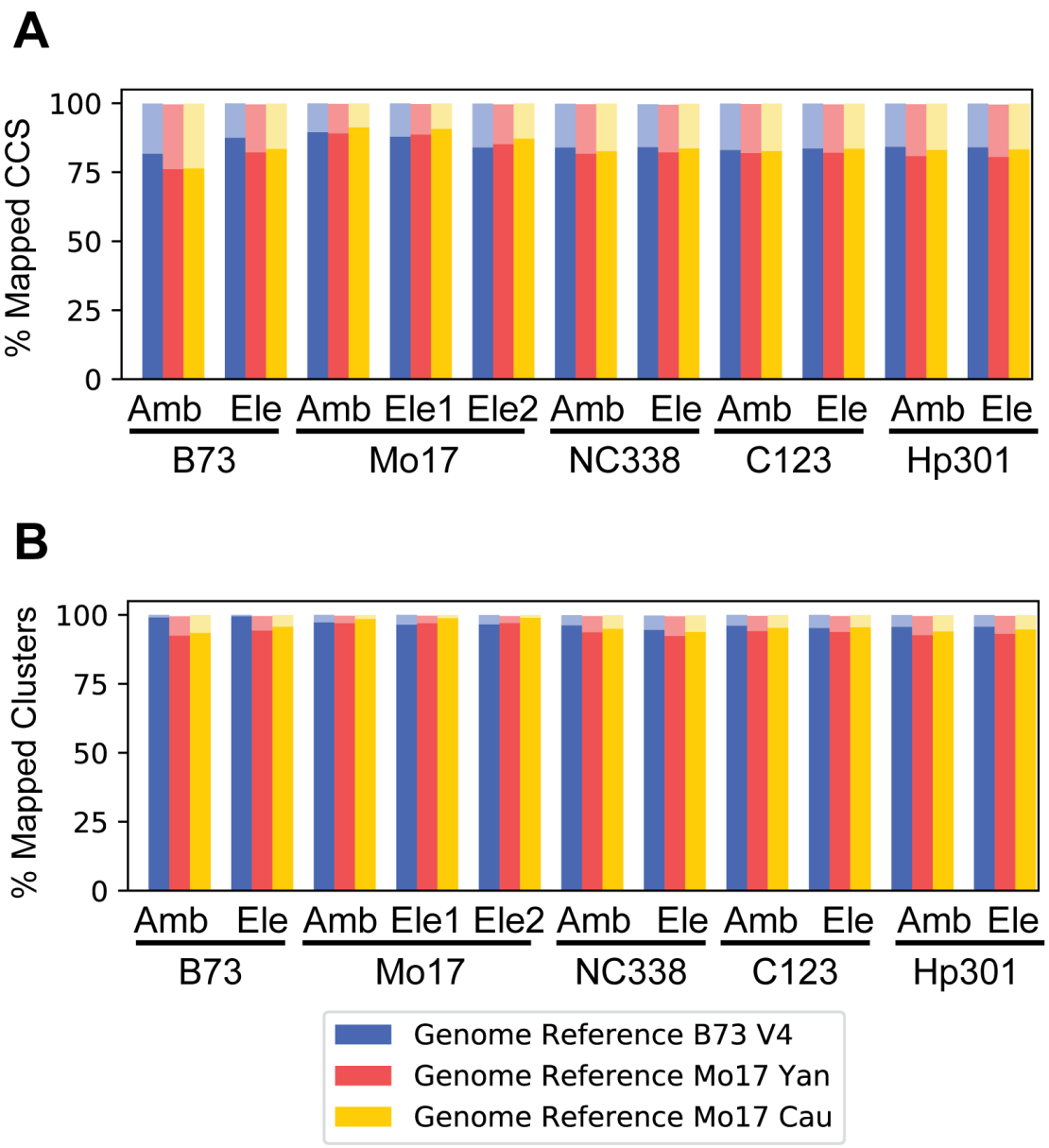
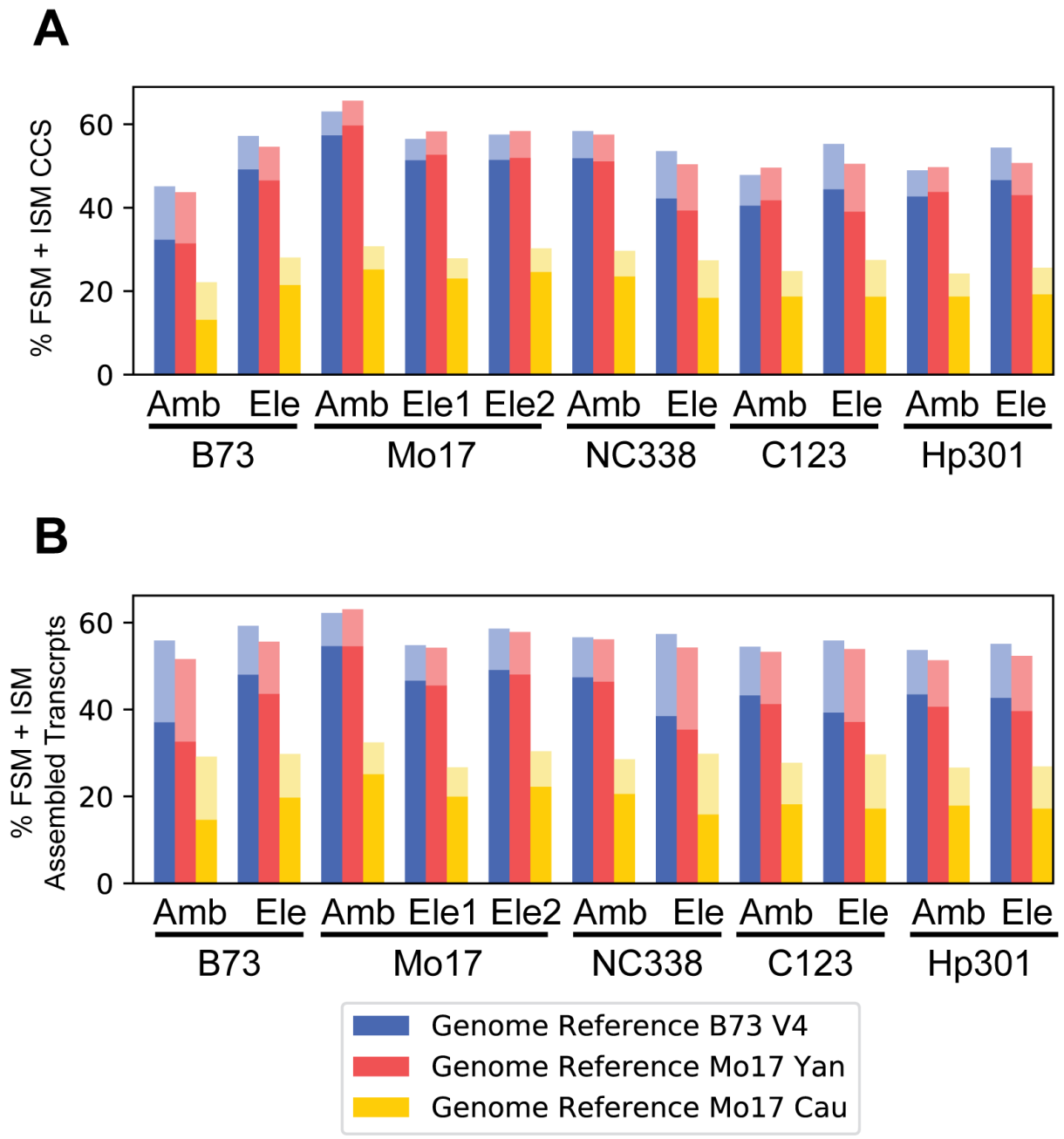
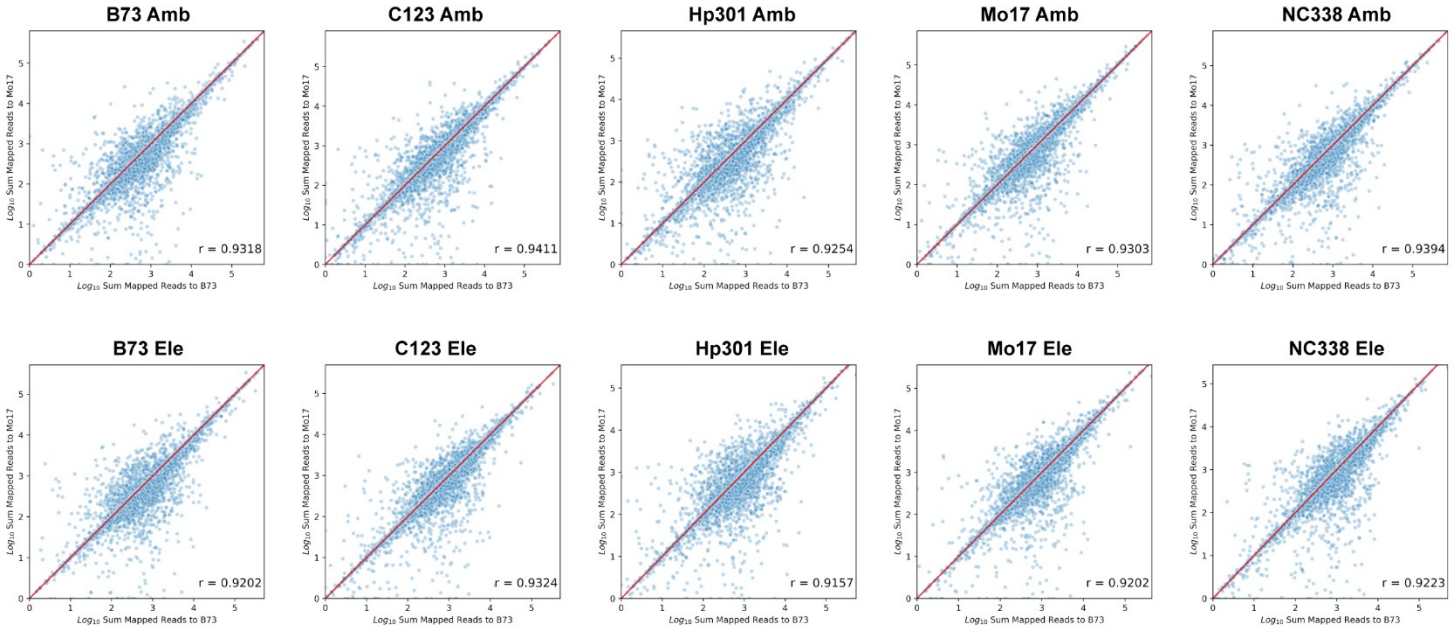


