Microsatellite analysis to rationalize grape germplasm in India and development of a molecular database

Anuradha Upadhyay*, Lalitkumar B. Aher[#], Manisha P. Shinde,
Kavita Y. Mundankar, Anuj Datre and G. S. Karibasappa

National Research Centre for Grapes, Manjri Farm Post, Solapur Road, Pune 412 307, Maharashtra, India

Abstract

Three hundred and seventeen grape accessions from the National Active Grape Germplasm Site in India were analysed with 25 microsatellite markers. A total of 411 alleles were detected, of which 42% were rare alleles. Unique alleles for 56 genotypes were also identified. The analysis of microsatellite data identified 63 duplicate accessions and only 254 accessions were unique genotypes. Several cases of misnomers, synonymy and homonymy were identified. Parental genotype for a few clonal selections was ascertained. Population structure analysis grouped 254 unique genotypes into four major clusters. The analysis also revealed the presence of admixtures with only 79% of pure ancestry. A core collection comprising 80 genotypes was identified, which represented all the alleles and genetic diversity. A user-friendly and interactive computer application software was developed for storage and the retrieval of molecular data. A molecular database for the 254 genotypes was created. This analysis will help in the rationalization and better management of germplasm. Information on genetic diversity and population structure will form the basis for varietal improvement programmes.

Keywords: core collection; genetic diversity; grape; microsatellite; molecular database

Introduction

Grapevine is a heterozygous, perennial fruit crop. The germplasm of such plants is maintained as an *ex situ* field collection. Considering the high input cost required for the management of field germplasm, accurate identification is essential to reduce redundancy. Ampelography, the traditional method of cultivar identification in grapevine, is time consuming, requires an experienced worker and often fails to differentiate closely related genotypes. DNA-based molecular markers circumvent the limitations of the ampelography technique and thus

play an important role in grape germplasm management. Among the different molecular marker systems, microsatellite markers have been proved to be the most suitable for the characterization and management of grape germplasm (Dangl *et al.*, 2001). The use of microsatellite markers for the analysis of grape germplasm has been reported by several researchers (Ibáñez *et al.*, 2009; Cipriani *et al.*, 2010; Santana *et al.*, 2010; Laucou *et al.*, 2011).

In India, the reference to grape cultivation is found in ancient text. Several species of *Vitis*, namely *V. lanata*, *V. parviflora* and *V. palmata*, are found in the Himalayan regions of the country. Native *V. vinifera* varieties, such as Rangspey, Choultu White, Choultu Black and Kannai locally belonging to *V. lanata*, are still grown in Himachal Pradesh. In modern history, commercial *V. vinifera* is believed to be introduced by Persians in 1300 AD and its cultivation spread and flourished during the

*E-mail: anu_upadhyay@yahoo.com;
#Present address: National Institute of Abiotic Stress Management, Baramati 413 115, Pune, Maharashtra, India.

Mogul empire (1526–1758 AD). Grape cultivation declined after the fall of the Mogul empire. Systematic research was started in the early 20th century. Presently, grape cultivation covers 111,000 ha with a production of 1.24 million MT mainly for table purpose. Grape research in India was scattered, and grape germplasm has been collected from different parts of the country and world by researchers at different institutes. After the establishment of the National Research Centre for Grapes in 1997, attempts were made to consolidate grape germplasm in India through its collection from different national and international research institutes, growers' field, wineries, private nurseries and exploration in the Himalayan regions of the country. Passport data and ampelography analysis has suggested the occurrence of several synonyms, probable duplicates and misnomers in the germplasm. However, these observations need to be confirmed with more precise and reliable identification provided by microsatellite markers.

Molecular analysis of a large number of genotypes results in extensive data. Consequently, the management of such data becomes complex. A computer-based information system and database offers the ease of handling of large datasets and their subsequent utilization. Several databases have been developed for the microsatellite

data of grape germplasm in different countries (Lefort and Roubelakis-Angelakis, 2000; Dettweiler and Eibach, 2003; Vouillamoz *et al.*, 2009; Veloso *et al.*, 2010), and their practical application has been demonstrated (Nunez *et al.*, 2004). In this paper, we report the microsatellite analysis of 317 grape accessions from germplasm maintained at the National Research Centre for Grapes, for genetic diversity and structure analysis, the rationalization of germplasm and the creation of a molecular database.

Materials and methods

Plant material and DNA extraction

A set of 317 grape accessions including rootstocks, hybrids and commercial varieties was selected from the National Active Germplasm Site for Grape located at the National Research Centre for Grapes, Pune, India. The details of these accessions are given in Table S1 (available online). DNA was extracted from fresh, young, tender leaves using the DNeasy® Plant Kit (Qiagen, Valencia, CA, USA). DNA was quantified spectrophotometrically by measuring absorbance at 260 nm and quality was assessed on 1% Tris−acetate−EDTA agarose gel electrophoresis.

