Supporting Information

Muchero et al. 10.1073/pnas.0905886106

SI Text

Growth Conditions for RNA Extraction for Libraries in Table S1. Seeds were surface-sterilized for 5 min in 1% sodium hypochlorite with 0.01% Tween 20 and rinsed thoroughly with deionized water before planting.

Developing axillary buds from cultivar California Blackeye no. 27′ (CB27) were collected from well-watered and drought-stressed greenhouse-grown seedlings. Seeds were germinated in pots filled with steam-sterilized University of California Riverside soil mix UCMIX-3 (http://agops.ucr.edu/pdfs/soil_mix_recipes.pdf). The temperature was set at 35 °C day and 18 °C night with natural light. Buds from 10-day-old well-watered seedlings were harvested from one set of pots. Another set of plants was grown under drought stress induced by withholding water from 6-day-old seedlings for 4 days before sampling.

For the remaining nine genotypes (Table S1), a more complex range of tissues was used to increase RNA complexity. Hypocotyls, primary roots, epicotyls, and foliage tissues were collected from seeds germinated on sterile filter paper with water containing 100 units/mL nystatin and 50 μg/mL cefotaxime and kept for 4–5 days in the dark. In addition, young trifoliate and apical meristem tissues were collected from greenhouse-grown well-watered and drought-stressed 6-week-old seedlings, essentially as for CB27. Seeds were germinated in pots and thinned to four plants per pot. For the stress treatment, plants were allowed to grow for 3 weeks after which irrigation was withheld for an additional 3 weeks. The nonstressed plants were watered daily for 6 weeks using an automated drip irrigation system. All samples were frozen on dry ice and kept at −80 °C before RNA isolation. The RNA isolated from genotype IT84S-2049 was used to construct two cDNA libraries by using different cloning methods.

RNA Extraction and mRNA Purification for Libraries in Table S1. The TRIzol method (Invitrogen) was used to extract RNA from all tissues except for hypocotyl and primary root samples for which the Qiagen RNeasy kit was used. The following modifications were made to the TRIzol protocol: (*i*) 10-min incubation was done in extraction buffer after homogenization, (*ii*) centrifugation was carried out with a Beckman GPKR/GH 3.7 rotor at 1,430 × *g* for 20 min, and (*iii*) extractions were performed twice with 10 mL of chloroform. RNA was quantified at 260 nm by using a spectrophotometer (Beckman DU650) and the quality was assessed by RNA denaturing gel electrophoresis.

For each genotype except CB27, RNA was pooled as follows; 1/3 from hypocotyls and primary root, 1/3 from epicotyl and foliage leaves, and 1/6 from each combination of young trifoliate leaves and apical meristem from unstressed and stressed plants, for a total of 400–500 μg of total RNA. The mRNAs were purified by using Poly(A) Track mRNA Isolation System IV (Promega) or an Absolutely mRNA purification kit (Stratagene) and quantified with a NanoDrop ND-1000 spectrophotometer. Poly(A) RNA quality was assessed by running 25–250 ng on a RNA Lab-On-A-Chip (Caliper Technologies) using the Bioanalyzer 2100 (Agilent Technologies).

cDNA Libraries and Sequencing. The eleven libraries in Table S1 were constructed by using between 1.7 and 5.0 μg of mRNA. Nine of these libraries, UCRVU04 to UCRVU12, were constructed by using the SuperScript Plasmid System with Gateway Technology for cDNA Synthesis and Cloning (Invitrogen) and the pCMV·SPORT6 vector. These cDNAs were directionally cloned with SalI on the 5′ end and NotI on 3′ end. Fragments >0.5 kb were selected after size fractionation via gel filtration and cDNA concentration was determined by using a NanoDrop ND-1000 spectrophotometer. Vector ligation was performed by using 100 ng of cDNA and 1 μg of vector and incubated at room temperature for 3 h. Electrocompetent *E. coli* strain DH10B cells were transformed by adding 4 μL of ligated DNA to 80 μL of DH10B cells and delivering a brief 1,850-V shock using an Electroporator 2510 (Eppendorf). A volume of 1.5–2.0 mL of SOC medium was added to the electroporated cells, which were then incubated 1 h for recovery. Yields in these libraries were between 1.1 and 3×10^7 cfu.

