**Supplementary Figures**

**Supplementary Figure S1.** The stepwise filtering of *G. max* and *G. soja* accessions held in the USDA Germplasm Collection for analysis of population structure and genome-wide association mapping.

**Supplementary Figure S2.** Percentage distribution of the 14,430 soybean accessions used in the population structure analysis according to world region (panel a) and maturity group class (panel b).

**Supplementary Figure S3.** Exploration of the optimal number of genetic subpopulations (K) using Δ cross-validation error values in the soybean germplasm collection. A solid line denotes the choice of K=5 which represents the most likely number of subpopulations within the soybean germplasm collection.

**Supplementary Figure S4.** Genome-wide association study for protein (panel a) and oil (panel b) within each world region class. Manhattan plot for association within each subpopulation (left) and quantile-quantile plots of -log10 (P value) (right) are vertically arranged in each panel. Markers are plotted on the x-axis according to their physical position on each chromosome. The solid horizontal line indicates the calculated threshold value for declaring a significant association. The dashed vertical lines indicate the same significant associations detected in two or more MGs that co-localized with the significant SNP detected for either protein or oil using 12, 116 *G. max* accessions.

**Supplementary Figure S5.** Genome-wide association study for protein (panel a) and oil (panel b) within each MG class. Manhattan plot for association within each subpopulation (left) and quantile-quantile plots of -log10 (P value) (right). Markers are plotted on the x-axis according to their physical position on each chromosome. The solid horizontal line indicates the calculated threshold value for declaring significant association. The dashed vertical lines indicate the same significant associations detected in two or more MGs that co-localized with the significant SNP detected for either protein or oil using 12, 116 *G. max* accessions.

**Supplementary Figure S6.** Genome-wide association scans for *G. max* accessions using adjusted phenotype data for seed oil and protein content. Manhattan plots show the associations for seed protein and oil with SNP markers that are plotted on the x-axis according to their physical position on each chromosome. The solid horizontal line denotes the calculated threshold value for declaring significant association. The dashed vertical lines indicate that the significant association positions on chromosome 15 and 20 for protein were the same as for those oil.

**Supplementary Tables**

**Supplementary Table S1.** Distribution of 14,430 soybean accessions among seven arbitrarily chosen world regions (column names) by collation of accessions originating from 13 different specific world sub-regions (row names).

**Supplementary Table S2.** Distribution of 14,430 soybean accessions among eight maturity group MG classes. Of the 13 individual MGs (i.e. row numbers), the earliest maturing three (000, 00, 0), and the latest maturing three (VIII, IX, X) were combined (because of low accession numbers) to create just eight MG classes.

**Supplementary Table S3.** Accession identification numbers and seed oil and protein content phenotypes for 12,116 G max accessions used for GWAS.

**Supplementary Table S4.** Number of accessions within each subpopulation assigned to a given species based on membership coefficient criterion of >0.8. Accessions with a membership coefficient <0.8 were considered admixed.

**Supplementary Table S5.** Number of accessions within each subpopulation assigned to each world region class based on membership coefficient criterion of >0.8. Accessions with a membership coefficient <0.8 were assigned to the admixed group.

**Supplementary Table S6.** Number of accessions within each subpopulation assigned to each MG class based on membership coefficient criterion of >0.8. Accessions with a membership coefficient <0.8 were assigned to the admixed group.

**Supplementary Table S7.** Mean ancestry estimates of accessions within each world region relative to the subpopulation assignment. Estimates also provided for the 34 accessions that comprise a group known as USA soybean ancestors (Gizlice et al., 1994; Li and Nelson, 2001; Ude et al., 2003).