Table 1. Details of genetic diversity and polymorphic information detected by the 25 microsatellite markers in the total collection (TC) and the core collection (CC)

Locus	No. of alleles	No. of rare alleles (count ≤ 5)	PIC		Heterozygosity		Allelic diversity	
			TC	CC	TC	CC	TC	CC
VMC4f8	14	3	0.790	0.793	0.672	0.692	0.808	0.807
VMC7b1	15	5	0.796	0.847	0.275	0.229	0.819	0.862
VMC7f2	8	1	0.747	0.729	0.701	0.658	0.776	0.753
VMC7g3	17	10	0.562	0.709	0.484	0.532	0.580	0.723
VMC8b5	12	5	0.788	0.815	0.781	0.781	0.813	0.835
VrZAG62	13	1	0.806	0.854	0.778	0.790	0.826	0.867
VrZAG79	24	10	0.876	0.891	0.798	0.818	0.886	0.899
VVIB01	23	15	0.753	0.796	0.702	0.785	0.778	0.815
VVIB23	23	16	0.724	0.829	0.723	0.735	0.744	0.839
VVIB63	12	7	0.597	0.653	0.540	0.532	0.636	0.681
VVII52	14	7	0.780	0.812	0.589	0.662	0.805	0.831
VVIN33	25	10	0.831	0.903	0.704	0.756	0.841	0.909
VVIP17A	21	9	0.839	0.864	0.738	0.681	0.853	0.875
VVIP22	19	8	0.841	0.874	0.671	0.551	0.855	0.884
VVIP25B	23	14	0.719	0.813	0.665	0.584	0.747	0.827
VVIP31	12	1	0.844	0.856	0.764	0.667	0.858	0.869
VVIP77	16	4	0.827	0.879	0.720	0.687	0.842	0.889
VVIS63	17	9	0.613	0.734	0.286	0.345	0.647	0.749
VVIV15B	17	5	0.767	0.839	0.806	0.800	0.783	0.852
VVIV16	7	1	0.490	0.562	0.456	0.457	0.564	0.624
VVIV33	9	4	0.577	0.750	0.110	0.163	0.638	0.780
VVIV37	14	5	0.825	0.858	0.420	0.523	0.844	0.872
VVMD7	18	3	0.847	0.870	0.785	0.740	0.861	0.881
VVMD21	19	11	0.728	0.842	0.631	0.714	0.750	0.855
VVMD27	19	10	0.820	0.881	0.792	0.821	0.839	0.890
Total/average	411	174	0.751	0.81	0.624	0.628	0.776	0.827

Microsatellite analysis

Twenty-five microsatellite markers belonging to the VVMD, VMC, VrZAG and VVI sets were used for the analysis. The PCR and microsatellite analysis was as described previously in Upadhyay et al. (2010a, b). In brief, the PCR amplification reaction mixture (10 μl) contained 10 ng DNA, 0.66 μM labelled (6-Carboxy fluorescein (6-FAM), VIC or NED) forward primer, 0.66 μM reverse primer, 100 μM of each dNTP, 3.0 mM MgCl$_2$ and 1.0 U Taq polymerase (Bangalore Genei Pvt Ltd, Bangaluru, India). The PCR was performed either on a PTC 200 gradient thermal cycler (MJ Research, Waltham, MA, USA) or a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). The temperature profile consisted of the following steps: 10 min at 94°C followed by 35 cycles of 1 min at 94°C, 1 min at 55°C and 1 min at 72°C. PCR products were diluted 50 times, and 1 μl (for FAM- and VIC-labelled) or 2 μl (for NED-labelled) of the diluted mix were added to a mixture of 10 μl HI-DI formamide and 0.10 μl of GeneScan 500 ROX internal size standard. The mix was denatured at 94°C for 5 min and analysed on an ABI 3130 genetic analyser using a 36 cm capillary filled with POP7 polymer. GeneMapper version 4.0 was used to determine the peak size using the local Southern method and allele call. Allele-call data were manually checked and any misallele call was corrected. The data were scored as allele size. Genotypes with one peak for a given locus were recorded as homozygous.

Microsatellite data analysis

Allelic data were first analysed with software CERVUS version 3.0 (Kalinowski et al., 2007) to identify duplicate accessions. A mismatch at a maximum of two loci was allowed for assigning any accession as duplicate. The allelic data of unique genotypes were used to estimate the allele number per locus, allelic diversity (Div = 1-), heterozygosity (Het = 1-), polymorphism information content (PIC =), overall genotype frequencies and allele frequencies, and fitness to Hardy–Weinberg equilibrium. The data were also subjected to principal component analysis. This analysis was performed using the software JMP Genomics 6.0 (SAS Institute Inc., Cary, NC, USA, 1989–2007). Cluster analysis based on the neighbour-joining method was carried out using the software DARwin 4.0 (Perrier and Jacquemoud-Collet, 2006), and a tree was constructed.

