Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the U.S. collection assessed with EST-SSR markers

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0116.R1
Manuscript Type:	Article
Date Submitted by the Author:	21-Aug-2016
Complete List of Authors:	Wang, Mingli; USDA-ARS, PGRCU Dzievit, Matthew; Iowa State University, Department of Agronomy Chen, Zhenbang; University of Georgia, Department of Crop and Soil Sciences Morris, John; USDA-ARS , PGRCU Norris, Jessica; Emory University Barkley, Noelle; USDA-ARS, PGRCU Tonnis, Brandon; USDA-ARS, PGRCU Pederson, Gary; USDA-ARS, PGRCU Yu, Jianming; Iowa State University, Department of Agronomy
Keyword:	Castor germplasm, EST-SSR marker, genetic diversity, population structure, core collection establishment

SCHOLARONE™
Manuscripts

M. L. Wang^{1*}, M. Dzievit², Z. Chen³, J. B. Morris¹, J. E. Norris⁴, N. A. Barkley¹, B. Tonn¹, G. A. Pederson¹ and J. Yu^{2*}

Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the U.S. collection assessed with EST-SSR markers

¹USDA-ARS

Plant Genetic Resources Conservation Unit
1109 Experiment Street
Griffin, GA 30223
USA

²Department of Agronomy

Iowa State University
2014 Agronomy Hall
Ames, IA 50011
USA

³Department of Crop and Soil Sciences

University of Georgia
1109 Experiment Street
Griffin, GA 30223
USA

⁴Emory University Hospital

Emory University
1364 Clifton Road
Atlanta, GA 30322
USA

Corresponding author:

*: e-mail: mingli.wang@ars.usda.gov
Tel: (770) 229-3342

*: e-mail: jmyu@iastate.edu
Tel: (515) 294-2757

Abstract

Castor is an important oilseed crop and although its oil is inedible, it has multiple industrial and pharmaceutical applications. The entire U.S. castor germplasm collection was previously screened for oil content and fatty acid composition, but its genetic diversity and population structure has not been determined. Based on the screening results of oil content, fatty acid composition, and country origins, 574 accessions were selected and genotyped with 22 polymorphic EST-SSR markers. The results from cluster analysis, population structure, and principal component analysis were consistent, and partitioned accessions into four subpopulations. Although there were certain levels of admixtures among groups, these clusters and subpopulations aligned with geographic origins. Both divergent and redundant accessions were identified in this study. The U.S. castor germplasm collection encompasses a moderately high level of genetic diversity (pairwise dissimilarity coefficient = 0.53). The results obtained here will be useful for choosing accessions as parents to make crosses in breeding programs and prioritizing accessions for regeneration to improve germplasm management. A subset of 230 accessions was selected and will be planted in the field for establishing a core collection of the U.S. castor germplasm. Further evaluation of the U.S. castor germplasm collection is also discussed.

Key words: Castor germplasm, EST-SSR marker, genetic diversity, population structure, core collection establishment.

Introduction

Castor (*Ricinus communis* L.) belongs to the Euphorbiaceae family and has been cultivated earlier than 4,000 B.C. as indicated by the castor seeds buried in Egyptian tombs (Purseglove 1981; Morris et al. 2011). Originated in tropical Africa, castor was taken to India and China, and is now grown worldwide from temperate to tropical zones (Martin and Leonard 1967; Purseglove 1981). Castor oil has multiple industrial and medicinal applications (Morris 2004). It has been used as an efficient lubricant for high-speed engines such as airplanes and racing cars and as an ingredient in shampoo, soap, hand lotion, shoe polish, candles, carbon paper, lipstick, and ointments. In medicinal applications, castor oil has been used since ancient times as a popular stimulant laxative and oil inducer for women to start labor (Ogunniyi 2006), as eye-drops for lipid-deficient dry eyes (Goto et al. 2002), and as an ingredient for reducing edema, scabbing of patient wounds, and skin disorders (Khanal et al. 2007). Castor oil can also be used as feedstock for biodiesel production, but it is more valued by the chemical industry than the petroleum industry (César and Batalha 2010; Severino et al. 2012).

The U.S. castor germplasm collection is maintained at Griffin, GA by the USDA-ARS, Plant Genetic Resources Conservation Unit (PGRUC). To determine the variability in oil content and fatty acid composition, the entire castor collection was analyzed using nuclear magnetic resonance (NMR) and gas chromatography (GC) (Wang et al. 2011), but the genetic diversity and population structure of the U.S. collection have not been assessed. In addition, a U.S. castor core collection has not been established. Determining the genetic diversity and population structure of the U.S. collection will be useful for establishing a U.S. castor core collection to facilitate further research.

Castor is a diploid species ($2n = 2x = 20$) with a relatively small nuclear genome size ($1C = 323$ Mb) (Armstrong and Earle 1991; Bennett and Leitch 1995). Recent draft genome sequence of castor (Chan et al. 2010) indicated that 50% of the haploid nuclear genome (~350 Mb) is repetitive sequences. Some candidate genes involved in ricin and oil biosynthesis were also identified by the whole genome and transcriptome sequencing (Chan et al. 2010; Brown et al. 2012). Before the draft genome sequence, genetic diversity of castor germplasm was assessed on 41 germplasm accessions using simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers, but no clear geographic structure was identified (Allan et al. 2008). Subsequently, a small set of genomic SSR markers were developed and used to assess the genetic diversity of 38 accessions from the Brazil castor germplasm collection (Bajay et al. 2009; 2011). After the genome sequence (Chan et al., 2010), single nucleotide polymorphism (SNP) markers were developed and became available. A set of 48 SNP markers were used for genotyping 488 castor germplasm samples from 45 countries. A low level of genetic diversity was revealed and the germplasm samples were classified into five distinct groups (Foster et al. 2010). From the expressed sequence tag (EST) database, 118 polymorphic EST-SSR markers were developed and used for genotyping 24 castor germplasm accessions. These accessions were classified into five distinct clusters showing the geographic pattern across genetic diversity centers of castor (Qiu et al. 2010). Recently, 144 near inbred lines derived from the India castor core collection through selfing and single seed descent were genotyped with 39 SSR markers and classified into four clusters (Senthilvel et al. 2016). Because 136

lines were from India and the remaining eight were from six other countries, this collection does not fully represent the worldwide distribution or country origin of castor accessions.

There are six major centers for castor germplasm collection in the world holding from 11,300 to 18,700 accessions (Anjani 2012; Severino et al. 2012). There may be many redundant accessions in these collections due to multiple germplasm introductions and exchanges. Identifying redundant accessions in the collection is a major challenge for germplasm management, especially for the establishment of a core-collection. A castor core collection (normally representing about 10% of the entire collection) has not been established for the U.S. castor germplasm collection to date. A core collection with large genetic variability and well representation can be established based on genotypic data, rather than phenotypic data (Hu et al. 2000). Since EST-SSR markers were developed from the transcribed gene regions (theoretically transcribed regions more conserved than non-transcribed regions), these markers may have high distinguishing power for discerning genetic diversity and identifying redundant accessions within the species (Wang et al. 2006; Wang et al. 2009a). Compared with using SNP markers to assess genetic diversity, using EST-SSR markers may cost less and be easier to access for a lab that does not have high throughput genotyping equipment. We took advantage of the developed EST-SSR markers (Qiu et al. 2010; Pranavi et al. 2011) and used them to evaluate the U.S. castor germplasm collection. The objectives of this study were to (1) assess the genetic diversity and population structure of the U.S. castor germplasm collection; (2) determine the relationships between genetic diversity, population structure, and geographic origins; and (3) choose a subset of divergent accessions to help establish the U.S. castor core collection via further morphological observation and analysis.

Materials and Methods

The entire USDA castor germplasm collection (1,033 accessions available) was previously screened and evaluated for the variability in oil content, fatty acid composition, and seed weight by our laboratory (Wang et al. 2011). The screening results showed that many accessions had similar oil content and fatty acid profile and were collected from the same countries in the same collection trips as evident by their continuous PI numbers. This would suggest that some redundant accessions were collected. In order to avoid genotyping those redundant accessions, 574 accessions were selected based on the screening results and country origins (Table S1), and used in this experiment to assess the genetic diversity and population structure. The range and representation of oil content and country origins before and after selection were compared and shown in Figure S1. There was no significant difference in the range and sample distributions for the oil content between these two panels. In the subset, the accessions represented 48 countries and the main countries were United States (317, 55.2%), India (60, 10.5%), Iran (41, 7.1%), Brazil (32, 5.6%), South Africa (27, 4.7%), Argentina (24, 4.2%), Peru (13, 2.3%), Pakistan (8, 1.4%), and other countries (52, 9.1%).

DNA was extracted from the leftover tissue (~150 mg) from three seeds that had been used for fatty acid analysis. The intention for this arrangement was to make sure the generated DNA data are reliable for future association analysis with oil and fatty acid traits. Since castor seed contains toxic ricin and allergen proteins (Vietta and Uhr 1985; Olsnes and Sandvig 1988), DNA isolation was carefully conducted in the laboratory fume hood (Two River, WI) by using E.Z.N.A. Plant DNA kit from Omega Bio-Tek (Doraville,

Georgia, USA). Extracted DNA was diluted in 0.1 X TE (1 mM Tris, 0.1 mM EDTA, pH 8.0) to 10 ng / μ l for PCR.

Polymorphic EST-SSR markers were developed (Qiu et al. 2010) from castor EST sequences. We took advantage of the publically available EST-SSR markers and, and used them to screen eight selected germplasm accessions. Among 118 EST-SSR markers, 22 were polymorphic and easy to score (Table 1) and used for genotyping 574 selected castor germplasm accessions. All PCR reactions, program, and product separation on a Mega-gel system (C-DASG-400-50, CBS Scientific Co., Del Mar, CA) followed the method described by Wang et al. (2009b).

Strong clear bands on the gel images were scored using bins and as either present (1) or absent (0). The data were entered into a binary matrix to determine the number of alleles present for each SSR marker. Accessions were coded according to which allele the accession possessed, which created a 574 x 22 data matrix for analysis.

Population structure for the 574 accessions was determined by using the program STRUCTURE, version 2.3.4 (Pritchard et al. 2000). Ten replications were run for each subpopulation number ($k = 1\sim 20$). The admixture model was used, with a burn-in period of 100,000, and 100,000 MCMC reps after burn-in. Result files were uploaded to Structure Harvester (Dent and vonHoldt 2012), which uses the Evanno method to determine the optimal number of subpopulations. From the ten replications for each k , the replication with the highest likelihood was chosen to extract membership coefficients for population structure assignment.

