File S2. The meta-analysis of *S. pimpinellifolium* SolCAP genotyping data.

MATERIALS AND METHODS

The SolCAP data of 214 samples of *S. pimpinellifolium* were downloaded from previous studies (Blanca *et al.* 2012, 2015; Sim *et al.* 2012a). A set of 2,934 bi-allelic polymorphic SNPs was extracted after being filtered with the criteria that minor allele frequency is more than 0.05 and the proportion of missing genotypes is less than 25%. Because 627 SNP markers were found to be reverse-complement allele designation among these three studies, we dropped these markers and obtained 2,307 SNPs with consistent allele designation. This set of 2,307 SNPs was utilized in the analyses of ADMIXTURE and isolation by distance following the same procedures described in Materials and Methods and File S1, respectively. Meanwhile, because some accessions were genotyped in more than one SolCAP studies, different suffixes—"_2012S," "_2012B," and "_2015B,"—were added to the sample name to indicate their original references Sim *et al.* 2012a, Blanca *et al.* 2012, and Blanca *et al.* 2015, respectively. Also, the percentage of identical SNP genotypes of the same accessions were calculated based on the 2,307 SNP genotypes without missing values.

RESULTS

Meta-analysis of SolCAP genotyping array resulted in 15 subpopulations

To compare with our analysis of the genetic differentiation of *S. pimpinellifolium* in the current study, we performed a meta-analysis of the

genetic differentiation of *S. pimpinellifolium* using combined SNP-marker genotypic data of SolCAP array from the previous studies. We downloaded the genotypes of 214 samples representing 126 accessions from three previous studies (Blanca *et al.* 2012, 2015; Sim *et al.* 2012a) and conducted the meta-analysis using our workflow (please see details in the "Materials and Methods" section). Initially, we extracted a marker set of 2,934 bi-allelic SNPs to investigate genetic diversity between samples from different studies but tagged the same name. The samples in Blanca *et al.*, 2012 separated from those of the other two studies in the PCA plot (Figure S9A), while most of the accessions in Blanca *et al.*, 2012 were involved in the study of Blanca *et al.*, 2015 (Table S5). It suggested that the batch effect occurred when these datasets merged. Considering the SolCAP genotyping array is an Illumina bead array, which uses the TOP/BOT strand and A/B allele designation to assign the actual polymorphism of samples, data merging might introduce reverse-complement allele designation (Illumina 2014). We resolved the problem of the batch effect after we removed the markers with inconsistent SNP assignment among these three datasets (Table S6 and Figure S9B). The genotypic data of 2,307 SNPs in 214 samples was remained (Table S5) and used to conduct further analyses. ADMIXTURE suggested the best K equaled to 15 (Figure S7 and Figure S8). Also, the correlation coefficient between the genetic distance and geographic distance was 0.55, and this correlation was statistically significant (p-value < 0.001) (Figure S6).

LITERATURE CITED

Blanca, J., J. Cañizares, L. Cordero, L. Pascual, M. J. Diez *et al.*, 2012

Variation Revealed by SNP Genotyping and Morphology Provides Insight

into the Origin of the Tomato. PLoS One 7(10): e48198.

Blanca, J., J. Montero-Pau, C. Sauvage, G. Bauchet, E. Illa *et al.*, 2015

Genomic variation in tomato, from wild ancestors to contemporary

breeding accessions. BMC Genomics 16: 257.

Illumina, 2014 Infinium ® Genotyping Data Analysis. Tech. Note 10p.

Sim, S. C., G. Durstewitz, J. Plieske, R. Wieseke, M. W. Ganal *et al.*, 2012b

Development of a large SNP genotyping array and generation of

high-density genetic maps in tomato. PLoS One 7: e40563.