The allelic data were also subjected to population structure analysis using the software STRUCTURE version 2.3.4 (Pritchard et al., 2000), a model-based Bayesian clustering method. Twenty iterations were run for K values ranging from 1 to 14 with a burn-in length of 10,000 generations and 10,000 Markov chain Monte Carlo replications. The analyses were run with correlated allele frequencies (Falush et al., 2003). Other parameters were set to their default setting. All the accessions were considered as having an unknown origin. To set the number of subpopulations, a graph of K value versus Ln P(D) as well as the ΔK method described by Evanno et al. (2005) were used. The accessions with a membership probability of ≥80% were considered as of pure ancestry.

Microsatellite data were also used to identify core collection representing the entire genetic diversity by means of the software PowerCore version 1.0 (www.genebank. go.kr/eng/PowerCore/powercore.jsp), which selects the entries of core sets by the advanced M (maximization) strategy implemented through a modified heuristic algorithm (Kim et al., 2007). The data of core collection accessions were subjected to statistical analysis, as above, to assess their level of genetic diversity and their distribution pattern.

Development of an information system for storing molecular data

A user-friendly information storage and retrieval system was designed to store the microsatellite data of grape germplasm. It uses SQL server 2008 for data storage in the database and a platform-independent front-end application that can be implemented on any PENTIUM computer with Windows XP/2000/Win 7. The front-end application was developed with the Microsoft dot net framework using Microsoft Visual Studio 2005.

Results

Identification of duplicates and misnomers

The microsatellite data analysis of the 317 grape accessions with CERVUS identified 63 accessions with an identical allele profile and 254 accessions as unique genotypes. The microsatellite data of the 254 unique genotypes at 15 loci are given in Table S2 (available online). Several accessions with an identical profile were known synonyms and clonal selections/somatic variants; however, the identification of several supposedly different accessions as redundant was interesting. For example, accessions maintained as Arki, Ceffer, E/2/7, Pierce, Goathe, Malagha and Azabella were found to have an identical genotype at 23 loci. Other sets of identical accessions were Chaouch, Convergent Large Red and Alden; Buckland Sweet and Dakh, E/5/20 and E/7/22; HY/17/54/4/17 and HY/23/14/23; PS/II/11/4 and PS/III/11/1; Gold and Fakri; and E/12/3 and

E/12/7. However, these results need corroboration with detailed ampelography and molecular analysis before terming them as duplicates in the germplasm.

We also identified several misnomers in the germplasm. The accession maintained as Merlot was long suspected to be a misnomer, as its ampelographic and performance attributes did not match with the documented attributes. The allele profile of this accession did not match with the allele profile of Merlot obtained from ENTAV, France at 13/20 loci. Instead, this accession had an allele profile similar to Cinsaut at 14/20 loci and shared at least one allele at another four loci, suggesting that this accession may be a sibling or a progeny of Cinsaut. Similarly, another accession maintained as Zinfandel in the germplasm was found to have an allele profile identical to that of Syrah at all the loci, suggesting it to be either Syrah or its clone. Bangalore Blue was developed at the Department of Horticulture, Bangaluru, Karnataka, and has been reported to be a selection from Isabella. However, its allele profile did not match with that of the accession maintained as Isabella in the germplasm. Instead, its profile matched with other two accessions maintained as Azabella and Pierce. Pierce is a known synonym of Isabella Royal, a mutant of Isabella (Vitis International Variety Catalogue), whereas the origin of Azabella is unknown. These results suggested that Azabella might be a misnomer for Isabella and the true identity of the accession maintained as Isabella needs to be established. The variety Cheema Sahebi has been reported to be derived from Pandhari Sahebi (Patil and Rane, 1974). However, data indicated that it is a selection of Spin Sahebi, a partial male sterile variety, as the allele profile of Cheema Sahebi was an exact match of Spin Sahebi. Two cases of homonymous accessions were detected. The accessions Gulabi1 and 2 showed a different allelic profile at several loci. Similarly, two accessions maintained as Cardinal differed in their allele profile.

Microsatellite marker information

The 25 microsatellite loci were consistent with good PCR efficiency and peak structure. Thus, the selected markers were useful for generating data for characterization and the creation of a database. These microsatellite markers detected 411 alleles in the 254 unique genotypes with an average of 16.5 alleles per marker (Table 1). The number of alleles detected for each marker varied between 7 (VVIV16) and 25 (VVIN33). The polymorphic information content ranged between 0.490 (VVIV16) and 0.876 (VrZAG79) with an average of 0.751. The marker heterozygosity varied between 0.110 (VVIV33) and 0.806 (VVIV15B) with an average of 0.624. The average allele diversity was 0.776 ranging from 0.564 (VVIV16)

to 0.886 (VrZAG79). All the loci were in Hardy–Weinberg equilibrium, thus confirming their suitability for identity analysis. Rare alleles (alleles with a count of ≤ 5) were observed for all the microsatellite markers. A total of 174 rare alleles were detected, which was 42% of the total number of alleles. The number of rare alleles detected by each marker ranged from 1 (VrZAG62, VVIP31 and VVIV16) to 16 (VVIB23). Unique alleles were detected in 56 accessions.

