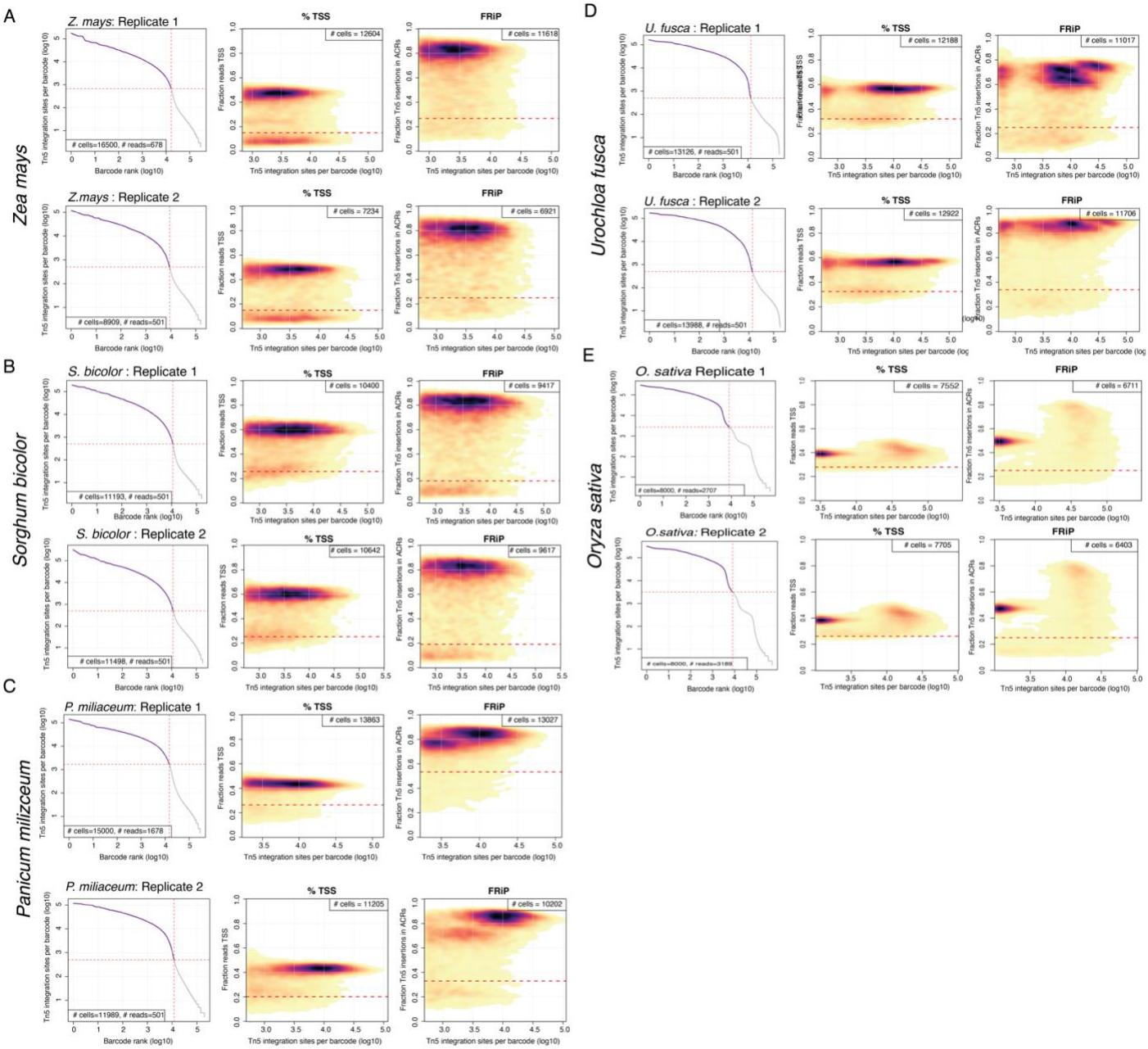
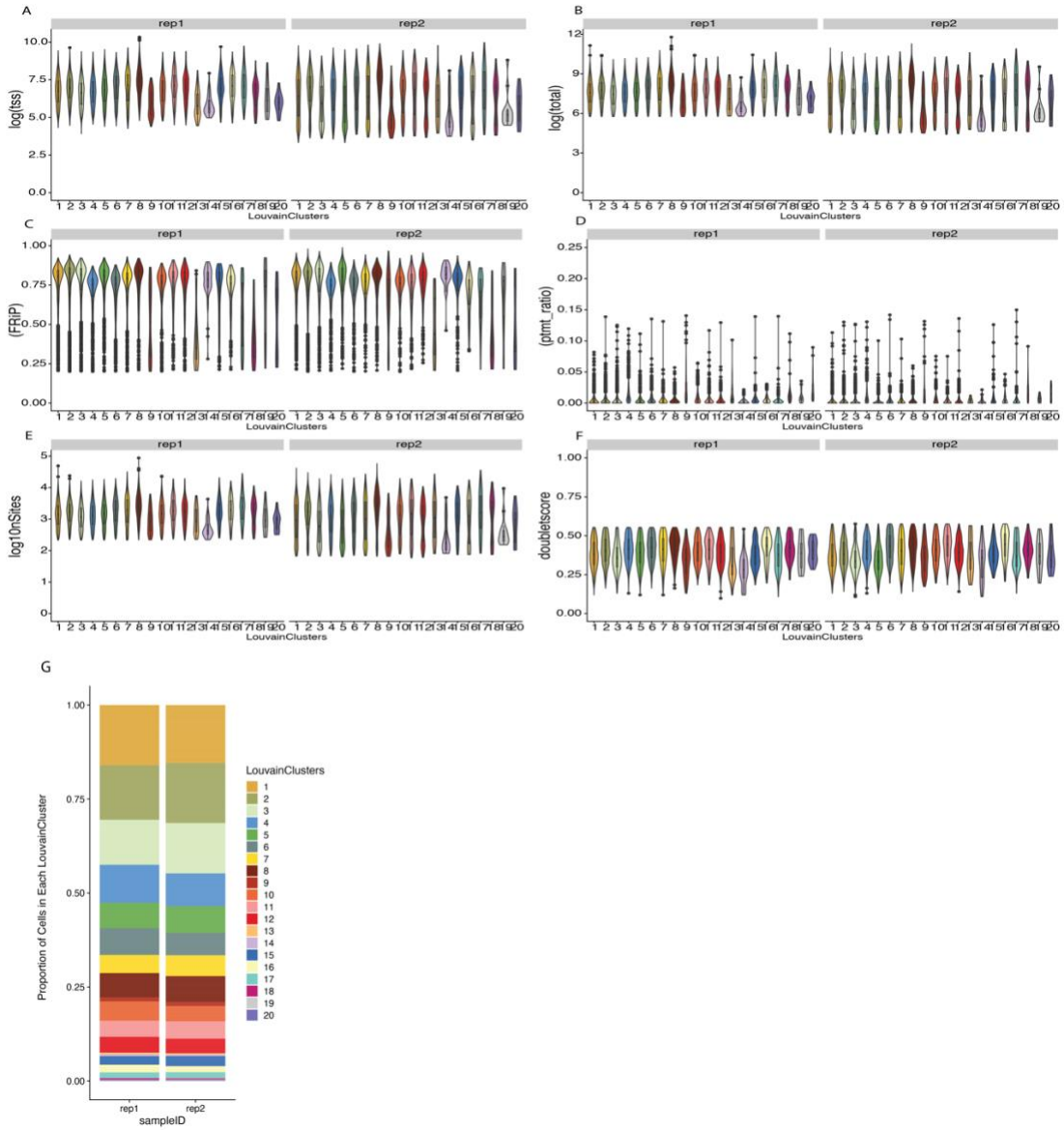
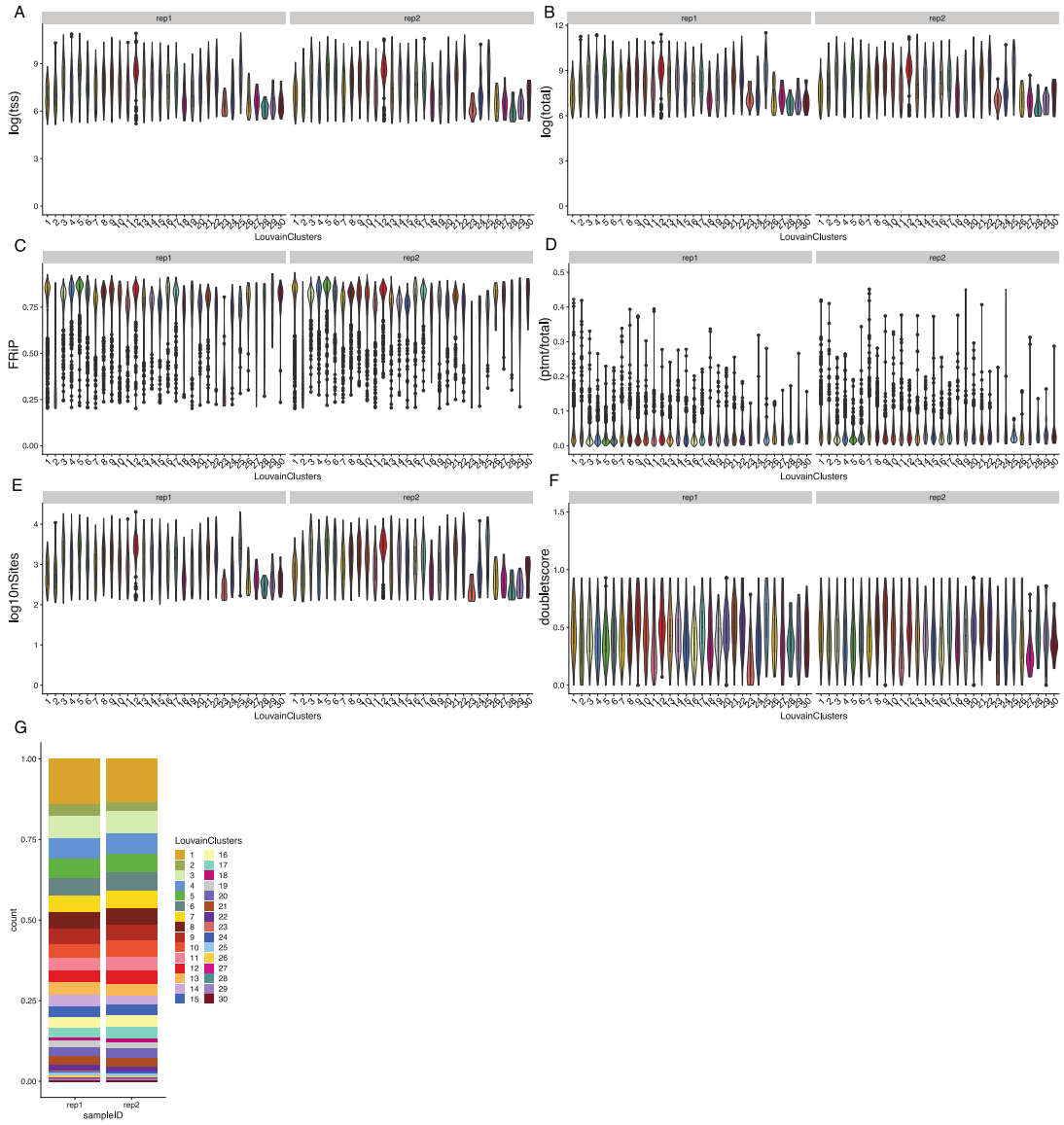


