Genotyping-by-sequencing diversity analysis of international Vanilla collections uncovers hidden diversity and enables plant improvement

Alan Chambers, Angélica Cibrián-Jaramillo, Adam P. Karremans, David Moreno Martinez, Juan Hernandez-Hernandez, Maria Brym, Marcio F.R. Resende Jr., Ruth Moloney, Sheryl N. Sierra, Tomas Hasing, Yasmin A. Alomia, Ying Hu, Vanilla Genotyping Consortium

Plant Science 311 (2021) 111019

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Research Article

Genotyping-By-Sequencing diversity analysis of international *Vanilla* collections uncovers hidden diversity and enables plant improvement

Alan Chambers [b,*], Angélica Cibrián-Jaramillo [c], Adam P. Karremans [g,h], David Moreno Martinez [i], Juan Hernandez-Hernandez [e], Maria Brym [b], Marcio F.R. Resende Jr. [f], Ruth Moloney [j], Sheryl N. Sierra [k], Tomas Hasing [d], Yasmin A. Alomia [a], Ying Hu [f], Vanilla Genotyping Consortium

[a] *Department of Biological Sciences, Faculty of Sciences, Universidad de Los Andes, Cra. 1 N° 18A - 12, Bogotá, Colombia*
[b] *Tropical Research and Education Center, Horticultural Sciences Department, University of Florida, 18905 SW 280th St, Homestead, FL, 33033, USA*
[c] *Ecological and Evolutionary Genomics Laboratory, Unidad de Genómica Avanzada (Langebio), CINVESTAV, Km 9.6 Carretera Irapuato-León, Guanajuato, CP 36824, Mexico*
[d] *Elo Life Systems, Durham, NC, 27709, USA*
[e] *Campo Experimental Ixtacuaco, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Km 4.5 Carretera Martínez de la Torre-Tlapacoyan, Veracruz, CP 93600, Mexico*
[f] *Horticultural Sciences Department, University of Florida, Gainesville, FL, USA*
[g] *Lankester Botanical Garden, University of Costa Rica, P.O. Box 302-7050, Cartago, Costa Rica*
[h] *Naturalis Biodiversity Center, Endless Forms Group, Sylviusweg 72, Leiden, 2333 BE, the Netherlands*
[i] *Posgrado en Ecología Tropical, Centro de Investigaciones Tropicales, Universidad Veracruzana, José María Morelos 44, Zona Centro, CP 91000, Xalapa, Veracruz, Mexico*
[j] *Corridgeree Belize Ltd, Mile 6, Southern Highway, Silk Grass, Stann Creek District, Belize*
[k] *College of Agriculture, Food, Environment and Natural Resources, Cavite State University, Indang, Cavite, 4122, Philippines*

ARTICLE INFO

This work is dedicated to Dr. Caroline Herron.

Keywords:
Vanilla planifolia
V. × *tahitensis*
Genotyping-by-sequencing
Diversity
Single nucleotide polymorphisms

ABSTRACT

Genomics-based diversity analysis of natural vanilla populations is important in order to guide conservation efforts and genetic improvement through plant breeding. Vanilla is a cultivated, undomesticated spice that originated in Mesoamerica prior to spreading globally through vegetative cuttings. Vanilla extract from the commercial species, mainly *V. planifolia* and *V.* × *tahitensis*, is used around the world as an ingredient in foods, beverages, cosmetics, and pharmaceuticals. The global reliance on descendants of a few foundational clones in commercial production has resulted in an industry at heightened risk of catastrophic failure due to extremely narrow genetic diversity. Conversely, national and institutional collections including those near the center of cultivation contain previously undiscovered diversity that could bolster the genetic improvement of vanilla and guide conservation efforts. Towards this goal, an international vanilla genotyping effort generated and analyzed 431,204 single nucleotide polymorphisms among 412 accessions and 27 species from eight collections. Phylogenetic and STRUCTURE analysis sorted vanilla by species and identified hybrid accessions. Principal Component Analysis and the Fixation Index (FST) were used to refine relationships among accessions and showed differentiation among species. Analysis of the commercial species split *V. planifolia* into three types with all *V.* × *tahitensis* accessions being most similar to *V. planifolia* type 2. Finally, an in-depth analysis of *V.* × *tahitensis* identified seven *V. planifolia* and six *V. odorata* accessions as most similar to the estimated parental genotypes providing additional data in support of the current hybrid theory. The prevalence of probable *V.* × *tahitensis* parental accessions from Belize suggests that *V.* × *tahitensis* could have originated from this area and highlights the need for vanilla conservation throughout Central and South America. The genetic groupings among

* Corresponding author at: Tropical Research and Education Center, Horticultural Sciences Department, University of Florida, 18905 SW 280th St, Homestead, FL, 33033, USA.

E-mail addresses: ac@ufl.edu (A. Chambers), angelica.cibrian@cinvestav.mx (A. Cibrián-Jaramillo), adam.karremans@ucr.ac.cr (A.P. Karremans), dmoreno29@hotmail.com (D. Moreno Martinez), hernandezh.juan@inifap.gob.mx (J. Hernandez-Hernandez), maria.brym@ufl.edu (M. Brym), mresende@ufl.edu (M.F.R. Resende), ruth.moloney@6ptp.com (R. Moloney), snsierra@gmail.com (S.N. Sierra), tengfang.huang@elolife.ag (T. Hasing), ya.alomia10@uniandes.edu.co (Y.A. Alomia), ying.hu@ufl.edu (Y. Hu).

https://doi.org/10.1016/j.plantsci.2021.111019
Received 2 June 2021; Received in revised form 4 August 2021; Accepted 7 August 2021
Available online 13 August 2021

accessions, particularly for *V. planifolia*, can now be used to focus breeding efforts on fewer accessions that capture the greatest diversity.

1. Introduction

Vanilla planifolia is an undomesticated, wild species that was originally cultivated in Mesoamerica. *V. planifolia* Andrews [1], the primary commercial species, was spread globally as part of the spice trade from the new world [2,3]. Vanilla was successfully propagated clonally by cuttings in European gardens in the 1800s and eventually spread to today's major commercial production areas including Madagascar [4]. In nature, the flowers of *V. planifolia* and its close relatives are dependent on the Neotropical, endemic Euglossini bees to become fertilized and form seed capsules commonly called beans [5–7]. Producing vanilla beans outside its native range was elusive until a method of artificial pollination was discovered in 1838 by Charles Morren in France with modifications in 1841 by Edmond Albius in Reunion, an island nation off the coast of Madagascar [8,9]. Vanilla thrived in new growing locations and artificial pollination enabled production of this high-value spice around the world. Today's commercial vanilla cultivation still heavily relies on vegetatively propagated descendants of those foundational clones, and, as a result, vanilla has often been reported as having limited genetic diversity [7]. The recent publication of a chromosome-scale *V. planifolia* genome and demonstration of genomics-based diversity analysis for vanilla have unlocked the potential to characterize vanilla diversity with unprecedented resolution [10,11]. Combining these tools to analyze vanilla diversity in Central and South America are key to understanding this commercially important genus and to guide conservation and plant breeding efforts.

The genus *Vanilla* comprises about 118 accepted species in the Orchidaceae family [12]. *Vanilla* species have a pantropic distribution, occurring naturally in the New World tropics, from southern Florida and Mexico, Central America and the Antilles, to South America; in tropical Africa, both on the mainland and on Madagascar; and in tropical Asia, from China and India to New Guinea and Southeast Asia [13]. Just over half of all vanilla species are endemic to tropical America, and only a subset of those new world species bear the highly esteemed aromatic fruits that contain vanillin. The aromatic vanilla species, which constitute the secondary gene pool of the vanilla crop [14], can be found growing in the wild from Florida to Argentina. The commercially important *V. planifolia* itself is only known to occur naturally in Mexico, Central America, and perhaps certain areas in northern South America [7,12].

The global spread of *V. planifolia* followed the early spice trade routes. The species arrived in Europe as early as 1510 with successful cultivation as early as 1807 [15]. Vanilla spread from Europe and into Africa, India, Reunion Island, and Madagascar throughout the 1800s [16]. A trans-Pacific migration through Hawaii and the Philippines included dispersion of *V. planifolia* and *V.* × *tahitensis* [17], though accurately and consistently identifying vanilla species has been a major challenge. Therefore, there has been increasing interest in accurately characterizing vanilla accessions within and among species.

The narrow genetic base of *V. planifolia* in commercial cultivation is a major risk to global supply chains as clones can all be similarly susceptible to prevalent pathogens including *Fusarium* and viruses [13,18]. Additionally, as of 2011, up to 50 % fruit drop has been recorded in Mexico possibly due to average increases in temperature, reduction in humidity, and/or the incidence of pathogens that exacerbate yield losses [19]. As a result, there is a great need to identify diversity with the commercial species in order to detect novel genetics and reduce production risks. These efforts can be achieved through the strategic cultivation of diverse, existing accessions and through plant breeding to create novel genetic combinations with improved plant performance. This approach is more efficient when guided by morphological, phenotypic, or genomics-based information.

Vanilla extract is the world's second most valuable spice and is made from cured vanilla beans. Green vanilla beans have little aroma and must be cured in order to develop their characteristic flavor profile. Curing methods include thermal treatments to disrupt cellular organization and support enzymatic activity with a final drying treatment to stabilize beans for shipping prior to extraction in aqueous ethanol. Today, the United States and European Union are the two largest importers of cured vanilla beans. *V. planifolia* and *V.* × *tahitensis* are the only permissible species for use in making vanilla extract [20]. *V. pompona* is an additional, edible vanilla species consumed locally in Central and South America, but was omitted from the original Standard of Identity. Hybrids of *V. planifolia* with *V. pompona* show good growth, resistance to higher temperatures, and fusarium resistance, but the best *V. pompona* accessions for hybrid populations still need to be elucidated. The xerophytic character of *V. pompona* could also be important in a genetic improvement program due to climatic changes in traditional growing areas where higher temperatures and lower rainfall are becoming more prevalent [21]. Additionally, the presence of vanillin, protocatechuic acid, p-hydroxybenzoic acid, vanillic acid, phydroxybenzaldehyde, p-anisyl in *V. pompona* give it a high potential for the perfumery industry as an alternative to the food industry [22].

The largest producers of *V. planifolia* beans include Madagascar, Indonesia, Mexico, Papua New Guinea, China, Turkey, and Tonga [23,24]. *V.* × *tahitensis* is primarily grown in Tahiti and Papua New Guinea, has a distinct flavor profile compared to *V. planifolia*, and is generally accepted as a hybrid between *V. planifolia* and *V. odorata*. This was supported by a report including 21 additive SNPs among *V. planifolia*, *V. odorata*, and *V.* × *tahitensis* accessions by sequencing the Internal Transcribed Spacer (ITS) region [25]. Unfortunately, incomplete data reporting and lack of access to verified, living *V.* × *tahitensis* material has inhibited in-depth analyses with this hybrid leaving many outstanding questions for this commercially-important genotype.

Various molecular tools have been used to identify species, characterize diversity, identify hybrids, and resolve genetic relationships among vanilla accessions. Methods have included isozymes [26,62], RAPDs [27–29], AFLPs [4,30,31], microsatellites [32–35], plastid-derived single gene sequences [36–40], and others [25,41–45]. While useful, many of the molecular tools in these studies had limitations including utility of the data generated, confidence in the correct species assignments of accessions tested, or the low number of accessions analyzed. The recent publication of a genomics-based single nucleotide polymorphism diversity study of 112 accessions generated the greatest resolution on vanilla diversity to date, but was limited by the diversity of material available for testing [11]

The purpose of the present study was primarily to leverage new genomics tools to advance our current understanding of vanilla diversity within international collections. This was accomplished by focusing on wild-collected, living collections from multiple countries primarily throughout Mexico, Central America, and South America. This is the first time that an international study of vanilla germplasm has been analyzed at such high resolution. A genomics-based approach overcomes many limitations of confounding morphological descriptors that can be affected by growing environment, advances our capability to assign species for living specimens, and identifies gaps in existing collections. This foundational information is critical for establishing and diversifying vanilla breeding programs and highlights the critical need to collaboratively conserve vanilla diversity within and across countries particularly in Mexico, Central America, and South America.

2. Materials and methods

2.1. Vanilla plant accessions

Accessions in this study came primarily from international collections (Table 1, Supplementary Table S1). Accessions from Belize were provided by Corridgeree Belize Ltd and the Belize National Herbarium. This collection was made both by collecting specimens growing in coastal tropical forests (with wildlife collection permits) and taking leaf samples from the vanilla collection in the Belize National Herbarium. Accessions from Colombia were provided by the Universidad de Los Andes. The material was collected directly in the field from wild populations that cover different ecosystems in Colombia. Identity was verified for most of the samples based on floral morphological features. In some cases, vegetative traits suggested the formation of hybrids. Costa Rican material from Jardin Botanico Lankester and the University of Costa Rica was collected from natural areas. Three collections from Mexico including Langebio Centro de Investigación y Estudios Avanzados, Unidad de Manejo para la conservación de la Vida Silvestre Xochitlcalli (SEMARNAT: UMA-IN-VIV-0203-VER/15), and the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) represent vanilla diversity collected over time with various levels of phenotypic characterization. INIFAP accessions were maintained on forest tutor trees similar to their natural habitat. Accessions from the Philippines were provided by Cavite State University and the collection was founded by purchasing vanilla material throughout the country from commercial vendors with one accession provided by Dr. Calixto M. Protacio at the Institute of Crop Science, College of Agriculture and Food Science, University of the Philippines Los Baños Laguna. These accessions could include cuttings from naturalized populations that are descendants of the early trans-Pacific migration from Central America. The University of Florida vanilla collection in the United States of America includes accessions from botanical gardens, online vendors, and hobbyists. The living collection is maintained on raised beds with 16 cm of mulch in a shade house under 70 % filtered light. Individual trellises are spaced 1.2 m apart and include a central post, a column of wire mesh as a climbing substrate, and horizontal wooden post for vine support at 1.2 m from the ground. Supplemental irrigation is supplied through drip lines during the dry season. Average temperatures in Homestead, Florida vary between 18 °C in the winter months to 27 °C in the summer. Average rainfall is ~4.4 cm per month in the dry season (November to March), and up to ~20.6 cm on average per month during the peak of the rainy season (June to September) (Florida Automated Weather Network, http://fawn.ifas.ufl.edu)[52]. Additional information on this collection has been published previously [11].

Table 1
Vanilla accessions from multiple international collections that were included in this study and genotyped for the first time. Country of origin, supporting institution, number of accessions provided, and the number of included species from each are shown. There were 291 accessions genotyped for the first time as part of this study with overlapping representatives from at least 26 species. Note, most collections are larger than the number of accessions and species that were selected for this study.

Country	Institution	Accessions	Species
Belize	Corridgeree Belize Ltd/Belize National Herbarium	18	1
Colombia	Universidad de Los Andes	14	10
Costa Rica	Jardin Botanico Lankester - University of Costa Rica	14	11
Mexico	Instituto Nacional de Investigaciones Forestales, Agrícolas, y Pecuarias (INIFAP)	12	3
Mexico	Langebio Centro de Investigación y Estudios Avanzados	30	2
Mexico	Unidad de Manejo para la conservación de la Vida Silvestre Xochitlcalli. SEMARNAT: UMA-IN-VIV-0203-VER/15	59	1
Multiple	Marie Selby Botanical Gardens	11	6
Philippines	Cavite State University	27	2
United States	University of Florida	106	12
	Total	291	26

2.2. DNA extraction

DNA extractions were performed using a modified CTAB method at the University of Florida. For fresh tissue, 250 mg of mature leaf was freeze dried for 24 h and ground to a fine powder in a modified paint shaker using two stainless steel beads in a 2 ml screw cap vial. For desiccated samples, tissues were freeze dried for 24 h upon arrival and then ground to a fine powder as for fresh samples. All ground tissue was mixed with 3 ml of DNA extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris, 2 mM EDTA, 1% β-mercaptoethanol, pH 8.0) in a 15 ml conical tube, vortexed to completely disperse the tissue, and then incubated at 65C° for 10 min. Samples were then cooled to room temperature for 5 min and 3 ml chloroform:isoamyl alcohol (24:1 v/v) was added to each sample, vortexed, and centrifuged for 10 min at 18,000×*g*. The supernatant was transferred to a fresh tube to precipitate DNA using two volumes of 95 % ethanol. Samples were centrifuged at 10,000×*g* for 10 min to pellet DNA, aspirated, washed once with one ml of 70 % ethanol, centrifuged at 10,000×*g* for 2 min, aspirated, and finally resuspended in 75 μl molecular biology grade water.

2.3. Sequencing vanilla accessions

Library preparation and sequencing for genotyping-by-sequencing (GBS) was performed as previously described [11]. Briefly, libraries were constructed using dual restriction enzymes BamHI and NsiI and sequencing was performed using NextSeq 1 × 150 bp with 300–744 bp size selection. Target sample depth was 0.5 million reads per sample. All reads were quality filtered by the University of Minnesota Genomics Center as previously reported and screened for contamination using FastQ Screen (default parameters).

2.4. SNP calling

SNP calling leveraged the phased *V. planifolia* 'Daphna' genome sequence [10]. Raw reads were processed with fastq-mcf v1.05 (-q 30 -l 50) and then aligned against the reference assembly of haplotype A with HISAT2 v2.1 (default parameters). The bam files were merged and read groups were added using bamaddrg 3fccbf0. Joint variant calling was performed with FreeBayes v1.3.1-19-g54bf409 (default parameters). Variants were then filtered with VCFtools v0.1.17 (–max-missing 0.3 –maf 0.05 –min-meanDP 10 –max-meanDP 1000). The resulting SNP set was used for phylogentic, STRUCTURE, and FST analyses. A separate SNP set was created that only included *V. planifolia* and *V.* × *tahitensis* accessions. SNPs from these accessions were filtered using the same parameters as for the complete dataset.

2.5. Phylogenetic analysis

The filtered VCF files were converted to Phylip format by concatenating the SNPs with PGDSpider v.2.0.9.0 [46] with IUPAC ambiguity codes for polymorphic data. The maximum likelihood (ML) phylogenetic tree with 1000 rapid bootstrap inference was constructed by using RAxML v.8.2.12 [47]. The analysis was run using an ascertainment bias correction (ASC) for the data set containing only concatenated informative SNP positions, and a general time-reversible substitution model accounting for among-site rate heterogeneity (ASC_GTRGAMMA model). The RAxML results were graphically visualized in FigTree v.1.4.3 (downloaded from http://tree.bio.ed.ac.uk/software/figtree/).