Population structure

The STRUCTURE analysis was performed to infer the number of gene pools in the germplasm. In the K versus $\mathrm{Ln}\,P(D)$ graph (Fig. S1(A), available online), a break in the slope was observed at $K = 4$ and $K = 8$ and a plateau was reached at $K = 11$, giving an inconclusive result on the exact number of subpopulations. However, when subjected to ΔK analysis, a sharp peak was obtained at $K = 4$ (Fig. S1(B), available online), indicating four major clusters. The mean F_{st} value for a cluster ranged from 0.083 (cluster 2) to 0.617 (cluster 3) with an average of 0.262 over all the clusters (Table S3, available online). The average intra-cluster distance ranged between 0.428 (cluster 3) and 0.846 (cluster 2). The data presented in Table 2 indicated that allele frequency divergence among the clusters varied between 0.061 (clusters 1 and 4) and 0.258 (clusters 1 and 3). The STRUCTURE analysis also revealed admixtures in the populations (Fig. S2, available online). Only 79% (201/254) of the accessions were of pure ancestry (membership probability $\geq 80\%$), whereas the remaining 53 accessions (21%) had ancestry from two or more groups (membership probability $\geq 10\%$ from each group). Among the clusters, cluster 4 had a maximum number of accessions with mixed ancestry and cluster 3 had most accessions with pure ancestry. The distribution of the accessions among the four clusters is given in Table S1 (available online). The accessions in each cluster were further subjected to STRUCTURE analysis to estimate subgroups for each cluster. The estimated number of subgroups for clusters 1, 2, 3 and 4 was 5, 2, 1 and 4, respectively. However, the structure was weak because the ΔK value was low and ranged from 3.7 to 9.0. Cluster analysis using the neighbour-joining

Table 2. Inter-cluster allelic frequency divergence among the clusters

	2	3	4
1	0.143	0.258	0.061
2	–	0.250	0.082
3		–	0.199

method and principal coordinate analysis were also conducted to complement these results. In the cluster analysis using the neighbour-joining method, several groups/subgroups were obtained, although it was difficult to decipher the exact level of grouping (Fig. S3, available online). In the principal component analysis, the first three axes accounted for only 24% of the total variation.

Identification of core collection

PowerCore version 1.0, which uses the advanced 'M' strategy with a modified heuristic algorithm, was adopted to analyse the microsatellite data for the identification of a core subset in the germplasm. The analysis identified a core subset of 80 accessions (Table S1, available online), which constituted 31.5% of the total accession. These core accessions included some commercially grown varieties, rootstocks and interspecies hybrids in addition to other genotypes. The microsatellite data of the 80 accessions were reanalysed to assess the extent of genetic diversity captured by the core subset. This analysis indicated that 100% genetic diversity of the 254 accessions was represented by the core subset (Table 1). Among the core accessions, the primer heterozygosity varied between 0.159 (VVIV33) and 0.823 (VVMD27) with an average of 0.630, whereas allelic diversity was 0.827 across all loci and varied between 0.623 (VVIV16) and 0.908 (VVIN33). The 80 core accessions were distributed in all the groups and subgroups of the unrooted neighbour -joining tree (Fig. S3, available online), thus confirming the representation of the entire genetic diversity.

Information system for storing and management of molecular data and creation of a database for microsatellite data

The information system was developed to meet the following requirements: (1) to store data generated from different molecular marker techniques such as microsatellite and amplified fragment length polymorphism (AFLP); (2) to allow the import of allelic microsatellite data as well as binary data generated by AFLP markers; (3) to maintain masters for data on varieties and markers with details; (4) to support basic functions such as data entry, editing, viewing, search and data retrieval; (5) to generate ready-to-print reports based on different criteria; (6) to export data from different file formats; (6) to facilitate identification of a variety and calculate similarity percentage with other varieties in the database for a set of selected marker data; (7) to build a multi-user system with security

and access control; (8) to develop front-end software using dot net technology that runs on a Windows platform. This software is available on request.