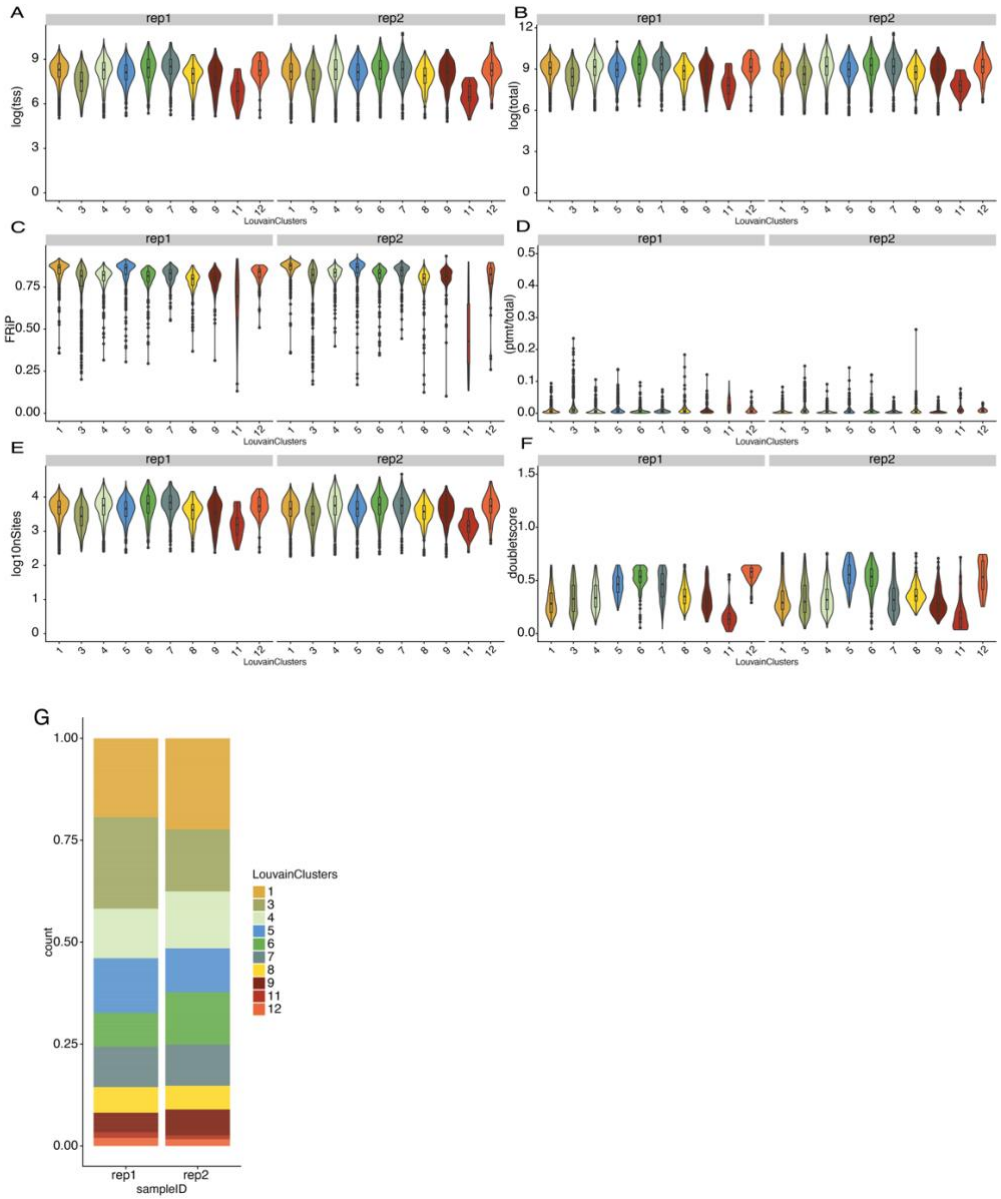
Supplemental Figure 1: Isolation of high-quality cells in sciATAC-seq libraries. Left) Knee plot of cells, where the x-axis is the rank of each individual UMI barcode based off of the number of Tn5 integrations found per nuclei, and the y-axis is the number of Tn5 integrations log10 scaled. Cut offs were drawn based on the number of expected nuclei per replicate, and only nuclei with greater than 500 Tn5 sites were considered. The red dashed lines pinpoint the cutoffs. **Middle)** Density plot of nuclei and the number of Tn5 sites found around the TSS of protein coding genes as a fraction of the total reads per nuclei. Red dashed line represents the applied cut offs. **Right)** Density plot of nuclei and the number of Tn5 sites found in peaks related to the number of total Tn5 sites per nuclei. Peaks were called on the pseudo-bulked library. Red dashed line represents the applied cut offs. **A**) The nuclei isolation metrics for *Z.mays* **B**) The nuclei isolation metrics for *S. bicolor* **C**) The nuclei isolation metrics for *P. miliaceum* **D**) The nuclei isolation metrics for *U. fusca* **E**) The nuclei isolation metrics for *O. sativa*



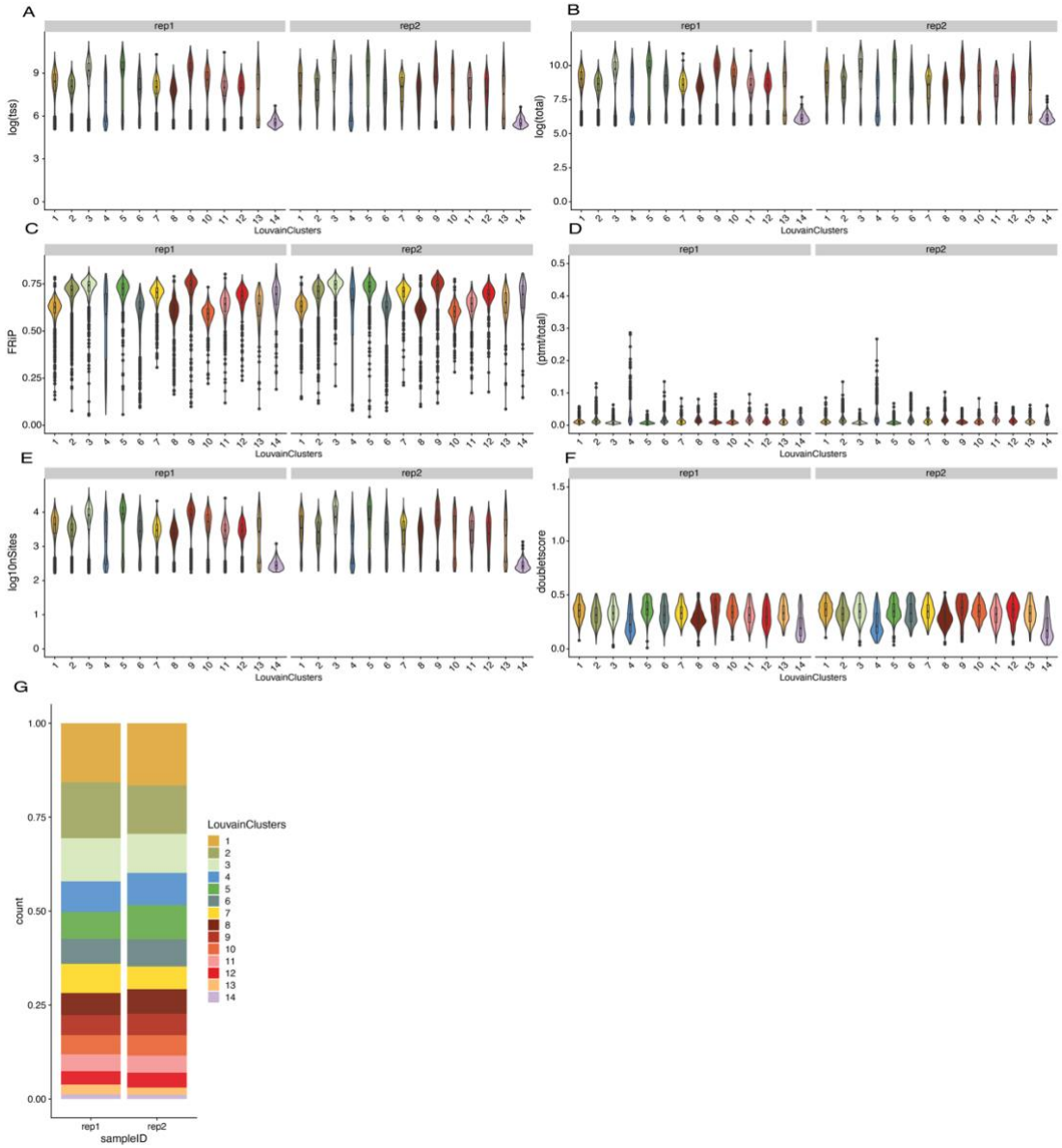
Supplemental Figure 2: **Quality control of sciATAC-seq replicates in *Z. mays*.** **A**) Violin plot with x-axis labeling the louvain clusters identified by clustering on genomic bins and generating a UMAP embedding (check methods). Y-axis, the number of Tn5 integration events occurring in the TSS of protein coding genes log transformed. **B**) Violin plot with louvain cluster number on the x-axis, with the y-axis showing the distribution of the total number of Tn5 integration events in each louvain cluster log transformed. **C**) Violin plot with louvain cluster number on the x-axis, and the y-axis being the Fraction of Reads In Peaks (FRiP) for each louvain cluster. **D**) Violin plot with louvain cluster number on the x-axis, and the y-axis representing the ratio of the number of organelle reads (here mitochondria or chloroplast) divided by the total number of reads in each nuclei. **E**) Violin plot with louvain cluster number on the x-axis, y-axis is log number of ACRs which a cell has a a Tn5 integration event in **F**) A violin plot with louvain cluster number on the x-axis, on the y-axis is doublet scores, with higher scores indicating a higher probability of a cell is a doublet. **G**) Barplot of the proportion of cells in each replicate in each louvain cluster.



