UNIVERSITÀ DEGLI STUDI DI PADOVA

Department of Comparative Biomedicine and Food Science – BCA

Second Cycle Degree (MSc)
in
Biotechnologies for Food Science

Genetic characterization of Mandevilla spp. experimental lines by gSSR genotyping and cpDNA barcoding

Supervisor
Prof. Gianni Barcaccia
Co-supervisor
Dr. Fabio Palumbo

Candidate
Chandana Mulagala

Academic year 2021-2022

INDEX

Abstract

Mandevilla is a genus comprehending several ornamental plant species with a brilliant future overall in view of its high business acknowledgment and added esteem. Its nomenclature also includes *Dipladenia*, while its common name is "Brazilian jasmine" and is a decorative plant that was brought into the European market around 150 years ago. Until 1955, its development was limited to experienced English landscapers, and it was then spread all through Europe by Danish horticulturists. At present, this plant can be tracked down covering overhangs, lattices, arbores, and scenes, adding a tropical style to any open-air space. *Mandevilla* is particularly valued for its remarkable protection from wind, dry season, and pungent air, making it an ideal flower for summer in the Mediterranean region.

However, like most of the ornamental crops, there are very few molecular tools that could be used for the purpose of cultivar innovation and breeding. In this research work, we report the development and analysis of 23 Simple Sequence Repeat (SSR) markers in *Mandevilla*. The microsatellites used in the work were developed and isolated from a newly arranged draft genome of the sample named "2001". The developed markers were thus evaluated on a *Mandevilla* core collection, consisting of 55 samples of breeding and pre-commercial value, to determine the discrimination capability of the molecular markers panel, to characterize the core collection genotypes, and for evaluating the genetic similarity and genetic structure of the core collection in exam. Moreover, a DNA barcoding approach was also performed to investigate the species ancestry, since most of the genotypes available are probably interspecific hybrids, characterized by complex genealogies.

The analysis parameters estimated after the amplification of the samples, a very garbled scenario, due to interspecific crosses. Based on the genetic similarity matrix, 7 comparisons were found to have 100% genetic similarity and the least genetic similarity was identified to be 50.2%. Genetic structure analysis revealed that the most probable number of ancestors constituting the core collection was equal to 26. The observation of admixed patterns was interpreted as a consequence of interspecific crosses that took place to obtain required morphological characters throughout the centuries. DNA barcoding and haplotyping analysis based on the *rbc*L and ITS1 genes divided samples into two clusters indicating their maternal inheritance patterns.

Overall molecular data highlighted that the core collection of *Mandevilla* analyzed in this study has a very complex genetic structure and is characterized by high genetic diversity. Thus, the molecular markers under study could find utility in breeding programs aimed at the selection of valuable experimental lines well adapted to the agronomic and environmental conditions, and in genetic traceability studies of the *Mandevilla* spp.

Introduction

Dipladenia or *Mandevilla* – Origin, Distribution, Domestication and species

Mandevilla, also known as *Dipladenia*, is a Brazilian ornamental plant that was introduced in the European market about 150 years ago under the common name "Brazilian Jasmine". Their growth was limited to experienced English gardeners until 1955, when it was spread across Europe by Danish horticulturists. This plant can currently be found on balconies, trellises, arbores and landscapes and gives every outdoor area a tropical flair. *Mandevilla* is particularly valued for its excellent resistance to wind, drought and salty air, which makes it an ideal flower for summer sales in the Mediterranean area (Oder, Lannes and Viruel, 2016). In a less tropical climate, it will require the warmth of a heated greenhouse or a cool conservatory. In the past decade, *Mandevilla's* commercial scene has changed a lot; the increasing demand in the European market goes hand in hand with an expansion to the American, Asian and Australian markets. The extraordinary increase in the number of available commercial varieties genotypes, from approximately 10 to 100 in the last 10 years (Oder, Lannes and Viruel, 2016), shows the growing interest of producers and consumers to this plant. The high commercial demand and the advantages for the growers put *Mandevilla* in a privileged position among the leading developers of ornamental plants in the new emerging markets.

Mandevilla Lindl. (Apocynaceae, Apocynoideae) is the largest genus of the Mesechiteae tribe, with around 170 species of lianas, vines and suffruticose herbs that are common in the neotropics of Mexico and the Antilles to northern Argentina. These plants are adapted to a wide variety of habitats, such as deserts, tepuis, open grasslands, and forests, resulting in a remarkable morphological variation that makes *Mandevilla* one of the most demanding and complex genera for taxonomists working on Neotropical Apocynaceae (Simões, Kinoshita and Endress, 2007). The currently accepted description in *Mandevilla* was defined in 1933 by Woodson (Woodson, 1933), who recognized 108 species that were divided into two subgenus: *Mandevilla* subgenus *Mandevilla* (= Eumandevilla) and *Mandevilla* subgenus *Exothostemon* (G.Don). He also proposed an inquisitive classification within the subgenus *Mandevilla* with five different sections. The greatest innovation in the Woodson classification was the creation of synonymy between the genera *Mandevilla* and *Dipladenia*, which differ in their flowers and leaves and, in

particular, in their ability to grow vines; *Mandevilla* is a longer vine, while *Dipladenia* is bushier and has smaller leaves (Oder, Lannes and Viruel, 2016).

Mandevilla sanderi (Hemsl.) Woodson (Woodson, 1933)is a plant originating from Brazil which, due to its decorative aspect, its profuse and extensive flowering, its persistent and glossy foliage, its tolerance to limited water availability and its resistance to many plant pests, is increasingly growing in ornamental market. It is characterized by voluble, woody stems, persistent foliage with decussate phyllotaxy, pink, infundibuliform corollas, and axillary racemose-type inflorescences.

In the current market, the name *Mandevilla* brings together genotypes that differ greatly not only in their phenotypic properties but also in their conditions of use, reflecting the western history of the introduction of *Mandevilla.* The first species introduced at the end of the 19th century were imported from their natural habitat in Brazil, added to nursery collections, and marketed without a license. These genotypes are commonly referred to as "native species". The most popular are *Mandevilla sanderi* (Hemsl) Woodson "Rosea", *Mandevilla × amabilis* (Beck & Backhf) "Alice du Pont", and *Mandevilla boliviensis* (Hooker F.) Woodson. Later introductions were hybrids (*Mandevilla hybrida*), which were developed in planned breeding programs and were subject to royalties (e.g., the Sundaville® and Diamantina® collections). The third type of plant material is somatic ("sport") mutations of cultivars or hybrids. Spontaneous mutations are common in the *Mandevilla* genome and can lead to genotypes that differ only in one commercial target trait, such as flower color (Oder, Lannes and Viruel, 2016).

Molecular markers

Molecular markers are of great importance in biological systems studies, including plants. They are used as crucial tools for plant identification and plant improvement (Cordeiro *et al.*, 2001). They are most applicable for the identification of plants and the relationships investigation between them. This is proven to be important in plant breeding, policing intellectual property rights, and forensic applications (J De Riek, 2009). The use of genetic markers has advanced quickly with early strategies dependent on phenotyping by DNA-based strategies for expanding refinement. The rapid development of technology has provided simple methods to assess the genetic variations in the genome. Early breeding strategies, that were not based on the analysis of DNA, were supplanted by genomic

approaches thanks to the development of the innovative technology of Polymerase Chain Reaction (PCR). The invention of PCR has incredibly expanded the feasibility of high-throughput marker screening. Early PCR-based techniques relied on restriction enzymes and universal primers due to the lack of sequence information for many species. This evolved, in the last decades, thanks to the widespread adoption of genomes sequencing and assembly, genomics information availability, technology improvement and the development of locus-specific molecular markers, such as the widely used microsatellite, or simple sequence repeat (SSR) markers, that are PCR-based, codominant and highly polymorphic (Henry, 2012).

SSR molecular markers and SSR genotyping

Microsatellites are nucleotide sequences, placed in both coding and non-coding genomic regions, consisting of tandemly repeated DNA motif of 1-6 nucleotides. They are inherited co-dominantly, highly variable in the number of motif's repetition, widely spread in the genome, and multiallelic. They are also referred to by different terms such as Simple Sequence Repeats (SSRs), or Short Tandem Repeats (STRs). In plants, SSRs are abundant and ubiquitous, and they are preferentially selected as extragenic, but associated to phenotype-related genes. Various types of SSRs like dimeric, trimeric, and tetrameric repeats are used as markers for molecular genetic studies (Senan *et al.*, 2014). The analysis involving SSRs requires knowledge about the DNA sequence of either side of the repeat region for the design of the specific primers needed for the PCR amplification (Ablett, Hill and Henry, 2006). The SSR markers' polymorphisms have been in the past identified through the use of polyacrylamide gel electrophoresis (PAGE), but nowadays, with the technological improvement of the last two decades, gels are replaced by the capillary electrophoresis systems, which allow the automated data capture and the analysis of a large number of samples reducing time and money costs (Rossetto, McNally and Henry, 2002). Analysis of a large number of samples is possible by effectively combining loci with different lengths and primers and by labelling them with different coloured fluorophores. Due to the high rates of mutations and polymorphisms, SSRs are widely used for genetic analysis, such as DNA genotyping for marker-assisted selection and breeding (MAS and MAB) purposes (Yang *et al.*, 2019), DUS testing, breeders' rights protection and genetic traceability.