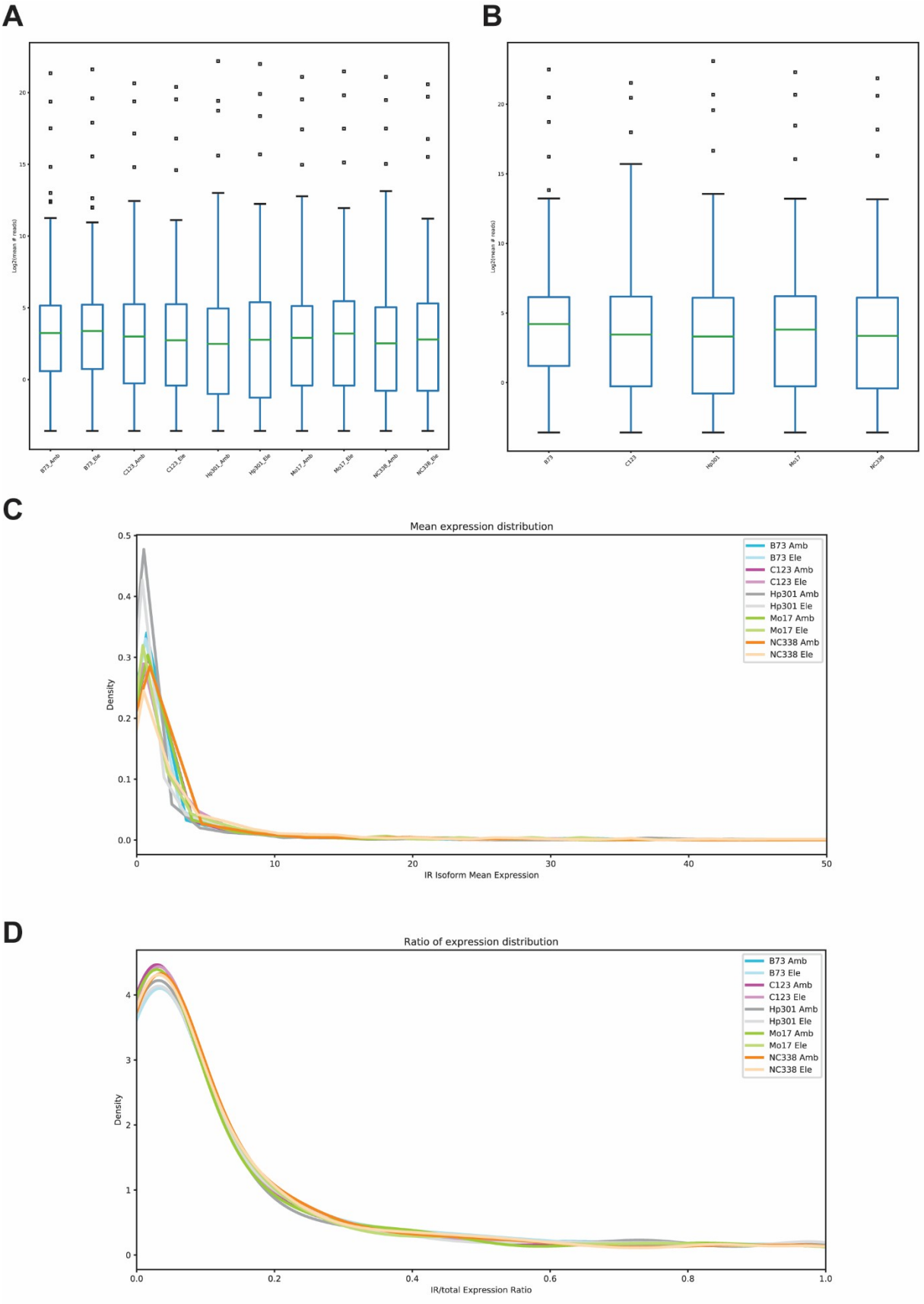
Supplementary Figure 1. A) CCS long reads and B) Clusters map equally well to the B73 v4 or Mo17 reference genomes for all 5 genotypes. The darker color (dark blue, dark red, dark yellow) indicates percent reads (A) or clusters (B) with alignment identity greater than or equal to 95% to the indicated reference sequence and alignment length greater than or equal to 95% of the read or cluster length. Percent reads or clusters with alignment identity less than 95% and/or alignment length less than 95% of the read or cluster length is indicated by a light blue, light red color or light yellow. The percent mapped CCS long reads ranges from 99.4% to 99.9% and percent mapped clusters range from 99.5% to 99.99% (Supplementary Table 1).



Supplementary Figure 2. Proportion of mapped A) CCS reads and B) assembled transcripts identified as full-splice match (FSM; dark red, dark blue, or dark yellow) or incomplete-splice match (ISM; light red, light blue, or light yellow) when mapped and compared to each of the 3 reference genomes. FSM and ISM represent a match to a complete reference junction chain or to a truncated reference junction chain, respectively. The percent of mapped CCS reads with annotated junctions (FSM or ISM) ranges from 44% to 66% with B73 or Mo17 Yan, but 22% to 31% with Mo17 Cau (Supplementary Table 1). Similarly mapped assembled transcripts range from 51% to 63% when mapped to B73 or Mo17 Yan, but from 27% to 30% with Mo17 Cau (Supplementary Table 1).

