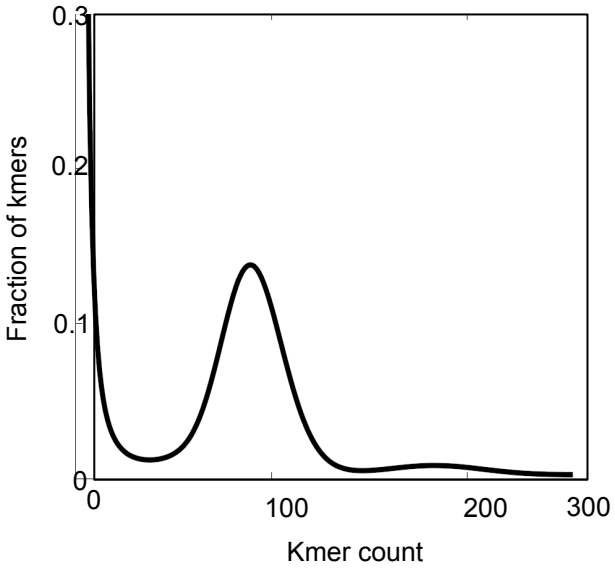
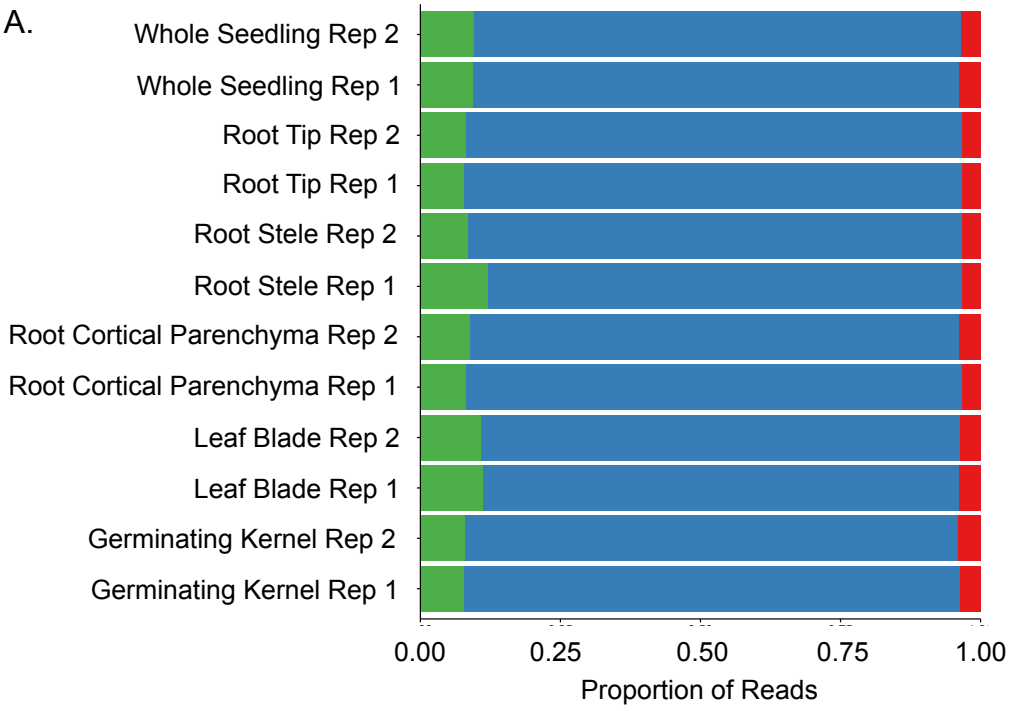