SSR-based DNA genotyping is a procedure that involves the use of locus-specific primers to amplify SSR regions. The resulting amplicons are visualized using either gel electrophoresis or capillary electrophoresis. SSRs have been one of the most favoured molecular markers for plant genotyping in the last 20 years due to their high levels of polymorphism, wide distribution across most plant genomes, and ease of use. For these reasons they will continue to be a useful tool in many species for years to come (Mason, 2015).

Mandevilla is one of the most understudied ornamental genus, information about its DNA content is lacking and the C value range for known Apocynaceae species is large (from 0.31 to 2.45 pg, according to the Kew Gardens database) ((Oder, Lannes and Viruel, 2016). The fundamental objective of this project was to genetically characterise the experimental lines of *Mandevilla* using SSR markers. Alongside, another important objective was the identification of the parental species relationships among the analysed lines of the core collection.

When developing an essay based on SSRs, progressive loss of putative markers is a part of the set development. This could be due to any of the following reasons:

- Presence of the chimeric clones that hamper the generation of clear genomic sequences
- Inability to design efficient or specific primers in the flanking regions of the microsatellite motif
- Poor quality of the amplification obtained with the primers designed.

Species differentiation in the genus *Mandevilla* and DNA barcoding

DNA barcoding can be defined as a process of identification of species based on the molecular diversity of short DNA segments of either nuclear or non-nuclear (chloroplast and mitochondria) genomes (Vijayan and Tsou, 2010). The process of DNA barcoding consists in the sequencing of barcodes (specific extra-nuclear genomic regions) that are expected to be conserved within the same genus and/or species. The universally used barcodes in plants are the maturase K (*matK*) and the rubisco large sub-unit (*rbc*L) genes, the *psb*A-*trn*H interspace and the *trn*L intron. All these barcoding regions are placed in the chloroplast genome. Moreover, the nuclear Internal Transcribed Spacer (ITS1) is also used (Li *et al.*, 2015). In particular, *rbc*L is a gene present in the chloroplast genome,

which is suggested as one of the best potential candidate plant barcodes based on the straightforward recovery of the gene sequence, the large amount of easily accessible data, and good discriminatory power. It is widely used in phylogenetic investigations with over 50,000 sequences available in Genbank (Li *et al.*, 2015) and there are very few *rbc*L records on BOLD (Meiklejohn KA *et al.*, 2019). The major advantage of using *rbc*L is that it is simple to amplify, sequence and align in most of the plants.

In addition to these plastidial genes, ITS1, the internal transcribed spacer of nuclear ribosomal DNA, became one of the most commonly used DNA markers in phylogenetic and DNA barcoding analyses, and it has been recommended as a core plant DNA barcode (Cheng *et al.*, 2016). The greater discriminatory power of ITS1 over plastid regions at low taxonomic levels has been widely studied leading to it also being suggested as a plant barcode ((Stoeckle, 2003) (Kress *et al.*, 2005) (Sass *et al.*, 2007)).

Genetic certification of plant materials has become a fundamental requirement in the ornamental industry due to the increase in the competitiveness. DNA barcoding represents to be the unique and effective tool for the unambiguous determination of nature of ornamental species (A Giovino., et al 2020).

Aim of the study

In today's ornamental market, the term *Mandevilla* consists of genotypes that are very different, not only in their morphological traits (phenotype) but also in their genetic conditions. The continuous introduction of new varieties in the market and the interest of breeders in protecting the intellectual property on the varieties they develop need an accurate genotyping method. Due to their high polymorphism, co-dominance and reproducibility, SSRs are considered to be suitable molecular markers for genotyping analyses. The application of molecular markers in the field of ornamental plants is delayed compared to that of agricultural crops or model species due to several reasons. First, the economic importance of individual crops is relatively low, which limits both public funding and industrial support for applied research projects. Second, many ornamentals have complex genomes (large size, high heterozygosity, polyploidy, interspecific origin, etc…), which makes the genetic analyses difficult. Third, the breeding of ornamental species is different from that performed in agricultural crops, being mostly vegetatively propagated, and thus resulting in a reduced need for advanced molecular breeding techniques.

As a result, most ornamental plants still lack the technological resources to meet the needs of the breeding industry. In recent years, the exponential increase in information about plant genomes and the rapid pace of technological development have created new resources for genetic research in ornamental plants, and molecular markers can be developed much more easily and cheaply than they were in the past. As a result, molecular markers such as SSR, but also SNP, are now becoming available in the most important ornamental plant species such as rose, carnation, gerbera, lilium, tulip or chrysanthemum, but little or no studies have been done on most ornamental species and mapped molecular markers are not yet available (Oder, Lannes and Viruel, 2016).

This said, the project had two major aims as following,

The first part of study was the genetic characterization of a *Mandevilla* core collection composed of 55 different samples of as many breeding or pre-commercial lines by means of SSR markers. For this task, it was necessary to identify a suitable set of genetic markers able to generate informative and representative data. Data coming from the

microsatellite analysis were used to calculate the genetic similarity among samples and to perform a genetic structure analysis of the core collection.

The second part then, consisted in the implementation of DNA barcoding approach for further identification of the samples origin in term of species composition due to the possibility of interspecies crosses among the genus *Mandevilla*. For this purpose, two barcoding regions, both nuclear and plastidial (ITS1 and *rbc*L, respectively) were adopted to molecularly identify the samples' taxa.

Materials and methods

Plant materials and genomic DNA extraction

A total of 55 samples of fresh young leaves from the lines belonging to *Mandevilla* genus was provided by a private company. Upon arrival, each sample was labelled and carefully stored in polyethylene bags at -20 ° C in a freezer until further use. The genomic DNA (gDNA) was extracted using CTAB (cetyltrimethylammonium bromide) method (Doyle, J.J., 2019) with a few modifications. After the extraction, quality and quantity of gDNA has been evaluated using a NanoDrop 2000c UV-Vis spectrophotometer (Thermo Scientific, Pittsburgh, PA). DNA samples integrity was estimated by gel electrophoresis on 1% agarose/1× TAE gels containing 1× Sybr® Safe DNA gel stain (Life Technologies, Carlsbad, CA). The extracted gDNA samples were then diluted to a 20 ng/µl concentration for further analyses steps.

Primer design and PCR testing

In the research work, we designed and analysed 104 primer pairs which were able to amplify as many as Simple Sequence Repeat (SSR) loci in *Mandevilla*. From a prior whole genome sequencing, carried out on one of the samples using Illumina® technology to assemble the draft genome of this species, the presence of microsatellite regions was investigated by means of MIcroSAtellite identification tool (MISA) software (Beier *et al.*, 2017). The primer pairs for the selected microsatellites were designed using Primer3 software (Untergasser *et al.*, 2012).

The designed primers were then tested on a reduced number of samples (2001, 2009, 2010, 2022, 2013, 2049, 2031, and 2055) following the phenotypic information given by the private company that provided the core collection. After the testing phase, efficient markers were arranged into multiplexes to be used in the further genotyping analyses. To reduce the time and costs of the molecular essay, four multiplexes were organized and PCRs were set and optimized for multiple loci amplification. Multiplex PCRs were carried out grouping markers with primers having similar annealing temperature and diverse predicted amplicon size. Moreover, the minimum tendency on dimers formation was considered and it was evaluated by using the "Multiple Primer Analyzer" software (Thermo Fisher Scientific, Inc., Waltham, MA, United States). The PCR amplification

reactions were carried out using a modified version of the M13-tailed primer method firstly reported by Schuelke (Schuelke, 2000) and lately described.

Multiplex PCRs were accomplished by using forward SSR primers anchored with four different tails named PAN1, PAN2, PAN3, and M13 that are complementary to as many fluorescinated primers dyed with different fluorophores (VIC, NED, PET, and 6-FAM, respectively). The amplification protocol was performed using a 9600 thermocycler (Applied Biosystem, Foster City, CA, USA) and 96-well plates following the tested optimal conditions (Table 1). PCR products were then visualized by 2% agarose/1× TAE gel containing 1× Sybr® Safe DNA staining. The fluorescent-labelled PCR products were capillary electrophoresed on an ABI 3730 DNA Analyzer (Applied Biosystems) by adopting the GeneScan 500 LIZ as molecular weight internal standard. Finally, the size of each peak was manually determined by using Peak Scanner software 1.0 (Applied Biosystems).

