Supplementary Information

Table S1 Information for the reads, read depth, variants and indels

Information	Total
Sequence used for exome capture probe design	112,626
Total Number of Samples	121
Read length	2 x 150 bp
Total reads for all samples	4,267,734,316
Average reads per sample	35,270,531
Average read depth per sample	9.564
Raw variant Number	13,474,028
Total SNPs	11,216,753
Total Insertions	1,652,481
Total Deletions	597,883
Total SNPs after filtering (within and outside gene)	591,919
Total SNPs within genes	313,985

Table S2 Individual heterozygosity of plants from different location, generation and health

Location	Health	Generation	Individual/ pool	N	Average	Std	Max	Min
Washington	Strong	S_1	pool	3	0.611	0.065	0.682	0.555
		S_2	individual	9	0.389	0.040	0.463	0.333
	Weak	S_1	pool	3	0.515	0.031	0.549	0.486
		S_2	individual	9	0.348	0.053	0.417	0.248
Utah	Strong	S_1	pool	2	0.541	0.141	0.642	0.441
		S_2	pool	11	0.362	0.086	0.468	0.196
	Weak	S_1	pool	2	0.419	0.007	0.425	0.414
		S_2	pool	12	0.519	0.065	0.608	0.433
Wisconsin	Parental clone	S_0	individual	5	0.417	0.056	0.481	0.362
	Strong	S_1	individual	56	0.316	0.089	0.486	0.115
	Top, low and No Seeded	S_2	pool	9	0.373	0.095	0.492	0.203

Table S3 Summary statistics from GWAS result using FarmCPU with 588,136 SNPs.

Chromosome	Position	P.value	MAF	Effect	SNP	PAE(%)
6	41916109	6.54E-20	0.1446	0.352	S6_41916109	8.556
8	68992917	2.02E-19	0.1529	0.296	S8_68992917	4.831
1	23794794	1.98E-16	0.0744	-0.319	S1_23794794	5.304
4	53208128	3.55E-12	0.0537	-0.242	S4_53208128	3.494
8	87604023	1.35E-11	0.0826	-0.208	S8_87604023	4.511
5	80794484	3.13E-11	0.0868	-0.216	S5_80794484	6.017
7	82129849	3.81E-11	0.0620	-0.211	S7_82129849	2.445
7	22499070	8.43E-11	0.0537	-0.274	S7_22499070	9.740
3	2154144	2.04E-10	0.1157	0.171	S3_2154144	1.026
7	31387748	1.69E-08	0.1116	0.150	S7_31387748	1.793
6	24909781	3.57E-08	0.0868	-0.228	S6_24909781	2.667

PAE - Phenotypic Variation Explained
MAF - Minor Allele Frequency
SNP - Single Nucleotide Polymorphism

Table S4 MAF of individual strong and weak group as well as pooled strong and weak groups over significant SNPs.

CHR	POS	IND	IND Strong	IND Weak	Pooled	Pooled Strong	Pooled Weak
6	41916109	0.031	0.021	0.111	0.154	0.045	0.275
8	68992917	0.088	0.092	0.055	0.166	0.181	0.050
1	23794794	0.025	0.014	0.111	0.107	0.045	0.175
4	53208128	0.044	0.014	0.277	0.166	0.136	0.200
8	87604023	0.013	0.007	0.055	0.226	0.159	0.300
5	80794484	0.088	0.079	0.166	0.250	0.227	0.125
7	82129849	0.013	0.014	0.001	0.130	0.023	0.250
7	22499070	0.056	0.064	0.001	0.214	0.227	0.150
3	2154144	0.025	0.014	0.111	0.130	0.136	0.125
7	31387748	0.063	0.064	0.055	0.321	0.272	0.225
6	24909781	0.056	0.036	0.222	0.130	0.090	0.175

CHR - Chromosomes

POS - Position

IND - Individual

Figure S1. Selection of pairs of strong and weak plants within self-pollinated inbred lines. Six seeds were planted individually for each self-pollinated individual (A). The individual (the middle column) with a significantly deficient progeny was identified for sampling. The deficient progeny and the strongest progeny within the individual lines were sampled (B). The deficient plants must have sufficient leaves to extract DNA (C).