2.6. Population STRUCTURE

The population structure was investigated to identify clusters of genetically related individuals using the Bayesian clustering method implemented in STRUCTURE, v.2.3.4 [48]. Ten independent STRUCTURE runs were performed for each of K = 2-16 (K = number of genetic clusters) with a burn-in of 10000 and 20000 iterations. The number of clusters that best fit the observed genotype data was inferred by examining the average and standard deviation (SD) of the natural log probability of each model [49] using the online version of STRUCTURE HARVESTER [50]. Additionally, the R package SNPRelate [51] was used to perform principal component analysis (PCA) and the R package adegenet [52] was used to perform discriminant analysis of principal components (DAPC). Loci were trimmed to linkage disequilibrium (LD < 0.2) using the command "snpgdsLDpruning". DAPC was performed on all individual genotypes by using the successive K-means approach, implemented by the find.clusters function, to identify the optimal number of groups based on Bayesian Information Criterion (BIC). PCA plots were created using JMP Pro v15.

To investigate the diversity of *V. planifolia* further, the *V. planifolia*-only PCA was performed on 27 accessions while excluding three diverse *V. planifolia* lines (AC108, AC181, and AC185) to improve resolution among the remaining accessions. Pairwise Fst were calculated with the Weir and Cockerham formula [53] with the SNPRelate package [51].

2.7. rbcL sequencing

Partial sequencing of the *rbcL* gene was used to confirm species identifications of selected accessions and to identify the maternal parent of probable hybrids. Primers *rbcL* forward 5' CTTCACAAGCAGCAGCTAGTTC 3' and *rbcL* reverse 5' ATGTCACCACAAACAGAAAC 3' were used to amplify an ~1,300 bp fragment of the *rbcL* gene in 25 μl reactions using Phusion High-Fidelity polymerase (F531, Thermo Scientific) according to the manufacturer's instructions. PCR amplification was conducted using an Applied Biosystems SimpliAmp Thermal Cycler with the following program: initial denaturation 5 min at 95 °C followed by 30 amplification cycles of 30 s at 95 °C, 30 s at 55 °C, and 60 s at 72 °C, with a final 10 min extension at 72 °C. Sanger Sequencing was conducted at GENEWIZ (GENEWIZ, South Plainfield, NJ) using *rbcL* forward as the sequencing primer. Sequences were verified against published vanilla *rbcL* sequences from NCBI (ncbi.nlm.nih.gov) using the default settings for MUSCLE alignment [54] and Geneious Tree Builder within the Geneious Software package (Geneious version 11.1, Newark, NJ). Accessions AC289_2, AC351_2, AC361_2, AC361_2, AC399_2, AC433_2, AC471_2, AC504_2, AC524_2, AC525_2, AC526_2, and AC534_2 were selected for sequencing. AC361_2 was sequenced twice with independent PCR and sequencing reactions to confirm the presence of a single SNP when compared to the typical sequence for *V. planifolia* type 1 accessions.

2.8. Parentage of V. × tahitensis

The 431,204 SNPs from 412 accessions both in this study and the previous vanilla GBS study were used to investigate the paternal parent of *V.* × *tahitensis*. The total SNPs were filtered using vcftools (–max-missing 0.5 –min-meanDP 10 –max-meanDP 1000 –remove-indels –max-alleles 2). The 143,122 resulting SNPs were then screened for homozygosity (0/0) in at least 9 out of 10 accessions in a subset of *V. planifolia* (AC275_2, AC277_2, AC281_2, AC285_2, AC293_2, AC314_2, AC368_2, AC372_2, AC375_2, AC504_2) and heterozygosity in 5 out of 6 *V.* × *tahitensis* accessions (AC421_2, AC432_2, AC433_2, AC509_2, AC526_2, AC237_2). The resulting SNPs were then used to identify other accessions with the alternative allele (0/1 or 1/1) to identify those that might have contributed the minor allele found in *V.* × *tahitensis*.

A similar approach to that described above was taken to identify the accessions most closely related to the maternal *V. planifolia* parent of *V.* × *tahitensis*. The same 143,122 SNPs were screened among *V. odorata* accessions (AC177, AC213, AC207, AC212_2, AC458_2, and AC341_2) for homozygous reference alleles (0/0) and that were heterozygous (0/1) in the *V.* × *tahitensis* accessions. There were 468 SNPs identified after filtering. All accessions were then screened for the presence of the alternative allele (0/1 or 1/1) for the 468 SNPs.

2.9. Data availability

Raw sequencing data is available on NCBI at part of Bioproject PRJNA733522. *rbcL* sequences have also been deposited at NCBI for AC289_2 (MZ391815), AC351_2 (MZ391816), AC361_2a), MZ391817), AC361_2b), MZ391818), AC399_2 (MZ391819), AC433_2 (MZ391820), AC471_2 (MZ391821), AC504_2 (MZ391822), AC524_2 (MZ391823), AC525_2 (MZ391824), AC526_2 (MZ391825), and AC534_2 (MZ391826).

3. Results

3.1. Vanilla accessions included in this study

There were 291 new accessions genotyped as part of this study as summarized in Table 1. Institutions from Belize, Colombia, Costa Rica, Mexico, the Philippines, and the United States submitted material from their domestic collections for genotyping. Library preparation and sequencing of accessions were similar to a previous *Vanilla* GBS study [11]. As a result, the 291 accessions were able to be combined with publicly available *Vanilla* GBS data resulting in a total of 412 accessions from 27 species available for combined analyses (Supplemental Table S1).

Species in this study included *V. atropogon*, *V. barbellata*, *V. bicolor*, *V. claviculata*, *V. dilloniana*, *V. dresslerii*, *V. griffithii*, *V. hartii*, *V. helleri*, *V. imperialis*, *V. insignis*, *V. karen-christianae*, *V. lindmaniana*, *V. madagascariensis*, *V. mexicana*, *V. odorata*, *V. palmarum*, *V. phaeantha*, *V. phalaenopsis*, *V. planifolia*, *V. poitaei*, *V. pompona*, *V. roscheri*, *V. siamensis*, *V. sotoarenasii*, *V.* × *tahitensis*, and *V. trigonocarpa*. Three additional designations including "low depth", "hybrid", and "species" were assigned based on sequencing depth, high likelihood of being a hybrid, and uncertain species assignments, respectively. *V. planifolia* represented ~41 % of the 412 accessions in this study, and hybrid and unknown accessions included 10.4 % and 3.9 % of accessions, respectively. Some species designations could not be fully resolved and will require follow up research for targeted morphological and genetic confirmation.

3.2. Single nucleotide polymorphism analysis

Target read depth was 0.5 million reads per sample for accessions in this study. After sequencing, there were 185,168,710 total reads with an average of 636,318 reads per sample generated from the 291 accessions. Overall, 3.83 % of the reads matched chloroplast sequences and there was only minor contamination as measured by FastQ Screen (0% *E. coli*, 0.15 % human, and 2.36 % mouse hits). All GBS reads from 412 accessions were mapped to the phased *V. planifolia* 'Daphna' genome resulting in detection of 431,204 Single Nucleotide Polymorphisms (SNPs) using Freebayes. There were 15,746 SNPs after filtering based on missing data, allele frequency, and depth that were used for all analyses unless otherwise specified.

3.3. Heterozygosity

Heterozygosity was analyzed among the sequenced accessions to look for patterns among groups of accessions and species (Supplemental Table S1). As expected, accessions annotated as putative hybrids were among the highest heterozygous genotypes with a maximum of 31.3 %

for a verified *V. planifolia* x *V. pompona* hybrid (AC445_2). The average heterozygosity of all accessions annotated as hybrids was 15.3 %. Heterozygosity can be impacted by reads that map or do not map to the reference sequence due to diversity, and this could be reflected in the lower heterozygosity in general among the most diverse species tested. Conversely, analysis of *V. planifolia* accessions using the 'Daphna' genome was not limited in this way, and thus the results could be more rigorously interpreted. Natural breaks in the data were noticed when analyzing *V. planifolia* heterozygosity. One group of *V. planifolia* accessions included most of the accessions for this species (*V. planifolia* type 1) and had an average heterozygosity of 22.7 %. Two other *V. planifolia* groups had average heterozygosity levels of 12.2 % (*V. planifolia* type 3: AC185, AC364_2, AC366_2, AC368_2, AC381_2, AC382_2, AC535_2, and AC536_2) and 2.6 % (*V. planifolia* type 2: AC205, AC206, AC289_2, AC293_2, AC295_2, AC399_2, AC400_2, AC504_2, AC517_2, AC525_2, and AC540_2). These groups were also generally distinguished using a focused phylogenetic analysis of just *V. planifolia* and *V.* × *tahitensis* accessions as described in 3.8.

3.4. Phylogenetic and STRUCTURE analyses

Phylogenetic and STRUCTURE analyses were conducted to investigate relationships among the 412 vanilla accessions (Fig. 1, Supplemental Fig. S1). In general, species were uniquely separated into clades. As found previously [11], there were a number of misidentified, hybrid, or unknown accessions indicating the need to verify accessions using genomics approaches prior to investing heavily in plant breeding. Phylogenetic and STRUCTURE analyses were suitable in many cases to confidently assign species or hybrids within clusters, but this was not possible in every case where "species" was assigned to unknowns (Table S1). *V. planifolia* accessions formed a number of branches in the phylogenetic tree. Some branches were country-specific. For example, four accessions (AC525_2, AC524_2, AC399_2, and AC400_2) from Colombia formed one branch and included both *V. planifolia* and one *V.* × *tahitensis* (AC524_2) accessions. There were many branches that included accessions from both Mexico and the Philippines spread throughout the *V. planifolia* accessions. Belize had fewer accessions available for this study, but these were also dispersed among the *V. planifolia* accessions.

The GBS results also revealed some relationships in improved detail compared to previous studies. The *V.* × *tahitensis* accessions were similarly located on the phylogenetic tree and distinct from *V. planifolia* in the STRUCTURE plot. Three *V. sotoarenasii* accessions were found within the *V. planifolia* group with STRUCTURE showing similarity to *V. planifolia* and *V.* × *tahitensis*. *V. karen-christianae* formed a distinct group compared to other accessions.

3.5. Principal component analysis

Principal component analysis was conducted on all 412 accessions (Fig. 2a, Supplemental Table S2). Principal component 1 (PC1) captured 21.85 % of the variance, and Principal Component 2 (PC2) captured 17.59 % of the variance. *V. insignis* heavily skewed PC1. *V. planifolia* accessions were generally located together except for some low depth accessions including AC310_2, AC317_2, AC323_2, AC325_2, AC339_2, AC342_2, AC354_2, AC363_2, AC365_2, AC371_2, and AC397_2.

V. × *tahitensis* was found roughly between the *V. planifolia* and *V. odorata* groups when plotting PC1 and PC2, but PC2 and Principal Component 4 (PC4) showed better separation of *V. planifolia*, *V.* × *tahitensis*, and *V odorata*. The PC2 and PC4 plot (Fig. 2b) identified 11 *V. planifolia* type 2 accessions (AC205, AC206, AC289_2, AC293_2, AC295_2, AC399_2, AC400_2, AC504_2, AC517_2, AC525_2, and AC540_2) that were separated from the main *V. planifolia* group that tended towards *V.* × *tahitensis* along PC4. This is the same group of accessions that had an overall heterozygosity of 2.6 %. The three *V. sotoarenasii* accessions were also located among the *V. planifolia* type 2 accessions in Fig. 2b.

3.6. Global genetic variance analysis

Total genetic variance contained within subpopulations was analyzed by FST to investigate population differentiation due to genetic structure (Fig. 3). Higher FST values indicate higher differentiation among populations. As expected, *V. planifolia*, *V.* × *tahitensis*, and *V. odorata* showed low FST values in comparisons among the three groups. High FST values were found among a number of diverse species including *V. palmarum*, *V. trigonocarpa*, *V. dresslerii* and *V. griffithii*, *V. siamemsis*, *V. atropogon*, *V. madagascariensis*, *V. phalaenopsis*, *V. roscheri*, *V. imperialis*, *V. potaei*, *V. claviculata*, *V. dilloniana*, and *V. barbellata*.

3.7. Species identification using rbcL sequencing

A few accessions were selected for *rbcL* (ribulose-1,5-bisphosphate

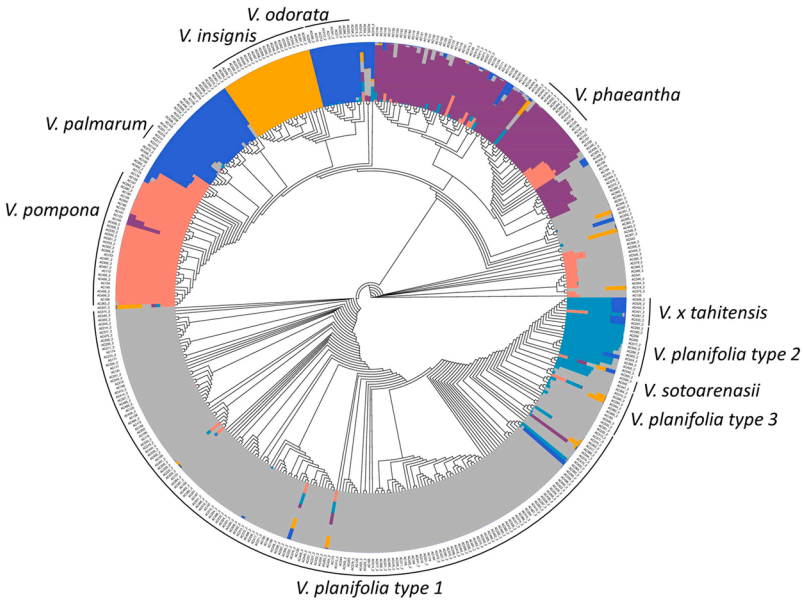

Fig. 1. Combined STRUCTURE (K = 6) and Circos plot of samples from this study (designated with "_2") and accessions from a previous study [11]. Selected species and types including *V.* × *tahitensis*, *V. planifolia* type 2, *V. sotoarenasii*, *V. planifolia* type 3, *V. planifolia* type 1, *V. pompona*, *V. palmarum*, *V. insignis*, *V. odorata*, and *V. phaeantha* are shown. Color coding in the STRUCTURE plot indicates similar SNPs patterns with gray indicating *V. planifolia* type 1 (as an example). See online version for color figure.

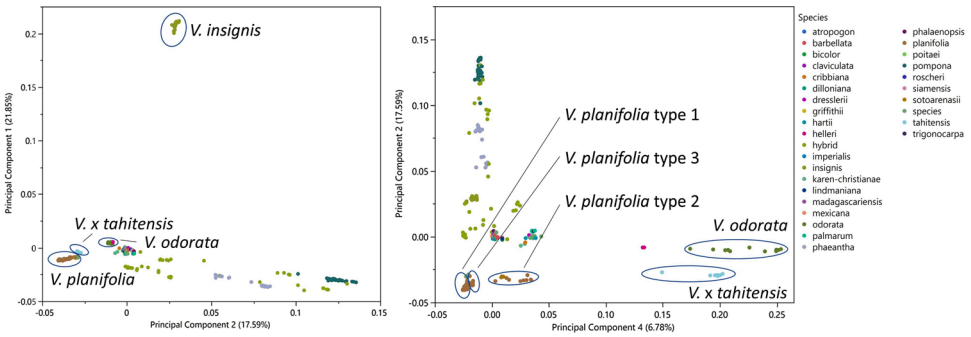

Fig. 2. Principal Component Analysis plots of 412 vanilla accessions for a) PC1/PC2 and b) PC2/PC4. Plots show multiple species including *V. planifolia*, *V.* × *tahitensis*, *V. odorata*, and *V. insignis* accessions within ovals. *V. planifolia* types 1, 2, and 3 are shown with greater differentiation in PC2/PC4. Species assignments are coded by color. See online version for color figure.

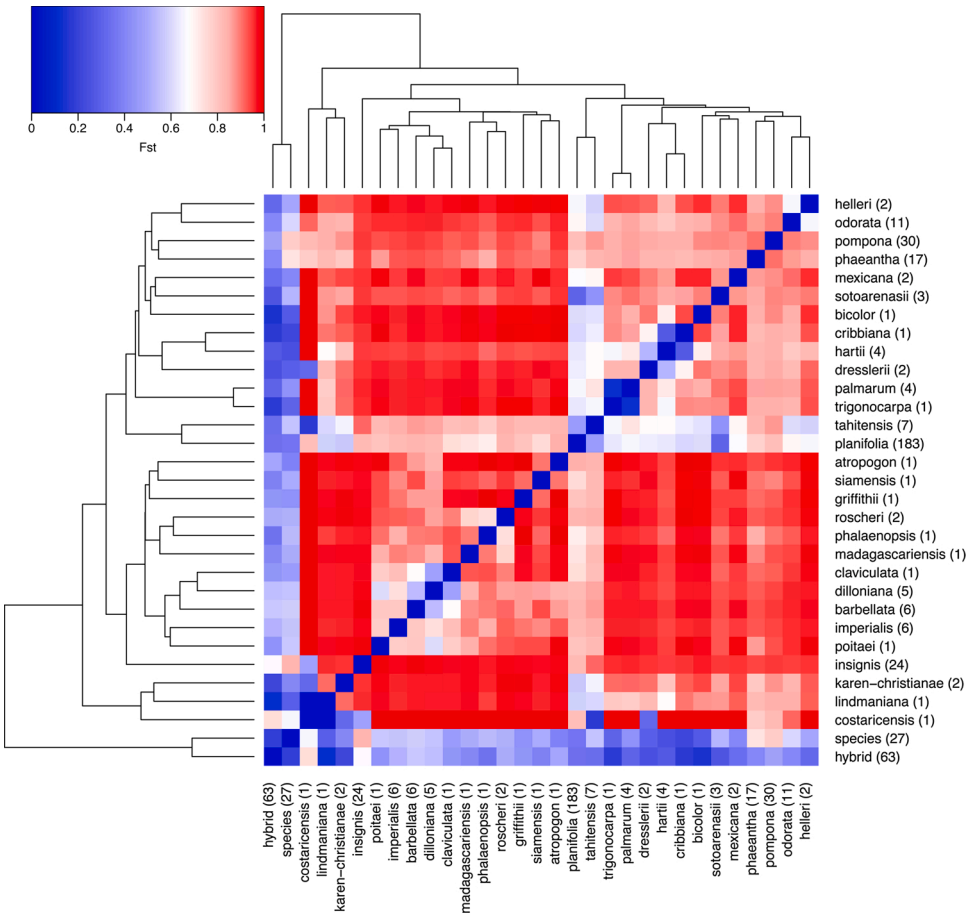

Fig. 3. FST analysis of species in this study. Higher FST values (1.0) are shown in red and indicate a higher degree of differentiation among populations. Lower Fst values are shown in blue and indicate lower degrees of differentiation among populations. See online version for color figure.

carboxylase) sequencing to confirm species assignments and to putatively identify maternal parents (Fig. 4). *rbcl* is a chloroplast-encoded, maternally-inherited gene. AC289_2, AC399_2, AC504_2, and AC525_2 were from *V. planifolia* type 2 and clustered closely with *V.* × *tahitensis*. These accessions had sequences identical to the most abundant *V. planifolia* type 1 sequences as expected. Two *V. planifolia*-like accessions, AC351_2 and AC361_2 were found on the phylogeny near the *V. planifolia/V. phaeantha* and *V. planifolia/V. pompona* hybrids. AC351_2 is from the Tuxtlas jungle in Veracruz Mexico where wild *V. planifolia* was collected more than 100 years ago and where crops were never established. This accession had the normal *V. planifolia rbcL* sequence, but had a much lower heterozygosity (2.2 %) than most *V. planifolia* (22.7 %) and was not remarkably deficient for sequencing depth or percent missing data. AC361_2 is from the jungle of Chimalapas Oaxaca, Mexico in the same region that included sterile plants that were morphologically similar to *V.* × *tahitensis* [55]. This accession had one SNP in the *rbcL* sequence distinguishing it from the typical *V. planifolia*, had low mean SNP depth (4.0) that could indicate poor quality sequencing, and is morphologically similar to *V. planifolia* type 1 accessions. AC433_2 and AC526_2 are *V.* × *tahitensis* accessions that matched the *V. planifolia rbcL* sequence as expected due to *V. planifolia* being the maternal parent. AC534 was identified as a *V. karen-christianae* and had an *rbcL* sequence identical to *V. odorata*, but was clearly separated from *V. odorata* on a separate branch in the phylogenetic tree.