SSR Genotyping

The amplifications of the 55 samples of the core collection were carried out adopting the following conditions: Platinum Master-Mix 2× (Applied Biosystems, Carlsbad, CA, USA), GC Enhancer 10× (Applied Biosystems, Carlsbad, CA, USA), 0.25 μM tailed forward primer, 0.75 μM reverse primer, 0.5 μM fluorophore-labelled oligonucleotides, 40 ng of gDNA, and H_2O to a final volume of 20 μL. Table 1 indicates the PCR conditions.

Table 1. The PCR conditions used for the amplification indicating the step, time, temperature and number of cycles.

Step	Time (min:sec)	Temperature °C	Number of cycles
Initial denaturation	05:00	95	1
Denaturation	00:30	95	
Annealing	00:45	54-50*	5
Elongation	00:45	72	
Denaturation	00:30	95	
Annealing	00:45	51	35
Elongation	00:45	72	
Final Elongation	10:00	72	1

*Touch-down method – The temperature decreases by 1°C at every cycle, starting temperature is depending on the multiplex.

SSR Data analysis

PopGene software package v. 1.32 (Yeh, 1997) was used for the statistical analysis of all the SSR marker loci. For each SSR locus and overall SSR markers, the observed number of alleles per locus (n_e), Nei's expected homozygosity (H_e; (Nei, 1973)), Levene's observed homozygosity (H_o; (Levene, 1949)), and the average homozygosity (H_a; Nei, 1978) were computed. The Shannon's information index (I) described by Lewontin (Lewontin, 1974) was used to estimate the phenotypic diversity of the allele profiles.

The genetic similarity (GS), calculated in all the pairwise comparisons using the Rohlf's genetic similarity coefficient, also known as Simple Matching coefficient, and a dendrogram based on the unweighted pair-group with arithmetic mean (UPGMA) method were obtained using the NTSYS software package v. 2.21c (Rohlf, 1998). Moreover, a Principal Coordinate Analysis (PCoA) was also developed, and samples were labelled with different colours according to those used in the UPGMA dendrogram to highlight the identified clusters.

The genetic structure of core collection in exam was analysed by using a Bayesian clustering algorithm implemented in STRUCTURE v2.2 software (Pritchard, Stephens and Donnelly, 2000). Using the admixture model with independent allele frequencies, ten replicate simulations were conducted for each value of K, with the number of founding groups ranging from 1 to 30, and a burn-in period of 2×10^5 and final run of 10^6 Markov chain Monte Carlo (MCMC) iterations. The method described by Evanno et al, (Evanno, Regnaut and Goudet, 2005) was used to evaluate the most probable estimation of K using the implemented algorithm of STRUCTURE Harvester web-software (Earl and vonHoldt, 2012). Results were then analysed, and a bar chart was created using an Excel spreadsheet with vertical bars for each identified group, based on the most probable number of K, labelled with different colours.

Geneious Prime software was then used to visualize the sequencing resulting chromatograms and the quality of the sequences was improved by trimming the low-quality sections in the 5′ and 3′ positions. Moreover, the resulting ITS1 chromatograms were also analysed with the "Heterozygote" plugin version 2.0.0 (Biomatters) to automatically identify the heterotic positions and then manually checked. Alignments of the resulting sequences were carried out based on the barcoding region for all the

samples. The resulting multiple alignments were used for the construction of a neighbour-joining tree using the Juke–Cantor algorithm, and polymorphic sites were used to create a LOGO graph.

DNA Barcoding through Sanger sequencing for species determination

DNA barcoding analysis to determine the maternal species of the 55 samples in exam was carried out and this was accomplished by using one plastidial DNA barcodes, namely the ribulose-1,5-bisphosphate carboxylase-oxygenase ribonuclease large subunit (*rbc*L). Being *rbcL* inherited by the maternal parent only, is not suitable for interspecific crosses determination. Thus, a nuclear region, namely the Internal Transcribed Space 1 (ITS1), was also considered along with the plastidial gene to verify the paternal species. The primer couples used for both regions are reported in Table 2

Table 2. List of primers used for each chloroplast (cpDNA) and nuclear (nuDNA) marker with their nucleotide sequence, and reference source.

Marker	Primer name	Primer Sequence (5′-3′)	Ta (°C) *	References
*rbc*L gene (cpDNA)	rbcL_F	GCAGCATTYCGAGTAASTCCYCA	55	(Nicolè, S et al., 2011)
	rbcL_R	GAAACGYTCTCTCCAWCGCATAAA	55	
ITS1 (nuDNA)	ITS5	GGAAGTAAAAGTCGTAACAAGG	55	(White, T.J et al., 2003)
	ITS2	GCTGCGTTCTTCATCGATGC	55	

* Ta: primers' annealing temperature.

DNA amplifications of all the samples were performed in a total volume of 20 µL of the reaction mixture. This included 10 µL of Platinum Master-Mix 2×with 1 µL of GC enhancer, 1 µL of DNA (50 ng/µL), 2 µL of each 10 µM diluted primer (Table 2) and sterile water to reach the final volume.

Amplifications were carried out using 9600 thermocycler (Applied Biosystem, Foster City, CA, USA) using the following thermal conditions: 95°C for 5 min, 40 cycles of 95°C for the 30s, 54°C for 45s, and then 72°C for 1 min, then 72 °C for 10 mins of final elongation. The obtained products were evaluated electrophoresis in 2% agarose/1× TAE gels containing 1 × SYBR Safe DNA Gel Stain (Life Technologies), and then purified with ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher) and forward sequenced on an ABI 3730XL Genetic Analyzer (Applied Biosystems).

Results

SSR marker analysis

The genome draft of one sample named "2001" was previously produced through Illumina® sequencing in order to identify thousands of suitable SSR regions to be used for genotyping analyses. The assembled contigs were screened for microsatellite regions and 5,408 potential SSRs, distributed over 4,806 contigs, were identified. 104 SSR markers, encompassing a dinucleotide or a trinucleotide repeat motif and showing a repeat motif length greater than 25 nucleotides, were chosen for further validation steps. Few failed the amplification of discrete PCR products whilst 23 (22%) were found to be efficient and used for the genetic characterization of the *Mandevilla* core collection. They were selected as they showed: i) good amplification success rates, ii) electropherogram peaks of high intensity and easily scorable, iii) marked attitude to be amplified in multiplex reactions, and iv) high polymorphism information content (PIC) coefficients. The selected SSR primer pairs were arranged into four PCR multiplexes based on the marker size, melting temperature and attached anchor (Table 3).

Table 3. List of the selected primer pairs. Marker name, expected fragment sizes, forward and reverse sequences along with the melting temperatures, anchor and the multiplex are reported.

Marker Name	Marker Size	Primer F	Primer R	Primer F Tm (°C)	Primer R Tm (°C)	Anchor	Multiplex
SSR_30	233	CAACACCTATACCTCACACC	GAGTTTGTAGTCTCCAACCTT	55.2	55.2	M13	1
SSR_47	157	TGCTGCATTAATCACCTACA	GGCAGAAGAAGATTTGTCCA	54.7	55.6	M13	1
SSR_02	327	ATTTGTTTGCAACCTCCATG	CCGCAACTCAAACTCAAATT	55.0	55.1	PAN1	1
SSR_34	202	TCTCCAATTAGCAGTACAAGG	TTAGACAGGGAGAGAGACAG	55.0	55.0	PAN1	1
SSR_41	171	GCCTCTCAAGTCATTAGGTG	AGGGTACTAAGGATGGTCTAA	55.6	54.5	PAN2	1
SSR_12	153	TGAAATAAAGGGTTAGGGCA	TCACTAATCCAGACAATCACA	54.2	54.3	PAN3	1
SSR_28	415	GAGATCAATGAGGATGGGAC	CACTTACAGTTTCAGGTCCT	55.0	54.6	M13	2
SSR_48	187	CCGTGCCTCCTATGATTTAC	CTGACCATGCAATTACTCCT	55.7	55.1	M13	2
SSR_50	138	CATTCAGCACACAGTTCTTC	AGTCATCGTTGTGAAATGGA	54.9	55.0	M13	2
SSR_40	300	TGGACGAACTTGATACTACG	TGTTGAAAATCCCAGTCCAA	54.7	55.1	PAN1	2
SSR_15	223	TGAGGCACATACCATAGAGA	AATTTCTTGTCGTGGGCTAT	55.0	55.0	PAN3	2
SSR_59	163	ATTCAGCACACAGTTCTTCT	GTCTATGACGGAGAGAAACC	54.9	55.0	M13	3
SSR_60	336	CCCTAGAGACCTTTTCATCC	CGAGTGTCTTCAAGCCATTA	54.5	55.5	M13	3
SSR_67	115	TACTAATTCGTCGTTTGGCT	CTTTTAGGTCATTTGGTCCAA	54.8	54.0	PAN1	3
SSR_55	185	TTTCAGCATAGGTTCGACAA	AAAGCCTGAATCTCCTCTTG	55.0	55.0	PAN2	3
SSR_76	272	AATAAACAGCCCAGTCTCAA	TTCTTCAATTTGCAGCCTTT	54.5	54.2	PAN3	3
SSR_89	424	AAACTGGGACCATACACATC	TTGACGTAACTGTTTGACCA	55.0	55.2	PAN3	3
SSR_64	206	GGCACCTGTTAATATCAGTG	GATGGATGTAGAGGATGGTG	53.8	54.7	M13	4
SSR_74	253	GACGGATGCTCTTAATTCCT	GTGTACAGATCCCTACTTCC	54.9	54.3	M13	4
SSR_70	345	TATTGAGGTTTGGCTTTCGA	CATTAACACCCCTCTTGTCA	55.0	55.0	PAN1	4
SSR_80	421	CTTTGGATTTGAAAGCGGAA	CAAAGGTATGTCTCTGGGTC	54.8	55.2	PAN2	4
SSR_61	413	ACAAAGCTTCTCCATCTCAG	GGGTGACTTTCCTGCTAATT	55.1	55.3	PAN3	4
SSR_95	139	ATTTTCCGTGAATCCAGATCT	TGAGAAGGGGTTGTTGTTG	55.0	55.2	PAN3	4