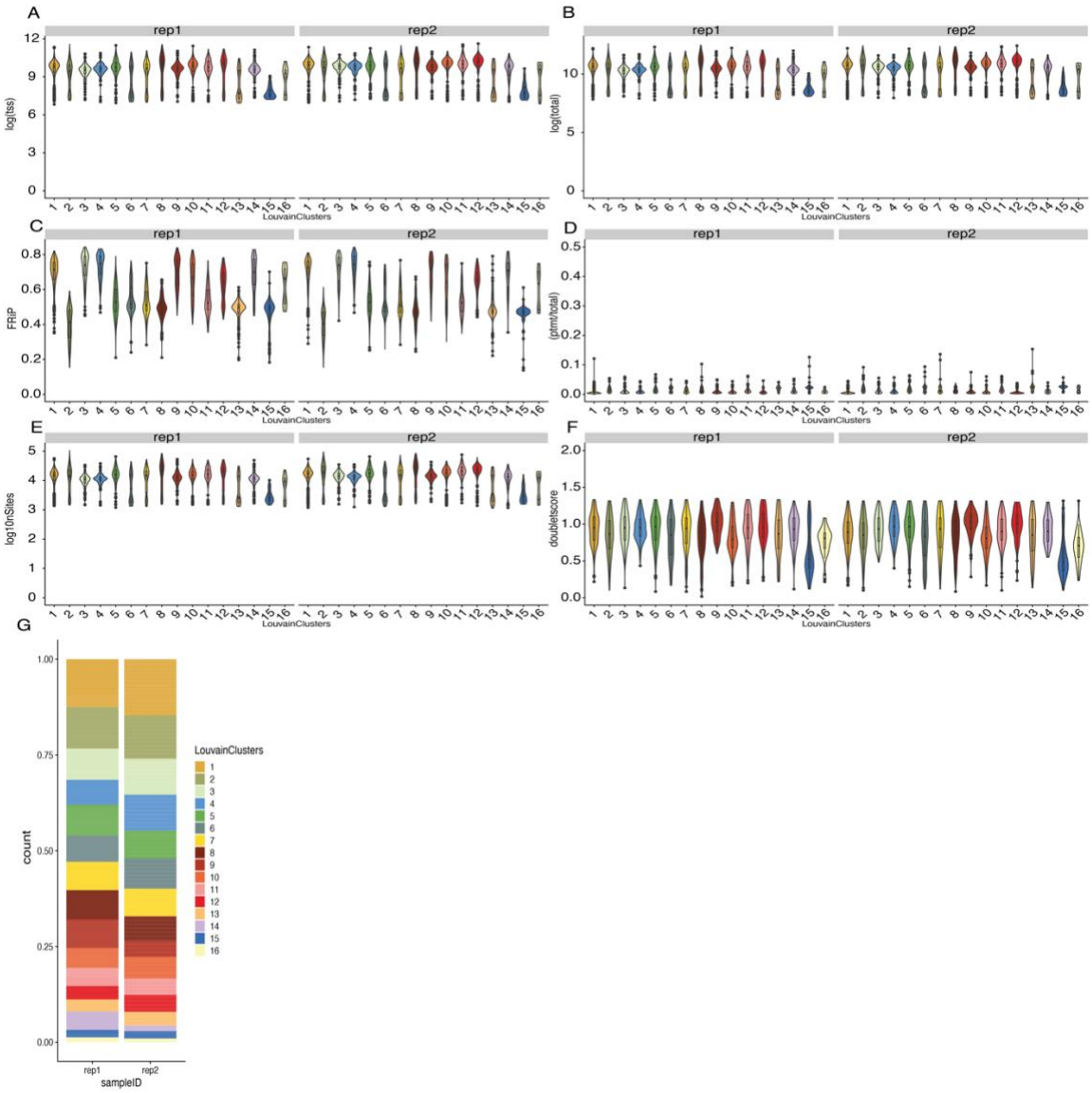
Supplemental Figure 3: **Quality control of sciATAC-seq replicates in *S. bicolor.* A**) Violin plot with x-axis labeling the louvain clusters identified by clustering on genomic bins and generating a UMAP embedding (check methods). Y-axis, the number of Tn5 integration events occurring in the TSS of protein coding genes log transformed. **B**) Violin plot with louvain cluster number on the x-axis, with the y-axis showing the distribution of the total number of Tn5 integration events in each louvain cluster log transformed. **C**) A violin plot with louvain cluster number on the x-axis, and the y-axis being the Fraction of Reads In Peaks (FRiP) for each louvain cluster. **D**) Violin plot with louvain cluster number on the x-axis, and the y-axis representing the ratio of the number of organelle reads (here mitochondria or chloroplast) divided by the total number of reads in the nuclei. **E**) A violin plot with louvain cluster number on the x-axis, y-axis is the log number of ACRs which a cell has found a Tn5 integration event in **F**) A violin plot with louvain cluster number on the x-axis, on the y-axis is doublet scores, with higher scores indicating a higher chance of a cell being a doublet. **G**) Barplot of the proportion of cells in each replicate in each louvain cluster.



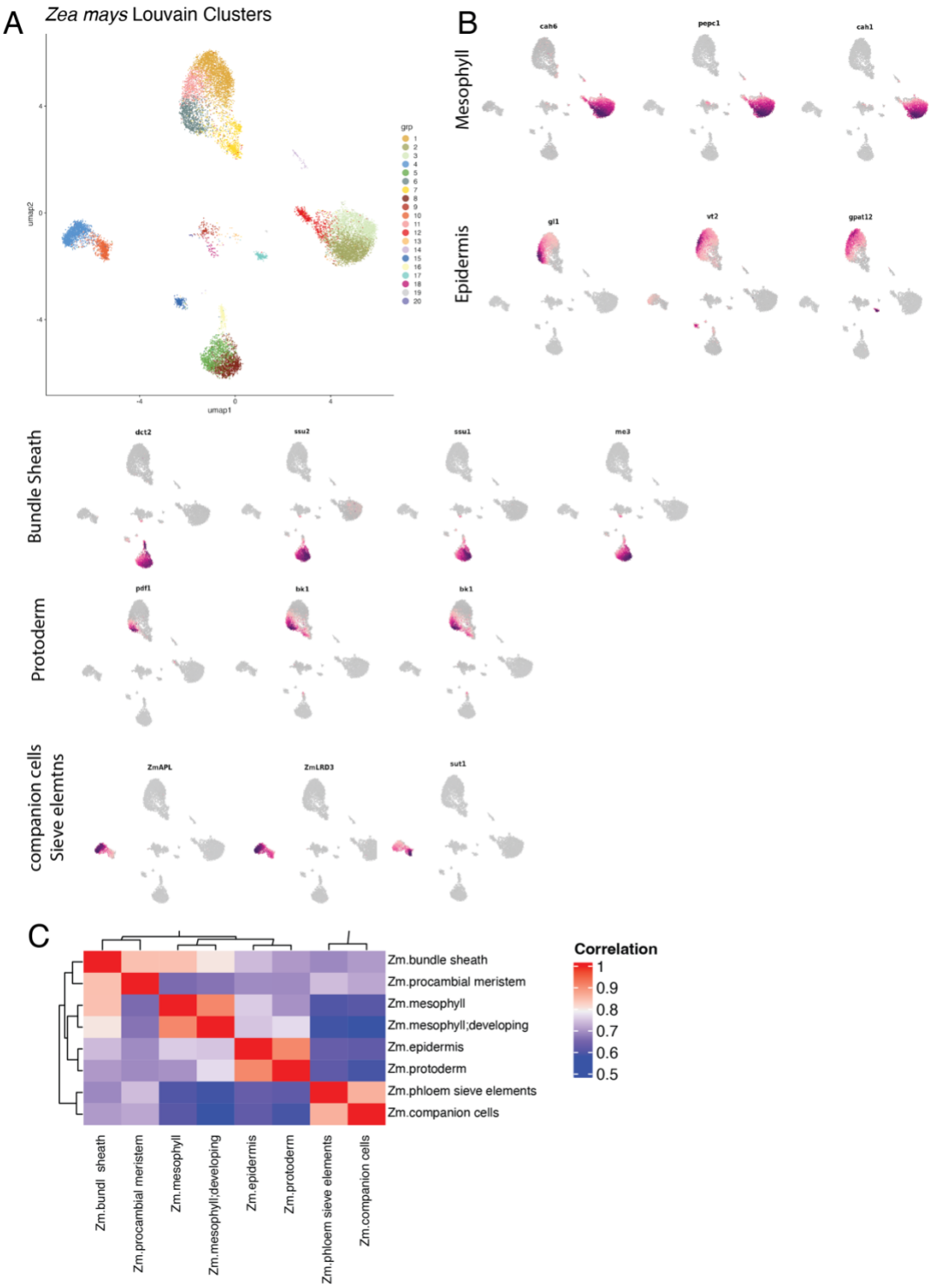
Supplemental Figure 4: **Quality control of sciATAC-seq replicates in *P. miliaceum.* A**) Violin plot with x-axis labeling the louvain clusters identified by clustering on genomic bins and generating a UMAP embedding (check methods). Y-axis, the number of Tn5 integration events occurring in the TSS of protein coding genes log transformed. **B**) Violin plot with louvain cluster number on the x-axis, with the y-axis showing the distribution of the total number of Tn5 integration events in each louvain cluster log transformed. **C**) A violin plot with louvain cluster number on the x-axis, and the y-axis being the Fraction of Reads In Peaks (FRiP) for each louvain cluster. **D**) Violin plot with louvain cluster number on the x-axis, and the y-axis representing the ratio of the number of organelle reads (here mitochondria or chloroplast) divided by the total number of reads in the nuclei. **E**) A violin plot with louvain cluster number on the x-axis, y-axis is the log number of ACRs which a cell has found a Tn5 integration event in **F**) A violin plot with louvain cluster number on the x-axis, on the y-axis is doublet scores, with higher scores indicating a higher chance of a cell being a doublet. **G**) Barplot of the proportion of cells in each replicate in each louvain cluster.