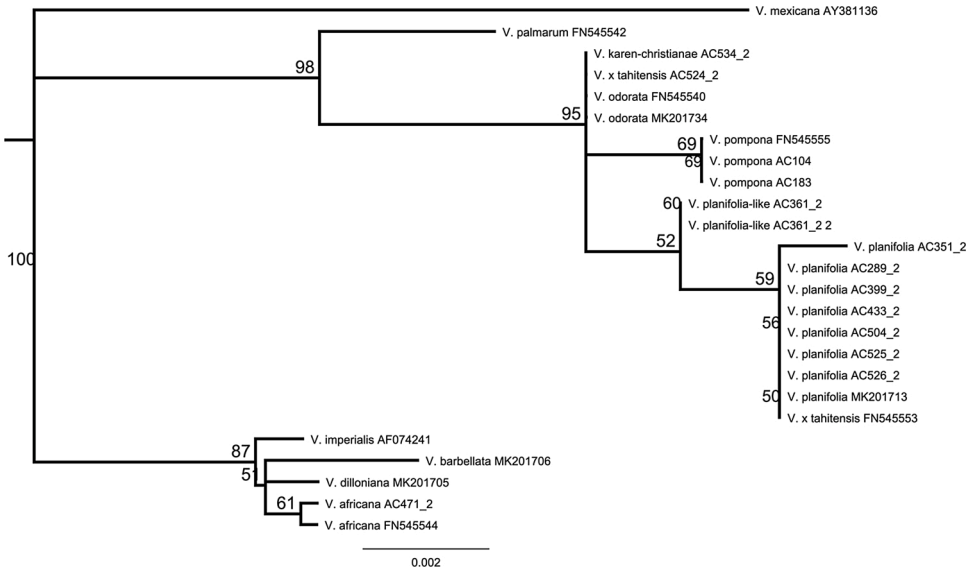

Fig. 4. Species analysis using the plastid-encoded, partial *rbcL* gene of selected accessions. Selected species from NCBI are also shown and include GenBank IDs. Accessions tested from this study mostly matched *V. planifolia* type 1 sequences except for AC524_2 (a unique *V.* × *tahitensis*) and AC534_2 (*V. karen-christianae*) that matched *V. odorata,* and AC471_2 that matched *V. africana*. AC361_2 had a single SNP differentiating it from *V. planifolia* type 1.

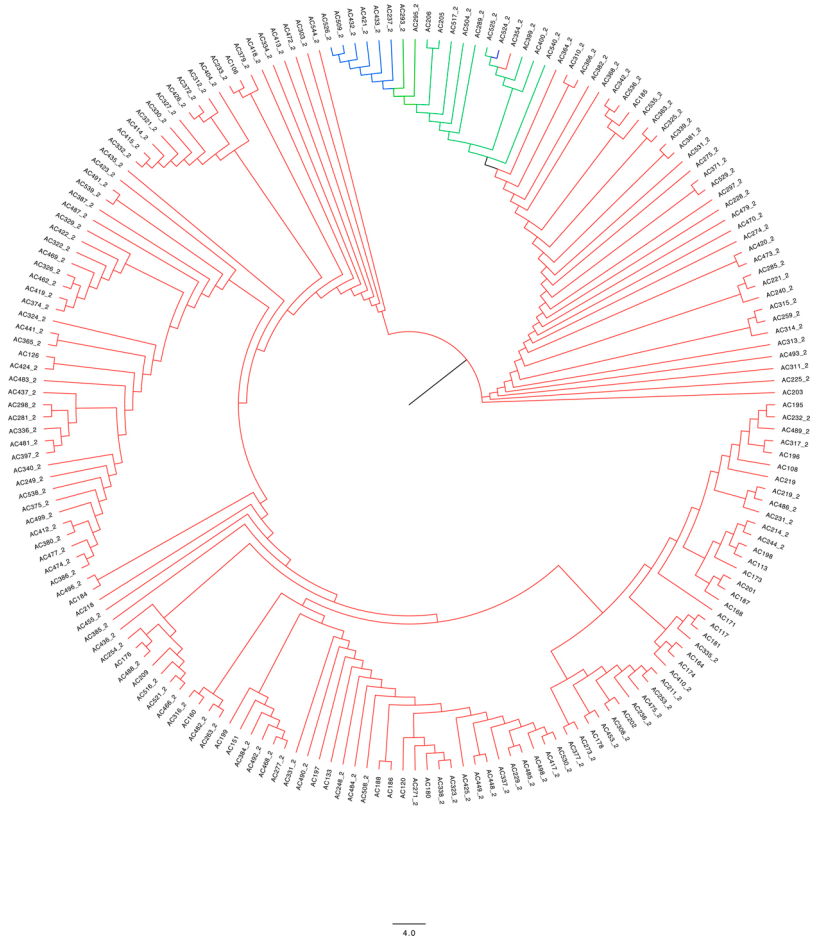

Fig. 5. Phylogenetic tree including only *V. planifolia* and *V.* × *tahitensis* accessions. The figure shows a major split among the *V. planifolia* accessions. *V.* × *tahitensis* (blue) accessions are grouped together and mostly distinct from *V. planifolia* type 1 (green). *V. planifolia* types 1 and 3 are shown in red. Accession 354_2 (red) is shown among *V. planifolia* group 2 accessions (green), but had low sequencing depth and might not be placed accurately in the tree.

AC471_2 from the Philippines was most similar to a *V. africana* sequence (GenBank FN545544) deposited on NCBI and was most similar to AC537_2 (a putative *V. havilandii*) on the phylogenetic tree. AC524_2 was identified as *V.* × *tahitensis* from the phylogenetic and STRUCTURE analyses, but the *rbcL* sequence matched *V. odorata* indicating this this accession is most likely a *V. odorata* x *V. planifolia* hybrid versus the more commonly cultivated *V.* × *tahitensis*, a *V. planifolia* x *V. odorata* hybrid.

3.8. Diversity within *V. planifolia* and *V.* × *tahitensis*

The diversity within *V. planifolia* and *V.* × *tahitensis* is especially valuable due to the current vanilla Standard of Identity that only permits these two types for inclusion in vanilla extract. Therefore, the most straightforward approach (from a regulatory perspective) for the genetic improvement of vanilla includes identifying and leveraging diversity within these two types of vanilla. A separate analysis was conducted that only included *V. planifolia* and *V.* × *tahitensis* accessions. The SNPs from these accessions were filtered using the same parameters as the complete dataset and resulted in 14,133 SNPs. Analysis of *V. planifolia* and *V.* × *tahitensis* using the 14,133 SNPs showed that the accessions split into two groups with one comprised of *V. planifolia* type 1 accessions and the other with *V. planifolia* type 2, *V. planifolia* type 3, and *V.* × *tahitensis* accessions (Fig. 5). The 11 *V. planifolia* type 2 accessions that were differentiated in Fig. 2b were located closest to the *V.* × *tahitensis* accessions in Fig. 5.

PCA analysis of 190 *V. planifolia*, *V.* × *tahitensis*, and *V. sotoarenasii* accessions showed a similar trend to the phylogenetic analysis (Fig. 6). PC1 explained 27.33 % of the variance and PC2 explained 7.98 %. The *V. planifolia* types 1, 2, and 3 showed some separation from each other. *V.* × *tahitensis* was separated from *V. planifolia* in the PCA plot, and *V. sotoarenasii* accessions were located close to the *V. planifolia* type 2 accessions.

3.9. Genetic variance within *V. planifolia* and *V.* × *tahitensis*

FST analysis was also performed on just the 190 *V. planifolia* and *V.* × *tahitensis* accessions along with three additional *V. sotoarenasii* accessions (Fig. 7). The *V. planifolia* accessions were split by type for this analysis. For *V.* × *tahitensis*, *V. planifolia* type 3 by *V.* × *tahitensis* had the highest FST (0.482) compared to *V. planifolia* type 1 by *V.* × *tahitensis* (0.478) or *V. planifolia* type 2 by *V.* × *tahitensis* (0.413). *V. planifolia* type 2 by *V. planifolia* type 3 had the highest FST value (0.488) among

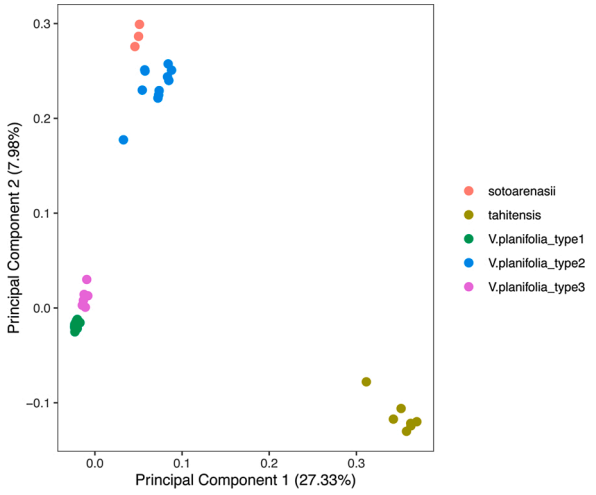

Fig. 6. PCA plot of 190 accessions including only *V. planifolia* and *V.* × *tahitensis*. *V. planifolia* types 1 and 3 are shown in red, *V. planifolia* type 2 is shown in green, and *V.* × *tahitensis* accessions are shown in blue. See online version for color figure.

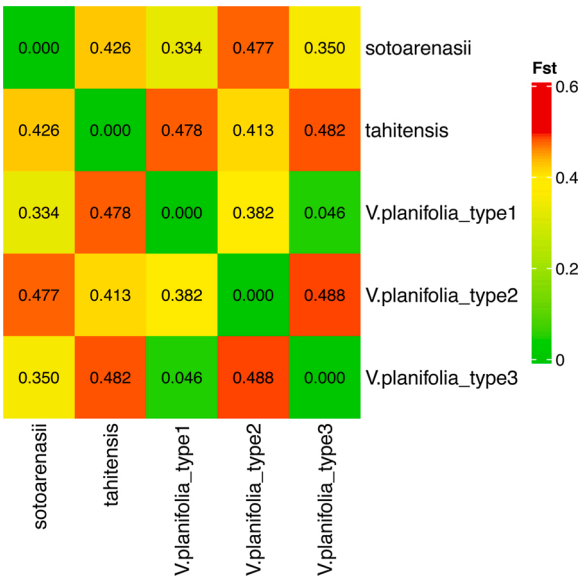

Fig. 7. FST analysis of *V. planifolia*, *V.* × *tahitensis*, and *V. sotoarenasii* accessions. Higher FST values imply greater differentiation among populations. *V. planifolia* type 1, *V. planifolia* type 2, *V. planifolia* type 3, *V.* × *tahitensis* (tahitensis), and *V. sotoarenasii* (sotoarenasii) are shown. See online version for color figure.

V. planifolia types compared to *V. planifolia* type 1 and *V. planifolia* type 2 (0.382) or *V. planifolia* type 1 and *V. planifolia* type 3 (0.046). *V. sotoarenasii* and *V. planifolia* type 2 had the highest FST (0.477) among other *V. planifolia* comparisons with *V. sotoarenasii*. Thus, *V. planifolia* group 2 could be considered to have the lowest degree of differentiation from *V.* × *tahitensis* among the other comparisons in Fig. 7.

3.10. Parentage of *V.* × *tahitensis*

The maternal parent of the *V.* × *tahitensis* hybrid is *V. planifolia* as evidenced by multiple studies showing *rbcL* sequences that are identical between the two types. Both types are also closely located in the phylogenetic trees leveraging genomics-based diversity analysis. It is generally hypothesized that the paternal parent of *V.* × *tahitensis* was *V. odorata*, but supporting information is generally lacking especially for vouchered, verified, and publicly available accessions. Additional supporting evidence would strengthen the hypothesis that *V.* × *tahitensis* is a *V. planifolia* x *V. odorata* hybrid. The original 431,204 SNPs were filtered by quality resulting in 143,122 SNPs that were then screened for 1) homozygosity (0/0) in a subset of *V. planifolia* accessions and 2) heterozygosity (0/1) in *V.* × *tahitensis*. This would identify minor alleles that were contributed from the paternal parent. The resulting filtered set of 1,544 SNPs were used to screen potential minor allele donors among the other sequenced species. As expected, *V.* × *tahitensis* accessions had the highest number of SNPs matching the above criteria with 1,485 to 1,535 of the 1,544 minor alleles. The highest number of SNP matches in accessions other than *V.* × *tahitensis* were all *V. odorata* AC212_2 (1,507 minor allele count, Costa Rica), 341_2 (1,402 minor allele count, Mexico), AC458_2 (1,371 minor allele count, Belize), AC213 (1,360 minor allele count, Belize), AC207 (1,355 minor allele count, Belize), and AC177 (1,106 minor allele count, unknown provenance). The next highest match was *V. helleri* (AC434_2) with a much lower minor allele count of 464. In terms of morphology, AC341_2 and AC350_2 flowers are similar but not identical to *V. odorata* accessions from Central and South America (Fig. 8). The area where these specimens were collected is relatively close to previously collected *V. odorata* accessions used in a report investigating the hybrid origin of *V.* × *tahitensis* [25].

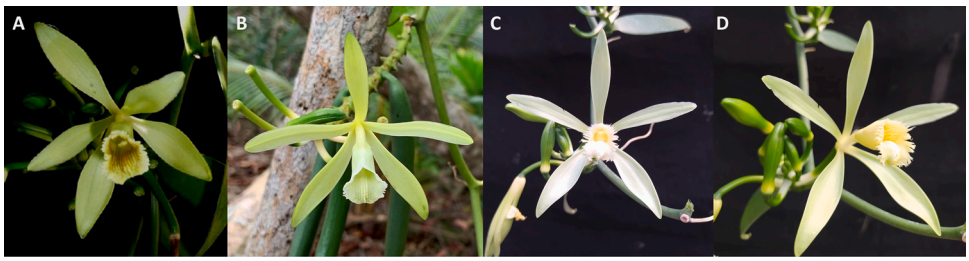

Fig. 8. Representative images from selected accessions. A) flower from a reverse *V.* × *tahitensis* accession (AC352_2) from Mexico, B) *V. sotoarenasii* flower at the type locality in Costa Rica (AC209_2), C) flower of *V. odorata* AC341_2, and D) flower of *V. odorata* AC350_2.

3.11. *V.* × *tahitensis maternal parent*

V. planifolia is the maternal species of the *V.* × *tahitensis* hybrid, but the improved resolution of genomics-based diversity analysis could identify a *V. planifolia* accession that is most similar to the *V. planifolia* genotype in *V.* × *tahitensis*. The hypothesis was that sampling among so many accessions from the center of vanilla cultivation could identify genotypes among *V. planifolia* accessions that are more similar to *V.* × *tahitensis* than others. Therefore, assuming *V. odorata* as the paternal parent, 431,204 SNPs were filtered for high major allele homozygosity (0/0) in *V. odorata* accessions but that were also heterozygous (0/1) in *V.* × *tahitensis*. The resulting 468 SNPs were then used to identify accessions with the highest minor allele count that could have been inherited by *V.* × *tahitensis* from the maternal parent. All six *V.* × *tahitensis* accessions were among the best candidates with 455–465 out of the 468 minor alleles as expected. There were seven *V. planifolia* accessions that most closely matched the *V.* × *tahitensis* minor alleles including AC293_2 (457 minor allele count, Belize), AC295_2 (456 minor allele count, Belize), AC517_2 (455 minor allele count, Belize), AC206 (447 minor allele count, Belize), AC504_2 (432 minor allele count, Nicaragua), AC205 (414 minor allele count, Belize), and AC289_2 (409 minor allele count, Belize). The next closest accession was AC313_2, a *V. planifolia* from the Northern Sierra region of Oaxaca with a 356 minor allele count. All of the accessions most closely matching *V.* × *tahitensis* were categorized as *V. planifolia* type 2 and primarily from Belize as measured both by global SNP analysis and by the reduced set of targeted SNPs.

4. Discussion and conclusions

Genomics-based diversity analyses are important for the conservation and genetic improvement of historically under-researched genera like *Vanilla*. Previous studies relying on few specimens to analyze the variation of morphological traits and limited molecular markers have been useful, but broader morphological assessment and genetic sampling with increased resolution and reduced cost per datapoint are leading to new discoveries. The potential to rapidly characterize wild populations with thousands of SNPs is transforming conservation approaches and is helping to resolve previously ambiguous species assignments. For example, *V. sotoarenasii* was entirely found within the *V.* × *tahitensis* and *V. planifolia* clades indicating a close relationship to these accessions. Further, accession AC209_2 from Costa Rica comes from the type locality (*locus typicus*) or type population for *V. sotoarenasii*. This supports the morphological overlap observed when studying multiple specimens belonging to these three taxa indicating that *V. sotoarenasii* as a heterotypic, nomenclatural synonym of *V. planifolia* [12].

Vanilla diversity studies tend to focus on accessions from Mexico and massively clonally propagated plants in commercial production [16,29,30]. These studies point to a few, foundational vanilla clones from Papantla, Veracruz, Mexico that have become the predominant vanilla type in commercial production. While Mexico is the likely orgin of today's commercial vanilla plants, our sampling of vanilla accessions beyond Mexico emphasizes the need to work across borders to capture a more complete representation of the depth of vanilla diversity. This is supported by modeling work that indicates the potential distribution of *V. planifolia* from North, Central, and South America [56]. Additionally, the potential of genomics-based diversity analyses to support plant breeding has been reviewed by many groups and for many crop species [57–59] including the use of wild relatives in crop breeding [60]. This underscores the potential synergy from expanding vanilla sampling, conservation, and genomics to support vanilla plant improvement.

The global results from this study closely match the previous vanilla GBS study [11]. This was not surprising given the similar sequencing approaches that both centered on mapping sequencing reads to a version of the *V. planifolia* 'Daphna' genome. One major benefit of the present study was the increased number of *V. planifolia* accessions from multiple international collections. Additionally, the inclusion of a number of *V.* × *tahitensis* accessions in this study finally unlocked this genotype for analysis. The previous genomics-based diversity study on vanilla had two accessions that were inaccurately assigned as *V.* × *tahitensis*, but are now considered to be *V. planifolia* that is closely related to *V.* × *tahitensis*. Additionally, the relative presentation of the more highly represented species was different between the two studies, but the overall species assignments were consistent. There were a number of accessions in both studies that will require follow up research to identify if they are distinct species or hybrids.

