lubinsky et al., 2008; Arenas & Dressler, 2010; Azofeifa-Bolaños et al., 2017; Villanueva-Viramontes et al., 2017), as it shows more variation at several taxonomic levels, including species and subspecies, and can be especially helpful for closely related species delimitation (Besse, 2014a; Besse et al., 2021). The ITS region was amplified via polymerase chain reaction (PCR) using the primers *ITS 17SE* and *ITS 26SE* (Besse, 2014b). Amplification reactions included 20 µL of Gotaq master Mix 2X (Promega, Madison, WI, USA), 0.8 µL each of forward and reverse primers (10 µM), 14.4 µL of H_2O and 4 µL of DNA (10 ng/µL). The PCR cycling profile started with an initial 3 min denaturation step at 95 °C, followed by 30 cycles each with 30 s of denaturation at 94 °C, 30 s of annealing at 60 °C and 45 s of elongation at 72 °C, and the PCR program ended with a final 10 min elongation step at 72 °C. The amplification products were sent to Genoscreen (Lille, France) for sequencing. Amplicons were sequenced in both directions (reverse/forward). The 54 sequences obtained have been submitted to the GenBank database under accession numbers MW879465- MW879518, 10 others are issued from previous work (Besse et al., 2021) (Table S1). Nucleotide sequences were aligned using the MEGAX software

(Kumar et al., 2018) and checked manually to avoid possible errors. Analyses were performed using four phylogenetic methods: Neighbor-joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Analysis. The first three were implemented with MEGAX software (Kumar et al., 2018) and the last one was done under BEAST 1.10.4 (Drummond et al., 2012). The nucleotide substitution model by Motoo Kimura + gamma distribution (K80 + G) was selected, as it was identified to be the best-fitting model of molecular evolution according to the Bayesian Information Criterion (BIC) and decision-theoretic performance-based approach (DT) calculated with jModelTest 2.1.10 software (Posada, 2008). Two separate runs of the Bayesian Monte Carlo Markov Chain (MCMC) were performed for 20 000 000 generations each, with a tree sampled every 2000 generations. The calibrated Yule of speciation (Heled & Drummond, 2012) was chosen as the speciation process for all analyses. Strict and uncorrelated (exponential and lognormal) relaxed clock models were tested with the K80 + G substitution model. But, according to the test of molecular clocks using the ML method in MEGAX, the null hypothesis of equal evolutionary rate throughout the tree was rejected at a 5% significance level ($P = 6.220E^{-0.75}$). Moreover, a comparison between the marginal likelihood of the three clock models tested (under Tracer 1.7.1 [Rambaut et al., 2018]) via path sampling approach in BEAST indicated that uncorrelated exponential relaxed clock was the most appropriate model. Therefore, the uncorrelated exponential relaxed molecular clock approach implemented in BEAST 1.10.4 was retained for divergence time estimation of leafless *Vanilla* species from the SWIO region. Log files were analyzed with Tracer 1.7.1. (Rambaut et al., 2018), to assess if each parameter converged on a stationary distribution and confirm that the combined effective sample sizes (ESS) for all parameters were larger than 200, as recommended by Drummond et al. (2007), to ensure that the MCMC chain has been run long enough to get a valid estimate of the parameter. For the calibration, we used the divergence date of the SWIO leafless *Vanilla* group inferred by Bouetard et al. (2010). This calibration point was set at a normal prior age of 4.4 Myr with a standard deviation of 2.3 Myr (95% HPD: 0.6–8.2) close to the estimated dates from Bouetard et al. (2010). LogCombiner v1.10.4 (BEAST package) was used to combine all resulting trees from the two runs. Then, TreeAnnotator 1.10.4 (BEAST package) allowed to summarize the sample of post-burning trees as a single tree after discarding the first 10% as burn-in. The summarized tree was finally viewed and constructed using Figtree 1.4.4 (Rambaut, http://tree.bio.ed.ac.uk/software/figtree/) and Inkscape (https://inkscape.org/fr/).

2.3 Taxonomic diagnosis aided by a population genetics approach

Clades obtained from phylogenetic analysis were further diagnosed using a population genetics approach. This method is often recommended for the delimitation of closely related species (Duminil & Di Michele, 2009; Wang et al., 2019) as they can be considered to be highly differentiated populations (Queloz et al., 2010).

For that, microsatellite genotyping data of the 49 leafless *Vanilla* samples were collected. A subset of the data was issued from our previous studies (Gigant et al., 2014, 2016; Andriamihaja et al., 2021) and 14 new samples were studied (Table S1) at seven microsatellite loci (*mVroCIR04*, *mVroCIR05*, *mVhuCIR03*, *mVhuCIR04*, *mVhuCIR06*, *mVhuCIR07*, *mVhuCIR08*) designed by Gigant et al. (2011a). PCR amplification and microsatellite genotyping were performed as described previously (Andriamihaja et al., 2021). To avoid bias due to batch effect, all microsatellite profiles generated from capillary electrophoresis were carefully checked and allele sizes were scored using GENEIOUS v.7.0 (Olsen et al., 2014), whether they were generated in previous studies or were specific to this study. Assignment of individuals to genetic clusters on the basis of multi-locus genotypes was performed using Bayesian clustering analyses implemented in the STRUCTURE software (Pritchard et al., 2000). The model with no admixture and correlated allele frequencies was chosen. Indeed, according to Porras-Hurtado et al. (2013), the correlated allele frequencies model is more powerful to discriminate closely related populations. In the case of low levels of correlation across populations, correlated allele frequencies should yield the same result as the independent allele frequencies model (Porras-Hurtado et al., 2013). The number of genetic clusters (*K*) was set from 1 to 11 with 10 replicate runs for each *K*-value, 100 000 burn-in and 1 000 000 Markov chain Monte Carlo (MCMC) replications. The most probable number of K was determined by plotting the ad hoc criteria Ln $P(X \mid K)$ (Pritchard et al., 2000) and the summary likelihood statistics ΔK (Evanno et al., 2005). The best *K* corresponds normally to the highest value of mean Ln $P(X \mid K)$ and ΔK (Pritchard et al., 2000; Evanno et al., 2005). Bar plot of assignment test obtained from STRUCTURE was generated via POPHELPER package in R (Francis, 2017). Population structure was also estimated using a Principal Coordinate Analysis (PCoA) implemented in the adegenet 2.0.0 package in R (Jombart, 2016). Gene variation of each genetic group was assessed by calculating three indices: the total number of alleles (Na), the mean number of alleles per locus (Al), and the number of private alleles (Pa) using GenAlex 6.5 (Peakall & Smouse, 2012). The mean allelic richness (Ar) was computed using FSAT 2.9.4 (Goudet, 1995).

2.4 Morphological measures in Madagascan species

In *Vanilla* species, the morphological differences are mainly based on floral traits (Arenas & Dressler, 2010; Soto Arenas & Cribb, 2010). So, floral characters of all Malagasy genetic entities identified from molecular analyses were compared. For that, the results of previous work on eight populations representing six genetic groups in Madagascar were re-examined (Andriamihaja et al., 2021). In addition, fieldwork was carried out in Manompana (North-East of Madagascar) in December 2021 to complete the floral data set. Flowers were sampled from homogeneous populations identified from the results in Andriamihaja et al. (2021), that is, populations where only one species can be observed. The 10 collected flowers were measured and scanned using a canoScan LiDE 120 in accordance with the previous methods (Andriamihaja et al., 2021), so as to make the comparison possible. Color variables were extracted after transforming scanned floral images into CIE (*L*a*b**) color space under R 3.5.1 software (R Core Team, 2019) according to the script of Kendal et al. (2013). The scanner was equipped with a

measuring scale that facilitated floral measurement. Consequently, 17 floral traits and nine color variables were assessed (Andriamihaja et al., 2021). The floral measurements were obtained either directly *in situ* using a scale and a caliper, or after processing the images under IMAGE J software (Abràmoff et al., 2004). The 26 morphological measurements obtained from the 10 collected samples were added to those used in the previous study (Andriamihaja et al., 2021) to reperform a Principal Component Analysis (PCA) using FACTOMINER (Lê et al., 2008). Means and standard deviations of the 10 most explanatory variables from PCA were calculated. Multivariate analysis of variance (MANOVA) was conducted using R software to test the statistical significance of floral distinctiveness based on the priori groupings of genetic structuring results. Differences between genetic groups for these floral characters were determined using a Bonferroni corrected pairwise t-test under R software. The stems of the different genetic groups were also observed in a general way in order to detect possible differences between them.

3 Results

3.1 Phylogenetic reconstruction based on ITS

The length of the aligned ITS region was 762 bp (including part of the ITS1, the 5.8*s* RNA gene, and part of the ITS2 regions). The topology of the trees obtained with the different phylogenetic analysis methods was generally similar. For this reason, only the result issued from Bayesian analyses with BEAST is presented in this paper (Fig. 2). As also demonstrated in a previous study on the *Vanilla* genus using chloroplast genes (Bouetard et al., 2010), leafless *Vanilla* species from the SWIO region were found to be monophyletic (Fig. 2).

Contrary to what was obtained by Bouetard et al. (2010), our results showed that the leafless group is positioned closer to *V. crenulata* Rolfe and *V. africana* Lindl. than to *V. imperialis* Kraenzl. and *V. polylepis* Summerh. (Figs. 2, S1). Nevertheless, the phylogenetic relationships of these African as well as Malagasy (*V. françoisii* H.Perrier) leafy species with the leafless group were poorly supported (Bayesian posterior probability of clade support (PP) < 50%), which precludes any further insight on the precise origin of the leafless species group before conducting a more thorough sampling of African leafy species.

ITS sequences were however very powerful to resolve species within the leafless monophyletic group, the main focus of our study. Our results showed, with the highest support (100%), a clear division of the leafless *Vanilla* group into two evolutionary lineages based on flower color rather than geographical origin: one clade (PP: 100%, Mean age: 1.92 Myr, 95% HPD: [0.04–4.98]) grouping all white

