Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of castor

S. Senthilvel*, Mobeen Shaik, K. Anjani, Ranjan K. Shaw, Poornima Kumari, C. Sarada, B. Usha Kiran

ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad – 500030, India

*Corresponding author

S. Senthilvel

e mail: senthilvel.senapathy@icar.gov.in

Telephone: +91 40 24598116

Fax: +91 40 24017969

The final publication is available at www.springerlink.com

https://link.springer.com/article/10.1007/s13562-016-0356-8

Abstract

Castor (*Ricinus communis* L.) is an industrially important oilseed crop and is the only source of an unusual fatty acid, ricinoleic acid in plant species. The castor oil and its products have numerous industrial uses including biofuel; hence, the demand for castor oil is ever increasing globally. Current productivity levels in castor are inadequate to meet the requirement, which underscores the need for breeding high yielding cultivars with better adaptability by exploiting diverse genetic resources. This study reports development and characterization of a set of inbred lines derived from a core germplasm collection of castor. The panel of 144 inbreds exhibited an excellent phenotypic diversity for morpho-agronomic traits related to plant architecture and yield components. However, SSR allelic diversity appears to be only moderate. The average number of alleles per SSR locus in the genotype panel was 3.0 and mean gene diversity was 0.38. Nevertheless, a majority of the inbred pairs (77%) had very less estimated kinship coefficients (<0.05) suggesting that they were not related by pedigree. A very low level of genetic relatedness among the genotypes and absence of population structure suggest that this genotype panel consists of ideal set of materials for association mapping studies aiming at molecular breeding of key traits in castor. To our knowledge, this is the first report on the development and characterization of a large inbred collection representing the bulk of genetic diversity in castor, which can be further exploited for genetic, physiological and molecular studies towards achieving higher productivity.

Keywords

Ricinus communis, core collection, molecular diversity, SSR markers

Introduction

Castor (*Ricinus communis* L.) is the only species of the genus *Ricinus* belonging to Euphorbiaceae family. It is predominantly a cross pollinated species (aided by wind) though self-pollination does occur (Moshkin 1986). East Africa is considered the probable origin of castor based on the prevalence of diversity (Vavilov 1951); however, it is widely distributed across the world. It is an economically important non-edible oilseed crop; mostly grown in marginal lands in arid and semi-arid regions contributing significantly to the livelihoods of the resource poor farmers. The castor seed contains unique oil with more than 80 per cent ricinoleic acid (an unusual, monounsaturated, 18-carbon fatty acid), which has many desirable industrial properties. The oil and its derivatives are used in manufacturing of lubricants, fuel, paints and coatings, plastics, cosmetics etc. Castor is also considered a potential crop for biodiesel production (Shrirame et al. 2011).

Collection, characterization and utilization of germplasm accessions are critical for genetic improvement of castor for higher productivity and quality. Global efforts in characterization of germplasm collections have shown tremendous variation for morphological traits in castor (Popova and Moshkin 1986; Webster 1994; Anjani 2012). Contrarily, molecular marker analyses have revealed only low to moderate level of diversity in castor (Allan et al. 2008; Qiu et al. 2010; Gajera et al. 2010; Foster et al. 2010). A concern is that limited sample size and/or poor representation of samples used for molecular marker analyses in most of the studies (except Foster et al. 2010) would have underestimated the actual genetic diversity present in castor. Hence, there is a need to assess the extent of molecular diversity in castor using a representative set of germplasm.

A global castor germplasm collection of more than 3,000 accessions are maintained at the ICAR-Indian Institute of Oilseeds Research (IIOR), Hyderabad, India from which, a core set has been developed based on agro-morphological traits. The core set represented almost the entire variability present in the whole collection (Sarada and Anjani 2013). This core collection can be used to assess the overall genetic diversity in the species as it fairly represents cultivars, land races, semi-wild and wild forms and to explore trait mapping through linkage disequilibrium (LD) based association analysis. However, direct use of germplasm accessions for trait mapping and breeding purposes is restricted in castor because the accessions may be highly heterogeneous due to outcrossing nature of the species. Development of highly homozygous lines (inbreds) from the accessions, which constitute the core set of germplasm would be helpful for immediate use in breeding programmes. Furthermore, it would facilitate accurate genotyping, replicated phenotyping and sharing of materials and data across research groups (Pang et al. 2015). In this context, this study was undertaken to generate near inbred lines from the accessions of castor core germplasm and analyze the extent of genetic variability and population structure in the inbred collection for its further utilization in genetic studies and breeding applications.

Materials and methods

Plant material

A subset of 144 accessions from a core germplasm set of 165 accessions developed at IIOR (Sarada and Anjani 2013) was used in the present study. The core set was originally extracted from 3,003 germplsam accessions, which consisted of indigenous and exotic sources. Data on 14 quantitative traits namely plant height up to primary raceme, number of nodes up to primary

raceme, total length of primary raceme, length of primary raceme covered by capsules, total number of racemes per plant, days to flowering, days to 50% maturity, 100-seed weight, oil content, seed yield (g/plant) at four dates of harvest after sowing (120, 150, 180, 210) and total seed yield (g/plant) were used in constructing the core set (Sarada and Anjani 2013). The core set displayed excellent diversity for agro-morphological traits related to plant architecture and yield components. The frequency distribution of 14 quantitative traits in the core set of castor accessions is provided in Supplementary Figure 1. In this study, a representative plant of the original accession, which was maintained through self-pollination by the Germplasm Management Unit of IIOR, was further advanced by single seed descent method for four generations to produce near homozygous lines (inbred lines) for each accession. The source/geographic origin and major morphological features of the germplasm accessions used in this study are given in Supplementary Table 1.

Genomic DNA extraction and SSR analysis

Total genomic DNA was extracted from fresh leaf samples of 30 day-old seedlings using the DNEasy Plant Mini kit (Qiagen, Germany). The quality and quantity was assessed through 0.8% agarose gel electrophoresis. All samples were adjusted to a uniform DNA concentration of approximately 10ng/µl using sterile distilled water. A set of 45 SSR primer pairs designed from the publically available expressed sequence tags and genome sequences of castor was used for genotyping. These SSRs (8 EST-SSRs and 37 genomic SSRs) were chosen based on their locations in scaffolds of the draft castor genome (Chan et al. 2010) as there is no linkage map available yet in castor. The primer sequences of the SSRs used in the analysis are given in Supplementary Table 2. PCR amplification was done in 10µl reaction volume containing 1X

PCR buffer (Merck Millipore) with 1.5mM $MgCl_2$, 0.2mM each of dNTPs 0.4µM each of forward and reverse primer, 0.5U Taq polymerase (Merck Millipore) and 10ng genomic DNA as template. Reaction conditions were as follows: initial denaturation at 94°C for 5min, followed by 35 cycles of denaturation at 94°C for 30s, annealing at 55°C for 30s and extension at 72°C for 30s followed by final extension at 72°C for 7min. PCR products were size-separated on 6% native polyacrylamide gels (PAGE) run on 0.5× TBE buffer at 600V for 3h using Seqi-Gen GT system (38 × 30cm) of Bio-Rad Laboratories Inc., USA. After electrophoresis, the banding patterns of the PCR products on PAGE gels were visualized by silver-staining as reported by Tegelstrom (1992) with little modifications.

Data analysis

The allelic data of SSR markers was analyzed using the software, PowerMarker version 3.25 (Liu and Muse, 2005). Level of polymorphism with respect to each marker based on allele frequencies, observed and expected heterozgosity and polymorphic information content (PIC) were measured. The pairwise linkage disequilibrium (LD) for all polymorphic markers was worked out using PowerMarker version 3.25. The LD between SSR loci was considered significant when r^2 values were >0.1 (Abdurakhmonov and Abdukarimov, 2008).

The genotypic data based on SSR markers were used to calculate the pairwise kinship coefficients among the genotypes as defined by Ritland (1996) using the software, SPAGeDi (Hardy and Vekemans 2002). All negative kinship values between individuals were set to zero. The SSR allelic data was also used to calculate the pairwise dissimilarity coefficients (simple matching) and construct a neighbor joining tree (Perrier et al. 2003) using DARwin5.0 software (Perrier and Jacquemoud-Collet 2006) to understand the relationship among the genotypes.

Principal coordinates analysis (PCoA) was performed to visualize the overall representation of diversity in the inbred collection.