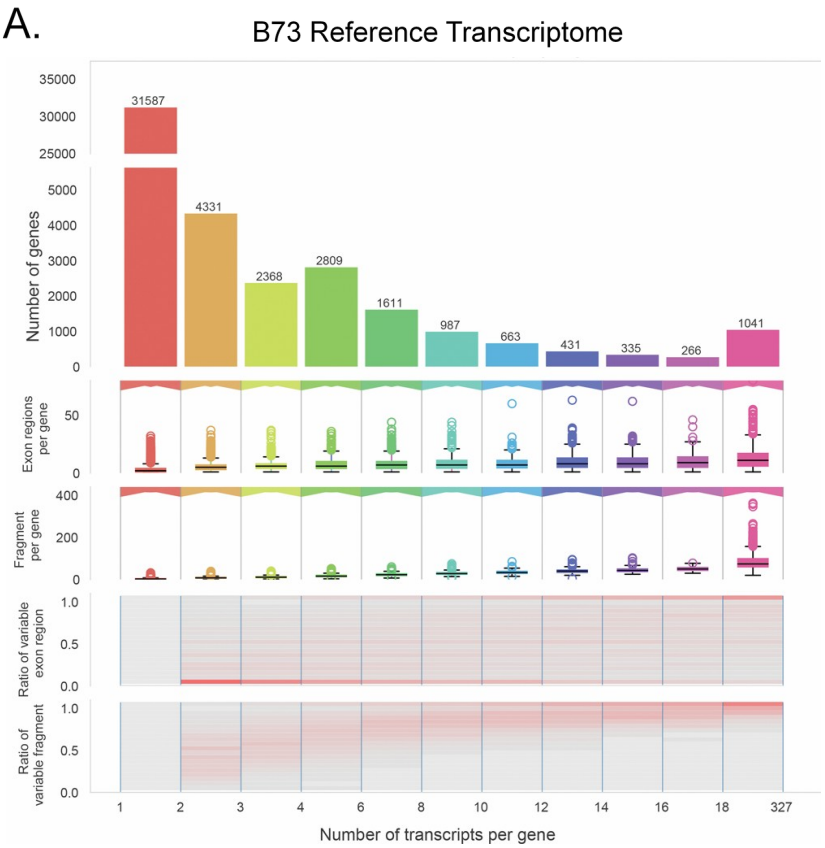


Supplementary Figure 3. Scatter plots of gene-level short-read counts (log_{10} scale) for each sample mapped to B73 v4 (X-axis) vs. Mo17 CAU (Y-axis) reference genomes. Genes included are in the assembled leaf transcriptome and are members of a one-to-one B73-Mo17 syntenic pair (n = 10,614 genes). Pearson correlation coefficients (r) are provided for each sample.

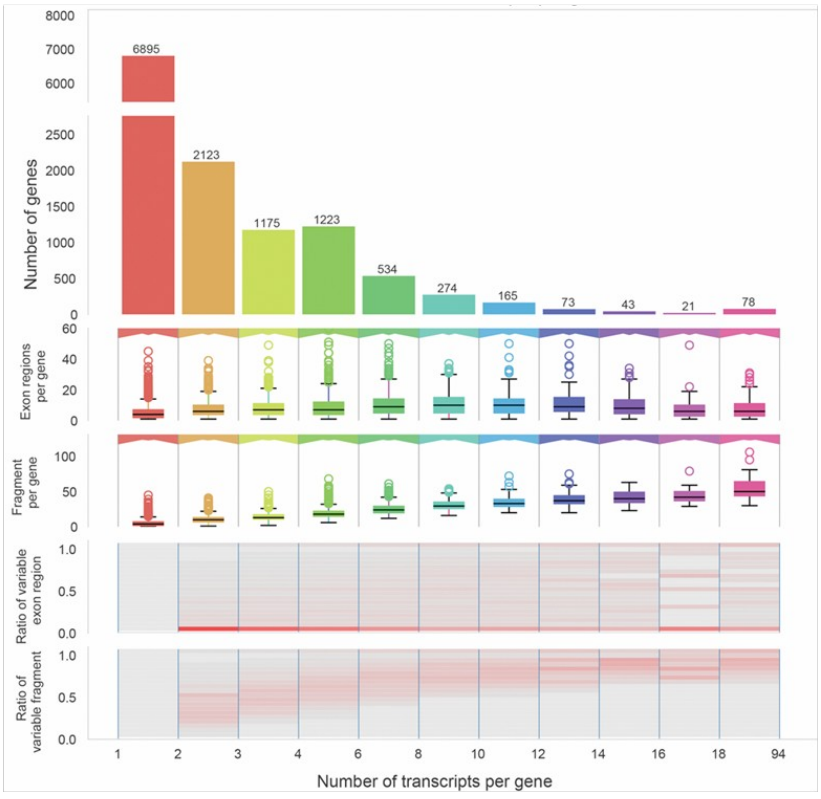

A
B
C
Mean expression distribution
Density
IR Isoform Mean Expression
B73 Amb
B73 Ele
C123 Amb
C123 Ele
Hp301 Amb
Hp301 Ele
Mo17 Amb
Mo17 Ele
NC338 Amb
NC338 Ele
D
Ratio of expression distribution
Density
IR/total Expression Ratio

Supplementary Figure 4. The log2 scaled distribution of the mean number of reads associated with novel, fusion, or antisense loci for (A) each genotype – treatment combination and (B) each genotype summed across treatments. For annotated genes with 2 transcripts, (C) the distribution of mean expression (TPM) of transcripts with intron retention (IR) within each genotype and treatment sample, and (D) the distribution of the proportional contribution of the IR transcript expression to the total gene expression (ratio of IR expression to the total gene expression).

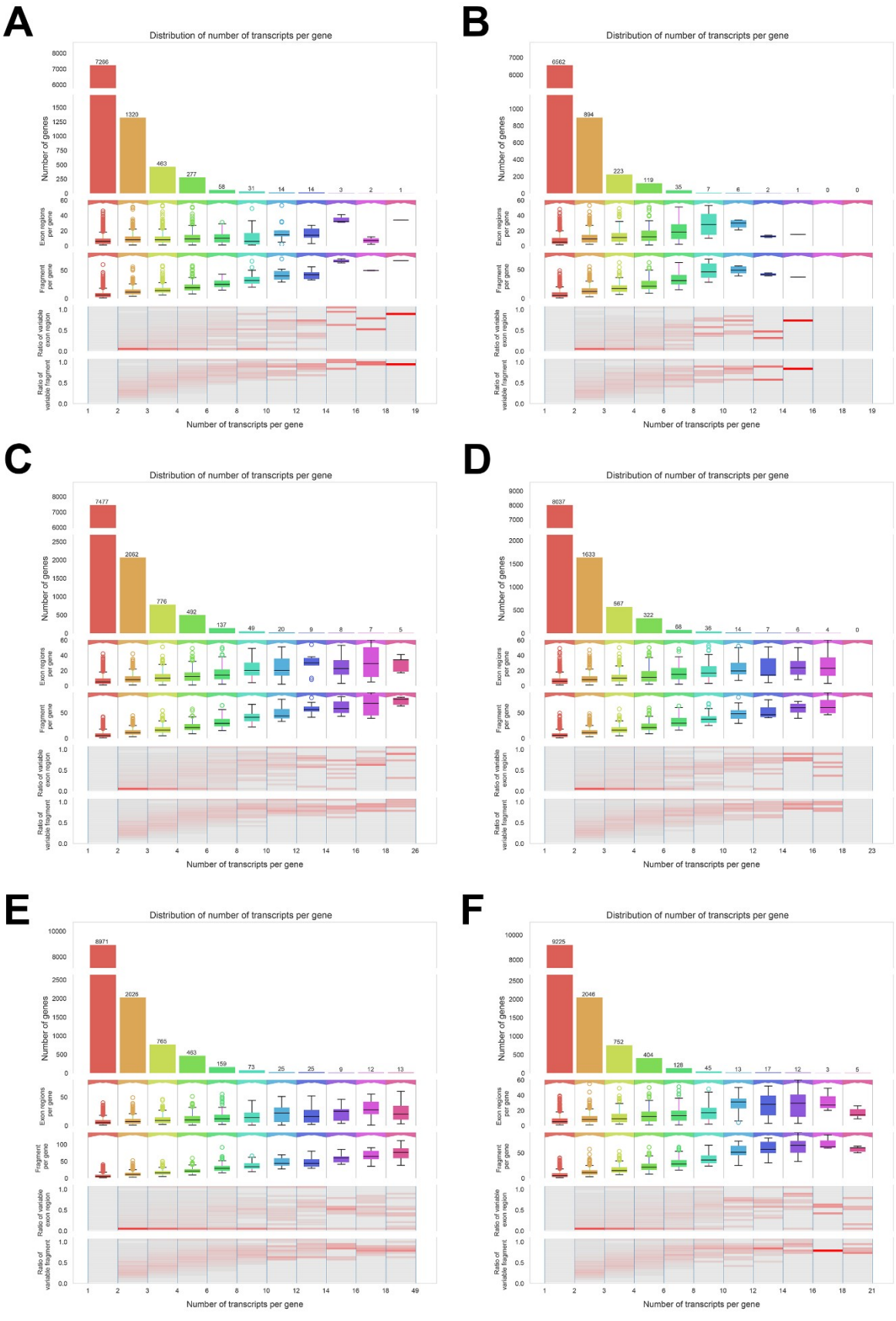
A.