PowerMarker version 3.25 (Liu and Muse 2005) was used to calculate molecular diversity statistics. Darwin version 6.0.13 (Perrier and Jacquemoud-Collet 2006) was used to calculate a pairwise dissimilarity coefficient matrix using simple matching to quantify the genetic diversity and an unweighted neighbor-joining tree with 100 replications of bootstrapping. One percent of the genotype data were missing, therefore, the dissimilarity matrix was used for principal component analysis (PCA). The first three principal components were extracted for analysis.

Results and discussion

Profile of SSR markers

In total, 22 EST-SSR markers were successfully used for genotyping 574 selected castor accessions. Sixty-six alleles were revealed with an average number of 3 alleles per marker ranging from 2 to 6 alleles and 0.57 for major allele frequency. The number of alleles per SSR marker was consistent with previously published results (3.1 ranging from 2 to 7, Allan et al. 2008; 3.3 ranging from 2 to 5, Baja et al. 2009; 3.0 ranging from 2 to 6, Qiu et al. 2010; 3.0 ranging from 2 to 5, Senthilvel et al. 2016). The mean polymorphism information content (PIC) was 0.44 (Table 2), the same as reported by Senthilvel et al. 2016, similar to 0.403 reported by Allan et al. (2008) but slightly higher than 0.36 reported by Qiu et al. (2010). The minor difference of the estimated PIC values may be explained by different sets of SSR markers assessed on different numbers of accessions in separate studies. A moderately high level of genetic diversity (0.53) was revealed from this study and comparable with other studies (0.60, Allan et al. 2008; 0.42, Qiu et al. 2010; 0.38, Senthilvel et al. 2016). More accessions were assessed in this study than the previous studies.

Population structure and genetic diversity

Structure analysis revealed four subpopulations with certain levels of admixture between subpopulations: G1, G2, G3, and G4 shown in red, green, blue, and yellow on Fig. 1a and each corresponding subpopulation contained 157, 108, 161, and 148 accessions, respectively (Table 3). The number of subpopulations ($k = 4$) was determined from the Structure Harvester output and Delta k analysis (Fig. S2). The level of genetic diversity within G2 (0.54) and G1 (0.49) was higher than that within G3 (0.39) and G4 (0.39) (Table 2). The genetic distance among these four subpopulations was measured by Nei's minimum distance. The genetic distance between G2 and G3 (0.25) was the largest and the one between G3 and G4 (0.12) was the smallest (Table 4). The results from principal component analysis (Figure 2) showed that the G1 (red) and G3 (blue) subpopulations were well separated by principal component 1 (PC1). The G3 (blue) and G4 (yellow) subpopulations were well separated by principal component 2 (PC2). The G4 (yellow) and G2 (green) subpopulations were well separated by principal component 3 (PC3). Overall, the results from STRUCTURE analysis and principal component analysis were consistent.

The UPGMA (unweighted pair group method with arithmetic average) tree analysis clustered 574 accessions into four branches (B1, B2, B3, and B4) using EST-SSR marker data (Fig. 1b). The results from UPGMA tree analysis were generally consistent with the results from STRUCTURE analysis. As expected from the algorithm difference between clustering analysis and STRUCTURE analysis, a few accessions were placed in different subpopulations or branches.

Population structure with collection site or conservation location

The results of population structure and country origin from the passport data in Germplasm Resources Information Network (GRIN) are summarized in Table S1 and projected onto the world map (Figure 3). Since most of the accessions (over 55%) in this subset were obtained from the United States, it is understandable that the accessions from the U.S. may be classified into different subpopulations. However, the number of accessions assigned into each subpopulation was significantly different. The number of the U.S. accessions in each subpopulation G1, G2, G3, and G4 were 41 (26.1%), 6 (5.6%), 154 (95.7%), and 116 (78.4%), respectively. Although these accessions were conserved in the U.S., this does not necessarily mean that all these accessions were collected from the U.S. Some of these accessions may be obtained from different donations. For example, some American researchers went to other countries for castor germplasm collection. Once finishing their studies, they donated their collection from the trip to the USDA germplasm collection. These accessions might be recorded with the researcher's affiliated University (such as Texas A&M University). However, these accessions were not originally collected from the U.S. There were 161 accessions in subpopulation G3 (blue dots on Figure 3) with 154 accessions from the U.S. and only 7 accessions from other countries. Therefore, the G3 subpopulation represented the genetic diversity from 26.8% (154/574) of the U.S. conservation. In subpopulation G4 (yellow diamonds on Figure 3), there were 148 accessions with 116 from the U.S., 12 from India, 6 each were from Peru and Argentina, and 5 from other individual countries. The G4 subpopulation mainly represented another type of genetic diversity from 20.2% (116/574) of the U.S. conservation but also contained accessions from India, Peru, and Argentina. In subpopulation G1 (red squares on Figure 3), there were 157 accessions with 41 from the U.S., 31 from India, 24 from Brazil, 17 from

South Africa, 6 from Pakistan, and 44 from other countries. The G1 subpopulation mainly represented the genetic diversity from the U.S., India, Brazil, and South Africa accessions. In subpopulation G2 (green triangles on Figure 3), there were 108 accessions with 37 from Iran, 14 each from Argentina and India, 8 from South Africa, 6 each from the U.S. and Peru, and 23 from other countries. The G2 subpopulation mainly represented the accessions from Iran, Argentina, and India. This subpopulation contained the higher genetic diversity (0.54) than other subpopulations (0.49, 0.39, and 0.39 in Table 2). Among these four subpopulations, the country origins for some accessions may be not accurate due to germplasm donations, multiple introductions, and exchanges from country to country.

In addition to nuclear DNA markers, non-nuclear SNP markers had also been developed from the chloroplast DNA sequences. Using 83 chloroplast SNP markers, 894 accessions had been classified into two main clades and four subclades (Rivarola et al. 2011). In the future, non-nuclear SNP markers may be used for genotyping the accessions within the U.S. castor germplasm collection. New classification (based on the non-nuclear SNPs) may help to clarify the country origin issues for some of the accessions.

Identification and selection of divergent accessions for the core collection establishment

There were 1,033 accessions in the U.S. castor germplasm collection. According to our chemical analysis results (including seed weight, oil content, and fatty acid composition), 574 accessions were selected and genotyped. From genetic analysis, these 574 accessions were classified into four subpopulations (Table S1). Each subpopulation contained different levels of genetic diversity. The genetic diversity level was not proportionally correlated with the number of accessions in each group. For example, G2 and G3 contained 108 and 161 accessions, respectively. Although G2 contained fewer accessions than G3, its genetic diversity (0.54) was much higher than G3 (0.39). For establishing core collection, we have to reduce the number of accessions (from 100% to 10%). In this process, we need to consider both the number of accessions and level of genetic diversity in each subpopulation. In theory, 103 accessions (10% of the original size) can be selected to represent the genetic diversity of the entire U.S. collection. Castor is an oilseed crop and its seed contains a high percentage of oil. Long-term storage of high oil seeds may lead to a low percentage of seed germination. After further considering of the potential reduction of viable accessions due to the declining germination rate, we selected 230 (about 20%) from the assessed accessions to form a core collection. The selected accessions are listed in Table S2. Cluster analysis was also performed on this selected set of 230 accessions and the result is shown in Fig. S3. The patterns of genetic diversity in two sets (574 versus 230 accessions) are similar. These selected accessions will be planted in the field for germination test and morphological observation. Based on germination test, field observation, and genetic diversity, the U.S. castor core collection will be established.

In addition, we also plan to obtain a critical piece of information for the selected set: ricin and *Ricinus communis* agglutinin (RCA) content in castor seeds. Reducing these toxic compounds in castor seeds is one of the main objectives in castor breeding programs. Previous research using the U.S. castor germplasm collection (263 accessions) identified significant variation in the amount of ricin and RCA (1.9 mg/g – 16.0 mg/g) in castor seeds (Pinkerton et al. 1999). Our next urgent research objective is to quantify the ricin and RCA

content in seeds of these 230 accessions and then to take the ricin and RCA content into consideration when the new final core collection is to be established.

Acknowledgements

The authors gratefully thank Drs. Paul Raymer and Jack Huang for useful suggestions and comments on the manuscript, Mr. David Pinnow and Brice McEver for assisting in DNA extraction and gel electrophoresis.

References

- Allan, G., Williams, A., Rabinowicz, P.D., Chan, A.P., Ravel, J., and Kleim, P. 2008. Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLP and SSRs. *Genet. Resour. Crop Evol.* **55**: 365-378.
- Anjani, K. 2012. Castor genetic resources: A primary gene pool for exploitation. *Ind. Crops Prod.* **35**: 1-14.
- Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**: 208-218.
- Bajay, M.M., Pinheiro, J.B., Batista, C.E.A., Nobrega, M.B.M., and Zucchi, M.I. 2009. Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. *Conservation Genet. Resour.* **1**: 237-239.
- Bajay, M.M., Zucchi, M.I., Kiihl, T.A.M., Batista, C.E.A., Monteiro, M., and Pinheiro, J.B. 2011. Development of a novel set of microsatellite markers for castor bean, *Ricinus communis* (Euphorbiaceae). *American J. Botany* **98**: e87-e89.
- Bennett, D.D., and Leitch, I.L. 1995. Nuclear DNA amounts in angiosperms. *J. Botany* **76**: 113-176.
- Brown, A.P., Kroon, J.T., Swarbreck, D., Febrer, M., Larson, T.R., Graham, I.A., Caccamo, M., and Slabas, A.R. 2012. Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol liquid biosynthetic pathways. *PLoS One* **7**: e30100.
- César, A., and Batalha, M.O. 2010. Biodiesel production from castor oil in Brazil: A difficult reality. *Energy Policy* **38**: 4031-4039.
- Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Frase-Liggett, C.M., Ravel, J., and Rabinowicz, P.D. 2010. Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnol.* **28**: 951-956.