The population structure in the genotype panel was also assessed using the STRUCTURE 2.3.4 software package (Pritchard et al. 2000; Falush et al. 2003). This software uses a model based on a Bayesian clustering approach to infer the population structure. To infer population structure, 15 runs for each K value from 1 to 15 were performed. For each run, a burn-in of 50,000 iterations to minimize the effect of the starting configuration was followed by an additional 100,000 iterations using a model with admixture (genotype might have mixed ancestry) and correlated allele frequencies. The likelihood of different K values was calculated and the value of K with the highest likelihood was interpreted to correspond to the number of subpopulations in the sample. The ideal value of K was determined from the uppermost hierarchical level of population structure, detected using an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values (Evanno et al. 2005). The online version of STRUCTURE HARVESTER (http://tayloro.biologyucla.edu/Struct_harvest) software (Earl and VonHoldt 2012) was used for calculating the ad hoc statistic ΔK.

Results and Discussion

A core set of germplasm was developed at IIOR for effective utilization of vast genetic resources available in castor. The core set was derived from over 3,000 germplasm accessions collected from across countries, which fairly represented the entire diversity available in the species (Sarada and Anjani 2013). The core set consisted of one accession each from Russia, Hungary, Nigeria, Australia and Brazil, three accessions from USA and the remaining were from India representing 18 states. As the accessions of the core set may contain certain level of

heterogeneity due to highly outcrossing nature of the species, a representative plant from each accession obtained from germplasm management unit of IIOR was advanced by selfing for four generations through single seed descent method to derive a near inbred line from each accession of the core set. A set of 144 inbreds thus obtained from the core set of germplasm was characterized for genetic diversity and population structure using 45 SSR markers.

Genotypic diversity

Molecular markers complement phenotypic descriptors in revealing genetic diversity effectively. Among different molecular marker systems, co-dominant markers are the most reliable for characterizing the genetic variability because of their capability to distinguish allelic types providing valuable information about the heterozygosity state of a given species. Currently, SNP and SSR markers are widely used for evaluating genetic diversity. SSRs are preferred over SNPs because of multi-allelic nature, which provides more information per locus (Remington et al. 2001). In castor, only a very few publications reported the use of molecular markers; SNP (Foster et al. 2010) and SSR markers (Allan et al. 2008; Bajay et al. 2009; Qiu et al. 2010) in characterizing genetic diversity. Invariably, all these studies indicate only a low to moderate level of molecular diversity.

In this study, a set of 45 SSR markers (Supplementary Table 2) were used to characterize 144 inbred lines out of which 39 were polymorphic (Table 1). The percentage of polymorphism observed in the inbred panel (87%) was higher than the polymorphism reported earlier (41%) for EST-SSR in castor (Qiu et al. 2010). Low proportion of polymorphic SSRs observed by Qiu et al. (2010) might perhaps be due to use of only small number (24) of castor genotypes. Also, Qiu et al. (2010) used EST-SSRs, which are less polymorphic than genomic SSRs. Since the EST-

SSRs are developed from the transcribed region of the genome, the level of polymorphism shown by this class of SSRs is low (Varshney et al., 2005).

A total of 116 alleles were observed at 39 polymorphic SSR loci across 144 inbred lines. The number of alleles per locus ranged from 2 to 5 with a mean of 2.97 (Table 1). The major allele frequency ranged from 0.379 to 0.993 with an average of 0.721. Twenty eight minor alleles (frequency <0.05) were found at 18 SSR loci. Overall the SSR allelic diversity was low in the inbred collection, which could be due to the use of many trinucleotide SSRs (41 out of 45). In general, tri-nucleotide repeats were found less polymorphic compared to di-nucleotide repeats especially in human (Chakraborty *et al*., 1997) and *Drosophila* (Schug *et al*., 1998). However, low allelic diversity was observed earlier in castor irrespective of number of repeat units of the microsatellite loci used. Qiu *et al*. (2010) have reported low number of alleles per locus (2.97) using a set of 118 EST based SSR markers consisted of 68 di-nucleotides, 42 tri-nucleotides and 8 tetra-nucleotides repeats. Bajay *et al*. (2009) detected 2 to 5 alleles, with an average of 3.3 alleles per locus using 12 genomic SSRs on 38 castor accessions from Brazilian Agricultural Research Corporation (EMBRAPA). Similarly, Allan et al. (2008) found that the average number of alleles per locus was 3.1, when they assessed genetic diversity in 200 individuals comprising 41 castor accessions using nine SSR markers.

The PIC values of SSR loci varied greatly (0.014 to 0.621) in this study but with a low mean value (0.329) (Table 1). Qiu *et al*. (2010) also reported higher range of PIC values (0.07-0.73) with low mean value (0.36) in castor. The level of observed heterozygosity for the marker loci was very low, as expected. It ranged from 0.000 to 0.140 with a mean of 0.046 (Table 1). Four SSR loci did not show heterozygosity, while the remaining loci showed only a narrow range. The accessions used in this study had undergone over seven generations of selfing

followed by four generations of advancement by single seed descent method; therefore, the lines would have reached high level of homozygosity. This set of inbred lines may be readily useful for breeding and genetic studies.

A modest level (0.382) of expected heterozygosity (gene diversity) was observed in the inbred collection. Previous studies also reported similar range of expected heterozygosity in castor. Bajay et al. (2009) noted that the expected heterozygosity ranged from 0.188 to 0.712 with an average of 0.416. Qiu et al. (2010) also reported similar level of gene diversity in castor using SSR markers. Contrary to these observations, Allan et al. (2008) reported very low gene diversity based on AFLP (0.126) and SSR (0.188) markers in a collection of 40 castor accessions. Foster et al. (2010) reported low gene diversity estimates in a worldwide castor collection of 488 samples (0.21) and in 13 wild populations (0.22) based on SNP markers. The differences in the gene diversity estimates in these studies might possibly have occurred due to differences in the sample size and poor representation of the diversity, which suggest that use of a representative set of germplasm is necessary to get the proper estimate of gene diversity in a species. In this study, a large collection of inbreds with a good representation of diversity in the core germplasm was used to obtain gene diversity estimates; hence, we hope that it would be more dependable.

Overall, this study strongly supports earlier findings that castor possess low genetic diversity at molecular level. Small number of SSR markers (45) used for evaluation of genetic diversity would not have impacted on the results of this study considering that SSRs are hyper-variable and a few loci are capable of resolving relatedness (Gustafsson and Lonn 2003; Hammerli and Reusch 2003). Studies on other oilseed crops: sunflower (Mandel et al. 2011), groundnut (Jiang et al. 2007), rapeseed-mustard (Vinu et al. 2013), sesame (Wu et al. 2014) and

safflower (Kiran et al. 2015) also have revealed low to moderate level of SSR allelic diversity. Interestingly, the SSR allelic diversity in the oilseed crops appears to be low compared to other crops particularly cereals: rice (Zhang et al. 2011), wheat (Balfourier et al. 2007), barely and oat (Leisova et al. 2007). It was also surprising to note that excellent phenotypic diversity in the castor core accessions was not reflected at molecular level. Hence, it would be important to study why the phenotypic diversity does not translate into high level of genetic diversity in castor. One possibility is that most of the phenotypic variations in castor may be due to the epigenetic mechanisms, as found in jatropha. Yi et al. (2010) reported significant epigenetic diversity within and among populations of *Jatropha curcas* L. collected from five different countries. More than half of CCGG sites surveyed by methylation sensitive florescence AFLP were methylated with significant difference in inner cytosine and double cytosine methylation among populations. Most epigenetic differential markers can be inherited as epialleles following Mendelian segregation. These results suggest possible involvement of epigenetics in jatropha development.

Genetic relatedness and structure

The chosen set of SSR markers are not in LD in the study population. Only two pairs of marker loci (RCGSSR4569 & RCGSSR3898: $r^2 = 0.6641$; RCGSSR337 & mRcDOR355: $r^2 = 0.4593$) were in significant LD among 741 possible combinations. The pairwise kinship estimates in the inbred collection ranged from 0.00 to 0.98 with an overall average of 0.08. Over 60 per cent of inbred pairs had zero estimated kinship values (Fig 1). Only about 4 per cent of genotype pairs had above 0.20 kinship values suggesting that majority of the inbreds were not related by pedigree. The weak or no kinship observed between inbreds in the panel could be attributed to the inclusion of a broad range of genotypes and the exclusion of closely related accessions during the process of core construction.