B73 Reference Transcriptome
Number of genes
31587
4331
2368
2809
1611
987
663
431
335
266
1041
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
1
2
3
4
6
8
10
12
14
16
18
327
Number of transcripts per gene

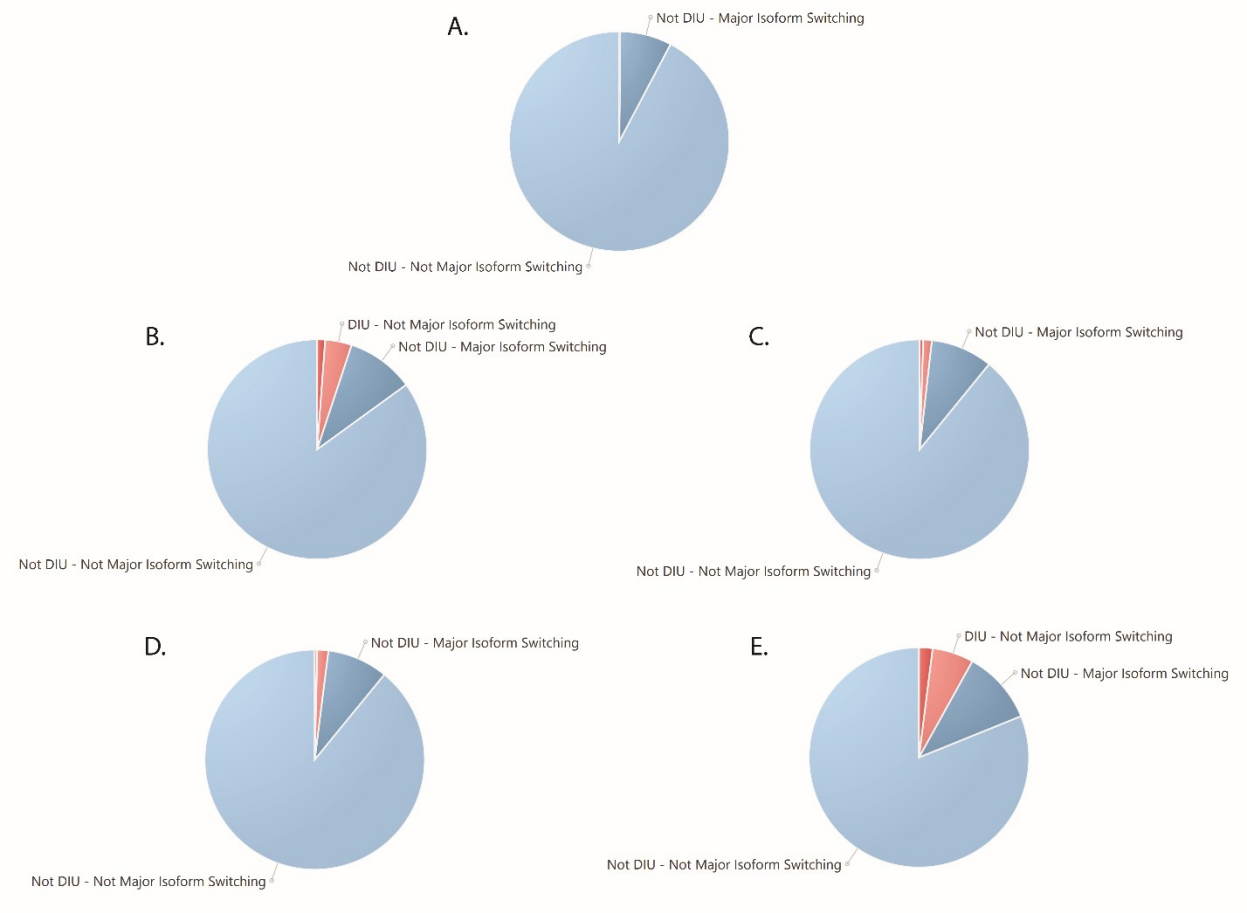
B.


Leaf Transcriptome
Number of genes
6895
2123
1175
1223
534
274
165
73
43
21
78
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
1
2
3
4
6
8
10
12
14
16
18
94
Number of transcripts per gene

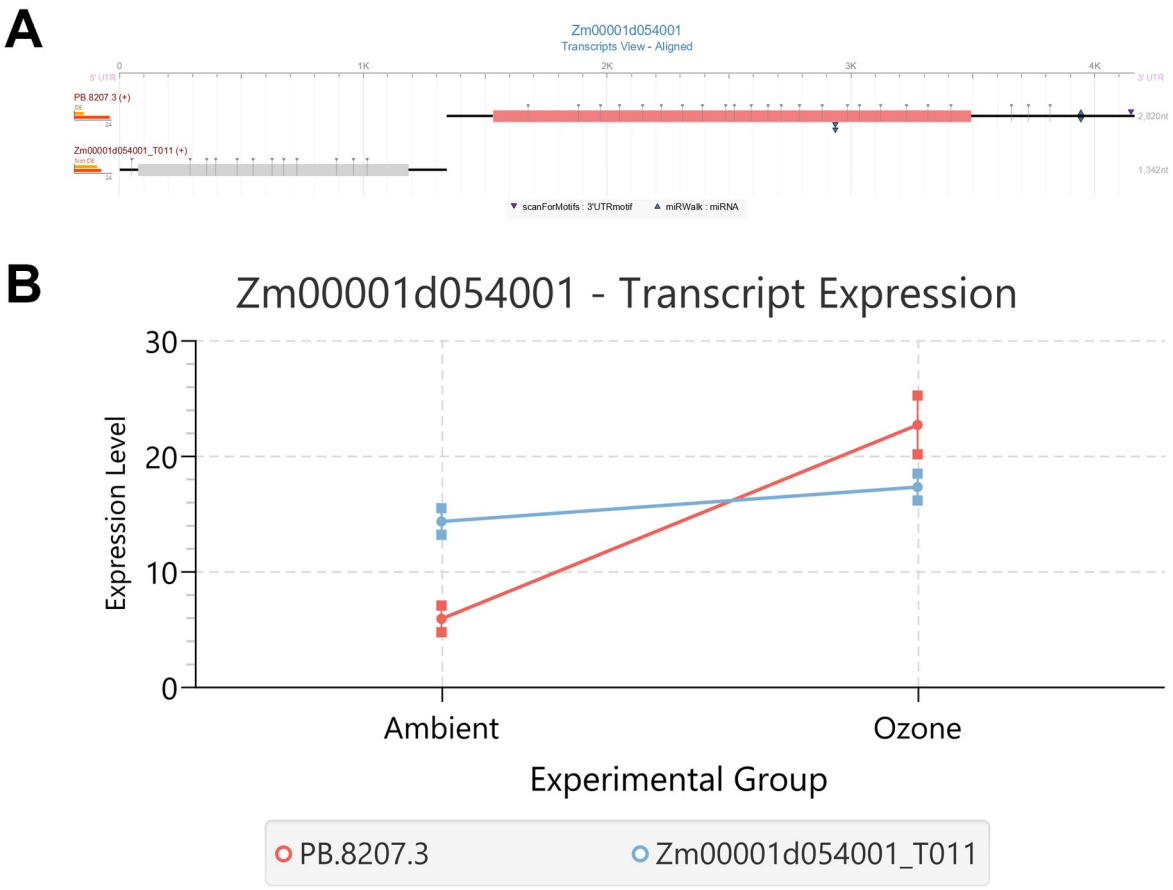
Supplementary Figure 5. Transcriptome complexity. The x-axis is the number of transcripts. Top row, a bar plot for the number of genes with *x* transcripts. Second row, a box and whisker plot for the distribution of the number of exon regions for the genes with *x* transcripts. Third row, a box and whisker plot for the distribution of the number of fragments for the genes with *x* transcripts. Fragments are defined as in (NEWMAN *et al.* 2018). Fourth row, for each transcript the proportion of variable/alternative exons in each gene with *x* transcripts. Fifth row, the proportion of variable fragments in each gene with *x* transcripts that summarize the variation in overlapping exons. A higher proportion of variable exon regions indicates a higher proportion of alternative exons. A higher proportion of variable exon fragments indicates a higher proportion of alternative donor and acceptor sites. Panel A) all transcripts with unique splice junctions for the B73 v4 reference transcriptome (133,204 transcripts, 46,430 genes) and Panel B) Assembled leaf transcriptome from 5 genotypes (B73, C123, Hp301, Mo17, NC338) and 2 treatments (ambient, elevated ozone) (31,388 transcripts, 12,604 genes). For comparison purposes, the B73 leaf transcriptome from Wang et al. (WANG *et al.* 2018) is presented in Supplemental Figure 6A.