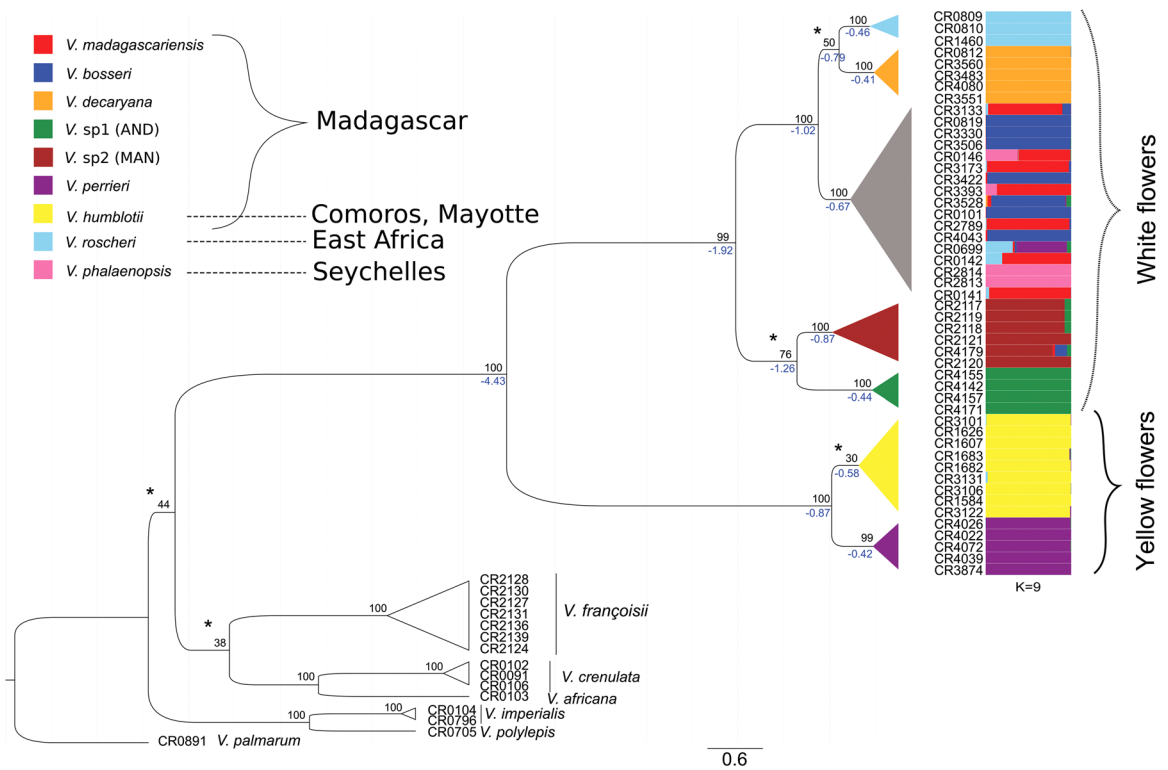

Fig. 2. Bayesian phylogenetic tree (BEAST) based on ITS region and genetic clustering analysis (STRUCTURE) using seven microsatellite loci ($K = 9$) of leafless *Vanilla* species from the South-West Indian Ocean (SWIO) region. Bayesian node support probability (black), as well as internal node time of divergence (blue), are displayed on the BEAST tree. Unreliable nodes (as confirmed by ML in Fig. S1) with low support probability are indicated by stars. In the Structure plot, each individual is represented by a horizontal bar with color proportional to their membership in the nine genetic groups computed from the Bayesian analysis. Species names (and geographical location) corresponding to the nine clusters are indicated.

flower species from Seychelles, East African coast and Madagascar, and one more recent clade (PP: 100%, Mean age: 0.87 Myr, 95% HPD: [0.01–2.74]) comprising yellow flower species from the Comoros archipelago and Madagascar (Fig. 2).

Within the white flower leafless group, two main clades were revealed from BEAST analysis: one comprising species from West of Madagascar (*V. madagascariensis* Rolfe, *V. bosseri* L.Allorge, *V. decaryana* H.Perrier), Seychelles (*V. phalaenopsis* Rchb.f.), and the East African coast (*V. roscheri* Rchb.f.); the second with accessions from East of Madagascar (*V*. sp1, *V*. sp2) (Fig. 2). The first one was further divided into three well-defined clades: *V. roscheri* (East African coast) (PP: 100%), *V. decaryana* (Madagascar) (PP: 100%) and a more ancient monophyletic group (PP: 100%) comprising accessions from species *V. madagascariensis, V. bosseri*, and *V. phalaenopsis*. Based on BEAST results, *V. roscheri* and *V. decaryana* formed a monophyletic group, but with very low support (PP: 50%), which did not hold in ML analysis (Fig. S1). The second clade within the white flower species was an early diverging monophyletic group (PP: 76%) comprising two leafless populations from East of Madagascar, with a group from the East coast (*V*. sp2, PP: 100%) being older (mean age: 0.87 Myr, 95% HPD: [0.02–2.77]) than the one from the South-East in the Malagasy highlands' domain (*V*. sp1, PP: 100%, mean age: 0.44 Myr, 95% HPD: [0–1.61]) (Fig. 2). In accordance with the relatively low support (PP: 76%) revealed in the Bayesian method, the ML method indicated that *V*. sp1 and *V*. sp2 did not form a monophyletic group (Fig. S1).

Within the yellow flower group, *V. perrieri* Schltr. from Madagascar formed a well-individualized group (PP: 100%) from the poorly defined (PP: 30%) *V. humblotii* Rchb.f. group (from Madagascar and the Comoros archipelago) (Fig. 2). The ML tree (Fig. S1) clearly showed that *V. perrieri* differentiated from *V. humblotii*.

3.2 Microsatellite genetic assignment

On the basis of the Ln $P(X \mid K)$ criteria (Pritchard et al., 2000), the most probable number of genetic clusters was $K = 9$. But considering the result of Evanno's method (Evanno et al., 2005), the best *K* was estimated at 4, with also a small peak at 9 (Fig. S2). Thus, we considered that the best *K* was nine combining the two methods. Among the $K = 9$ clusters estimated by both Evanno and Pritchard methods, seven matched the previously defined Malagasy genetic groups in Andriamihaja et al. (2021): four corresponded to already described species endemic to Madagascar (*V. madagascariensis*, *V. bosseri*, *V. decaryana*, *V. perrieri*), two represented the two populations from the East of Madagascar (*V*. sp1, *V*. sp2), and one corresponded to a species supposedly endemic to Comoros archipelago but also present in Madagascar (*V. humblotii*). Microsatellites confirmed the ITS result by showing that *V. humblotii* accessions from the Comoros archipelago and Madagascar were grouped in the same genetic cluster. In addition, two new genetic clusters were revealed in the present study, corresponding to the newly studied species from the SWIO (*V. phalaenopsis* from Seychelles and *V. roscheri* from East Africa) (Fig. 2). Moreover, microsatellite genetic assignment allowed to separate three species (*V. madagascariensis*, *V. bosseri* and *V. phalaenopsis*) that were unresolved within a monophyletic group revealed by ITS phylogenetic analysis (Fig. 2). Assignment probabilities were generally high (>80%) for almost all specimens, including unknown accessions, suggesting a clear clustering. Only two accessions, issued from botanical garden collections, did not match clearly to a species cluster, CR0146 and CR0699, and could be interspecific hybrid species. As for ITS analysis, the two Malagasy eastern groups (*V*. sp1 and *V*. sp2) were well-differentiated genetically based on microsatellite analyses. The PCoA results showed in general the same genetic structure pattern as the STRUCTURE results (Fig. S3). The first three axes retained in the analysis explained 12.60%, 11.29%, and 8.46% of the overall variance in microsatellite data, respectively (Fig. S3). The genetic variation of each genetic group is presented in Table S2. *V. bosseri* was the species with the highest genetic diversity with an average allelic richness of 3 alleles/locus and it also showed the highest number of private alleles ($Pa = 16$). The number of private alleles was also high in *V. humblotii* and *V. perrieri* (Table S2).

3.3 Morphological identification of Malagasy species

Morphological analyses were performed to compare the eastern *V*. sp2 population with other previously studied Malagasy species (Andriamihaja et al., 2021). The two principal axes of the PCA, representing 70.7% of the total variation, indicated a clear separation of four major non-overlapping floral groups, three of which were already described previously (Andriamihaja et al., 2021): (i) yellow flower group (*V. humblotii* and *V. perrieri*), (ii) western Malagasy species (*V. decaryana*, *V. bosseri*, *V. madagascariensis*), (iii) population from the South-East of Madagascar living in high elevation (*V*. sp1), and (iv) a new group corresponding to the population from Manompana in the North-East of Madagascar (*V*. sp2) (Fig. 3).

In general, the different genetic groups overlapped in the space of the first two principal components (PCs), with the exception of the two eastern groups, which showed a clear discontinuity (Fig. 3). The first PCA component explained 49.8% of the total floral variation and was associated with the size of the flower. PC1 allowed to clearly separate the three western white flower species (*V. decaryana*, *V. madagascariensis*, *V. bosseri*) from the four others (two yellow flowers species and two white flower genetic groups from the East). The color of the petal (*a*b** of the inside face), the color of the sepal (*b** of the outside face), and the color of the labellum (*L*a*b** of the inside face) contributed most strongly to the PC2, which explained 20.9% of the total floral variation. PC2 mainly differentiated white flower species from yellow flower species.

As already demonstrated by our previous study, six of the genetic groups showed specific morphological features (Andriamihaja et al., 2021) (Table 2). Indeed, MANOVA demonstrated a significant floral differentiation of grouping ($P < 0.001$). Reusing the morphological data sets of our previous study (Andriamihaja et al., 2021), we found the same discriminating features, but we were here able to determine that accessions from the 7th genetic group (*V*. sp2) formed a distinct morphological group from all other previously studied species: (i) the amount of shine and yellow of the

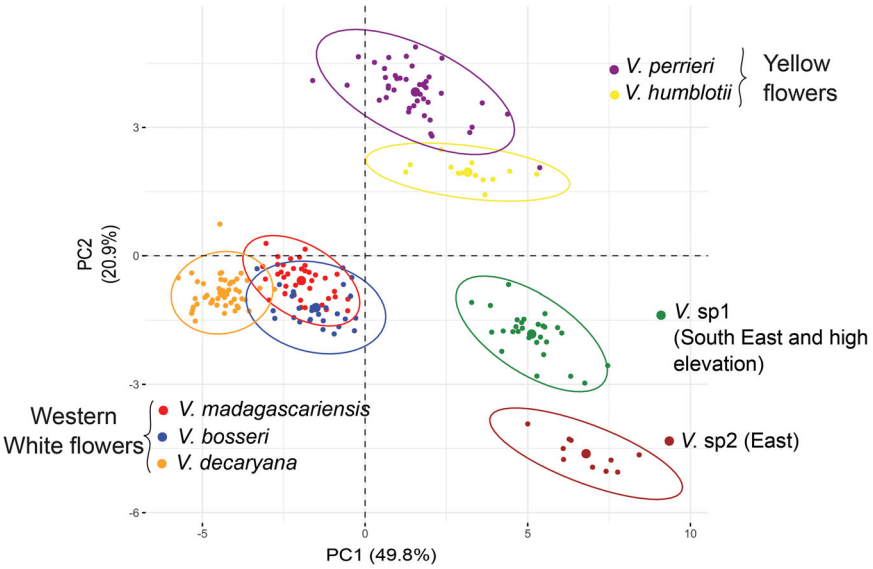

Fig. 3. PCA using 26 floral traits of the seven Malagasy genetic groups. The large dots inside the circles represent the centroids of each group. Flowers were collected from homogeneous populations identified in Andriamihaja et al. (2021), that is, populations where only one species can be observed. The colors of the dots indicate the assumed species to which the individual belongs. PCA, Principal Component Analysis.