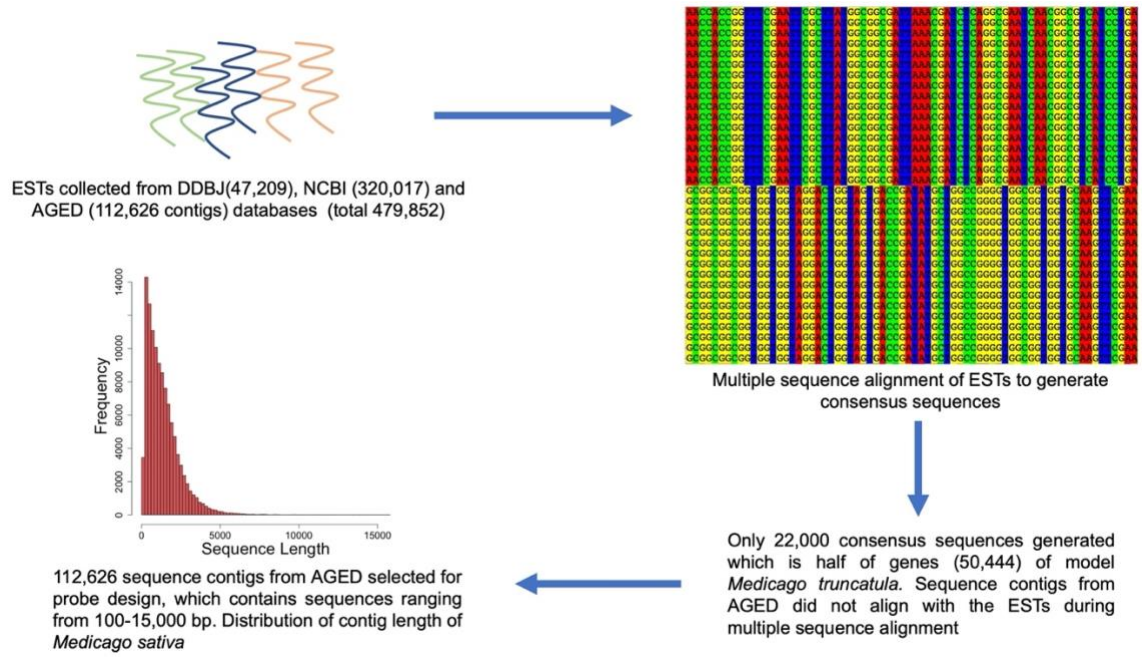

Figure S2. Design of exome capture chips for genotyping alfalfa accession in this study. DDBJ (DNA Data Bank of Japan), NCBI (National Center for Biotechnology Information), and AGED (Alfalfa Gene Index and Expression Database). ESTs are the Expressed Sequence Tags collected from three public databases.