Similarly, the pairwise dissimilarity coefficients, calculated from SSR allelic data, ranged from 0.027 to 0.671 with a mean of 0.384, which indicated that most of the inbreds were considerably diverse except a few. Cluster analysis showed three major clusters and many sub groups within the major clusters but the grouping was not supported by high bootstrap values. A neighbor joining tree depicting the genetic relationship of 144 castor inbreds based on pairwise dissimilarity matrix is shown in Fig 2. The results of PCoA showed that the first two axes captured only 5.6 per cent and 5.2 per cent of total variance, respectively and did not show strong groupings (Supplementary Figure 2).

To further verify the results of cluster and PCoA analyses, the programme STRUCTURE was used. The STRUCTURE uses a model based on a Bayesian clustering approach to infer the population structure (Pritchard et al. 2000) wherein the accessions are classified into a predefined number of clusters in such a way that linkage disequilibrium (LD) does not occur within the cluster but present between clusters. The structure analysis was performed by setting the number of clusters (K) from 1 to 15 with 15 replications for each K. The average logarithm of the probability of likelihood [LnP(D)] and standard deviations for K = 1 to 15 are presented in Supplementary Table 3. The LnP(D) showed a constant increase with increasing subpopulation number (K) and no significant clear cut-off was observed based on the LnP(D) and Delta K plots, as shown in Supplementary Figure 3. The barplots (Fig. 3) showing the population structure for K=2, K=4 and K=8 also indicated clear admixture in the individuals. From these results, it can be inferred that there are no clear-cut subpopulations within the collection of inbreds.

The results of neighbor joining clustering, PCoA and model based STRUCTURE analyses clearly suggested that there was no marked genetic structure in the inbred panel. The

lack of strong genetic structure in the collection could be attributed to the following reasons: (i) the inbreds were derived from the core collection, where closely related genotypes were removed during constitution of the core and (ii) castor being an outcrossing species, very low level of population differentiation is expected due to extensive gene flow among individuals. Absence of genetic structure in the genotype panel is a desirable feature for association analysis to avoid spurious marker-trait associations (Flint-Garcia et al. 2003).

In conclusion, a set of 144 inbred lines was established from a global germplasm collection of castor. The molecular characterization data indicated that the inbred collection contained minimum repetitiveness and a reasonable level of genetic diversity with no marked population structure. Hence, this core set may be ideal genetic material for inclusion in the association mapping panel. Furthermore, the genetic diversity information generated in this study coupled with the agronomic data would assist in selection of suitable genotypes for breeding as well as physiological and molecular studies in castor.

Acknowledgment

This study was funded by Biotechnology Industry Research Assistance Council (BIRAC), New Delhi through a grant (Project No. BT/CRS00013/CRS-01/12) and ICAR-Indian Institute of Oilseeds Research, Hyderabad. The authors thank Dr. K.S. Varaprasad, Director, IIOR for the facilitation and encouragement and Dr Kadirvel Palchamy for editing the manuscript.

References

Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics 2008:574927. doi:10.1155/2008/574927

Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. Genet Resour Crop Ev 55:365-378

Anjani K (2012) Castor genetic resources: A primary gene pool for exploitation. Ind Crop Prod 35:1-14

Bajay MM, Pinheiro JB, Batista CEA, Nobrega MBD, Zucchi MI (2009) Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. Conserv Genet Resour 1:237-239

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265-1275

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041-1046.

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Rave J, Rabinowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951-956

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics 164:1567-1587

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol l10:13

Gajera BB, Kumar N, Singh AS, Punvar BS, Ravikiran R, Subhash N, Jadeja GC (2010) Assessment of genetic diversity in castor (*Ricinus communis* L.) using RAPD and ISSR markers. Ind Crop P*rod* 32:491-498

Gustafsson S, Lonn M (2003) Genetic differentiation and habitat preference of flowering-time variants within *Gymnadeniaconopsea*. Heredity 91:284–292

Hammerli A, Reusch T (2003) Genetic neighbourhood of clone structures in eelgrass meadows quantified by spatial auto correlation of microsatellite markers. Heredity 91:448–455

Hardy OJ, Vekeman X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618-620

Jiang HF, Liao BS, Ren XP, Lei Y, Emma M, Fu TD, Crouch JH (2007) Comparative assessment of genetic diversity of peanut (*Arachis hypogaea* L.) genotypes with various

levels of resistance to bacterial wilt through SSR and AFLP analyses. J Genet Genomics 34:544–554

Kiran BU, Mukta N, Kadirvel P, Alivelu K, Kishore P, Senthilvel S, Varaprasad KS (2015) Genetic diversity in safflower (*Carthamus tinctorius* L.) germplasm as revealed by SSR markers. Plant Genet Resour C. doi:10.1017/S1479262115000295

Leisova L, Kucera L, Dotlacil L (2007) Genetic resources of barley and oat characterised by microsatellites. Czech J Genet Plant Breeding 43:97–104

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129

Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theor Appl Genet 123:693–704

Moshkin VA (1986) Flowering and pollination. In: Moshkin VA (ed) Castor, Oxonian Press Pvt. Ltd., New Delhi, pp 43-49

Pang W, Li X, Choi SR, Dhandapani V, Im S, Park MY, Jang CS, Yang MS, Ham IK, Lee EM, Kim W, Lee SS, Bonnema AB, Park S, Piao Z, Lim YP (2015) Development of a leafy *Brassica rapa* fixed line collection for genetic diversity and population structure analysis. Mol Breeding doi:10.1007/sl 1032-015-0221-9

Perrier X, Flori A, Bonnot F (2003) Data analysis methods In: Hamon P, Seguin M, Perrier X, Glaszmann JC (ed) Genetic diversity of cultivated tropical plants, Enfield, Science Publishers, Montpellier, pp 43-76

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

Popova GM, Moshkin VA (1986) Botanical classification. In: Moshkin VA (ed) Castor, Oxonian Press Pvt. Ltd., New Delhi, pp11-27

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

Qiu L, Yang C, Tian B, Yang, JB, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis*L.). BMC Plant Biol l10:278

Remington DL, Thornsberry JM, Matsuola Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. U.S.A. 98:11479–11484

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res 67:175-185

Sarada C, Anjani K (2013) Establishment of castor core collection utilizing self-organizing mapping (SOM) networks. J Indian Soc Agricultural Statist 67:71-78

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 15(12):1751-1760.

Shrirame HY, Panwar NL, Bamniya BR (2011) Bio Diesel from Castor Oil - A Green Energy Option. Low Carbon Economy 2:1-6

Tegelstrom H (1992) Detection of mitochondrial DNA fragments. In: Hoelzel R (ed) Molecular Genetic Analysis of Populations: A Practical Approach, IRL Press, Oxford, UK, pp 89-114

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48-55

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Selected writings translated from the Russian by Chester KS, Chronica Botanica, Waltham MA

Vinu V, Singh N, Vasudev S, Yadava, DK, Kumar S, Naresh S, Bhat SR, Prabhu KV (2013) Assessment of genetic diversity in *Brassica juncea* (Brassicaceae) genotypes using phenotypic differences and SSR markers. Revista de biología tropical 61:1919-1934

Webster GL (1994) Synopsis of the genera and supra generic taxa of Euphorbiaceae. Ann Mo Botl Gard 8:33-144

Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y (2014) Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. BMC Genet 15:35

Yi CX, Zhang SL, Liu XK, Bui HTN, Hong Y (2010) Does epigenetic polymorphism contribute to phenotypic variances in *Jatropha curcas* L.? BMC Plant Biol 10:259

Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y (2011) Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. PLoS ONE 12, p.e27565

Table 1. Diversity measures of SSR loci used in characterization of inbreds derived from the castor core germplasm

