ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2014) **12**(1); 129–139
doi:10.1017/S1479262113000415

Microsatellite marker analysis reveals the events of the introduction and spread of cultivated mulberry in the Indian subcontinent

R. Ramesh Krishnan[1], V. Girish Naik[1]*, S. R. Ramesh[2] and S. M. H. Qadri[1]

[1]*Molecular Biology Laboratory – 1, Central Sericultural Research and Training Institute, Srirampura, Manandavadi Road, Mysore 570 008, India and* [2]*Central Sericultural Germplasm Resources Centre, P.B. No. 44, Thally Road, Hosur 635 109, India*

Received 6 June 2013; Accepted 11 August 2013 – First published online 4 September 2013

Abstract

The secret art of silk culture along with mulberry seeds is presumed to have spread from China to other parts of the world including India through the famous Silk Road. In this study, we investigated a set of 36 important mulberry genotypes (designated as 'breeders' collection') of historical importance that have been frequently used in Indian crop improvement programmes over the last five decades. This study is the first to employ a large number of microsatellite markers (140 screened; 70 used for profiling) to elucidate the diversity, structure and breeding history of mulberry. The cluster and STRUCTURE analysis corroborated with the known genetic relationships and origin. The groupings by STRUCTURE ($k = 4$) confirm parallel breeding efforts undertaken in the eastern, southern and northern regions of the country. The cultivar 'Mysore Local' shares a common parent with 'Berhampore Local', and this supports the historical records of its introduction from eastern India to the Kingdom of Mysore by the erstwhile ruler Tippu Sultan. The popular variety 'Kanva-2' shared a common male parent with the Japanese variety 'Kousen', supporting the contribution of exotic progenitors in Indian cultivars. The findings of this study will be useful in formulating new strategies for mulberry improvement and reveals the historical events of the introduction and spread of cultivated mulberry in the Indian subcontinent.

Keywords: breeders' collection; breeding history; diversity and structure; microsatellite marker; mulberry

Introduction

Mulberry (*Morus* spp.) is a perennial crop cultivated mainly in the Asian continent for sericulture and is also exploited for its valued timber and fruits. Mulberry foliage forms the only source of food for the domesticated silkworm *Bombyx mori* L. for the production of natural silk – 'the queen of textiles'. Silk production is a sustainable activity for labour-rich rural India with attractive financial returns at regular intervals. India earns a foreign exchange of about US$480 million (CSB, 2012) through the export of silk yarn and finished products.

The genus *Morus* belongs to the order Rosales, the family *Moraceae* and the tribe Moreae (Zhang *et al*., 2011) and is distributed mainly in the temperate and subtropical regions of the Northern Hemisphere. A considerable difference of opinion exists among the taxonomists with regard to the precise number of species under the genus *Morus*. However, the widely accepted classification by Koidzumi (1917, 1923) recognizes 24 mulberry species and a variety under a species. In India, most of the cultivated mulberry belongs to *Morus alba* and *Morus indica*.

* Corresponding author. E-mail: vgirishnaik@yahoo.com

The centre of origin of mulberry is in the region bordering China–Korea–Japan (Vavilov, 1926). Cultivated mulberry is not native to India and is believed to be introduced by Buddhist monks around 140 BCE through the Silk Road (Hyde, 1984). It is presumed that mulberry was first introduced in the form of seeds to north-eastern India through the Gangtok pass and the Kashmir valley through Ladak. Sericulture was introduced to the south from eastern India during 1720–1740 by the ruler Tippu Sultan of the Kingdom of Mysore (Datta and Nanavaty, 2005). Significant diversity in wild mulberry has been recorded in the Indo-Himalayan region (Ravindran *et al*., 1997). The wild mulberry *Morus serrata* is restricted to the sub-Himalayan regions and *Morus laevigata* is distributed in the mainland as well as Andaman Islands (Ravindran *et al*., 1998).

The characterization and evaluation of mulberry germplasm using phenotypic markers is mostly unreliable due to high genotype × environmental interactions and specific cultivation procedures practised in the crop (Naik and Dandin, 2005). In the last decade, the dominant marker systems such as random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) (Bhattacharya and Ranade, 2001; Naik *et al*., 2002; Awasthi *et al*., 2004; Srivastava *et al*., 2004), amplified fragment length polymorphism (AFLP) (Sharma *et al*., 2000; Botton *et al*., 2005) and sequence-related amplified polymorphism marker (Zhao *et al*., 2009) have been utilized for mulberry genetic characterization. The development of mulberry-specific microsatellite (SSR, simple sequence repeat) markers was first reported by Aggarwal *et al*. (2004) and then by Zhao *et al*. (2005). Unlike other agriculturally important crops, genomic resources are limited in mulberry.

Over the last five decades, mulberry breeders in India invariably maintained a small set of germplasm with alleles of interest, precisely termed as 'breeders' collection'. The present study aims to utilize mulberry-specific SSR markers to estimate the genetic diversity and structure of the breeders' collection that has been extensively used in mulberry crop improvement programmes in India. This study also reconstructs the pedigree and kinship of some of the important mulberry cultivars that was hitherto mostly not available or uncertain.

Materials and methods

Plant material

Breeders' collection consisting of 36 mulberry genotypes (Table S1, available online) was maintained in an experimental plot at the Central Sericultural Research and Training Institute (CSRTI), Mysore. These genotypes were maintained as medium bushes by regular pruning in a paired row system and replicated clonally with a spacing of 150 cm between a paired row and 90 cm within a paired row. The plants of the same genotype were replicated 60 cm apart in a row. The plants were grown under the standard package of practices recommended for mulberry cultivation in South India (Sarkar *et al*., 2000).

Morphological characterization

A total of 11 key morphological characters that are least affected by the environmental factors were recorded according to the procedure followed by Thangavelu *et al*. (1997). The characteristics such as branching nature, curvedness/straightness of the branch, phyllotaxy, and nature, lobation, colour, apex, base, margin, shape and surface of the leaf were recorded. The characteristic states were conveniently coded and used for the analysis.

DNA isolation

High-molecular-weight genomic DNA was extracted from fresh young leaves using the HiPurA Plant Genomic DNA Miniprep Purification Spin Kit (HiMedia, Mumbai, Maharashtra, India). Genomic DNA was assayed on 1% agarose gel stained with ethidium bromide for quality check and quantification of the stock solution. An aliquot of each of the stock solution was diluted to a uniform concentration of 10 ng/μl for the amplification of the polymerase chain reaction (PCR).