A
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
B
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
C
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
D
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
E
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
F
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene

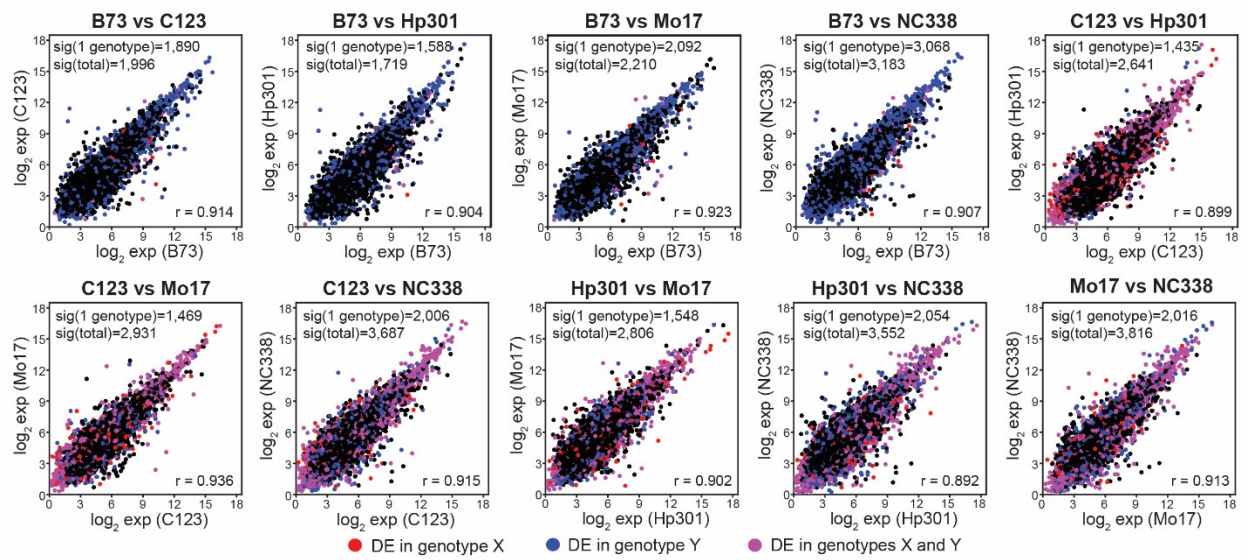
Supplementary Figure 6. Transcriptome complexity for Wang et al. data (WANG *et al.* 2018). See Supplementary Figure 5 for description. Transcriptome complexity is shown for (A) leaf, (B) silk, (C) pericarp, (D) bract, (E) shoot and (F) seedling. In the single tissue leaf transcriptome, there are 13,580 assembled transcripts and 9,451 genes. Approximately 77% of genes in the assembled leaf transcriptome have a single transcript.



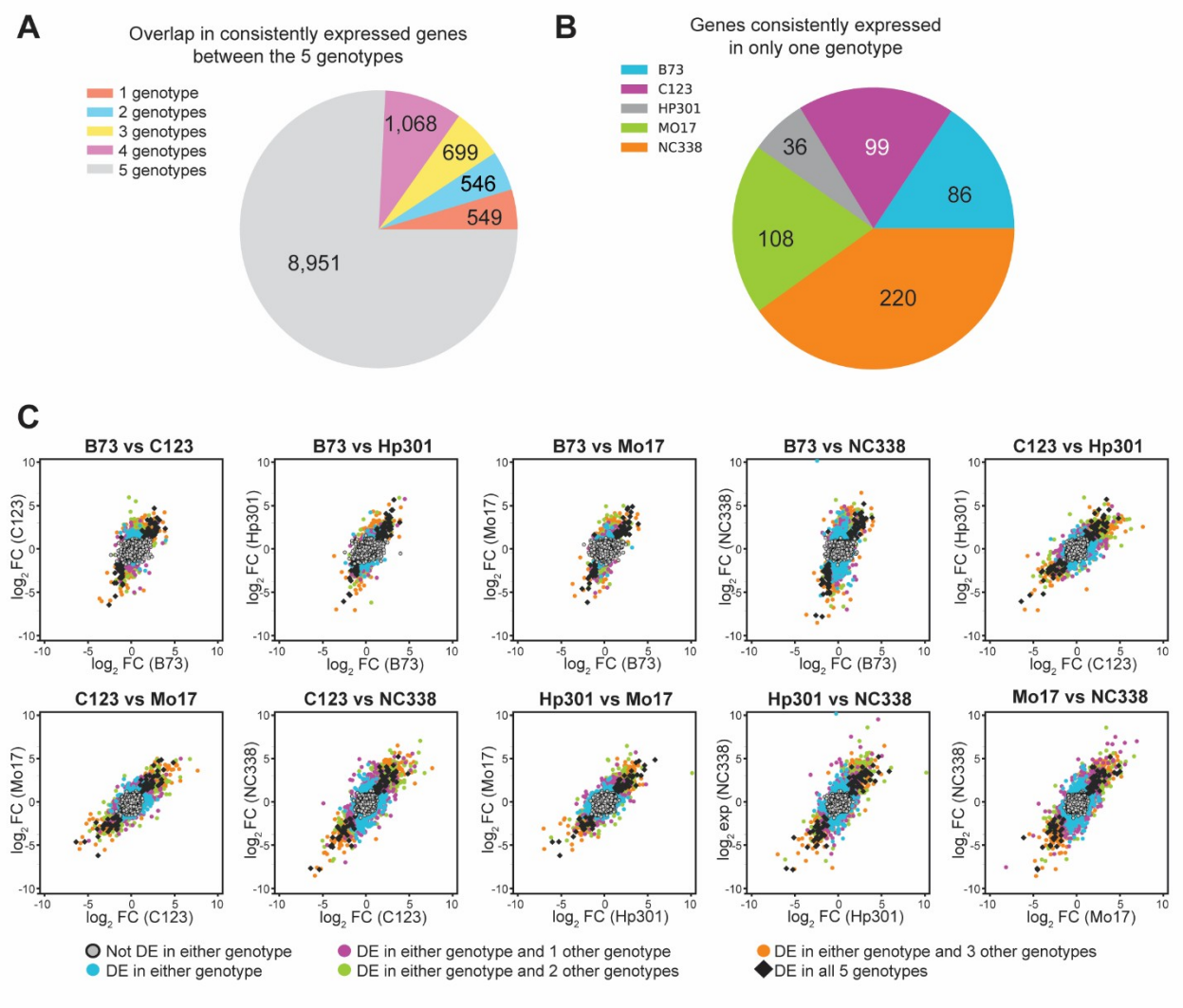
Supplementary Figure 7. Summary of tappAS (DE LA FUENTE *et al.* 2019) differential isoform usage (DIU) results for the 5 genotypes in this study. The number of gene with DIU and major isoform switching (n, dark red) for (A) B73 (n=1, 2874 genes tested), (B) C123 (n=34, 2941 genes tested), (C) Hp301 (n=14, 2696 genes tested), (D) Mo17 (n=11, 2961 genes tested), and (E) NC338 (n=59, 2966 genes tested). Few of the genes with DIU and major isoform switching are shared across multiple genotypes. Only 15 genes are shared by 2 genotypes and 1 gene (Zm00001d054001) is shared by Hp301, Mo17, and NC338.



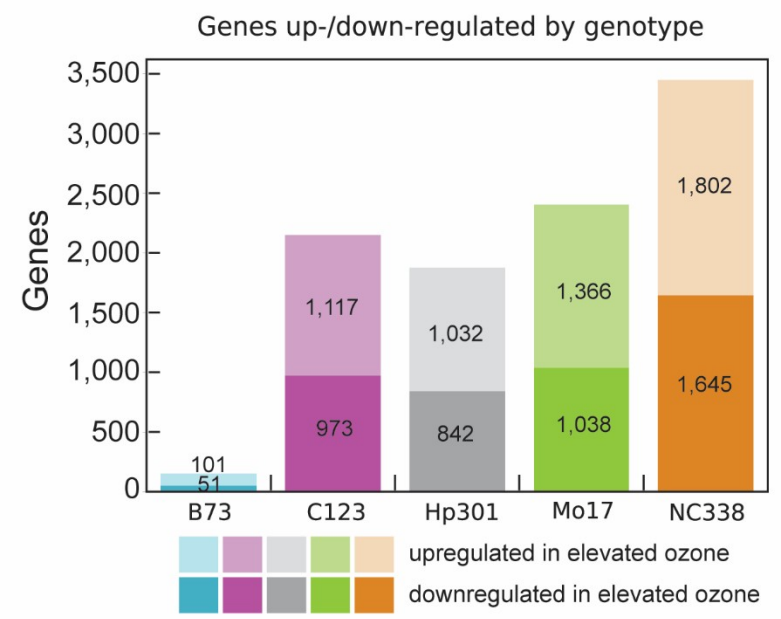
Supplementary Figure 8. An example gene (Zm00001d054001) identified with DIU and major isoform switching by tappAS (DE LA FUENTE *et al.* 2019) in Hp301, Mo17 and NC338 genotypes. Transcript models are shown in (A) and transcript expression in ambient and ozone conditions for genotype NC338 in (B).