Table 2 Bonferroni corrected pairwise *t* test of 10 floral parameters (mean ± SD) among 7 structure genetic groups

Floral Traits	*V. humblotii*	*V. madagascariensis*	*V. perrieri*	*V. bosseri*	*V. decaryana*	*V. allorgeae*	*V. atsinananensis*
LS	16.6 ± 2.3^{a}	7.2 ± 0.8^{b}	13.7 ± 1.9^{c}	9.4 ± 1.2^{d}	4.8 ± 0.4^{e}	21.4 ± 2.5^{f}	20.5 ± 3.4^{f}
TFW	6.1 ± 1.1^{a}	2.9 ± 0.4^{b}	4.2 ± 0.9^{c}	3.3 ± 0.5^{b}	1.8 ± 0.5^{d}	7.9 ± 1.0^{e}	13.1 ± 1.5^{f}
BIFP	41.4 ± 5.9^{a}	-2.6 ± 0.7^{b}	58.5 ± 8.8^{c}	-2.2 ± 0.8^{b}	-2.4 ± 0.7^{b}	-1.8 ± 0.6^{b}	0.5 ± 0.03^{b}
AIFP	-8.1 ± 1.1^{a}	1.8 ± 0.4^{b}	-10.3 ± 1.7^{c}	2.3 ± 0.4^{b}	1.8 ± 0.7^{b}	1.8 ± 0.5^{d}	-1.5 ± 0.1^{b}
OW	4.9 ± 0.5^{a}	3.6 ± 0.3^{b}	4.1 ± 0.5^{c}	3.8 ± 0.3^{b}	3.3 ± 0.3^{d}	5.9 ± 0.6^{e}	7 ± 1.2^{f}
PL	7.5 ± 0.5^{a}	5.8 ± 0.5^{b}	6.5 ± 0.6^{c}	5.9 ± 0.6^{b}	4.2 ± 0.3^{d}	8.8 ± 0.5^{a}	7.5 ± 0.4^{e}
BIL	40.5 ± 1.9^{a}	2.5 ± 0.9^{b}	39.8 ± 6.6^{a}	2.2 ± 1.3^{b}	-0.8 ± 0.5^{c}	$0.8 \pm 0.6^{b,c}$	$0.6 \pm 0.04^{b,c}$
CL	2.5 ± 0.2^{a}	1.8 ± 0.1^{b}	2.4 ± 0.2^{a}	2.2 ± 0.2^{c}	1.6 ± 0.2^{d}	3.3 ± 0.1^{e}	3.1 ± 0.1^{f}
ASL	7.1 ± 0.6^{a}	5.9 ± 0.6^{b}	6.6 ± 0.7^{a}	5.9 ± 0.5^{b}	4.2 ± 0.3^{c}	8.5 ± 0.5^{a}	7.2 ± 0.4^{d}
OWE	1.2 ± 0.4^{a}	0.6 ± 0.1^{b}	0.8 ± 0.2^{c}	0.6 ± 0.1^{b}	0.5 ± 0.1^{b}	1.7 ± 0.3^{d}	2.8 ± 0.8^{e}

Means followed by the same letter in the same row are not significantly different ($P > 0.05$). Morphological data of *V. humblotii* Rchb.f. ($n = 13$), *V. madagascariensis* Rolfe ($n = 38$), *V. perrieri* Schltr. ($n = 44$), *V. bosseri* L.Allorge ($n = 29$), *V. decaryana* H.Perrier ($n = 46$), and *V. allorgeae* (*V.* sp1) ($n = 28$) were taken from Andriamihaja et al. (2021). *V. atsinananensis* (*V.* sp2) ($n = 10$) floral samples are specific to this study. AIFP, *a** of inside face of petal; ASL, adaxial sepal length (cm); BIFP, *b** of inside face of petal; BIL, *b** of inside of labellum; CL, column length (cm); LS, labellum surface (cm^2); *n*, number of samples; OW, ovary width (mm); OWE, ovary weight (g); PL, petal length (cm); TFW, total floral weight (g).

inside face of petals that allowed to distinguish the two yellow flower species, *V. humblotii* and *V. perrieri*; (ii) the small size of flowers (petals < 5 cm long), characteristic of *V. decaryana*; (iii) the absence of hair on the labellum allowing to discriminate *V. bosseri* from *V. madagascariensis;* and iv) the large size of flowers (petals > 7 cm long) enabling to discriminate *V.* sp1 (AND), located in the South-East, from the western species with white flowers. The newly studied eastern genetic group (*V.* sp2, MAN) had the largest floral parts, the heaviest flowers with an average flower weight of 13.1 g (Fig. 4; Table 2) and floral parts with a more rounded apex, a result specific to this study. The 10 most discriminating floral traits are presented in Table 2, with 9 characters already raised in a previous study (Andriamihaja et al., 2021) and one new (Column length) specific to this study.

4 Discussion

4.1 Phylogenetic relationships and phylogeography of Malagasy leafless species and sister species from the SWIO region

Based on the ITS region, which is the most used sequence marker in plant classification and in orchid phylogeny (Tsai et al., 2005; Arenas & Dressler, 2010; Sharma et al., 2012; Azofeifa-Bolaños et al., 2017), this study provides further support for the monophyly of the leafless *Vanilla* group from

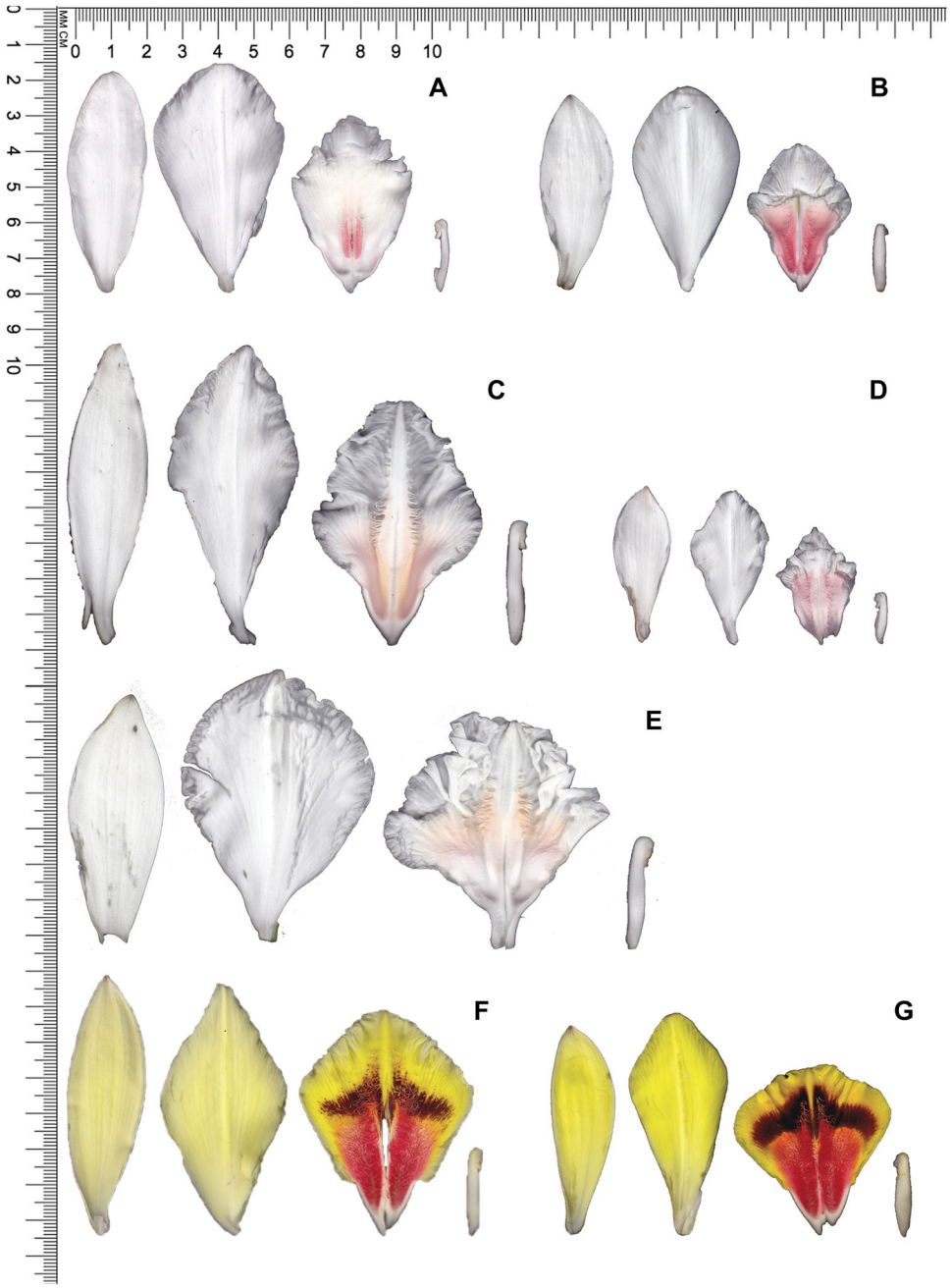

Fig. 4. Scanned flower parts (sepal, petal, labellum, column) of the different leafless *Vanilla* species from Madagascar: **A**, *V. bosseri*. **B**, *V. madagascariensis*. **C**, *V.* sp1. **D**, *V. decaryana*. **E**, *V.* sp2. **F**, *V. humblotii*. **G**, *V. perrieri*.

the SWIO region, as already found by Bouetard et al. (2010) using chloroplast DNA.

In contrast to chloroplast DNA (Bouetard et al., 2010), the ITS region clearly resolved the phylogeny of this leafless *Vanilla* group with generally highest Bayesian posterior probability support values and the separation of many known species into several monophyletic groups (Fig. 2). Although the precise relationships between African leafy species and the leafless *Vanilla* group from the SWIO region were poorly resolved in our study, the African related origin of the common ancestor of the group and its dispersion to Madagascar was clearly demonstrated by Bouetard et al. (2010). Similarly, several Malagasy plants originated from historical dispersions from Africa (Forrest & Hollingsworth, 2003; Micheneau et al., 2008; Anthony et al., 2010; Wikström et al., 2010; Strijk et al., 2012; Callmander et al., 2016). Ali & Huber (2010) demonstrated that rather than land bridges (such as the Davie Ridge in the Mozambique channel), strong winds and currents flowing from the East coast of the African continent eastward toward Madagascar during the Paleogene period (about 66–23 Ma [Luterbacher et al., 2004]) may have facilitated

colonization events from Africa to Madagascar. Our phylogenetic reconstruction based on different methods strongly supports an initial split of the leafless group into two main lineages: a clade of yellow flower species and a clade of white flower species. Both leafless groups have differentiated in Madagascar, from where they have then colonized nearby islands (Seychelles, Comoros archipelago) and backwards to the eastern African coast by transoceanic dispersion, as will be discussed below. Long-distance colonization events have already occurred several times in the evolution of the genus *Vanilla* and specifically at the origin of leafless species (Bouetard et al., 2010).