Optimization of SSR marker amplification

Microsatellite marker profiling on polyacrylamide gel electrophoresis was optimized using different approaches including touchdown PCR and standardization of PCR conditions (annealing temperature (T_a), primer and template concentration and number of PCR cycles). PCR conditions for 140 mulberry-specific SSR markers (Table S2, available online) were optimized and marker polymorphism was assessed using a set of diverse mulberry germplasm which included distinct species and varieties (data not shown).

SSR marker profiling

PCR amplifications were carried out in 10 μl reaction volumes, containing 10 ng of template DNA, 0.5 pmol of each primer, 2 mM $MgCl_2$, 100 μM of each deoxyribonucleotide triphosphates, 1× PCR buffer and 0.5 U *Taq*

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

DNA polymerase (HiMedia, India) on DNA Engine PTC – 200 (MJ Research, Waltham, Massachusetts, USA) programmed to the following cycling profile: initial denaturation at 94°C for 5 min followed by 25 cycles of 94°C for 45 s denaturation, primer-specific annealing temperature (Table S2, available online) for 30 s and 72°C for 45 s extension followed by the final extension step of 72°C for 5 min. The PCR products were resolved on 8% non-denaturing polyacrylamide gels using the Sequi-Gen GT electrophoresis system (Bio-Rad, Hercules, California, USA). The gels were silver-stained according to the method of Bassam and Gresshoff (2007) with modification and scanned. The allele sizes were estimated with reference to the pBR322 DNA-*Msp*I digest marker (New England Biolabs, Ipswich, Massachusetts, USA) and recorded in base pairs.

Data analysis

Genetic diversity

Morphological traits were analysed by calculating simple matching dissimilarity coefficients using DARwin v. 5.0.158 (Perrier and Jacquemound-Collet, 2006). Number of alleles, observed heterozygosity (H_O), expected heterozygosity (H_E) and polymorphic information content (PIC) were calculated for SSR marker data using PowerMarker v. 3.25 (Liu and Muse, 2005). A simple matching dissimilarity coefficient was calculated for all pairs of individuals, a hierarchical tree based on the unweighted pair group method with arithmetic mean (UPGMA) method was constructed using DARwin v. 5.0.158 and the tree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The principal coordinate analysis (PCoA) was carried out using GeneAlex v. 6.5b3 (Peakall and Smouse, 2006), and the first two principal components were plotted using ggplot2 (Wickham, 2009) in the R program (R Development Core Team, 2008). Rare and private allelic richness were calculated for each individual using HP-Rare v. 1.1 (Kalinowski, 2005).

Structure

The structure of the breeders' collection was analysed using Bayesian model-based clustering implemented through the program STRUCTURE v. 2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). The admixture model was chosen to run the program STRUCTURE for ten independent simulations for each subpopulation (k) value ranging from 1 to 10 having iterations of 10,000 burn-in length and 10,000 Markov chain Monte Carlo. The optimal subpopulation size was determined by considering the following: (1) maximum-likelihood value on the plot; (2) *ad hoc* statistic (Δk) (Evanno *et al.*, 2005); (3) pattern of clustering realized by the UPGMA method; (iv) grouping depicted by PCoA. Among the ten replicated runs, the run with the maximum-likelihood value was chosen to assign the posterior membership quotient (Q) to each genotype. Nei's genetic distance among the subpopulations was estimated and pairwise F_{ST} was calculated using Alrequin v. 3.1 software (Excoffier *et al.*, 2006).

The analysis of molecular variance (AMOVA) was used to assess the genetic variance within and among the STRUCTURE-derived subpopulations and was calculated using Alrequin v. 3.1 with 10,000 bootstrapping to obtain a reliable estimate.

Pedigree reconstruction

The Micro-Checker program (Oosterhout *et al.*, 2004) was used to identify the null alleles and to exclude them from parentage analysis. Allelic marker data were computed along with the known pedigree information to reconstruct the pedigree of the breeders' collection using the program COLONY v. 2.0.2.3 (Wang and Santure, 2008). COLONY assigns parentage and estimates full- and half-sibs based on the maximum-likelihood method. The pedigree chart was drawn using Pedigraph v. 2.4 (Garbe and Da, 2008) based on the assignments by COLONY.

Mantel test

The significance of the correlation of clusters generated by genotypic and phenotypic marker data was tested using GeneAlex v. 6.5b3 by the Mantel test (Mantel, 1967). The Mantel test was also performed to check the correlation between the results of pairwise Nei's genetic distance and F_{ST} for the STRUCTURE-defined subgroups.

Results

Morphological analysis

The dissimilarity coefficient matrix based on simple matching using morphological characters was generated and presented in the upper diagonal of the heat map (Fig. S1, available online). The mean pairwise dissimilarity was 0.408. The dissimilarity coefficient was maximum (0.909) between Mysore Local and Kanva-2 × Kousen (Sahana) and was minimum (0) between three pairs of genotypes, namely (1) S-34 and S-30, (2) S-799 and C-776 and (3) S-523 and RFS-135. The frequency distribution of morphological characteristics in the breeders' collection is shown in the stacked bar plot (Fig. S2, available online).

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

Mulberry-specific SSR markers and polymorphism

Initially, 140 SSR primer pairs were screened using a diverse set of 12 genotypes. However, based on marker resolution and polymorphism, only 70 SSR primer combinations (Table S3, available online) were utilized to profile the breeders' collection. Among the selected SSR markers, 26 (37.1%) represented di-repeat motifs followed by 13 (18.6%) each of tri- and compound repeats. Hexa- and penta-repeat motifs were found in 12 (17.1%) and 4 (5.7%) markers, respectively. Tetra-repeat motifs were rare and found only in one (1.4%) among the markers employed in the study. The 70 SSR loci generated a total of 215 alleles with a mean of 3.1 alleles per primer pair. Among the SSR loci, 15 were monomorphic and the rest produced alleles ranging from 2 to 9. A total of 66 rare alleles ($A_R \leq 10\%$) were recorded among the 36 SSR markers and the maximum (5) was observed in MulSTR 3. The H_O varied from 0 to 0.84 and the H_E ranged from 0 to 1. The PIC was maximum (0.82) in MESTSSR 31 and minimum (0) in the monomorphic markers.

