Supplementary file

Supplementary Table 1. Statistics of SNPs identified from the tomato reference genome using own raw data.

SNP type	Pileup (default)
Homo-type	10,699 (12.2%)
ref = N	4,647
SNP	6,052
Hetero-type	77,230 (87.8%)
ref = N	120
SNP	77,110
Total # of SNPs	87,929

Supplementary Table 2. Comparison of commonly detected (A) homo-type and (B) hetero-type SNPs across the sequence coverage between pileup and new_pileup. The first number in each cell corresponds to SNP callers produced by the pileup data set. Data produced by the new_pileup is indicated in parentheses.

(A)

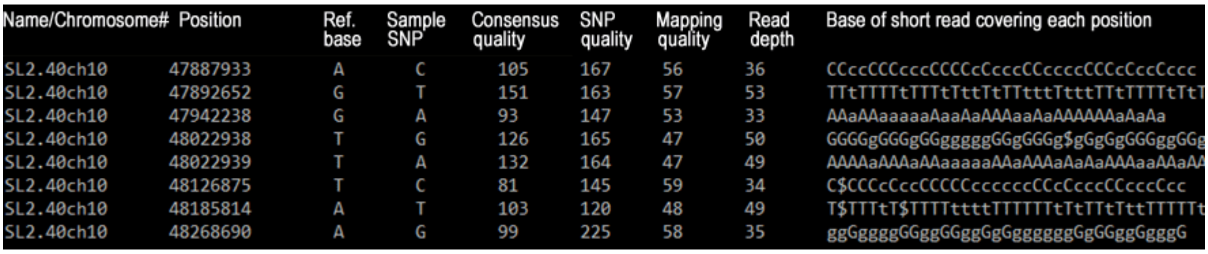
coverage	2×	10×	20×	40×
2×	3,099 (7,469)[a]			
10×	1,360 (5,046)	388 (1,477)[a]		
20×	1,172 (4,502)	4,905 (9,300)	108 (575)[a]	
40×	1,148 (4,076)	4,844 (8,573)	5,526 (9,475)	471 (980)[a]

(B)

coverage	2×	10×	20×	40×
2×	51,244 (23,695)[a]			
10×	32,153 (34,151)	39,665 (103,348)[a]		
20×	24,050 (29,357)	79,515 (121,458)	13,373 (23,871)[a]	
40×	12,158 (2,959)	51,114 (107,940)	64,112 (127,867)	11,102 (23,924)[a]

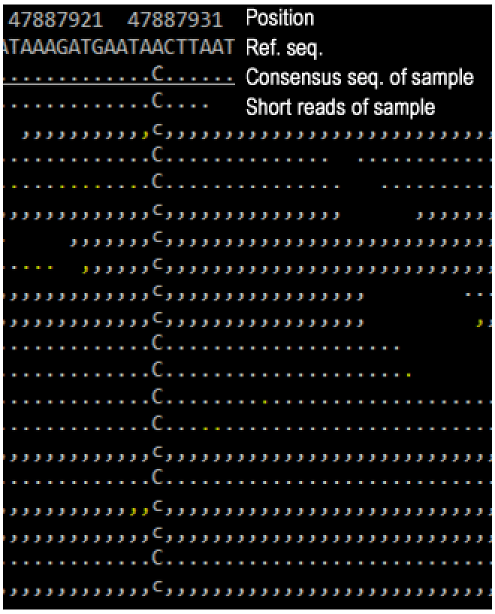
[a] The number of SNPs found in one sample.

(A)



Name/Chromosome#	Position	Ref. base	Sample SNP	Consensus quality	SNP quality	Mapping quality	Read depth	Base of short read covering each position
SL2.40ch10	47887933	A	C	105	167	56	36	CCccCCCccccCCCCcCcccCCccccCCCcCccCccc
SL2.40ch10	47892652	G	T	151	163	57	53	TTtTTTTtTTTtTttTtTTtttTtttTTtTTTTtTtT
SL2.40ch10	47942238	G	A	93	147	53	33	AAaAAaaaaaAaaAaAAAaaAaAAAAAAaAaAa
SL2.40ch10	48022938	T	G	126	165	47	50	GGGGgGGGgGGgggggGGgGGGg$gGgGgGGGggGGg
SL2.40ch10	48022939	T	A	132	164	47	49	AAAAaAAAaAAaaaaaAAaAAAaAaAaAAAaaAAaAA
SL2.40ch10	48126875	T	C	81	145	59	34	C$CCCcCccCCCCCccccccCCcCcccCCcccCcc
SL2.40ch10	48185814	A	T	103	120	48	49	T$TTTtT$TTTTttttTTTTTTtTtTTtTttTTTTTt
SL2.40ch10	48268690	A	G	99	225	58	35	ggGggggGGggGGggGgGggggggGgGGggGgggG