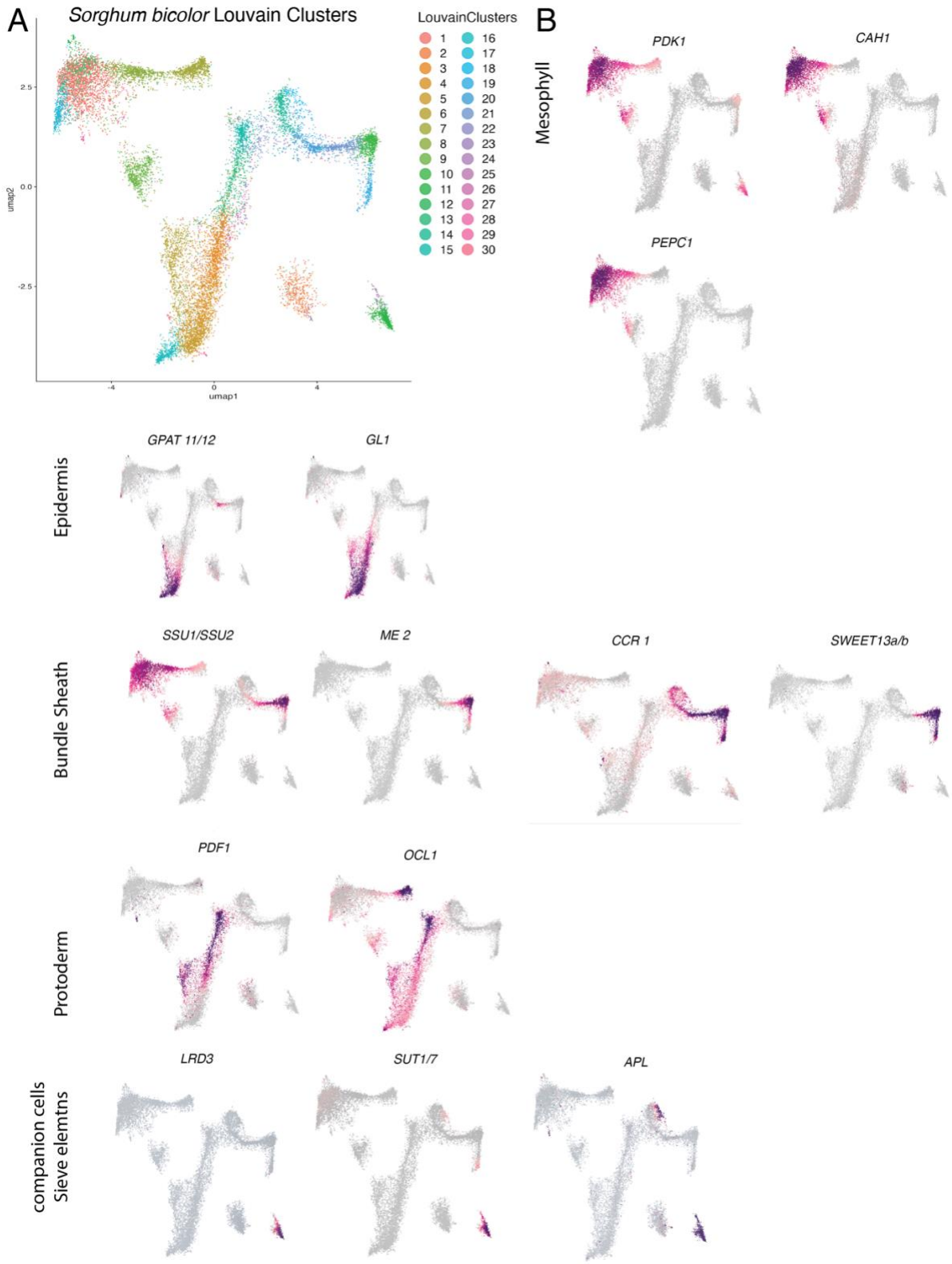
Supplemental Figure 5: **Quality control of sciATAC-seq replicates in *U. fusca.*** **A**) Violin plot with x-axis labeling the louvain clusters identified by clustering on genomic bins and generating a UMAP embedding (check methods). Y-axis, the number of Tn5 integration events occurring in the TSS of protein coding genes log transformed. **B**) Violin plot with louvain cluster number on the x-axis, with the y-axis showing the distribution of the total number of Tn5 integration events in each louvain cluster log transformed. **C**) A violin plot with louvain cluster number on the x-axis, and the y-axis being the Fraction of Reads In Peaks (FRiP) for each louvain cluster. **D**) Violin plot with louvain cluster number on the x-axis, and the y-axis representing the ratio of the number of organelle reads (here mitochondria or chloroplast) divided by the total number of reads in the nuclei. **E**) A violin plot with louvain cluster number on the x-axis, y-axis is the log number of ACRs which a cell has found a Tn5 integration event in **F**) A violin plot with louvain cluster number on the x-axis, on the y-axis is doublet scores, with higher scores indicating a higher chance of a cell being a doublet. **G**) Barplot of the proportion of cells in each replicate in each louvain cluster.



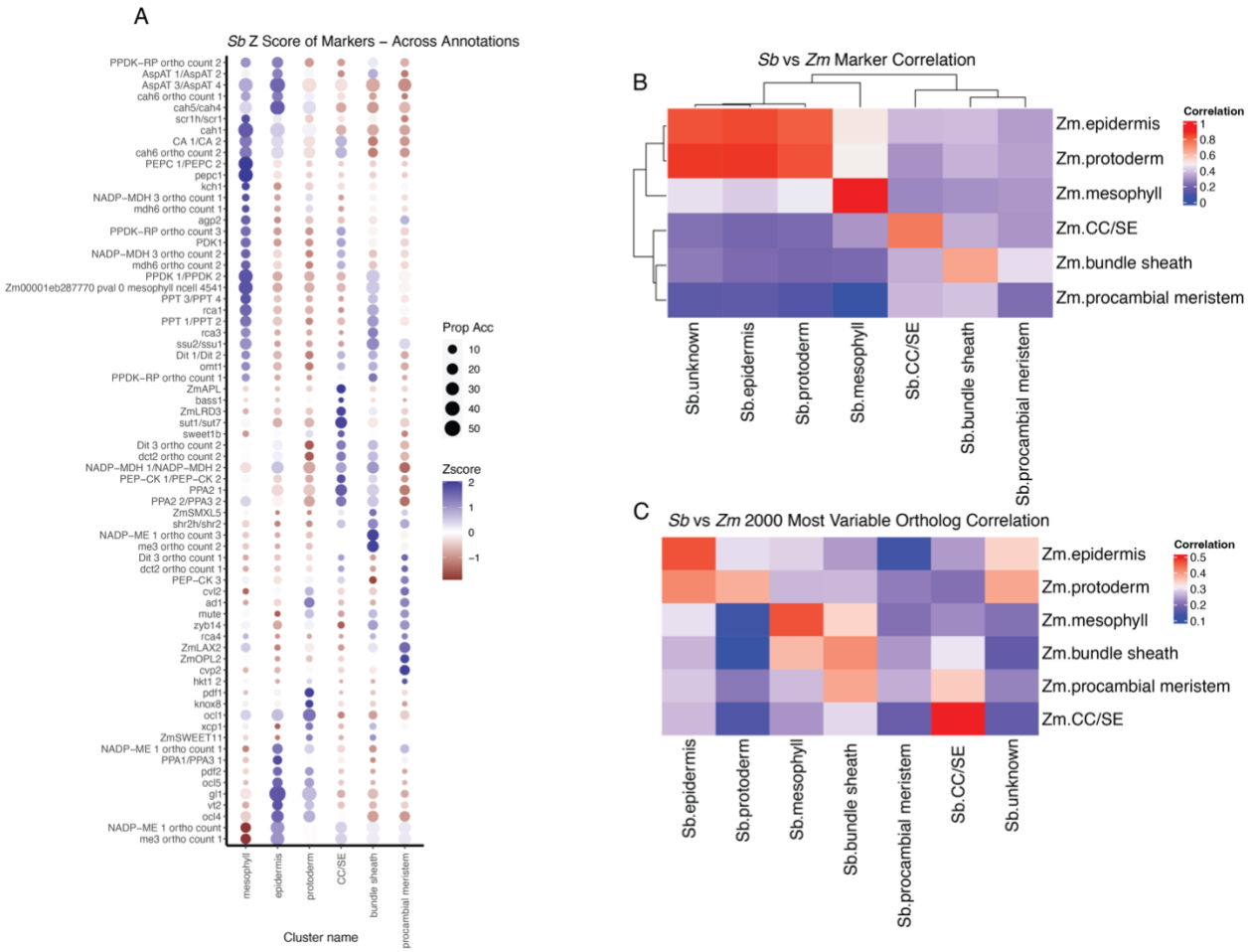
Supplemental Figure 6: **Quality control of sciATAC-seq replicates of *O. sativa.* A**) Violin plot with x-axis labeling the louvain clusters identified by clustering on genomic bins and generating a UMAP embedding (check methods). Y-axis, the number of Tn5 integration events occurring in the TSS of protein coding genes log transformed. **B**) Violin plot with louvain cluster number on the x-axis, with the y-axis showing the distribution of the total number of Tn5 integration events in each louvain cluster log transformed. **C**) A violin plot with louvain cluster number on the x-axis, and the y-axis being the Fraction of Reads In Peaks (FRiP) for each louvain cluster. **D**) Violin plot with louvain cluster number on the x-axis, and the y-axis representing the ratio of the number of organelle reads (here mitochondria or chloroplast) divided by the total number of reads in the nuclei. **E**) A violin plot with louvain cluster number on the x-axis, y-axis is the log number of ACRs which a cell has found a Tn5 integration event in **F**) A violin plot with louvain cluster number on the x-axis, on the y-axis is doublet scores, with higher scores indicating a higher chance of a cell being a doublet. **G**) Barplot of the proportion of cells in each replicate in each louvain cluster.