- Dent, A.E., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER; a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resour.* **4**: 359-361.
- Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Foster, J.T., Allan, G.J., Chan, A., Rabinowicz, P.D., Ravel, J., Jackson, P.J., and Kleim, P. 2010. Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). *BMC Plant Biol.* **10**: 13.
- Goto, E., Shimazaki, J., Monden, Y., Takano, Y., Yagi, Y., Shimmura, S., Tsuba, K. 2002. Low-concentration homogenized castor oil eye drops for noninflamed obstructive meibomian gland dysfunction. *Ophthalmology* **109**: 2030-2035.
- Hu, J., Zhu, J., and Xu, H.M. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **101**: 264-268.
- Khanal, S., Tomlinson, A., Pearce, E.I., Simmons, P.A. 2007. Effect of an oil-in-water emulsion on the tear physiology of patients with mild to moderate dry eye. *Cornea* **26**: 175-181.
- Liu, K., and Muse, S.V. 2005. PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* **21**: 2128-2129.
- Martin, J.K. and Leonard, W.H. 1967. Principles of field crop production. Macmillan, New York.
- Morris, B. 2004. Phytochemical traits in the genetic resources of castorbean. *Current Topics Plant Biol.* **5**: 63-67.
- Morris, J.B., Wang, M.L., and Morse, S.A. 2011. *Ricinus*. In Wild crop relatives: Genomic and breeding resources, oilseeds. *Edited by C. Kole*. Springer, Heidelberg, Germany. pp. 251-260.
- Ogunniyi, D.S. 2006. Castor oil: a vital industrial raw material. *Bioresour. Technol.* **97**: 1086-1091.
- Olsnes, S. and Pihl, A. 1973. Different biological properties of the two constituent peptide chains of ricin, a toxic protein inhibiting protein synthesis. *Biochemistry* **12**: 3121-3126.
- Olsnes, S., and Sandvig, K. 1988. How protein toxins enter and kill cells. In *Immunotoxins*. *Edited by A.E. Frankel*. Kluwer, Boston, MA. pp 39-73.

- Page, R.D.M. 1996. TREEVIEW. Version 1.6.6. An application to display phylogenetic trees on personal computers. *Computer Appl. Biosci.* **12**: 357-358.
- Perrier, X., Jacquemoud-Collet, J.P. 2006. **DARwin** software <http://darwin.cirad.fr/>
- Pinkerton, S.D., Rolfe, R., Auld, D.L., Ghetie, V., and Lauterbach, B.F. 1999. Selection of castor for divergent concentrations of ricin and *Ricinus communis* agglutinin. *Crop Sci.* **39**: 353-357.
- Pranavi, B., Sitaram, G., Yamini, K.N., and Kumar, V.D. 2011. Development of EST-SSR markers in castor bean (*Ricinus communis* L.) and their utilization for genetic purity testing of hybrids. *Genome* **54**: 684-691.
- Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. 2000. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**: 170-181.
- Purseglove, J.W. 1981. Tropical crops, dicotyledons. Longman, Essex, U.K. pp 1-719.
- Qiu, L., Yang, C., Tian, B., Yang, J.B., and Liu, A. 2010. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol.* **10**: 278.
- Rivarola, M., Foster, J.T., Chan, A.P., Williams, A.L., Rice, D.W., Liu, X., Melake Berhan, A., Creasy H.H., Puiu D., Rosovitz M.J., Khouri H.M. Beckstrom Sternberg S.M., Allan G.J., Keim P., Ravel J., Rabinowicz P.D. 2011. Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLoS One* **6**: e21743.
- Senthilvel, S., Shaik, M., Anjani, K., Shaw, R.K., Kumari, P., Sarada, C., Kiran, B.U. 2016. Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of castor. *J. Plant Biochem. Biotechnol.* DOI 10.1007/s13562-016-0356-8.
- Severino, L.S., Auld, D.L., Baldanzi, M., Cândido, M.J.D., Chen, G., Crosby, W., Tan, D., He, X., Lakshamma, P., Lavanya, C., Machado, O.L.T., Mielke, T., Milani, M., Miller, T.D., Moris, J.B., Morese, M.S., Navas, A.A., Soares, D.J., Sofiatti, V., Wang, M.L., Zanutto, M.D., and Zieler, H. 2012. A review on the challenges for increased production of castor. *Agronomy J.* **104**: 853-880.
- Vietta, E.S., and Uhr, J.W. 1985. Immunotoxins: redirecting nature's poisons. *Cell* **41**: 653-654.
- Wang, M.L., Barkley, N.A., and Jenkins, T.M. 2009a. Microsatellite markers in plants and insects. Part I: Applications and biotechnology. *Genes Genomes Genomics* **3**: 54- 67.
- Wang, M.L., Morris, J.B., Tonniss, B., Pinnow, D., Davis, J., Raymer, P., and Pederson, G.A. 2011. Screening of the entire USDA castor germplasm collection for oil content

and fatty acid composition for optimum biodiesel. J. Agric. Food Chem. **59**: 9250-9256.

Wang, M.L., Mosjidis, J.A., Morris, J.B., Dean, R.E., Jenkins, T.M., and Pederson, G.A. 2006. Genetic diversity of *Crotalaria* germplasm assessed through phylogenetic analysis of EST-SSR markers. Genome **49**: 707-75.

Wang, M.L., Zhu, C., Barkley, N.A., Chen, Z., Erpelding, J.E., Murray, S.C., Tuinstra, M.R., Tesso, T., Pederson, G.A., and Yu, J. 2009b. Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor. Appl. Genet. **120**: 13-23.

Draft

Table 1. Information on selected polymorphic EST-SSR markers

EST ID	Code	Forward primer (5'-3')	Reverse Primer (5'-3')	Repeats	Size (bp)	Alleles
EG662973	Rc02	CCTCCATTGGTTTCTCCATTTA	TGATTCGAGATCAGAGTTGACG	(GT) ₉	269	4
EE255832	Rc05	CATCACCTCTATCCATCCGTTT	CCTTCATCATTGAACTCCCTTC	(AC) ₈	200	4
EE259292	Rc16	CATGAACCCCACAATTTACAGA	GACAGACCCATCTCAAGATTCC	(TC) ₁₅	254	4
EE259262	Rc20	CGTTTAAGGATCCGCATATCTC	GCTCAACCACCTCAAGATTTTC	(AG) ₂₀	244	2
GE633816	Rc25	ACCCCTGCAAAACCTAATAAT	TTCGGTGTAAGAATCCGACTT	(CT) ₂₃	259	3
EE260187	Rc35	AAGCGAATTCATACTCCTTCA	AAACCCACGTCTATCCTCAAAA	(AG) ₁₃	238	2
EG698508	Rc45	TATTACAATACCGCCCTATGC	ACCGCTGTTATTCTGTCAAAT	(CAC) ₆	280	2
EG662586	Rc61	GCACTGAGGGTTAATTCTGGAC	GTCGTAGACCACCTTAGCATCC	(GGC) ₇	232	2
EG701975	Rc65	CGTAGCTGAAGCTGAACAAAGA	GCTTGACGGAGAGTAGAGCAAT	(AGA) ₈	261	3
EG696213	Rc76	GGAGGGATATGGATTTAGATG	CGGAAGATGAGAGAACAGGTT	(ATA) ₉	225	3
EG666840	Rc90	CACCTCCACAAGAACAAGAACA	CGGACGACGAACTTTAGATTC	(GAA) ₁₁	245	2
EG699176	Rc95	TAGACCCAGCAGTGCTCATAAA	CCACCTTTAGGTGGATGTTGT	(CAA) ₁₉	235	3
EG693883	Rc158	CCCCGACGATAACTACCATAAA	AGGTCATCGAATAACGACCAAG	(ACC) ₇	224	3
GE632520	Rc167	GTCTGTGACAACAACCACGAAC	GAGATGGGAGAGATGAAGAGGA	(TTC) ₇	167	2
GE633829	Rc176	TACCCAATGACTGCAATCTTCA	GGGTTTGTGAGAAGGAGAAAA	(CTT) ₈	222	3
EG683300	Rc186	AGATCAAATCGGGGAAGAGAAT	CCATCATCAGTAGGTGGCATT	(GGA) ₁₀	175	2
EE258383	Rc193	CTACAAGAACAAAACGCCACCT	CGGACGACGAACTTTAGATTC	(GAA) ₁₁	227	2
GE634515	Rc206	TGAGTTGTGTGCTACTCTTTGGT	CATCGGACGGTGCTAAAATAAT	(CT) ₆	193	2
EG676447	Rc239	CCAACAAATTCCTCCCTCTCT	CCAAGATTGAAGCAAAAACACA	(TC) ₁₃	203	3
EG661770	Rc251	AAACAAACAACACCGAAATCCT	CGTAGCGACTCATTTACACCAA	(CT) ₁₅	187	4
EG662055	Rc258	GCAGTTAGGCATCAAAATCCTC	GAAGTGCAGCTGTTAGGGTTTC	(GAAA) ₅	201	3
EV520791	Rc374	GATCTCGTAATGAATGTTGG	CTCTCTCTTGAGGCTCTTC	(GCT) ₅	143	3

Table 2. Summary statistics for the selected castor accessions and subpopulations detected by structure analysis based on 22 EST-SSR markers

Statistics	Overall	G1	G2	G3	G4
Sample size	574	157	108	161	148
Total number of alleles	66	66	65	64	63
Number alleles per locus	3.00	3.00	2.96	2.91	2.86
Major allele frequency	0.57	0.62	0.56	0.70	0.71
Genetic diversity	0.53	0.49	0.54	0.39	0.39
PIC [†]	0.44	0.41	0.46	0.33	0.33

[†]Polymorphism information content

Table 3. Genetic distance* between castor groups from structure analysis

Group	G1	G2	G3	G4
G1	0	0.22	0.23	0.17
G2		0	0.25	0.24
G3			0	0.12
G4				0

*Nei's minimum distance

Table 4. Number of accessions in each subpopulation.

Country	G1	G2	G3	G4	Total
Afghanistan	0	1	0	1	2
Algeria	2	0	0	0	2
Argentina	3	14	1	6	24
Bahamas	0	1	0	0	1
Benin	0	1	0	0	1
Botswana	0	1	0	0	1
Brazil	24	4	1	3	32
Bulgaria	1	0	0	0	1
Cambodia	0	0	0	0	0
China	2	0	0	0	2
Colombia	3	0	0	0	3
Costa Rica	0	1	0	0	1
Cuba	0	3	1	0	4
Ecuador	1	0	0	0	1
Egypt	2	0	0	0	2
El Salvador	0	0	0	0	0
Ethiopia	0	0	0	0	0
Former Soviet Union	2	2	0	0	4
Greece	1	1	0	0	2
Guatemala	0	0	0	0	0
Hungary	3	1	0	0	4
India	31	14	3	12	60
Indonesia	0	0	0	1	1
Iran	3	37	0	1	41
Israel	1	0	0	0	1
Jamaica	0	0	0	0	0
Jordan	0	1	0	0	1
Kenya	0	0	0	0	0
Madagascar	0	0	0	0	0
Mexico	2	0	0	0	2
Morocco	1	1	0	0	2
Nepal	0	0	0	0	0
Nigeria	0	0	0	1	1
Pakistan	6	2	0	0	8
Panama	0	1	0	0	1
Paraguay	2	0	0	0	2
Peru	1	6	0	6	13
Portugal	2	0	0	0	2
Puerto Rico	0	1	0	0	1
Serbia	0	0	0	0	0
South Africa	17	8	1	1	27
Sri Lanka	0	0	0	0	0
Syria	0	0	0	0	0
Taiwan	1	0	0	0	1
Turkey	1	0	0	0	1
United States	41	6	154	116	317
Uruguay	2	0	0	0	2

Virgin Islands (U.S.)	0	1	0	0	1
Zaire	2	0	0	0	2
Total	157	108	161	148	574

Draft

215x184mm (150 x 150 DPI)

178x113mm (107 x 141 DPI)

216x127mm (107 x 126 DPI)

Supplementary materials

Supplementary Tables

Table S1. Selected 574 accessions from the U.S. castor germplasm collection, country origin, and subpopulation membership coefficients.