SSR loci	Number of alleles	Major allele frequency	Gene diversity	Observed heterozygosity	PIC
RcDES140	5	0.732	0.431	0.041	0.396
mRcDOR55	2	0.844	0.263	0.072	0.229
mRcDOR59	5	0.379	0.685	0.008	0.621
mRcDOR103	2	0.788	0.334	0.058	0.278
mRcDOR228	5	0.446	0.599	0.067	0.516
mRcDOR313	3	0.528	0.577	0.133	0.494
mRcDOR355	3	0.888	0.201	0.032	0.183
mRcDOR385	3	0.838	0.279	0.044	0.253
RCGSSR130	3	0.737	0.401	0.029	0.341
RCGSSR157	4	0.493	0.563	0.125	0.468
RCGSSR317	2	0.620	0.471	0.113	0.360
RCGSSR337	3	0.864	0.241	0.096	0.223
RCGSSR603	3	0.891	0.201	0.029	0.191
RCGSSR954	3	0.483	0.545	0.140	0.441
RCGSSR1014	3	0.718	0.442	0.079	0.396
RCGSSR1030	3	0.592	0.559	0.088	0.493
RCGSSR1129	2	0.923	0.142	0.022	0.132
RCGSSR1230	2	0.719	0.404	0.058	0.322
RCGSSR1252	3	0.781	0.346	0.065	0.291
RCGSSR1434	2	0.553	0.494	0.092	0.372
RCGSSR9595	3	0.647	0.509	0.056	0.447
RCGSSR3831	2	0.899	0.182	0.014	0.166
RCGSSR3898	2	0.993	0.014	0.000	0.014
RCGSSR3956	4	0.884	0.214	0.000	0.206
RCGSSR4569	2	0.989	0.022	0.007	0.021
RCGSSR4947	4	0.654	0.515	0.021	0.464
RCGSSR5329	3	0.489	0.596	0.007	0.513
RCGSSR5646	3	0.617	0.530	0.016	0.460
RCGSSR5772	3	0.935	0.124	0.000	0.120
RCGSSR6564	2	0.770	0.355	0.035	0.292
RCGSSR6765	3	0.724	0.429	0.030	0.379
RCGSSR6813	3	0.958	0.081	0.000	0.079
RCGSSR7841	3	0.769	0.378	0.023	0.340
RCGSSR9310	2	0.660	0.449	0.028	0.348
RCGSSR10101	4	0.572	0.610	0.029	0.566
RCGSSR10187	4	0.859	0.252	0.021	0.238
RCGSSR10527	3	0.626	0.474	0.000	0.368
RCGSSR10548	3	0.610	0.526	0.059	0.448
RCGSSR12022	2	0.645	0.458	0.072	0.353
Mean	2.974	0.721	0.3819	0.0464	0.3288

Supplementary Table1. Source/geographic origin of 144 accessions of castor core subset used in the study

S. No.	Accession number	Morphological features			Source/Origin
		Stem colour[1]	Wax coating[2]	Spine on capsule[3]	
1	RG43	R	3	D	Tindivanam,Tamilnadu, India
2	RG61	R	1	D	Dantiwada, Gujarat, India
3	RG72	R	2	D	Hyderabad, Telangana, India
4	RG94	R	2	D	Hyderabad, Telangana, India
5	RG111	R	2	D	USSR
6	RG178	G	2	D	Hungery
7	RG193	R	2	A	USA
8	RG220	M	2	D	Dantiwada, Gujarat, India
9	RG224	M	2	D	Tindivanam,Tamilnadu, India
10	RG249	G	3	S	Tindivanam,Tamilnadu, India
11	RG252	R	1	D	Tindivanam,Tamilnadu, India
12	RG260	R	3	D	Tindivanam,Tamilnadu, India
13	RG264	G	2	D	Tindivanam,Tamilnadu, India
14	RG289	R	2	D	Tindivanam,Tamilnadu, India
15	RG294	R	3	D	Tindivanam,Tamilnadu, India
16	RG297	G	3	D	Tindivanam,Tamilnadu, India
17	RG408	R	1	D	Dantiwada, Gujarat, India
18	RG426	G	1	D	Dantiwada, Gujarat, India
19	RG430	G	1	D	Dantiwada, Gujarat, India
20	RG433	R	2	A	Dantiwada, Gujarat, India
21	RG489	R	1	D	Dantiwada, Gujarat, India
22	RG537	G	3	D	Dantiwada, Gujarat, India
23	RG551	R	1	D	Dantiwada, Gujarat, India
24	RG558	R	0	D	Dantiwada, Gujarat, India
25	RG565	G	2	D	Dantiwada, Gujarat, India
26	RG566	R	2	A	Dantiwada, Gujarat, India
27	RG589	R	2	D	Dantiwada, Gujarat, India
28	RG607	G	2	D	Dantiwada, Gujarat, India
29	RG673	G	2	D	Dantiwada, Gujarat, India
30	RG714	R	3	S	Hyderabad, Telangana, India
31	RG732	R	2	D	Hyderabad, Telangana, India
32	RG735	R	2	D	Hyderabad, Telangana, India
33	RG784	R	2	D	Dantiwada, Gujarat, India
34	RG790	R	1	D	Raichur, Karnataka, India
35	RG829	R	2	D	Hyderabad, Telangana, India
36	RG886	G	2	D	Tindivanam,Tamilnadu, India
37	RG892	R	2	A	Tindivanam,Tamilnadu, India
38	RG905	M	1	D	Tindivanam,Tamilnadu, India
39	RG908	M	3	D	Tindivanam,Tamilnadu, India

40	RG941	G	3	D	Dantiwada, Gujarat, India
41	RG969	R	2	D	Dantiwada, Gujarat, India
42	RG999	R	3	D	Dantiwada, Gujarat, India
43	RG1068	M	3	D	Dantiwada, Gujarat, India
44	RG1114	R	1	D	Tindivanam,Tamilnadu, India
45	RG1125	R	1	D	Tindivanam,Tamilnadu, India
46	RG1142	R	1	D	Tindivanam,Tamilnadu, India
47	RG1146	G	3	D	Tindivanam,Tamilnadu, India
48	RG1149	G	3	D	Nigeria
49	RG1173	R	3	D	Australia
50	RG1180	G	3	A	Exotic-unknown origin
51	RG1274	P_U	1	D	Tindivanam,Tamilnadu, India
52	RG1289	M	3	D	Unknown origin
53	RG1305	R	2	D	Tindivanam,Tamilnadu, India
54	RG1313	M	2	D	Tindivanam,Tamilnadu, India
55	RG1340	M	2	A	Tindivanam,Tamilnadu, India
56	RG1354	G	1	D	Tindivanam,Tamilnadu, India
57	RG1364	M	1	D	Tindivanam,Tamilnadu, India
58	RG1383	G	3	D	Tindivanam,Tamilnadu, India
59	RG1406	M	2	D	Tindivanam,Tamilnadu, India
60	RG1507	R	1	D	Palem, Telangana, India
61	RG1523	R	2	D	S.K. Nagar, Gujarat, India
62	RG1545	M	3	D	Palem, Telangana, India
63	RG1579	M	3	D	Palem, Telangana, India
64	RG1627	R	2	D	Bihar, India
65	RG1647	G	3	D	Bihar, India
66	RG1654	G	2	D	Bihar, India
67	RG1669	R	2	D	Bihar, India
68	RG1689	M	0	D	Brazil
69	RG1696	M	3	A	S.K. Nagar, Gujarat
70	RG1707	M	2	D	Hyderabad, Telangana, India
71	RG1709	M	2	D	Hyderabad, Telangana, India
72	RG1759	G	1	D	Hyderabad, Telangana, India
73	RG1849	G	2	D	Hyderabad, Telangana, India
74	RG1864	G	3	D	Hyderabad, Telangana, India
75	RG1904	M	2	D	Hyderabad, Telangana, India
76	RG1952	R	3	A	Meghalaya, India
77	RG1963	R	0	D	Meghalaya, India
78	RG1978	M	3	D	Mizoram, India
79	RG1981	M	2	D	Mizoram, India
80	RG1999	G	2	D	Manipur, India
81	RG2014	R	3	D	Nagaland, India
82	RG2022	M	0	D	Assam, India
83	RG2024	M	3	D	Assam, India
84	RG2035	R	3	D	Assam, India