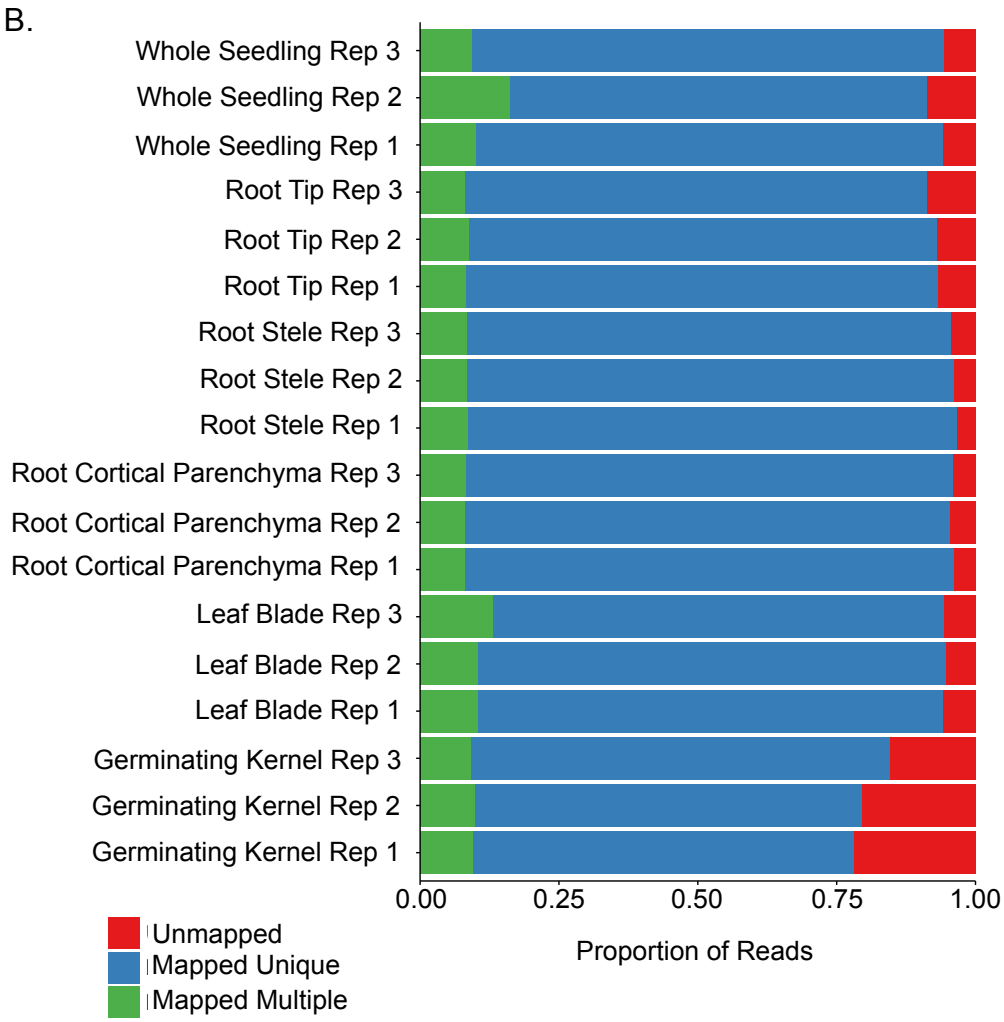
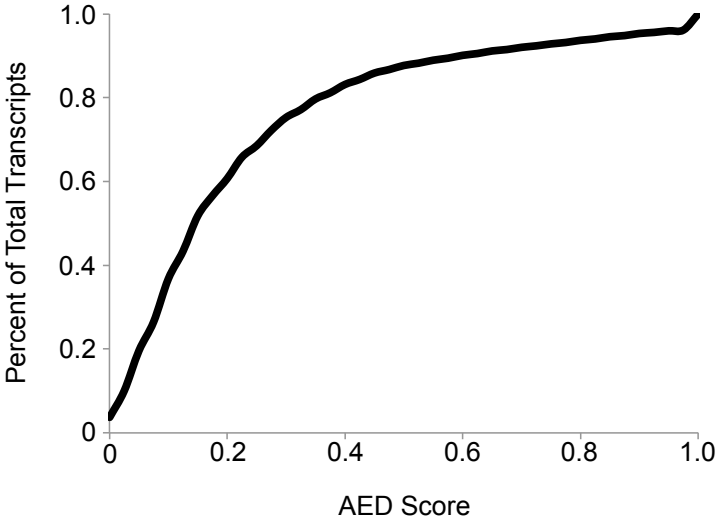