Accession number	Country	G1	G2	G3	G4	Gx
217557	India	0.374	0.162	0.254	0.21	G1
217558	India	0.495	0.086	0.27	0.149	G1
223409	Iran	0.391	0.249	0.131	0.23	G1
240673	Uruguay	0.563	0.33	0.043	0.064	G1
240674	Uruguay	0.66	0.267	0.025	0.048	G1
241362	Brazil	0.671	0.259	0.022	0.048	G1
241363	Brazil	0.622	0.229	0.101	0.048	G1
241364	Brazil	0.679	0.262	0.038	0.021	G1
241365	Brazil	0.674	0.259	0.029	0.038	G1
241366	Brazil	0.72	0.201	0.046	0.033	G1
241368	Brazil	0.583	0.197	0.084	0.136	G1
241369	Brazil	0.547	0.326	0.035	0.093	G1
241371	Brazil	0.507	0.323	0.1	0.071	G1
243062	Brazil	0.404	0.255	0.035	0.305	G1
243207	Brazil	0.513	0.096	0.248	0.143	G1
248390	India	0.669	0.043	0.06	0.228	G1
248391	India	0.499	0.06	0.042	0.399	G1
248424	Zaire	0.528	0.12	0.029	0.324	G1
248441	Zaire	0.646	0.311	0.008	0.034	G1
248500	Hungary	0.951	0.025	0.01	0.014	G1
248501	Hungary	0.653	0.233	0.036	0.078	G1
248502	Hungary	0.635	0.208	0.07	0.087	G1
248926	India	0.95	0.015	0.011	0.024	G1
248927	India	0.828	0.031	0.01	0.131	G1
248929	India	0.898	0.022	0.039	0.041	G1
248931	India	0.6	0.286	0.023	0.09	G1
248935	India	0.779	0.016	0.021	0.185	G1
248937	India	0.496	0.067	0.019	0.418	G1
248938	India	0.491	0.083	0.054	0.372	G1
248939	India	0.876	0.008	0.024	0.092	G1
248940	India	0.754	0.006	0.021	0.218	G1
248941	India	0.942	0.032	0.012	0.013	G1
248942	India	0.896	0.006	0.017	0.081	G1
248943	India	0.452	0.03	0.373	0.144	G1
248944	India	0.747	0.016	0.102	0.135	G1
248946	India	0.557	0.258	0.056	0.128	G1
248951	India	0.837	0.009	0.121	0.032	G1

248956	India	0.965	0.01	0.009	0.016	G1
248958	India	0.455	0.043	0.304	0.199	G1
248961	India	0.909	0.015	0.017	0.059	G1
248964	India	0.434	0.059	0.429	0.078	G1
248969	India	0.334	0.332	0.019	0.315	G1
248970	India	0.702	0.021	0.016	0.262	G1
250226	Pakistan	0.884	0.029	0.037	0.05	G1
250397	Pakistan	0.566	0.124	0.177	0.133	G1
250398	Pakistan	0.913	0.06	0.011	0.016	G1
250574	Egypt	0.698	0.256	0.014	0.031	G1
250575	Egypt	0.521	0.44	0.025	0.013	G1
250622	Pakistan	0.597	0.216	0.069	0.118	G1
250742	Iran	0.546	0.21	0.203	0.041	G1
250884	Iran	0.871	0.05	0.041	0.038	G1
252001	Turkey	0.956	0.018	0.008	0.017	G1
253422	Israel	0.95	0.026	0.011	0.013	G1
253620	Morocco	0.961	0.007	0.009	0.023	G1
254405	India	0.527	0.374	0.024	0.075	G1
254406	India	0.601	0.098	0.08	0.221	G1
255238	Mexico	0.929	0.043	0.01	0.017	G1
257442	South Africa	0.72	0.21	0.014	0.057	G1
257443	South Africa	0.668	0.283	0.021	0.029	G1
257444	South Africa	0.839	0.044	0.04	0.077	G1
257446	South Africa	0.57	0.04	0.017	0.372	G1
257447	South Africa	0.818	0.013	0.061	0.108	G1
257448	South Africa	0.833	0.034	0.036	0.097	G1
257449	South Africa	0.537	0.026	0.018	0.419	G1
257457	South Africa	0.548	0.254	0.098	0.1	G1
257458	South Africa	0.769	0.033	0.135	0.063	G1
257459	South Africa	0.567	0.092	0.286	0.055	G1
257461	South Africa	0.735	0.015	0.087	0.162	G1
257653	Former Soviet Union	0.532	0.434	0.012	0.022	G1
258368	Former Soviet Union	0.492	0.486	0.011	0.011	G1
258388	Algeria	0.941	0.035	0.011	0.014	G1
258390	Algeria	0.838	0.011	0.113	0.039	G1
262553	Bulgaria	0.901	0.009	0.049	0.041	G1
265504	United States	0.897	0.022	0.037	0.044	G1
265505	United States	0.87	0.015	0.083	0.031	G1
265506	Colombia	0.821	0.017	0.026	0.136	G1
265507	Colombia	0.764	0.075	0.049	0.112	G1
265508	United States	0.949	0.014	0.017	0.02	G1
265509	Colombia	0.634	0.139	0.078	0.15	G1
267801	Brazil	0.922	0.042	0.013	0.023	G1
267802	Brazil	0.942	0.03	0.008	0.02	G1
267804	Brazil	0.95	0.017	0.014	0.018	G1

267806	Brazil	0.553	0.094	0.025	0.328	G1
267807	Brazil	0.934	0.021	0.03	0.015	G1
267936	India	0.58	0.064	0.322	0.035	G1
271861	Ecuador	0.911	0.017	0.026	0.046	G1
274768	South Africa	0.624	0.357	0.011	0.008	G1
274769	South Africa	0.934	0.037	0.008	0.021	G1
274770	South Africa	0.884	0.014	0.024	0.078	G1
274771	South Africa	0.687	0.059	0.063	0.191	G1
274772	South Africa	0.812	0.133	0.023	0.033	G1
274773	South Africa	0.944	0.011	0.011	0.035	G1
277021	Argentina	0.538	0.052	0.017	0.393	G1
277024	Argentina	0.94	0.012	0.009	0.039	G1
277025	Argentina	0.602	0.204	0.011	0.183	G1
289220	Greece	0.597	0.012	0.02	0.371	G1
298598	Brazil	0.686	0.104	0.029	0.181	G1
298599	Brazil	0.536	0.356	0.024	0.084	G1
298600	Brazil	0.608	0.325	0.043	0.024	G1
298601	Brazil	0.617	0.283	0.027	0.073	G1
306724	Taiwan	0.696	0.283	0.009	0.013	G1
323304	Pakistan	0.489	0.358	0.117	0.036	G1
323589	Portugal	0.818	0.103	0.024	0.055	G1
323591	Portugal	0.953	0.013	0.012	0.022	G1
326198	Mexico	0.461	0.434	0.046	0.058	G1
326343	Paraguay	0.951	0.008	0.029	0.012	G1
326344	Paraguay	0.962	0.014	0.008	0.016	G1
384014	Brazil	0.617	0.295	0.021	0.067	G1
384015	Brazil	0.564	0.295	0.063	0.078	G1
426929	Pakistan	0.882	0.034	0.03	0.054	G1
436592	China	0.866	0.012	0.059	0.064	G1
436593	China	0.777	0.049	0.08	0.094	G1
439590	Brazil	0.811	0.172	0.008	0.009	G1
486318	Peru	0.563	0.306	0.028	0.103	G1
599751	United States	0.502	0.019	0.311	0.167	G1
599752	United States	0.858	0.009	0.103	0.03	G1
631156	United States	0.784	0.055	0.092	0.069	G1
642000	United States	0.46	0.017	0.454	0.069	G1
4685	United States	0.554	0.013	0.413	0.019	G1
4686	United States	0.936	0.017	0.013	0.033	G1
4687	United States	0.67	0.069	0.228	0.032	G1
7352	United States	0.423	0.396	0.022	0.16	G1
7353	United States	0.611	0.011	0.35	0.028	G1
8532	United States	0.632	0.209	0.018	0.141	G1
15620	United States	0.799	0.022	0.166	0.013	G1
15621	United States	0.94	0.024	0.017	0.019	G1
15622	United States	0.956	0.013	0.013	0.017	G1