The correlation of repeat motifs with the PIC was studied. The mean PIC value ranged from a maximum of 0.37 for di-repeats to a minimum of 0.06 for hexa-repeat motifs. The compound repeat motifs with a mean PIC value of 0.35 were excluded from the analysis. Pearson's correlation analysis was carried out to establish the correlation of repeat motif size with the PIC value. Pearson's correlation coefficient ($r = -0.44$) indicated a negative correlation between the two parameters. The analysis suggests that the di-repeat motifs were more polymorphic and the PIC value decreased with an increase in the repeat motif size.

Genetic diversity and structure

The mean pairwise dissimilarity (Fig. S1, available online) computed based on simple matching was 0.26 among the breeders' collection. The least dissimilarity of 0.01 was recorded between RFS-135 and RFS-175, two cultivars evolved from the open-pollinated hybridization of Kanva-2. The next pair having minimum genetic dissimilarity was Punjab Local and Himachal Local with a dissimilarity coefficient value of 0.02. The maximum genetic divergence was recorded between Lamia Bay and Muki with a dissimilarity coefficient value of 0.61. Lamia Bay showed significantly higher dissimilarity coefficient values with all the collections except Doomar Nali. Lamia Bay is a wild mulberry collection belonging to *M. laevigata* from the Andaman Islands. Hierarchical clustering based on the UPGMA method (Fig. 1) showed four major groups (C1–C4). The first group (C1) consisted of 31 genotypes, which can be further

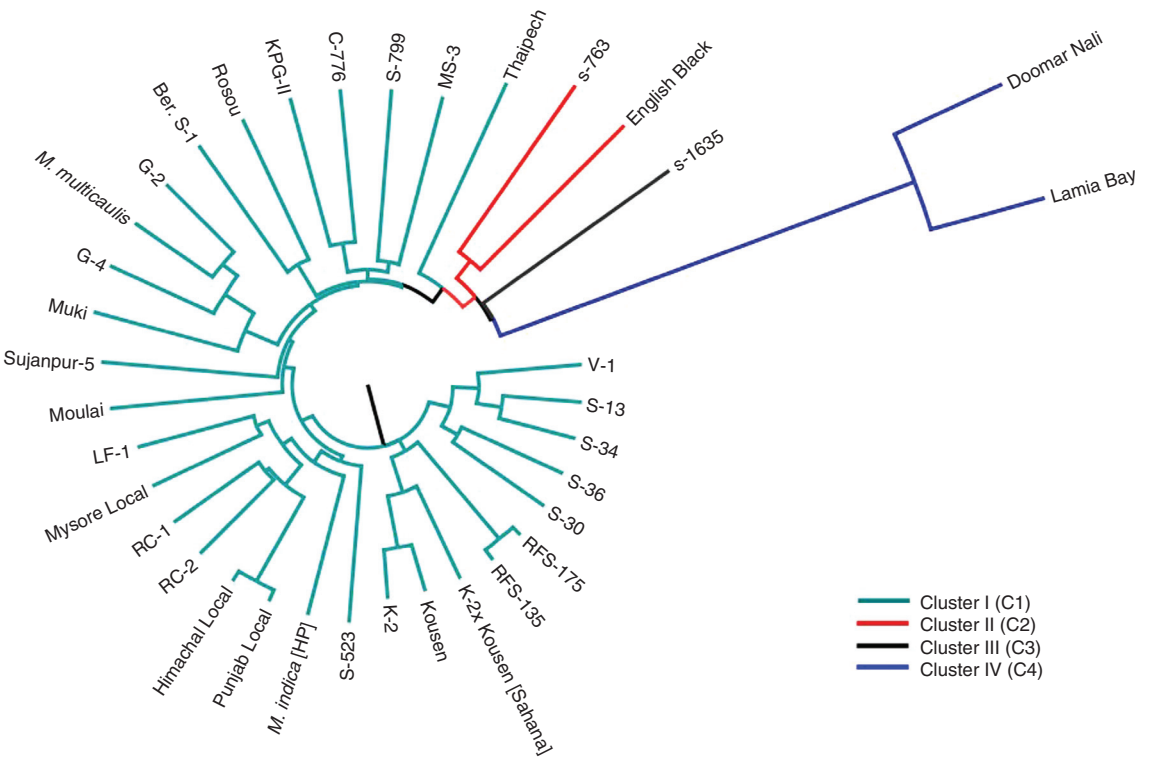

Fig. 1. (colour online) UPGMA clustering of the breeders' collection based on the microsatellite allelic data.

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

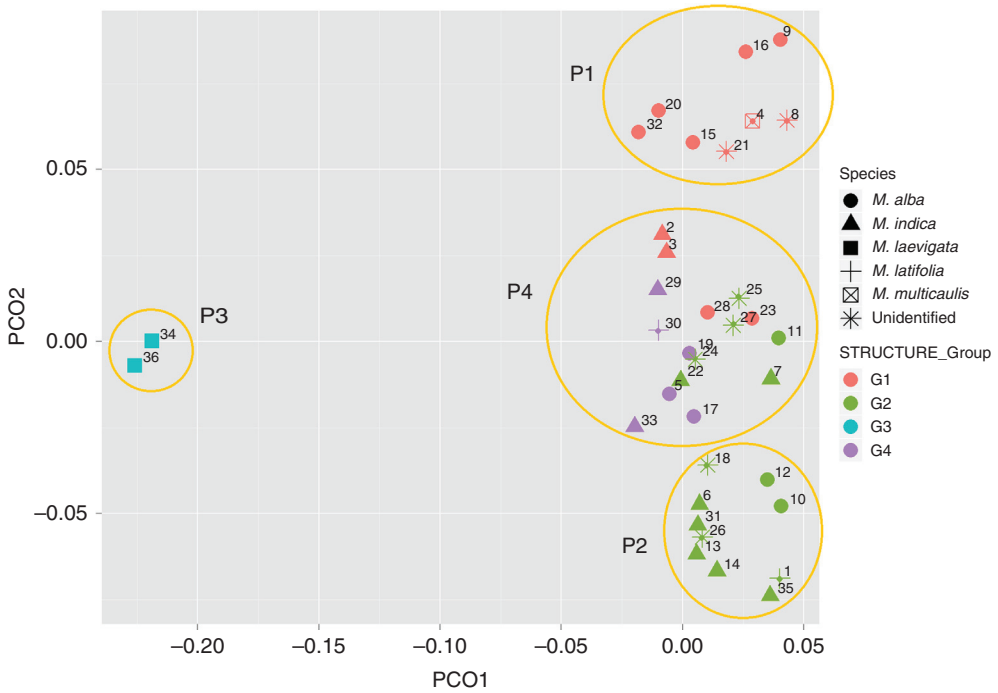

Fig. 2. (colour online) Scatter plot of the first two principal coordinates of mulberry genotypes. Numbers correspond to the serial number of the mulberry genotype given in Table S1 (available online).