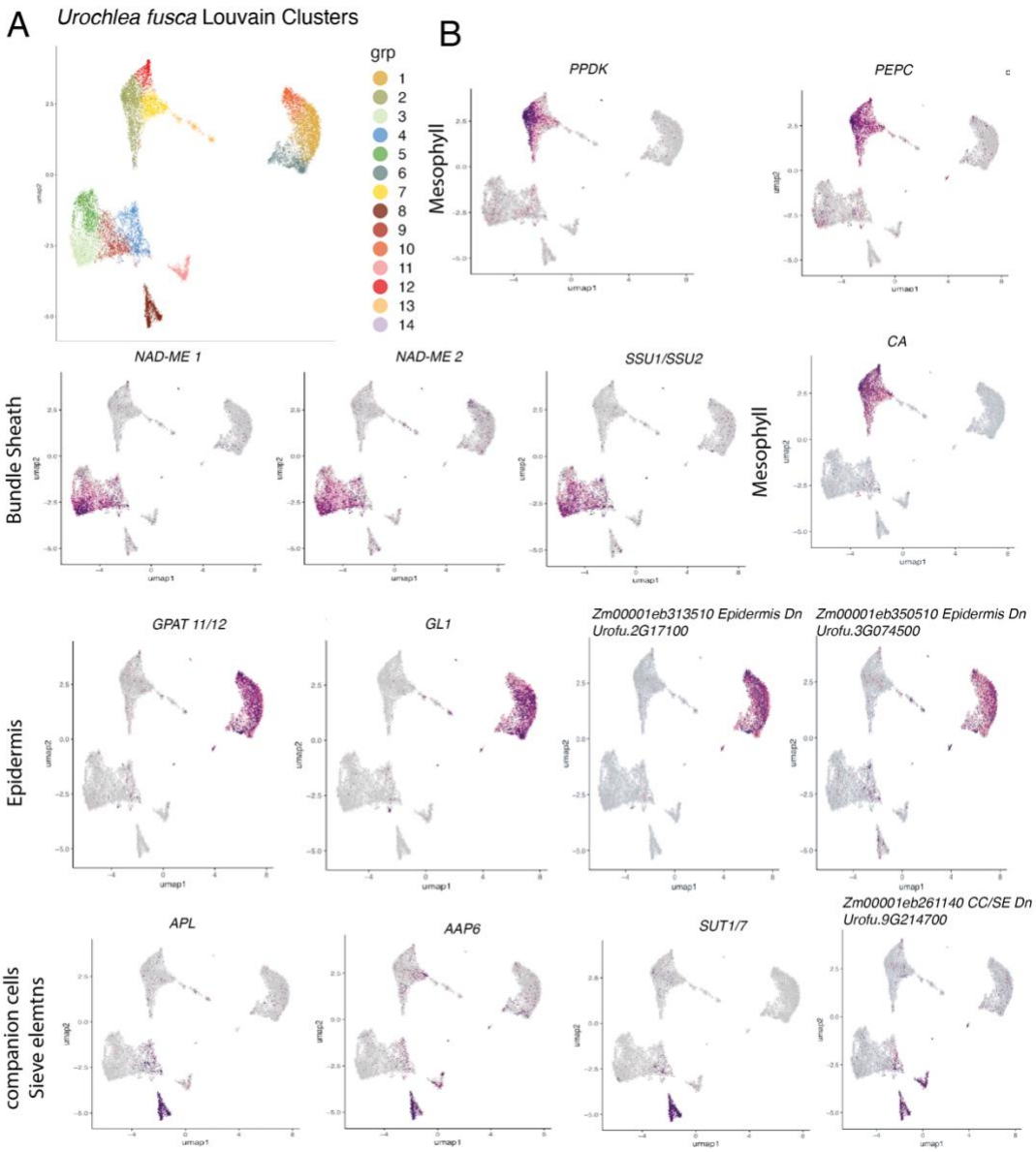
Supplemental Figure 7: Annotation of the *Z. mays* dataset using markers. A) The louvain clusters identified in *Z. mays* in their UMAP embeddings (Methods) **B**) Selected marker genes used to annotate various cell types. The cell type label can be found on the left side, and the marker name can be found on the top of each individual UMAP. **C)** Correlation of *Z. mays* gene body chromatin accessibility of all cell types using the top 2,000 most variable genes. Cell types were ordered using hierarchical clustering.



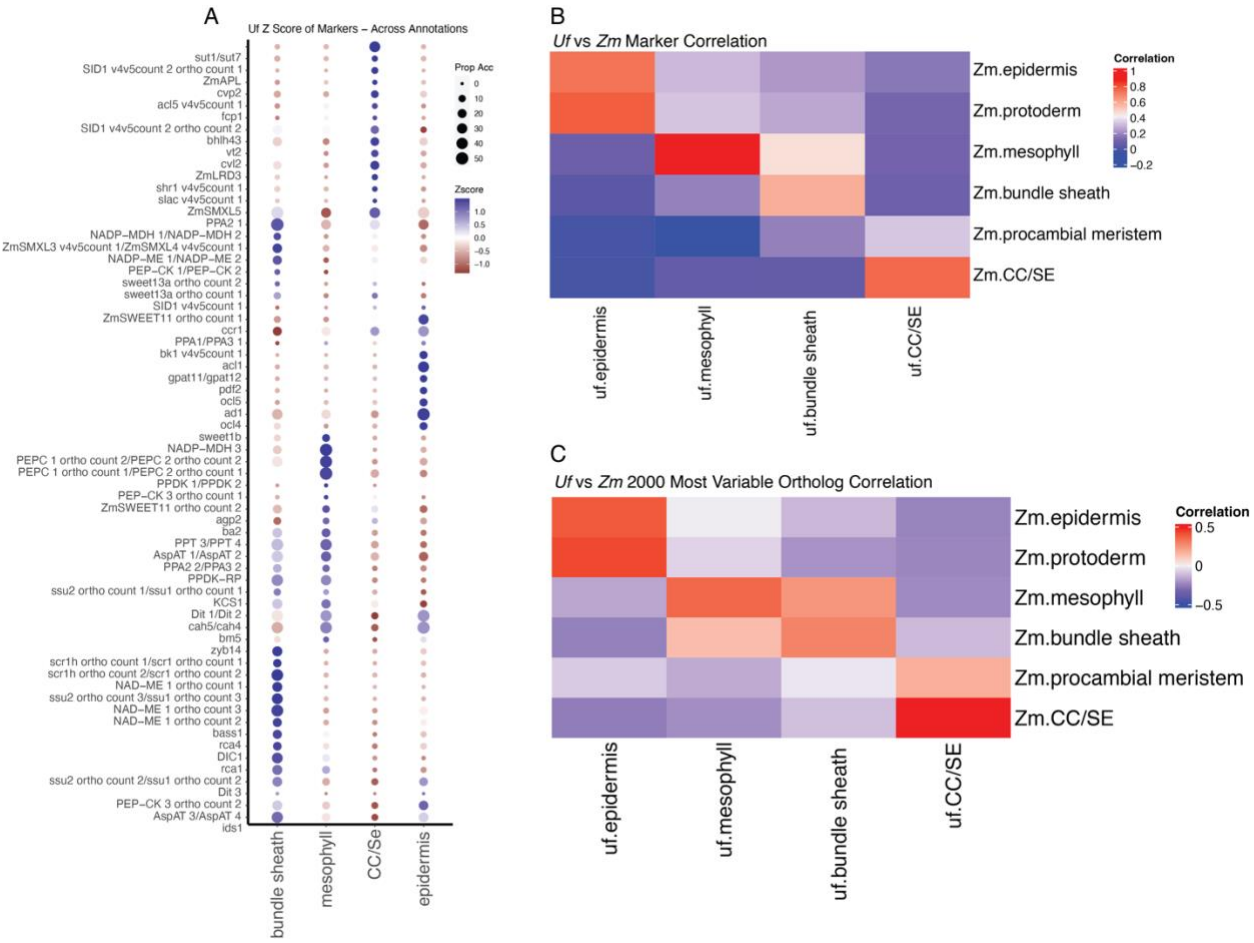
Supplemental Figure 8: Annotation of the *S. bicolor* dataset using markers. A) The louvain clusters identified in *S. bicolor* in their corresponding UMAP embeddings (Methods) **B)** Selected marker genes used to annotate various cell types. The cell type label can be found on the left side, and the marker name can be found on the top of each individual UMAP.



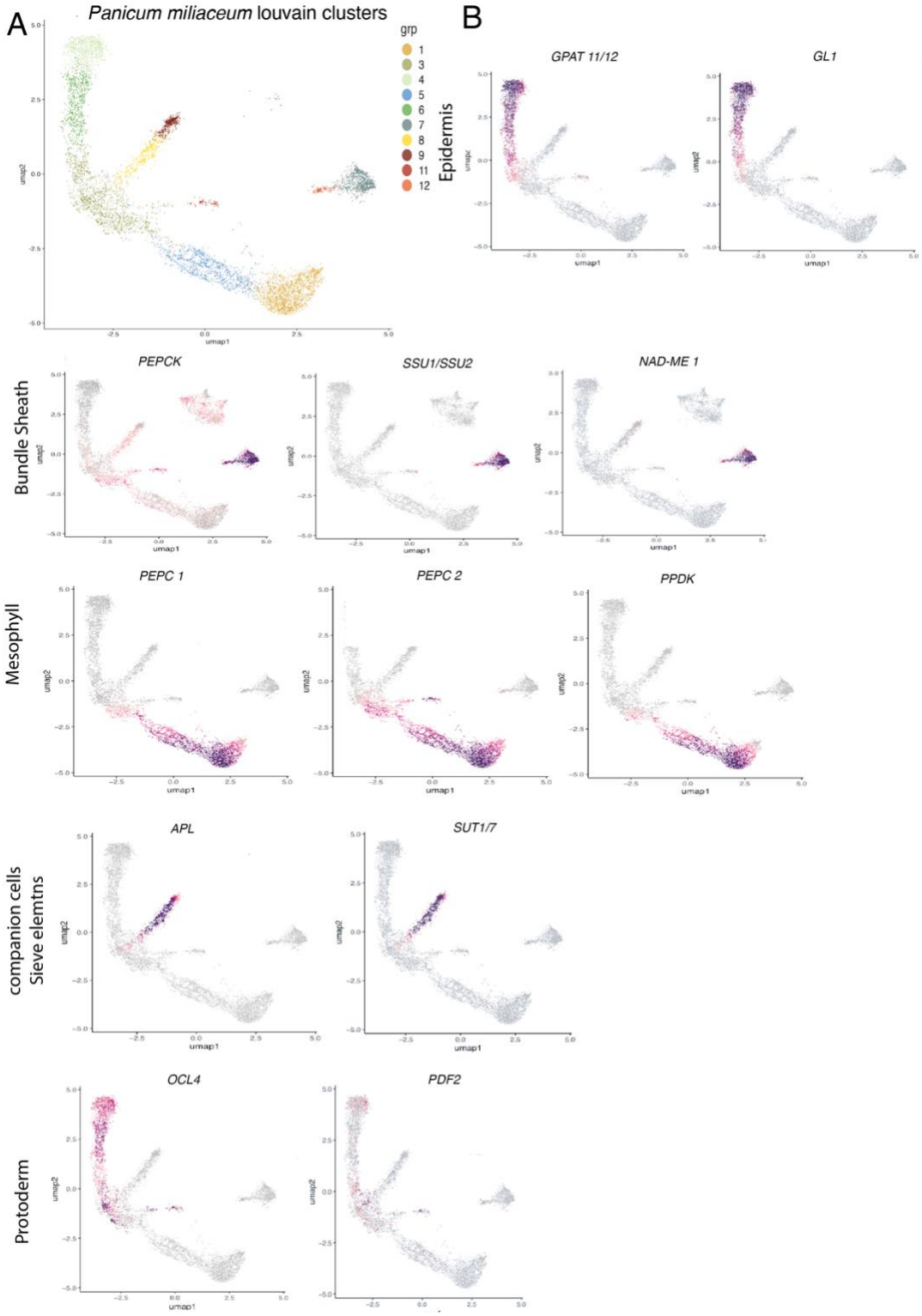
Supplemental Figure 9: Additional annotation data for *S. bicolor* A) Dotplots for various marker genes used to annotate cell types. The y-axis represents cell types, and the x-axis is a list marker genes. The size of each circle is proportional to the number of cells within that cell type that showed chromatin accessibility of the marker. Each dot is colored by is z-score transformed values across the clusters. **B**) Correlation of marker genes between *Z. mays* and *S. bicolor*. Correlations were done using gene body chromatin accessibility, and a linear model was used to control for species **C**) Correlation of the 2,000 most variable single copy orthologs between *Z. mays* and *S. bicolor*. Orthologs were selected by taking the average variation between both datasets. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species specific effects.