SSR marker statistical analysis

Table 4 summarizes the number of detected alleles per locus (Na), the effective number of alleles (Ne), the observed homozygosity (Ho), expected homozygosity (He), and the polymorphism information content (PIC) along with the mean and standard deviation values resulting from the 23 selected loci analysed among the core collection's samples that presented no triple peaks after the electropherograms screening.

Table 4. Descriptive statistics of genetic diversity calculated across SSR markers. Sample size, observed and expected homozygosis, number of observed alleles (Na) and number of effective alleles (Ne) are reported for each SSR locus investigated. Shannon's index, observed and expected homozygosity was calculated too. The overall values and standard deviations are also reported for each parameter.

Marker Name	Sample size	na	ne	H_o	H_e	I	PIC
SSR_02	106	3	2.04	0.49	0.49	0.74	0.51
SSR_12	104	4	2.99	0.12	0.33	1.18	0.67
SSR_15	90	5	3.72	0.09	0.26	1.41	0.73
SSR_28	94	3	1.55	0.70	0.64	0.60	0.36
SSR_30	90	4	1.78	0.80	0.56	0.78	0.44
SSR_34	100	7	3.88	0.24	0.25	1.59	0.74
SSR_40	92	6	3.66	0.33	0.27	1.41	0.73
SSR_41	100	10	4.18	0.30	0.23	1.69	0.76
SSR_47	110	6	1.64	0.64	0.61	0.86	0.39
SSR_48	98	5	2.11	0.57	0.47	0.95	0.53
SSR_50	102	9	4.99	0.10	0.19	1.80	0.80
SSR_55	110	4	3.32	0.16	0.30	1.27	0.70
SSR_59	110	8	4.92	0.07	0.20	1.75	0.82
SSR_60	80	3	1.81	0.93	0.55	0.68	0.45
SSR_61	106	6	2.39	0.57	0.41	1.20	0.58
SSR_64	110	4	2.15	0.40	0.46	0.87	0.53
SSR_67	108	4	1.83	0.63	0.54	0.83	0.45
SSR_70	106	6	3.59	0.34	0.27	1.50	0.72
SSR_74	104	5	3.15	0.54	0.31	1.30	0.68
SSR_76	108	6	3.13	0.33	0.31	1.38	0.68
SSR_80	106	4	2.72	0.30	0.36	1.11	0.63
SSR_89	98	6	2.04	0.47	0.48	1.01	0.51
SSR_95	108	8	4.23	0.13	0.23	1.62	0.76
Mean	102	5.48	2.95	0.40	0.38	1.20	0.62
St. dev.		1.93	1.06	0.49	0.14	0.37	0.14

Among all the selected SSR primer pairs the sample size varied between 80 to 110 with a mean value of 102. In total 126 alleles were detected among the samples with several observed alleles per locus ranging between 3 (SSR_02, SSR_28 and SSR_60) and 10 (SSR_41) and an average of 5.48 among all. Effective number of alleles per locus ranged

between 1.55 (SSR_28) and 4.99 (SSR_50) with a mean value of 2.95. Each microsatellite locus showed highly variable levels of observed homozygosis (H_o), ranging from 0.07 (SSR_59) to 0.93 (SSR_60) with an average observed homozygosis of 0.40. Moreover, the expected homozygosis (H_e) showed a mean value equal to 0.38, ranging from 0.19 (SSR_50) to 0.64 (SSR_28). Shannon's index was also calculated, and it was considered to determine the discriminative ability of each marker locus among the different samples. With a minimum value of 0.60 (SSR_28) and a maximum one of 1.80 (SSR_50) with an average value equal to 1.20, which was considered relatively high. The mean PIC value was observed to be 0.62 with the values ranging between 0.36 (SSR_28) to 0.82 (SSR_59).

Genetic Similarity analysis

The analysis of the average genetic similarity (GS), which was calculated in all pair-wise comparisons among all the analysed samples, is reported in Figure 1. Considering the genetic similarity matrix, the lowest value scored was 52% in the comparison of samples 2009 and 2016. On the other hand, samples 2053 and 2051, 2001, 2002 and 2008, and 2033 and 2034 were found to have the highest possible genetic similarity of 100%. Considering other high GS values detected, sample 2038 was found to have a percentage of 99.2% with samples 2030 and 2033, and the same was also true for samples 2001 and 2003, and 2028 compared to 2029.In the matrix the percentage values and the deep green colour indicated the highest percentage of GS, while the higher intensity of red showed the lower percentage of GS (Figure 1).

The UPGMA dendrogram (Figure 2) grouped the 55 samples into two main branches, one including samples labelled in yellow and one including those labelled with the other set of colours. This second main branch, was sub-categorised into four distinct and separate sub-clusters coloured in grey, purple, green and blue. The samples belonging to the grey (80%), purple (80%) and green (75%) clusters were found to have high GS within them, while those of blue cluster (2015 and 2016 had higher GS (72%) than 2047 (70%)) had low GS values (70%), both among themselves and with the other clusters' samples. From the dendrogram, samples of clusters coloured in grey, purple and green were most similar within clusters, with values over 76%, while samples in the yellow branch were on average more dissimilar, with an internal mean GS value of around 70%. Samples 2051 and 2053, 2001, 2002 and 2008 from purple coloured branch and samples 2033 and 2034 from green coloured branch had 100% genetic similarity, respectively.

Genetic relationships among samples were further studied using a Principal Coordinate Analysis (Figure 3). The PCoA plot revealed a clustering scheme, similar to that obtained from the UPGMA dendrogram. From the principal coordinate analysis (PCoA), the percentage of variation explained by dimension 1 is 47.0%, whereas that of dimension 2 is 22.1%. Overall, the PCoA dimensions contributed for the 69.1% of the total variation.