85	RG2184	M	2	D	Aurangabad, Maharashtra, India
86	RG2195	R	2	D	Chindwara, Madhya Pradesh, India
87	RG2266	M	3	D	Hyderabad, Telangana, India
88	RG2269	R	2	D	Hyderabad, Telangana, India
89	RG2288	M	1	D	Tindivanam,Tamilnadu, India
90	RG2320	R	2	D	Tindivanam,Tamilnadu, India
91	RG2326	R	2	D	Tindivanam,Tamilnadu, India
92	RG2375	M	3	D	Tindivanam,Tamilnadu, India
93	RG2377	M	2	D	Tindivanam,Tamilnadu, India
94	RG2378	R	2	D	Tindivanam,Tamilnadu, India
95	RG2390	R	2	D	Tindivanam,Tamilnadu, India
96	RG2430	M	3	D	Andhra Pradesh, India
97	RG2451	M	3	D	Maryland, USA
98	RG2454	R	1	D	Hiriyur, Karnataka, India
99	RG2457	M	3	D	Visakhapatnam, Andhra Pradesh, India
100	RG2465	R	1	D	Barimada, Andhra Pradesh, India
101	RG2473	M	3	D	Kanpur, Uttar Pradesh, India
102	RG2474	R	3	D	Kanpur, Uttar Pradesh, India
103	RG2481	G	2	A	Kanpur, Uttar Pradesh, India
104	RG2498	M	3	D	Kanpur, Uttar Pradesh, India
105	RG2582	R	2	D	Kangra, Himachal Pradesh, India
106	RG2588	M	2	D	Kangra, Himachal Pradesh, India
107	RG2593	M	1	D	Doda, Jammu & Kashmir, India
108	RG2676	M	3	D	USA
109	RG2681	-	-	D	Bijapur, Karnataka, India
110	RG2685	G	3	D	Orissa, India
111	RG2705	R	0	D	Andaman & Nicobar Islands, India
112	RG2717	G	0	D	Andaman & Nicobar Islands, India
113	RG2719	R	2	D	Andaman & Nicobar Islands, India
114	RG2725	R	2	D	Andaman & Nicobar Islands, India
115	RG2789	R	2	D	Villipuram,Tamilnadu, India
116	RG2810	G	3	D	Coimbatore,Tamilnadu, India
117	RG2818	R	2	D	Coimbatore,Tamilnadu, India
118	RG2819	G	3	D	Dindigal,Tamilnadu, India
119	RG2821	G	0	D	Dindigal,Tamilnadu, India
120	RG2839	G	2	D	Kanyakumari,Tamilnadu, India
121	RG2866	R	2	A	Warangal, Telangana, India
122	RG2874	R	3	S	Warangal, Telangana, India
123	RG2902	R	2	D	Warangal, Telangana, India
124	RG2944	R	1	D	West Godavari, Andhra Pradesh, India
125	RG2958	R	3	D	West Godavari, Andhra Pradesh, India
126	RG2980	R	2	A	Guntur, Andhra Pradesh, India
127	RG2991	R	1	D	Ranchi, Jarkhand, India
128	RG3005	R	2	D	East Godavari, Andhra Pradesh, India
129	RG3013	R	2	D	Chamranagar, Karnataka, India

130	RG3037	G	3	D	Kutch, Gujarat, India
131	RG3048	G	3	D	Kutch, Gujarat, India
132	RG3063	G	3	D	Kutch, Gujarat, India
133	RG3067	G	3	D	Kutch, Gujarat, India
134	RG3073	R	3	D	Kutch, Gujarat, India
135	RG3088	G	1	D	Gujarat, India
136	RG3102	G	3	D	Salem, Tamilnadu, India
137	RG3116	R	2	A	Jodhpur, Rajastan, India
138	RG3126	G	1	D	Jodhpur, Rajastan, India
139	RG3134	G	2	D	Jodhpur, Rajastan, India
140	RG3195	R	3	D	NBPGR, New Delhi, India
141	RG3198	G	2	D	NBPGR, New Delhi, India
142	RG3224	R	2	D	Banskantha, Gujarat, India
143	RG3233	R	2	A	Sirohi, Rajasthan, India
144	RG3283	R	2	A	NBPGR , New Delhi, India

[1]Stem colour of the plant: R – Red, G – Green, M – Mahogany, Pu – Purple

[2]Waxi coating in the plant:1 – Present on stem, 2 – present on stem and lower surface of the leaf, 3 – present on stem, lower and upper surface of leaves

[3]Presence of spine on the capsule: A – Absent, S – Sparse, D – Dense

Supplementary Table 2. Details of SSR markers used in characterizing the inbreds derived from the castor core germplasm

S. No.	SSR ID*	Motif	Primer sequence (5' – 3')	Expected product size
1	RcDES140	(CCACCG)4	F: AATTACATTACTGCTGCCAACC R: TCAGCAGATGCATAGTTCTCAG	380
2	mRcDOR55	(TC)12	F: TCCTCTTCCTCTTCCTCGTT R: CGTCAGCCATGGTTAGAGAC	196
3	mRcDOR59	(TC)15	F: ATGGGTAGATTGAGCTGCTG R: GACTGAAATTAAGTGCGGGA	209
4	mRcDOR103	(CAG)8	F: AATGACAGCGAGTTCAGGAG R: GCCATAAACTCACCACAACC	171
5	mRcDOR228	(AC)11	F: CTGGAGCTTTATTCAAGCCA R: ACATCCATGCCAACTTCATC	334
6	mRcDOR313	(TGC)7	F: CTTCAACAACAACACCATCG R: CTGGCCACAAAGCTGTAGAT	133
7	mRcDOR355	(TGT)7	F: GAAAAGGGTGCTTCTCCTTC R: GTGATCACAACCTACGAGGG	242
8	mRcDOR385	(AGA)7	F: TTAGTTCCTCAATCTCCCCC R: CAAAAGACCGAGGAGTCTCA	231
9	RCGSSR119	(GAT7	F: GTATTCGTTGCTAAAGCGGA R: TAAATCCACCGTCTTCCTCA	258
10	RCGSSR130	(TAA)8	F: ATCGACCACTTCGCAGAATA R: CTGAACCCAACCATGAAAAG	266
11	RCGSSR157	(TTA)7	F: ATTCAGGCCAGATAACCCAT R: TGTTGGAACCGTACAATGAG	211
12	RCGSSR317	(TTA)8	F: GTCGGTCTTTTGCCTCATTA R: TGGTCACTCACCCTGTTTCT	296
13	RCGSSR337	(GAA)7	F: CAAAATCAGCCTTACAGGGA R: CAACCCGCATAAGTTTAGGA	230
14	RCGSSR603	(AAG)9	F: GCTCAACAACCGAGTGACTT R: GAGGCCAAAATGTCAGTGTC	292
15	RCGSSR954	(ATC)7	F: AGGACAGGGAAATCACACAG R: GTGGGATTCTGCAGGTTATG	261
16	RCGSSR1014	(TCT)11	F: AAACCTTAGGTGTTGACCGC R: AAGCCCAATTATCTGGAAGC	233
17	RCGSSR983	(AAT7)	F: CGTATAATCCACCATGCGAT R: GGCAGCCTATCCATACCAAT	279
18	RCGSSR1030	(TAT)9	F: CATCCCATTTGCTTCTTCAC R: TTGTAGCTCAGCTGCCTTCT	229
19	RCGSSR1129	(TAC)7	F: CTTGTATGGGCTCAATTTGG R: TTTACGGCAAACTTCTGAGG	205
20	RCGSSR1230	(AAG)9	F: GAGGCAGCCAACATCTTAAA R: AGAGACAAGTGAGCTGGGTG	206
21	RCGSSR1252	(AAT)13	F: CAAGCTTCACGTCCTCAGAT R: GCAGCTCGAGGATATGGTTA	269
22	RCGSSR1368	(TGC)7	F: ACGCTGCTTCTACTGCTGTT R: TCCGGATCATTCTCCTTACA	225
23	RCGSSR1434	(ATA)10	F: GTTAAAAGCCAAAGAAGGGG R: GCCTTTTTAGTGGGCCTAAC	293