The clade of species with white flowers was possibly formed earlier than the one with yellow flower species based on the relaxed molecular clock analysis. It was divided into two main groups showing a clear East/West Madagascar biogeographic disjunction, with the most ancestral lineage (estimated about 1.26 Myr, 95% HPD: [0.04–3.78]) represented by the two eastern Malagasy populations of which the Northeastern Manompana population (*V.* sp2) was the most ancient (Mean age: 0.87 Myr, 95% HPD: [0.02–2.77]) (Fig. 2). The second monophyletic group (PP: 100%) within the clade of white flower species was divided into three clades: (i) a strongly supported clade (PP: 100%) consisting of accessions from Madagascar and Seychelles (*V. madagascariensis, V. bosseri, V. phalaenopsis*), appeared about 0.67 Ma (95% HPD: [0.01-2.01]), (ii) *V. roscheri* group from East African coast (PP: 100%) and (iii) the most recent clade (Mean age: 0.41 Myr, 95% HPD: [0-1.26]) formed by *V. decaryana* from South of Madagascar (PP: 100%). This structuration in the white flower clade suggests a Malagasy origin. The common ancestor of this clade differentiated in Madagascar, most probably first in the East, followed by dispersion and differentiation in the West by a non-adaptive mechanism as inferred in the "western forest refugia" and "Montane refugia" scenarios (Vences et al., 2009), as suggested in our previous study (Andriamihaja et al., 2021). However given the uncertainty of the divergence dates estimates (Table S3), it is also possible that a leafless ancestor could have been widespread in eastern and western Madagascar and then has adapted to humid versus dry conditions, giving rise to two phylogenetic lineages, under the "ecogeographic constraint" scenario (Vences et al., 2009). Cases of diversification along West/East environmental gradients have already been addressed several times in Madagascan animals (e.g., Vences & Glaw, 2003; Yoder & Heckman, 2006; Glaw et al., 2009; Vences et al., 2009; Younger et al., 2019). In Malagasy plants, analogous cases of differentiation between western versus eastern species have been also reported within *Canarium* L. genus (Federman et al., 2018) and several genera of the Sarcolaenaceae (Randrianasolo & Miller, 1999; Aubriot et al., 2016). Nevertheless, it is noteworthy that the group of western Malagasy species has then given rise, by transoceanic dispersion, to *V. roscheri* in the Southeastern coast of Africa and to *V. phalaenopsis* in Seychelles. Cameron (2011) suggested that *V. roscheri, V. phalaenopsis* and *V. madagascariensis* were probably morphological variations of the same species. The ITS phylogeny clearly showed that this is not the case for *V. roscheri*. On the other hand, *V. phalaenopsis* from Seychelles and *V. madagascariensis* (and *V. bosseri*) from Madagascar appeared as conspecific species on the ITS tree. They nevertheless represent genotypic species that have recently diverged within the leafless *Vanilla* group from SWIO as demonstrated using microsatellite genetic structuring (Figs. 2, S3). Furthermore, according to the USC of de Queiroz (2007), it is sufficient for lineages to evolve separately to be classified as species. So, we considered that *V. madagascariensis*, *V. bosseri* and *V. phalaenopsis* form distinct recent species as they form distinct genetic groups that evolve separately. Since *V. madagascariensis* and *V. phalaenopsis* are geographically isolated from each other on different islands (Madagascar versus Seychelles), their separate evolution can be explained via genetic drift and/or selection, particularly for *V. phalaenopsis* in Seychelles. Gaining access to more numerous *V. phalaenopsis* accessions would allow confirming this hypothesis.

The leafless *Vanilla* group with yellow flowers comprises two known botanical species, *V. humblotii* and *V. perrieri* (Portères, 1954). Despite ancient records of *V. humblotii* as endemic to the Comoros archipelago (Portères, 1954), the present phylogenetic analysis confirmed its presence also in Madagascar, a fact that was already noticed by Cribb & Hermans (2009) and previously suggested on the basis of population genetics approaches (Andriamihaja et al., 2021). Indeed, we demonstrate here that *V. humblotii* accessions from Madagascar and the Comoros archipelago belong to the same genotypic and phylogenetic group.

According to Soto Arenas & Cribb (2010), *V. perrieri* may well be conspecific to the Malagasy leafy species *V. françoisii*. But based on our results, *V. perrieri* has differentiated from *V. humblotii* present in northern Madagascar, most probably as a consequence of isolation by ecology (Andriamihaja et al., 2021). Indeed, if *V. humblotii* was only observed in the North of Madagascar, *V. perrieri* has extended along the West coast from the North to the South, toward much more dryer ecological niches (Allorge-Boiteau, 2005; Cribb & Hermans, 2009). In fact, the western coast of Madagascar is made up of four ecoregions ranging from the deciduous forests of the North with average annual rainfall between 1000 and 1500 mm to the spiny thickets in the arid zones of the extreme South with a maximum annual rainfall of 500 mm (Burgess et al., 2004; Andriamihaja et al., 2020). The existence of geographically separated populations of *V. humblotii* raises major questions on the origin and biogeographic history of this species in the SWIO that will require further genetic diversity studies in the different islands to be fully resolved. Nevertheless, as in the case of white flower species, the ITS tree topology suggests a Malagasy origin for yellow flower species; therefore, implying that *V. humblotii* has colonized the Comoros archipelago from Madagascar, adding another trans-oceanic dispersion event to those already demonstrated for the white flower group from Madagascar to Seychelles and the Southeastern African coast. Based on paleogeographic reconstructions and paleoceanographic modeling, it was demonstrated that since the Miocene (about 23 to 6 Ma [Sluiter et al., 2016]), currents and winds have been, like in the present times, oriented from Madagascar to Africa (Ali & Huber, 2010), thus the Malagasy origin of *V. roscheri* (East Africa) ancestor is possible. Likewise, an ancestry of *V. humblotii* from Madagascar is therefore more plausible

than from Comoros. On the basis of molecular, geological, and bathymetric analyses, Warren et al. (2010) showed that the low sea level of 0.5 Ma resulted in the enlargement of currently small islands, such as Seychelles, Maldives, and the Chagos Archipelago. Also, additional islands appeared between Seychelles and the Mascarene Islands (Warren et al., 2010), probably facilitating the recent dispersal of the ancestor of *V. phalaenopsis* from Madagascar. Inter-islands dispersions from Madagascar to the other islands of the SWIO region have also been reported many times in plants including orchids (Yuan et al., 2005; Micheneau et al., 2008; Wikström et al., 2010; Jaros et al., 2016; Kainulainen et al., 2017; Le Péchon et al., 2020). In *Vanilla*, these transoceanic dispersal events may have occurred through floating vegetation mats moved by sea currents and winds (Renner, 2004; Zhang et al., 2007), or by frugivorous migratory animals (zoochory) of which birds and bats have been suspected to be the main dispersal vectors for *Vanilla* species in the SWIO region and elsewhere (Lubinsky et al., 2006; Bory et al., 2008; Bouetard et al., 2010; Gigant et al., 2011b), facilitated here by the short distance between the different territories (<2000 km).

4.2 Leafless *Vanilla*: How many species are there in Madagascar?

Morphological taxonomy was the first method used in plant classification (Mallet, 1995; de Queiroz, 2005, 2007; Householder et al., 2010; Aldhebiani, 2018; Rouhan & Gaudeul, 2021). From the description of the first species in the *Vanilla* genus (Rolfe, 1896) to the recent new discoveries (Pansarin, 2010; Flanagan & Mosquera-Espinosa, 2016; Pansarin & Miranda, 2016; Fraga et al., 2017; Parizaca, 2019), floral traits have been the basis of *Vanilla* taxonomy. *Vanilla* is one taxonomically complex plant group (Gigant et al., 2011b; Andriamihaja et al., 2020), with the number of species increasing at a significant rate, which has led to frequent taxonomic revisions (Rolfe, 1896; Portères, 1954; Arenas & Dressler, 2010; Soto Arenas & Cribb, 2010; Karremans et al., 2020). Multivariate grouping of morphological traits allowed us to distinguish four non-overlapping groups of leafless *Vanilla* species from Madagascar, with flower size and color being the discriminating features (Fig. 3; Table 2). Although the characteristics of the flowers are the most discriminating in *Vanilla* genus, visual comparisons we have made between stems of the different genetic groups, as well as an ongoing study, showed differences in the abundance and forms of warts on the stems (Andrianantenaina JG, pers com., 2020). However, *Vanilla* species may show phenotypic plasticity and intraspecific variation (Arenas & Dressler, 2010; Bory et al., 2010; Soto Arenas & Cribb, 2010) and sometimes different species may have similar morphological traits (Portères, 1954; Soto Arenas & Cribb, 2010) requiring specialist expertise to distinguish species. We therefore used a complementary approach within an integrative taxonomic framework combining PheSC, PhySC, and GSC. The use of several sources of data has already proven its effectiveness in the definition of *Vanilla* species in Costa Rica and in the revision of the genus (Arenas & Dressler, 2010; Soto Arenas & Cribb, 2010; Azofeifa-Bolaños et al., 2017).

The combination of phylogeny using ITS, morphological analysis, and ISSR clustering methods have already been used in the *Vanilla* systematic analysis and allowed to delimit *V. planifolia* and its relatives in the Mexican Yucantan peninsula (Villanueva-Viramontes et al., 2017). In our study, the genotypic method approach based on microsatellite data, combined with floral morphological data, identified all five species previously described in Madagascar (*V. decaryana, V. bosseri, V. madagascariensis, V. humblotii, V. perrieri*) (Portères, 1954; Allorge-Boiteau, 2005; Cribb & Hermans, 2009). In addition, two new genotypic clusters were detected in the East of Madagascar (Fig. 2). The seven Malagasy genetic clusters identified matched those previously revealed by Andriamihaja et al. (2021), showing the power of a population assignment analysis despite the small intraspecific sampling used here. Although morphological groups sometimes overlap, each matched in general one specific genotypic cluster (Figs. 3, 4).

The use of morphological and genotypic data was effective to discriminate plant species in several genera such as *Ancistrocladus* Wall. (Ancistrocladaceae) (Turini et al., 2014), *Phoenix* L. (Arecaceae) (Pintaud et al., 2010), *Carapa* Aubl. (Meliaceae) (Duminil et al., 2006), and *Epidendrum* L. (Orchidaceae) (Pessoa et al., 2012).

When considering the PhySC approach using ITS, the results matched those from the PheSC and GSC for three leafless *Vanilla* species identified in Madagascar (*V. decaryana, V. humblotii, V. perrieri*) and for the two new species from East of Madagascar (*V*. sp1, *V*. sp2). However, although ITS analysis was consistent with the hypothesis that *V. madagascariensis* and *V. bosseri* could be conspecific species based on PhySC (Fig. 2), microsatellites revealed that these two species are nevertheless genotypic distinct species and have recently diverged in Madagascar. Furthermore, they also show a clear morphological differential feature (no hair on the labellum of *V. bosseri*) (Fig. 4), showing that they can be considered also a distinct phenetic species.

In conclusion, this integrative taxonomy approach therefore identifies seven leafless species indigenous to Madagascar (*V. decaryana, V. bosseri, V. madagascariensis, V. perrieri, V. humblotii, V*. sp1, *V*. sp2) that have distinct morphological and genotypic traits. All species form distinct monophyletic groups, apart from *V. bosseri* and *V. madagascariensis* that are recently diverged from each other (and therefore only differentiated using rapidly evolving markers such as microsatellites), and also present a marked morphological difference.

The present study is the first comprehensive contribution to understanding the relationships among some leafless species within the genus *Vanilla*. This study provides also a clear taxonomic placement of Malagasy *Vanilla* species using integrative taxonomy and multi-territorial samples. However, the low resolution of the relationships between leafless species from the SWIO region and their leafy sister species from Africa and Madagascar, as well as the lack of comparative flower samples from Seychelles, East Africa and the Comoros archipelago are limitations and require further sampling and study. The recent availability of genomic methods and resources for the genus *Vanilla* (Hasing et al., 2020; Favre et al., 2021; Hu et al., 2021) will also allow the use of these powerful new techniques in future integrative taxonomic studies. Nevertheless, our study demonstrates, according to various complementary evidence

(morphological, phylogenetic, genotypic), a major result with the discovery of two new leafless *Vanilla* species endemic to Madagascar, and distributed in the East of the island, outside the generally admitted distribution range of leafless *Vanilla* species (Portères, 1954; Allorge-Boiteau, 2005; Cribb & Hermans, 2009) (Fig. 1). These two species are further described and referred to as *V. allorgeae* sp. nov. (*V.* sp1) and *V. atsinananensis* sp. nov. (*V.* sp2).

