**Figure S1. Analysis and classification of DN+|MTP- and DN-|MTP+ nonreference pan-genes in *A. thaliana*.** The Venn diagram shows the overlap between the sets of nonreference pan-genes detected by the DN and MTP pan-genome construction approaches. Transcript sequences of DN+|MTP- pan-genes (black bars) were mapped to the reference genome sequence (grey bars) and the MTP nonreference sequences (purple bars). DN-|MTP+ genes were analyzed by mapping their protein sequences (dotted bars) to reference proteins (red striped bars) and to DN nonreference proteins (striped bars).

**Figure S2. Comparison of soybeanpan-genomes constructed using the DN and MTP approaches.** The same input data were used for constructing two pan-genomes containing eight cultivated soybean accessions, one using the DN approach and the other using the MTP approach. **(A)** The overall DN and MTP pan-genome sizes and compositions – number of reference (grey) and nonreference (purple) pan-genes, as well as core (green), shell (light blue), and singletons (orange). **(B)** Number of reference and nonreference pan-genes detected by the two approaches in each ecotype.

**Figure S3. Meta-analysis of published plant pan-genomes**. Fifteen published plant pan-genomes constructed using the Map-to-pan (orange) and De novo (green) approaches were analyzed. **(A)** Correlation between two pan-genome composition measures: the percentage of core pan-genes (present in 95% of pan-genome accessions), and the overall gene occupancy. **(B)** Correlation between the number of accessions included in a pan-genome and the percentage of core pan-genes. **(C)** Correlation between the number of accessions included in a pan-genome and the overall gene occupancy.

**Figure S4. Subsampling simulation of *B. napus* and rice MTP pan-genomes.** Previously published *B. napus* (**A-C**) and rice (**D-F**) pan-genomes constructed using the DN and MTP approaches were compared. We controlled for the difference in the number of accessions by randomly subsampling the MTP pan-genomes to match the number of accessions in the corresponding DN pan-genomes (9 for *B. napus* and 67 for rice), and then computed the pan-genome composition metrics from the subsampled data sets. This procedure was repeated 100 times to produce distributions of the percentage of core pan-genes (**A,D**), the overall gene occupancy (**B,E**), and the percentage of nonreference pan-genes (**C,F)**. Red arrows indicate the values observed for the DN pan-genomes.

**Figure S5. Analysis of *A. thaliana* gene presence-absence detection thresholds in the Map-to-pan approach. (A)** Distributions of the fractions of genes covered by mapped reads in a single *A. thaliana* ecotype (An-1) using different depth thresholds on 50× sequencing data. **(B)** The percentage of genes determined as present as a function of the coverage threshold, using different depth thresholds on 50× sequencing data of a single *A. thaliana* ecotype (An-1). **(C)** The percentage of core pan-genes as a function of the coverage threshold in an *A. thaliana* pan-genome containing eight accession, using different depth thresholds on 50× sequencing data. **(D)** The overall gene occupancy as a function of the coverage threshold in an *A. thaliana* pan-genome containing eight accession, using different depth thresholds on 50× sequencing data. **(E)** The percentage of genes determined as present as a function of the sequencing depth using different depth fraction thresholds on 50× sequencing data of a single *A. thaliana* ecotype (An-1).