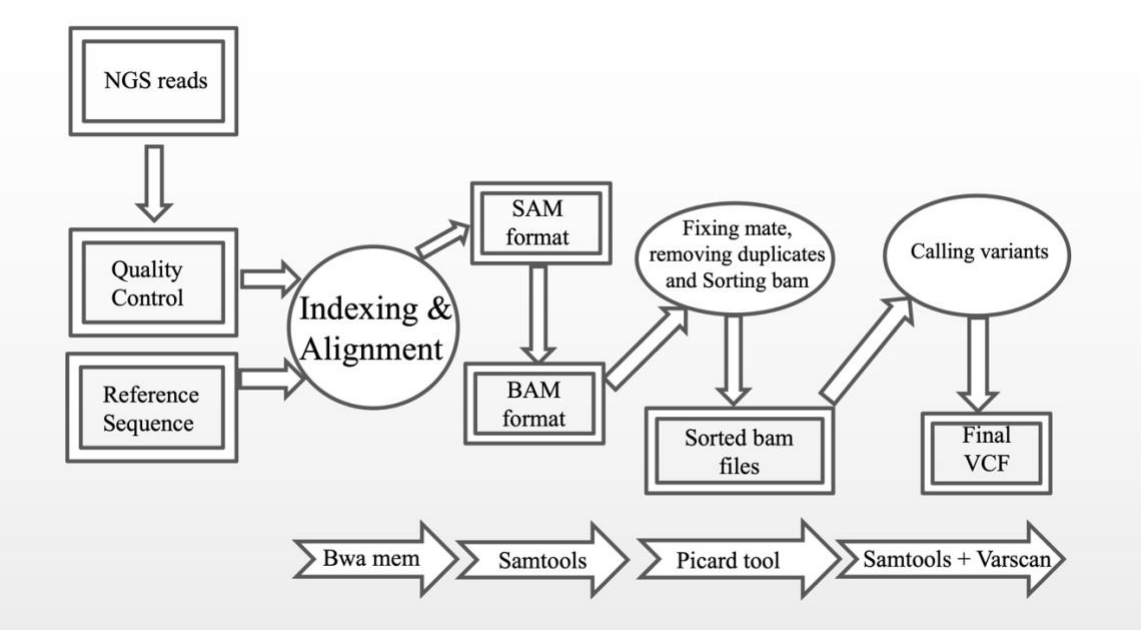

Figure S3. SNP calling pipeline for exome capture sequencing. The SNP calling pipeline is presented in the order it is performed. The bold arrow represents the workflow, and the circle and rectangular boxes represent the activities performed during the workflow. The gray arrow at the bottom with the arrow facing from left to right includes all the tools that were used to generate the respective results in the analysis.

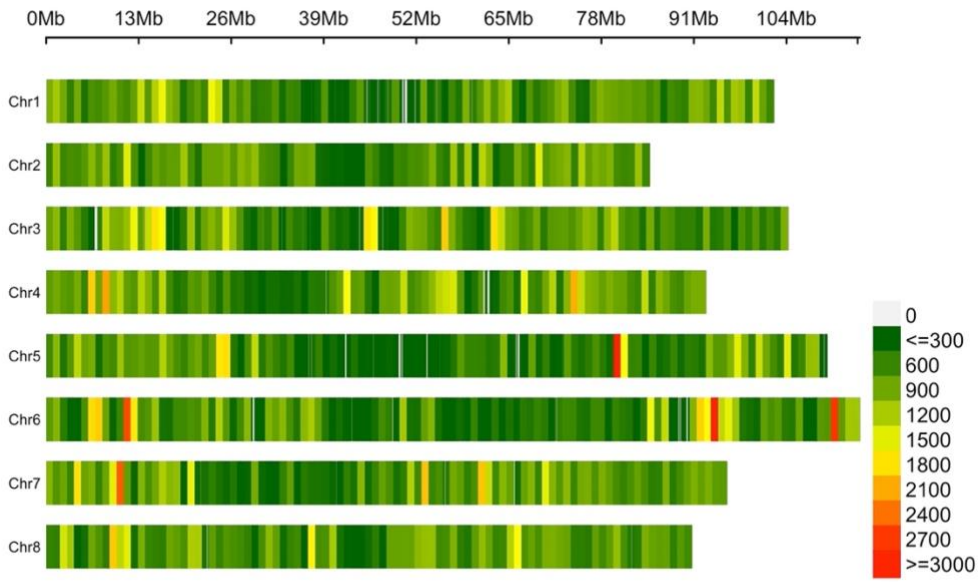

Figure S4. **Genome-wide distribution of SNPs on all the Chromosomes.** The horizontal axis shows the chromosome length while the colors within each tile represent SNPs density within a 1 Mb window size.