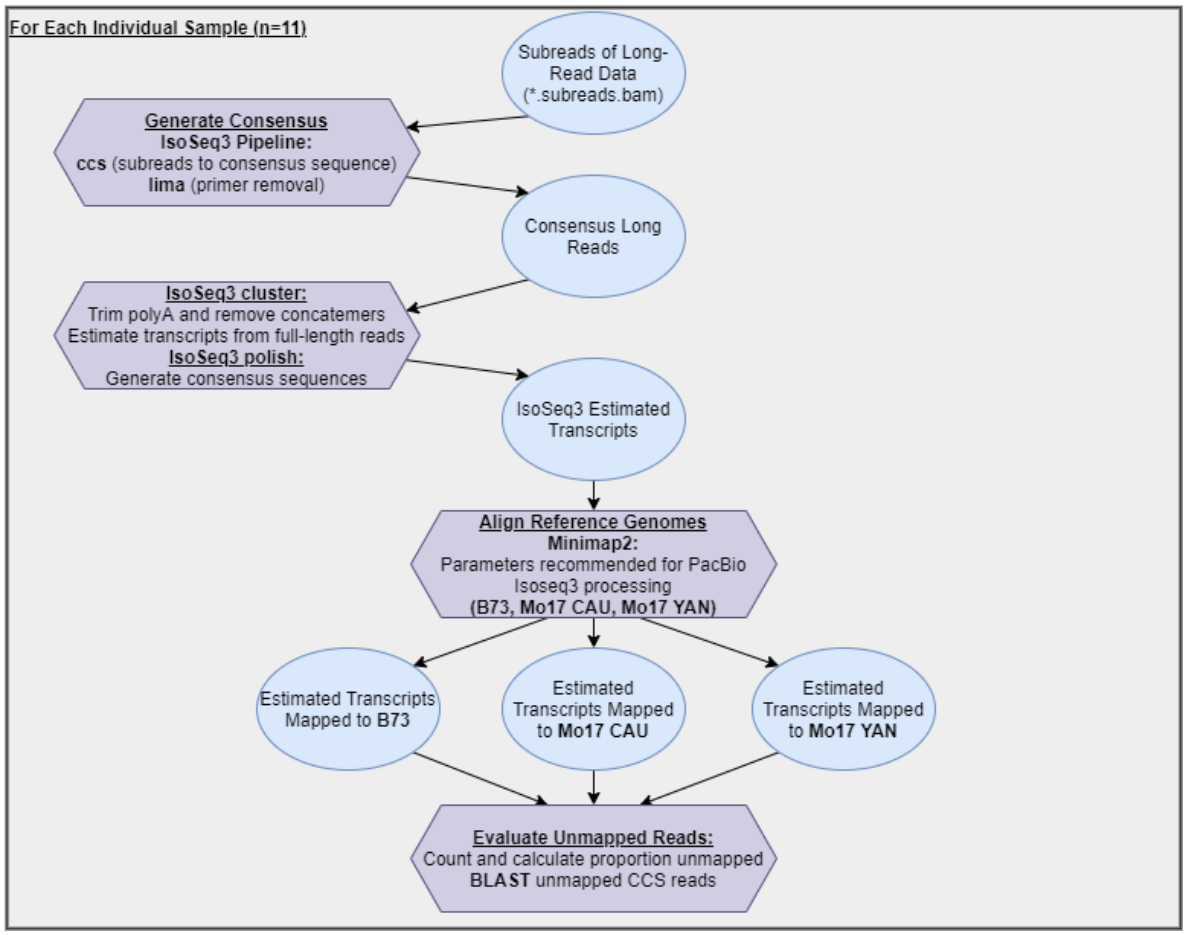
Supplementary Figure 9. Differentially expressed genes**.** A scatterplot for $\log_2$(TPM) for each pair of genotypes in the ambient condition. Genes in each plot are colored based on differential gene expression between ambient and ozone: (1) red genes are differentially expressed in the genotype on the x-axis; (2) genes colored blue are differentially expressed in the genotype on the y-axis; (3) genes differentially expressed in both genotypes are colored purple; (4) genes not differentially expressed in either genotype are colored black. The relatively small transcriptional response of B73 can be seen in the first 4 panels (B73 on x-axis): the proportion of genes colored blue compared to genes colored red or purple is higher for plots with B73 that in plots without B73. Other genotype combinations show a larger degree of a shared response (e.g., a greater number of genes colored purple). We note that there is an even distribution of differentially expressed genes across the range of expression in the ambient condition.



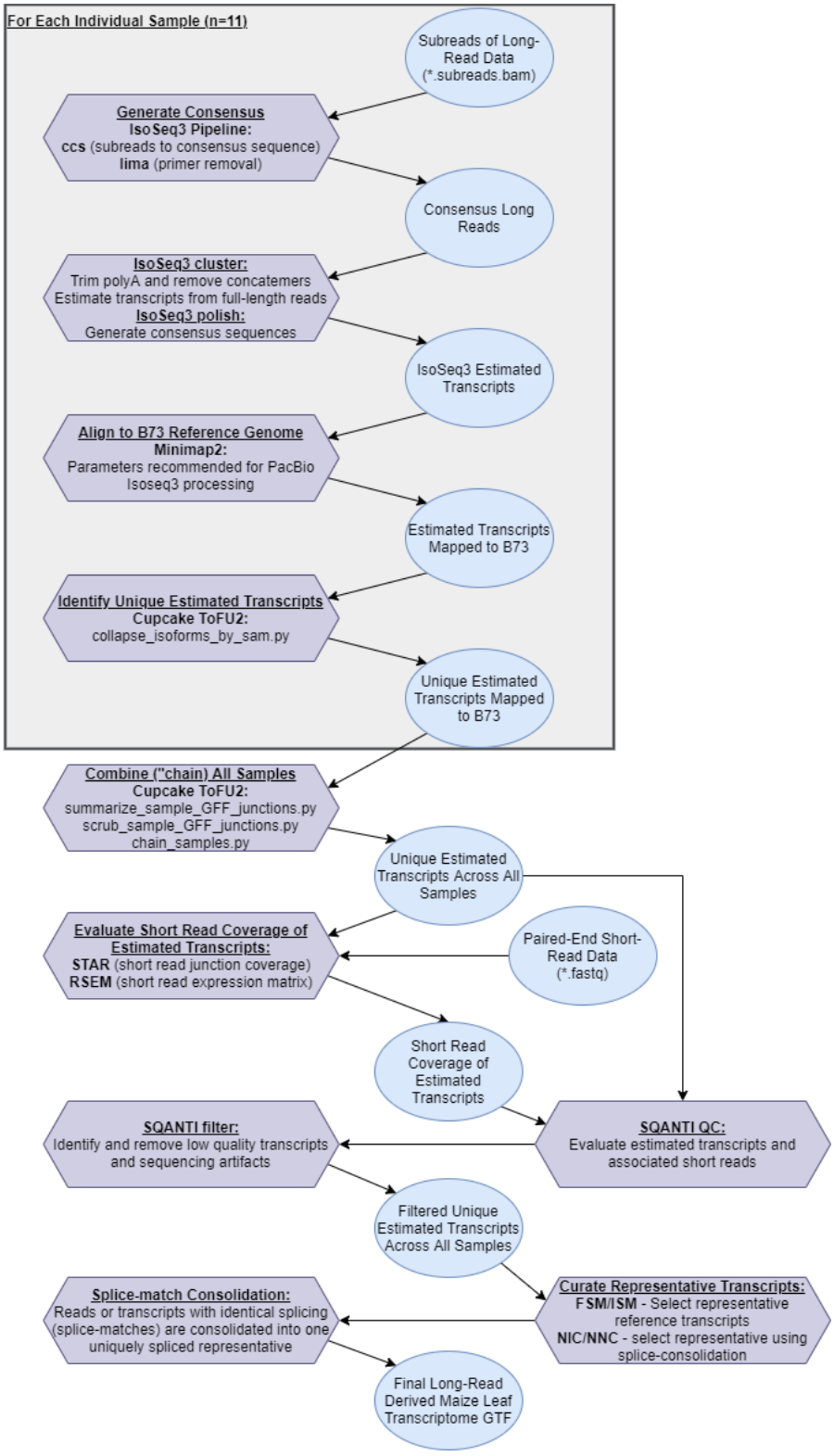
Supplementary Figure 10. Maize leaf transcriptome expression. (A) Overlap of genes analyzable across genotypes. (B) Number of genes analyzable in only a single genotype. (C) A scatterplot of the $\log_2$(Fold Change) for each indicated pair of genotypes. Genes in each scatterplot are colored based on the number of genotypes for which that gene is differentially expressed between ambient and ozone: (1) genes colored gray with black outline are not differentially expressed in the genotypes indicated on the axes; (2) genes colored pink are differentially expressed in either genotype indicated on the axes; (3) genes differentially expressed in either genotype indicated on the axes plus 1 additional genotype are colored blue; (4) genes differentially expressed in either genotype indicated on the axes plus 2 additional genotypes are colored yellow; (5) genes differentially expressed in either genotype indicated on the axes plus 3 additional genotypes are colored purple; (6) genes differentially expressed in all 5 genotypes are shown as gray diamonds.