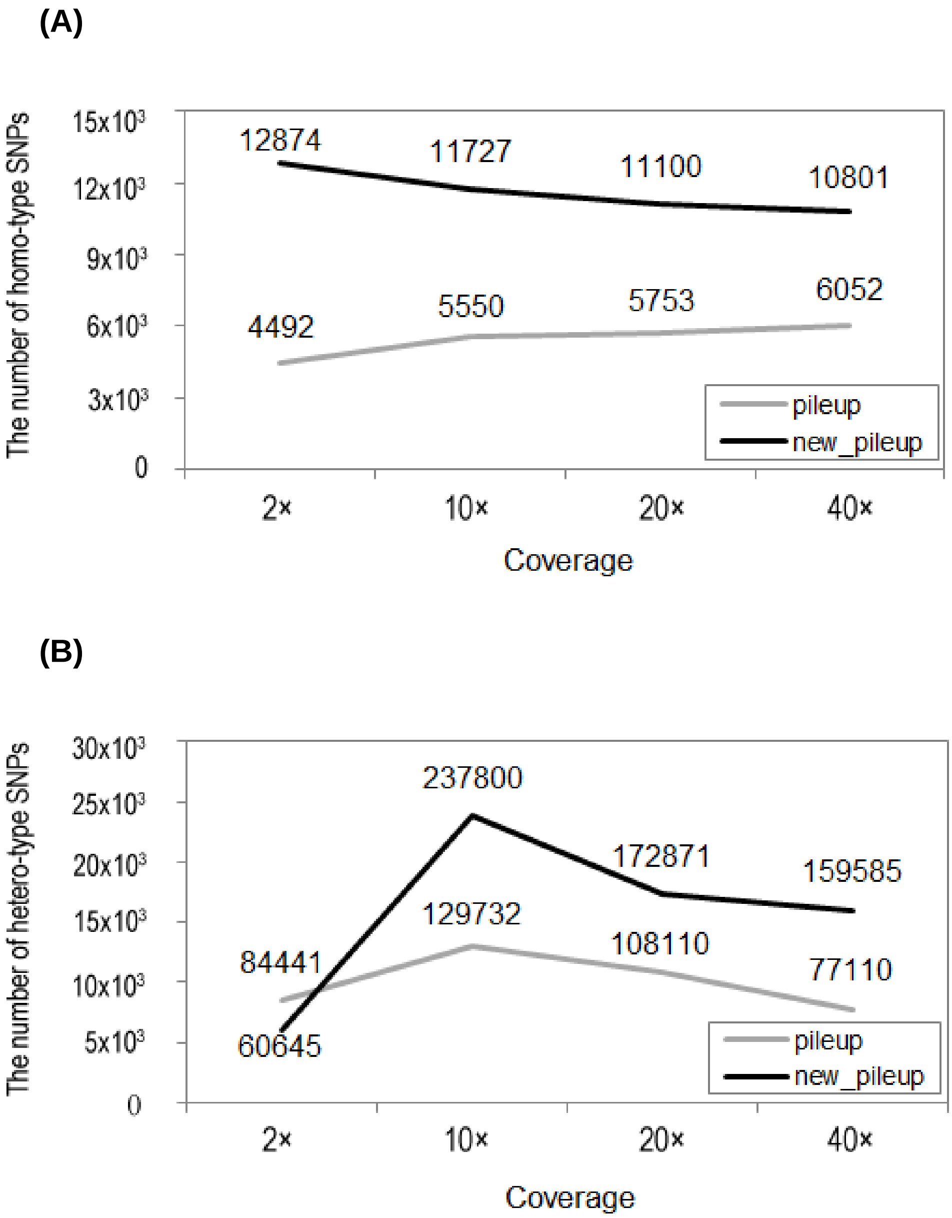
(B)



Supplementary Figure 1. Snapshot of alignment with BWA sequence-based SNP validation. (A) Eight SNPs positioned between 47,887,933 and 48,268,690 were discovered on chromosome 10. (B) T-view display of the result of pileup mapping of SNP SL2-40ch10_47,887,933. In (A), Name, name of the project; Chromosome #, the number of the chromosome; Position, position within the chromosome; Ref. base, nucleotide of the reference genome (Heinz 1706); Sample SNP, the consensus base of sample (Heinz 1706); Consensus quality, the phred-scaled probability that the genotype call is correct; SNP quality, the phred-scaled probability that an inferred SNP is identical to the reference; Mapping quality, root mean square mapping quality;

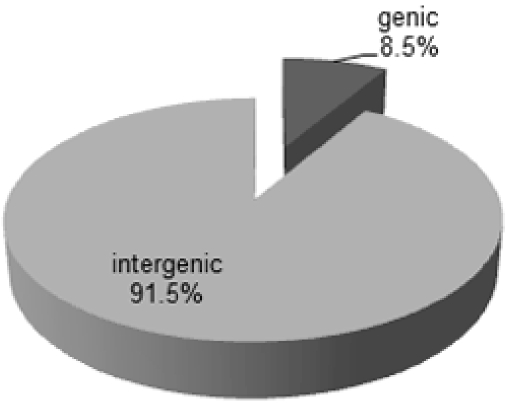
Read depth, the number of reads covering the position; Base of short read covering each SNP position, the aligned bases of all reads covering this genomic position.