4.3 Taxonomic treatment

4.3.1 Description of new species

Vanilla allorgeae Andriamihaja & Pailler, sp. nov. (Figs. 4C, 5, 6)

Type Madagascar, Fianarantsoa province, Ambalavao district, Tsaranoro valley, E 46°46′37.93″, S 22°05′4.77″, alt. ca. 900 m; December 2018; N.H. Nambinintsoa, A. Botomanga and C. F. Andriamihaja 7 (TAN!), Holotype: Hic designatus.

Diagnosis *Vanilla allorgeae* Andriamihaja & Pailler is similar to *Vanilla madagascariensis* Rolfe but differs from the latter having larger petals, sepals and lip. *V. allorgeae* also differs from *V. magascariensis* having a lip with the orange throat (versus red throat in *V. madagascariensis*) and thickened keels with two rows of hairs extended to the lateral lobes (versus abundant red hairs from the base to the middle of the lobe). *V. allorgeae* Andriamihaja & Pailler differs from *V. atsinananensis* Andriamihaja & Pailler having longer sepals and petals.

Description Hemi-ephiphytic branching aphyllous vine, more than 5 m long. Terrestrial roots brown, pubescent, and ramified. Aerial roots are whitish to brown, glabrous, flattened, and non ramified. Stems, cylindric (8–10 mm thick), fleshy, green, smooth when old and sometimes verrucose when young, flexuous, with two opposite longitudinal grooves along the internode. Internodes 3–7 cm long. Leaves green, early deciduous, ca 6–8 cm long, triangular, acuminate. Inflorescence sub-terminal or axial, racemose, 20–25 cm long, ca 15–35 flowered. Floral bracts, ca 1–3 cm long, ovate, apex acute. Flowers white, ephemeral lasting all day; pedicel and ovary green, 50.0–80.0 × 4.8–7.5 mm, terete, smooth, curved; buds green at the base and dark green at the apex, apex acute. Sepal (adaxial: 7.4–9.9 × 1.7–2.7 cm, lateral: 7.3–9.1 × 1.6–2.6 cm) white (internally), brownish-green with white border margin (externally), lanceolate, smooth margin, reflexed, apex acute. Petals (7.6–9.7 × 2.0–5.1 cm) fused to the labellum for ca 1 cm, white, with a green and conspicuous median longitudinal line on the external side, ovate, reflexed, margin undulate, apex acute. Labellum trilobed (6–8 × 4–6 cm), fused to the dorsal side of the column for 1.0—1.5 cm, with orange peach throat from the base to the middle of the labellum then white up to the apex, margin fringed; the palate with two longitudinal median keels, subtubular in the third basal part then canaliculate and marginally hairy, thinner, and glabrous at the apex. Keels hairs long ca 2–4 mm, simple or bi-(tri-)furcate. Column subterete, 30–37 × 4–6 mm, white, flattened on the ventral part; stigma transversely broadly oblong. Rostellum quadrangular, 2–4 × 3–6 mm. Anther basifixed and oblong, 4–6 × 1–3 mm.

Distribution This species is known from the Tsaranoro valley forest (E 46°46′15.45″, S 22° 5′10.97″) fragment near Andringitra National Park in the South-East and central highlands' domain of Madagascar, district of Ambalavao. According to flower pictures on Tropicos (Charles Rakotovao, n°7236!), it is also present in forests in the district of Ivohibe, region of Ihorombe (South-East).

Ecology and Habitat Tsaranoro valley is accessed via the National Road 7 (RN7) south of Ambalavao by taking the road to Vohitsaoka and then to the village of Morarano. The place is located at an elevation of 850-945 m in the western valleys of the Andringitra mountains. This natural barrier retains the humidity of blowing winds from the eastern coast; therefore, Tsaranoro has a warmer and dryer climate. The average annual temperature is 19.4 °C (over a 21-year period between 1999 and 2019) and the average annual rainfall is 1071 mm (over a 21-year period between 1999 and 2019) (https://fr.climate-data.org/). Population of *V. allorgeae* was discovered in a sacred grove, a burial place for the local communities. Vegetation is a remnant of a highland humid forest surrounded by wooded grassland mosaic (Moat & Smith, 2007). Forest cover ranged from 20% to 60%. The most common species are *Dombeya* Lam. species (Sterculiaceae), *Euphorbia tetrapera* Baker (Euphorbiaceae), and *Tragia tiverneana* Leandri (Euphorbiaceae). The forest canopy is 8 m in height. *V. allorgeae* is a root-climber, found clinging either to trees or to rocks for light interception.

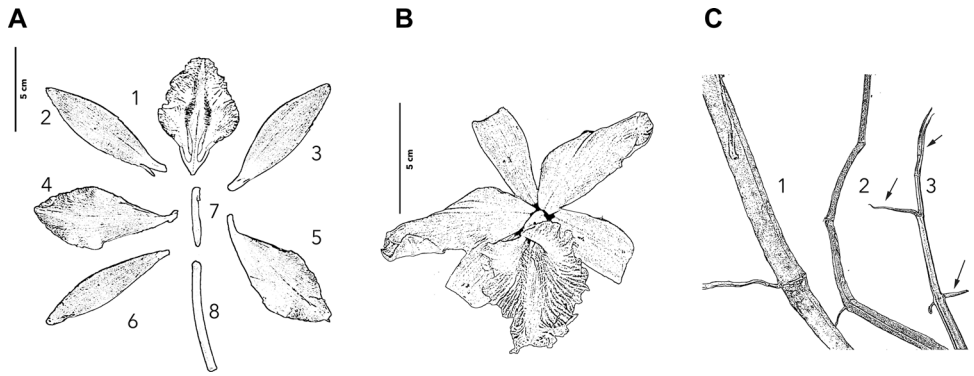

Fig. 5. *Vanilla allorgeae* Andriamihaja & Pailler (*V.* sp1). **A**, Dissected perianth: labellum (1), left lateral sepal (2) right lateral sepal (3), left petal (4), right petal (5), adaxial sepal (6), column (7), pedicel and ovary (8). **B**, Flowers. **C**, Portion of stem (1), rudimentary leaves (2, 3) and root (image by P. Besse).

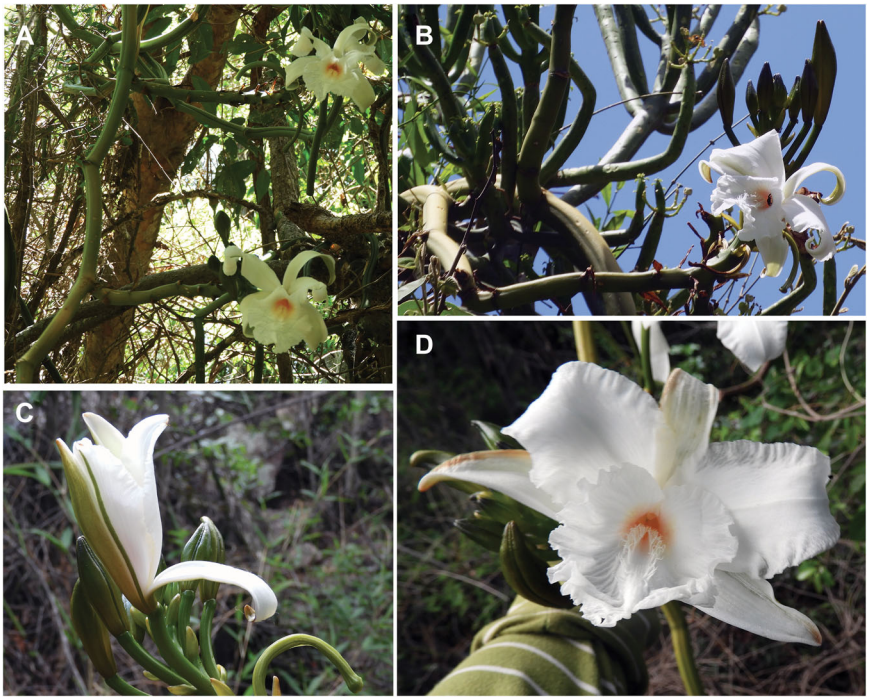

Fig. 6. *Vanilla allorgeae* Andriamihaja & Pailler (*V.* sp1). **A, B**, Vegetative habit of plant. **C**, Inflorescence with opening flower showing the abaxial surface of sepals and petals. **D**, Fully opened flower with front view of labellum. Photographs: A and B by N. H. Nambinintsoa, C and D by C. F. Andriamihaja.

Phenology Initiation of flowering has been observed in the month of October 2018, corresponding to the beginning of the rainy season in Madagascar. Flowering has ended in December 2018.

Etymology This species is named in honor of Dr. Lucile Allorge, a botanist who has been working for a long time on the plants of Madagascar, especially on Malagasy endemic *Vanilla*.

Proposed conservation status The Tsaranoro Valley Forest is remnant of sacred forests of less than 60 ha, managed by a local organization. Very few fruits were observed during our fieldwork at Tsaranoro indicating a loss of interaction with pollinators. Based on a previous genetic analysis (Andriamihaja et al., 2021), *Vanilla* population in Tsaranoro has a high clonality rate, thus a very low genetic diversity that may reduce the adaptive capacity of the population to climate change and threaten the species in the long term. Given these threats explained above and the restricted distribution of the species, although the forests are protected, it seems appropriate to include *V. allorgeae* as endangered [EN]. However, additional studies of other populations are needed to validate this conservation status. Then, a formal assessment should be carried out according to the IUCN Red list criteria (IUCN, 2001) to confirm the status.

Additional specimens examined MADAGASCAR. 1—District of Ivohibe. Commune of Ivohibe. Longoza Fokontany. Forest with tall trees and grassy savannah of Vohibory. Efanôla river. E 46°41′03″, S 22°35′15″, alt. 733 m, December 2016, C. Rakotovao 7236 [flower photographs on Tropicos!] (MO, P, TAN). 2—District of Ivohibe and Ihosy. Large crassulescent vine, 8–10 m covering coralliform Euphorbiaceae. October 1992, L. Allorge, P (P01801828!). 3—Mananara basin, Menarahaka valley, xerophilic forest remains near Analavoka. Alt. 700–800 m, October 1924, H. Humbert, P (P00102694!).

Vanilla atsinananensis Andriamihaja & Pailler, sp. nov. (Figs. 4E, 7, 8).

Type Madagascar**,** Toamasina province, Analanjirofo region, Soanierana Ivongo district, Manompana commune, Ambodiriana protected forest, E 49°41′46.7″, S 16°40′33.0″, alt. ca. 200 m; December 2020; M. Grisoni, N. H. Nambinintsoa, A. Botomanga, C. F. Andriamihaja 8 (TAN!), Holotype: Hic designatus.

Diagnosis *Vanilla atsinananensis* Andriamihaja & Pailler is similar to *V. allorgeae* Andriamihaja & Pailler but differs from the latter having obovate and larger petals (versus ovate in *V. allorgeae*). *V. atsinananensis* also differs from *V. allorgeae* Andriamihaja & Pailler having a lip with an orange throat only in the middle part of the labellum (versus from the base to the tip of the labellum in *V. allorgeae* Andriamihaja & Pailler), verrucose old stems. *V. atsinananensis* Andriamihaja & Pailler differs from *V. madagascariensis* Rolfe having larger sepals, petals and deeply fringed lips.