15623	United States	0.949	0.029	0.009	0.013	G1
15624	United States	0.939	0.027	0.012	0.022	G1
15625	United States	0.94	0.022	0.027	0.011	G1
15627	United States	0.943	0.013	0.032	0.012	G1
42843	United States	0.936	0.036	0.018	0.009	G1
85785	United States	0.766	0.014	0.196	0.024	G1
85789	United States	0.944	0.008	0.029	0.019	G1
85791	United States	0.905	0.011	0.07	0.014	G1
85795	United States	0.926	0.016	0.04	0.018	G1
85804	United States	0.951	0.012	0.02	0.017	G1
85805	United States	0.928	0.01	0.033	0.028	G1
85806	United States	0.934	0.009	0.039	0.018	G1
85808	United States	0.911	0.008	0.066	0.014	G1
85817	United States	0.899	0.009	0.08	0.012	G1
85840	United States	0.949	0.018	0.016	0.018	G1
85856	United States	0.854	0.011	0.045	0.089	G1
85912	United States	0.407	0.224	0.308	0.061	G1
85918	United States	0.511	0.212	0.049	0.229	G1
85919	United States	0.538	0.319	0.034	0.109	G1
85920	United States	0.896	0.035	0.019	0.049	G1
85938	United States	0.518	0.062	0.399	0.021	G1
85988	United States	0.517	0.018	0.442	0.023	G1
86004	United States	0.518	0.017	0.441	0.024	G1
86070	United States	0.441	0.023	0.403	0.133	G1
86307	United States	0.402	0.025	0.245	0.327	G1
183346	India	0.627	0.162	0.179	0.033	G1
243059	Brazil	0.809	0.054	0.075	0.062	G1
243200	Brazil	0.513	0.295	0.035	0.157	G1
248928	India	0.507	0.434	0.02	0.039	G1
254404	India	0.341	0.308	0.092	0.259	G1
208843	Cuba	0.017	0.935	0.028	0.02	G2
208844	Cuba	0.017	0.951	0.018	0.014	G2
209132	Puerto Rico	0.26	0.58	0.109	0.051	G2
209326	Virgin Islands (U.S.)	0.231	0.52	0.054	0.195	G2
209436	Panama	0.032	0.738	0.033	0.197	G2
209623	Cuba	0.031	0.939	0.012	0.017	G2
212115	Afghanistan	0.01	0.95	0.02	0.02	G2
213021	India	0.223	0.427	0.208	0.143	G2
215768	Peru	0.048	0.641	0.046	0.265	G2
215769	Peru	0.037	0.656	0.047	0.26	G2
215771	Peru	0.043	0.588	0.069	0.3	G2
215772	Peru	0.275	0.572	0.053	0.1	G2
215773	Peru	0.069	0.757	0.098	0.076	G2
215774	Peru	0.191	0.516	0.219	0.074	G2
217539	Pakistan	0.072	0.471	0.192	0.265	G2

217559	India	0.297	0.603	0.06	0.041	G2
219766	Botswana	0.237	0.727	0.015	0.021	G2
219767	Argentina	0.036	0.935	0.01	0.018	G2
219768	Argentina	0.021	0.916	0.019	0.044	G2
219769	Argentina	0.011	0.948	0.025	0.017	G2
219771	Argentina	0.319	0.353	0.029	0.299	G2
219772	Argentina	0.026	0.909	0.033	0.033	G2
219773	Argentina	0.271	0.675	0.026	0.028	G2
219774	Argentina	0.059	0.704	0.05	0.188	G2
219775	Argentina	0.036	0.876	0.026	0.062	G2
221048	Argentina	0.02	0.948	0.014	0.018	G2
221049	Argentina	0.068	0.773	0.064	0.095	G2
221051	Argentina	0.032	0.704	0.048	0.215	G2
221053	Argentina	0.02	0.494	0.249	0.238	G2
221054	Argentina	0.052	0.816	0.052	0.08	G2
222265	Iran	0.01	0.961	0.018	0.011	G2
222745	Iran	0.008	0.979	0.008	0.006	G2
222828	Iran	0.006	0.978	0.009	0.007	G2
222829	Iran	0.008	0.981	0.006	0.005	G2
222830	Iran	0.007	0.979	0.006	0.007	G2
223013	Iran	0.287	0.676	0.013	0.023	G2
223408	Iran	0.007	0.957	0.017	0.019	G2
226669	Iran	0.009	0.931	0.043	0.017	G2
226670	Iran	0.073	0.843	0.054	0.029	G2
227869	Iran	0.4	0.501	0.032	0.067	G2
229541	Iran	0.007	0.955	0.022	0.016	G2
229620	Iran	0.008	0.956	0.016	0.021	G2
229785	Iran	0.01	0.751	0.128	0.111	G2
232858	Costa Rica	0.313	0.547	0.07	0.07	G2
240310	Benin	0.045	0.392	0.221	0.342	G2
241367	Brazil	0.441	0.451	0.015	0.093	G2
243061	Brazil	0.117	0.622	0.142	0.12	G2
243081	Jordan	0.036	0.608	0.098	0.257	G2
244573	Brazil	0.021	0.658	0.023	0.297	G2
245105	Iran	0.011	0.962	0.019	0.008	G2
246996	Bahamas	0.024	0.902	0.026	0.048	G2
247094	Iran	0.016	0.889	0.041	0.054	G2
247095	Iran	0.367	0.566	0.038	0.029	G2
247096	Iran	0.012	0.958	0.017	0.013	G2
247097	Iran	0.016	0.959	0.016	0.009	G2
247098	Iran	0.016	0.941	0.023	0.019	G2
247099	Iran	0.011	0.966	0.012	0.011	G2
247100	Iran	0.027	0.615	0.308	0.05	G2
247101	Iran	0.194	0.781	0.01	0.014	G2
248392	India	0.366	0.498	0.023	0.113	G2

248393	India	0.136	0.7	0.135	0.029	G2
248933	India	0.067	0.675	0.03	0.228	G2
248947	India	0.017	0.573	0.013	0.397	G2
248954	India	0.096	0.461	0.015	0.428	G2
248955	India	0.202	0.616	0.058	0.124	G2
248959	India	0.174	0.746	0.013	0.067	G2
248963	India	0.363	0.536	0.016	0.084	G2
248967	India	0.167	0.589	0.028	0.216	G2
250026	Iran	0.009	0.965	0.012	0.013	G2
250027	Iran	0.013	0.945	0.016	0.026	G2
250028	Iran	0.036	0.927	0.013	0.024	G2
250743	Iran	0.037	0.774	0.173	0.017	G2
250874	Iran	0.014	0.944	0.026	0.017	G2
250876	Iran	0.012	0.772	0.198	0.018	G2
250877	Iran	0.011	0.949	0.016	0.025	G2
250878	Iran	0.023	0.803	0.164	0.01	G2
250879	Iran	0.207	0.667	0.072	0.053	G2
250880	Iran	0.016	0.664	0.307	0.013	G2
250881	Iran	0.017	0.664	0.304	0.015	G2
250883	Iran	0.074	0.521	0.391	0.014	G2
250938	Iran	0.174	0.74	0.071	0.015	G2
250939	Iran	0.109	0.799	0.066	0.025	G2
250942	Iran	0.107	0.868	0.018	0.007	G2
253621	Morocco	0.353	0.574	0.059	0.015	G2
257445	South Africa	0.388	0.552	0.026	0.035	G2
257450	South Africa	0.02	0.857	0.042	0.082	G2
257451	South Africa	0.161	0.744	0.043	0.052	G2
257452	South Africa	0.325	0.524	0.024	0.127	G2
257453	South Africa	0.254	0.451	0.022	0.273	G2
257454	South Africa	0.325	0.642	0.011	0.021	G2
257455	South Africa	0.216	0.696	0.013	0.076	G2
257462	South Africa	0.365	0.531	0.048	0.056	G2
257652	Former Soviet Union	0.287	0.497	0.016	0.2	G2
257654	Former Soviet Union	0.104	0.536	0.265	0.095	G2
267937	India	0.097	0.779	0.024	0.1	G2
269668	Pakistan	0.073	0.878	0.018	0.031	G2
277020	Argentina	0.309	0.643	0.026	0.021	G2
280219	Greece	0.042	0.754	0.095	0.11	G2
370042	India	0.181	0.776	0.015	0.027	G2
461404	Brazil	0.207	0.638	0.036	0.12	G2
508054	Hungary	0.21	0.524	0.011	0.255	G2
85850	United States	0.326	0.607	0.042	0.025	G2
85879	United States	0.098	0.374	0.286	0.241	G2
85922	United States	0.198	0.415	0.098	0.29	G2
85923	United States	0.125	0.444	0.044	0.387	G2

85937	United States	0.011	0.492	0.471	0.026	G2
86243	United States	0.063	0.478	0.034	0.425	G2
179731	India	0.277	0.404	0.065	0.254	G2
209622	Cuba	0.069	0.096	0.782	0.053	G3
219770	Argentina	0.198	0.361	0.381	0.06	G3
243211	Brazil	0.182	0.163	0.597	0.057	G3
248936	India	0.133	0.277	0.385	0.204	G3
248965	India	0.231	0.157	0.321	0.292	G3
248968	India	0.025	0.381	0.429	0.164	G3
257460	South Africa	0.324	0.027	0.602	0.046	G3
42844	United States	0.028	0.338	0.611	0.023	G3
85858	United States	0.054	0.031	0.527	0.389	G3
85859	United States	0.372	0.017	0.583	0.027	G3
85860	United States	0.025	0.029	0.916	0.029	G3
85861	United States	0.046	0.407	0.473	0.074	G3
85862	United States	0.016	0.022	0.946	0.016	G3
85863	United States	0.023	0.055	0.907	0.014	G3
85865	United States	0.198	0.065	0.708	0.03	G3
85866	United States	0.01	0.031	0.935	0.024	G3
85871	United States	0.033	0.283	0.578	0.106	G3
85872	United States	0.428	0.053	0.467	0.052	G3
86873	United States	0.43	0.038	0.504	0.028	G3
85885	United States	0.01	0.063	0.904	0.023	G3
85886	United States	0.092	0.045	0.826	0.037	G3
85888	United States	0.227	0.053	0.711	0.009	G3
85889	United States	0.174	0.093	0.709	0.024	G3
85892	United States	0.218	0.174	0.589	0.019	G3
85894	United States	0.014	0.014	0.942	0.03	G3
85895	United States	0.01	0.26	0.671	0.058	G3
85896	United States	0.019	0.269	0.518	0.194	G3
85897	United States	0.023	0.065	0.864	0.049	G3
85898	United States	0.016	0.241	0.716	0.027	G3
85899	United States	0.057	0.044	0.641	0.258	G3
95902	United States	0.019	0.009	0.544	0.429	G3
85906	United States	0.032	0.017	0.91	0.041	G3
85908	United States	0.027	0.011	0.791	0.171	G3
85909	United States	0.023	0.039	0.783	0.155	G3
85917	United States	0.024	0.077	0.815	0.085	G3
85927	United States	0.01	0.024	0.866	0.099	G3
85928	United States	0.054	0.061	0.748	0.136	G3
85929	United States	0.021	0.048	0.91	0.021	G3
85930	United States	0.023	0.008	0.593	0.375	G3
85934	United States	0.019	0.041	0.907	0.033	G3
85935	United States	0.009	0.011	0.949	0.031	G3
85936	United States	0.018	0.222	0.666	0.093	G3