Supplemental Figure 1. Distribution of 23-mers in the PH207 sequence reads.

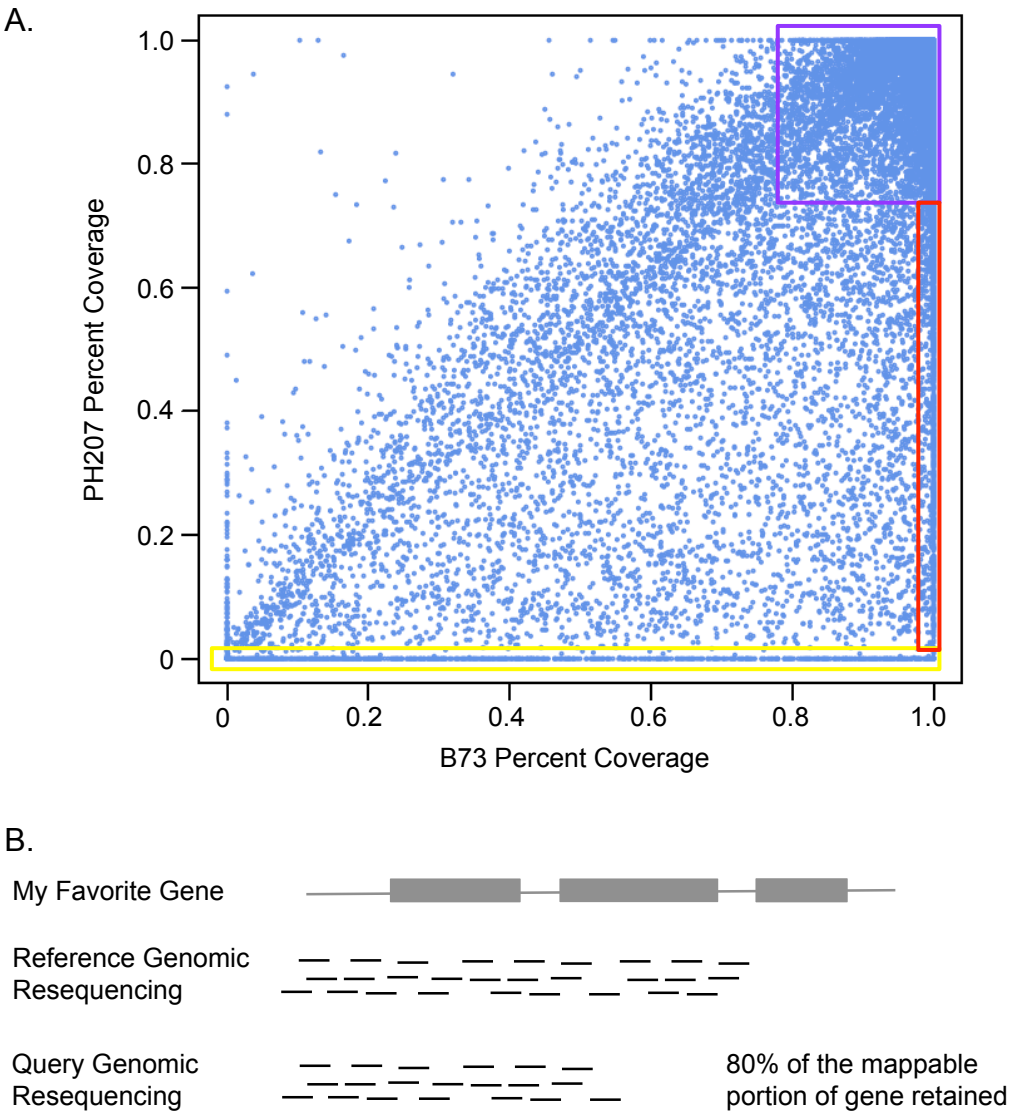




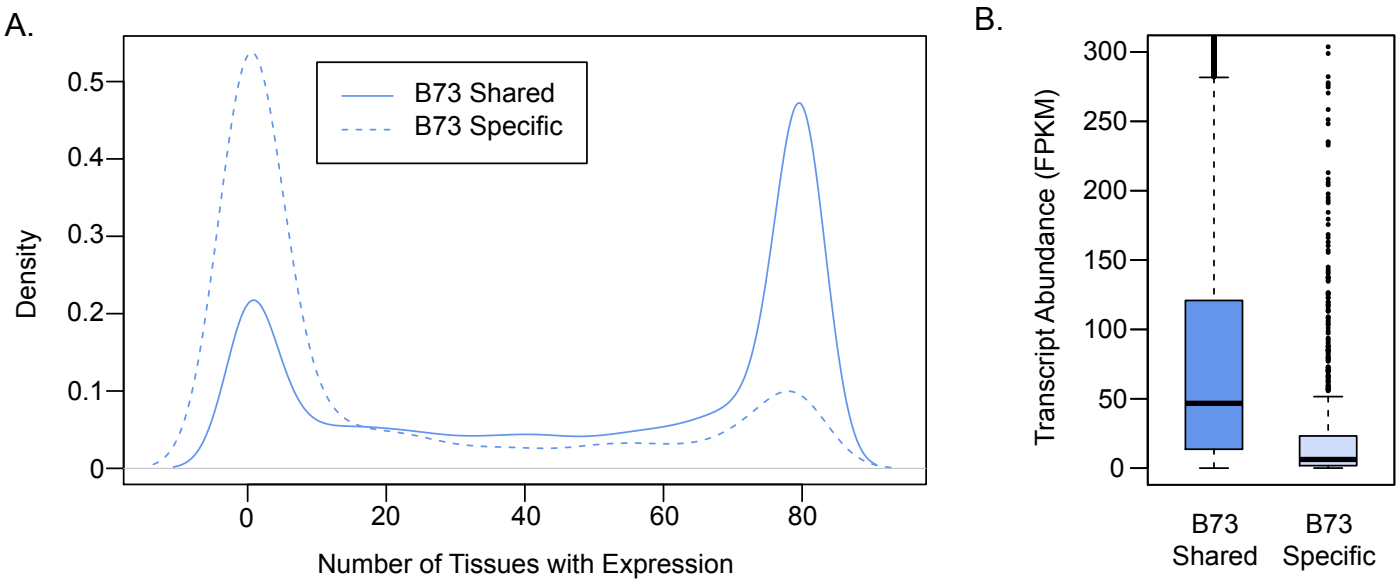
Supplemental Figure 2. PH207 (A) and B73 (B) RNAseq read mapping statistics.



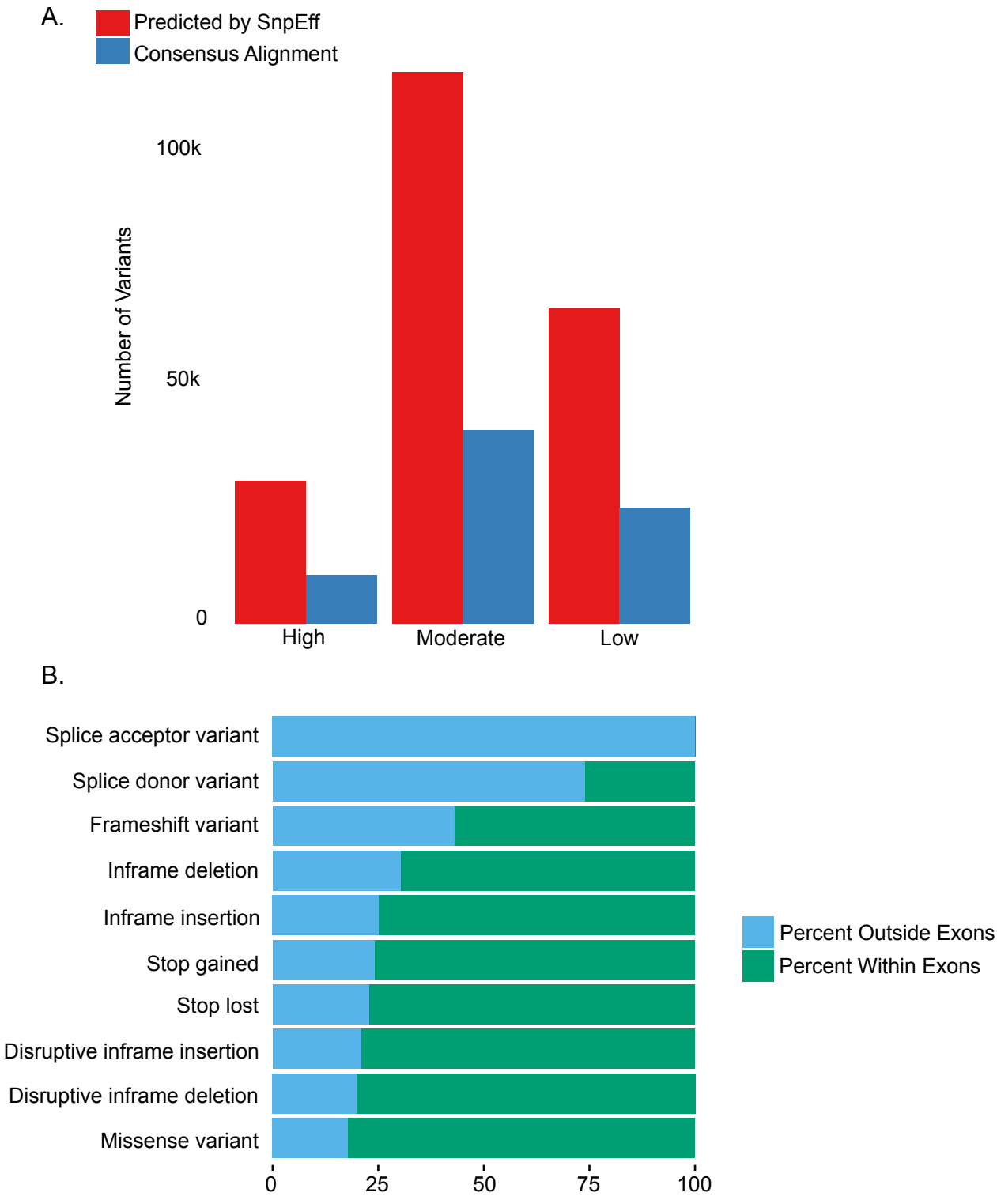
Supplemental Figure 3. Annotation edit distance curve for the PH207 MakerP annotation. Annotation Edit Distance (AED) is a measure of the congruency between an annotation and the evidence that was provided to generate the evidence and is measured as one minus accuracy. Thus, an AED score closer to zero indicates better support from the evidence. This curve show the proportion of transcripts that have support at a given AED score or better such that all transcripts have an AED score of one or better and few transcripts have a perfect AED score of zero.