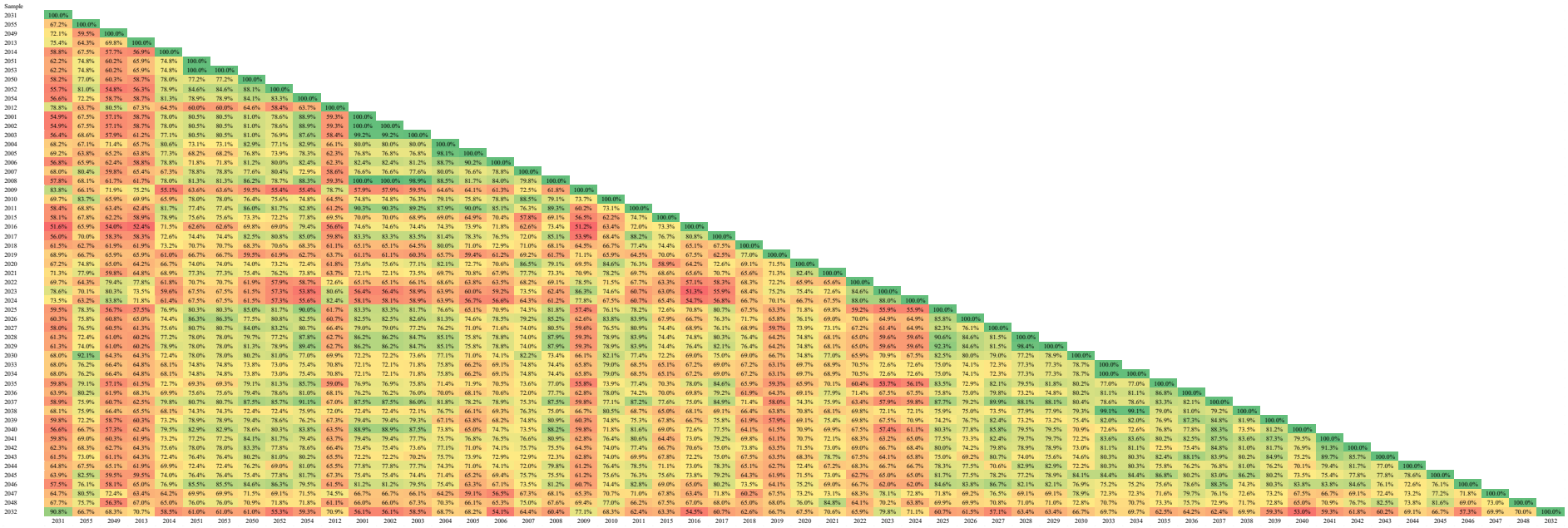

Figure 1. General overview of the genetic similarity matrix (55x55 samples) calculated in all pair-wise combinations and based on the analysis of 23 SSRs. Highest GS values are highlighted in green, while lowest in red varying through yellow

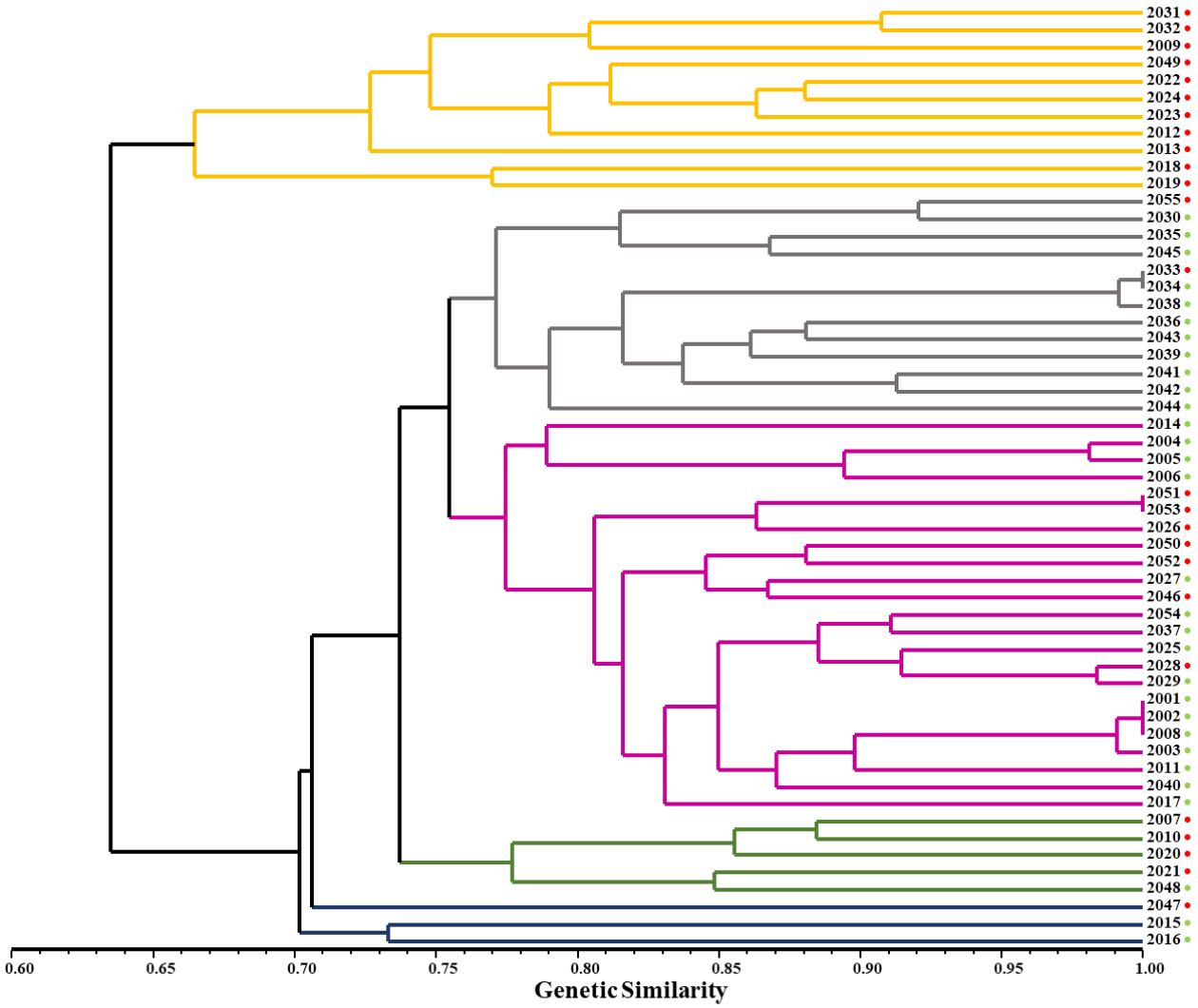

Figure 2. UPGMA dendrogram of the *Mandevilla* core collection based on the genetic similarity matrix. Different coloured branches represent different identified clusters of samples. The red and green dots on the extreme right indicate the clusters to which samples belong based on the Neighbour Joining (NJ) tree based on the polymorphic sites among ITS1 nuclear region and *rbc*L chloroplast barcoding regions.

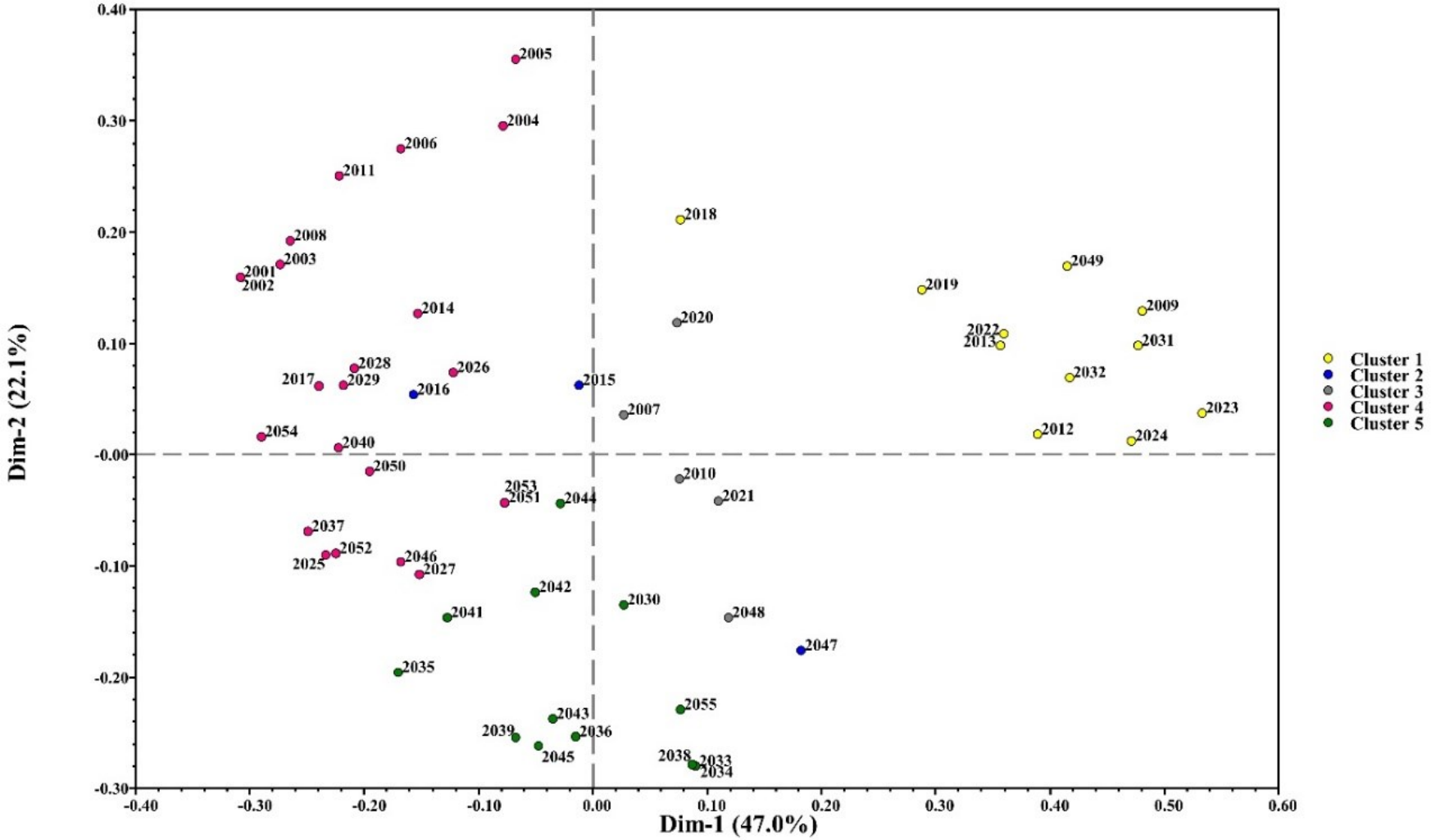

Figure 3. PCoA diagram derived from the genetic similarity estimates. Different colours are used to label samples according to the clusters identified in the UPGMA dendrogram.