In (B), Position, position within the chromosome; Ref. seq., sequence of the reference genome (Heinz 1706); Consensus seq., the consensus sequence of sample (Heinz 1706); Short reads of sample, the aligned all reads to reference sequence.

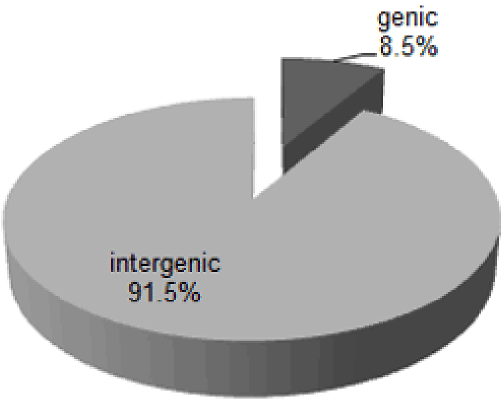


Supplementary Figure 2. Comparison of the number of homo- and hetero-type SNPs detected by pileup versus new_pileup. Line graph showing the number of (A) homo-type and (B) hetero-type SNPs identified by pileup and new_ pileup analysis according to the genome coverage of raw data.

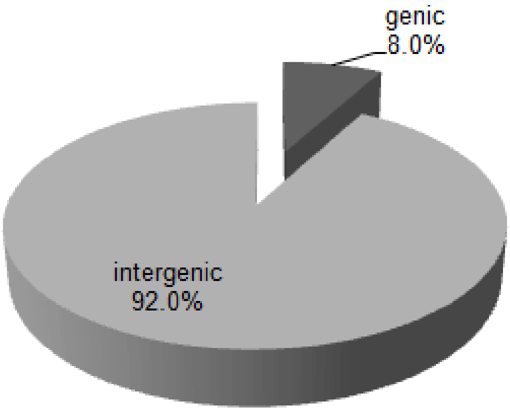
(A)


genic
8.5%
intergenic
91.5%

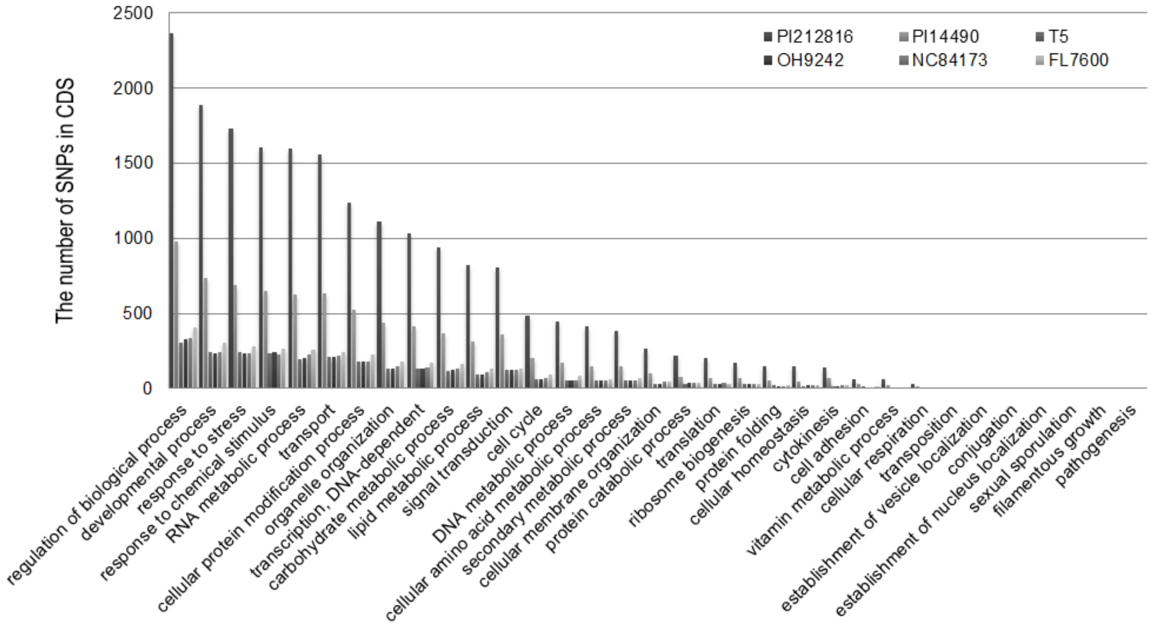
(B)


genic
8.5%
intergenic
91.5%

(C)


genic
8.0%
intergenic
92.0%

Supplementary Figure 3. The distribution of identified SNPs according to genome structure. The percentage of (A) total SNPs, (B) homo-type SNPs, and (C) hetero-type SNPs detected in *S. pimpinellifolium* are shown. More than 90 percent of all SNPs were located within intergenic regions.



Supplementary Figure 4. Gene ontology classifications for six transcriptome data sets. SNPs identified within the coding region (CDS) corresponded to the biological process (BP) category. The y-axis indicates the number of SNPs represented from six transcriptomes (except M82) while the x-axis indicates the GO term of BP with depth equal to 2.