divided into six subgroups. The majority of the genotypes under the first group (C1) are important cultivated varieties of India with region-specific characteristics. Only two collections, namely *Morus multicaulis* and Rosou, are of exotic origin and extensively used as donor parents in mulberry breeding programmes. Group II (C2) was represented by two accessions: English Black and S-763. Cluster III (C3) was represented only by S-1635, a national check for mulberry improvement programmes in India until recently. The last group (C4) consisted of two wild mulberry genotypes from the Andaman Islands known to harbour genes for resistance to biotic stress and tolerance to abiotic stress.

The scatter plot (Fig. 2) based on the PCoA revealed four major groups (P1–P4) among the breeders' collection. The two collections of wild mulberry from the Andaman Islands represented a distinct group (P3) of the PCoA and corroborated with grouping obtained by the UPGMA method. The important mulberry varieties of India, which were mostly confined to group I of the UPGMA clustering, formed two groups in the PCoA scatter plot (P2 and P4). Two genotypes (English Black and S-763) from group II of the UPGMA clustering formed a separate group (P1) in the PCoA along with six other genotypes from groups I and III of UPGMA. The PCoA results were broadly in concurrence with UPGMA analysis.

The genetic structure of the breeders' collection was estimated using 70 locus-specific SSR markers by the STRUCTURE (Fig. 3). The LnP(D) was maximum at $k = 5$ and decreased gradually until $k = 10$. The *ad hoc*

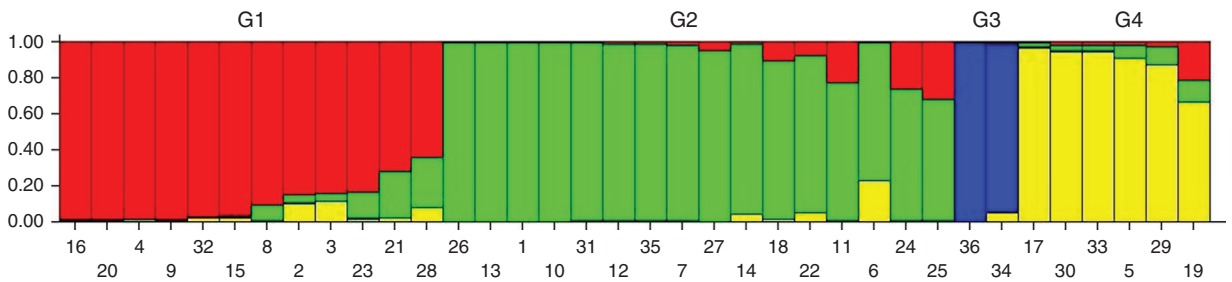

Fig. 3. (colour online) Genetic structure analysis of the breeders' collection based on SSR markers with subpopulation size (k) = 4. The *Y*-axis indicates the membership coefficients and the *X*-axis corresponds to the serial number of mulberry genotypes in Table S1 (available online).

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

Table 1. Nei's genetic distance (upper diagonal) and F_{ST} (lower diagonal) between the STRUCTURE-inferred subgroups

Subgroup	G1	G2	G3	G4
G1	0	0.0303	0.2107	0.0138
G2	0.0832	0	0.2170	0.0173
G3	0.4748	0.5019	0	0.1739
G4	0.0543	0.0483	0.4464	0

statistic (Δk) was found to be maximum at $k = 4$ (Fig. S3(a) and (b), available online). However, UPGMA clustering, PCoA, geographical origin and known pedigree of the genotypes supported the assumption of $k = 4$. In the STRUCTURE analysis also, the wild mulberry genotypes formed a distinct group with very little sharing of genomes with the other genotypes. However, the organization of the rest of the genotypes represented a hybrid structure (admixtures). The grouping obtained by the structure analysis with $k = 4$ was designated as G1, G2, G3 and G4, respectively.

Nei's genetic distance and F_{ST} were calculated (Table 1) based on SSR allelic data corresponding to the groups ($k = 4$) of the STRUCTURE analysis. Among the groups, Nei's genetic distance (0.217) and pairwise F_{ST} (0.502) were found to be maximum between G2 and G3. G2 consisted of the genotypes mainly belonging to *M. alba* from South India with exceptions of Kousen and Sujanpur-5. However, G3 consisted of two collections of *M. laevigata* and the divergence estimated by the two methods was well explained. However, minimum Nei's genetic distance (0.014) was recorded between G1 and G4 and the minimum pairwise F_{ST} (0.049) was between G2 and G4. The pairwise F_{ST} was marginally higher between G1 and G4 (0.543). The minimum values recorded for both the parameters can be appreciated from the fact that these genotypes belonging to different groups (G1 – North India, G2 – South India and G3 – exotic with exceptions) belonged to cultivated mulberry, namely *M. alba* and species with close affinity to this taxon. The AMOVA revealed that about 83% of the total genetic variance was accounted for within the population and the rest 17% was accounted for among the populations (Table 2).

Pedigree reconstruction

Genotypes with well-recorded pedigree information were used for anchoring and parentage prediction based on SSR allelic inheritance. The Micro-Checker program detected null alleles in three SSR primers (MESTSSR 13, MESTSSR 31 and MESTSSR 42) and were excluded from the parentage analysis. COLONY was run independently three times and the program was able to consistently assign the relationships. The program assigned half-, full-sibs and parents based on microsatellite marker exclusion. A total of six full-sib and 37 half-sib relationships were established. The evolution and genetic relationships among the breeders' collection was determined (Fig. S4, available online).

Mantel test

The correlation of microsatellite markers with important morphological (qualitative) characteristics was tested and resulted in significantly low correlation ($R^2 = 0.004$). Nei's genetic distance showed a high correlation ($R^2 = 0.993$) with F_{ST}.