Genetic structure analysis

Regarding the investigation of the genetic structure of the core collection, to ascertain the likely number of genetic groups (K) within the collection, STRUCTURE software was utilized. The population structure was estimated with the most likely number of K equal to 26 and, to a lesser extent, 2. Each sample's resulting memberships were plotted as a vertical histogram divided into K = 2 and K = 26 (Figure 4).

At the first grouping level, the 55 samples of the core collection were separated into two main ancestral clusters (K = 2). The samples formed three distinct membership patterns: one composed of 11 samples having high membership values to one ancestral group (orange cluster), 20 samples showing the same results for the other one (blue cluster), and 24 samples being admixed. On the other level of separation (K = 26), samples were categorised into various clusters, with a low number of individuals presenting high membership to one specific ancestral group, while the majority of them showed high admixture levels.

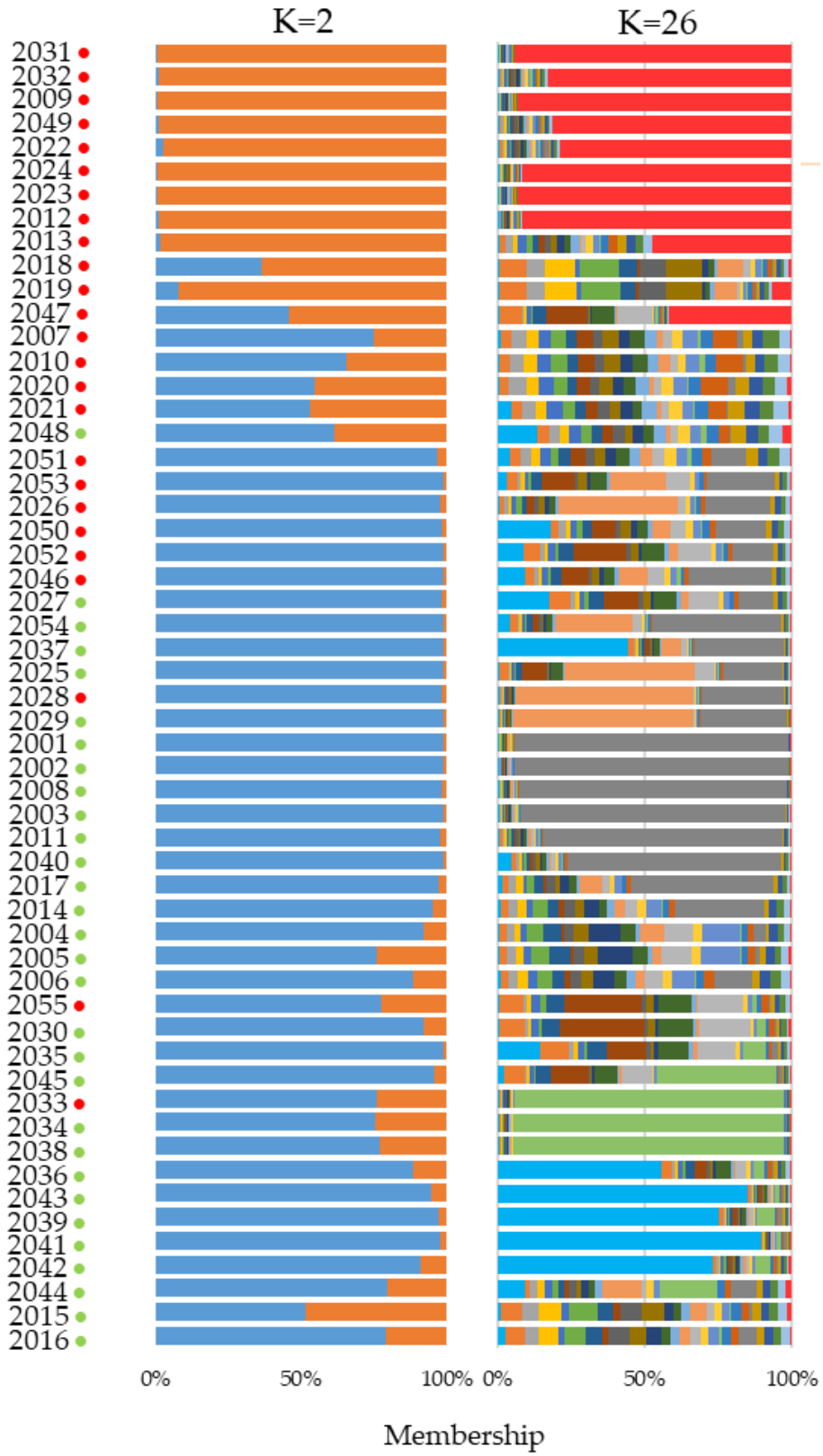

Figure 4. Population structure of the core collection as estimated by STRUCTURE software where all the samples are represented by vertical histograms portioned into K = 2 and K = 26, distinct segments corresponding to the estimated membership. The red and green dots on the left indicate the indicate the clusters to which samples belong based on the Neighbour Joining (NJ) tree based on the polymorphic sites among ITS1 nuclear region and *rbc*L chloroplast barcoding regions.

DNA barcoding analysis

The analysis of DNA barcoding sequences used in the molecular taxonomy was conducted to investigate the phylogenetic origin of the germplasm collection and to possibly correlate the results with the findings obtained through SSR analysis. The length of the sequences obtained were 632 bp (*rbc*L) and 306 bp (ITS1). Most of the aligned sites were conserved, but few SNPs were observed.

In particular, *rbc*L was polymorphic only at position 575 (A>C), splitting the core collection into two groups of 30 and 25 accessions. Both *rbc*L sequences were used to interrogate the BOLD database, and the best match was *Mandevilla sanderi* (100% query coverage), with a slight difference in terms of identity values (100%, first sequence and 99.84% second sequence). ITS1, despite being shorter than *rbc*L, was polymorphic in 15 positions (15/306). All ITS1 sequences, searched in GenBank through blastN, had *Mandevilla atroviolacea* as the best match with full query coverage (always 100%) and an identity score ranging from 98.52% to 95.45%. The results obtained from the neighbour-joining tree-based on a multiple alignment of the two concatenated sequences revealed that samples were clustered in two main subgroups, coloured in red and green in Figure 5. The red-coloured cluster was further sub-categorised into other branches with different bootstrap values of 95.5, 85.3, 53.9 and 71.9, respectively (Figure 5).