Supplemental Figure 4. Resequencing based approach to identify partial gene deletions. A) Coverage of B73 genes based on alignment of B73 and PH207 resequencing reads to the B73 reference genome assembly. Yellow box highlights genes that are putatively absent in PH207. Red box highlights genes that are putative partial deletions in PH207. Purple box highlights genes that are mappable based on alignment of B73 resequencing reads and are completely, or nearly completely present in PH207. B) Schematic for identification of partial gene deletions based on mapping of resequencing reads from the cognate and reciprocal genome.



Supplemental Figure 5. Density distribution of expression in B73 and PH207 shared and genotype specific genes. Expression abundances were obtained from the maize B73 atlas. For each tissue used in this analysis the average of three biological replicates was used. Each biological replicate consisted of pooled tissue from three plants. A) Distribution of number of tissues with expression. B) Distribution of average transcript abundance measured in fragments per kilobase of exon models per million fragments mapped (FPKM). The whiskers are plotted at a maximum of 1.5 times the interquartile range away from the end of the box in each direction.



Supplemental Figure 6. Distribution of predicted impact of PH207 variants. A) Distribution of variant impacts based on PH207 resequencing reads aligned to the B73 reference genome and impact of variants predicted based on B73 annotated gene models using SnpEff. Blue bars show the proportion of the total variants for which the exact position in PH207 was determined based on BLAST alignments. B) Distribution of effects in genes showing a one-to-one relationship in the OrthoMCL analysis for variants that fall within or outside of a PH207 annotated exon.

Supplemental Table 1. Reads generated from the PH207 genome assembly.

Library	Estimated insert size	# Input reads	Read length (nt)	Genome coverage	# Filtered reads
300bp paired-end	409 bp	1,831,978,758	100	74.8	1,689,637,208
450bp paired-end	330 bp	123,052,983	250	12.6	55,023,524
800bp paired-end	651 bp	1,100,100,327	100	44.9	1,014,705,636
3kbp mate-pair	3689 bp	314,672,106 393,819,988	100 150	12.8 24.1	98,092,362
8kbp mate-pair	8414 bp	378,800,717 469,536,724	100 150	15.5 28.7	119,883,428
15kbp mate-pair	12379 bp	215,860,425 154,541,907	100 150	8.8 9.5	42,333,362
Illumina/Moleculo TruSeq Synthetic Long Reads	N/A	1,263,629	>1500	2.2	1,263,629

Supplemental Table 2. Assembly statistics at each progressive assembly step.

Progressive assembly steps	Number	N50 (bp)	NG50 (bp)	N10 (bp)	NG10 (bp)	Max size (bp)	Total size (Mb)	Genomic sequence (Mb)
Contigs	21,997,957	5,250	2,280	22,320	19,483	102,924	1,848	1,848
Scaffolds	127,488	654,385	525,655	2,031,813	1,886,273	5,223,973	2,102	1,718
Corrected Scaffolds	132,022	630,443	504,775	1,978,224	1,810,400	5,223,973	2,084	1,893