Description Hemi-ephiphytic branching aphyllous vine, more than 7 m long. Terrestrial roots brown, verrucose, and ramified. Aerial roots whitish to brown, glabrous, flattened, and non ramified. Stems, cylindric (5–15 mm thick), fleshy, green, verrucose, flexuous, with two opposite longitudinal grooves along the internode. Internodes 5–18 cm long. Leaves green, early decidious, ca 2–4 cm long, triangular, acuminate. Inflorescence sub-terminal or axial, racemose, ca 10-30 flowered. Floral bracts ovate, apex acute. Flowers white,

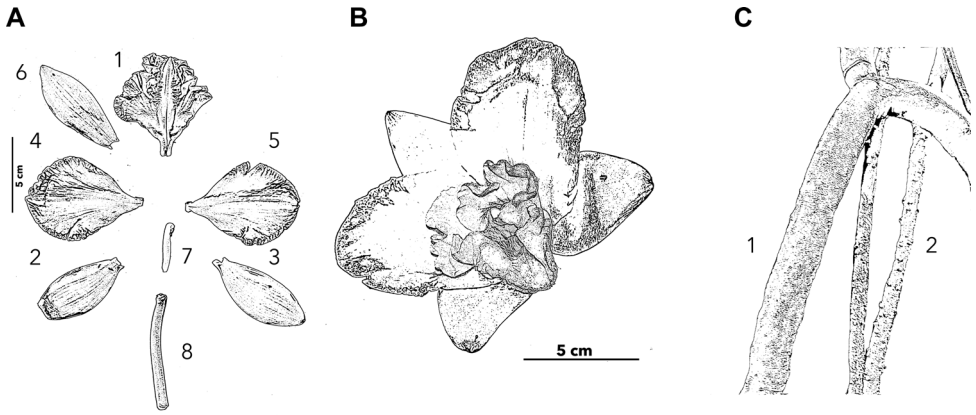

Fig. 7. *Vanilla atsinananensis* Andriamihaja & Pailler (*V.* sp2). **A,** Dissected perianth: labellum (1), left lateral sepal (2) right lateral sepal (3), left petal (4), right petal (5), adaxial sepal (6), column (7), pedicel and ovary (8). **B**, Flowers. **C**, Parts of stem (1) and root (2) (Image by P. Besse).

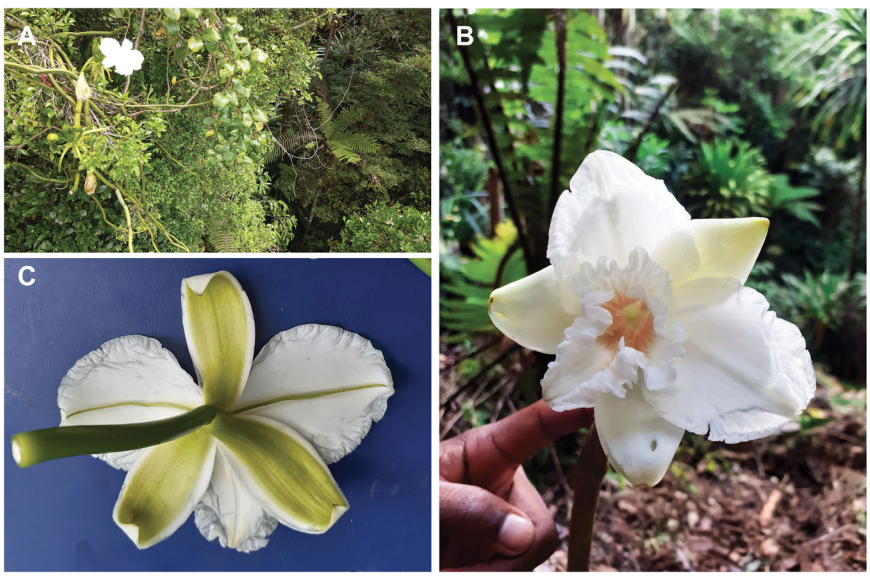

Fig. 8. *Vanilla atsinananensis* Andriamihaja & Pailler (V. sp2). **A,** Vegetative habit of the plant. **B,** Fully opened flower with front view of the labellum. **C,** Back view of flower showing the abaxial surface of sepals and petals. Photographs: M. Grisoni.

ephemeral; pedicel and ovary green, 67–81 × 6–8 mm, terete, smooth, curved; buds green, apex obtuse. Sepal (adaxial: 6.5–7.9 × 2.2–3.0 cm, lateral: 5.8–7.2 × 2.3–3.0 cm) white (internally), green with white margin (externally), ovate, smooth margin, apex acute. Petals (6.6–8.1 × 3.8–5.6 cm), white, with green and conspicuous median longitudinal line on the external side, obovate, margin undulate, apex obtuse. Labellum trilobed (5–7 × 4–7 cm), fused to the dorsal side of the column for 1.0–1.5 cm, with orange peach throat in the middle part and white in the other parts, margin deeply fringed; midlobe with two longitudinal median keels, sub-tubular in the third basal part then canaliculate, marginally hairy in the central part, thinner and glabrous at the base and the apex. Keel hairs, ca 2–4 mm, simple or bifurcate. Column subterete, 29–32 × 5–6 mm, white, flattened on the ventral part; stigma transversely broadly oblong. Rostellum quadrangular, 4–6 × 3–4 mm. Anther basifixed and oblong, 3–6 × 2–3 mm.

Distribution The species has been firstly reported in Ambodiriana reserve in the district of Soanierana Ivongo (North-East of Madagascar) managed by NGO ADAFAM delegated site manager. One population was also observed in the South of Toamasina (E 48°46′27.07″, S 18°26′47.50″), Madagascar. The species was also seen in 2017 in the forest flap of Ambatoharagnana, commune of Sahambala, district of Toamasina II according to the photos of Benjamina Ralaijaona, Giovanni A. Rakotonirina, Anthony R. Syde, Harisandy M. Rasoanindriana & P. Antilahimena on Tropicos (photo n°51!).

Ecology and Habitat The Ambodiriana forest (240 ha) is located two hours hike from the town of Manompana, North of Toamasina (eastern Madagascar) at an elevation of 100–250 m. It is exposed to the Alizé, with 2937 mm of annual rainfall (average over a 21-year period between 1999 and 2019) and 23.1 °C of average annual temperature (over a

21-year period between 1999 and 2019) (https://fr.climate-data.org). Vegetation is a lowland humid forest (Moat & Smith, 2007) with trees reaching more than 30 m in height such as *Canarium madagascariense* Engl. (Burseraceae), *Anthostema madagascariensis* Baill. (Euphorbiaceae) and *Uapaca ferruginea* Baill. (Euphorbiaceae). Forest cover ranged from 40% to 80%. Ambodiriana is home to about a hundred orchid species (https://www.adafam.org/) including *V. atsinananensis*, which grows as hemiepiphyte on supporting trees.

Phenology Species was seen flowering from November. Flowers have been collected in December.

Etymology The specific epithet (atsinananensis) comes from the word "atsinanana", which means "East" in the Malagasy language. It makes reference to the East coast distribution of the species.

Proposed conservation status Given the low number of individuals in the Ambodiriana forest, *V. atsinananensis* might be considered as endangered (EN) or vulnerable (VU). A formal assessment according to IUCN red list criteria (IUCN, 2001) should however be done to confirm this status.

Additional specimens examined MADAGASCAR. 1—Toamasina II district, Sahambala commune, Sahavongo fokontany, nearest village Sahavongo, Ambatoharagnana forest flap, Ambinanin'ny Bikôka. E 49°05′25″, S 18°01′28″, alt. 515 m, February 2017, B. Ralaijaona, G. A. Rakotonirina, A. R. Syde, H. M. Rasoanindriana & P. Antilahimena 51 [flower photographs on Tropicos!] (MO, TAN). 2—Fort-Carnot road (5 km from the bifurcation), January 1964, J. Bosser 19049, P (P00334809!).

4.3.2 Identification key to leafless *Vanilla* species from Madagascar

In addition to including *V. bosseri* described by Allorge (Allorge-Boiteau, 2013) and the two new species discovered from our analyses (*V. allorgeae* and *V. atsinananensis*), we also provide a new revised key to the leafless *Vanilla* species from Madagascar based on Soto Arenas & Cribb (2010).

Key to leafless *Vanilla* species from Madagascar

1a. White flowers……2
1b. Yellow flowers……6
2a. Petals 70–98 mm long. Floral pedicel 51–82 mm long. Lip with fringed margin and peach orange throat: species from East of Madagascar……3
2b. Petals 30–70 mm long. Floral pedicel 34–65 mm long. Lip with pinkish, reddish or yellowish throat: species from West of Madagascar……4
3a. Buds green at the base and dark green at the apex. Sepals brownish green on the external surface, lanceolate, inrolled and measuring 7–10 cm long. Petals ovate, inrolled with apex acute. Labellum with fringed margin and orange peach throat from the base to the middle of the labellum and white in the middle of the labellum to the apex ……*V. allorgeae*
3b. Buds green with apex obtuse. Sepals green on the external surface, ovate, and measuring 5–8 cm long. Petals obovate with apex obtuse. Labellum with deeply fringed margin and orange peach throat in the middle of the labellum and white in the other parts of the labellum……*V. atsinananensis*
4a. Column of 9–19 mm long. Sepals 25–49 mm long, lip with pinkish color……*V. decaryana*
4b. Column 14–30 mm long. Sepals 48–73 mm long……5
5a. Lip obovate, obtuse, the margins not undulate, with two lines of long papillae or fleshy hair and pink-reddish throat……*V. madagascariensis*
5b. Lip with yellow-pink throat, without fleshy hair and with red protrusion between two crests……*V. bosseri*
6a. Petals and sepals with clear yellow color. Petals 65–85 mm × 30–41 mm. Lip ca. 62 mm long × 42 mm maximum width. Column 20–28 mm long……*V. humblotii*
6b. Petals and sepals with intense yellow color. Petals 51–79 mm × 18–37 mm. Lip ca. 50 mm long and 49 mm maximum width. Column 18–30 mm……*V. perrieri*

Acknowledgements

We acknowledge all the institutions that have granted us research and collect permits: DREED of Madagascar, Madagascar National Parks, SAGE, Asity, Tsaranoro VOI, Bongolava Maitso, CFPF, ONG ADAFAM Ambodiriana funded by American Orchids Society (AOS) and Univet Nature; and BRC Vatel for plant material. This research was financially and logistically supported by the University of Antananarivo, University of Reunion Island, CIRAD, and the UE-Feder-Interreg Project GerminationII-phase2. It was funded by a PhD grant to C. F. Andriamihaja from the European Regional Development Fund attributed by the Regional Council and from an Erasmus+ grant. A. V. Ramarosandratana and A. Botomanga also received a grant from Chanel Parfum Beauté for a conservation program of *Vanilla* species. We thank Adrien Rieux, a researcher from CIRAD, for his important advice on phylogenetic analysis with BEAST and Jean-Michel Hervouet, president of the France Orchid Federation, for his valuable recommendations. We are also grateful to Hoby N. Nomenjanahary, Johnson G. Andrianantenaina, Onja M. Ramilijaona and all guides for special assistance during fieldwork and Rodolphe L. Gigant for pictures.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Microsatellite data used to produce genetic structuring (Fig. 2) and PCoA (Fig. S3) are accessible on the online UMR PVBMT Cirad dataverse (https://doi.org/10.18167/DVN1/3OFMLH).