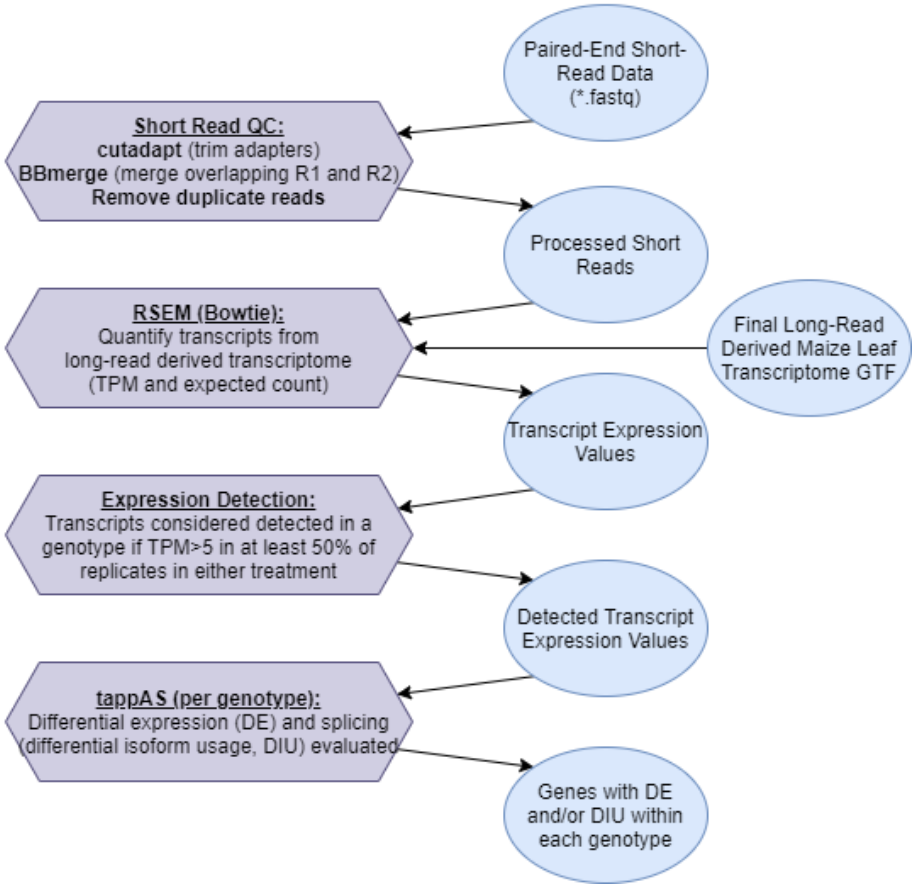
Supplementary Figure 11. For all genotypes, differential expression was roughly balanced between over- and under-expression with B73 showing a relatively minor response and NC338 showing the most extreme response to ozone stress.



Supplementary Figure 12. Expression presence/absence evaluation using long read data.



Supplementary Figure 13. Building an unbiased leaf transcriptome from 10 independent long-read libraries for 5 genotypes (B73, Mo17, C123, HP301, NC338) in ozone and ambient conditions. The Mo17 ambient library was sequenced twice, once for calibration and then together with the other 9 libraries.



Supplementary Figure 14. Quantifying transcripts from the unbiased leaf transcriptome using short reads.

Supplementary Table 1. (A) Quality control information for PacBio read processing, and (B) transcriptome mapping and performance metrics. High-quality transcripts from each sequenced PacBio library (genotype × ozone treatment) were mapped to 3 reference genomes: (i) B73 RefGen_v4 (JIAO *et al.* 2017), (ii) Mo17 Yan (YANG *et al.* 2017) and (iii) Mo17 Cau (SUN *et al.* 2018). For each row parameter x sample combination, PacBio assembled transcript numbers generated via mapping (minimap2, v2.12, (LI 2018)) the PacBio polished transcripts to the B73, Mo17 Yan and Mo17 Cau references, respectively, are separated by forward slashes. Cupcake ToFU2 (version 20180629) supporting scripts were used to generate unique isoforms. The number if PacBio transcripts 'ignored' by ToFU2 due to low coverage, low identify or non-mapping are indicated for each reference. Results from the QC function of SQANTI (TARDAGUILA *et al.* 2018) are noted for each library and for the combined chained samples: FSM, full-splice match; ISM, incomplete-splice match; NIC, novel-in-catalog; NNC, novel-not-in-catalog. Chained transcriptome, SQANTI filtering, and curation are also included.

Supplementary Table 2. RNA-seq quality control information: (A) read counts from sequencer and after processing, and (B) summary of unmapped paired-end (PE) and single-end (SE) short reads when aligned to 3 reference genomes: (i) B73 RefGen_v4 (JIAO *et al.* 2017), (ii) Mo17 Yan (YANG *et al.* 2017) and (iii) Mo17 Cau (SUN *et al.* 2018).

Supplementary Table 3. Novel and antisense loci with at least one read in any replicate in B73 only, including the 3 novel loci with an average of at least 5 reads across replicates in both treatments (novelGene_372, novelGene_344, and novelGene_769). Read counts for each genotype by treatment combination is given.

Supplementary Table 4. The 12 genes annotated in the NAM population as private to B73 (HUFFORD *et al.* 2021) and expressed only in the B73 leaf samples.

Supplementary Table 5. Genes with evidence of differential isoform usage and major isoform switching in at least one genotype. The DIU_qval_*genotype* columns contain the q-value for the indicated DIU test in tappAS. The flag_DIU_*genotype* variables = 1 if the tappAS DIU test was significant (q-value < 0.05) for the indicated genotype. The flag_DIU_majorisoformSwitch_*genotype* variables = 1 for the indicated genotype if the DIU test was significant and the test showed a major isoform switch. The total_usage_change_*genotype* columns contain the total change in expression across transcripts for the indicated genotype. The mean_TPM_*condition_genotype* columns contain the mean TPM values for the indicated genotype – condition. The flag_analyze_DIU_majorIsoformSwitch_*genotype* columns = 1 if the gene is significant in DIU test, contains a major switch and is analyzable in the given genotype. Sum_analyze_DIU_majorisoformswitch is the sum of the flag_analyze_DIU_majorIsoformSwitch_*genotype* flags.

Supplementary Table 6. GO enrichment results of RSEM quantified DE analysis. The flag_analyze_DE_all5 variable = 1 for genes differentially expressed in all 5 genotypes. The flag_analyze_DE_*genotype* variables = 1 for genes differentially expressed in the indicated genotype. The flag_analyze_DE_all_noB73 variable = 1 for genes differentially expressed in C123, Hp301, Mo17, and NC338, but not B73. The flag_B73Zero_restUp variable = 1 for genes with a fold change between ambient and ozone of 0 for B73 and > 0 for all other genotypes in genes differentially expressed in at least one genotype.