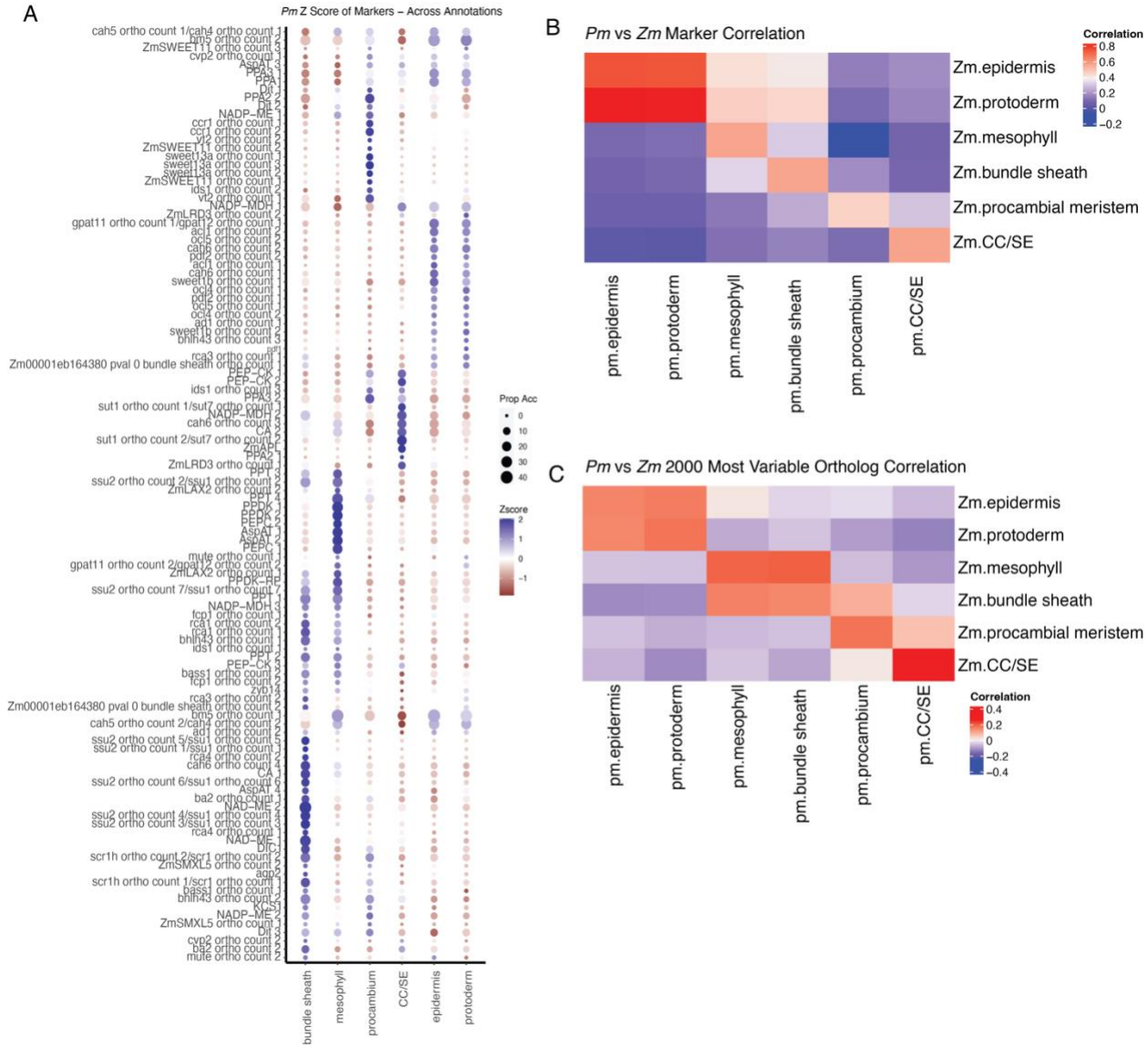
Supplemental Figure 10: Annotation of the *U. fusca* dataset using markers. A) The louvain clusters identified in *U. fusca* in their corresponding UMAP embeddings (Methods) **B**) Selected marker genes used to annotate various cell types. The cell type label can be found on the left side, and the marker name can be found on the top of each individual UMAP.



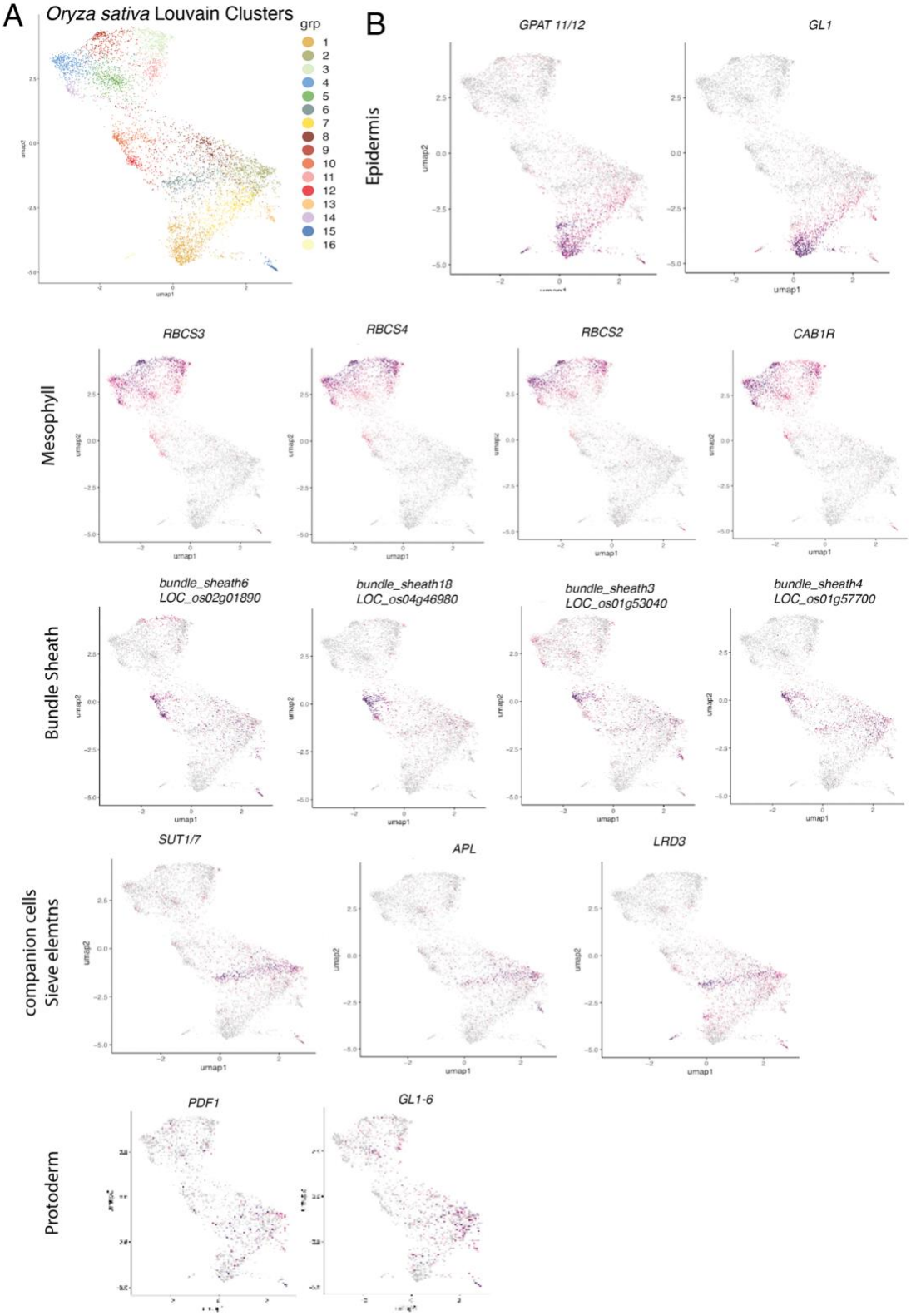
Supplemental Figure 11: Additional annotation data for *U. fusca* A) Dotplots for various marker genes used to annotate cell types. The y-axis represents cell types, and the x-axis is a list marker genes. The size of each circle is proportional to the number of cells within that cell type that showed chromatin accessibility of the marker. Each dot is colored by is z-score transformed values across the clusters. **B**) Correlation of marker genes between *Z. mays* and *U. fusca*. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species **C**) Correlation of the 2,000 most variable single copy orthologs between *Z. mays* and *U. fusca*. Orthologs were selected by taking the average variation between both datasets. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species specific effects.