Discussion

SSR marker polymorphism

The present investigation is the first to utilize a large number of SSR markers (genic and genomic) in mulberry. The study revealed a negative correlation ($r = -0.44$) of repeat motif size with PIC values. Di-repeat motifs were more informative in mulberry compared with tri- and tri-repeats which in turn were more informative than tetra-repeats and so on. The mean repeat motif number decreased from 13.6 (di) to 6 (tri), 4 (tetra), 3.5 (penta) and 2.6 (hexa). Studies on the distribution of microsatellites in plant genomes have suggested that di-repeats are more prevalent in non-coding regions and expressed sequence tag (EST) sequences are comparatively richer in tri-repeats (Victoria *et al.*, 2011). We analysed 3311 mulberry-specific ESTs in the GenBank database and

Table 2. AMOVA among and within the STRUCTURE-inferred subgroups ($k = 4$)

Source of variation	df[a]	Sum of squares	Variance components	Percentage of variation
Among the populations	3	108.243	1.74	17.10
Within the population	68	574.729	8.45	82.90
Total	71	682.972	10.19	100.0

[a] Degrees of freedom.

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

found that the majority of SSRs (42%) were tri-repeats (unpublished data). It has been found that the variation in repeat motif number is higher in di-repeat motifs than in other repeat motifs (Yu *et al*., 2011), and allelic polymorphism is directly proportional to the number of repeat units (Sung *et al*., 2010). This offers a plausible explanation for the negative correlation of repeat motif size with the PIC value in the study.

Genetic diversity and structure

The breeders' collection mainly composed of an assembly of genotypes which are commercially important, having special attributes such as higher leaf yield, quality, combining ability, resistance to biotic stress, tolerance to alkalinity, moisture stress, etc. (Table S1, available online). The majority of these genotypes belonged to the cultivated species of mulberry (*M. alba*, *M. indica*, *M. multicaulis* and *M. latifolia*). Only two recent collections, namely Lamia Bay (♂) and Doomar Nali (♀), belonging to the wild species (*M. laevigata*) have been included for the enhancement of gene pool availability in breeding programmes. These two genotypes have been selected for possessing some unique characteristics such as reproductive compatibility with cultivated types, resistance to termites, large-sized leaves and high regeneration ability (Ravindran *et al*., 1998). The two wild genotypes are diploids unlike many of the *M. laevigata* collections from the mainland which are mostly triploids or higher polyploids. The overall mean dissimilarity based on SSR alleles among the breeders' collection was estimated to be 0.26, which is low. The mean dissimilarity was further reduced to 0.22 if the two wild genotypes were removed from the analysis. Until recently, most of the efforts in mulberry improvement in India were towards increasing the leaf yield and many of the other commercial traits were often neglected (Tikader and Kamble, 2008). In addition, mulberry breeders have preferred to use the same set of gene pool for achieving the objectives of crop improvement programmes resulting in a narrow genetic base of the evolved varieties. Therefore, there is a strong need for broadening the genetic base of mulberry by the assessment of total diversity available in the mulberry species and the construction of a core subset for effective utilization of the natural gene pool in crop improvement.

Even though the heterozygous nature of mulberry has often been reported in the literature (Naik *et al*., 2002; Naik and Dandin, 2005), no quantifiable assessment has been made so far. The present study is the first to estimate heterozygosity in mulberry based on microsatellite allelic richness (A_R; Table 3). The mean A_R was calculated as 1.36 and all the genotypes recorded a value >1, indicating that the mulberry varieties analysed were heterozygous. S-13, a variety recommended for rain-fed areas, was most heterozygous ($A_R = 1.79$) of all the genotypes. The assessment of private allelic richness (A_{PR}) indicated that the wild genotype Doomar Nali recorded the maximum number of unique alleles among the breeders' collection. This genotype has several unique characteristics and is a useful donor parent in breeding with the largest sized leaves in mulberry (Ravindran *et al*., 1997). Among the STRUCTURE-derived four subgroups (G1–G4), G3 having wild genotypes showed high private allelic richness than the other subgroups (Table S4, available online).

Many studies have used dominant marker systems such as RAPD (Chatterjee *et al*., 2004; Orhan *et al*., 2007), ISSR (Kar *et al*., 2008) and AFLP (Sharma *et al*., 2000;

Table 3. Allele and private allele statistics of the mulberry genotypes

Sl. no.	Genotype	A_R[a]	A_{PR}[b]
1	Kousen	1.49	0.00
2	Punjab Local	1.40	0.00
3	Himachal Local	1.39	0.00
4	*M. multicaulis*	1.41	0.00
5	Ber. S-1	1.24	0.03
6	Mysore Local	1.30	0.01
7	S-13	1.79	0.00
8	G-4	1.40	0.00
9	Muki	1.33	0.00
10	S-36	1.36	0.00
11	S-34	1.36	0.00
12	S-30	1.30	0.01
13	RFS-175	1.40	0.00
14	Sujanpur-5	1.30	0.00
15	S-763	1.29	0.01
16	English Black	1.34	0.00
17	Thaipech	1.41	0.03
18	RC-2	1.34	0.00
19	S-799	1.46	0.01
20	KPG-II	1.27	0.00
21	C-776	1.33	0.00
22	LF-1	1.27	0.00
23	Moulai	1.39	0.00
24	RC-1	1.23	0.00
25	G-2	1.54	0.00
26	K-2 × Kousen (Sahana)	1.69	0.00
27	V-1	1.26	0.00
28	S-523	1.27	0.00
29	MS-3	1.39	0.03
30	Rosou	1.17	0.03
31	RFS-135	1.36	0.00
32	S-1635	1.17	0.03
33	*M. indica* (HP)	1.30	0.00
34	Doomar Nali	1.34	0.06
35	Kanva-2	1.39	0.00
36	Lamia Bay	1.29	0.03
	Mean	1.36	0.007778

[a] A_R, allelic richness. [b] A_{PR}, private allelic richness.

https://doi.org/10.1017/S1479262113000415 Published online by Cambridge University Press