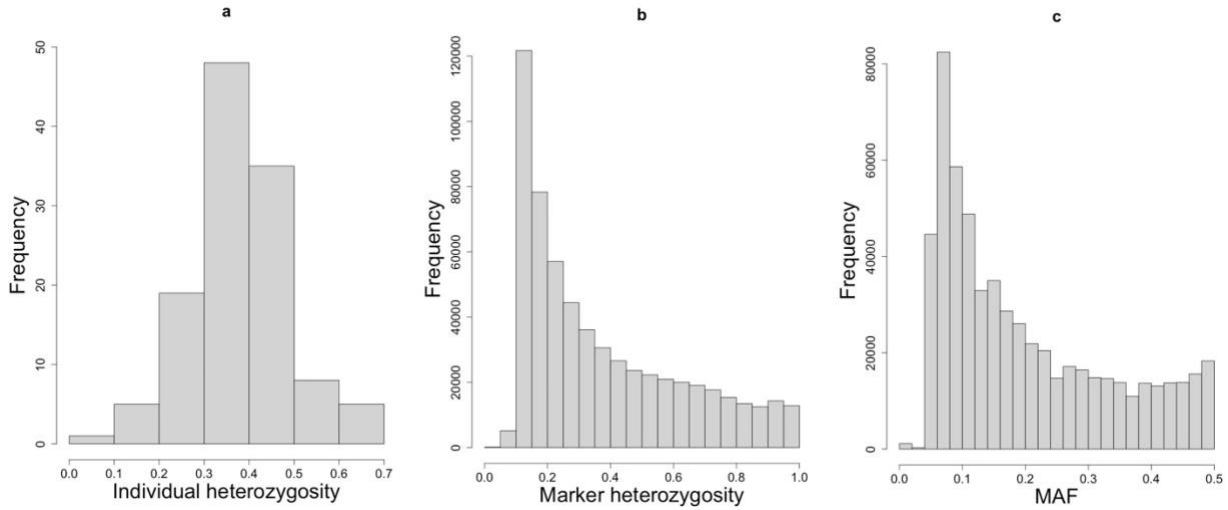

Figure S5. The numbers of sequencing reads, read depth, and its distribution. The total reads and average read depth shows increasing trends with an average of 30 million for total reads and around 9.5 for average read depth. The distribution of average sequencing depth shows most of the samples with an average depth of 8-12.