85940	United States	0.065	0.135	0.79	0.01	G3
85941	United States	0.082	0.152	0.757	0.009	G3
85942	United States	0.073	0.011	0.856	0.06	G3
85943	United States	0.008	0.017	0.956	0.019	G3
85944	United States	0.022	0.12	0.565	0.293	G3
85945	United States	0.136	0.201	0.579	0.084	G3
85946	United States	0.02	0.051	0.881	0.047	G3
85947	United States	0.408	0.018	0.529	0.045	G3
85948	United States	0.009	0.017	0.953	0.021	G3
85949	United States	0.009	0.024	0.916	0.052	G3
85950	United States	0.058	0.105	0.796	0.042	G3
85951	United States	0.038	0.033	0.802	0.127	G3
85952	United States	0.021	0.313	0.64	0.026	G3
85953	United States	0.014	0.015	0.912	0.059	G3
85855	United States	0.018	0.032	0.733	0.217	G3
85958	United States	0.027	0.044	0.787	0.142	G3
85959	United States	0.014	0.013	0.919	0.054	G3
85961	United States	0.222	0.153	0.54	0.085	G3
85962	United States	0.037	0.268	0.668	0.027	G3
85963	United States	0.03	0.049	0.78	0.141	G3
85964	United States	0.031	0.023	0.783	0.164	G3
85965	United States	0.032	0.05	0.772	0.146	G3
85966	United States	0.087	0.208	0.461	0.244	G3
85967	United States	0.029	0.043	0.667	0.261	G3
85969	United States	0.059	0.009	0.876	0.056	G3
85970	United States	0.009	0.007	0.97	0.014	G3
85971	United States	0.009	0.016	0.962	0.013	G3
85972	United States	0.072	0.019	0.839	0.07	G3
85976	United States	0.012	0.01	0.943	0.036	G3
85978	United States	0.013	0.01	0.941	0.037	G3
85979	United States	0.02	0.01	0.901	0.069	G3
85980	United States	0.025	0.01	0.819	0.146	G3
86006	United States	0.241	0.015	0.697	0.047	G3
86008	United States	0.028	0.011	0.834	0.127	G3
86010	United States	0.013	0.01	0.963	0.013	G3
86011	United States	0.031	0.011	0.819	0.139	G3
86013	United States	0.014	0.009	0.955	0.021	G3
86020	United States	0.019	0.025	0.767	0.19	G3
86021	United States	0.011	0.136	0.824	0.029	G3
86022	United States	0.026	0.012	0.795	0.167	G3
86025	United States	0.014	0.014	0.714	0.258	G3
86030	United States	0.01	0.014	0.944	0.032	G3
86035	United States	0.052	0.007	0.878	0.063	G3
86042	United States	0.014	0.023	0.889	0.073	G3
86045	United States	0.093	0.151	0.639	0.117	G3

86051	United States	0.014	0.009	0.639	0.338	G3
86056	United States	0.013	0.01	0.916	0.06	G3
86058	United States	0.013	0.017	0.956	0.015	G3
86059	United States	0.008	0.008	0.965	0.018	G3
86061	United States	0.011	0.027	0.813	0.149	G3
86062	United States	0.011	0.024	0.938	0.027	G3
86064	United States	0.01	0.011	0.932	0.047	G3
86065	United States	0.018	0.016	0.915	0.051	G3
86067	United States	0.031	0.016	0.888	0.065	G3
86069	United States	0.02	0.014	0.941	0.024	G3
86072	United States	0.033	0.01	0.759	0.199	G3
86079	United States	0.043	0.014	0.686	0.256	G3
86080	United States	0.009	0.033	0.921	0.037	G3
86083	United States	0.01	0.014	0.858	0.119	G3
86087	United States	0.019	0.017	0.908	0.056	G3
86088	United States	0.014	0.013	0.952	0.022	G3
86100	United States	0.011	0.112	0.854	0.023	G3
86101	United States	0.013	0.021	0.838	0.127	G3
86102	United States	0.047	0.011	0.889	0.053	G3
86108	United States	0.018	0.103	0.847	0.032	G3
86110	United States	0.016	0.009	0.948	0.027	G3
86112	United States	0.072	0.027	0.565	0.337	G3
86113	United States	0.012	0.008	0.964	0.016	G3
86114	United States	0.016	0.025	0.935	0.023	G3
86115	United States	0.012	0.007	0.964	0.017	G3
86116	United States	0.014	0.006	0.965	0.015	G3
86118	United States	0.03	0.052	0.893	0.026	G3
86119	United States	0.009	0.007	0.964	0.02	G3
86120	United States	0.009	0.007	0.964	0.019	G3
86121	United States	0.007	0.008	0.976	0.009	G3
86122	United States	0.01	0.009	0.958	0.023	G3
86123	United States	0.013	0.013	0.96	0.014	G3
86125	United States	0.017	0.107	0.836	0.04	G3
86129	United States	0.013	0.006	0.96	0.022	G3
86130	United States	0.008	0.006	0.973	0.013	G3
86131	United States	0.007	0.01	0.972	0.011	G3
86132	United States	0.008	0.006	0.974	0.012	G3
86133	United States	0.017	0.027	0.946	0.011	G3
86134	United States	0.133	0.044	0.805	0.018	G3
86136	United States	0.012	0.141	0.822	0.026	G3
86137	United States	0.023	0.018	0.925	0.034	G3
86138	United States	0.011	0.008	0.955	0.025	G3
86140	United States	0.01	0.129	0.583	0.278	G3
86146	United States	0.011	0.009	0.961	0.019	G3
86148	United States	0.011	0.042	0.632	0.315	G3

86152	United States	0.01	0.011	0.959	0.02	G3
86153	United States	0.013	0.058	0.707	0.222	G3
86154	United States	0.011	0.007	0.969	0.013	G3
86157	United States	0.011	0.007	0.959	0.023	G3
86160	United States	0.016	0.011	0.949	0.023	G3
86164	United States	0.019	0.013	0.92	0.048	G3
86165	United States	0.018	0.013	0.904	0.065	G3
86166	United States	0.104	0.014	0.611	0.271	G3
86167	United States	0.058	0.008	0.723	0.211	G3
86168	United States	0.009	0.009	0.952	0.031	G3
86170	United States	0.01	0.009	0.8	0.181	G3
86172	United States	0.028	0.008	0.734	0.23	G3
86173	United States	0.012	0.009	0.957	0.022	G3
86176	United States	0.009	0.016	0.944	0.03	G3
86177	United States	0.059	0.011	0.869	0.062	G3
86181	United States	0.024	0.044	0.843	0.089	G3
86182	United States	0.054	0.011	0.878	0.058	G3
86183	United States	0.01	0.012	0.954	0.025	G3
86184	United States	0.009	0.009	0.966	0.016	G3
86185	United States	0.012	0.018	0.862	0.108	G3
86186	United States	0.012	0.012	0.948	0.027	G3
86187	United States	0.014	0.01	0.95	0.026	G3
86188	United States	0.01	0.035	0.94	0.015	G3
86189	United States	0.065	0.033	0.836	0.066	G3
86190	United States	0.01	0.008	0.954	0.027	G3
86191	United States	0.016	0.012	0.508	0.464	G3
86193	United States	0.014	0.012	0.488	0.486	G3
86196	United States	0.028	0.108	0.439	0.424	G3
86308	United States	0.068	0.045	0.579	0.309	G3
215720	Peru	0.18	0.152	0.193	0.474	G4
215721	Peru	0.307	0.171	0.127	0.395	G4
215722	Peru	0.118	0.415	0.02	0.447	G4
215770	Peru	0.029	0.436	0.029	0.505	G4
215775	Peru	0.206	0.182	0.098	0.514	G4
215776	Peru	0.157	0.304	0.052	0.487	G4
219776	Argentina	0.066	0.385	0.054	0.495	G4
221046	Argentina	0.057	0.295	0.075	0.573	G4
221047	Argentina	0.03	0.376	0.021	0.573	G4
221050	Argentina	0.056	0.291	0.063	0.59	G4
221052	Argentina	0.045	0.234	0.025	0.696	G4
221055	Argentina	0.024	0.308	0.019	0.649	G4
221471	Afghanistan	0.049	0.246	0.044	0.661	G4
221698	Indonesia	0.021	0.423	0.015	0.541	G4
229786	Iran	0.31	0.293	0.034	0.363	G4
238194	Brazil	0.037	0.437	0.033	0.492	G4

241370	Brazil	0.051	0.421	0.068	0.46	G4
248930	India	0.453	0.013	0.015	0.519	G4
248932	India	0.484	0.009	0.019	0.489	G4
248934	India	0.135	0.407	0.026	0.432	G4
248945	India	0.05	0.263	0.01	0.678	G4
248948	India	0.276	0.077	0.043	0.605	G4
248949	India	0.452	0.021	0.027	0.5	G4
248950	India	0.108	0.317	0.017	0.557	G4
248952	India	0.074	0.375	0.105	0.446	G4
248953	India	0.122	0.203	0.139	0.536	G4
248962	India	0.39	0.167	0.011	0.431	G4
257456	South Africa	0.373	0.038	0.087	0.502	G4
4692	United States	0.243	0.16	0.011	0.587	G4
85857	United States	0.085	0.19	0.253	0.471	G4
85877	United States	0.06	0.15	0.257	0.533	G4
85880	United States	0.015	0.008	0.451	0.526	G4
85884	United States	0.028	0.255	0.056	0.66	G4
85900	United States	0.028	0.101	0.156	0.715	G4
85903	United States	0.018	0.011	0.462	0.509	G4
85905	United States	0.082	0.011	0.099	0.808	G4
85907	United States	0.02	0.037	0.219	0.725	G4
85910	United States	0.012	0.01	0.384	0.594	G4
85911	United States	0.118	0.106	0.327	0.45	G4
85913	United States	0.017	0.007	0.353	0.623	G4
85914	United States	0.019	0.011	0.052	0.918	G4
85915	United States	0.014	0.008	0.021	0.957	G4
85916	United States	0.02	0.019	0.266	0.695	G4
85924	United States	0.014	0.008	0.326	0.652	G4
85925	United States	0.01	0.015	0.25	0.725	G4
85926	United States	0.279	0.028	0.299	0.394	G4
85931	United States	0.054	0.191	0.138	0.618	G4
85932	United States	0.036	0.014	0.255	0.695	G4
85933	United States	0.071	0.019	0.417	0.493	G4
86047	United States	0.019	0.007	0.416	0.558	G4
86075	United States	0.024	0.01	0.406	0.56	G4
86078	United States	0.048	0.014	0.4	0.537	G4
86095	United States	0.013	0.023	0.326	0.637	G4
86197	United States	0.025	0.011	0.024	0.939	G4
86199	United States	0.146	0.022	0.037	0.795	G4
86200	United States	0.16	0.024	0.041	0.776	G4
86201	United States	0.01	0.013	0.012	0.966	G4
86202	United States	0.031	0.191	0.284	0.494	G4
86203	United States	0.019	0.064	0.098	0.819	G4
86204	United States	0.071	0.011	0.067	0.851	G4
86205	United States	0.058	0.028	0.232	0.683	G4

