Supplementary Information

The online version contains supplementary material available at

Table S1: List of 305 samples collection information in this study.

Table S2: BLAST results of *SNP-associated* contigs of 159,372 against the Chines pine reference genome with identity ≥ 95% and unique location.

Table S3: The identified 20,055 SNPs set.

Table S4: BLAST results of *SNP-associated* contigs compared with the non-redundant (nr) protein database.

Table S5: Gene Ontology (GO) annotations results from the 159,372 *SNP-associated* contigs.

Table S6 The Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations results based on the 159,372 *SNP-associated* contigs.

Figure S1. Workflow for the application of genotyping-by-sequencing in combination with technical replicates filtering and homologous reference genome blasted and validated for reliable, informative SNPs and contigs in the development of genomic resources in Masson pine genomic research.

DNA samples with replicates → Enzyme-based DNA libraries preparation and sequencing

de novo assembly using MEGAHIT ← GBS sequenced reads quality control using FastQC and Trimmomatic

Blasted against reference genome of *Pinus taeda* with identiy > 95% → Generating bam files with *de novo* reference using ref-ANGSD

SNP filtering missing < 10% *GT1_del_dupl_missing.pl* ← SNP calling using ANGSD

Plink format and filtered with the MAF > 0.01, missing < 10% → *SNP-associated* contigs blasted against reference genome of *Pinus tabuliformis* with identity >=95% and unique location remained

Reference-located-annotated SNP and replicates filtered SNPs ← *Reference-located-SNP-associated* contigs functional annotation and validation

MAF > 0.05, Missing < 5% and mC >= 4 (These parameters can be changed) → SNPs with 35bp distance filtered

Genetic analysis validated the informativeness of *functional-related* SNP ← *Blasted-reference-located-annotated* contigs *and functional-related* SNPs with 35bp distance

Developing genetic resources in Masson pine with reliable, informative SNP and contigs obtained via GBS approach
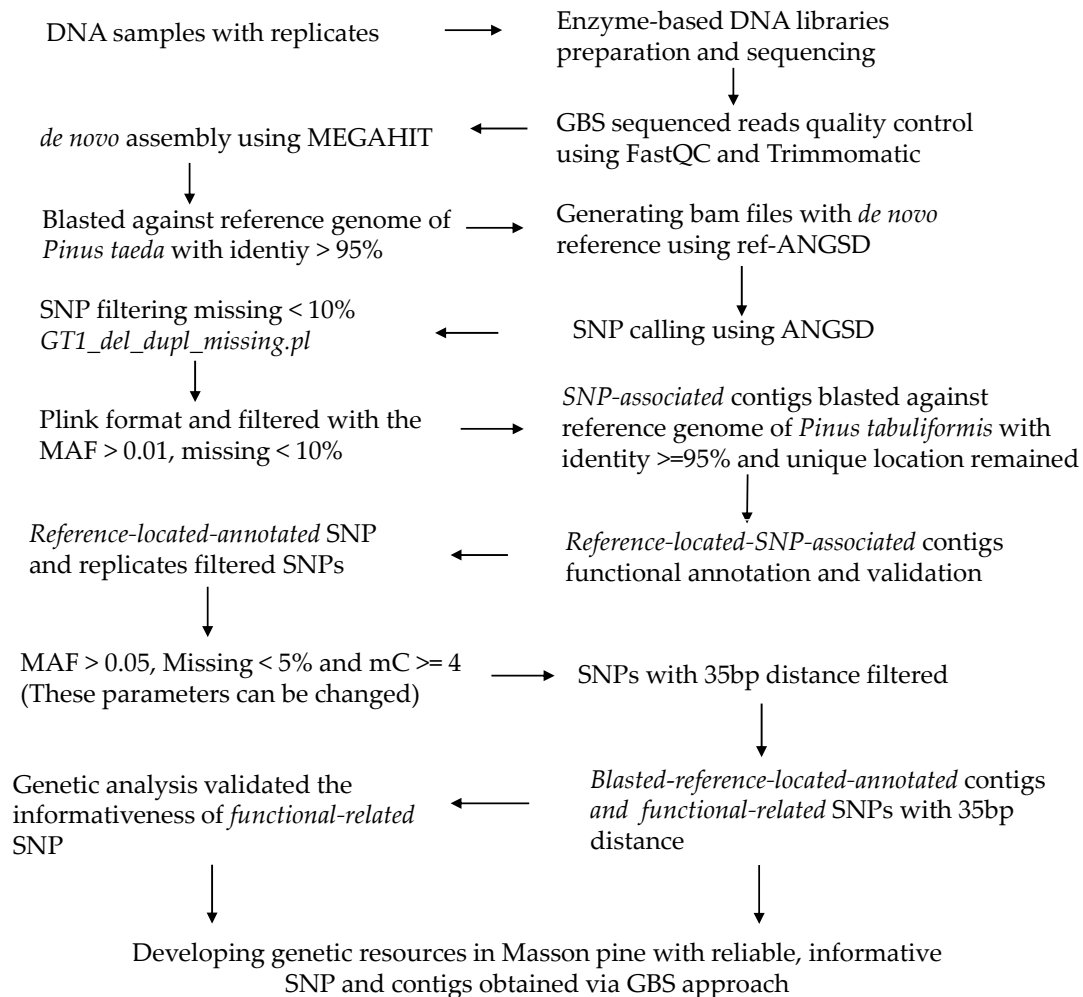
Figure S2. Gene Ontology classification of *SNP-associated* contigs identified from GBS in Masson pine (A), KEGG Pathway details of annotated contigs in Masson pine (B).
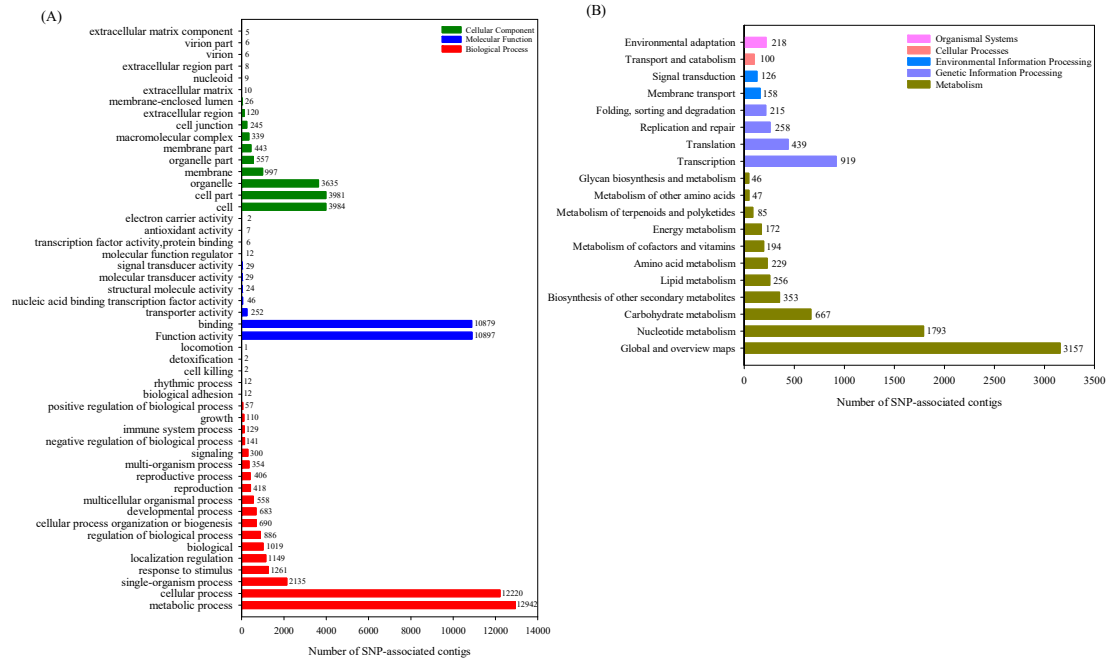
(A)