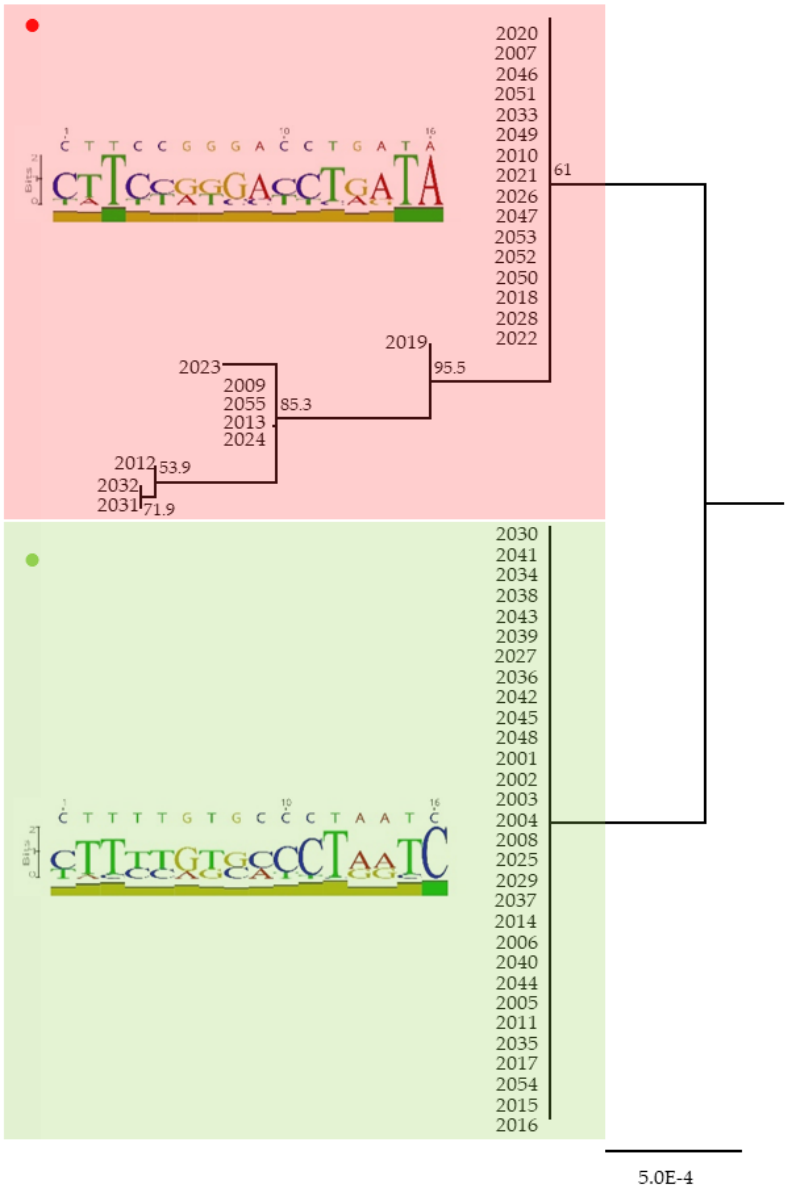

Figure 5. Neighbour Joining tree based on the polymorphic sites among ITS1 nuclear region and *rbc*L chloroplast barcoding regions. Bootstrap values are reported next to each node of the NJ-tree.

Discussion

The main goal of this study was the genetic characterisation of a *Mandevilla* spp. core collection, constituted of experimental and pre-commercial lines using SSRs molecular markers. A total number of 104 SSR markers were tested during this study, 23 (22%) of which were found to be suitable for the DNA genotyping analysis of core collection. Various statistics such as the genetic similarity, the UPGMA dendrogram, the PCoA diagram, the genetic structure, and the DNA barcoding were used to molecularly characterize the samples of the core collection in exam that are further discussed in this session.

The SSR data have been used to access the genetic similarity of the analysed samples. Shannon's index, expected homozygosity (H_e), considered as equal to the polymorphism information content (PIC) following the interpretation of Borin et al. (Borin *et al.*, 2021) is one of the most common parameters for describing the genetic informativeness of a molecular marker locus. H_e is often taken into consideration for the diversity analysis while PIC is more often used for the linkage studies (Kalinowski, Taper and Marshall, 2007).

The values of both the parameters, H_e and PIC are related to the number and frequency of the detected alleles and reflect the compendium of different factors: the ability of markers to detect variability, the characteristics of the genotype set analysed (relatedness of the samples), and the reproductive system of the species.

The selection of most suitable SSR markers was based on their efficiency, adaptability to be used in multiplex reactions and ease of scoring. As a result, the 23 markers reported in Table 3, were used to characterize the core collection. The resulting SSR markers dataset was used to compute the Polymorphism Index Content (PIC), which according to Botstein et al. (Botstein et al., 1980), defines a molecular marker locus as highly informative for values of PIC>0.50, those with 0,25<PIC<0,50 as informative, and PIC>0.25 as poorly informative. The high PIC values observed for the 23 selected SSR markers reflected in their considerable discriminative capability in the genotyping analyses of the samples considered in this study. These results were confirmed by the subsequent analysis carried out.

The records found in this research work proved that some SSRs had high levels of homozygosity (0.93), indicating a repeated self-pollination leading to high homozygous state while those with low levels of homozygosity (0.07) must be a result of hybridization due crossing between highly homozygous parents or clonal propagation (maintaining the heterozygous state in the samples at each generation). Overall, the main values of observed homozygosity of SSRs obtained from the samples of core collection are closer to those found in other species that are, similar to *Mandevilla*, which are self-fertile but can also be cross-pollinated, such as lychee (Ho = 0.57) or cherimoya (Ho= 0.40)) (Escribano, Viruel and Hormaza, 2008)(Viruel and Hormaza, 2004), and lower than those found in predominantly allogamous species, such as avocado (Ho= 0.83) (Gross-German & Viruel, 2013).

The PCoA and the UPGMA dendrogram grouped samples according to their genetic similarity calculated in all pair-wise comparisons within the core collection. From our findings, several common genetic backgrounds were hypothesized for samples showing high GS values among them. The results of the dendrogram indicate that branches in grey, purple, and green colours are the most closely related ones, because of a shared high genetic similarity, whilst the blue cluster is poorly genetically linked to the first three. Finally, the branch in yellow is placed separately from the others, indicating a poor genetic similarity with the rest of the germplasm and, probably, also a different origin. Moreover, some degree of genetic dissimilarity is also present within the yellow cluster, as demonstrated by the long arms observed in this branch.

The results from the PCoA were comparable to those obtained from the UPGMA dendrogram, as samples were placed according to the clustering previously discussed. The green, purple and grey labelled samples were grouped close to each other, but still distinguishable in the PCoA and were separated from the others, mainly based on the first dimension of the plot (left quadrants of Figure 3). The yellow-coloured samples, on the other hand, were clustered far from the other clusters (right quadrant of the plot) indicating their strong distinctiveness. The samples belonging to respective colours were placed close to each other forming the clusters and those with low GS were placed far from each other. For instance, 2001 and 2052 belonged to the purple-coloured samples forming a cluster but they were placed on the extreme ends of the clusters indicating that the GS between these samples was less when compared with GS between 2001 and 2002

which were placed right after each other. Most of the samples formed defined clusters in the PCoA analysis except for those coloured in blue.

Regarding the genetic structure reconstruction of the core collection under study, STRUCTURE software analysis was used to categorize the 55 samples of *Mandevilla* into the most probable number of ancestral groups. The most probable results showed two (K = 2) or twenty-six (K = 26) putative ancestors, with samples membership percentages reflecting, in general, the clustering observed in the genetic similarity-based analysis. It is worth pointing out how, for K = 2, all individuals grouping together in the UPGMA tree under the yellow cluster showed an individual membership to the same founding group (orange) higher than 92% (except for 2018, 63%). This was partially confirmed also analysing the same collection for K = 26 (e.g., see founding group in red), confirming the groupings identified for K = 2 and strengthening the common origin of some accessions. K = 26 also revealed, in comparison to K = 2, a higher number of admixed accessions. This is in agreement with the breeding strategies carried out in several ornamental species, where interspecific crosses are accomplished to maximize the phenotypic variability.

The DNA barcoding analysis, based on Sanger sequencing, of one cytoplasmic region (*rbc*L) and one nuclear region (ITS1) was carried out on all the 55 samples of the core collection. The two variants of the *rbc*L sequence found within the entire collection differed only for one SNP and, from the BOLD alignment, they both perfectly matched (100% query coverage, 99.84%-100% identity value) with *Mandevilla sanderi* that can be therefore considered the maternal lineage of all the samples. This would also explain why the SSR panel worked in all the interspecific hybrids of the germplasm: probably the 23 SSR chosen for the analysis were (casually) selected on the portion of the maternal genome common to all the accessions. Moreover, it is worth emphasizing that one of the two variants of the *rbc*L locus was shared by all the samples belonging to the SSR-based yellow cluster (Figure 2), in line with what emerged from the genetic structure analysis (Figure 4). This would support the hypothesis that these samples have a very distinct origin from the others. As a side note, it should be acknowledged that a single barcoding sequence is probably not enough to undoubtedly ascertain the maternal origin of the germplasm.

ITS1, despite being shorter than *rbc*L, was polymorphic in 15 positions. All the ITS1 sequences, searched in GenBank through blastN, had *Mandevilla atroviolacea* as the best

match with full query coverage (always 100%) but the identity score ranged from 98.52% to 95.45%. This low degree of genetic identity must be interpreted considering the very limited number of ITS1 sequences available in GenBank for the *Mandevilla* genus. It is extremely likely that the species/hybrids to which the germplasm samples belong are not represented in GenBank and thus cannot be properly identified.

Finally, from the NJ-tree produced through a multiple alignment of the two concatenated regions, two major branches were observed. Few agreements were detected between the DNA barcoding analysis and the SSR-based analyses: samples of the yellow cluster (UPGMA and PCoA) also clustered together in the same NJ-based branch (red coloured cluster).