86206	United States	0.03	0.009	0.069	0.892	G4
86207	United States	0.024	0.111	0.025	0.84	G4
86208	United States	0.015	0.025	0.031	0.929	G4
86209	United States	0.024	0.012	0.027	0.937	G4
86210	United States	0.071	0.007	0.171	0.751	G4
86212	United States	0.013	0.017	0.052	0.919	G4
86213	United States	0.046	0.016	0.027	0.911	G4
86214	United States	0.353	0.018	0.187	0.442	G4
86215	United States	0.012	0.012	0.021	0.955	G4
86217	United States	0.01	0.013	0.01	0.967	G4
86218	United States	0.011	0.016	0.021	0.952	G4
86219	United States	0.015	0.18	0.02	0.785	G4
86220	United States	0.01	0.014	0.018	0.958	G4
86221	United States	0.008	0.017	0.015	0.961	G4
86223	United States	0.012	0.025	0.022	0.941	G4
86224	United States	0.05	0.018	0.017	0.915	G4
86225	United States	0.047	0.034	0.013	0.906	G4
86228	United States	0.012	0.089	0.186	0.713	G4
86230	United States	0.113	0.295	0.034	0.558	G4
86233	United States	0.023	0.032	0.023	0.922	G4
86236	United States	0.012	0.015	0.029	0.943	G4
86238	United States	0.011	0.014	0.073	0.902	G4
86245	United States	0.031	0.011	0.045	0.913	G4
86246	United States	0.038	0.017	0.022	0.922	G4
86253	United States	0.085	0.056	0.04	0.818	G4
86261	United States	0.248	0.09	0.078	0.583	G4
86266	United States	0.024	0.019	0.012	0.945	G4
86268	United States	0.026	0.01	0.011	0.953	G4
86270	United States	0.011	0.187	0.011	0.791	G4
86280	United States	0.257	0.009	0.018	0.716	G4
86281	United States	0.015	0.2	0.016	0.769	G4
86282	United States	0.014	0.012	0.097	0.877	G4
86295	United States	0.016	0.012	0.061	0.91	G4
86313	United States	0.043	0.024	0.068	0.865	G4
86314	United States	0.012	0.01	0.035	0.943	G4
86315	United States	0.009	0.011	0.034	0.946	G4
86316	United States	0.03	0.06	0.016	0.895	G4
86319	United States	0.013	0.012	0.09	0.885	G4
86320	United States	0.009	0.009	0.01	0.972	G4
86322	United States	0.011	0.025	0.015	0.949	G4
86323	United States	0.01	0.009	0.033	0.948	G4
86325	United States	0.034	0.031	0.015	0.92	G4
86330	United States	0.082	0.101	0.011	0.805	G4
86331	United States	0.071	0.009	0.057	0.862	G4
86333	United States	0.022	0.016	0.157	0.805	G4

86337	United States	0.21	0.016	0.22	0.554	G4
86338	United States	0.051	0.014	0.027	0.907	G4
86340	United States	0.027	0.036	0.023	0.914	G4
86346	United States	0.08	0.014	0.04	0.865	G4
86351	United States	0.032	0.239	0.115	0.615	G4
86352	United States	0.186	0.012	0.024	0.778	G4
86354	United States	0.031	0.007	0.042	0.92	G4
86356	United States	0.027	0.01	0.018	0.946	G4
86359	United States	0.015	0.021	0.094	0.871	G4
86360	United States	0.014	0.019	0.091	0.876	G4
86361	United States	0.011	0.008	0.016	0.966	G4
86363	United States	0.02	0.065	0.065	0.85	G4
86363	United States	0.01	0.012	0.031	0.946	G4
86369	United States	0.03	0.007	0.025	0.938	G4
86370	United States	0.025	0.013	0.327	0.636	G4
86373	United States	0.012	0.017	0.054	0.917	G4
86374	United States	0.01	0.015	0.033	0.942	G4
86376	United States	0.037	0.054	0.169	0.74	G4
86386	United States	0.024	0.007	0.024	0.944	G4
86387	United States	0.012	0.015	0.189	0.783	G4
86389	United States	0.009	0.016	0.028	0.946	G4
86390	United States	0.029	0.008	0.036	0.927	G4
86391	United States	0.032	0.009	0.015	0.945	G4
86392	United States	0.009	0.02	0.038	0.933	G4
86394	United States	0.024	0.007	0.014	0.955	G4
86398	United States	0.219	0.109	0.048	0.624	G4
86403	United States	0.022	0.008	0.015	0.954	G4
86404	United States	0.04	0.01	0.028	0.922	G4
86405	United States	0.026	0.009	0.015	0.95	G4
86406	United States	0.033	0.013	0.01	0.943	G4
86407	United States	0.096	0.041	0.016	0.847	G4
86410	United States	0.114	0.035	0.017	0.834	G4
86414	United States	0.016	0.012	0.063	0.909	G4
86418	United States	0.016	0.323	0.023	0.638	G4
86419	United States	0.036	0.118	0.124	0.723	G4
86420	United States	0.024	0.09	0.024	0.862	G4
86421	United States	0.036	0.107	0.04	0.817	G4
183010	United States	0.016	0.018	0.06	0.906	G4
243214	Brazil	0.259	0.26	0.021	0.459	G4
248957	India	0.303	0.163	0.135	0.399	G4
248960	India	0.147	0.365	0.023	0.465	G4
298597	Nigeria	0.047	0.059	0.014	0.879	G4

Table S2. Selected 230 accessions for core collection establishment. Country origin and subpopulation membership coefficients are based on the analysis of the 574 accessions.

Accession number	Country	G1	G2	G3	G4	Gx
217558	India	0.495	0.086	0.27	0.149	G1
223409	Iran	0.391	0.249	0.131	0.23	G1
240673	Uruguay	0.563	0.33	0.043	0.064	G1
241362	Brazil	0.671	0.259	0.022	0.048	G1
241364	Brazil	0.679	0.262	0.038	0.021	G1
241366	Brazil	0.72	0.201	0.046	0.033	G1
241369	Brazil	0.547	0.326	0.035	0.093	G1
243062	Brazil	0.404	0.255	0.035	0.305	G1
248424	Zaire	0.528	0.12	0.029	0.324	G1
248500	Hungary	0.951	0.025	0.01	0.014	G1
248502	Hungary	0.635	0.208	0.07	0.087	G1
248926	India	0.95	0.015	0.011	0.024	G1
248935	India	0.779	0.016	0.021	0.185	G1
248938	India	0.491	0.083	0.054	0.372	G1
248940	India	0.754	0.006	0.021	0.218	G1
248942	India	0.896	0.006	0.017	0.081	G1
248944	India	0.747	0.016	0.102	0.135	G1
248958	India	0.455	0.043	0.304	0.199	G1
250226	Pakistan	0.884	0.029	0.037	0.05	G1
250397	Pakistan	0.566	0.124	0.177	0.133	G1
250574	Egypt	0.698	0.256	0.014	0.031	G1
250742	Iran	0.546	0.21	0.203	0.041	G1
252001	Turkey	0.956	0.018	0.008	0.017	G1
253422	Israel	0.95	0.026	0.011	0.013	G1
253620	Morocco	0.961	0.007	0.009	0.023	G1
255238	Mexico	0.929	0.043	0.01	0.017	G1
257444	South Africa	0.839	0.044	0.04	0.077	G1
257449	South Africa	0.537	0.026	0.018	0.419	G1
257458	South Africa	0.769	0.033	0.135	0.063	G1
257653	Former Soviet Union	0.532	0.434	0.012	0.022	G1
258388	Algeria	0.941	0.035	0.011	0.014	G1
262553	Bulgaria	0.901	0.009	0.049	0.041	G1
265504	United States	0.897	0.022	0.037	0.044	G1
265506	Colombia	0.821	0.017	0.026	0.136	G1
265508	United States	0.949	0.014	0.017	0.02	G1
267801	Brazil	0.922	0.042	0.013	0.023	G1
267806	Brazil	0.553	0.094	0.025	0.328	G1
274768	South Africa	0.624	0.357	0.011	0.008	G1
274773	South Africa	0.944	0.011	0.011	0.035	G1
277021	Argentina	0.538	0.052	0.017	0.393	G1
277024	Argentina	0.94	0.012	0.009	0.039	G1

289220	Greece	0.597	0.012	0.02	0.371	G1
298599	Brazil	0.536	0.356	0.024	0.084	G1
306724	Taiwan	0.696	0.283	0.009	0.013	G1
323589	Portugal	0.818	0.103	0.024	0.055	G1
326198	Mexico	0.461	0.434	0.046	0.058	G1
326343	Paraguay	0.951	0.008	0.029	0.012	G1
384014	Brazil	0.617	0.295	0.021	0.067	G1
436592	China	0.866	0.012	0.059	0.064	G1
486318	Peru	0.563	0.306	0.028	0.103	G1
631156	United States	0.784	0.055	0.092	0.069	G1
4686	United States	0.936	0.017	0.013	0.033	G1
7353	United States	0.611	0.011	0.35	0.028	G1
15620	United States	0.799	0.022	0.166	0.013	G1
15624	United States	0.939	0.027	0.012	0.022	G1
85785	United States	0.766	0.014	0.196	0.024	G1
85806	United States	0.934	0.009	0.039	0.018	G1
85817	United States	0.899	0.009	0.08	0.012	G1
85918	United States	0.511	0.212	0.049	0.229	G1
85920	United States	0.896	0.035	0.019	0.049	G1
85988	United States	0.517	0.018	0.442	0.023	G1
86307	United States	0.402	0.025	0.245	0.327	G1
243200	Brazil	0.513	0.295	0.035	0.157	G1
208843	Cuba	0.017	0.935	0.028	0.02	G2
209132	Puerto Rico	0.26	0.58	0.109	0.051	G2
209326	Virgin Islands (U.S.)	0.231	0.52	0.054	0.195	G2
209436	Panama	0.032	0.738	0.033	0.197	G2
212115	Afghanistan	0.01	0.95	0.02	0.02	G2
213021	India	0.223	0.427	0.208	0.143	G2
215769	Peru	0.037	0.656	0.047	0.26	G2
215773	Peru	0.069	0.757	0.098	0.076	G2
217539	Pakistan	0.072	0.471	0.192	0.265	G2
219766	Botswana	0.237	0.727	0.015	0.021	G2
219767	Argentina	0.036	0.935	0.01	0.018	G2
219771	Argentina	0.319	0.353	0.029	0.299	G2
219775	Argentina	0.036	0.876	0.026	0.062	G2
221053	Argentina	0.02	0.494	0.249	0.238	G2
222745	Iran	0.008	0.979	0.008	0.006	G2
223013	Iran	0.287	0.676	0.013	0.023	G2
227869	Iran	0.4	0.501	0.032	0.067	G2
229785	Iran	0.01	0.751	0.128	0.111	G2
232858	Costa Rica	0.313	0.547	0.07	0.07	G2
240310	Benin	0.045	0.392	0.221	0.342	G2
241367	Brazil	0.441	0.451	0.015	0.093	G2
243081	Jordan	0.036	0.608	0.098	0.257	G2