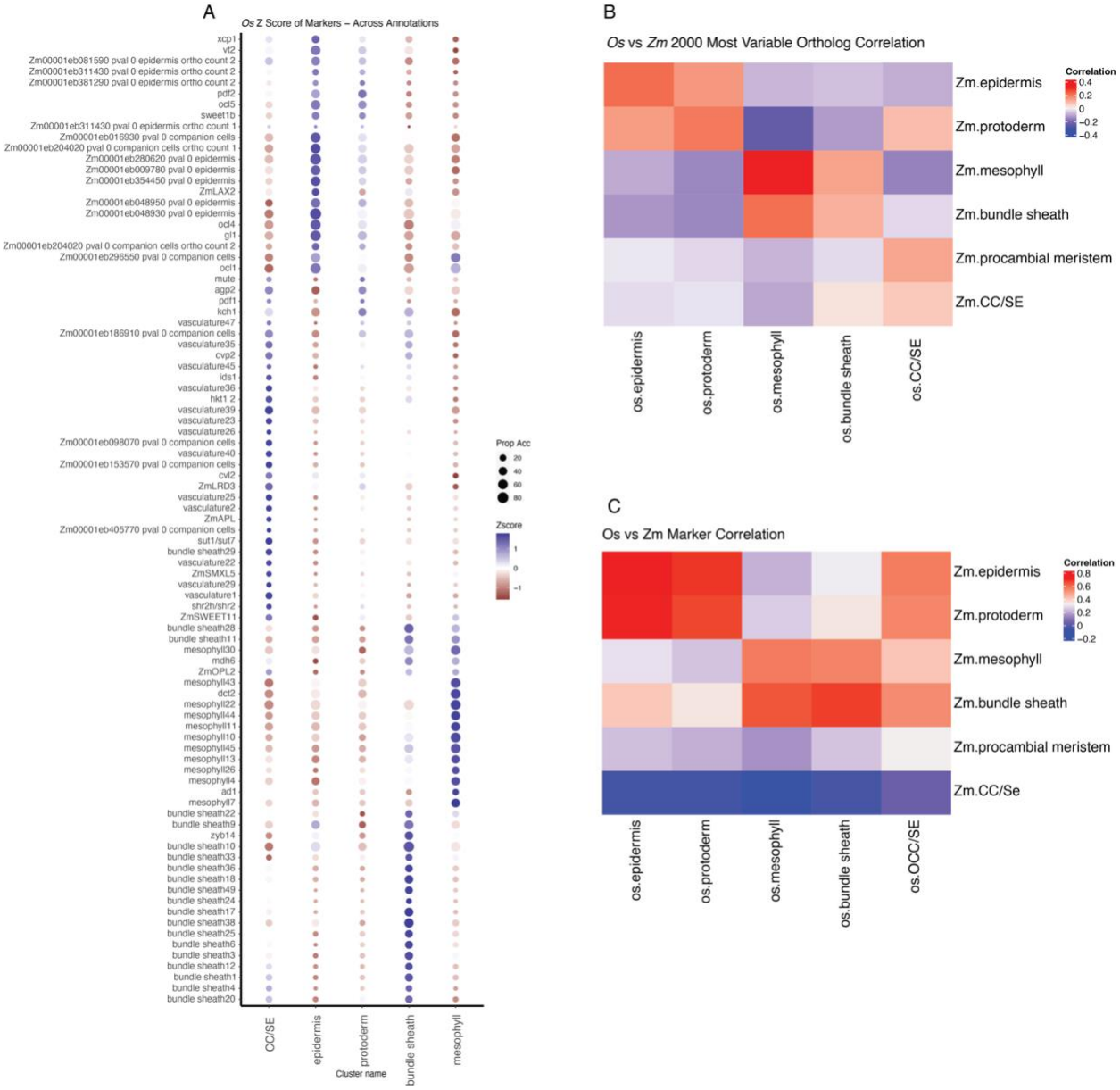
Supplemental Figure 12: Annotation of the *P. miliaceum* dataset using markers. A) The louvain clusters identified in *P. miliaceum* in their corresponding UMAP embeddings (Methods) **B**) Selected marker genes used to annotate various cell types. The cell type label can be found on the left side, and the marker name can be found on the top of each individual UMAP.



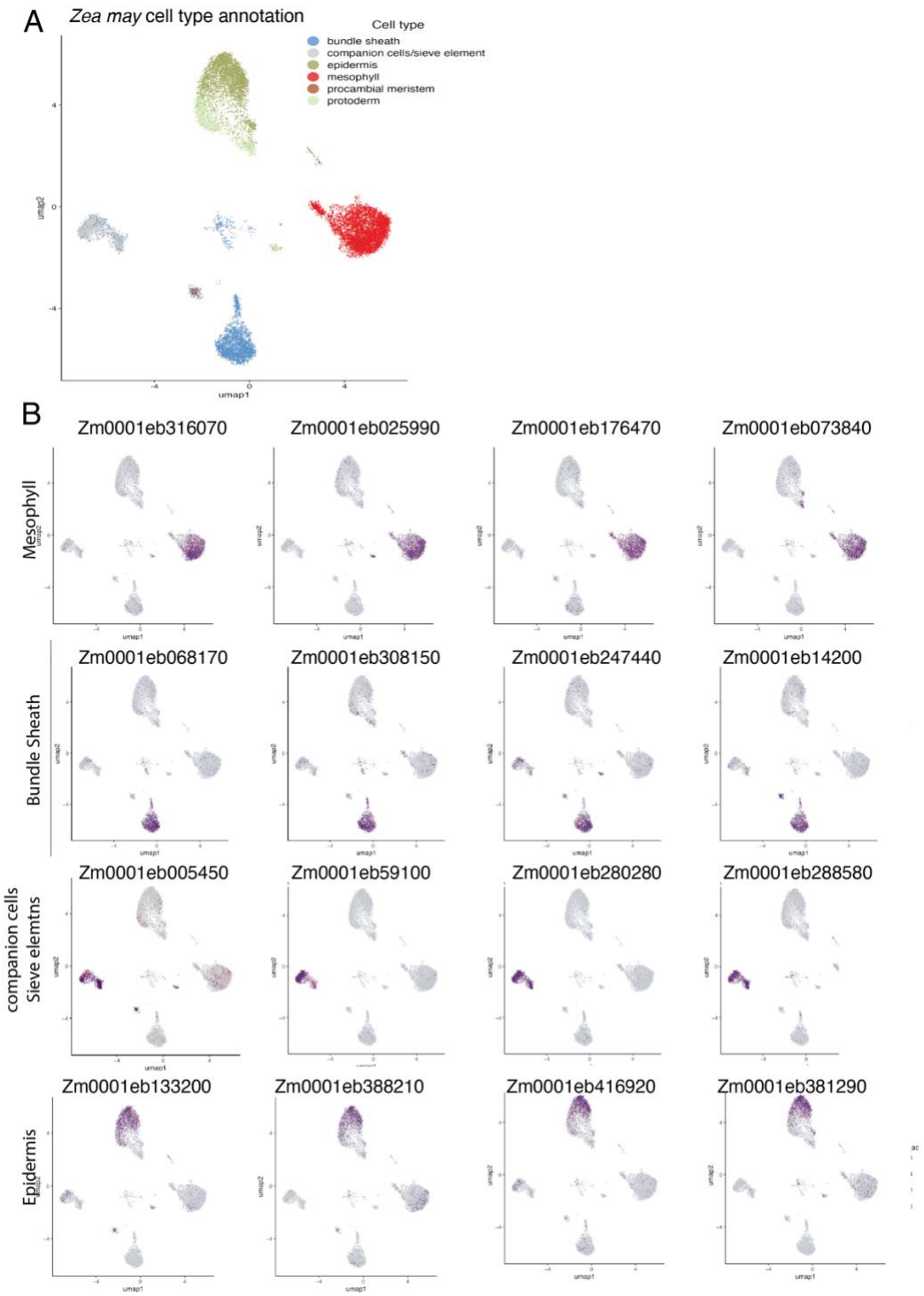
Supplemental Figure 13: Additional annotation data for *P. miliaceum* **A**) Dotplots for various marker genes used to annotate cell-types. The y-axis represents cell types, and the x-axis is a list marker genes. The size of each circle is proportional to the number of cells within that cell type that showed chromatin accessibility of the marker. Each dot is colored by is z-score transformed values across the clusters. **B**) Correlation of marker genes between *Z. mays* and *P. miliaceum*. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species **C**) Correlation of the 2,000 most variable single copy orthologs between *Z. mays* and *P. miliaceum*. Orthologs were selected by taking the average variation between both datasets. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species specific effects.



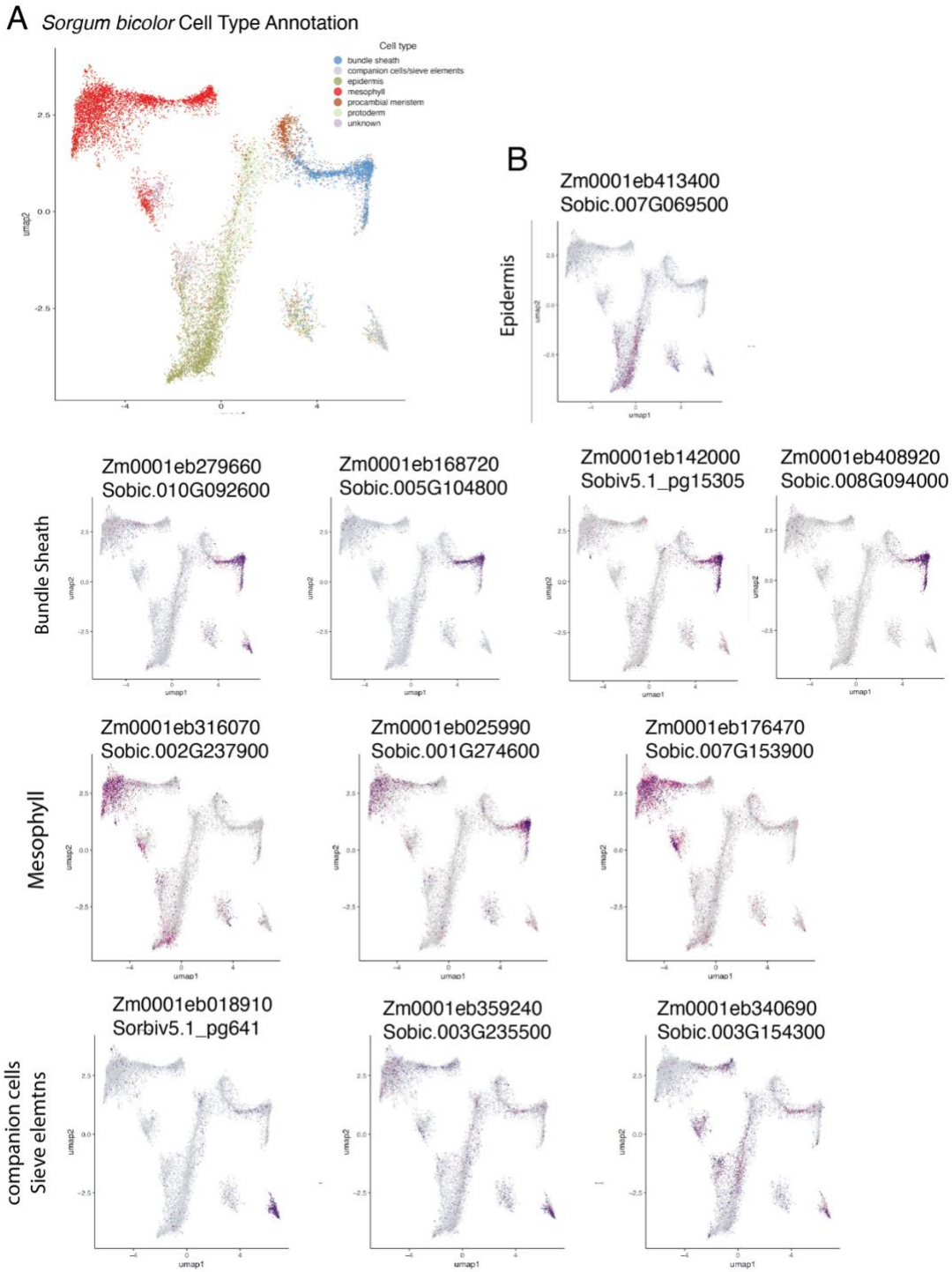
Supplemental Figure 14: Annotation of the *O. sativa* dataset using markers. A) The louvain clusters identified in *O. sativa* in their corresponding UMAP embeddings (Methods) **B**) Selected marker genes used to annotate various cell types. The cell type label can be found on the left side, and the marker name can be found on the top of each individual UMAP.