Botton *et al.*, 2005; Kafkas *et al.*, 2008) to characterize indigenous and exotic mulberry. Based on fluorescent-AFLP marker analysis of 45 mulberry germplasm accessions across different eco-geographic regions of Japan and other parts of the world, Sharma *et al.* (2000) inferred that clustering of these genotypes into distinct groups was not apparent. Furthermore, the study could not establish any correlation between estimated genetic relationships and the geographical origin of the genotypes. On the contrary, Vijayan (2004) analysed 18 mulberry genotypes from India and Japan, and clustering based on ISSR and RAPD markers clearly demarcated the accessions based on the countries of origin. However, studies on cultivated mulberry (Bhattacharya *et al.*, 2001; Botton *et al.*, 2005) utilizing the dominant markers could not establish the correlation between genetic divergence and geographical distribution. These inconsistencies in the performance of random markers have limited their use. To the best of our knowledge, to date, there is no report substantiating the known mulberry genetic relationships based on the pedigree using dominant markers. The lack of availability of sufficient co-dominant markers in mulberry has limited its use until now. Aggarwal *et al.* (2004) first reported the development of six mulberry-specific SSR markers and the analysis was limited to the testing of cross-species affinity using a set of 43 mulberry genotypes. Similarly, Zhao *et al.* (2005) validated the microsatellite polymorphism in mulberry using 27 genotypes of Chinese origin. Furthermore, Zhao *et al.* (2007) compared the performance of SSR and ISSR using the same set of Chinese mulberry genotypes and concluded that the former detected higher polymorphism, indicating higher discriminatory capacity than the latter. In the present study, the UPGMA clustering precisely supported the known genetic relationships of varieties: Sahana, G-2, G-4, RFS-135, RFS-175, etc. (Table S1, available online).

The clustering derived from the UPGMA method (Fig. 1) could be broadly categorized into four groups and in concordance with genetic relationships based on the known pedigree. Cluster I (C1) was the largest with the maximum number (31) of genotypes. The genotypes in this cluster belonged to different species of cultivated mulberry (*M. alba*, *M. indica*, *M. multicaulis* and *M. latifolia*) and used extensively in breeding without an inter-specific sexual barrier. However, high reproductive compatibility does not recommend the species distinction. This has also been supported by Hirano's (1980) observation based on isozymes and several sap proteins. The clustering of the large number of cultivated genotypes should be inferred in this background. Cluster I could be further resolved to smaller subgroups to establish accurate genetic relationships among the mulberry varieties. RFS-135 and RFS-175 (both Open Pollinated Hybrid (OPH) of Kanva-2) were the closest followed by Himachal Local and Punjab Local. The high-yielding 'S' series of mulberry varieties suitable for both irrigated and rain-fed conditions of South India formed a distinct subgroup, suggesting the utilization of a common gene pool for the evolution of these varieties by the breeders. In early mulberry crop improvement efforts, OPH seeds were utilized for selecting better recombinants and development of varieties (Kesavacharyalu *et al.*, 2006). Due to this reason, many of the popular cultivars developed in the 1980s lack clear pedigree records. The exotic collection (English Black) and an elite variety (S-763) in cluster II (C2) were significantly divergent with maximum intra-cluster distance. S-1635 in cluster III (C3) is a unique and widely adaptable variety to different agro-climatic conditions in the sericulture zones of India. The grouping of two wild genotypes from the Andaman Islands was distinct (C4), which also supported the relationship based on geographical distribution and taxonomic status (Naik, 2007). PCoA (Fig. 2) based on SSR allelic data further supported genetic relationships among the genotypes and was very similar to the one depicted by the UPGMA method. Broadly, the PCoA also plotted four groups (P1–P4) across the two axes on the Cartesian plane. The wild genotypes were clearly separated in the PCoA and formed a distinct group (P3). The rest of the mulberry genotypes were distributed in three other groups according to the known or recorded genetic relationships of the mulberry genotypes. The species distinction of cultivated mulberry could not be achieved both in the UPGMA and PCoA, suggesting the use of gene pools of different origin in Indian breeding programmes and inferring that most of them are hybrids.

The STRUCTURE analysis with assumed subpopulation size ($k = 4$) is shown in Fig. 3. Wild mulberry genotypes from the Andaman Islands formed a distinct group (G3), structured with little genome sharing with other groups. The grouping of these wild genotypes can be appreciated in the background of taxonomic distinction as well as reproductive isolation. The varieties evolved in the eastern and northern parts of the country were grouped (G1) and separated from southern genotypes (G2). The exotic genotypes with a few exceptions formed the last group (G4). The STRUCTURE analysis throws insight into the breeding and evolutionary history of mulberry varieties of South India. The varieties developed in the 1970s and early 1980s were from a narrow gene pool of Mysore Local (introduced to Mysore region during 1720–40 by Tippu Sultan), Kousen (Japanese variety) and Kanva-2 (traditional cultivar) along with Sujanpur-5 (Guruprasad, 2011). In the late 1980s and early 1990s, there was a shift in the approach of South Indian breeders to utilize more exotic genotypes and varieties

originating from eastern India in breeding programmes for leaf quality improvement to produce the international grade bivoltine silk (2A–4A). As a consequence, a high-yielding variety, Victory-1 (V-1), with superior leaf quality was developed. In addition, the breeding of two mulberry varieties, namely G-2 and G-4, using *M. multicaulis* (an exotic donor parent of the eastern Indian breeding programme) resulted in the broadening of the genetic base of South Indian cultivars and the genotypes being grouped with eastern and northern Indian varieties in the STRUCTURE analysis with $k = 4$.

Nei's genetic distance ($k = 4$) indicated that wild genotypes of the Andaman Islands are divergent from the rest of the groups. The mulberry varieties developed in the eastern and northern regions of the country showed least genetic distance. F_{ST} was highest (0.502) among the Andaman and South Indian genotypes, and they share least genetic diversity. The AMOVA revealed a high percentage of variation within the subpopulation (82.9) compared with a low percentage of variation among the subpopulations.

Breeding history and pedigree reconstruction

Systematic mulberry breeding was initiated in India during the early 1960s, utilizing few exotic varieties from Japan (Kousen and Rosou), France (*M. multicaulis* and *M. rotundiloba*), Italy (Cattaneo) and Myanmar (Mandalay) (Das and Krishnaswami, 1965). However, the introduction and breeding history of mulberry in India have been poorly recorded and have few valid reasons: (1) introduction of cultivated mulberry to India during early historical times lacked proper records; (2) high heterozygous nature of the plant coupled with dioecy and outbreeding; (3) clonal propagation of mulberry in tropical climate and high phenotypic plasticity. In the absence of credible breeding records, different approaches for deciphering the kinship and pedigree using molecular techniques have been suggested (Jones *et al.*, 2010). Lacombe *et al.* (2012) successfully reconstructed the pedigree of grapevine cultivars with 20 nuclear SSR markers and a large number of markers increased the confidence on the assignments (Jones *et al.*, 2010). Parentage analysis without parents (Waples and Waples, 2011) and polymorphic nuclear microsatellite markers enabled us to reconstruct parentage for most of the mulberry genotypes without having all the potential parents in the analysis.