The major strength of this study includes being the first genomics-enabled characterization of multiple international collections around the center of diversity for *V. planifolia* specifically. Multiple branches were formed among *V. planifolia* accessions, and these will be useful for selecting representative types for future phenotyping, in-depth characterization, and plant breeding. The split among *V. planifolia* into types based on phylogenetic, PCA, and STRUCTURE analyses was reported here for the first time and was only able to be resolved due to the abundance of SNPs used for these analyses. Phenotypic characterization of representatives from these types in the future could guide parental selection for breeding work. The results from this study also demonstrate the need for conservation of vanilla species throughout Central and South America especially for the commercial species as truly wild vanilla in Mexico is exceptionally rare. For example, there were *V. planifolia* accessions from the remote, native forests of Colombia with interesting genetic relationships to other *V. planifolia* and *V.* × *tahitensis* accessions, yet are from an under-researched geography. Additional research could identify novel traits from these accessions, or help further our understanding of the native or migratory distribution of *V. planifolia*.

The origin of *V.* × *tahitensis* has intrigued the vanilla industry for over 100 years due to its distinct flavor profile. This hybrid origin of *V.* × *tahitensis* is confirmed and expanded in the present study. A number of *V. planifolia* type 2 accessions were identified that share a more similar genotype with *V.* × *tahitensis* than the most abundant *V. planifolia* type 1 accessions in general. The *V. planifolia* and *V. odorata* accessions with the genotypes most closely matching *V.* × *tahitensis* point towards an origin in what is now Belize. Additional research in the future could help confirm the Belize origin hypothesis, but will require additional international support especially within Central America and in French

Polynesia where *V.* × *tahitensis* is commonly grown. Future work should also include re-creating *V.* × *tahitensis* not by chance in a forest ecosystem but by selecting superior *V. planifolia* and *V. odorata* parents and creating new hybrid plants. This will enable the analysis of segregating populations of sufficient size to further investigate genetic contributions to the unique aroma profile and non-dehiscing pods of *V.* × *tahitensis*. The distinct morphology of *V. sotoarenasii* and its close genetic relationship with *V.* × *tahitensis* could warrant further investigation as a variant of *V. planifolia*.

Analyzing a large number of samples from multiple countries has provided a unique view of vanilla across borders as shown in this study and in others [4,29,30]. In Colombia, for example, it is common to find individuals with intermediate morphological characters between sympatric species, which suggests high natural hybridization [14]. Colombian researchers have thought that *V. odorata* from the tropical moist forest (bio-geographic Chocó) is not the same as the *V. odorata* from the dry forests. The results of the present study suggest that vanilla growing in these moist forests also includes *V.* × *tahitensis*. This was the case for AC524_2 that was collected in Colombian moist forest and originally designated as *V. odorata* because of its morphology. The *rbcL* results indicating that AC524_2 is a reverse (*V. odorata* x *V. planifolia*) *V.* × *tahitensis* hybrid is interesting. This could further support the hypothesis that *V.* × *tahitensis* is the result of a natural, although rare, hybridization event between species with overlapping native ranges. This is not the first time this has been suggested. AC352_2 was originally collected in January 2009 in Mexico south of Jalisco. Miguel Soto-Arenas, a well-known and respected vanilla researcher, indicated that it was almost certainly *V.* × *tahitensis* 'Haapape' (Fig. 8) based on pictures, and the accession was included in a published, intermediate report [55]. AC352_2 was also previously found to have a *V. odorata rbcL* sequence (data not published), and was sourced from the same area as AC341_2 (*V. odorata*) and AC350_2 (*V. odorata*, low depth). Additional analysis of vanilla specimens collected in Jalisco, Mexico and the moist forests of Chocó, Colombia could support or refine the current migratory theory that *V.* × *tahitensis* was introduced from Mexico to the Philippines and then Tahiti [7], and could have been moved from Tahiti or from Mexico to the wet areas of the Colombian Pacific naturally or through commercial shipping routes.

Resolving the classification of vanilla species and hybrids is benefiting from genomics. For example, accession AC396_2 was compatible with *V. dressleri* and turned out to be a hybrid with information mainly from *V. planifolia*. In nature, there is difficulty in distinguishing among hybrids, and it is possible to find individuals that are not necessarily a first-generation hybrid. These unknown accessions can possess intermediate traits that make precise phenotypic determination difficult. An encouraging result from this study includes AC401_2, which has some morphological characters similar to *V. dressleri* (eg. hard and rigid stems), as having a unique genotype compared to other accessions of *V. dressleri* included in this study. This finding is promising if its organoleptic qualities are taken into account, since it shares some of its aroma profile with *V. planifolia* [61].

In conclusion, the major challenge for vanilla research in general, and for *V.* × *tahitensis* specifically, includes the lack of germplasm availability and thoroughly characterized germplasm. This creates a dearth of consensus among research groups, prevents follow up studies to confirm research results, and reduces overall synergy. Vanilla is a spice with global demand and could benefit from increased research across borders as demonstrated in this study. Even though vanilla is a commodity traded spice, the international movement of vanilla material is currently restricted by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) that has historically created roadblocks for advancing research and genetic improvement of this species. Protecting endangered species while facilitating the safe movement of vanilla germplasm are perfectly compatible with the long-term strategy of vanilla conservation, research, and improvement through plant breeding and could be facilitated through a global, collaborative network of vanilla researchers.

Author contributions statement

AHC: Conceptualization, Methodology, Data curation, Writing – original draft, Supervision, Project administration, Funding acquisition. YAA: Formal analysis, Visualization. MB: Investigation, Methodology. All authors: Investigation, Resources, Writing – Review & Editing.

Funding

Funding for this research was provided to AHC by a University of Florida SEEDIT grant. The sponsor did not have a role in study design, data collection, interpretation, or writing of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks are expressed to Dr. Robert T. González for his valuable comments and contributions on the Colombian species and for his help in the collection of specimens. We also thank Dr. Tupac Otero of the Universidad Nacional de Colombia for his collaboration in collecting the samples and Dr. Pablo R. Stevenson of the Universidad de Los Andes for the academic support. The material from Colombia was legally exported through CITES permit 44706 from December 2019, given by Ministerio de Ambiente y Desarrollo Sostenible.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.plantsci.2021.111019.

References

[1] H.C. Andrews, The Botanist's Repository for New and Rare Plants, London, 1797.
[2] H. Bruman, The culture history of Mexican vanilla, Hisp. Am. Hist. Rev. 28 (1948) 360–376.
[3] M.A. Soto-Arenas, Filogeografía y recursos genéticos de las vainillas de México, Instituto Chinoin AC, 1999. Informe final SNIB-Conabio Proyecto J. 101.
[4] S. Bory, P. Lubinsky, A.M. Risterucci, J.L. Noyer, M. Grisoni, M.F. Duval, P. Besse, Patterns of introduction and diversification of Vanilla planifolia (Orchidaceae) in Reunion Island (Indian Ocean), Am. J. Bot. 95 (2008) 805–815.
[5] P. Lubinsky, M. Van Dam, A. Van Dam, Pollination of Vanilla and evolution in Orchidaceae, Lindleyana 75 (2006) 926–929.
[6] E. Householder, J. Janovec, A.B. Mozambite, J.H. Maceda, J. Wells, R. Valega, H. Maruenda, E. Christenson, Diversity, natural history, and conservation of Vanilla (Orchidaceae) in Amazonian wetlands of Madre De Dios, Peru, J. Bot. Res. Inst. Tex. (2010) 227–243.
[7] M.A. Soto-Arenas, R.L. Dressler, 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 (2010) 285–354.
[8] C. Morren, Sur la fructification de la vanille obtenue au moyen de la fécondation artificielle, Comptes Rendues Acad. Sci. 6 (1838) 489–492.
[9] T. Ecott, Vanilla: Travels in Search of the Luscious Substance, Penguin UK, 2005.
[10] T. Hasing, H. Tang, M. Brym, F. Khazi, T. Huang, A.H. Chambers, A phased Vanilla planifolia genome enables genetic improvement of flavour and production, Nat. Food 1 (2020) 811–819.
[11] Y. Hu, M. Resende, A. Bombarely, M. Brym, E. Bassil, A. Chambers, Genomics-based diversity analysis of Vanilla species using a Vanilla planifolia draft genome and Genotyping-By-Sequencing, Sci. Rep. 9 (2019) 3416.
[12] A.P. Karremans, I.F. Chinchilla, G. Rojas-Alvarado, M. Cedeño-Fonseca, A. Damián, G. Léotard, A reappraisal of Neotropical Vanilla. With a note on taxonomic inflation and the importance of alpha taxonomy in biological studies, Lankesteriana 20 (3) (2020) 395–497.
[13] K.M. Cameron, Vanilla phylogeny and classification, in: D. Havkin-Frenkel, F. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2011, pp. 241–255.
[14] N.S. Flanagan, P. Chavarriaga, A.T. Mosquera-Espinosa, Conservation and sustainable use of Vanilla crop wild relatives in Colombia, in: D. Havkin-Frenkel,

F. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2018, pp. 85–110.
[15] N. Childers, H. Cibes, in: C. Withner (Ed.), The Orchids: A Scientific Study, The Ronald Press Company, New York, 1959, pp. 477–508.
[16] S. Bory, M. Grisoni, M.-F. Duval, P. Besse, Biodiversity and preservation of vanilla: present state of knowledge, Genet. Resour. Crop Evol. 55 (2008) 551–571.
[17] D.S. Correll, Vanilla-Its botany, history, cultivation and economic import, Econ. Bot. 7 (1953) 291–358.
[18] S. Koyyappurath, T. Atuahiva, R. Le Guen, H. Batina, S. Le Squin, N. Gautheron, V. E. Hermann, J. Peribe, M. Jahiel, C. Steinberg, E.C.Y. Liew, C. Alabouvette, P. Besse, M. Dron, I. Sache, V. Laval, M. Grisoni, Fusarium oxysporum f. sp radicis-vanillae is the causal agent of root and stem rot of vanilla, Plant Pathol. 65 (2016) 612–625.
[19] J. Hernandez-Hernández, Mexican vanilla production, in: D. Havkin-Frenkel, F. C. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2018, pp. 3–25.
[20] United States Food and Drug Administration, Code of Federal Regulations Title 21 Volume 2 Part 169.3, US Food and Drug Administration, 2021.
[21] R. Menchaca, J. Ramos, D. Moreno, M. Luna, M. Mata, L. Vázquez, M. Lozano, In vitro germination of Vanilla planifolia and V. pompona hybrids, Revista Colombiana de Biotecnología 13 (2011) 80–84.
[22] D. Ehlers, M. Pfister, Compounds of vanillons (Vanilla pompona Schiede), J. Essent. Oil Res. 9 (1997) 427–431.
[23] FAOSTAT, Food and Agriculture Organization of the United Nations, in, 2021.
[24] FAO, Vanilla: Post-harvest Operations, 2009.
[25] P. Lubinsky, K.M. Cameron, M.C. Molina, M. Wong, S. Lepers-Andrzejewski, A. Gómez-Pompa, S.C. Kim, Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, Vanilla tahitensis (Orchidaceae), Am. J. Bot. 95 (2008) 1040–1047.
[26] L.R. Nielsen, H.R. Siegismund, Interspecific differentiation and hybridization in Vanilla species (Orchidaceae), Heredity 83 (1999) 560–567.
[27] P. Besse, D. Da Silva, S. Bory, M. Grisoni, F. Le Bellec, M.F. Duval, RAPD genetic diversity in cultivated vanilla: Vanilla planifolia, and relationships with V. x tahitensis and V. pompona, Plant Sci. 167 (2004) 379–385.
[28] D. Minoo, V.N. Jayakumar, S.S. Veena, J. Vimala, A. Basha, K.V. Saji, K.N. Babu, K. V. Peter, Genetic variations and interrelationships in Vanilla planifolia and few related species as expressed by RAPD polymorphism, Genet. Resour. Crop Evol. 55 (2008) 459–470.
[29] P.M. Schlüter, M.A.S. Arenas, S.A. Harris, Genetic variation in Vanilla planifolia (Orchidaceae), Econ. Bot. 61 (2007) 328.
[30] P. Lubinsky, S. Bory, J. Hernandez, S.C. Kim, A. Gomez-Pompa, Origins and dispersal of cultivated vanilla (Vanilla planifolia jacks. [Orchidaceae]), Econ. Bot. 62 (2008) 127–138.
[31] S. Lepers-Andrzejewski, S. Causse, B. Caromel, M. Wong, M. Dron, Genetic linkage map and diversity analysis of Tahitian vanilla (Vanilla× tahitensis, Orchidaceae), Crop Sci. 52 (2012) 795–806.
[32] A.L. Ramos-Castellá, L.G. Iglesias-Andreu, J. Martínez-Castillo, M. Ortíz-García, R. H. Andueza-Noh, P. Octavio-Aguilar, M. Luna-Rodríguez, Evaluation of molecular variability in germplasm of vanilla (Vanilla planifolia G. Jackson in Andrews) in Southeast Mexico: implications for genetic improvement and conservation, Plant Genet. Resour. 15 (2017) 310–320.
[33] S. Bory, D. Da Silva, A.-M. Risterucci, M. Grisoni, P. Besse, M.-F. Duval, Development of microsatellite markers in cultivated vanilla: polymorphism and transferability to other vanilla species, Sci. Hortic. 115 (2008) 420–425.
[34] V.B. Perez, L.G.I. Andreu, E.A.E. Manzano, J.M. Castillo, M.M.O. Garcia, P. O. Aguilar, Molecular and microclimatic characterization of two plantations of Vanilla planifolia (Jacks ex Andrews) with divergent backgrounds of premature fruit abortion, Sci. Hortic. 212 (2016) 240–250.
[35] R.L. Gigant, N. Rakotomanga, C. Goulié, D. Da Silva, N. Barre, G. Citadelle, D. Silvestre, M. Grisoni, P. Besse, Microsatellite markers confirm self-pollination and autogamy in wild populations of Vanilla mexicana Mill. (syn. V. inodora) (Orchidaceae) in the island of Guadeloupe. Microsatellite Markers, InTech, 2016, pp. 73–94.
[36] M.A. Soto Arenas, R.L. Dressler, 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 Int. J. Orchid 9 (2010) 285–354.
[37] K.M. Cameron, Utility of plastid psaB gene sequences for investigating intrafamilial relationships within Orchidaceae, Mol. Phylogenet. Evol. 31 (2004) 1157–1180.
[38] A. Bouetard, P. Lefeuvre, R. Gigant, S. Bory, M. Pignal, P. Besse, M. Grisoni, Evidence of transoceanic dispersion of the genus Vanilla based on plastid DNA phylogenetic analysis, Mol. Phylogenet. Evol. 55 (2010) 621–630.
[39] K.M. Cameron, M. Carmen Molina, Photosystem II gene sequences of psbB and psbC clarify the phylogenetic position of Vanilla (Vanilloideae, Orchidaceae), Cladistics 22 (2006) 239–248.
[40] K.M. Cameron, M.W. Chase, W.M. Whitten, P.J. Kores, D.C. Jarrell, V.A. Albert, T. Yukawa, H.G. Hills, D.H. Goldman, A phylogenetic analysis of the Orchidaceae: evidence from rbcL nucleotide sequences, Am. J. Bot. 86 (1999) 208–224.
[41] M. Divakaran, K.N. Babu, P.N. Ravindran, K.V. Peter, Interspecific hybridization in vanilla and molecular characterization of hybrids and selfed progenies using RAPD and AFLP markers, Sci. Hortic. 108 (2006) 414–422.
[42] P.C. Verma, D. Chakrabarty, S.N. Jena, D.K. Mishra, P.K. Singh, S.V. Sawant, R. Tuli, The extent of genetic diversity among Vanilla species: comparative results for RAPD and ISSR, Ind. Crop Prod. 29 (2009) 581–589.
[43] R. Sreedhar, L. Venkatachalam, K. Roohie, N. Bhagyalakshmi, Molecular analyses of Vanilla planifolia cultivated in India using RAPD and ISSR markers, Orchid Sci. Biotechnol. 1 (2007) 29–33.
[44] K.M. Cameron, On the value of nuclear and mitochondrial gene sequences for reconstructing the phylogeny of vanilloid orchids (Vanilloideae, Orchidaceae), Ann. Bot. 104 (2009) 377–385.
[45] S. Villanueva-Viramontes, M. Hernández-Apolinar, G.C. Fernández-Concha, A. Dorantes-Euán, G.R. Dzib, J.M. Castillo, Wild Vanilla planifolia and its relatives in the Mexican Yucatan Peninsula: systematic analyses with ISSR and ITS, Bot. Sci. 95 (2017) 169–187.
[46] H.E.L. Lischer, L. Excoffier, PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs, Bioinform 28 (2012) 298–299.
[47] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[48] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[49] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[50] D.A. Earl, B.M. Vonholdt, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conserv. Genet. Resour. 4 (2012) 359–361.
[51] X.W. Zheng, D. Levine, J. Shen, S.M. Gogarten, C. Laurie, B.S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data, Bioinformatics 28 (2012) 3326–3328.
[52] T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new method for the analysis of genetically structured populations, BMC Genet. 11 (2010).
[53] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population-structure, Evolution 38 (1984) 1358–1370.
[54] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[55] M.A. Soto Arenas, Recopilación y análisis de la información existente sobre las especies mexicanas del género Vanilla, Reporte, Herbario de la Asociación Mexicana de Orquideología, AC, Instituto Chinoín, AC, México, DF, 2009.
[56] P. Ellestad, F. Forest, M. Serpe, S.J. Novak, S. Buerki, Harnessing large-scale biodiversity data to infer the current distribution of Vanilla planifolia (Orchidaceae), Bot. J. Linn. Soc. 196 (2021) 407–422.
[57] J.F. He, X.Q. Zhao, A. Laroche, Z.X. Lu, H.K. Liu, Z.Q. Li, Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, Front. Plant Sci. 5 (2014) 484.
[58] S. Getachew, N. Bille, J. Bell, W. Gebreselassie, Genotyping by sequencing for plant breeding-a review, Adv. Biotechnol. Microbiol. 14 (2020), 555891.
[59] C. Kim, H. Guo, W.Q. Kong, R. Chandnani, L.S. Shuang, A.H. Paterson, Application of genotyping by sequencing technology to a variety of crop breeding programs, Plant Sci. 242 (2016) 14–22.
[60] H. Dempewolf, G. Baute, J. Anderson, B. Kilian, C. Smith, L. Guarino, Past and future use of wild relatives in crop breeding, Crop Sci. 57 (2017) 1070–1082.
[61] R.T. González Mina, Plasticidad morfofisiológica comparativa de especies de vanilla silvestres del pacífico Colombiano, Doctorado en Ciencias Agrarias, 2016.
[62] J.A. Cibrián, *Variación genética de Vanilla planifolia en México*, Undergraduate Thesis. Facultad de Ciencias, Universidad Nacional Autónoma de México, 1999.

