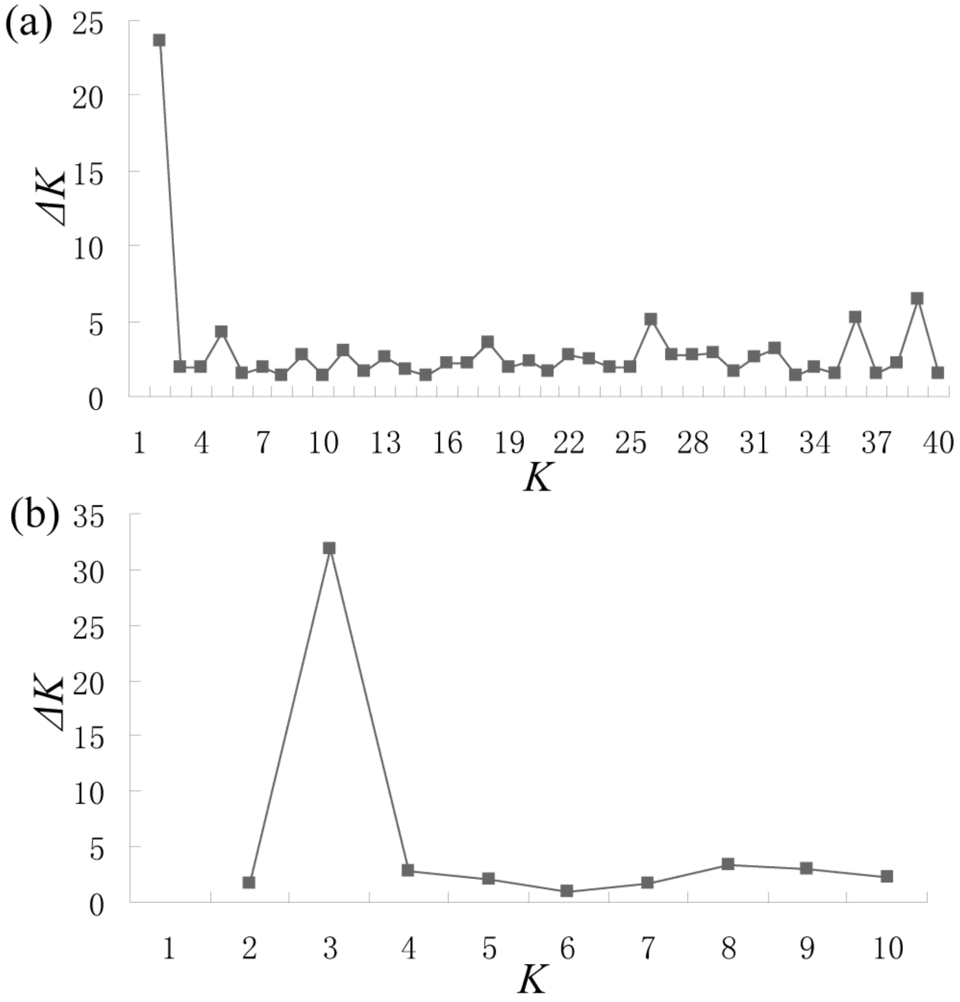
SUPPLEMENTARY DATA

TABLE S1. Genetic statistics of the 20 microsatellite loci used in the study

Locus	Repeat motif	Chromosome number	N_A	H_E
Satt174	(ATT)10	A1	11	0.817
Satt429	(ATT)25	A2	23	0.894
Sat_128	(ATT)31	B1	22	0.892
Satt556	(ATT)14	B2	21	0.920
Sat_140	(AT)13	C1	27	0.901
Satt227	(ATT)9	C2	11	0.800
Satt179	(ATT)25	D1a	26	0.928
BE021153	(GT)10	D1b	5	0.407
Satt082	(ATT)13	D2	6	0.707
Satt369	(ATT)17	E	17	0.810
Satt114	(ATT)17	F	15	0.793
Satt324	(ATT)19	G	12	0.806
U08405	(AT)10	H	16	0.754
Satt571	(ATT)14	I	21	0.913
AW310961	(ATT)16	J	21	0.911
Satt240	(ATT)11	K	12	0.770
Satt166	(ATT)19	L	19	0.814
Satt175	(ATT)16	M	18	0.902
Sct_195	(CT)11	N	6	0.783
BE801128	(CAA)13	O	8	0.736
Overall		20	317	0.813

N_A: number of detected alleles. H_E: expected heterozygosity.

FIG. S1. Inference of genetic cluster (*K*) of wild soybean based on the Δ*K* value. Δ*K* value was estimated as per Evanno *et al*. (2005) as a function of *K*: (a) for the 40 wild soybean populations from China based on 20 microsatellites; and (b) for representative wild soybean individuals from south China, north-east China, Russian Far East, South Korea and Japan, based on 56 microsatellites.



REFERENCE

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of the number of clusters of indiciduals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.