The parentage analysis confirmed the available pedigree records of eastern Indian varieties. Ber. S-1 (OPH of Mandalay) was evolved from an unknown parent 2 (UKP2), predicted to be Mandalay and shares a parent (UKP3) with Rosou. The analysis also revealed that Thaipech (variety from Thailand) was developed by the hybridization of Mandalay (UKP2) and UKP1. The allelic inheritance pattern reconfirmed the origin of an important eastern variety, C-776, by the hybridization of *M. multicaulis* and Black Cherry. The traditional cultivar 'Mysore Local' shares a common parent (UKP17) with Rosou. UKP17 is also the progenitor for mulberry varieties: *M. indica* and 'Berhampore Local'. This supports the introduction of 'Mysore Local' to the south particularly to the Kingdom of Mysore.

The development of a mulberry variety, Kanva-2 (K-2) or M-5, had significantly improved sericulture in the Mysore State (present-day Karnataka) during the late 1960s. According to the pedigree record, Kanva-2 originated from the OPH seeds of Mysore Local in a Kanva farm near Mysore. However, the male parent that contributed to the development of this variety had remained elusive. The present investigation suggests that the varieties Kanva-2 and Kousen have a common paternal progenitor (UKP13). The finding is very important in terms of understanding the gene flow in mulberry and planning future breeding programmes. The present observation has been further substantiated by the close genetic relationship among Kanva-2 and Kousen as also inferred by the random marker analysis (Naik and Dandin, 2005). It is also interesting to note that the hybrid of Kanva-2 × Kousen (Sahana) was identified for cultivation under shaded conditions, especially in coconut plantations of Kerala (Das *et al.*, 2010).

The three popular varieties of South India for cultivation under rain-fed conditions, namely S-13, RFS-135 and RFS-175, are all OPHs of Kanva-2. However, the male parent of these varieties is so far unknown. The SSR allelic inheritance reveals a common male parent (UKP10), a mulberry genotype which was not included in the present investigation. The study has further validated the pedigree of many mulberry cultivars of India and specifically of South India. Furthermore, the present investigation has also provided insight into the genetic relationships of many other mulberry genotypes that were hitherto unknown. This is the first report on cultivated mulberry varieties of India wherein the pedigree records have been tested and validated based on microsatellite allele sharing and full parentages have been uncovered for most of the varieties. The course of mulberry breeding has been documented for possible revision and improvisation of breeding strategies in India.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262113000415

Acknowledgements

The authors acknowledge the financial support received from the Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi (sanction order no. BT/PR13197/PBD/19/205/2009, dated 20 September 2010) for undertaking the research work. R. R. K. received a Junior Research Fellowship under the project grant. The authors thank Dr Rajeev K. Varshney and Dr Pawan Khera of the International Crop Research Institute for Semi-Arid Tropics, Hyderabad for their useful comments for improving the draft of the manuscript. V. G. N. is grateful to Dr M. Udayakumar and Dr M. S. Sheshashayee, University of Agriculture Sciences, Bangalore for providing SSR primers of MESTSSR and Mul3SSR series. The authors also thank Mr Marian Vincent Pinto, CSRTI, Mysore for critically checking the manuscript.

References

Aggarwal RK, Udaykumar D, Hendre PS, Sarkar A and Singh LI (2004) Isolation and characterization of six novel microsatellite markers for mulberry (*Morus indica*). *Molecular Ecology Resources* 4: 477–479.

Awasthi AK, Nagaraja GM, Naik GV, Kanginakudru S, Thangavelu K and Nagaraju J (2004) Genetic diversity and relationship in mulberry (genus *Morus*) as revealed by RAPD and ISSR marker assays. *BMC genetics* 5: 1.

Bassam BJ and Gresshoff PM (2007) Silver staining DNA in polyacrylamide gels. *Nature Protocols* 2: 2649–2654.

Bhattacharya E and Ranade SA (2001) Molecular distinction amongst varieties of mulberry using RAPD and DAMD profiles. *BMC Plant Biology* 1: 3.

Botton A, Barcaccia G, Cappellozza S, Tos RD, Bonghi C and Ramina A (2005) DNA fingerprinting sheds light on the origin of introduced mulberry (*Morus* spp.) accessions in Italy. *Genetic Resources and Crop Evolution* 52: 181–192.

Central Silk Board (CSB) (2012) *Annual Report 2011–12*. Bangalore: Central Silk Board.

Chatterjee SN, Nagaraja GM, Srivastava PP and Naik G (2004) Morphological and molecular variation of *Morus laevigata* in India. *Genetica* 121: 133–143.

Das BC and Krishnaswami S (1965) Some observations on interspecific hybridization in mulberry. *Indian Journal of Sericulture* 4: 1–8.

Das PK, Balakrishna R, Mathur VB, Vedavyasa K and Dandin SB (2010) Studies on the performance of improved mulberry variety Sahana as intercrop under coconut plantation. *Indian Journal of Sericulture* 49: 199–202.

Datta RK and Nanavaty M (2005) *Global Silk Industry: A Complete Source Book*. Boca Raton, FL: Universal Publishers, pp. 18–20.

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.

Excoffier L, Laval G and Schneider S (2006) *An Integrated Software Package for Population Genetics Data Analysis. Alrequin V. 3.1*. Computational and Molecular Population Genetics Lab, Institute of Zoology, University of Berne. Available at http://cmpg.unibe.ch/software/arlequin3/

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.

Garbe JR, Da Y (2008) *Pedigraph User Manual Version 2.4*. Department of Animal Science, University of Minnesota. Available at http://animalgene.umn.edu/pedigraph/

Guruprasad (2011) Development of core collection of mulberry (*Morus* spp.) germplasm by molecular marker aided analysis. PhD Thesis, University of Mysore, pp. 145–151.