Plant Science 311 (2021) 111019

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Research Article

Genotyping-By-Sequencing diversity analysis of international *Vanilla* collections uncovers hidden diversity and enables plant improvement

Alan Chambers [b,*], Angélica Cibrián-Jaramillo [c], Adam P. Karremans [g,h], David Moreno Martinez [i], Juan Hernandez-Hernandez [e], Maria Brym [b], Marcio F.R. Resende Jr. [f], Ruth Moloney [j], Sheryl N. Sierra [k], Tomas Hasing [d], Yasmin A. Alomia [a], Ying Hu [f], Vanilla Genotyping Consortium

[a] *Department of Biological Sciences, Faculty of Sciences, Universidad de Los Andes, Cra. 1 N° 18A - 12, Bogotá, Colombia*
[b] *Tropical Research and Education Center, Horticultural Sciences Department, University of Florida, 18905 SW 280th St, Homestead, FL, 33033, USA*
[c] *Ecological and Evolutionary Genomics Laboratory, Unidad de Genómica Avanzada (Langebio), CINVESTAV, Km 9.6 Carretera Irapuato-León, Guanajuato, CP 36824, Mexico*
[d] *Elo Life Systems, Durham, NC, 27709, USA*
[e] *Campo Experimental Ixtacuaco, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Km 4.5 Carretera Martínez de la Torre-Tlapacoyan, Veracruz, CP 93600, Mexico*
[f] *Horticultural Sciences Department, University of Florida, Gainesville, FL, USA*
[g] *Lankester Botanical Garden, University of Costa Rica, P.O. Box 302-7050, Cartago, Costa Rica*
[h] *Naturalis Biodiversity Center, Endless Forms Group, Sylviusweg 72, Leiden, 2333 BE, the Netherlands*
[i] *Posgrado en Ecología Tropical, Centro de Investigaciones Tropicales, Universidad Veracruzana, José María Morelos 44, Zona Centro, CP 91000, Xalapa, Veracruz, Mexico*
[j] *Corridgeree Belize Ltd, Mile 6, Southern Highway, Silk Grass, Stann Creek District, Belize*
[k] *College of Agriculture, Food, Environment and Natural Resources, Cavite State University, Indang, Cavite, 4122, Philippines*

ARTICLE INFO

This work is dedicated to Dr. Caroline Herron.

Keywords:
Vanilla planifolia
V. × *tahitensis*
Genotyping-by-sequencing
Diversity
Single nucleotide polymorphisms

ABSTRACT

Genomics-based diversity analysis of natural vanilla populations is important in order to guide conservation efforts and genetic improvement through plant breeding. Vanilla is a cultivated, undomesticated spice that originated in Mesoamerica prior to spreading globally through vegetative cuttings. Vanilla extract from the commercial species, mainly *V. planifolia* and *V.* × *tahitensis*, is used around the world as an ingredient in foods, beverages, cosmetics, and pharmaceuticals. The global reliance on descendants of a few foundational clones in commercial production has resulted in an industry at heightened risk of catastrophic failure due to extremely narrow genetic diversity. Conversely, national and institutional collections including those near the center of cultivation contain previously undiscovered diversity that could bolster the genetic improvement of vanilla and guide conservation efforts. Towards this goal, an international vanilla genotyping effort generated and analyzed 431,204 single nucleotide polymorphisms among 412 accessions and 27 species from eight collections. Phylogenetic and STRUCTURE analysis sorted vanilla by species and identified hybrid accessions. Principal Component Analysis and the Fixation Index (FST) were used to refine relationships among accessions and showed differentiation among species. Analysis of the commercial species split *V. planifolia* into three types with all *V.* × *tahitensis* accessions being most similar to *V. planifolia* type 2. Finally, an in-depth analysis of *V.* × *tahitensis* identified seven *V. planifolia* and six *V. odorata* accessions as most similar to the estimated parental genotypes providing additional data in support of the current hybrid theory. The prevalence of probable *V.* × *tahitensis* parental accessions from Belize suggests that *V.* × *tahitensis* could have originated from this area and highlights the need for vanilla conservation throughout Central and South America. The genetic groupings among

* Corresponding author at: Tropical Research and Education Center, Horticultural Sciences Department, University of Florida, 18905 SW 280th St, Homestead, FL, 33033, USA.

E-mail addresses: ac@ufl.edu (A. Chambers), angelica.cibrian@cinvestav.mx (A. Cibrián-Jaramillo), adam.karremans@ucr.ac.cr (A.P. Karremans), dmoreno29@hotmail.com (D. Moreno Martinez), hernandezh.juan@inifap.gob.mx (J. Hernandez-Hernandez), maria.brym@ufl.edu (M. Brym), mresende@ufl.edu (M.F.R. Resende), ruth.moloney@6ptp.com (R. Moloney), snsierra@gmail.com (S.N. Sierra), tengfang.huang@elolife.ag (T. Hasing), ya.alomia10@uniandes.edu.co (Y.A. Alomia), ying.hu@ufl.edu (Y. Hu).

https://doi.org/10.1016/j.plantsci.2021.111019
Received 2 June 2021; Received in revised form 4 August 2021; Accepted 7 August 2021
Available online 13 August 2021

accessions, particularly for *V. planifolia*, can now be used to focus breeding efforts on fewer accessions that capture the greatest diversity.

1. Introduction

Vanilla planifolia is an undomesticated, wild species that was originally cultivated in Mesoamerica. *V. planifolia* Andrews [1], the primary commercial species, was spread globally as part of the spice trade from the new world [2,3]. Vanilla was successfully propagated clonally by cuttings in European gardens in the 1800s and eventually spread to today's major commercial production areas including Madagascar [4]. In nature, the flowers of *V. planifolia* and its close relatives are dependent on the Neotropical, endemic Euglossini bees to become fertilized and form seed capsules commonly called beans [5–7]. Producing vanilla beans outside its native range was elusive until a method of artificial pollination was discovered in 1838 by Charles Morren in France with modifications in 1841 by Edmond Albius in Reunion, an island nation off the coast of Madagascar [8,9]. Vanilla thrived in new growing locations and artificial pollination enabled production of this high-value spice around the world. Today's commercial vanilla cultivation still heavily relies on vegetatively propagated descendants of those foundational clones, and, as a result, vanilla has often been reported as having limited genetic diversity [7]. The recent publication of a chromosome-scale *V. planifolia* genome and demonstration of genomics-based diversity analysis for vanilla have unlocked the potential to characterize vanilla diversity with unprecedented resolution [10,11]. Combining these tools to analyze vanilla diversity in Central and South America are key to understanding this commercially important genus and to guide conservation and plant breeding efforts.

The genus *Vanilla* comprises about 118 accepted species in the Orchidaceae family [12]. *Vanilla* species have a pantropic distribution, occurring naturally in the New World tropics, from southern Florida and Mexico, Central America and the Antilles, to South America; in tropical Africa, both on the mainland and on Madagascar; and in tropical Asia, from China and India to New Guinea and Southeast Asia [13]. Just over half of all vanilla species are endemic to tropical America, and only a subset of those new world species bear the highly esteemed aromatic fruits that contain vanillin. The aromatic vanilla species, which constitute the secondary gene pool of the vanilla crop [14], can be found growing in the wild from Florida to Argentina. The commercially important *V. planifolia* itself is only known to occur naturally in Mexico, Central America, and perhaps certain areas in northern South America [7,12].

The global spread of *V. planifolia* followed the early spice trade routes. The species arrived in Europe as early as 1510 with successful cultivation as early as 1807 [15]. Vanilla spread from Europe and into Africa, India, Reunion Island, and Madagascar throughout the 1800s [16]. A trans-Pacific migration through Hawaii and the Philippines included dispersion of *V. planifolia* and *V.* × *tahitensis* [17], though accurately and consistently identifying vanilla species has been a major challenge. Therefore, there has been increasing interest in accurately characterizing vanilla accessions within and among species.

The narrow genetic base of *V. planifolia* in commercial cultivation is a major risk to global supply chains as clones can all be similarly susceptible to prevalent pathogens including *Fusarium* and viruses [13,18]. Additionally, as of 2011, up to 50 % fruit drop has been recorded in Mexico possibly due to average increases in temperature, reduction in humidity, and/or the incidence of pathogens that exacerbate yield losses [19]. As a result, there is a great need to identify diversity with the commercial species in order to detect novel genetics and reduce production risks. These efforts can be achieved through the strategic cultivation of diverse, existing accessions and through plant breeding to create novel genetic combinations with improved plant performance. This approach is more efficient when guided by morphological, phenotypic, or genomics-based information.

Vanilla extract is the world's second most valuable spice and is made from cured vanilla beans. Green vanilla beans have little aroma and must be cured in order to develop their characteristic flavor profile. Curing methods include thermal treatments to disrupt cellular organization and support enzymatic activity with a final drying treatment to stabilize beans for shipping prior to extraction in aqueous ethanol. Today, the United States and European Union are the two largest importers of cured vanilla beans. *V. planifolia* and *V.* × *tahitensis* are the only permissible species for use in making vanilla extract [20]. *V. pompona* is an additional, edible vanilla species consumed locally in Central and South America, but was omitted from the original Standard of Identity. Hybrids of *V. planifolia* with *V. pompona* show good growth, resistance to higher temperatures, and fusarium resistance, but the best *V. pompona* accessions for hybrid populations still need to be elucidated. The xerophytic character of *V. pompona* could also be important in a genetic improvement program due to climatic changes in traditional growing areas where higher temperatures and lower rainfall are becoming more prevalent [21]. Additionally, the presence of vanillin, protocatechuic acid, p-hydroxybenzoic acid, vanillic acid, phydroxybenzaldehyde, p-anisyl in *V. pompona* give it a high potential for the perfumery industry as an alternative to the food industry [22].

The largest producers of *V. planifolia* beans include Madagascar, Indonesia, Mexico, Papua New Guinea, China, Turkey, and Tonga [23,24]. *V.* × *tahitensis* is primarily grown in Tahiti and Papua New Guinea, has a distinct flavor profile compared to *V. planifolia*, and is generally accepted as a hybrid between *V. planifolia* and *V. odorata*. This was supported by a report including 21 additive SNPs among *V. planifolia*, *V. odorata*, and *V.* × *tahitensis* accessions by sequencing the Internal Transcribed Spacer (ITS) region [25]. Unfortunately, incomplete data reporting and lack of access to verified, living *V.* × *tahitensis* material has inhibited in-depth analyses with this hybrid leaving many outstanding questions for this commercially-important genotype.

Various molecular tools have been used to identify species, characterize diversity, identify hybrids, and resolve genetic relationships among vanilla accessions. Methods have included isozymes [26,62], RAPDs [27–29], AFLPs [4,30,31], microsatellites [32–35], plastid-derived single gene sequences [36–40], and others [25,41–45]. While useful, many of the molecular tools in these studies had limitations including utility of the data generated, confidence in the correct species assignments of accessions tested, or the low number of accessions analyzed. The recent publication of a genomics-based single nucleotide polymorphism diversity study of 112 accessions generated the greatest resolution on vanilla diversity to date, but was limited by the diversity of material available for testing [11]

The purpose of the present study was primarily to leverage new genomics tools to advance our current understanding of vanilla diversity within international collections. This was accomplished by focusing on wild-collected, living collections from multiple countries primarily throughout Mexico, Central America, and South America. This is the first time that an international study of vanilla germplasm has been analyzed at such high resolution. A genomics-based approach overcomes many limitations of confounding morphological descriptors that can be affected by growing environment, advances our capability to assign species for living specimens, and identifies gaps in existing collections. This foundational information is critical for establishing and diversifying vanilla breeding programs and highlights the critical need to collaboratively conserve vanilla diversity within and across countries particularly in Mexico, Central America, and South America.

2. Materials and methods

2.1. Vanilla plant accessions

Accessions in this study came primarily from international collections (Table 1, Supplementary Table S1). Accessions from Belize were provided by Corridgeree Belize Ltd and the Belize National Herbarium. This collection was made both by collecting specimens growing in coastal tropical forests (with wildlife collection permits) and taking leaf samples from the vanilla collection in the Belize National Herbarium. Accessions from Colombia were provided by the Universidad de Los Andes. The material was collected directly in the field from wild populations that cover different ecosystems in Colombia. Identity was verified for most of the samples based on floral morphological features. In some cases, vegetative traits suggested the formation of hybrids. Costa Rican material from Jardin Botanico Lankester and the University of Costa Rica was collected from natural areas. Three collections from Mexico including Langebio Centro de Investigación y Estudios Avanzados, Unidad de Manejo para la conservación de la Vida Silvestre Xochitlcalli (SEMARNAT: UMA-IN-VIV-0203-VER/15), and the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) represent vanilla diversity collected over time with various levels of phenotypic characterization. INIFAP accessions were maintained on forest tutor trees similar to their natural habitat. Accessions from the Philippines were provided by Cavite State University and the collection was founded by purchasing vanilla material throughout the country from commercial vendors with one accession provided by Dr. Calixto M. Protacio at the Institute of Crop Science, College of Agriculture and Food Science, University of the Philippines Los Baños Laguna. These accessions could include cuttings from naturalized populations that are descendants of the early trans-Pacific migration from Central America. The University of Florida vanilla collection in the United States of America includes accessions from botanical gardens, online vendors, and hobbyists. The living collection is maintained on raised beds with 16 cm of mulch in a shade house under 70 % filtered light. Individual trellises are spaced 1.2 m apart and include a central post, a column of wire mesh as a climbing substrate, and horizontal wooden post for vine support at 1.2 m from the ground. Supplemental irrigation is supplied through drip lines during the dry season. Average temperatures in Homestead, Florida vary between 18 °C in the winter months to 27 °C in the summer. Average rainfall is ~4.4 cm per month in the dry season (November to March), and up to ~20.6 cm on average per month during the peak of the rainy season (June to September) (Florida Automated Weather Network, http://fawn.ifas.ufl.edu)[52]. Additional information on this collection has been published previously [11].

Table 1
Vanilla accessions from multiple international collections that were included in this study and genotyped for the first time. Country of origin, supporting institution, number of accessions provided, and the number of included species from each are shown. There were 291 accessions genotyped for the first time as part of this study with overlapping representatives from at least 26 species. Note, most collections are larger than the number of accessions and species that were selected for this study.

Country	Institution	Accessions	Species
Belize	Corridgeree Belize Ltd/Belize National Herbarium	18	1
Colombia	Universidad de Los Andes	14	10
Costa Rica	Jardin Botanico Lankester - University of Costa Rica	14	11
Mexico	Instituto Nacional de Investigaciones Forestales, Agrícolas, y Pecuarias (INIFAP)	12	3
Mexico	Langebio Centro de Investigación y Estudios Avanzados	30	2
Mexico	Unidad de Manejo para la conservación de la Vida Silvestre Xochitlcalli. SEMARNAT: UMA-IN-VIV-0203-VER/15	59	1
Multiple	Marie Selby Botanical Gardens	11	6
Philippines	Cavite State University	27	2
United States	University of Florida	106	12
	Total	291	26

2.2. DNA extraction

DNA extractions were performed using a modified CTAB method at the University of Florida. For fresh tissue, 250 mg of mature leaf was freeze dried for 24 h and ground to a fine powder in a modified paint shaker using two stainless steel beads in a 2 ml screw cap vial. For desiccated samples, tissues were freeze dried for 24 h upon arrival and then ground to a fine powder as for fresh samples. All ground tissue was mixed with 3 ml of DNA extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris, 2 mM EDTA, 1% β-mercaptoethanol, pH 8.0) in a 15 ml conical tube, vortexed to completely disperse the tissue, and then incubated at 65C° for 10 min. Samples were then cooled to room temperature for 5 min and 3 ml chloroform:isoamyl alcohol (24:1 v/v) was added to each sample, vortexed, and centrifuged for 10 min at 18,000×*g*. The supernatant was transferred to a fresh tube to precipitate DNA using two volumes of 95 % ethanol. Samples were centrifuged at 10,000×*g* for 10 min to pellet DNA, aspirated, washed once with one ml of 70 % ethanol, centrifuged at 10,000×*g* for 2 min, aspirated, and finally resuspended in 75 μl molecular biology grade water.

2.3. Sequencing vanilla accessions

Library preparation and sequencing for genotyping-by-sequencing (GBS) was performed as previously described [11]. Briefly, libraries were constructed using dual restriction enzymes BamHI and NsiI and sequencing was performed using NextSeq 1 × 150 bp with 300–744 bp size selection. Target sample depth was 0.5 million reads per sample. All reads were quality filtered by the University of Minnesota Genomics Center as previously reported and screened for contamination using FastQ Screen (default parameters).

2.4. SNP calling

SNP calling leveraged the phased *V. planifolia* 'Daphna' genome sequence [10]. Raw reads were processed with fastq-mcf v1.05 (-q 30 -l 50) and then aligned against the reference assembly of haplotype A with HISAT2 v2.1 (default parameters). The bam files were merged and read groups were added using bamaddrg 3fccbf0. Joint variant calling was performed with FreeBayes v1.3.1-19-g54bf409 (default parameters). Variants were then filtered with VCFtools v0.1.17 (–max-missing 0.3 –maf 0.05 –min-meanDP 10 –max-meanDP 1000). The resulting SNP set was used for phylogentic, STRUCTURE, and FST analyses. A separate SNP set was created that only included *V. planifolia* and *V.* × *tahitensis* accessions. SNPs from these accessions were filtered using the same parameters as for the complete dataset.

2.5. Phylogenetic analysis

The filtered VCF files were converted to Phylip format by concatenating the SNPs with PGDSpider v.2.0.9.0 [46] with IUPAC ambiguity codes for polymorphic data. The maximum likelihood (ML) phylogenetic tree with 1000 rapid bootstrap inference was constructed by using RAxML v.8.2.12 [47]. The analysis was run using an ascertainment bias correction (ASC) for the data set containing only concatenated informative SNP positions, and a general time-reversible substitution model accounting for among-site rate heterogeneity (ASC_GTRGAMMA model). The RAxML results were graphically visualized in FigTree v.1.4.3 (downloaded from http://tree.bio.ed.ac.uk/software/figtree/).

2.6. Population STRUCTURE

The population structure was investigated to identify clusters of genetically related individuals using the Bayesian clustering method implemented in STRUCTURE, v.2.3.4 [48]. Ten independent STRUCTURE runs were performed for each of K = 2-16 (K = number of genetic clusters) with a burn-in of 10000 and 20000 iterations. The number of clusters that best fit the observed genotype data was inferred by examining the average and standard deviation (SD) of the natural log probability of each model [49] using the online version of STRUCTURE HARVESTER [50]. Additionally, the R package SNPRelate [51] was used to perform principal component analysis (PCA) and the R package adegenet [52] was used to perform discriminant analysis of principal components (DAPC). Loci were trimmed to linkage disequilibrium (LD < 0.2) using the command "snpgdsLDpruning". DAPC was performed on all individual genotypes by using the successive K-means approach, implemented by the find.clusters function, to identify the optimal number of groups based on Bayesian Information Criterion (BIC). PCA plots were created using JMP Pro v15.