Category	Count
extracellular matrix component	5
virion part	6
virion	6
extracellular region part	8
nucleoid	9
extracellular matrix	10
membrane-enclosed lumen	26
extracellular region	120
cell junction	245
macromolecular complex	339
membrane part	443
organelle part	557
membrane	997
organelle	3635
cell part	3981
cell	3984
electron carrier activity	2
antioxidant activity	7
transcription factor activity,protein binding	6
molecular function regulator	12
signal transducer activity	29
molecular transducer activity	29
structural molecule activity	24
nucleic acid binding transcription factor activity	46
transporter activity	252
binding	10879
Function activity	10897
locomotion	1
detoxification	2
cell killing	2
rhythmic process	12
biological adhesion	12
positive regulation of biological process	57
growth	110
immune system process	129
negative regulation of biological process	141
signaling	300
multi-organism process	354
reproductive process	406
reproduction	418
multicellular organismal process	558
developmental process	683
cellular process organization or biogenesis	690
regulation of biological process	886
biological	1019
localization regulation	1149
response to stimulus	1261
single-organism process	2135
cellular process	12220
metabolic process	12942

Legend: Cellular Component, Molecular Function, Biological Process

Number of SNP-associated contigs

(B)

Category	Count
Environmental adaptation	218
Transport and catabolism	100
Signal transduction	126
Membrane transport	158
Folding, sorting and degradation	215
Replication and repair	258
Translation	439
Transcription	919
Glycan biosynthesis and metabolism	46
Metabolism of other amino acids	47
Metabolism of terpenoids and polyketides	85
Energy metabolism	172
Metabolism of cofactors and vitamins	194
Amino acid metabolism	229
Lipid metabolism	256
Biosynthesis of other secondary metabolites	353
Carbohydrate metabolism	667
Nucleotide metabolism	1793
Global and overview maps	3157

Legend: Organismal Systems, Cellular Processes, Environmental Information Processing, Genetic Information Processing, Metabolism

Number of SNP-associated contigs

Figure S3. Highlighting of YQ and ZB subpopulations in the PCoA plots of 299 Masson pines based on the SNPs obtained without or with replicates in different filtering scenarios from zero to six pairs. Panel (A) to (F) were the PCoA plots in SNPs of *rep0* (443,571) (A), *rep1* (435,039) (B), *rep6* (153,942) (C), *rep6blast* (134,611) (D), *rep6blastAnnot* (60,143) (E) and *rep6blastAnnot35bp* (F) (20,055). The different signal of colored circles with the abbreviations representrepresents the original resource of YQ and ZB subpopulation, the blank circles represent all other samples from four local populations displayed in Table S1.