Conclusion

Regardless of the wide appropriateness of the SSR markers in plant genomics, their improvement stays to be of significant importance in understudied crops, especially ornamental ones. In this research work, the set of 23 SSR markers adopted, proved to be suitable for the genetic characterization of a *Mandevilla* spp. core collection. The microsatellite markers development in this research has given a helpful insight concerning the *Mandevilla* genetic information relevant to the management and breeding of *Mandevilla* genotypes. The SSR marker-based genotyping technique presented in this study and firstly applied to *Mandevilla* has been successfully used and demonstrated a consistent discriminative ability of this kind of molecular marker in assessing the genetic relationships between different lines of this genus. In general, this approach is of crucial importance in estimating the levels of genetic diversity, not only to avoid its loss or promote its conservation, but also to allow its potential utilization in several fields. Future studies will certainly be needed to improve the molecular knowledge about this genus and its species, especially for the identification of interspecific crosses between them, but, nevertheless, this study's results give a first insight on the genetic complexity of this important ornamental plant and provide a suitable method for future genotyping analyses in *Mandevilla*.

References

A Giovino, F.M.A.P. (2020) "The technique of Plant DNA Barcoding: potential application in floriculture," *International Journal of Cytology, Cytosystematics and Cytogenetics*, 73-n. 2, pp. 29–40, Available at: http://digital.casalini.it/4710875.

Ablett, G., Hill, H. and Henry, R.J. (2006) "Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing," *Molecular Breeding*, 17(3), pp. 281–289, doi:https://doi.org/10.1007/s11032-006-6262-3.

Beier, S. *et al.* (2017) "MISA-web: a web server for microsatellite prediction," *Bioinformatics*, 33(16), pp. 2583–2585, doi:https://doi.org/10.1093/bioinformatics/btx198.

Borin, M. *et al.* (2021) "Developing and testing molecular markers in cannabis sativa (Hemp) for their use in variety and dioecy assessments," *Plants*, 10(10), doi:10.3390/plants10102174.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics, 32(3), p.314, doi: doi.org/0002-9297/80/3203-0013$01.58.

Cheng, T. *et al.* (2016) "Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity," *Molecular ecology resources*, 16(1), pp. 138–149, doi: https://doi.org/10.1111/1755-0998.12438

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M. and Henry, R.J., 2001. Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum. Plant science, 160(6), pp.1115-1123, doi: https://doi.org/10.1016/S0168-9452(01)00365-X

Doyle, J.J., (1990), Isolation of plant DNA from fresh tissue. Focus, 12, pp.13-15, Google scholar

Earl, D.A. and vonHoldt, B.M. (2012) "STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method," *Conservation Genetics Resources*, 4(2), pp. 359–361, doi:10.1007/s12686-011-9548-7.

Escribano, P., Viruel, M.A. and Hormaza, J.I. (2008) "Development of 52 new polymorphic SSR markers from cherimoya (Annona cherimola Mill.): Transferability to related taxa and selection of a reduced set for DNA fingerprinting and diversity studies," *Molecular Ecology Resources*, 8(2), pp. 317–321, doi:10.1111/j.1471-8286.2007.01941.x.

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

Gross-German, E. and Viruel, M.A. (2013) "Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: Genetic diversity, population

structure, and identification of race-specific markers in a group of cultivated genotypes," *Tree Genetics and Genomes*, 9(2), pp. 539–555, doi:10.1007/s11295-012-0577-5.

Henry, R.J. (2012) "*Molecular markers in plants*," John Wiley & Sons, Oxford, UK, ISBN: 978-0-470-95951-0

J De Riek, T.D. (2009) "Present use of molecular markers in ornamental breeding," *In XXIII International EUCARPIA Symposium, Section Ornamentals, Colourful Breeding and Genetics-Part II*, Part II(855), pp. 77–84, doi:10.17660/ActaHortic.2010.855.9.

Kalinowski, S.T., Taper, M.L. and Marshall, T.C. (2007) "Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment," *Molecular Ecology*, 16(5), pp. 1099–1106, doi:10.1111/j.1365-294X.2007.03089.x.

Kress, W.J. *et al.* (2005) "Use of DNA barcodes to identify flowering plants," PNAS, 102 (23), doi: https://doi.org/10.1073/pnas.0503123102

Levene, Howard, (1949) "On a Matching Problem Arising in Genetics." The Annals of Mathematical Statistics, vol. 20, no. 1, pp. 91–94, doi: http://www.jstor.org/stable/2236806.

Lewontin, R.C. (1974) *The genetic basis of evolutionary change*. Columbia University Press New York, ISBN: 0-231-08318-1(pbk.)

Li, X. *et al.* (2015) "Plant DNA barcoding: from gene to genome," *Biological Reviews*, 90(1), pp. 157–166, doi: https://doi.org/10.1111/brv.12104

Mason A.S. (2015) SSR Genotyping. In: Batley J. (eds) Plant Genotyping. Methods in Molecular Biology (Methods and Protocols), vol 1245. Humana Press, New York, NY, https://doi.org/10.1007/978-1-4939-1966-6_6.

Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. PLOS ONE 14(6): e0217084, doi: https://doi.org/10.1371/journal.pone.0217084

Nei, M. (1973) "Analysis of gene diversity in subdivided populations," *Proceedings of the National Academy of Sciences*, 70(12), pp. 3321–3323, doi: https://doi.org/10.1073/pnas.70.12.3321

Oder, A., Lannes, R. and Viruel, M.A. (2016) "A set of 20 new SSR markers developed and evaluated in mandevilla Lindl," *Molecules*, 21(10), doi:10.3390/molecules21101316.

Jonathan K Pritchard, Matthew Stephens, Peter Donnelly, (2000), Inference of Population Structure Using Multilocus Genotype Data, Genetics, Volume 155, Issue 2, Pages 945–959, doi: https://doi.org/10.1093/genetics/155.2.945

Rohlf, F.J. (1998) "NTSYSpc numerical taxonomy and multivariate analysis system version 2.0 user guide."

Rossetto, M., McNally, J. and Henry, R.J. (2002) "Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae," *Theoretical and Applied Genetics*, 104(1), pp. 61–66, doi: https://doi.org/10.1007/s001220200007

Sass, C. *et al.* (2007) "DNA barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of Cycads," *PLoS ONE*, 2(11), doi:10.1371/journal.pone.0001154.

Schuelke, M. (2000) "An economic method for the fluorescent labeling of PCR fragments," *Nature biotechnology*, 18(2), pp. 233–234, doi: https://doi.org/10.1038/72708

Senan S, Kizhakayil D, Sasikumar B, Sheeja TE (2014) Methods for development of microsatellite markers: an overview. Notulae Scientia Biologicae 6(1):1–13, doi: https://doi.org/10.15835/nsb619199

Simões, A.O., Kinoshita, L.S. and Endress, M.E. (2007) "New combinations in Mandevilla Lindley (Apocynaceae)," *Novon*, 17(1), pp. 87–90, doi:10.3417/1055-3177(2007)17[87:NCIMLA]2.0.CO;2.

Stoeckle, M. (2003) "Taxonomy, DNA, and the bar code of life," *BioScience*, 53(9), pp. 796–797, doi: https://doi.org/10.1641/0006-3568(2003)053[0796:TDATBC]2.0.CO;2

Untergasser, A. *et al.* (2012) "Primer3—new capabilities and interfaces," *Nucleic Acids Research*, 40(15), pp. e115–e115, doi:10.1093/nar/gks596.

Vijayan, K., and C. H. Tsou, (2010), "DNA Barcoding in Plants: Taxonomy in a New Perspective." Current Science, vol. 99, no. 11, Temporary Publisher, pp. 1530–41, doi: http://www.jstor.org/stable/24069450.

Viruel, M.A. and Hormaza, J.I. (2004) "Development, characterization and variability analysis of microsatellites in lychee (Litchi chinensis Sonn., Sapindaceae)," *Theoretical and Applied Genetics*, 108(5), pp. 896–902, doi:10.1007/s00122-003-1497-4.

Woodson, Robert E, (1933) "Studies in the Apocynaceae. IV. The American Genera of Echitoideae." Annals of the Missouri Botanical Garden, vol. 20, no. 4, Missouri Botanical Garden Press, pp. 605–790, doi: https://doi.org/10.2307/2394198.

Yang, J. *et al.* (2019) "Target SSR-seq: A novel SSR genotyping technology associate with perfect SSRS in genetic analysis of cucumber varieties," *Frontiers in Plant Science*, 10. doi:10.3389/fpls.2019.00531.

Yeh, F.C. (1997) "Population genetic analysis of co-dominant and dominant marker and quantitative traits," *Belgian J. Bot.*, 130, pp. 129–157, Available at: https://ci.nii.ac.jp/naid/10019157093/en/.