244573	Brazil	0.021	0.658	0.023	0.297	G2
246996	Bahamas	0.024	0.902	0.026	0.048	G2
247095	Iran	0.367	0.566	0.038	0.029	G2
247099	Iran	0.011	0.966	0.012	0.011	G2
247100	Iran	0.027	0.615	0.308	0.05	G2
248392	India	0.366	0.498	0.023	0.113	G2
248947	India	0.017	0.573	0.013	0.397	G2
250027	Iran	0.013	0.945	0.016	0.026	G2
250743	Iran	0.037	0.774	0.173	0.017	G2
250878	Iran	0.023	0.803	0.164	0.01	G2
250879	Iran	0.207	0.667	0.072	0.053	G2
250938	Iran	0.174	0.74	0.071	0.015	G2
253621	Morocco	0.353	0.574	0.059	0.015	G2
257445	South Africa	0.388	0.552	0.026	0.035	G2
257451	South Africa	0.161	0.744	0.043	0.052	G2
257453	South Africa	0.254	0.451	0.022	0.273	G2
257652	Former Soviet Union	0.287	0.497	0.016	0.2	G2
280219	Greece	0.042	0.754	0.095	0.11	G2
508054	Hungary	0.21	0.524	0.011	0.255	G2
85850	United States	0.326	0.607	0.042	0.025	G2
85937	United States	0.011	0.492	0.471	0.026	G2
209622	Cuba	0.069	0.096	0.782	0.053	G3
219770	Argentina	0.198	0.361	0.381	0.06	G3
243211	Brazil	0.182	0.163	0.597	0.057	G3
248936	India	0.133	0.277	0.385	0.204	G3
248965	India	0.231	0.157	0.321	0.292	G3
248968	India	0.025	0.381	0.429	0.164	G3
257460	South Africa	0.324	0.027	0.602	0.046	G3
42844	United States	0.028	0.338	0.611	0.023	G3
85859	United States	0.372	0.017	0.583	0.027	G3
85862	United States	0.016	0.022	0.946	0.016	G3
85865	United States	0.198	0.065	0.708	0.03	G3
85866	United States	0.01	0.031	0.935	0.024	G3
85871	United States	0.033	0.283	0.578	0.106	G3
86873	United States	0.43	0.038	0.504	0.028	G3
85888	United States	0.227	0.053	0.711	0.009	G3
85894	United States	0.014	0.014	0.942	0.03	G3
85897	United States	0.023	0.065	0.864	0.049	G3
95902	United States	0.019	0.009	0.544	0.429	G3
85909	United States	0.023	0.039	0.783	0.155	G3
85917	United States	0.024	0.077	0.815	0.085	G3
85928	United States	0.054	0.061	0.748	0.136	G3
85934	United States	0.019	0.041	0.907	0.033	G3
85940	United States	0.065	0.135	0.79	0.01	G3

85943	United States	0.008	0.017	0.956	0.019	G3
85946	United States	0.02	0.051	0.881	0.047	G3
85949	United States	0.009	0.024	0.916	0.052	G3
85950	United States	0.058	0.105	0.796	0.042	G3
85952	United States	0.021	0.313	0.64	0.026	G3
85958	United States	0.027	0.044	0.787	0.142	G3
85962	United States	0.037	0.268	0.668	0.027	G3
85965	United States	0.032	0.05	0.772	0.146	G3
85969	United States	0.059	0.009	0.876	0.056	G3
85972	United States	0.072	0.019	0.839	0.07	G3
85979	United States	0.02	0.01	0.901	0.069	G3
86008	United States	0.028	0.011	0.834	0.127	G3
86013	United States	0.014	0.009	0.955	0.021	G3
86022	United States	0.026	0.012	0.795	0.167	G3
86025	United States	0.014	0.014	0.714	0.258	G3
86035	United States	0.052	0.007	0.878	0.063	G3
86051	United States	0.014	0.009	0.639	0.338	G3
86059	United States	0.008	0.008	0.965	0.018	G3
86064	United States	0.01	0.011	0.932	0.047	G3
86069	United States	0.02	0.014	0.941	0.024	G3
86080	United States	0.009	0.033	0.921	0.037	G3
86088	United States	0.014	0.013	0.952	0.022	G3
86102	United States	0.047	0.011	0.889	0.053	G3
86112	United States	0.072	0.027	0.565	0.337	G3
86115	United States	0.012	0.007	0.964	0.017	G3
86119	United States	0.009	0.007	0.964	0.02	G3
86122	United States	0.01	0.009	0.958	0.023	G3
86129	United States	0.013	0.006	0.96	0.022	G3
86132	United States	0.008	0.006	0.974	0.012	G3
86136	United States	0.012	0.141	0.822	0.026	G3
86140	United States	0.01	0.129	0.583	0.278	G3
86152	United States	0.01	0.011	0.959	0.02	G3
86157	United States	0.011	0.007	0.959	0.023	G3
86165	United States	0.018	0.013	0.904	0.065	G3
86168	United States	0.009	0.009	0.952	0.031	G3
86173	United States	0.012	0.009	0.957	0.022	G3
86181	United States	0.024	0.044	0.843	0.089	G3
86184	United States	0.009	0.009	0.966	0.016	G3
86187	United States	0.014	0.01	0.95	0.026	G3
86190	United States	0.01	0.008	0.954	0.027	G3
86196	United States	0.028	0.108	0.439	0.424	G3
86308	United States	0.068	0.045	0.579	0.309	G3
215720	Peru	0.18	0.152	0.193	0.474	G4
215721	Peru	0.307	0.171	0.127	0.395	G4

215770	Peru	0.029	0.436	0.029	0.505	G4
215775	Peru	0.206	0.182	0.098	0.514	G4
219776	Argentina	0.066	0.385	0.054	0.495	G4
221047	Argentina	0.03	0.376	0.021	0.573	G4
221052	Argentina	0.045	0.234	0.025	0.696	G4
221055	Argentina	0.024	0.308	0.019	0.649	G4
221471	Afghanistan	0.049	0.246	0.044	0.661	G4
221698	Indonesia	0.021	0.423	0.015	0.541	G4
229786	Iran	0.31	0.293	0.034	0.363	G4
238194	Brazil	0.037	0.437	0.033	0.492	G4
241370	Brazil	0.051	0.421	0.068	0.46	G4
248930	India	0.453	0.013	0.015	0.519	G4
248934	India	0.135	0.407	0.026	0.432	G4
248945	India	0.05	0.263	0.01	0.678	G4
248948	India	0.276	0.077	0.043	0.605	G4
248950	India	0.108	0.317	0.017	0.557	G4
248953	India	0.122	0.203	0.139	0.536	G4
248962	India	0.39	0.167	0.011	0.431	G4
257456	South Africa	0.373	0.038	0.087	0.502	G4
4692	United States	0.243	0.16	0.011	0.587	G4
85880	United States	0.015	0.008	0.451	0.526	G4
85905	United States	0.082	0.011	0.099	0.808	G4
85911	United States	0.118	0.106	0.327	0.45	G4
85916	United States	0.02	0.019	0.266	0.695	G4
85926	United States	0.279	0.028	0.299	0.394	G4
86047	United States	0.019	0.007	0.416	0.558	G4
86095	United States	0.013	0.023	0.326	0.637	G4
86201	United States	0.01	0.013	0.012	0.966	G4
86204	United States	0.071	0.011	0.067	0.851	G4
86208	United States	0.015	0.025	0.031	0.929	G4
86210	United States	0.071	0.007	0.171	0.751	G4
86215	United States	0.012	0.012	0.021	0.955	G4
86220	United States	0.01	0.014	0.018	0.958	G4
86225	United States	0.047	0.034	0.013	0.906	G4
86233	United States	0.023	0.032	0.023	0.922	G4
86246	United States	0.038	0.017	0.022	0.922	G4
86266	United States	0.024	0.019	0.012	0.945	G4
86280	United States	0.257	0.009	0.018	0.716	G4
86313	United States	0.043	0.024	0.068	0.865	G4
86316	United States	0.03	0.06	0.016	0.895	G4
86323	United States	0.01	0.009	0.033	0.948	G4
86331	United States	0.071	0.009	0.057	0.862	G4
86340	United States	0.027	0.036	0.023	0.914	G4
86352	United States	0.186	0.012	0.024	0.778	G4

86360	United States	0.014	0.019	0.091	0.876	G4
86363	United States	0.01	0.012	0.031	0.946	G4
86374	United States	0.01	0.015	0.033	0.942	G4
86387	United States	0.012	0.015	0.189	0.783	G4
86392	United States	0.009	0.02	0.038	0.933	G4
86403	United States	0.022	0.008	0.015	0.954	G4
86407	United States	0.096	0.041	0.016	0.847	G4
86418	United States	0.016	0.323	0.023	0.638	G4
183010	United States	0.016	0.018	0.06	0.906	G4
243214	Brazil	0.259	0.26	0.021	0.459	G4
248957	India	0.303	0.163	0.135	0.399	G4
248960	India	0.147	0.365	0.023	0.465	G4
298597	Nigeria	0.047	0.059	0.014	0.879	G4

Draft

Supplementary Figures

a

b

b. Number of accessions in the subset from main countries

c

Figure S1. Oil content and country origin. **a)** Sample distribution of oil content for the entire collection and the subset collection. **b)** Numbers of samples from main countries in the 574-accession set. **c)** Numbers of samples from main countries in the entire collection.

Figure S2

Figure S2. Population structure analysis. **a)** Determination of appropriate subpopulation numbers by Structure Harvester analysis. The tested numbers of subpopulations were shown on *x*-axis. The Delta *k* output from Structure Harvester (measures rate of changes of probability of data given that *k*-value) was shown on *y*-axis. The highest probability is at *k* = 4. **b)** The population structure pattern with different *k* values. Numbers on the *y*-axis show the subpopulation membership percentage, and each column represents one of the 574 accessions.

Figure S3. Neighbor-joining tree analysis for the 230 selected accessions. Branch color indicates the subpopulation membership identified for each accession by population structure analysis.

Draft