To investigate the diversity of *V. planifolia* further, the *V. planifolia*-only PCA was performed on 27 accessions while excluding three diverse *V. planifolia* lines (AC108, AC181, and AC185) to improve resolution among the remaining accessions. Pairwise Fst were calculated with the Weir and Cockerham formula [53] with the SNPRelate package [51].

2.7. rbcL sequencing

Partial sequencing of the *rbcL* gene was used to confirm species identifications of selected accessions and to identify the maternal parent of probable hybrids. Primers *rbcL* forward 5' CTTCACAAGCAGCAGCTAGTTC 3' and *rbcL* reverse 5' ATGTCACCACAAACAGAAAC 3' were used to amplify an ~1,300 bp fragment of the *rbcL* gene in 25 μl reactions using Phusion High-Fidelity polymerase (F531, Thermo Scientific) according to the manufacturer's instructions. PCR amplification was conducted using an Applied Biosystems SimpliAmp Thermal Cycler with the following program: initial denaturation 5 min at 95 °C followed by 30 amplification cycles of 30 s at 95 °C, 30 s at 55 °C, and 60 s at 72 °C, with a final 10 min extension at 72 °C. Sanger Sequencing was conducted at GENEWIZ (GENEWIZ, South Plainfield, NJ) using *rbcL* forward as the sequencing primer. Sequences were verified against published vanilla *rbcL* sequences from NCBI (ncbi.nlm.nih.gov) using the default settings for MUSCLE alignment [54] and Geneious Tree Builder within the Geneious Software package (Geneious version 11.1, Newark, NJ). Accessions AC289_2, AC351_2, AC361_2, AC361_2, AC399_2, AC433_2, AC471_2, AC504_2, AC524_2, AC525_2, AC526_2, and AC534_2 were selected for sequencing. AC361_2 was sequenced twice with independent PCR and sequencing reactions to confirm the presence of a single SNP when compared to the typical sequence for *V. planifolia* type 1 accessions.

2.8. Parentage of V. × tahitensis

The 431,204 SNPs from 412 accessions both in this study and the previous vanilla GBS study were used to investigate the paternal parent of *V.* × *tahitensis*. The total SNPs were filtered using vcftools (–max-missing 0.5 –min-meanDP 10 –max-meanDP 1000 –remove-indels –max-alleles 2). The 143,122 resulting SNPs were then screened for homozygosity (0/0) in at least 9 out of 10 accessions in a subset of *V. planifolia* (AC275_2, AC277_2, AC281_2, AC285_2, AC293_2, AC314_2, AC368_2, AC372_2, AC375_2, AC504_2) and heterozygosity in 5 out of 6 *V.* × *tahitensis* accessions (AC421_2, AC432_2, AC433_2, AC509_2, AC526_2, AC237_2). The resulting SNPs were then used to identify other accessions with the alternative allele (0/1 or 1/1) to identify those that might have contributed the minor allele found in *V.* × *tahitensis*.

A similar approach to that described above was taken to identify the accessions most closely related to the maternal *V. planifolia* parent of *V.* × *tahitensis*. The same 143,122 SNPs were screened among *V. odorata* accessions (AC177, AC213, AC207, AC212_2, AC458_2, and AC341_2) for homozygous reference alleles (0/0) and that were heterozygous (0/1) in the *V.* × *tahitensis* accessions. There were 468 SNPs identified after filtering. All accessions were then screened for the presence of the alternative allele (0/1 or 1/1) for the 468 SNPs.

2.9. Data availability

Raw sequencing data is available on NCBI at part of Bioproject PRJNA733522. *rbcL* sequences have also been deposited at NCBI for AC289_2 (MZ391815), AC351_2 (MZ391816), AC361_2a), MZ391817), AC361_2b), MZ391818), AC399_2 (MZ391819), AC433_2 (MZ391820), AC471_2 (MZ391821), AC504_2 (MZ391822), AC524_2 (MZ391823), AC525_2 (MZ391824), AC526_2 (MZ391825), and AC534_2 (MZ391826).

3. Results

3.1. Vanilla accessions included in this study

There were 291 new accessions genotyped as part of this study as summarized in Table 1. Institutions from Belize, Colombia, Costa Rica, Mexico, the Philippines, and the United States submitted material from their domestic collections for genotyping. Library preparation and sequencing of accessions were similar to a previous *Vanilla* GBS study [11]. As a result, the 291 accessions were able to be combined with publicly available *Vanilla* GBS data resulting in a total of 412 accessions from 27 species available for combined analyses (Supplemental Table S1).

Species in this study included *V. atropogon*, *V. barbellata*, *V. bicolor*, *V. claviculata*, *V. dilloniana*, *V. dresslerii*, *V. griffithii*, *V. hartii*, *V. helleri*, *V. imperialis*, *V. insignis*, *V. karen-christianae*, *V. lindmaniana*, *V. madagascariensis*, *V. mexicana*, *V. odorata*, *V. palmarum*, *V. phaeantha*, *V. phalaenopsis*, *V. planifolia*, *V. poitaei*, *V. pompona*, *V. roscheri*, *V. siamensis*, *V. sotoarenasii*, *V.* × *tahitensis*, and *V. trigonocarpa*. Three additional designations including "low depth", "hybrid", and "species" were assigned based on sequencing depth, high likelihood of being a hybrid, and uncertain species assignments, respectively. *V. planifolia* represented ~41 % of the 412 accessions in this study, and hybrid and unknown accessions included 10.4 % and 3.9 % of accessions, respectively. Some species designations could not be fully resolved and will require follow up research for targeted morphological and genetic confirmation.

3.2. Single nucleotide polymorphism analysis

Target read depth was 0.5 million reads per sample for accessions in this study. After sequencing, there were 185,168,710 total reads with an average of 636,318 reads per sample generated from the 291 accessions. Overall, 3.83 % of the reads matched chloroplast sequences and there was only minor contamination as measured by FastQ Screen (0% *E. coli*, 0.15 % human, and 2.36 % mouse hits). All GBS reads from 412 accessions were mapped to the phased *V. planifolia* 'Daphna' genome resulting in detection of 431,204 Single Nucleotide Polymorphisms (SNPs) using Freebayes. There were 15,746 SNPs after filtering based on missing data, allele frequency, and depth that were used for all analyses unless otherwise specified.

3.3. Heterozygosity

Heterozygosity was analyzed among the sequenced accessions to look for patterns among groups of accessions and species (Supplemental Table S1). As expected, accessions annotated as putative hybrids were among the highest heterozygous genotypes with a maximum of 31.3 %

for a verified *V. planifolia* x *V. pompona* hybrid (AC445_2). The average heterozygosity of all accessions annotated as hybrids was 15.3 %. Heterozygosity can be impacted by reads that map or do not map to the reference sequence due to diversity, and this could be reflected in the lower heterozygosity in general among the most diverse species tested. Conversely, analysis of *V. planifolia* accessions using the 'Daphna' genome was not limited in this way, and thus the results could be more rigorously interpreted. Natural breaks in the data were noticed when analyzing *V. planifolia* heterozygosity. One group of *V. planifolia* accessions included most of the accessions for this species (*V. planifolia* type 1) and had an average heterozygosity of 22.7 %. Two other *V. planifolia* groups had average heterozygosity levels of 12.2 % (*V. planifolia* type 3: AC185, AC364_2, AC366_2, AC368_2, AC381_2, AC382_2, AC535_2, and AC536_2) and 2.6 % (*V. planifolia* type 2: AC205, AC206, AC289_2, AC293_2, AC295_2, AC399_2, AC400_2, AC504_2, AC517_2, AC525_2, and AC540_2). These groups were also generally distinguished using a focused phylogenetic analysis of just *V. planifolia* and *V.* × *tahitensis* accessions as described in 3.8.

3.4. Phylogenetic and STRUCTURE analyses

Phylogenetic and STRUCTURE analyses were conducted to investigate relationships among the 412 vanilla accessions (Fig. 1, Supplemental Fig. S1). In general, species were uniquely separated into clades. As found previously [11], there were a number of misidentified, hybrid, or unknown accessions indicating the need to verify accessions using genomics approaches prior to investing heavily in plant breeding. Phylogenetic and STRUCTURE analyses were suitable in many cases to confidently assign species or hybrids within clusters, but this was not possible in every case where "species" was assigned to unknowns (Table S1). *V. planifolia* accessions formed a number of branches in the phylogenetic tree. Some branches were country-specific. For example, four accessions (AC525_2, AC524_2, AC399_2, and AC400_2) from Colombia formed one branch and included both *V. planifolia* and one *V.* × *tahitensis* (AC524_2) accessions. There were many branches that included accessions from both Mexico and the Philippines spread throughout the *V. planifolia* accessions. Belize had fewer accessions available for this study, but these were also dispersed among the *V. planifolia* accessions.

The GBS results also revealed some relationships in improved detail compared to previous studies. The *V.* × *tahitensis* accessions were similarly located on the phylogenetic tree and distinct from *V. planifolia* in the STRUCTURE plot. Three *V. sotoarenasii* accessions were found within the *V. planifolia* group with STRUCTURE showing similarity to *V. planifolia* and *V.* × *tahitensis*. *V. karen-christianae* formed a distinct group compared to other accessions.

3.5. Principal component analysis

Principal component analysis was conducted on all 412 accessions (Fig. 2a, Supplemental Table S2). Principal component 1 (PC1) captured 21.85 % of the variance, and Principal Component 2 (PC2) captured 17.59 % of the variance. *V. insignis* heavily skewed PC1. *V. planifolia* accessions were generally located together except for some low depth accessions including AC310_2, AC317_2, AC323_2, AC325_2, AC339_2, AC342_2, AC354_2, AC363_2, AC365_2, AC371_2, and AC397_2.

V. × *tahitensis* was found roughly between the *V. planifolia* and *V. odorata* groups when plotting PC1 and PC2, but PC2 and Principal Component 4 (PC4) showed better separation of *V. planifolia*, *V.* × *tahitensis*, and *V odorata*. The PC2 and PC4 plot (Fig. 2b) identified 11 *V. planifolia* type 2 accessions (AC205, AC206, AC289_2, AC293_2, AC295_2, AC399_2, AC400_2, AC504_2, AC517_2, AC525_2, and AC540_2) that were separated from the main *V. planifolia* group that tended towards *V.* × *tahitensis* along PC4. This is the same group of accessions that had an overall heterozygosity of 2.6 %. The three *V. sotoarenasii* accessions were also located among the *V. planifolia* type 2 accessions in Fig. 2b.

3.6. Global genetic variance analysis

Total genetic variance contained within subpopulations was analyzed by FST to investigate population differentiation due to genetic structure (Fig. 3). Higher FST values indicate higher differentiation among populations. As expected, *V. planifolia*, *V.* × *tahitensis*, and *V. odorata* showed low FST values in comparisons among the three groups. High FST values were found among a number of diverse species including *V. palmarum*, *V. trigonocarpa*, *V. dresslerii* and *V. griffithii*, *V. siamemsis*, *V. atropogon*, *V. madagascariensis*, *V. phalaenopsis*, *V. roscheri*, *V. imperialis*, *V. potaei*, *V. claviculata*, *V. dilloniana*, and *V. barbellata*.

3.7. Species identification using rbcL sequencing

A few accessions were selected for *rbcL* (ribulose-1,5-bisphosphate

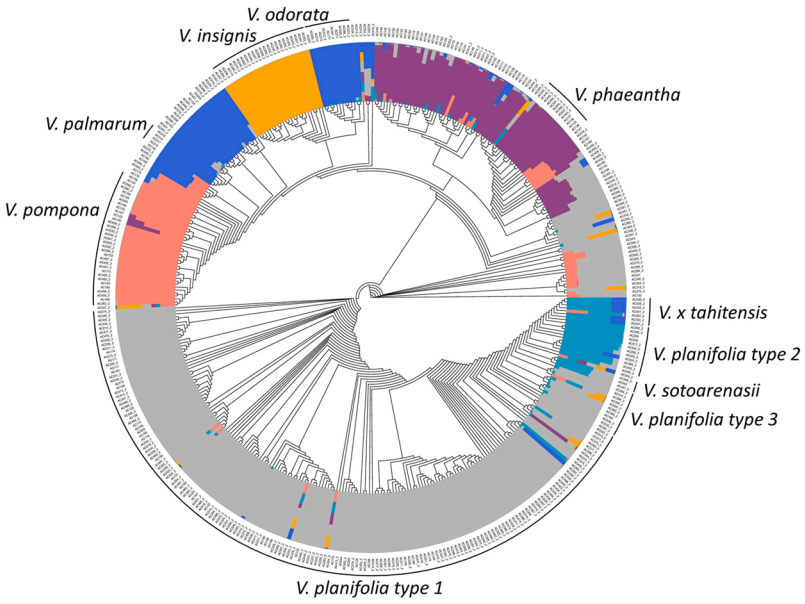

Fig. 1. Combined STRUCTURE (K = 6) and Circos plot of samples from this study (designated with "_2") and accessions from a previous study [11]. Selected species and types including *V.* × *tahitensis*, *V. planifolia* type 2, *V. sotoarenasii*, *V. planifolia* type 3, *V. planifolia* type 1, *V. pompona*, *V. palmarum*, *V. insignis*, *V. odorata*, and *V. phaeantha* are shown. Color coding in the STRUCTURE plot indicates similar SNPs patterns with gray indicating *V. planifolia* type 1 (as an example). See online version for color figure.

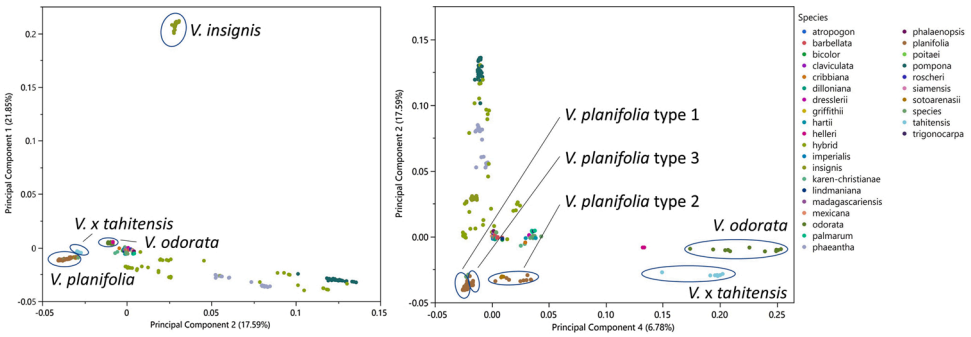

Fig. 2. Principal Component Analysis plots of 412 vanilla accessions for a) PC1/PC2 and b) PC2/PC4. Plots show multiple species including *V. planifolia*, *V.* × *tahitensis*, *V. odorata*, and *V. insignis* accessions within ovals. *V. planifolia* types 1, 2, and 3 are shown with greater differentiation in PC2/PC4. Species assignments are coded by color. See online version for color figure.

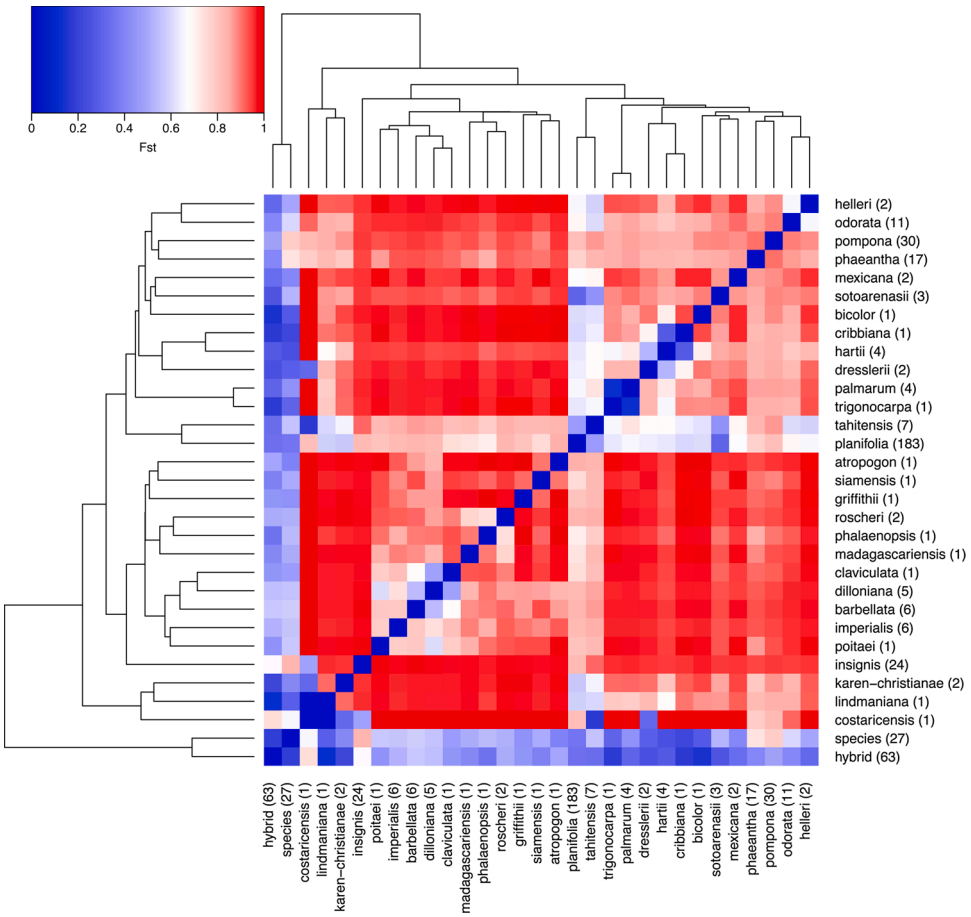

Fig. 3. FST analysis of species in this study. Higher FST values (1.0) are shown in red and indicate a higher degree of differentiation among populations. Lower Fst values are shown in blue and indicate lower degrees of differentiation among populations. See online version for color figure.

carboxylase) sequencing to confirm species assignments and to putatively identify maternal parents (Fig. 4). *rbcl* is a chloroplast-encoded, maternally-inherited gene. AC289_2, AC399_2, AC504_2, and AC525_2 were from *V. planifolia* type 2 and clustered closely with *V.* × *tahitensis*. These accessions had sequences identical to the most abundant *V. planifolia* type 1 sequences as expected. Two *V. planifolia*-like accessions, AC351_2 and AC361_2 were found on the phylogeny near the *V. planifolia/V. phaeantha* and *V. planifolia/V. pompona* hybrids. AC351_2 is from the Tuxtlas jungle in Veracruz Mexico where wild *V. planifolia* was collected more than 100 years ago and where crops were never established. This accession had the normal *V. planifolia rbcL* sequence, but had a much lower heterozygosity (2.2 %) than most *V. planifolia* (22.7 %) and was not remarkably deficient for sequencing depth or percent missing data. AC361_2 is from the jungle of Chimalapas Oaxaca, Mexico in the same region that included sterile plants that were morphologically similar to *V.* × *tahitensis* [55]. This accession had one SNP in the *rbcL* sequence distinguishing it from the typical *V. planifolia*, had low mean SNP depth (4.0) that could indicate poor quality sequencing, and is morphologically similar to *V. planifolia* type 1 accessions. AC433_2 and AC526_2 are *V.* × *tahitensis* accessions that matched the *V. planifolia rbcL* sequence as expected due to *V. planifolia* being the maternal parent. AC534 was identified as a *V. karen-christianae* and had an *rbcL* sequence identical to *V. odorata*, but was clearly separated from *V. odorata* on a separate branch in the phylogenetic tree.