24	RCGSSR1453	(AGA)7	F: CGATTTTCTACAACCCTTGC R: ACGAATCATCAAAACGACGG	214
25	RCGSSR1475	(GAA)8	F: CCAATAATCTCCCAATGCAC R: TGATTCGAGTCAGACAGCCT	294
26	RCGSSR9595	(AAG)9	F: GGGTATTGGAGGGAAAGAGA R: GTACCATTTGGCATATTGGG	299
27	RCGSSR3831	(AAT)9	F: TAGCTTGAATTTCCGAGCAG R: GGGCTTTACAATTCCCATCT	265
28	RCGSSR3898	(CTT)8	F: CTCTGATTACCTGCTTGGGA R: TACAGGAGTCACCCAATCGT	201
29	RCGSSR3956	(TGC)8	F: ATGATGCTGTTGGTGAAGGT R: AAGATCAATTTCTCCGACCC	237
30	RCGSSR4569	(CAC)7	F: ATCGCCATAAGCTGTGAGTT R: GGTTATCCAGGTTACCCACC	256
31	RCGSSR4947	(TTA)7	F: ATGGAAAGTAGTTTGCCTGG R: ATTGCCAAGGACTGACTGAG	249
32	RCGSSR5329	(CAC)9	F: TTCTTTCGCTCTCTCACACC R: TGTTGCAGCTTGACACATCT	232
33	RCGSSR5646	(TCC)12	F: AAACAAACCTTGGAGAACCC R: TGAGAGGTTGCAAGGTAAGG	224
34	RCGSSR5772	(TTA)13	F: GAGAGTGAAAGTGTCAAACACC R: CTTATTGGGCACAGGAAAAG	211
35	RCGSSR6564	(GAA)10	F: TGCTTTAGTCACGTGTAGCG R: CTGTGTCTAGATCCCCATGC	205
36	RCGSSR6765	(GAA)11	F: TTCTTCTTCTTCATCGTCCG R: TTTCACCCTCTCAACAGACC	282
37	RCGSSR6813	(AAG)8	F: AAAGAGAGAGAGAAGGGGCA R: ATCATCATCCCACACACACA	219
38	RCGSSR7841	(CTT)12	F: AAGGCAACCTTCATTAGCCT R: TCAACCCTTCACATATGGCT	207
39	RCGSSR7947	(TAT)7	F: GGACTTTCATTGTCATTGCC R: TTCCCTTTCCCTTTCTCTTC	260
40	RCGSSR9310	(AAG)8	F: AGTGGACGTGCAAACAAAGT R: AAGCAGATTGGCATGTTCTC	282
41	RCGSSR10101	(AAT)13	F: CAAATCAACATTAGGCCCAC R: TTGTTACATGTGGCACGAAC	279
42	RCGSSR10187	(AGA)10	F: TGTATGGGAGATGGGAAAGA R: CAAGTGGGTCCTTGAAGATG	274
43	RCGSSR10527	(CTT)7	F: GCAGAGCAATTCCACATCAT R: GTCGAGCGTTTAAAACAAGG	281
44	RCGSSR10548	(TGC)7	F: TTGGTTGCTGCATACTCTGA R: AACCAACAGCAGTCACCAGT	256
45	RCGSSR12022	(CTT)7	F: CTTCTGTTGCTGCTGCTTCT R: CAAGATCCAACGACCAAATC	276

*S.No. 1 to 8 are EST-SSRs and S.No. 9 to 45 are genomic SSRs

Supplementary Table 3. Mean and standard deviations of the logarithm of the probability of data for different number of sub population (k) tested in the core subset

K	Mean LnP(k)	S.D LnP(k)	Ln'(k)	|Ln''(K)|	Delta K
1	-6836.65	0.28	-	-	-
2	-6511.93	2.76	324.71	83.16	30.12
3	-6270.38	1.55	241.55	17.14	11.07
4	-6045.97	3.45	224.41	79.69	23.08
5	-5901.25	3.91	144.72	17.44	4.46
6	-5773.97	19.99	127.28	25.42	1.27
7	-5672.11	39.11	101.86	19.11	0.49
8	-5551.13	5.10	120.97	57.73	11.32
9	-5487.89	25.81	63.25	25.87	1.00
10	-5398.77	20.54	89.12	19.93	0.97
11	-5289.72	17.19	109.05	36.35	2.11
12	-5217.02	15.85	72.70	7.73	0.49
13	-5152.05	44.63	64.97	56.27	1.26
14	-5143.34	297.32	8.71	108.40	0.36
15	-5026.23	163.23	117.11	-	-

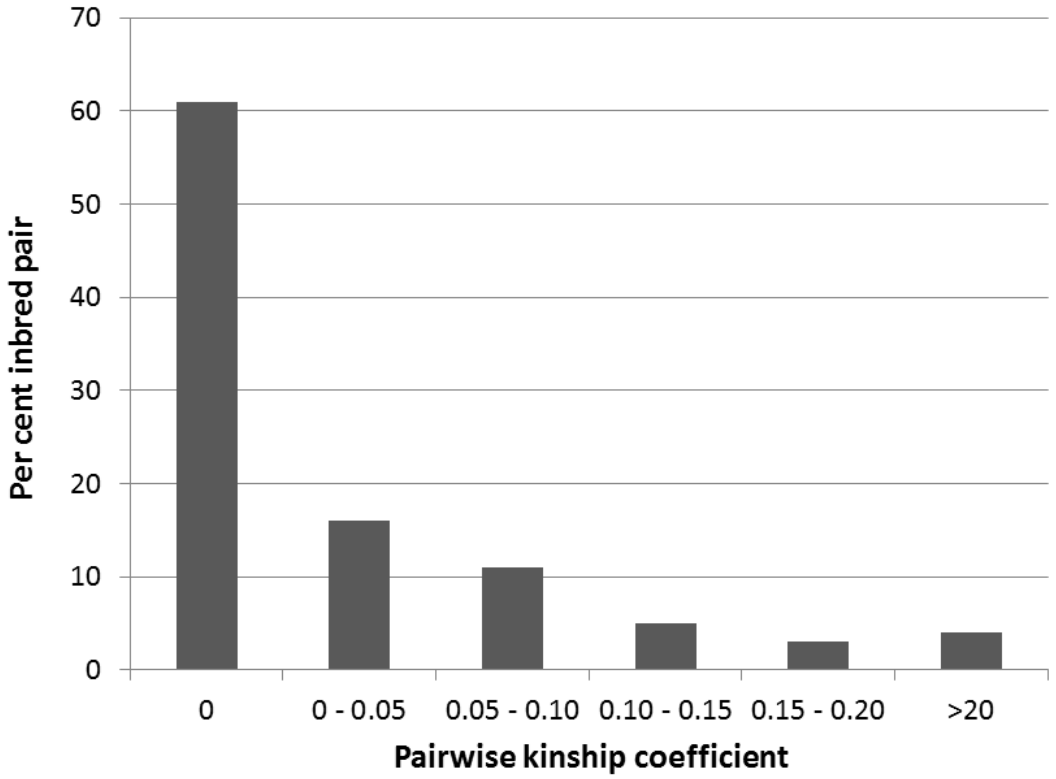


Fig. 1 Distribution of pairwise kinship coefficients among the pairs of inbred lines calculated from SSR genotypic data