Two libraries, UCRVU02 (by Chinta Sudhakar) and UCRVU03, were constructed from CB27 and IT84S-2049, respectively (Table S1), using a λ ZAP cDNA Synthesis Kit (Stratagene) and a Uni-ZAP XR vector. For these, cDNAs >0.5 kb were selected as described above and directionally cloned in pBluescript SK (phagemid) with EcoRI on the 5′ end and XhoI on the 3′ end. One hundred nanograms of cDNA and 1 μg of vector were used in the ligation for UCRVU02, whereas 25 ng of cDNA and 1 μg of vector were used for the UCRVU03 library. Ligated DNA was packaged into phage lambda particles and used to transduce *E. coli* strain XL 1-Blue-MRF cells. Phagemids were then mass excised for 3 h from 1×10^6 primary recombinants for library UCRVU02 or 2.5×10^5 primary recombinants for UCRVU03. Total phagemid yields in these libraries were 2.9×10^8 and 2.72×10^6 cfu, respectively.

In total 77,601 clones from the 11 cDNA libraries were sequenced by using the Sanger dideoxy chain termination method. Clones from cDNA libraries UCRVU02 and UCRVU03 were sequenced at the University of California Riverside Institute for Integrative Genome Biology. For libraries UCRVU04 to UCRVU12, plating and sequencing was performed at the U.S. Department of Energy Joint Genome Institute (JGI) Sequence Production Facility, Walnut Creek, CA. At JGI, each library was plated onto agarose plates containing ampicillin (254-mm plates from Teknova) at a density of ≈1,000 colonies per plate. Plates were grown at 37 °C for 18 h, then individual colonies were picked and used to inoculate a well containing LB media with ampicillin in a 384-well plate (Nunc). The 384-well plates were grown at 37 °C for 18 h. Plasmid DNA for sequencing was produced by rolling circle amplification (Templiphi; GE Healthcare). Subclone inserts were sequenced from both ends using primers complementary to the flanking pCMV-SPORT6 vector sequence and Big Dye terminator chemistry and then run on ABI 3730 instruments. Trace files were downloaded from JGI and then processed to remove vector and cloning oligo sequences and various contaminants and to trim to a high-quality region.

Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank and used further in this work. Accession numbers are as follows: UCRVU02, DR068280-DR068450 and EG594208-EG594295; UCRVU03, ES884082-884239; UCRVU04 through UcRVU12, FC456601-FC462285, FG807582-FG942979 and GE649954-650046.

SNP Selection for the Illumina GoldenGate Assay. Prioritization of SNPs for inclusion in the 1,536-SNP cowpea GoldenGate assay was in the following order: 1,248 SNPs identified in more than one pairwise cowpea genotype comparison involving only African breeding germplasm, or from the 3 × 3 SNP list; 19 SNPs found only in a comparison between RIL parent lines IT84S-

2049 and IT97K-461-4; 71 SNPs found only once in the African germplasm SNP lists and matching a soybean gene model targeted in a project led by D. Cook; 19 SNPs found only once in the African germplasm SNP lists and at least once in any other pairwise genotype comparison; 23 SNPs found only in a comparison between the parents of the standard mapping population IT84S-2246 × IT97K-461-4; 22 SNPs found only once in the African germplasm SNP lists and a best BLASTX located within a low marker-density region on *M. truncatula* chromosomes 3S, 6L, or 7S; and 134 SNPs found only once in the African germplasm SNP lists and with the best Illumina SNP scores among the remaining SNPs.

DNA Isolation and GoldenGate Genotyping Assay. Briefly, genomic DNA was isolated from 759 parental genotypes and RILs by using Plant DNeasy (Qiagen) starting with 100 mg of young trifoliate leaves. DNA concentration was determined by using a Quant-iT dsDNA Assay Kit Broad Range (Q33130) (Invitrogen) and fluorescence (485 nm/535 nm, 1.0 s) measured with a microtiter plate reader (PerkinElmer/Life Sciences; Wallac Victor^2, 1420 Multilabel counter). Samples were adjusted to 80 ng/μL in Tris-EDTA buffer. The GoldenGate genotyping assay was performed at the University of California Los Angeles genotyping facility by Joe DeYoung, Maricel Almonte, and Oi-Wa Choi using 250 ng of DNA for each assay.