The database consists of different tables. The details of genotypes including their location in the germplasm are stored in the table 'Variety'. The information on the marker class and the marker name is stored in the table 'Technique' and 'Primer', respectively. The table 'Primer' also contains the details of the PCR and amplification condition. The table 'Band' stores marker data such as allele size, whereas the table 'Login' is used to store different user IDs and allows the generation of password for access control. The logical model of the database is given in Fig. 1. The graphical user interface of the application menu is given in Fig. S4 (available online). The application main menu screen allows the user to access/navigate among the different functionalities provided by the program grouped under four menus, namely Data management, Operations, Reports and Help, each with a different submenu. The submenu 'Variety data' allows the user to add a new variety name and to edit or delete the variety already stored in the database. The molecular data for a variety cannot be added to the database if the variety name is not present in the database. A list of varieties already present in the database is provided on the screen and allows the selection of the variety name for modification and deletion. The submenu 'Primer data' allows to add a new marker name along with its details and to edit the information. The molecular data for a marker can be added only if the marker name is entered into the database. A detailed list of the marker data already present in the database is also provided. An important feature of the system is the comparison of imported data with the data of reference varieties. The standard data for the selected reference varieties for all the markers is first entered. A few of the reference varieties are always included in each set analysed on the automated system to check for run-to-run variation in allele size. Each Excel file to be imported into the database has the data for reference varieties in the first few rows. The system compares the data for reference varieties in the Excel file with that stored in the database. If some variation is detected, it adjusts the allele size of other varieties before importing and storing into the database. Under the 'Operation' menu, the submenu 'Identify variety' allows the user to compare the allelic data of an unknown material with the data of other varieties in the database, which generates a list of varieties with a matching profile. Similarly, the submenus 'Search' and 'Similarity percentage', respectively, allow the user to search the database for different criteria and to calculate the similarity between two genotypes. Different submenus under the menu 'Reports' are used to generate reports for variety, markers and allelic data
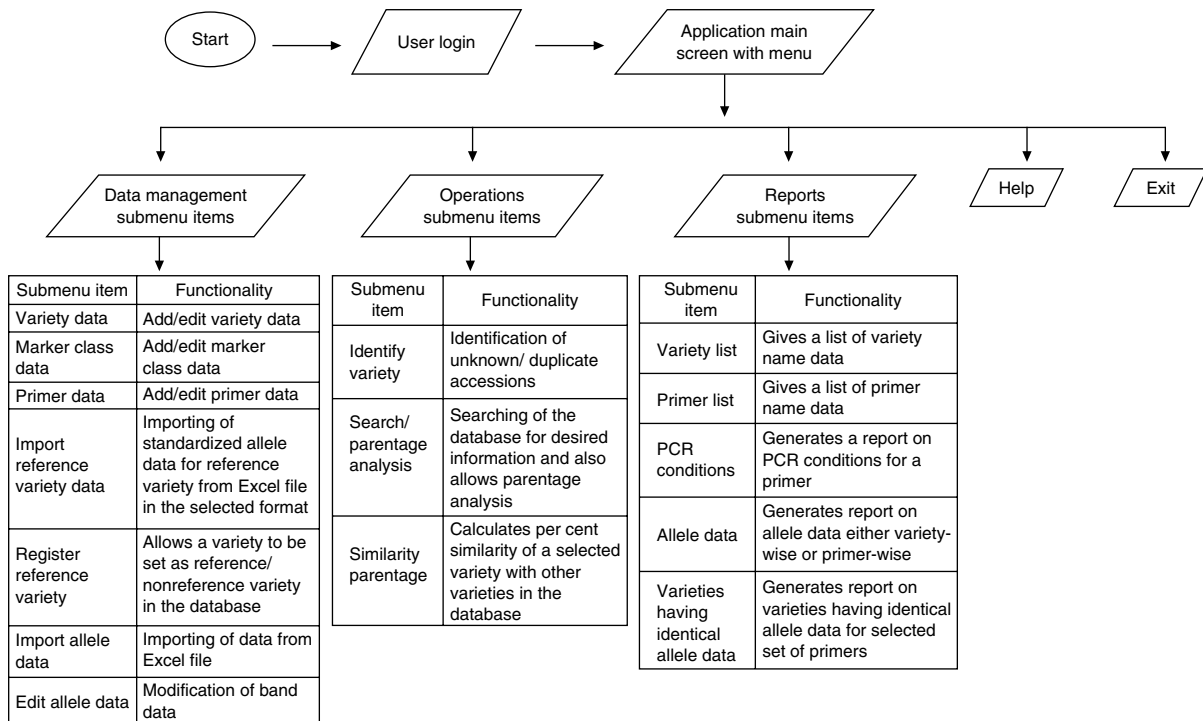
Fig. 1. Logical model of the database.

using different criteria. The menu 'Help' describes in detail the functioning of each user interface, thus assisting the user in working with the program. The creation of a new user ID and password management for a secure access to the data is done using SQL Server Management Studio 2008. The database is supported by SQL server 2008 and run on any PENTIUM computer with Windows XP/2000/Win7.