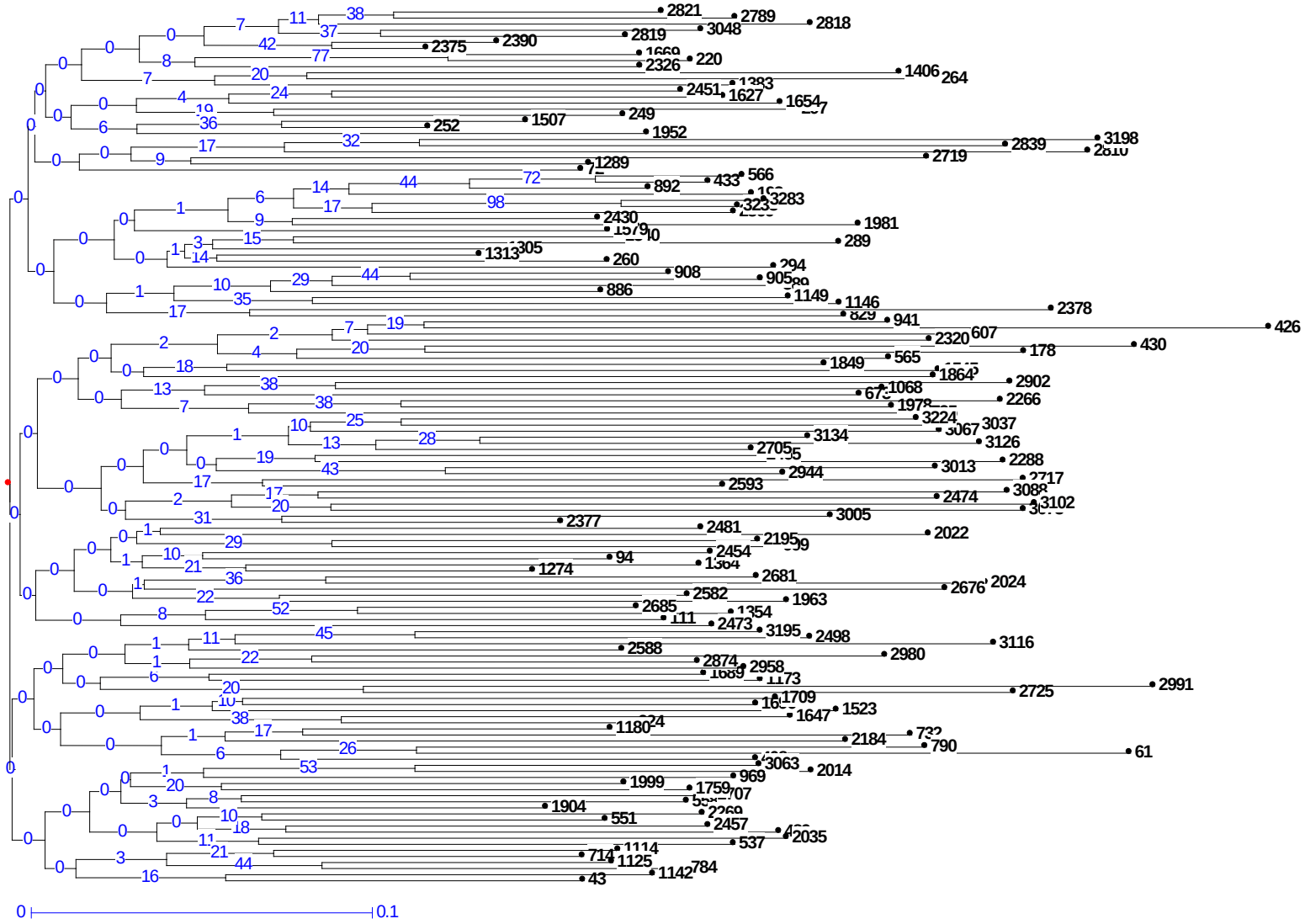
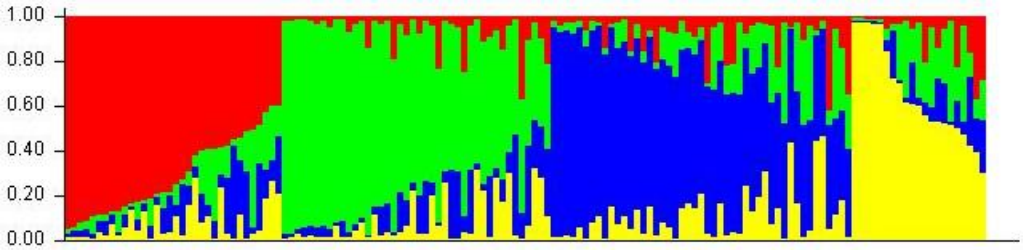


Fig. 2 Neighbour joining tree showing relationship of 144 inbred lines derived from the castor core germplasm

A

B



C

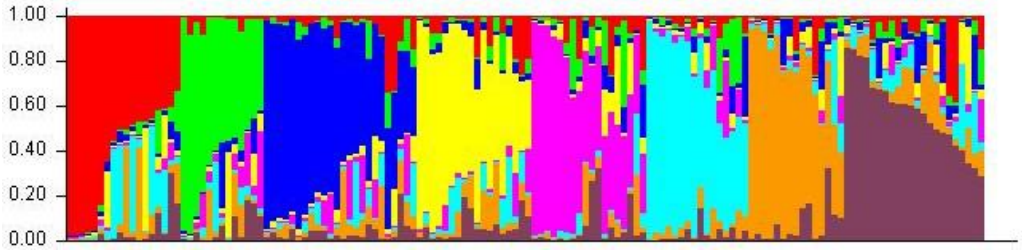
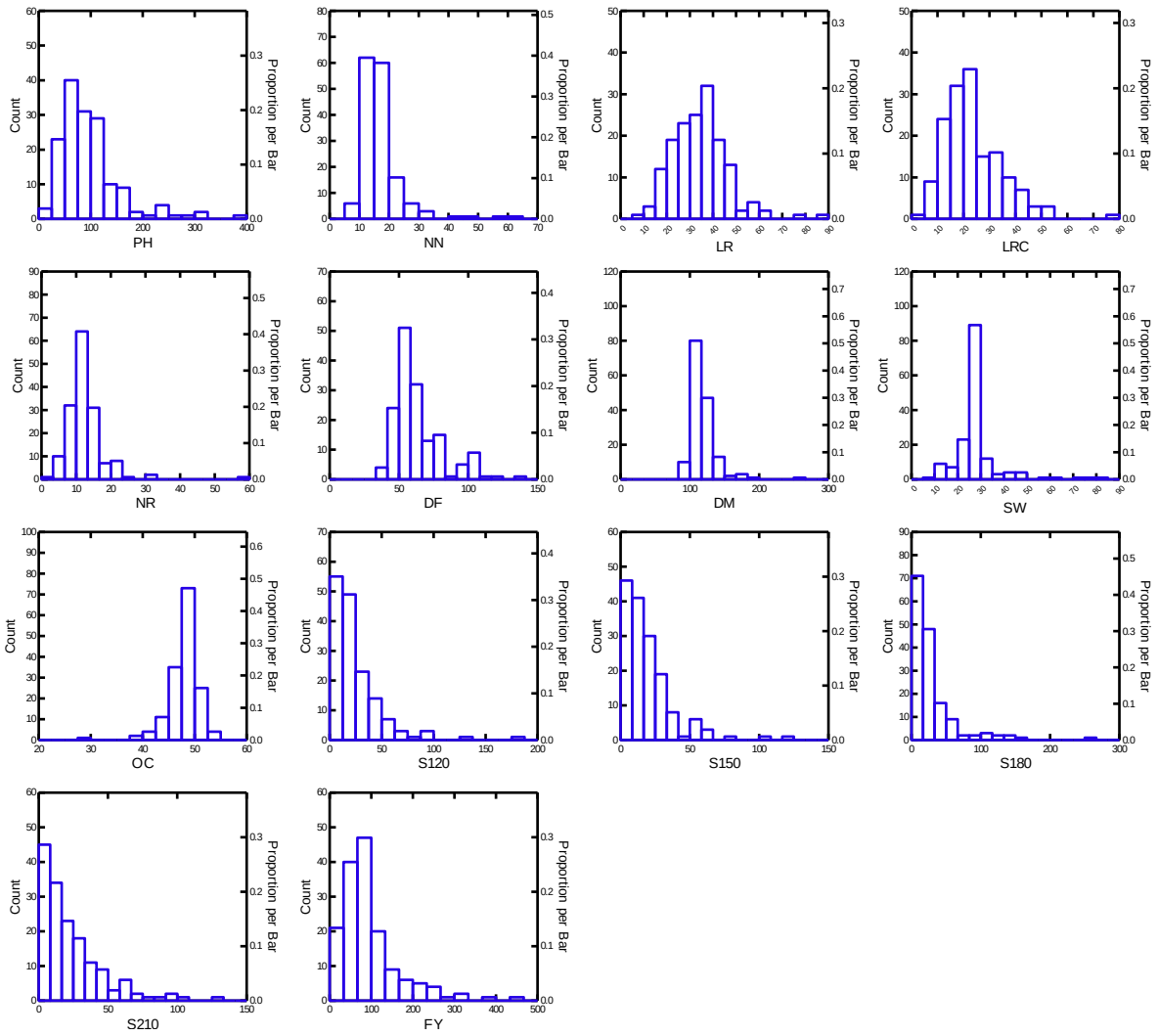
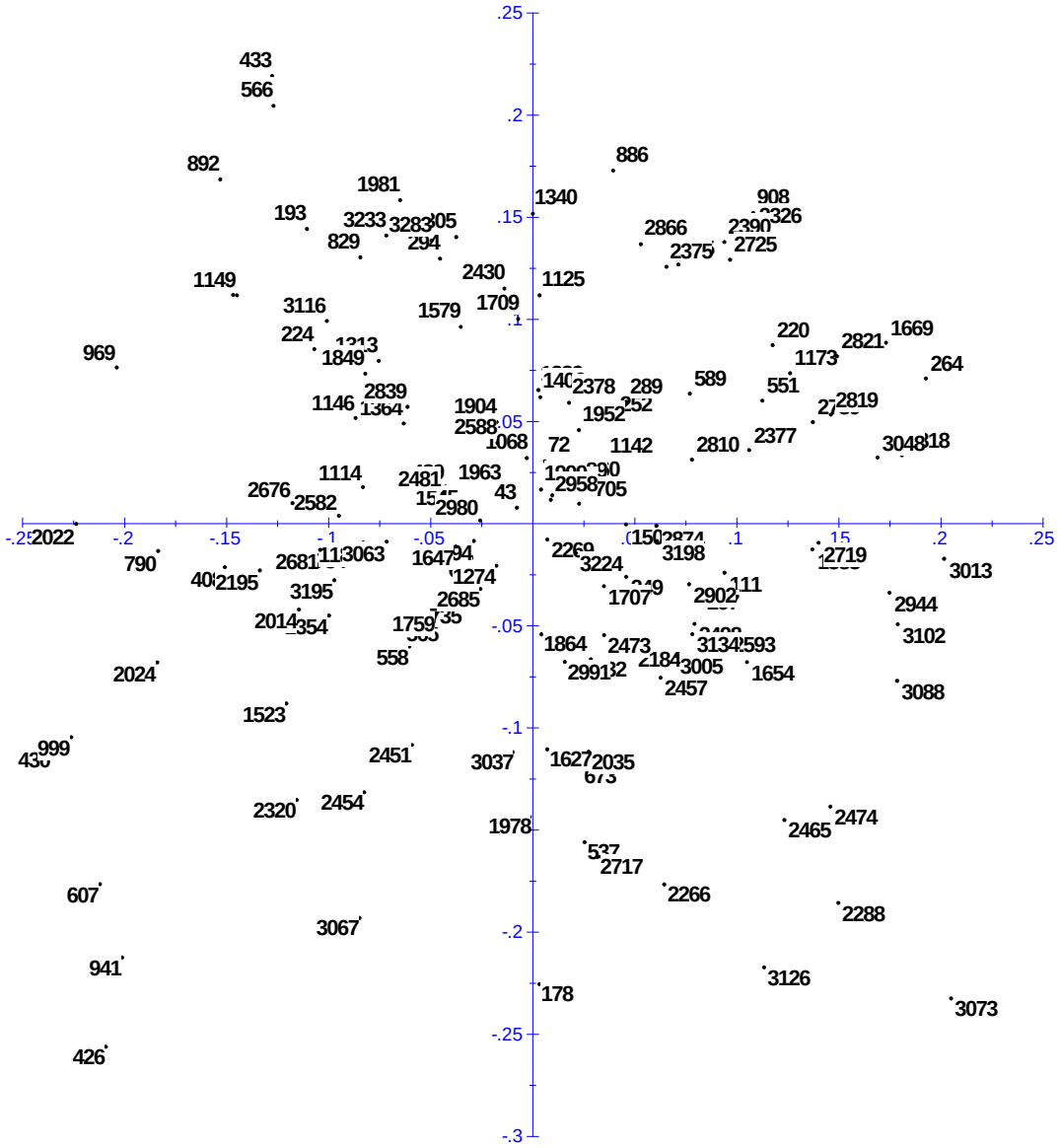