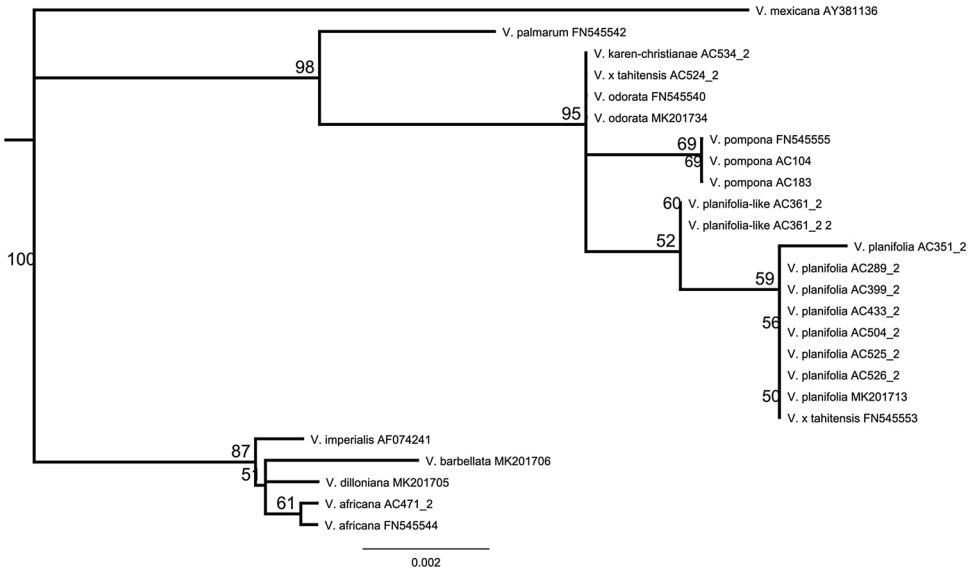

Fig. 4. Species analysis using the plastid-encoded, partial *rbcL* gene of selected accessions. Selected species from NCBI are also shown and include GenBank IDs. Accessions tested from this study mostly matched *V. planifolia* type 1 sequences except for AC524_2 (a unique *V.* × *tahitensis*) and AC534_2 (*V. karen-christianae*) that matched *V. odorata,* and AC471_2 that matched *V. africana*. AC361_2 had a single SNP differentiating it from *V. planifolia* type 1.

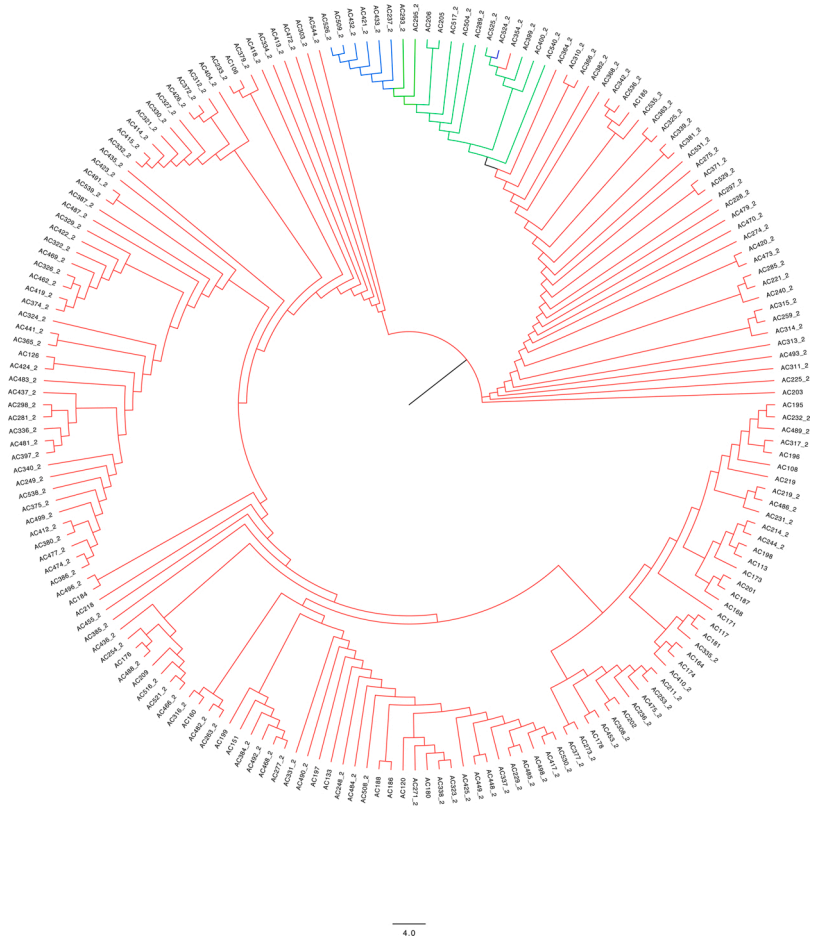

Fig. 5. Phylogenetic tree including only *V. planifolia* and *V.* × *tahitensis* accessions. The figure shows a major split among the *V. planifolia* accessions. *V.* × *tahitensis* (blue) accessions are grouped together and mostly distinct from *V. planifolia* type 1 (green). *V. planifolia* types 1 and 3 are shown in red. Accession 354_2 (red) is shown among *V. planifolia* group 2 accessions (green), but had low sequencing depth and might not be placed accurately in the tree.

AC471_2 from the Philippines was most similar to a *V. africana* sequence (GenBank FN545544) deposited on NCBI and was most similar to AC537_2 (a putative *V. havilandii*) on the phylogenetic tree. AC524_2 was identified as *V.* × *tahitensis* from the phylogenetic and STRUCTURE analyses, but the *rbcL* sequence matched *V. odorata* indicating this this accession is most likely a *V. odorata* x *V. planifolia* hybrid versus the more commonly cultivated *V.* × *tahitensis*, a *V. planifolia* x *V. odorata* hybrid.

3.8. Diversity within *V. planifolia* and *V.* × *tahitensis*

The diversity within *V. planifolia* and *V.* × *tahitensis* is especially valuable due to the current vanilla Standard of Identity that only permits these two types for inclusion in vanilla extract. Therefore, the most straightforward approach (from a regulatory perspective) for the genetic improvement of vanilla includes identifying and leveraging diversity within these two types of vanilla. A separate analysis was conducted that only included *V. planifolia* and *V.* × *tahitensis* accessions. The SNPs from these accessions were filtered using the same parameters as the complete dataset and resulted in 14,133 SNPs. Analysis of *V. planifolia* and *V.* × *tahitensis* using the 14,133 SNPs showed that the accessions split into two groups with one comprised of *V. planifolia* type 1 accessions and the other with *V. planifolia* type 2, *V. planifolia* type 3, and *V.* × *tahitensis* accessions (Fig. 5). The 11 *V. planifolia* type 2 accessions that were differentiated in Fig. 2b were located closest to the *V.* × *tahitensis* accessions in Fig. 5.

PCA analysis of 190 *V. planifolia*, *V.* × *tahitensis*, and *V. sotoarenasii* accessions showed a similar trend to the phylogenetic analysis (Fig. 6). PC1 explained 27.33 % of the variance and PC2 explained 7.98 %. The *V. planifolia* types 1, 2, and 3 showed some separation from each other. *V.* × *tahitensis* was separated from *V. planifolia* in the PCA plot, and *V. sotoarenasii* accessions were located close to the *V. planifolia* type 2 accessions.

3.9. Genetic variance within *V. planifolia* and *V.* × *tahitensis*

FST analysis was also performed on just the 190 *V. planifolia* and *V.* × *tahitensis* accessions along with three additional *V. sotoarenasii* accessions (Fig. 7). The *V. planifolia* accessions were split by type for this analysis. For *V.* × *tahitensis*, *V. planifolia* type 3 by *V.* × *tahitensis* had the highest FST (0.482) compared to *V. planifolia* type 1 by *V.* × *tahitensis* (0.478) or *V. planifolia* type 2 by *V.* × *tahitensis* (0.413). *V. planifolia* type 2 by *V. planifolia* type 3 had the highest FST value (0.488) among

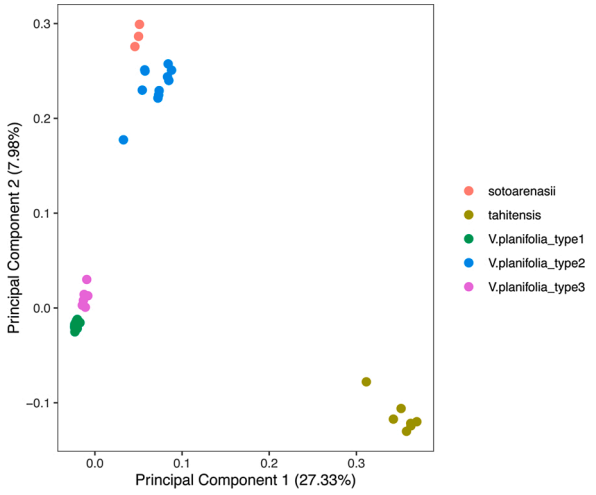

Fig. 6. PCA plot of 190 accessions including only *V. planifolia* and *V.* × *tahitensis*. *V. planifolia* types 1 and 3 are shown in red, *V. planifolia* type 2 is shown in green, and *V.* × *tahitensis* accessions are shown in blue. See online version for color figure.

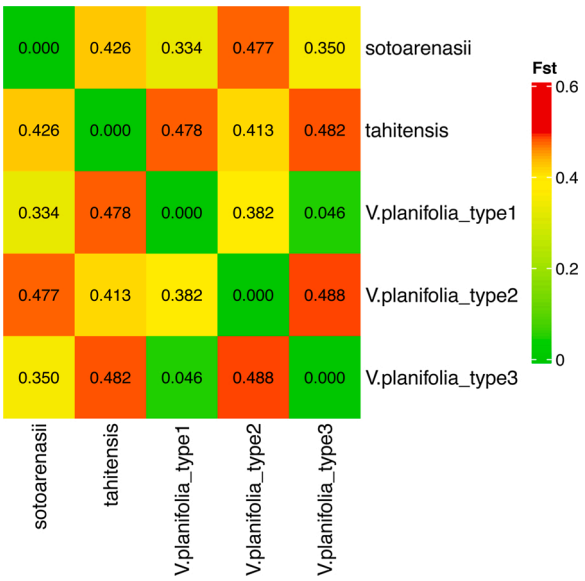

Fig. 7. FST analysis of *V. planifolia*, *V.* × *tahitensis*, and *V. sotoarenasii* accessions. Higher FST values imply greater differentiation among populations. *V. planifolia* type 1, *V. planifolia* type 2, *V. planifolia* type 3, *V.* × *tahitensis* (tahitensis), and *V. sotoarenasii* (sotoarenasii) are shown. See online version for color figure.

V. planifolia types compared to *V. planifolia* type 1 and *V. planifolia* type 2 (0.382) or *V. planifolia* type 1 and *V. planifolia* type 3 (0.046). *V. sotoarenasii* and *V. planifolia* type 2 had the highest FST (0.477) among other *V. planifolia* comparisons with *V. sotoarenasii*. Thus, *V. planifolia* group 2 could be considered to have the lowest degree of differentiation from *V.* × *tahitensis* among the other comparisons in Fig. 7.

3.10. Parentage of *V.* × *tahitensis*

The maternal parent of the *V.* × *tahitensis* hybrid is *V. planifolia* as evidenced by multiple studies showing *rbcL* sequences that are identical between the two types. Both types are also closely located in the phylogenetic trees leveraging genomics-based diversity analysis. It is generally hypothesized that the paternal parent of *V.* × *tahitensis* was *V. odorata*, but supporting information is generally lacking especially for vouchered, verified, and publicly available accessions. Additional supporting evidence would strengthen the hypothesis that *V.* × *tahitensis* is a *V. planifolia* x *V. odorata* hybrid. The original 431,204 SNPs were filtered by quality resulting in 143,122 SNPs that were then screened for 1) homozygosity (0/0) in a subset of *V. planifolia* accessions and 2) heterozygosity (0/1) in *V.* × *tahitensis*. This would identify minor alleles that were contributed from the paternal parent. The resulting filtered set of 1,544 SNPs were used to screen potential minor allele donors among the other sequenced species. As expected, *V.* × *tahitensis* accessions had the highest number of SNPs matching the above criteria with 1,485 to 1,535 of the 1,544 minor alleles. The highest number of SNP matches in accessions other than *V.* × *tahitensis* were all *V. odorata* AC212_2 (1,507 minor allele count, Costa Rica), 341_2 (1,402 minor allele count, Mexico), AC458_2 (1,371 minor allele count, Belize), AC213 (1,360 minor allele count, Belize), AC207 (1,355 minor allele count, Belize), and AC177 (1,106 minor allele count, unknown provenance). The next highest match was *V. helleri* (AC434_2) with a much lower minor allele count of 464. In terms of morphology, AC341_2 and AC350_2 flowers are similar but not identical to *V. odorata* accessions from Central and South America (Fig. 8). The area where these specimens were collected is relatively close to previously collected *V. odorata* accessions used in a report investigating the hybrid origin of *V.* × *tahitensis* [25].

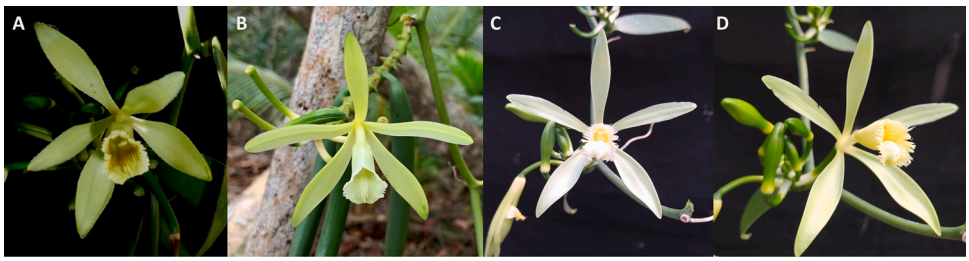

Fig. 8. Representative images from selected accessions. A) flower from a reverse *V.* × *tahitensis* accession (AC352_2) from Mexico, B) *V. sotoarenasii* flower at the type locality in Costa Rica (AC209_2), C) flower of *V. odorata* AC341_2, and D) flower of *V. odorata* AC350_2.

3.11. *V.* × *tahitensis maternal parent*

V. planifolia is the maternal species of the *V.* × *tahitensis* hybrid, but the improved resolution of genomics-based diversity analysis could identify a *V. planifolia* accession that is most similar to the *V. planifolia* genotype in *V.* × *tahitensis*. The hypothesis was that sampling among so many accessions from the center of vanilla cultivation could identify genotypes among *V. planifolia* accessions that are more similar to *V.* × *tahitensis* than others. Therefore, assuming *V. odorata* as the paternal parent, 431,204 SNPs were filtered for high major allele homozygosity (0/0) in *V. odorata* accessions but that were also heterozygous (0/1) in *V.* × *tahitensis*. The resulting 468 SNPs were then used to identify accessions with the highest minor allele count that could have been inherited by *V.* × *tahitensis* from the maternal parent. All six *V.* × *tahitensis* accessions were among the best candidates with 455–465 out of the 468 minor alleles as expected. There were seven *V. planifolia* accessions that most closely matched the *V.* × *tahitensis* minor alleles including AC293_2 (457 minor allele count, Belize), AC295_2 (456 minor allele count, Belize), AC517_2 (455 minor allele count, Belize), AC206 (447 minor allele count, Belize), AC504_2 (432 minor allele count, Nicaragua), AC205 (414 minor allele count, Belize), and AC289_2 (409 minor allele count, Belize). The next closest accession was AC313_2, a *V. planifolia* from the Northern Sierra region of Oaxaca with a 356 minor allele count. All of the accessions most closely matching *V.* × *tahitensis* were categorized as *V. planifolia* type 2 and primarily from Belize as measured both by global SNP analysis and by the reduced set of targeted SNPs.

4. Discussion and conclusions

Genomics-based diversity analyses are important for the conservation and genetic improvement of historically under-researched genera like *Vanilla*. Previous studies relying on few specimens to analyze the variation of morphological traits and limited molecular markers have been useful, but broader morphological assessment and genetic sampling with increased resolution and reduced cost per datapoint are leading to new discoveries. The potential to rapidly characterize wild populations with thousands of SNPs is transforming conservation approaches and is helping to resolve previously ambiguous species assignments. For example, *V. sotoarenasii* was entirely found within the *V.* × *tahitensis* and *V. planifolia* clades indicating a close relationship to these accessions. Further, accession AC209_2 from Costa Rica comes from the type locality (*locus typicus*) or type population for *V. sotoarenasii*. This supports the morphological overlap observed when studying multiple specimens belonging to these three taxa indicating that *V. sotoarenasii* as a heterotypic, nomenclatural synonym of *V. planifolia* [12].

Vanilla diversity studies tend to focus on accessions from Mexico and massively clonally propagated plants in commercial production [16,29,30]. These studies point to a few, foundational vanilla clones from Papantla, Veracruz, Mexico that have become the predominant vanilla type in commercial production. While Mexico is the likely orgin of today's commercial vanilla plants, our sampling of vanilla accessions beyond Mexico emphasizes the need to work across borders to capture a more complete representation of the depth of vanilla diversity. This is supported by modeling work that indicates the potential distribution of *V. planifolia* from North, Central, and South America [56]. Additionally, the potential of genomics-based diversity analyses to support plant breeding has been reviewed by many groups and for many crop species [57–59] including the use of wild relatives in crop breeding [60]. This underscores the potential synergy from expanding vanilla sampling, conservation, and genomics to support vanilla plant improvement.

The global results from this study closely match the previous vanilla GBS study [11]. This was not surprising given the similar sequencing approaches that both centered on mapping sequencing reads to a version of the *V. planifolia* 'Daphna' genome. One major benefit of the present study was the increased number of *V. planifolia* accessions from multiple international collections. Additionally, the inclusion of a number of *V.* × *tahitensis* accessions in this study finally unlocked this genotype for analysis. The previous genomics-based diversity study on vanilla had two accessions that were inaccurately assigned as *V.* × *tahitensis*, but are now considered to be *V. planifolia* that is closely related to *V.* × *tahitensis*. Additionally, the relative presentation of the more highly represented species was different between the two studies, but the overall species assignments were consistent. There were a number of accessions in both studies that will require follow up research to identify if they are distinct species or hybrids.