Discussion

Establishing the identity of accessions in the germplasm

As expected, several cases of misnomers, duplications and synonymy were detected in this study. This is due to the fact that collection comprises genotypes obtained from multiple sources that were maintained by several research workers. Mis-naming and mis-labelling due to change of hand and during transport might have resulted in some of the confusion. Local names given to the varieties also result in redundancy. Cipriani *et al.* (2010) allowed a mismatch at two loci for terming two genotypes as synonymous. With these criteria, the occurrence of many synonymies was found in this study. While some were known synonyms, several synonymies were due to the presence of clonal selections and/or

somatic variants. Microsatellite markers often fail to distinguish somatic variants and clones and thus can be used to confirm their genetic identity. Genetic identity of clonal selections from Kishmish Chernyi was confirmed recently by Upadhyay *et al.* (2011). In this study, genetic identity of Kishmish Rozavis White, a somatic variant of Kishmish Rozavis, identified at the National Research Centre for Grapes was confirmed. Though clonal selections showed an identical allele profile for the studied loci, they differ in bunch characters that are commercially important for table grapes. The clones of Kishmish Chernyi differ significantly in their bunch characters (Upadhyay *et al.*, 2011). The somatic variant of Kishmish Rozavis differs in berry colour, whereas the clones of Centennial Seedless differ in characters such as self-thinned bunches, naturally bold berries and early ripeness. Transposable elements have been found to be the major cause for such somatic polymorphism (Carrier *et al.*, 2012). Several researchers have reported the inability of microsatellite markers to differentiate clones, as well as limited success of microsatellite markers to detect variations among clones (Regner *et al.*, 2000; Vignani *et al.*, 2002; Moncada *et al.*, 2006; Moncada and Hinrichsen, 2007; Upadhyay *et al.*, 2010b). By and large, the detection of clonal variation remains unpredictable and the development of a robust marker system for unique identification of clones remains a challenge for researchers.

Detection of rare and unique alleles

A large number of rare alleles accounting for 42% of the total alleles were detected, several of them being unique for a genotype. A closer look at the data indicated that the majority of unique alleles were detected in rootstocks and genotypes belonging to different species, although some unique alleles for wine and table varieties were also detected. Previous studies have also reported the presence of distinct alleles in rootstocks (Sefc *et al.*, 1998; Crespan *et al.*, 2009). Sefc *et al.* (1998) reported that one-third of the alleles detected by ten loci occurred only in rootstock species, one-third in *V. vinifera* and only one-third were shared by two groups. Due to their interspecific nature, rootstocks have higher allelic diversity, which could also be one of the reasons for the detection of a relatively high number of alleles (16.4 alleles/primer). A high number of alleles (26 alleles/primer) have also been reported by Laucou *et al.* (2011) when analysing 4370 accessions including rootstocks and interspecies hybrids. In contrast, Cipriani *et al.* (2010) detected 274 alleles from 34 primers (eight alleles/primer) in 1004 accessions belonging to only *V. vinifera*. Lin and Walker (1998) reported the detection of 17.6 alleles/primer while analysing only 56 rootstocks with six primers. Another reason for higher polymorphism detected could be the fact that these primers were selected based on earlier reports on their high polymorphic nature.

Population structure analysis

The results from three different analyses to decipher the genetic relationship among the accessions indicate heterogeneity in the collection. The Bayesian method of analysis, which is found to be the most informative and useful method for inferring structure in the population, identified four clusters. Cluster 1 contained hybrids, their parents and accessions belonging to *V. vinifera*. All the rootstocks and interspecies hybrids grouped together in cluster 2. Within cluster 2, two subgroups were identified: one containing all the rootstocks and the other containing interspecies hybrids. Cluster 3 had a majority of the accessions belonging to *V. labrusca* species. Cluster 4 had the maximum number of accessions (108) belonging to *V. vinifera*, both wine and table type. Among the *V. vinifera* genotypes, the majority of the table and wine varieties formed a separate subcluster within clusters 2 and 4. Chenin Blanc and Sauvignon Blanc, identified as a sibling in a recent study (Myles *et al.*, 2011), were in the same cluster. Other wine varieties that share Traminer as the parent were in the same cluster. However genotypes collected from different regions and having different uses were distributed in different clusters. Cipriani *et al.* (2010) could not divide

745 grape cultivars into subpopulations when K values 1–10 were used. We could divide 254 unique genotypes into clusters using the *ad hoc* procedure of K value estimation developed by Evanno *et al.* (2005). This method was successfully used for estimating population structure in several heterogeneous and perennial plant species (Raina *et al.*, 2012; Du *et al.*, 2012). The Bayesian method of grouping also allowed the detection of admixtures and genotypes with pure ancestry. Since grouping is based on log-likelihood probabilities, this method is considered unbiased as against distance-based clustering (Lopez-Gartner *et al.*, 2009; Raina *et al.*, 2012). Data on allele frequency divergence will help in deciding parents for breeding, whereas information on population structure will be useful for selecting genotypes for association mapping studies.