Hirano H (1980) Thremmetological studies of protein variation in mulberry. *Bulletin Sericulture Experimental Station* 28: 67–186.

Hyde N (1984) The queen of textiles. *National Geographic Magazine* 165: 2–49.

Jones AG, Small CM, Paczolt KA and Ratterman NL (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources* 10: 6–30.

Kafkas S, Ozgen M, Dogan Y, Ozcan B, Ercisli S and Serce S (2008) Molecular characterization of mulberry accessions in Turkey by AFLP markers. *Journal of the American Society for Horticultural Science* 133: 593–597.

Kalinowski ST (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187–189.

Kar PK, Srivastava PP, Awasthi AK and Urs SR (2008) Genetic variability and trait association of ISSR markers with some biochemical traits in mulberry (*Morus* spp.) genetic resources available in India. *Tree Genetics and Genomes* 4: 75–83.

Kesavacharyalu K, Balakrishna R, Dandin SB and Sarkar A (2006) Development of promising mulberry (*Morus* Spp.) genotypes for sub-optimal irrigated conditions through advanced generation breeding approach. *Indian Journal Sericulture* 45: 155–159.

Koidzumi G (1917) Taxonomical discussion on *Morus* plants. *Bulletin Sericulture Experimental Station* 3: 1–62.

Koidzumi G (1923) Synopsis specierum generis Mori. *Bulletin Sericulture Experimental Station* 11: 1–50.

Lacombe T, Boursiquot JM, Laucou V, Vecchi-Staraz MD, Peros JP and This P (2012) Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). *Theoretical and Applied Genetics* 126: 401–414.

Liu KJ and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220.

Naik VG (2007) Characterization of mulberry (*Morus* spp.) varieties using DNA markers. PhD Thesis, University of Mysore, pp. 149-154.

Naik VG and Dandin SB (2005) Molecular characterization of some improved and promising mulberry varieties (*Morus* spp.) of India by RAPD and ISSR markers. *Indian Journal of Sericulture* 44: 59–68.

Naik VG, Sarkar A and Sathyanarayana N (2002) DNA fingerprinting of Mysore Local and V-1 cultivars of mulberry (*Morus* spp.) with RAPD markers. *Indian Journal of Genetics and Plant Breeding* 62: 193–196.

Oosterhout CV, Hutchinson WF, Wills DPM and Shipley P (2004) Micro-checker: software for identifying and correcting

genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

Orhan E, Ercisli S, Yildirim N and Aga G (2007) Genetic variations among mulberry genotypes (*Morus alba*) as revealed by random amplified polymorphic DNA (RAPD) markers. *Plant Systematics and Evolution* 265: 251–258.

Peakall R and Smouse PE (2006) GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Perrier X and Jacquemound-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.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ravindran S, Rao AA, Naik VG, Tikader A, Mukherjee P and Thangavelu K (1997) Distribution and variation in mulberry germplasm. *Indian Journal of Plant Genetic Resources* 10: 233–242.

Ravindran S, Tikader A, Naik VG, Rao AA and Mukherjee P (1998) Distribution of mulberry species in India and its utilization. *Indian Journal of Plant Genetic Resources* 12: 163–168.

Sarkar A, Kumar JS and Datta RK (2000) Gradual improvement of mulberry varieties under irrigated condition in South India and optimal programme for varietal selection in the tropics. *Sericologia* 40: 449–461.

Sharma A, Sharma R and Machii H (2000) Assessment of genetic diversity in a *Morus* germplasm collection using fluorescence-based AFLP markers. *Theoretical and Applied Genetics* 101: 1049–1055.

Srivastava PP, Vijayan K, Awasthi AK and Saratchandra B (2004) Genetic analysis of *Morus alba* through RAPD and ISSR markers. *Indian Journal of Biotechnology* 3: 527–532.

Sung W, Tucker A, Bergeron RD, Lynch M and Thomas WK (2010) Simple sequence repeat variation in the *Daphnia pulex* genome. *BMC Genomics* 11: 691.

Thangavelu K, Mukherjee P, Tikader A, Ravindran S, Goel AK, Ananda Rao A, Naik VG and Shekar S (1997) *Catalogue on Mulberry (*Morus *spp.) Germplasm*, vol. I. Hosur: Central Sericultural Germplasm Resources Centre.

Tikader A and Kamble CK (2008) Mulberry wild species in India and their use in crop improvement – a review. *Australian Journal of Crop Science* 2: 64–72.

Vavilov NI (1926) Studies on the origin of cultivated plants. *Trudy Byuro po Prikladnoi Botanike* 16: 139–248.

Victoria FC, Maia LCD and Oliveira ACD (2011) *In silico* comparative analysis of SSR markers in plants. *BMC Plant Biology* 11: 15.

Vijayan K (2004) Genetic relationships of Japanese and Indian mulberry (*Morus* spp.) genotypes revealed by DNA fingerprinting. *Plant Systematics and Evolution* 243: 221–232.

Wang J and Santure AW (2008) Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* 181: 1579–1594.

Waples RS and Waples RK (2011) Inbreeding effective population size and parentage analysis without parents. *Molecular Ecology Resources* 11: 162–171.

Wickham H (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York.

Yu JN, Won C, Jun J, Lim YM and Kwak M (2011) Fast and cost-effective mining of microsatellite markers using NGS technology: an example of a Korean water deer *Hydropotes inermis* argyropus. *PLoS ONE* 6: e26933.

Zhang SD, Soltis DE, Yang Y, Li DZ and Yi TS (2011) Multi-gene analysis provides a well-supported phylogeny of Rosales. *Molecular Phylogenetics and Evolution* 60: 21–28.

Zhao W, Fang R, Pan Y, Chung J, Chung I and Park Y (2009) Analysis of genetic relationships of mulberry (*Morus* L.) germplasm using sequence-related amplified polymorphism (SRAP) markers. *African Journal of Biotechnology* 8: 2604–2610.

Zhao W, Miao X, Jia S, Pan Y and Huang Y (2005) Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. *Plant Science* 168: 519–525.

Zhao W, Zhou Z, Xuexia M, Zhang Y, Wang S, Huang J, Xiang H, Pan Y and Huang Y (2007) A comparison of genetic variation among wild and cultivated *Morus* species (Moraceae: *Morus*) as revealed by ISSR and SSR markers. *Biodiversity and Conservation* 16: 275–290.