Data Production. A Visual FoxPro script was developed to automate phase assignment of RIL genotype calls based on parental genotypes, tally no-calls, heterozygotes, and nonparental alleles for each RIL and calculate allele frequencies.

Data Processing. Data processing before mapping included: exclusion of SNPs having poor technical performance in the Golden Gate assay; exclusion of SNPs exhibiting segregation distortion, defined as having a MAF <0.30; exclusion of RILs with excessive heterozygosity, which suggested cross-contaminated DNA samples or recent intercrossing or outcrossing; exclusion of RILs with excessive nonparental alleles, which suggested outcrossing or errors during seed handling causing stray genotypes to be propagated as RILs; and exclusion of RILs with excessive no-calls, indicating poor-quality DNA samples.

An acceptable standard for each of these parameters was determined for each mapping population empirically by visual inspection of the distribution of the above quality metrics and looking for obvious break points. In addition, individual SNP markers with no-calls >5% were discarded from individual mapping datasets, even if they passed the initial SNP technical quality assessment. More specifically for MAF, we attempted to avoid bias in the consensus map from regions showing segregation distortion. This was based on the observation that markers with lower MAF tended to map in the same regions within individual RIL populations. Exclusion of such markers based on χ^2 analysis would result in significant gaps in these maps. However, use of MAFs <0.30 caused some conflicts in consensus map building; therefore, we adopted the 0.30 threshold as an optimum cut-off point.

Individual Maps. LG assignments were determined with LOD grouping thresholds between 4 and 8, depending on the population. Within-LG marker order was calculated by using a mapping LOD threshold of 7 or higher. Markers determined by the JoinMap software to have the wrong phase assignment were inverted to the opposite phase. Markers with the wrong phase assignment were encountered mostly in the Dan Ila × TVu-7778 and Yacine 819 × 58–77 mapping populations where the genotype of one or both of the parental genotypes was not determined directly but only inferred. Maximum recombination was set at 45%.

Consensus Map. A consensus map based on the six mapping populations described above was first produced by using JoinMap 3.0 (11). Consensus LGs were generated one at a time based on integration of manually inspected individual-map LGs for anchor markers. Raw segregation data for LGs from each mapping population were combined by using the "join" command. At this point the same mapping LOD of 7, the Kosambi mapping function and 45% maximum recombination were used to generate consensus LGs. The use of the Haldane function compared with the Kosambi function was explored. In general the Haldane function resulted in the same marker order but slightly longer LGs. The Haldane function effect on cM distances was not consistent over all consensus LGs. This function was more sensitive to changes in mapping LODs from 1 to 7, resulting in shorter VuLG8, VuLG9, and VuLG11 compared with the Kosambi function. The Haldane function resulted in a net increase of 12 cM in overall map length with no major changes in the marker order.

Draft consensus LGs were compared with each other and individual-population maps to identify marker duplications and LG conflicts, respectively. Marker order was not fixed at this stage because JoinMap is known to introduce significant errors of marker order in dense consensus maps (13). The software uses raw segregation data to calculate pairwise recombination frequencies and LOD scores. This consensus building process assumes homogeneous recombination frequencies across populations, but does not account for presence of recombinational "hot and cold spots" and is hindered by the presence of missing data points within the pooled data. This results in significant local marker rearrangement (13).

Therefore, MergeMap, which is principally a marker ordering program, was used preferentially to assign final marker order in the consensus map. Briefly, MergeMap takes into account marker order from individual populations and does not use raw segregation data to build a consensus map. Rather, consensus marker order is determined by converting individual maps into directed acyclic graphs (DAGs) and based on shared markers (anchor markers), individual DAGs are merged into a consensus map using the shared markers to infer marker positions. Conflicts in marker assignment between different maps are presented as circles within these DAGs that highlight ambiguities in marker placement. This approach allows for the identification of conflicts between individual maps that may require additional markers to resolve or may represent bona fide biological processes such as translocations or duplication (13).