References

Abràmoff MD, Magalhães PJ, Ram SJ. 2004. Image processing with ImageJ. *Biophotonics International* 11: 36–42.

Aldhebiani AY. 2018. Species concept and speciation. *Saudi Journal of Biological Sciences* 25: 437–440.

Ali JR, Huber M. 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* 463: 653–656.

Allorge-Boiteau L. 2005. Les vanilles succulentes de Madagascar. *Succulentes* 2: 3–11.

Allorge-Boiteau L. 2013. Une nouvelle espèce de vanille à Madagascar. *Hommes et Plantes* 8: 4–5.

Andriamihaja CF, Ramarosandratana AV, Grisoni M, Jeannoda V, Besse P. 2020. The leafless *Vanilla* species-complex from the South-West Indian Ocean region: A taxonomic puzzle and a model for orchid evolution and conservation. *Research. Diversity* 12: 443.

Andriamihaja CF, Ramarosandratana AV, Grisoni M, Jeannoda VH, Besse P. 2021. Drivers of population divergence and species differentiation in a recent group of indigenous orchids (*Vanilla* spp.) in Madagascar. *Ecology and Evolution* 11: 2681–2700.

Anthony F, Diniz LEC, Combes MC, Lashermes P. 2010. Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Systematics and Evolution* 285: 51–64.

Arenas MAS, Dressler RL. 2010. A revision of the Mexican and central American species of *Vanilla* Plumier ex Miller with a characterization of their ITS region of the nuclear ribosomal DNA. *Lankesteriana* 9: 285–354.

Aubriot X, Soulebeau A, Haevermans T, Schatz GE, Cruaud C, Lowry PP. 2016. Molecular phylogenetics of Sarcolaenaceae (Malvales), Madagascar's largest endemic plant family. *Botanical Journal of the Linnean Society* 182: 729–743.

Azofeifa-Bolaños JB, Gigant LR, Nicolás-García M, Pignal M, Tavares-González FB, Hágsater E, Salazar-Chávez GA, Reyes-López D, Archila-Morales FL, García-García JA, Da Silva D, Allibert A, Solano-Campos F, Rodríguez-Jimenes GDC, Paniagua-Vásquez A, Besse P, Pérez-Silva A, Grisoni M. 2017. A new *Vanilla* species from Costa Rica closely related to *V. planifolia* (Orchidaceae). *European Journal of Taxonomy* 284: 1–26.

Barley AJ, White J, Diesmos AC, Brown RM. 2013. The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. *Evolution* 67: 3556–3572.

Besse P. 2014a. Guidelines for the choice of sequences for molecular plant taxonomy. In: Besse P ed. *Methods in molecular biology. Molecular plant taxonomy*. New York, NY: Humana. 39–51.

Besse P. 2014b. Nuclear ribosomal RNA genes: ITS Region. In: Besse P ed. *Methods in molecular biology. Molecular plant taxonomy*. New York, NY: Humana. 141–149.

Besse P, Da Silva D, Grisoni M. 2021. Plant DNA barcoding principles and limits: A case study in the genus *Vanilla*. In: Besse P ed. *Methods in molecular biology. Molecular plant taxonomy: Methods and protocols*. New York, NY: Humana. 131–148.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* 22: 148–155.

Bory S, Brown S, Duval MF, Besse P. 2010. Evolutionary processes and diversification in the genus *Vanilla*. In: Odoux E, Grisoni M eds. *Vanilla*. Boca Raton, FL: CRC Press. 15–29.

Bory S, Grisoni M, Duval MF, Besse P. 2008. Biodiversity and preservation of *Vanilla*: Present state of knowledge. *Genetic Resources and Crop Evolution* 55: 551–571.

Bouetard A, Lefeuvre P, Gigant RL, Bory S, Pignal M, Besse P, Grisoni M. 2010. Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Molecular Phylogenetics and Evolution* 55: 621–630.

Burgess ND, Underwood E, Dinerstein E, Olson D, Itoua I, Schipper J, Ricketts T, Newman K. 2004. *Terrestrial ecoregions of Africa and Madagascar: A conservation assessment*. Washington, DC: Island Press.

Callmander MW, Phillipson PB, Plunkett GM, Edwards MB, Buerki S. 2016. Generic delimitations, biogeography and evolution in the tribe Coleeae (Bignoniaceae), endemic to Madagascar and the smaller islands of the western Indian Ocean. *Molecular Phylogenetics and Evolution* 96: 178–186.

Cameron KM. 2011. *Vanilla orchids: Natural history and cultivation*. Portland, OR: Timber Press.

Chauhan T, Rajiv K. 2010. Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology* 1: 281–291.

China Plant BOL, Group LiDZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19641–19646.

Cribb P, Hermans J. 2009. *Field guide to the orchids of Madagascar*. Kew: Royal Botanical Gardens.

Darwin CMA. 1859. *The origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. London: John Murray.

Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–415.

de Candolle MAP. 1813. *Théorie élémentaire de la botanique, ou exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux*. Paris: Déterville.

de Queiroz K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences* 56: 196–245.

de Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Divakaran M, Babu KN, Ravindran PN, Peter KV. 2006. Interspecific hybridization in *Vanilla* and molecular characterization of hybrids and selfed progenies using RAPD and AFLP markers. *Scientia Horticulturae* 108: 414–422.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A. 2007. *A rough guide to BEAST 1.4*. United Kingdom: The University of Auckland, University of Edinburgh.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ. 2006. Blind population genetics survey of tropical rainforest trees. *Molecular Ecology* 15: 3505–3513.

Duminil J, Di Michele M. 2009. Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosystems* 143: 528–542.

Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ. 2012. Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa spp*. (Meliaceae). *Molecular Phylogenetics and Evolution* 62: 275–285.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology* 14: 2611–2620.

Favre F, Jourda C, Besse P, Charron C. 2021. Genotyping-by-sequencing technology in plant taxonomy and phylogeny. In: Besse P ed. *Methods in molecular biology. Molecular plant taxonomy: Methods and protocols*. New York, NY: Humana. 167–178.

Federman S, Donoghue MJ, Daly DC, Eaton DAR. 2018. Reconciling species diversity in a tropical plant clade (*Canarium*, Burseraceae). *PLoS ONE*. 13: e0198882.

Flanagan NS, Mosquera-Espinosa AT. 2016. An integrated strategy for the conservation and sustainable use of native *Vanilla* species in Colombia. *Lankesteriana* 16: 201–218.

Forrest LL, Hollingsworth PM. 2003. A recircumscription of *Begonia* based on nuclear ribosomal sequences. *Plant Systematics and Evolution* 241: 193–211.

Fraga CND, Couto DR, Pansarin ER. 2017. Two new species of *Vanilla* (Orchidaceae) in the Brazilian Atlantic Forest. *Phytotaxa* 296: 63–72.

Francis RM. 2017. POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* 17: 27–32.

Gigant RL, Brugel A, De Bruyn A, Risterucci AM, Guiot V, Viscardi G, Humeau L, Grisoni M, Besse P. 2011a. Nineteen polymorphic microsatellite markers from two African *Vanilla* species: Across-species transferability and diversity in a wild population of *V. humblotii* from Mayotte. *Conservation Genetics Resources* 4: 121–125.

Gigant RL, De Bruyn A, Church B, Humeau L, Gauvin-Bialecki A, Pailler T, Grisoni M, Besse P. 2014. Active sexual reproduction but no sign of genetic diversity in range-edge populations of *Vanilla roscheri* Rchb. f. (Orchidaceae) in South Africa. *Conservation Genetics* 15: 1403–1415.

Gigant RL, De Bruyn A, M'sa T, Viscardi G, Gigord L, Gauvin-Bialecki A, Pailler T, Humeau L, Grisoni M, Besse P. 2016. Combining pollination ecology and fine-scale spatial genetic structure analysis to unravel the reproductive strategy of an insular threatened orchid. *South African Journal of Botany* 105: 25–35.

Gigant RL, Severine B, Michel G, Pascale B. 2011b. Biodiversity and evolution in the *Vanilla* genus Grillo O. In: Oscar G, Gianfranco V ed. *The dynamical processes of biodiversity—Case studies of evolution and spatial distribution*. Rijeka: InTech. 1–26.

Glaw F, Nagy ZT, Köhler J, Franzen M, Vences M. 2009. Phylogenetic relationships of a new species of pseudoxyrhophiine snake (Reptilia: Lamprophiidae: Thamnosophis) suggest a biogeographical link between western and northern Madagascar. *Organisms Diversity & Evolution* 9: 13–22.

Goudet J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86: 485–486.

Hasing T, Tang H, Brym M, Khazi F, Huang T, Chambers AH. 2020. A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nature Food* 1: 811–819.

Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology* 61: 138–149.

Householder E, Janovec J, Mozambite AB, Maceda JH, Wells J, Valega R. 2010. Diversity, natural history, and conservation of *Vanilla* (orchidaceae) in Amazonian wetlands of Madre de Dios, Peru. *Journal of Botanical Research Institute of Texas* 4: 227–243.

Hu T, Chitnis N, Monos D, Dinh A. 2021. Next-generation sequencing technologies: An overview. *Human Immunology* 82: 801–811.

IUCN. 2001. *IUCN red list categories and criteria version 3.1*. Oxford: Information Press.

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

Jombart T. 2016. *Introduction to genetic data analysis using R*. London: Imperial College—MRC Centre for Outbreak Analysis and Modelling.

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

Karremans AP, Chinchilla IF, Rojas-Alvarado G, Cedeño-Fonseca M, Damián A, Léotard G. 2020. A reappraisal of Neotropical *Vanilla*. With a note on taxonomic inflation and the importance of alpha taxonomy in biological studies. *Lankesteriana* 20: 395–497.

Kendal D, Hauser CE, Garrard GE, Jellinek S, Giljohann KM, Moore JL. 2013. Quantifying plant colour and colour difference as perceived by humans using digital images. *PLoS ONE* 8: e72296.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.

Larson JL. 1968. The species concept of Linnaeus. *Isis* 59: 291–299.

Lê S, Josse J, Husson F. 2008. FactoMineR: An *R* package for multivariate analysis. *Journal of Statistical Software* 25: 1–28.

Le Péchon T, Traclet S, Barthelat F, Dimassi A, Gigord LDB, Hamidou A, Jourdain-Fievet L, Viscard G, Boullet V. 2020. Bridging the Mozambique Channel: First record of *Dombeya rosacea* (Dombeyoideae, Malvaceae) for Mayotte (Comoros Archipelago, Indian Ocean). *Phytotaxa* 460: 75–88.

Lubinsky P, Cameron KM, Molina MC, Wong M, Lepers-Andrzejewski S, Gomez-Pompa A, Kim SC. 2008. Neotropical roots of a Polynesian spice: The hybrid origin of Tahitian *Vanilla*, *Vanilla tahitensis* (Orchidaceae). *American Journal of Botany* 95: 1040–1047.

Lubinsky P, Van Dam M, Van Dam A. 2006. Pollination of *Vanilla* and evolution in the Orchidaceae. *Lindleyana* 75: 926–929.

Luterbacher HP, Ali JR, Gradstein FM, Hooker JJ, Monechi S, Ogg JG, Powell J, Röhl U, Sanfilippo A. 2004. The Paleogene period. In: Gradstein FM, Ogg JG, Smith A eds. *A geological time scale*. United Kingdom: Cambridge University Press. 384–408.