Fig. 3 Inferred population structure in collection of the inbred lines as per the model-based program STRUCTURE for different sub populations (K): (A) K=2, (B) K=4, (C) K=8 (Each single vertical line represents an inbred line and different colours represent different sub populations. The length of the coloured segment is the estimated membership probability of inbred line in the corresponding sub population)

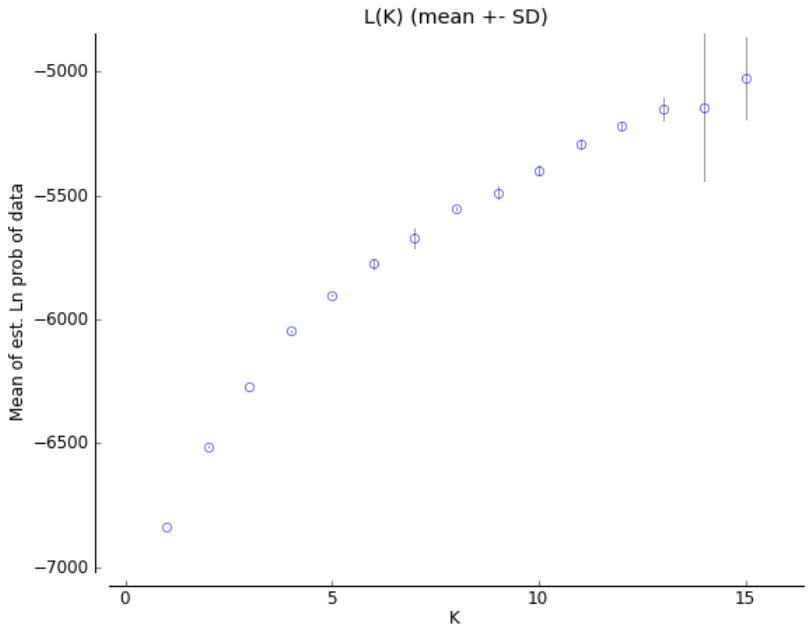


Supplementary Figure 1. Frequency distribution of morphological traits in the castor core set (PH: plant height up to primary raceme, NN: number of nodes up to primary raceme, LR: total length of primary raceme, LRC: length of primary raceme covered by capsules, NR: total number of racemes per plant, DF: days to flowering, DM: days to 50% maturity, SW: 100-seed weight, OC: oil content, S120: seed yield at 120 days after sowing, S150: seed yield at 150 days after sowing, S180: seed yield at 180 days after sowing, S210: seed yield at 210 days after sowing, FY: total seed yield)

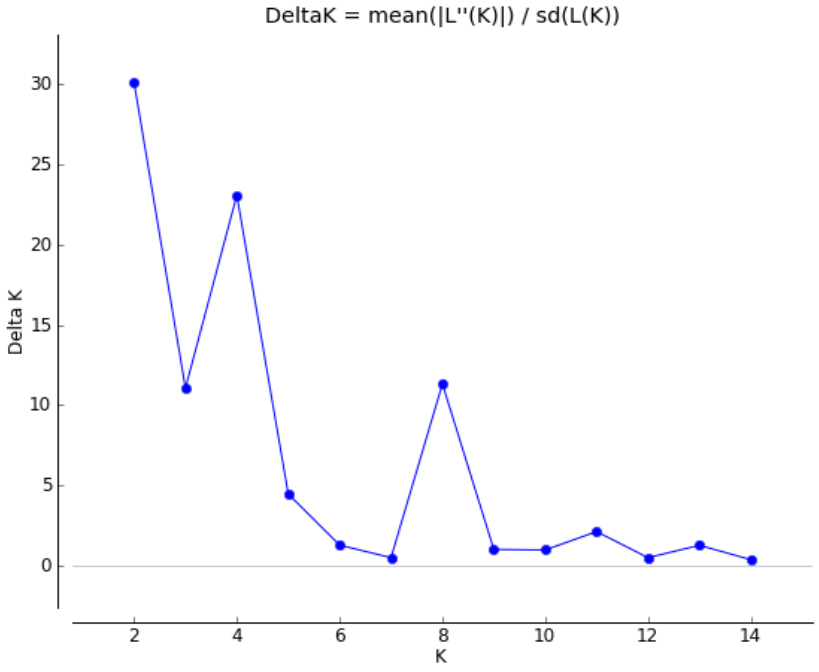


Supplementary Figure 2. Principal coordinate analysis (PCoA) of castor core sub set: Axes-1 (5.6%) and Axes-2 (5.2%) did not separate the genotypes into major groups

A



B



Supplementary Figure 3. Determination of the optimal value of K for the genotype panel: (A) Log probability of data, L(K) averaged over the replicates (B) Plot of Delta K calculated as the mean of the second-order rate of change in likelihood of K divided by the standard deviation of the likelihood of K as per Evanno et al. (2005)