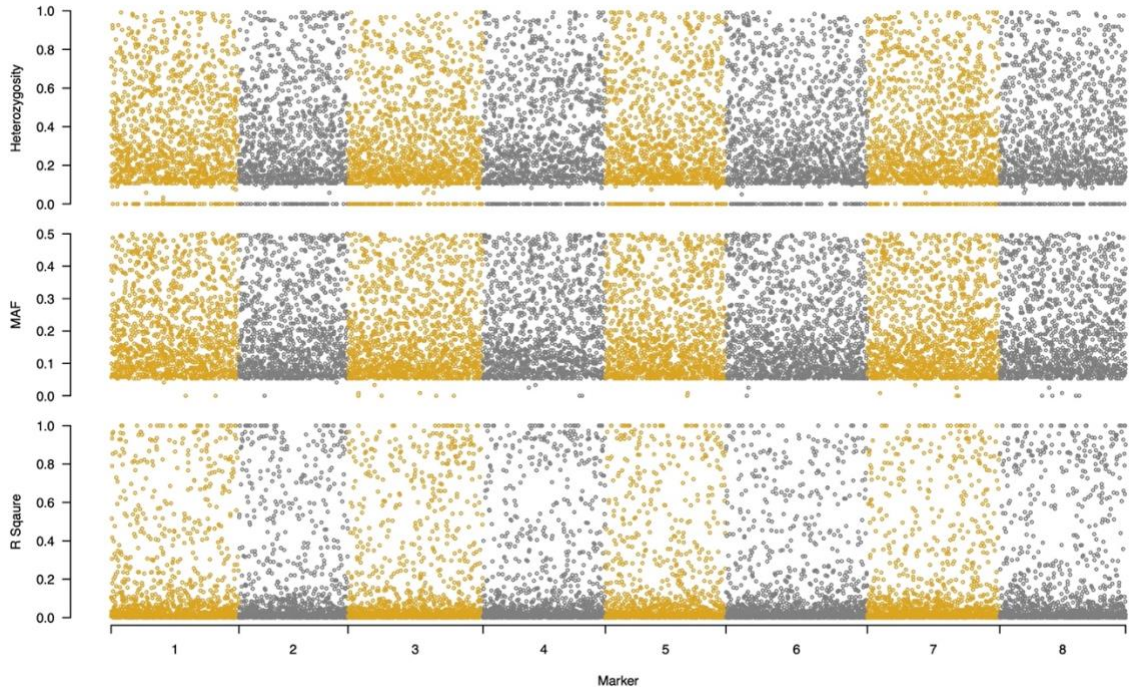

Figure S6. Heterozygosity, minor allele frequency, and linkage disequilibrium across alfalfa genome. The SNPs exhibit low heterozygosity across the genome. Each SNP is represented as a dot with heterozygosity as the vertical axis and genome position as the horizontal axis (a). Minor allele frequency (b) and linkage disequilibrium (c) are illustrated similarly. Linkage disequilibrium was calculated as the squared Pearson correlation coefficient between the adjacent SNPs.

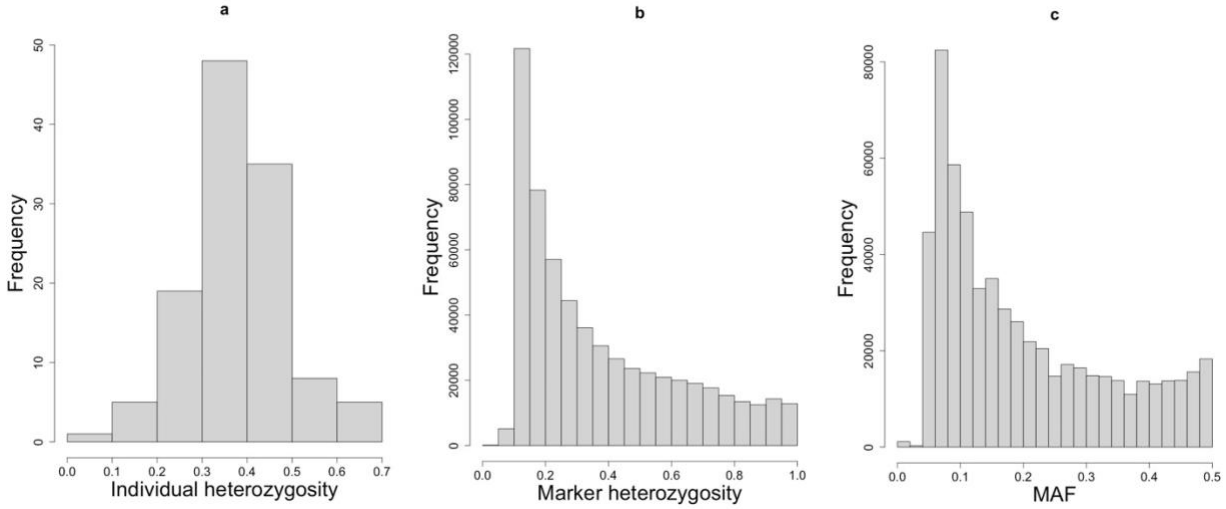

Figure S7. Heterozygosity and minor allele frequency of single nucleotide polymorphisms (SNPs). Heterozygosity was calculated for individual varieties (a) and markers (b). The MAF distribution is a skewed uniform distribution with more density on the 0 sides than on the 0.5 sides (c).