Because JoinMap calculates pairwise linkages from raw data, the software is able to resolve multiple LGs from a common pool of data generated from combining individual maps. This was useful when insufficient recombination resulted in markers belonging to two distinct LGs being mapped onto the same LG. Such an event could lead to spurious linkage of distinct LGs if marker order is exclusively considered in consensus map construction. After iteratively merging LGs, the resulting 11 showing no further LG conflicts or marker duplication were compared with the individual maps and inspected for marker placement conflicts. LG conflicts resulting from marker duplications in the consensus map were resolved by merging corresponding LGs and recalculating the map until no further duplications were observed. Finally, within-LG marker order was determined using the MergeMap software (12).

Consensus LGs were numbered VuLG1 to VuLG11 according to decreasing cM distances. These were used as reference LGs for comparison with the previously published cowpea genetic maps (6, 8). For comparison purposes, individual maps were calculated in JoinMap as described above using combined segregation data from previous and current studies for the 524B × IT84S-2049 and CB46 × IT93K-503–1 RIL populations. SNP markers were used as reference points for comparison between previous and new consensus maps.

PNAS PNAS PNAS

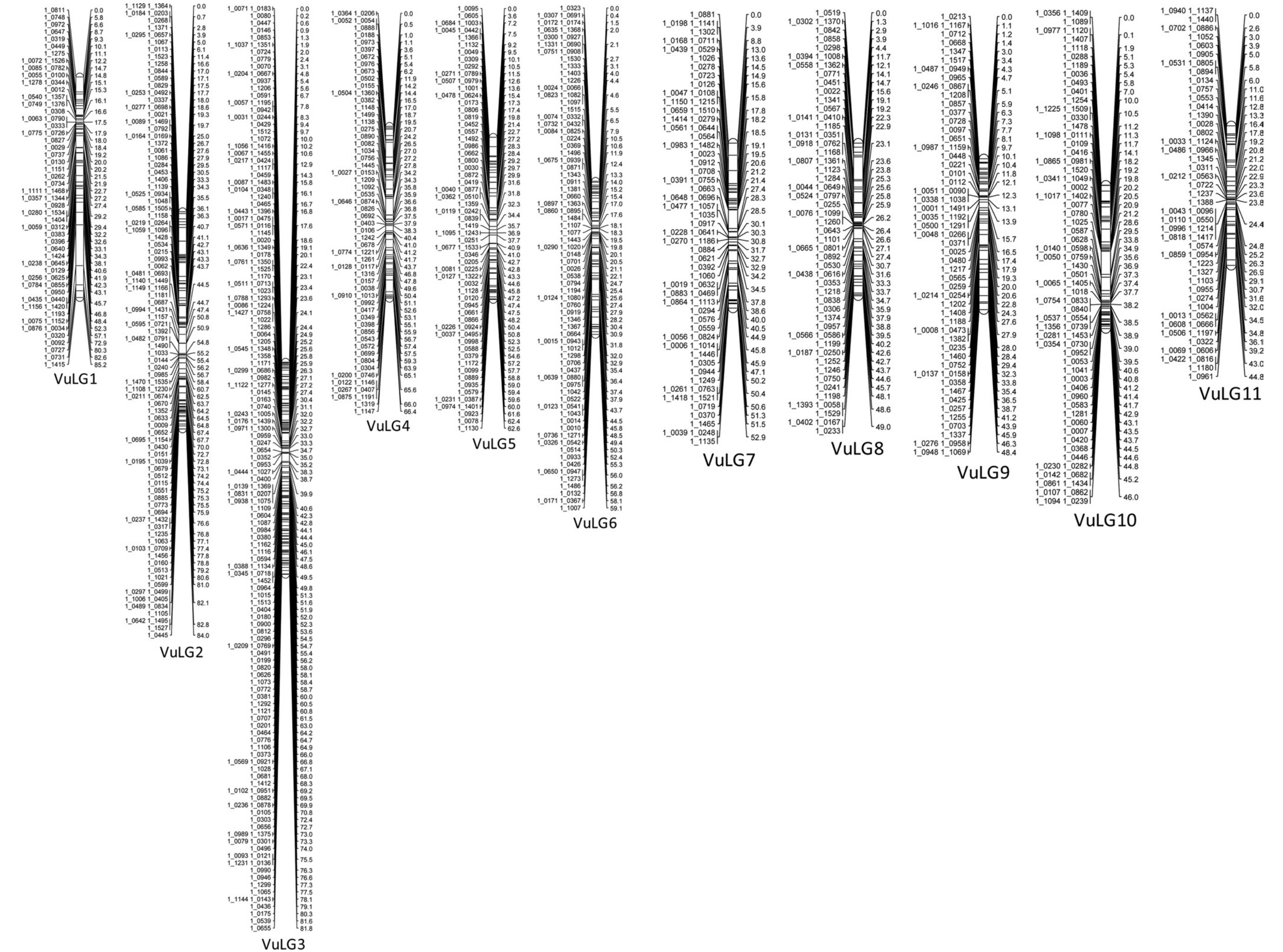

Fig. S1. Graphical representation of the consensus cowpea genetic linkage map constructed by using 928 EST-derived SNP markers segregating in six recombinant inbred populations.