The major strength of this study includes being the first genomics-enabled characterization of multiple international collections around the center of diversity for *V. planifolia* specifically. Multiple branches were formed among *V. planifolia* accessions, and these will be useful for selecting representative types for future phenotyping, in-depth characterization, and plant breeding. The split among *V. planifolia* into types based on phylogenetic, PCA, and STRUCTURE analyses was reported here for the first time and was only able to be resolved due to the abundance of SNPs used for these analyses. Phenotypic characterization of representatives from these types in the future could guide parental selection for breeding work. The results from this study also demonstrate the need for conservation of vanilla species throughout Central and South America especially for the commercial species as truly wild vanilla in Mexico is exceptionally rare. For example, there were *V. planifolia* accessions from the remote, native forests of Colombia with interesting genetic relationships to other *V. planifolia* and *V.* × *tahitensis* accessions, yet are from an under-researched geography. Additional research could identify novel traits from these accessions, or help further our understanding of the native or migratory distribution of *V. planifolia*.

The origin of *V.* × *tahitensis* has intrigued the vanilla industry for over 100 years due to its distinct flavor profile. This hybrid origin of *V.* × *tahitensis* is confirmed and expanded in the present study. A number of *V. planifolia* type 2 accessions were identified that share a more similar genotype with *V.* × *tahitensis* than the most abundant *V. planifolia* type 1 accessions in general. The *V. planifolia* and *V. odorata* accessions with the genotypes most closely matching *V.* × *tahitensis* point towards an origin in what is now Belize. Additional research in the future could help confirm the Belize origin hypothesis, but will require additional international support especially within Central America and in French

Polynesia where *V.* × *tahitensis* is commonly grown. Future work should also include re-creating *V.* × *tahitensis* not by chance in a forest ecosystem but by selecting superior *V. planifolia* and *V. odorata* parents and creating new hybrid plants. This will enable the analysis of segregating populations of sufficient size to further investigate genetic contributions to the unique aroma profile and non-dehiscing pods of *V.* × *tahitensis*. The distinct morphology of *V. sotoarenasii* and its close genetic relationship with *V.* × *tahitensis* could warrant further investigation as a variant of *V. planifolia*.

Analyzing a large number of samples from multiple countries has provided a unique view of vanilla across borders as shown in this study and in others [4,29,30]. In Colombia, for example, it is common to find individuals with intermediate morphological characters between sympatric species, which suggests high natural hybridization [14]. Colombian researchers have thought that *V. odorata* from the tropical moist forest (bio-geographic Chocó) is not the same as the *V. odorata* from the dry forests. The results of the present study suggest that vanilla growing in these moist forests also includes *V.* × *tahitensis*. This was the case for AC524_2 that was collected in Colombian moist forest and originally designated as *V. odorata* because of its morphology. The *rbcL* results indicating that AC524_2 is a reverse (*V. odorata* x *V. planifolia*) *V.* × *tahitensis* hybrid is interesting. This could further support the hypothesis that *V.* × *tahitensis* is the result of a natural, although rare, hybridization event between species with overlapping native ranges. This is not the first time this has been suggested. AC352_2 was originally collected in January 2009 in Mexico south of Jalisco. Miguel Soto-Arenas, a well-known and respected vanilla researcher, indicated that it was almost certainly *V.* × *tahitensis* 'Haapape' (Fig. 8) based on pictures, and the accession was included in a published, intermediate report [55]. AC352_2 was also previously found to have a *V. odorata rbcL* sequence (data not published), and was sourced from the same area as AC341_2 (*V. odorata*) and AC350_2 (*V. odorata*, low depth). Additional analysis of vanilla specimens collected in Jalisco, Mexico and the moist forests of Chocó, Colombia could support or refine the current migratory theory that *V.* × *tahitensis* was introduced from Mexico to the Philippines and then Tahiti [7], and could have been moved from Tahiti or from Mexico to the wet areas of the Colombian Pacific naturally or through commercial shipping routes.

Resolving the classification of vanilla species and hybrids is benefiting from genomics. For example, accession AC396_2 was compatible with *V. dressleri* and turned out to be a hybrid with information mainly from *V. planifolia*. In nature, there is difficulty in distinguishing among hybrids, and it is possible to find individuals that are not necessarily a first-generation hybrid. These unknown accessions can possess intermediate traits that make precise phenotypic determination difficult. An encouraging result from this study includes AC401_2, which has some morphological characters similar to *V. dressleri* (eg. hard and rigid stems), as having a unique genotype compared to other accessions of *V. dressleri* included in this study. This finding is promising if its organoleptic qualities are taken into account, since it shares some of its aroma profile with *V. planifolia* [61].

In conclusion, the major challenge for vanilla research in general, and for *V.* × *tahitensis* specifically, includes the lack of germplasm availability and thoroughly characterized germplasm. This creates a dearth of consensus among research groups, prevents follow up studies to confirm research results, and reduces overall synergy. Vanilla is a spice with global demand and could benefit from increased research across borders as demonstrated in this study. Even though vanilla is a commodity traded spice, the international movement of vanilla material is currently restricted by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) that has historically created roadblocks for advancing research and genetic improvement of this species. Protecting endangered species while facilitating the safe movement of vanilla germplasm are perfectly compatible with the long-term strategy of vanilla conservation, research, and improvement through plant breeding and could be facilitated through a global, collaborative network of vanilla researchers.

Author contributions statement

AHC: Conceptualization, Methodology, Data curation, Writing – original draft, Supervision, Project administration, Funding acquisition. YAA: Formal analysis, Visualization. MB: Investigation, Methodology. All authors: Investigation, Resources, Writing – Review & Editing.

Funding

Funding for this research was provided to AHC by a University of Florida SEEDIT grant. The sponsor did not have a role in study design, data collection, interpretation, or writing of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks are expressed to Dr. Robert T. González for his valuable comments and contributions on the Colombian species and for his help in the collection of specimens. We also thank Dr. Tupac Otero of the Universidad Nacional de Colombia for his collaboration in collecting the samples and Dr. Pablo R. Stevenson of the Universidad de Los Andes for the academic support. The material from Colombia was legally exported through CITES permit 44706 from December 2019, given by Ministerio de Ambiente y Desarrollo Sostenible.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.plantsci.2021.111019.

References

[1] H.C. Andrews, The Botanist's Repository for New and Rare Plants, London, 1797.
[2] H. Bruman, The culture history of Mexican vanilla, Hisp. Am. Hist. Rev. 28 (1948) 360–376.
[3] M.A. Soto-Arenas, Filogeografía y recursos genéticos de las vainillas de México, Instituto Chinoin AC, 1999. Informe final SNIB-Conabio Proyecto J. 101.
[4] S. Bory, P. Lubinsky, A.M. Risterucci, J.L. Noyer, M. Grisoni, M.F. Duval, P. Besse, Patterns of introduction and diversification of Vanilla planifolia (Orchidaceae) in Reunion Island (Indian Ocean), Am. J. Bot. 95 (2008) 805–815.
[5] P. Lubinsky, M. Van Dam, A. Van Dam, Pollination of Vanilla and evolution in Orchidaceae, Lindleyana 75 (2006) 926–929.
[6] E. Householder, J. Janovec, A.B. Mozambite, J.H. Maceda, J. Wells, R. Valega, H. Maruenda, E. Christenson, Diversity, natural history, and conservation of Vanilla (Orchidaceae) in Amazonian wetlands of Madre De Dios, Peru, J. Bot. Res. Inst. Tex. (2010) 227–243.
[7] M.A. Soto-Arenas, R.L. Dressler, 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 (2010) 285–354.
[8] C. Morren, Sur la fructification de la vanille obtenue au moyen de la fécondation artificielle, Comptes Rendues Acad. Sci. 6 (1838) 489–492.
[9] T. Ecott, Vanilla: Travels in Search of the Luscious Substance, Penguin UK, 2005.
[10] T. Hasing, H. Tang, M. Brym, F. Khazi, T. Huang, A.H. Chambers, A phased Vanilla planifolia genome enables genetic improvement of flavour and production, Nat. Food 1 (2020) 811–819.
[11] Y. Hu, M. Resende, A. Bombarely, M. Brym, E. Bassil, A. Chambers, Genomics-based diversity analysis of Vanilla species using a Vanilla planifolia draft genome and Genotyping-By-Sequencing, Sci. Rep. 9 (2019) 3416.
[12] A.P. Karremans, I.F. Chinchilla, G. Rojas-Alvarado, M. Cedeño-Fonseca, A. Damián, G. Léotard, A reappraisal of Neotropical Vanilla. With a note on taxonomic inflation and the importance of alpha taxonomy in biological studies, Lankesteriana 20 (3) (2020) 395–497.
[13] K.M. Cameron, Vanilla phylogeny and classification, in: D. Havkin-Frenkel, F. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2011, pp. 241–255.
[14] N.S. Flanagan, P. Chavarriaga, A.T. Mosquera-Espinosa, Conservation and sustainable use of Vanilla crop wild relatives in Colombia, in: D. Havkin-Frenkel,

F. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2018, pp. 85–110.
[15] N. Childers, H. Cibes, in: C. Withner (Ed.), The Orchids: A Scientific Study, The Ronald Press Company, New York, 1959, pp. 477–508.
[16] S. Bory, M. Grisoni, M.-F. Duval, P. Besse, Biodiversity and preservation of vanilla: present state of knowledge, Genet. Resour. Crop Evol. 55 (2008) 551–571.
[17] D.S. Correll, Vanilla-Its botany, history, cultivation and economic import, Econ. Bot. 7 (1953) 291–358.
[18] S. Koyyappurath, T. Atuahiva, R. Le Guen, H. Batina, S. Le Squin, N. Gautheron, V. E. Hermann, J. Peribe, M. Jahiel, C. Steinberg, E.C.Y. Liew, C. Alabouvette, P. Besse, M. Dron, I. Sache, V. Laval, M. Grisoni, Fusarium oxysporum f. sp radicis-vanillae is the causal agent of root and stem rot of vanilla, Plant Pathol. 65 (2016) 612–625.
[19] J. Hernandez-Hernández, Mexican vanilla production, in: D. Havkin-Frenkel, F. C. Belanger (Eds.), Handbook of Vanilla Science and Technology, Wiley-Blackwell, Singapore, 2018, pp. 3–25.
[20] United States Food and Drug Administration, Code of Federal Regulations Title 21 Volume 2 Part 169.3, US Food and Drug Administration, 2021.
[21] R. Menchaca, J. Ramos, D. Moreno, M. Luna, M. Mata, L. Vázquez, M. Lozano, In vitro germination of Vanilla planifolia and V. pompona hybrids, Revista Colombiana de Biotecnología 13 (2011) 80–84.
[22] D. Ehlers, M. Pfister, Compounds of vanillons (Vanilla pompona Schiede), J. Essent. Oil Res. 9 (1997) 427–431.
[23] FAOSTAT, Food and Agriculture Organization of the United Nations, in, 2021.
[24] FAO, Vanilla: Post-harvest Operations, 2009.
[25] P. Lubinsky, K.M. Cameron, M.C. Molina, M. Wong, S. Lepers-Andrzejewski, A. Gómez-Pompa, S.C. Kim, Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, Vanilla tahitensis (Orchidaceae), Am. J. Bot. 95 (2008) 1040–1047.
[26] L.R. Nielsen, H.R. Siegismund, Interspecific differentiation and hybridization in Vanilla species (Orchidaceae), Heredity 83 (1999) 560–567.
[27] P. Besse, D. Da Silva, S. Bory, M. Grisoni, F. Le Bellec, M.F. Duval, RAPD genetic diversity in cultivated vanilla: Vanilla planifolia, and relationships with V. x tahitensis and V. pompona, Plant Sci. 167 (2004) 379–385.
[28] D. Minoo, V.N. Jayakumar, S.S. Veena, J. Vimala, A. Basha, K.V. Saji, K.N. Babu, K. V. Peter, Genetic variations and interrelationships in Vanilla planifolia and few related species as expressed by RAPD polymorphism, Genet. Resour. Crop Evol. 55 (2008) 459–470.
[29] P.M. Schlüter, M.A.S. Arenas, S.A. Harris, Genetic variation in Vanilla planifolia (Orchidaceae), Econ. Bot. 61 (2007) 328.
[30] P. Lubinsky, S. Bory, J. Hernandez, S.C. Kim, A. Gomez-Pompa, Origins and dispersal of cultivated vanilla (Vanilla planifolia jacks. [Orchidaceae]), Econ. Bot. 62 (2008) 127–138.
[31] S. Lepers-Andrzejewski, S. Causse, B. Caromel, M. Wong, M. Dron, Genetic linkage map and diversity analysis of Tahitian vanilla (Vanilla× tahitensis, Orchidaceae), Crop Sci. 52 (2012) 795–806.
[32] A.L. Ramos-Castellá, L.G. Iglesias-Andreu, J. Martínez-Castillo, M. Ortíz-García, R. H. Andueza-Noh, P. Octavio-Aguilar, M. Luna-Rodríguez, Evaluation of molecular variability in germplasm of vanilla (Vanilla planifolia G. Jackson in Andrews) in Southeast Mexico: implications for genetic improvement and conservation, Plant Genet. Resour. 15 (2017) 310–320.
[33] S. Bory, D. Da Silva, A.-M. Risterucci, M. Grisoni, P. Besse, M.-F. Duval, Development of microsatellite markers in cultivated vanilla: polymorphism and transferability to other vanilla species, Sci. Hortic. 115 (2008) 420–425.
[34] V.B. Perez, L.G.I. Andreu, E.A.E. Manzano, J.M. Castillo, M.M.O. Garcia, P. O. Aguilar, Molecular and microclimatic characterization of two plantations of Vanilla planifolia (Jacks ex Andrews) with divergent backgrounds of premature fruit abortion, Sci. Hortic. 212 (2016) 240–250.
[35] R.L. Gigant, N. Rakotomanga, C. Goulié, D. Da Silva, N. Barre, G. Citadelle, D. Silvestre, M. Grisoni, P. Besse, Microsatellite markers confirm self-pollination and autogamy in wild populations of Vanilla mexicana Mill. (syn. V. inodora) (Orchidaceae) in the island of Guadeloupe. Microsatellite Markers, InTech, 2016, pp. 73–94.
[36] M.A. Soto Arenas, R.L. Dressler, 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 Int. J. Orchid 9 (2010) 285–354.
[37] K.M. Cameron, Utility of plastid psaB gene sequences for investigating intrafamilial relationships within Orchidaceae, Mol. Phylogenet. Evol. 31 (2004) 1157–1180.
[38] A. Bouetard, P. Lefeuvre, R. Gigant, S. Bory, M. Pignal, P. Besse, M. Grisoni, Evidence of transoceanic dispersion of the genus Vanilla based on plastid DNA phylogenetic analysis, Mol. Phylogenet. Evol. 55 (2010) 621–630.
[39] K.M. Cameron, M. Carmen Molina, Photosystem II gene sequences of psbB and psbC clarify the phylogenetic position of Vanilla (Vanilloideae, Orchidaceae), Cladistics 22 (2006) 239–248.
[40] K.M. Cameron, M.W. Chase, W.M. Whitten, P.J. Kores, D.C. Jarrell, V.A. Albert, T. Yukawa, H.G. Hills, D.H. Goldman, A phylogenetic analysis of the Orchidaceae: evidence from rbcL nucleotide sequences, Am. J. Bot. 86 (1999) 208–224.
[41] M. Divakaran, K.N. Babu, P.N. Ravindran, K.V. Peter, Interspecific hybridization in vanilla and molecular characterization of hybrids and selfed progenies using RAPD and AFLP markers, Sci. Hortic. 108 (2006) 414–422.
[42] P.C. Verma, D. Chakrabarty, S.N. Jena, D.K. Mishra, P.K. Singh, S.V. Sawant, R. Tuli, The extent of genetic diversity among Vanilla species: comparative results for RAPD and ISSR, Ind. Crop Prod. 29 (2009) 581–589.
[43] R. Sreedhar, L. Venkatachalam, K. Roohie, N. Bhagyalakshmi, Molecular analyses of Vanilla planifolia cultivated in India using RAPD and ISSR markers, Orchid Sci. Biotechnol. 1 (2007) 29–33.
[44] K.M. Cameron, On the value of nuclear and mitochondrial gene sequences for reconstructing the phylogeny of vanilloid orchids (Vanilloideae, Orchidaceae), Ann. Bot. 104 (2009) 377–385.
[45] S. Villanueva-Viramontes, M. Hernández-Apolinar, G.C. Fernández-Concha, A. Dorantes-Euán, G.R. Dzib, J.M. Castillo, Wild Vanilla planifolia and its relatives in the Mexican Yucatan Peninsula: systematic analyses with ISSR and ITS, Bot. Sci. 95 (2017) 169–187.
[46] H.E.L. Lischer, L. Excoffier, PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs, Bioinform 28 (2012) 298–299.
[47] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[48] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[49] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[50] D.A. Earl, B.M. Vonholdt, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conserv. Genet. Resour. 4 (2012) 359–361.
[51] X.W. Zheng, D. Levine, J. Shen, S.M. Gogarten, C. Laurie, B.S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data, Bioinformatics 28 (2012) 3326–3328.
[52] T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new method for the analysis of genetically structured populations, BMC Genet. 11 (2010).
[53] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population-structure, Evolution 38 (1984) 1358–1370.
[54] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[55] M.A. Soto Arenas, Recopilación y análisis de la información existente sobre las especies mexicanas del género Vanilla, Reporte, Herbario de la Asociación Mexicana de Orquideología, AC, Instituto Chinoín, AC, México, DF, 2009.
[56] P. Ellestad, F. Forest, M. Serpe, S.J. Novak, S. Buerki, Harnessing large-scale biodiversity data to infer the current distribution of Vanilla planifolia (Orchidaceae), Bot. J. Linn. Soc. 196 (2021) 407–422.
[57] J.F. He, X.Q. Zhao, A. Laroche, Z.X. Lu, H.K. Liu, Z.Q. Li, Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, Front. Plant Sci. 5 (2014) 484.
[58] S. Getachew, N. Bille, J. Bell, W. Gebreselassie, Genotyping by sequencing for plant breeding-a review, Adv. Biotechnol. Microbiol. 14 (2020), 555891.
[59] C. Kim, H. Guo, W.Q. Kong, R. Chandnani, L.S. Shuang, A.H. Paterson, Application of genotyping by sequencing technology to a variety of crop breeding programs, Plant Sci. 242 (2016) 14–22.
[60] H. Dempewolf, G. Baute, J. Anderson, B. Kilian, C. Smith, L. Guarino, Past and future use of wild relatives in crop breeding, Crop Sci. 57 (2017) 1070–1082.
[61] R.T. González Mina, Plasticidad morfofisiológica comparativa de especies de vanilla silvestres del pacífico Colombiano, Doctorado en Ciencias Agrarias, 2016.
[62] J.A. Cibrián, *Variación genética de Vanilla planifolia en México*, Undergraduate Thesis. Facultad de Ciencias, Universidad Nacional Autónoma de México, 1999.