Core collection

Core collection is a representative sample of the whole collection with minimum repetitiveness of genetic diversity of a crop and its relatives (Frankel, 1984; Frankel and Brown, 1984). Identification of a core collection is recommended for better management and easier access to the germplasm and its effective utilization. Core size should be at least 10% of the whole collection (Brown, 1989). Analysis of the microsatellite data of 254 unique genotypes identified 80 genotypes as core collection, which constituted 31% of the total genotypes. Several strategies have been used to identify core collections in the germplasm of different crop plants using molecular markers. While the H strategy proposed by Schoen and Brown (1993) samples accessions from groups based on their within-group genetic diversity to maximize the alleles in core collection, the M (maximization) strategy examines all the possible core collections and selects those that maximize the number of observed alleles at the marker loci (Schoen and Brown, 1993). The software PowerCore uses the modified M strategy to choose a core collection, and has been efficiently used in the identification of core collection in tea (Raina *et al.*, 2012), soybean (Cho *et al.*, 2008) and mungbean (Moe *et al.*, 2012). For grape, Le Cunff *et al.* (2008) and Cipriani *et al.* (2010) reported a representation of about 240 alleles in 48 and 30 genotypes, respectively. These two studies have analysed genotypes belonging to *V. vinifera* only. In our analysis, core germplasm collection constituted 31% of the total genotypes and captured all the alleles generated by 25 loci including rare alleles and unique alleles detected in rootstocks and other species of *Vitis*. Thus, the entire genetic diversity of the germplasm was represented by the core collection. Genetic diversity data of different markers for the core collection

and the total collection were comparable. Since these 80 genotypes provide a preliminary estimate of diversity in the germplasm, they can be used for assessing the polymorphism of new markers.

Molecular database for grape germplasm in India

Enormous data generated by molecular marker analysis require considerable efforts for proper data handling. An effective data storage and management system ensures minimum handling errors, avoids duplication of data and allows optimum utilization for varied purposes. The database structure is designed to accept data from different types of marker systems. Presently, the database contains microsatellite data of 254 unique genotypes. The stored information can be accessed easily and updated by importing the data from the Excel format. The data can be retrieved in different formats for specific requirements. Thus, the database will assist in ascertaining the true-to-typeness of planting material, avoid duplicates in the collection and enhance the utility of the data by better data storage, management and information retrieval.

Germplasm characterization is important for its conservation and efficient utilization. Assessment of genetic diversity is required for the selection of suitable parents for breeding programmes. Germplasm for a vegetatively propagated perennial crop such as grape is maintained as an active field collection, which requires immense labour, space and cost. Thus, it is prudent to minimize redundancy in collection by accurate characterization and identification of all the accessions. In this study, we could analyse the majority of grape germplasm maintained at the National Research Centre for Grapes in India. Several cases of misnomers, duplicates, synonymy and homonymy were detected. The analysis of population structure and the identification of core collection will be useful for devising breeding strategies. A molecular database will be useful not only for efficient storage and retrieval of data, but also for varietal identification and further germplasm enhancement and management.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262113000117

Acknowledgements

References

Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* 31: 818–824.

Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, Audeguin L, Boursiquot J-M and This P (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS One* 7: e32973.

Cho GT, Yoon MS, Lee J, Baek HJ, Kang JH, Kim TS and Paek NC (2008) Development of a core set of Korean soybean landraces (*Glycine max* (L.) Merr). *Journal of Crop Sciences and Biotechnology* 11: 157–162.

Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A and Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theoretical and Applied Genetics* 121: 1569–1585.

Crespan M, Meneghetti S and Cancellier S (2009) Identification and genetic relationship of the principal rootstocks cultivated in Italy. *American Journal of Enology and Viticulture* 60: 349–356.

Dangl GS, Mendum ML, Prins BH, Walker MA, Meredith CP and Simon CJ (2001) Simple sequence repeat analysis of a clonally propagated species: a tool for managing a grape germplasm collection. *Genome* 44: 432–438.

Dettweiler E and Eibach R (2003) The two *Vitis* databases as tools for germplasm management: *Vitis* international variety catalogue and European *Vitis* database. *Acta Horticulturae (ISHS)* 603: 505–509.

Du Q, Wang B, Wei Z, Zhang D and Li B (2012) Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *Journal of Heredity* 103: 853–862.

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: 2611–2620.

Falush D, Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Frankel OH (1984) Genetic perspective of germplasm conservation. In: Arber W, Illmensee K, Peacock WJ and Starlinger P (eds) *Genetic Manipulation: Impact on Man and Society*. Cambridge: Cambridge University Press, pp. 161–170.

Frankel OH and Brown AHD (1984) Current plant genetic resources – a critical appraisal. In: Chopra VL, Joshi BC, Sharma RP and Bansal HC (eds) *Genetics. New Frontiers Vol. IV*. New Delhi: Oxford & IBH Publishing Co., pp. 3–13.

Ibáñez J, Vélez MD, Andrés MT and Borrego J (2009) Molecular markers for establishing distinctness in vegetatively propagated crops: a case study in grapevine. *Theoretical and Applied Genetics* 119: 1213–1222.

Kalinowski ST, Taper ML and Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG and Park YJ (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing allele mining sets. *Bioinformatics* 23: 2155–2162.

Laucou V, Lacombe T, Dechesne F, *et al.*, (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theoretical and Applied Genetics* 122: 1233–1245.

Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon A-F, Boursiquot J-M and This P (2008) Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp sativa. *BMC Plant Biology* 8: 31.

Lefort F and Roubelakis-Angelakis KA (2000) The Greek *Vitis* database: a multimedia web-backed genetic database for germplasm management of *Vitis* resources in Greece. *Journal of Wine Research* 11: 233–242.

Lin H and Walker MA (1998) Identifying grape rootstocks with simple sequence repeat (SSR) DNA markers. *American Journal of Enology and Viticulture* 49: 403–407.

Lopez-Gartner G, Cortina H, McCouch SR and Moncada MDP (2009) Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fuorescent SSR markers. *Tree Genetics and Genomes* 5: 435–446.

Moe KT, Gwag J-G and Park Y-J (2012) Efficiency of PowerCore in core set development using amplified fragment length polymorphic markers in mungbean. *Plant Breeding* 131: 110–117.

Moncada X and Hinrichsen P (2007) Limited genetic diversity among clones of red wine cultivar 'Carmenère' as revealed by microsatellite and AFLP markers. *Vitis* 46: 174–180.

Moncada X, Pelsy F, Merdinoglu D and Hinrichsen P (2006) Genetic diversity and geographical dispersal in grapevine clones revealed by microsatellite markers. *Genome* 49: 1459–1472.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Jer-Ming Chia, Ware D, Bustamante CD and Buckler ES (2011) Genetic structure and domestication history of the grape. *Proceedings of National Academy of Science USA* 108: 3530–3535.

Nunez Y, Fresno J, Torres V, Ponz F and Gallego FJ (2004) Practical use of microsatellite markers to manage *Vitis vinifera* germplasm: molecular identification of grapevine samples collected blindly in D.O. "El Bierzo" (Spain). *Journal of Horticultural Science and Biotechnology* 79(3): 437–440.

Patil AV and Rane DA (1974) A short note on a mutation of Cheema Sahebi (syn Selection 7). *Vitis* 13: 21–214.

Perrier X, Jacquemoud-Collet JP. (2006) DARwin software. Available at http://darwin.cirad.fr/darwin.

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

Raina SN, Ahuja PS, Sharma RK, Das SC, Bhardwaj P, Negi R, Sharma V, Singh SS, Sud RK, Kaur R, Pandey V, Banik J, Razdan V, Sehgal D, Dar TUH, Kumar A, Bali S, Bhat V, Prasanna BM, Goel S, Negi MS, Vijayan P, Tripathi SB, Bera B, Hazarika M, Mandal AKA, Kumar RR, Vijayan D, Ramkumar S, Chowdhury BR and Mandi SS (2012) Genetic structure and diversity of India hybrid tea. *Genetic Resources and Crop Evolution* 59: 1527–1541.

Regner F, Wiedeck E and Stadlbauer A (2000) Differentiation and identification of White Riesling clones by genetic markers. *Vitis* 39: 103–107.

Santana JC, Heuertz M, Arranz C, Rubio JA, Martinez-Zapater JM and Hidalgo H (2010) Genetic structure, origins, and relationships of grapevine cultivars from the Castilian plateau of Spain. *American Journal of Enology and Viticulture* 61: 214–224.

Schoen DJ and Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of National Academy of Science USA* 90(10): 623–10627.

Sefc K, Regner F, Glössl J and Steinkellner H (1998) Genotyping of grapevine and rootstock cultivars using microsatellite markers. *Vitis* 37: 15–20.

Upadhyay A, Kadam US, Chacko PM and Karibasappa GS (2010a) Microsatellite and RAPD analysis of grape (*Vitis* spp.) accessions and identification of duplicates/misnomers in germplasm collection. *Indian Journal of Horticulture* 67: 8–15.

Upadhyay A, Kadam US, Chacko PM, Aher LB and Karibasappa GS (2010b) Microsatellite analysis to differentiate clones of Thompson seedless grapevine. *Indian Journal of Horticulture* 67: 260–263.

Upadhyay A, Aher LB and Karibasappa GS (2011) Detection of variation among clonal selections of grapevine (*Vitis vinifera* L.) cv. Kishmish Chernyi by AFLP analysis. *Journal of Horticultural Science and Biotechnology* 86: 230–234.

Veloso MM, Almandanim MC, Baleiras-Couto M, Pereira HS, Carneiro LC, Fevereiro P and Eiras-Dias J (2010) Microsatellite database of grapevine (*Vitis vinifera* L.) cultivars used for wine production in Portugal. *Ciência e Técnica Vitivinicola* 25: 53–61.

Vignani R, Scali M, Masi E and Cresti M (2002) Genomic variability in *Vitis vinifera* L. Sangiovese assessed by microsatellite and no radioactive AFLP test. *Electron Journal of Biotechnology* 5: 3–4.

Vouillamoz JF, Arnold C and Frei A (2009) Swiss *Vitis* microsatellite database. *Acta Horticulturae* 827: 477–480.