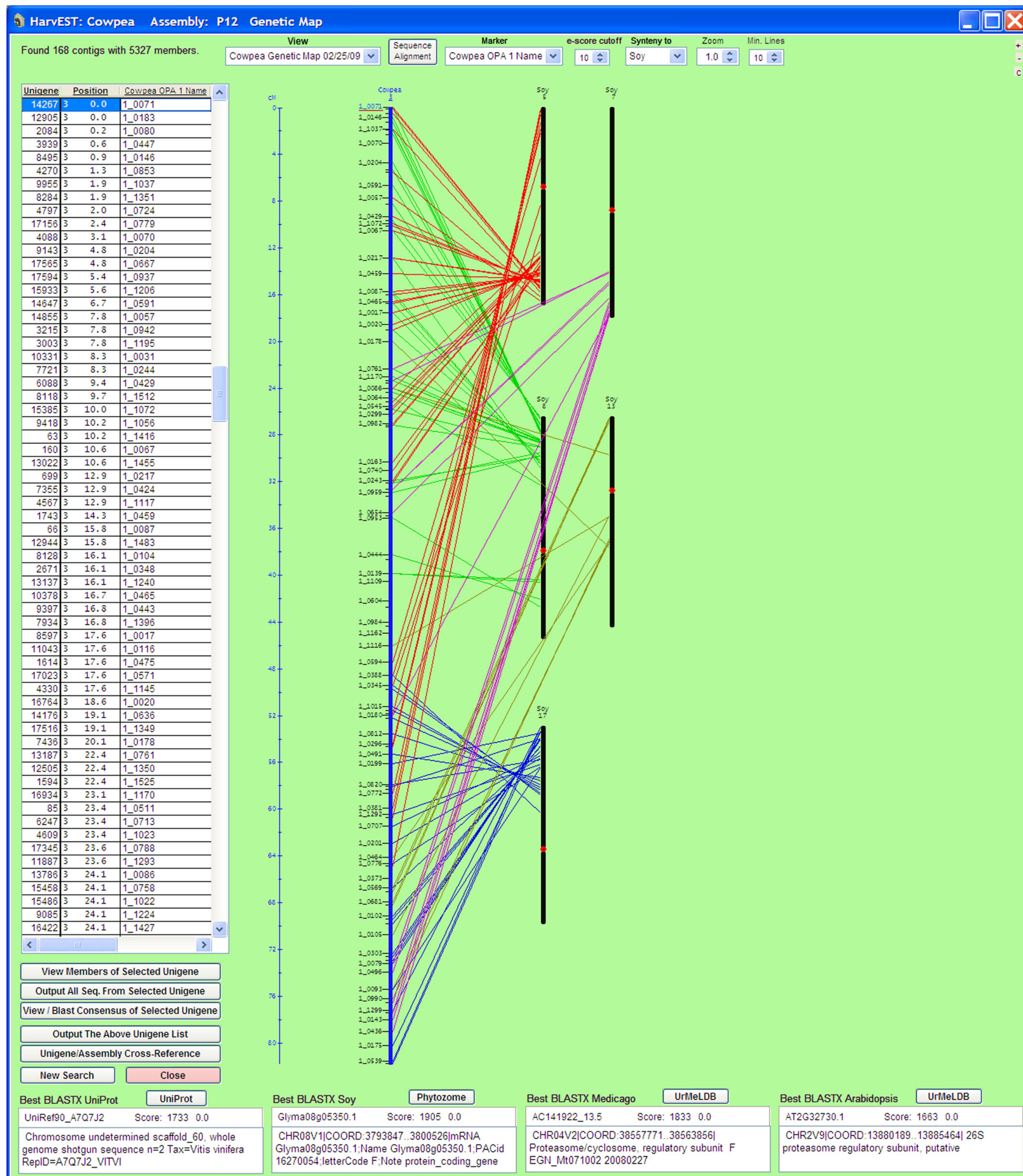

Fig. S2. Synteny between cowpea LG VuLG3 and soybean chromosomes. Screen shot from genetic map display in HarvEST:Cowpea (http://harvest.ucr.edu).

PNAS PNAS PNAS

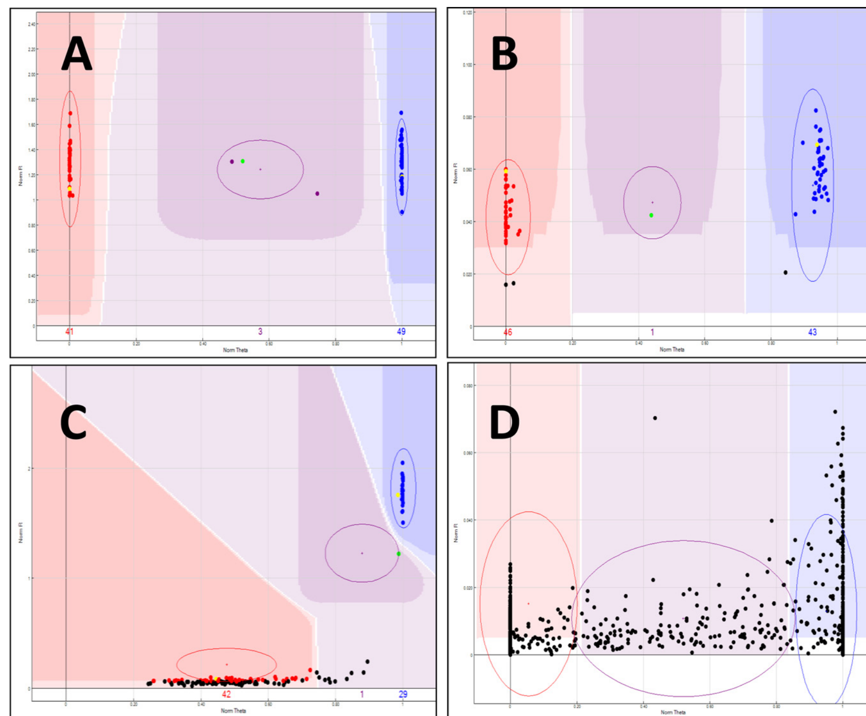

Fig. S3. BeadStudio representation of GoldenGate Assay genotyping data. (*A*) High-quality SNP showing clearly separate clusters of homozygous alleles (blue and red), synthetic heterozygote (green), and biological heterozygotes (purple). (*B*) Satisfactory quality SNP requiring minor manual adjustments of the default no calls (black). (*C*) A SNP with just one minor allele in a homozygous state and one biological heterozygote. (*D*) A technically unsatisfactory SNP.

Table S1. Sequencing results for cDNA libraries

Library	Germplasm ID	Clones	ESTs
UCRVU02	CB27	129	245
UCRVU03	IT84S-2049	83	158
UCRVU04	524B	15,282	28,404
UCRVU05	IT84S-2049	14,689	27,244
UCRVU06	UCR 5301	327	572
UCRVU07	UCR 779	15,304	26,746
UCRVU08	IT97K-461–4	15,403	28,257
UCRVU09	UCR 707	15,303	27,903
UCRVU10	PI 418979	361	637
UCRVU11	UCR 41	351	633
UCRVU12	UCR 2563	369	654
Totals		77,601	141,453

Other Supporting Information Files

Dataset S1 (XLS)