Supplementary Table 7. For each genotype and treatment, one-to-one B73-Mo17 Cau syntenic gene pairs in the 12,604 genes in the assembled leaf transcriptome (n=10,614 genes) are evaluated for short read expression. The gene counts consist of the number of genes with mapped reads when mapped to the B73 v4 reference genome, the number of genes with mapped reads when mapped to the Mo17 Cau reference genome, and the number of B73-Mo17 gene pairs with mapped reads when mapped to either reference. Also included are the number of genes where the difference in mapped read counts when mapped to the B73 v4 reference genome vs. Mo17 Cau is less than or equal to 5% and the number of genes with a difference greater than 5% with either the B73 v4 genome or the Mo17 Cau genome resulting in a greater number of mapped reads.

Supplementary Table 8. GO enrichment results of genome alignment DE analysis. The flag_ DE_all variable = 1 for genes differentially expressed in all 5 genotypes. The flag_ DE_*genotype* variables = 1 for genes differentially expressed in the indicated genotype. The flag_ DE_all_noB73 variable = 1 for genes differentially expressed in C123, Hp301, Mo17, and NC338, but not B73. The flag_B73Zero_restUp variable = 1 for genes with a fold change between ambient and ozone of 0 for B73 and > 0 for all other genotypes in genes differentially expressed in at least one genotype.

Database	Link
UniProt (Swiss-Prot)	https://www.uniprot.org/uniprot/?query=reviewed:yes%20AND%20organism:%22Zea%20mays%20[4577]%22#
UniProt (TrEMBL)	https://www.uniprot.org/uniprot/?query=reviewed:no%20AND%20organism:%22Zea%20mays%20[4577]%22#
RefSeq	ftp://ftp.ncbi.nlm.nih.gov/genomes/Zea_mays/protein/protein.fa.gz
Ensembl	ftp://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/pep/Zea_mays.B73_RefGen_v4.pep.all.fa.gz

Supplementary Table 9. List of public databases used for creating functionally annotated GFF-like file

Supplementary File 1. Functional annotation (in GFF-like format) mapped to the B73 genome of the curated maize leaf transcriptome generated from the data described in this study.

Supplementary File 2. GTF of the curated maize leaf transcriptome mapped to the B73 genome.

Supplementary File 3. The complete expression matrix of the curated maize leaf transcriptome. Expression values (TPM) for each transcript are given in the indicated *genotype_plant_chamber_condition* columns. The flag_*genotype_condition* variables are 0/1 binary indicator flags for analyzability, where the flag is equal to 1 if TPM > 5 in > 50% of the replicates. The mean TPM for each genotype–condition is given in the mean_*gentoype_condition* columns.

Supplementary File 4. Gene-level differential expression results for the assembled maize leaf transcriptome. The flag_*genotype_condition* variables are 0/1 binary indicator flags for analyzability where the flag is equal to 1 if at least 1 transcript for the indicated gene was analyzable (TPM > 5 in at least 50% of replicates). The sum_flag variable is the sum of the flag_*genotype_condition* variables for a given gene. The DE_pval_*genotype* columns contain the p-value for the test of DE. The flag_DE_*genotype* variable = 1 if the test of DE was significant (p-value < 0.05). The Log2FC_*genotype* columns contain the TPM fold change between ambient and ozone conditions. The mean TPM for each genotype–condition is given in the mean_TPM_*genotype_condition* columns. The flag_analyze_DE_*genotype* variables = 1 for genes differentially expressed and analyzable in the indicated genotype. The flag_analyze_DE_all5 variable = 1 for genes differentially expressed and detected in all 5 genotypes. GO IDs and GO terms are included for each gene.

Supplementary File 5. Gene-level flags and mean expression values for the assembled maize leaf transcriptome. The orthogroup_hoopes variable lists ortholog groups identified in Hoopes, et al. 2019 (HOOPES *et al.* 2019), and flag_zea_mays_paralog_hoopes is a 0/1 binary indicator flag for if the gene has more than one *Z. mays* V4 reference gene associated. The freq_pav_hoopes variable is the number of genotypes observed as PAV in previous studies (BROHAMMER *et al.* 2018; HOOPES *et al.* 2019) for the indicated gene, and flag_pav is a 0/1 binary indicator for if freq_pav_hoopes > 0. The flag_detect_shrtRd_genomeAln_*genotype* and flag_detect_ccs_*genotype* variables are 0/1 binary indicator flags for detection if at least one short read or CCS long read was present in genome alignments quantified with HTSeq (ANDERS *et al.* 2015). The flag_detect_shrtRd_rsem_*genotype* is a 0/1 binary flag for detection if short

read TPM > 0 for RSEM (LI and DEWEY 2011) quantified transcript alignments. The flag_detect_*genotype* variables are 1 if the sum of the long and short read detection flags is greater than 0. The variable num_detect_genotype is the sum of the flag_detect_*genotype* flags, and flag_detect_num_geno_gt0_lt5 is a 0/1 binary indicator for if the number of detected genotypes is greater than 0 and less than 5. The flag_analyze_*genotype* is a 0/1 binary flag for analyzability if at least one transcript of the indicated gene has short read TPM > 5 in at least 50% of replicates for either condition, and sum_analyze is the sum of flag_analyze_*genotype* flags. The mean_*genotype_condition* variables are the sum of transcript-level mean TPM values across replicates. The variable flag_no_assoc_v5_gene = 1 if no B74 v5 gene identifier could be associated with the B73 v4 gene using maizeGDB (https://maizegdb.org/) (WOODHOUSE *et al.* 2021). The variable flag_no_assoc_pangene = 1 if a NAM pan-gene (HUFFORD *et al.* 2021) could not be associated. Each gene is flagged for if there is an associated core pan-gene (flag_core = 1), near-core pan-gene (flag_near_core = 1), dispensable pan-gene (flag_dispensable = 1), or private pan-gene (flag_private = 1). Note that due to the associations of v4 to v5 genes and the possibility of multiple genes in a pan-gene, these flags are not mutually exclusive. The highest level of pan-gene classification (core > near-core > dispensable > private) is provided in the pangene_class variable. Additionally, the variable flag_any_tandem_dup = 1 when the associated pan-genes have at least one NAM founder with tandem duplication, and flag_temperate_tandem_dup = 1 where the associated pan-genes have at least one of the temperate NAM founders.

References

Anders, S., P. T. Pyl and W. Huber, 2015 HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31**:** 166-169.

Brohammer, A. B., T. J. Y. Kono, N. M. Springer, S. E. McGaugh and C. N. Hirsch, 2018 The limited role of differential fractionation in genome content variation and function in maize (Zea mays L.) inbred lines. Plant Journal 93**:** 131-141.

de la Fuente, L., Á. Arzalluz-Luque, M. Tardáguila, H. del Risco, C. Martí *et al.*, 2019 tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. bioRxiv**:** 690743.

Hoopes, G. M., J. P. Hamilton, J. C. Wood, E. Esteban, A. Pasha *et al.*, 2019 An updated gene atlas for maize reveals organ-specific and stress-induced genes. Plant Journal 97**:** 1154-1167.

Hufford, M. B., A. S. Seetharam, M. R. Woodhouse, K. M. Chougule, S. J. Ou *et al.*, 2021 De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 373**:** 655-+.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. Stitzer *et al.*, 2017 Improved maize reference genome with single-molecule technologies. Nature 546**:** 524-+.

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12**:** 323.

Li, H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34**:** 3094-3100.

Newman, J., P. Concannon, M. Tardaguila, A. Conesa and L. McIntyre, 2018 Event Analysis: Using Transcript Events To Improve Estimates of Abundance in RNA-seq Data. G3-Genes Genomes Genetics 8**:** 2923-2940.

Sun, S., Y. Zhou, J. Chen, J. Shi, H. Zhao *et al.*, 2018 Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nature Genetics 50**:** 1289-+.

Tardaguila, M., L. de la Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios *et al.*, 2018 SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res.

Wang, B., M. Regulski, E. Tseng, A. Olson, S. Goodwin *et al.*, 2018 A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. Genome Res 28**:** 921-932.

Woodhouse, M. R., E. K. Cannon, J. L. Portwood, L. C. Harper, J. M. Gardiner *et al.*, 2021 A pan-genomic approach to genome databases using maize as a model system. Bmc Plant Biology 21.

Yang, N., X. Xu, R. Wang, W. Peng, L. Cai *et al.*, 2017 Contributions of Zea mays subspecies mexicana haplotypes to modern maize. Nature Communications 8.