Mallet J. 1995. A species definition for the Modern Synthesis. *Trends in Ecology & Evolution* 10: 294–299.

Marchelli P, Gallo LA. 2001. Genetic diversity and differentiation in a southern beech subjected to introgressive hybridization. *Heredity* 87: 284–293.

Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

Micheneau C, Carlsward BS, Fay MF, Bytebier B, Pailler T, Chase MW. 2008. Phylogenetics and biogeography of Mascarene angraecoid orchids (Vandeae, Orchidaceae). *Molecular Phylogenetics and Evolution* 46: 908–922.

Minoo D, Jayakumar VN, Veena SS, Vimala J, Basha A, Saji KV, Nirmal Babu K, Peter KV. 2008. Genetic variations and interrelationships in *Vanilla planifolia* and few related species as expressed by RAPD polymorphism. *Genetic Resources and Crop Evolution* 55: 459–470.

Moat J, Smith P. 2007. *Atlas of the vegetation of Madagascar*. Kew: Royal Botanic Gardens.

Nielsen LR. 2000. Natural hybridization between *Vanilla claviculata* (W.Wright) Sw. and *V. barbellata* Rchb.f. (Orchidaceae): genetic, morphological, and pollination experimental data. *Botanical Journal of the Linnean Society* 133: 285–302.

Nielsen LR, Siegismund HR. 1999. Interspecific differentiation and hybridization in *Vanilla* species (Orchidaceae). *Heredity* 83: 560–567.

Olsen C, Qaadri K. 2014. *Geneious R7: A bioinformatics platform for biologists*. Newark, USA: Biomatters, Inc.

Pansarin ER. 2010. Taxonomic Notes on Vanilleae (Orchidaceae: Vanilloideae): *Vanilla dietschiana*, a Rare South American Taxon Transferred from *Dictyophyllaria*. *Selbyana* 30: 203–207.

Pansarin ER, Ferreira AWC. 2019. Elucidating cryptic sympatric speciation in terrestrial orchids. bioRxiv preprint. https://doi.org/10.1101/828491

Pansarin ER, Miranda MR. 2016. A new species of *Vanilla* (Orchidaceae: Vanilloideae) from Brazil. *Phytotaxa* 267: 84.

Parizaca AD. 2019. *Vanilla yanesha* (Orchidaceae), a new species of the membranaceous-leaved group from the central rainforest of Peru. *Willdenowia* 49: 5–9.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* 28: 2537–2539.

Pessoa EM, Alves M, Alves-Araújo A, Palma-Silva C, Pinheiro F. 2012. Integrating different tools to disentangle species complexes: A case study in *Epidendrum* (Orchidaceae). *Taxon* 61: 721–734.

Pinheiro F, Dantas-Queiroz MV, Palma-Silva C. 2018. Plant species complexes as models to understand speciation and evolution: A review of South American studies. *Critical Reviews in Plant Sciences* 37: 54–80.

Pintaud JC, Zehdi S, Couvreur T, Barrow S, Henderson S, Aberlenc-Bertossi F, Tregear J, Billotte N. 2010. Species delimitation in the genus *Phoenix* (Arecaceae) based on SSR markers, with emphasis on the identity of the date palm (*Phoenix dactylifera*). *Diversity, Phylogeny, and Evolution* 267–286.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. 2013. An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics* 4: 1–13.

Portères R. 1954. Le genre *Vanilla* et ses espèces. In: Bouriquet G ed. *Le vanillier et la vanille dans le monde*. Paris: P. Lechevalier. 94–920.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Queloz V, Duò A, Sieber TN, Grünig CR. 2010. Microsatellite size homoplasies and null alleles do not affect species diagnosis and population genetic analysis in a fungal species complex. *Molecular Ecology Resources* 10: 348–367.

R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.

Randrianasolo A, Miller JS. 1999. Taxonomic revision of the genus *Sarcolaena* (Sarcolaenaceae). *Annals of the Missouri Botanical Garden* 86: 702.

Renner S. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* 165: S23–S33.

Rieseberg LH, Carney SE. 1998. Plant hybridization. *New Phytologist* 140: 599–624.

Rolfe RA. 1896. A revision of the genus *Vanilla*. *Botanical Journal of the Linnean Society* 32: 439–478.

Rouhan G, Gaudeul M. 2021. Plant taxonomy: A historical perspective, current challenges, and perspectives. In: Besse P ed. *Molecular plant taxonomy: Methods and protocols*. New York: Springer. 1–38.

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 421–438.

Sharma SK, Dkhar J, Kumaria S, Tandon P, Rao SR. 2012. Assessment of phylogenetic inter-relationships in the genus *Cymbidium* (Orchidaceae) based on internal transcribed spacer region of rDNA. *Gene* 495: 10–15.

Simpson GG. 1951. The species concept. *Evolution* 5: 285–298.

Sluiter IRK, Blackburn DT, Holdgate GR. 2016. Fire and Late Oligocene to Mid-Miocene peat mega-swamps of south-eastern Australia: A floristic and palaeoclimatic interpretation. *Australian Journal of Botany* 64: 609.

Sokal RR. 1986. Phenetic taxonomy: Theory and methods. *Annual Review of Ecology and Systematics* 17: 423–442.

Soto Arenas MA, Cribb P. 2010. A new infrageneric classification and synopsis of the genus *Vanilla* Plum. ex mill. (Orchidaceae: Vanillinae). *Lankesteriana* 9: 355–398.

Stamos DN. 2013. Darwin's species concept revisited. In: Pavlinov I ed. *The species problem—Ongoing issues*. Rijeka, Croatia: InTech. 251–280.

Strijk JS, Noyes RD, Strasberg D, Cruaud C, Gavory F, Chase MW, Abbott RJ, Thébaud C. 2012. In and out of Madagascar: Dispersal to peripheral islands, insular speciation and diversification of Indian Ocean daisy trees (*Psiadia*, Asteraceae). *PLoS ONE* 7: e42932.

Tsai CC, Huang SC, Chou CH. 2005. Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcribed spacer of the nuclear ribosomal DNA. *Plant Systematics and Evolution* 256: 1–16.

Turini FG, Steinert C, Heubl G, Bringmann G, Lombe BK, Mudogo V, Meimberg H. 2014. Microsatellites facilitate species delimitation in Congolese *Ancistrocladus* (Ancistrocladaceae), a genus with pharmacologically potent naphthylisoquinoline alkaloids. *Taxon* 63: 329–341.

Van Valen L. 1976. Ecological species, multispecies, and oaks. *Taxon* 25: 233–239.

Vences M, Glaw F. 2003. Phylogeography, systematics and conservation status of boid snakes from Madagascar (*Sanzinia* and *Acrantophis*). *Rheinbach* 39: 181–206.

Vences M, Wollenberg KC, Vieites DR, Lees DC. 2009. Madagascar as a model region of species diversification. *Trends in Ecology & Evolution* 24: 456–465.

Villanueva-Viramontes S, Hernández-Apolinar M, Fernández-Concha GC, Dorantes-Euán A, Dzib GR, Martínez-Castillo J. 2017. Wild *Vanilla planifolia* and its relatives in the Mexican Yucatan Peninsula: Systematic analyses with ISSR and ITS. *Botanical Sciences* 95: 169.

Wang J, Luo J, Ma Y, Mao X, Liu J. 2019. Nuclear simple sequence repeat markers are superior to DNA barcodes for identification of closely related *Rhododendron* species on the same mountain. *Journal of Systematics and Evolution* 57: 278–286.

Warren BH, Strasberg D, Bruggemann JH, Prys-Jones RP, Thébaud C. 2010. Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* 26: 526–538.

Wikström N, Avino M, Razafimandimbison SG, Bremer B. 2010. Historical biogeography of the coffee family (Rubiaceae, Gentianales) in Madagascar: Case studies from the tribes Knoxieae, Naucleeae,

Paederieae and Vanguerieae: Biogeography of the coffee family in Madagascar. *Journal of Biogeography* 37: 1094–1113.

Wiley EO. 1978. The evolutionary species concept reconsidered. *Systematic Zoology* 27: 17.

Yoder AD, Heckman KL. 2006. Mouse lemur phylogeography revises a model of ecogeographic constraint in Madagascar. In: Fleagle J, Lehman SM eds. *Primate biogeography: progress and prospects*. New York: Springer. 255–268.

Younger JL, Dempster P, Nyári ÁS, Helms TO, Raherilalao MJ, Goodman SM, Reddy S. 2019. Phylogeography of the *Rufous Vanga* and the role of bioclimatic transition zones in promoting speciation within Madagascar. *Molecular Phylogenetics and Evolution* 139: 106535.

Yuan YM, Wohlhauser S, Möller M, Klackenberg J, Callmander MW, Küpfer P. 2005. Phylogeny and biogeography of *Exacum* (Gentianaceae): A disjunctive distribution in the Indian Ocean Basin resulted from long distance dispersal and extensive radiation. *Systematic Biology* 54: 21–34.

Zhang LB, Simmons MP, Renner SS. 2007. A phylogeny of Anisophylleaceae based on six nuclear and plastid loci: Ancient disjunctions and recent dispersal between South America, Africa, and Asia. *Molecular Phylogenetics and Evolution* 44: 1057–1067.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12858/suppinfo:

Fig. S1. Maximum Likelihood tree of leafless *Vanilla* species from SWIO region based on ITS region. Bootstrap support values (%) are displayed on the tree.

Fig. S2. A, Mean posterior probabilities of *K*, Ln $P(X \mid K)$, as implemented in STRUCTURE software (Pritchard et al., 2000). B, Distribution of ΔK values from STRUCTURE algorithm based on Evanno's method (Evanno et al., 2005).

Fig. S3. Principal coordinate analysis (PCoA) based on the mean squared Euclidean distance values between individuals for microsatellite data. A, PCoA plot of the two first axis. B, PCoA plot of the second and third axis. Colors represent the genetic groups of each individual based on STRUCTURE results. ALL (green), *V. allorgeae*; ATS (brown), *V. atsinananensis*; BOS (blue), *V. bosseri*; DEC (orange), *V. decaryana*; HUM (yellow), *V. humblotii*; HYB (black), hybrid individuals; MAD (red), *V. madagascariensis*; PER (magenta), *V. perrieri*; PHA (pink), *V. phalaenopsis*.

Table S1. List of specimens used in this study. Species names were defined according to the genetic structure based on a previous study (Andriamihaja et al., 2021), morphological description, and geographical location in the literature (Portères, 1954; Cribb & Hermans, 2009; Allorge-Boiteau, 2013). New samples specific to the microsatellite study are indicated by *. The remaining microsatellite data of leafless *Vanilla* species were taken from previous studies: Andriamihaja et al. (2021)[a], Gigant et al. (2016)[b], and Gigant et al. (2014)[d]. GenBank codes for new 54 ITS sequences are indicated by **. ITS sequence data of the 10 other samples were obtained from Besse et al. (2021).

Table S2. Summary of genetic diversity at seven microsatellite loci for each genetic group obtained by STRUCTURE analysis.

Table S3. Estimated divergence times from Bayesian relaxed clock analyses based on ITS markers and using an uncorrelated exponential relaxed clock. Mean and 95% HPD of the posterior probability distribution are in Myr. Node date of the common ancestor of the leafless *Vanilla* species from the SWIO region is taken from Bouetard et al. (2010).