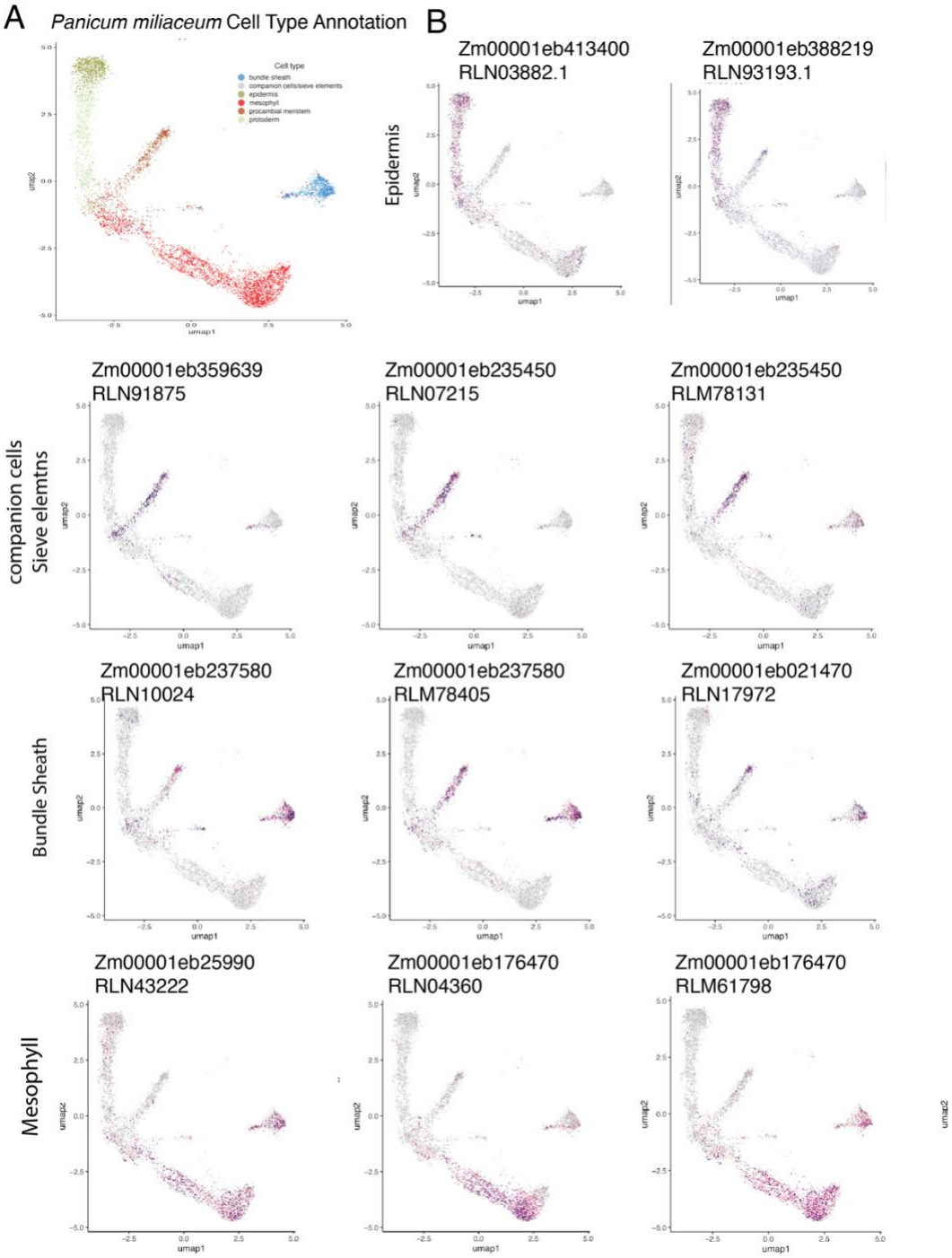
Supplemental Figure 15: Additional annotation data for *O. sativa* **A**) Dotplots for various marker genes used to annotate cell types. The y-axis represents cell types, and the x-axis is a list marker genes. The size of each circle is proportional to the number of cells within that cell type that showed chromatin accessibility of the marker. Each dot is colored by is z-score transformed values across the clusters. **B**) Correlation of marker genes between *Z. mays* and *O.sativa*. Correlations were done using gene body chromatin accessibility, and a linear model was used to control for species **C**) Correlation of the 2,000 most variable single copy orthologs between *Z. mays* and *O.sativa*. Orthologs were selected by taking the average variation between both datasets. Correlations were done using gene body chromatin accessibility, and a linear model was used in order to control for species specific effects.



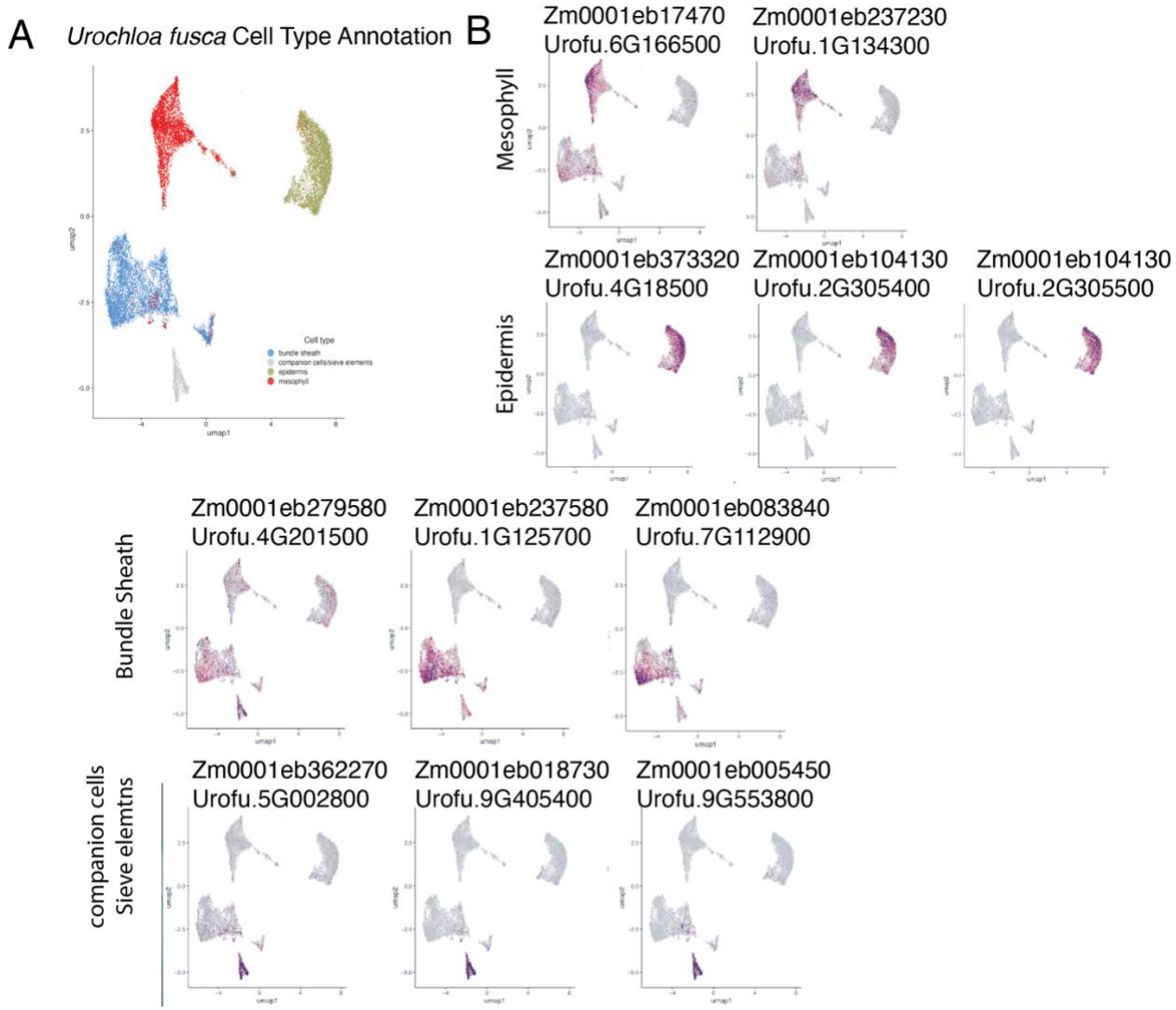
Supplemental Figure 16: ***De novo* marker discovery in diverse cell-types in *Z. mays*.** **A**) UMAP embeddings of the *Z. mays* annotation used for de novo discovery of marker genes. Each dot represents a cell, and the annotated cell type is labeled by the color. **B**) Representative examples of marker genes discovered using De-seq2, and gene body accessibility. Each row corresponds to a cell-type with the geneID in question at the top. Chromatin accessibility patterns were imputed using only a step size of one, and a KNN value of 25 in order to be conservative in visualization. Cells which are more purple indicate that the gene model in question is more accessible in this cell population.



Supplemental Figure 17: *De novo* markers from *Z. mays* used in the annotation of *S. bicolor.* **A**) UMAP embeddings of the *S. bicolor* annotation. Each dot represents a cell, and the annotated cell type is labeled by the color. **B**) Examples of marker genes from *Z. mays* and their chromatin accessibility patterns in *S. bicolor.* Each row corresponds to a cell-type with the *Z. mays* gene ID at the top, and the ortholog from *S. bicolor* below. Orthologues were assigned by OrthoFinder (see methods). Chromatin accessibility patterns were imputed using only a step size of one, and a KNN value of 25 in order to be conservative in visualization. Cells which are more purple indicate that the gene model in question is more accessible in this cell population.

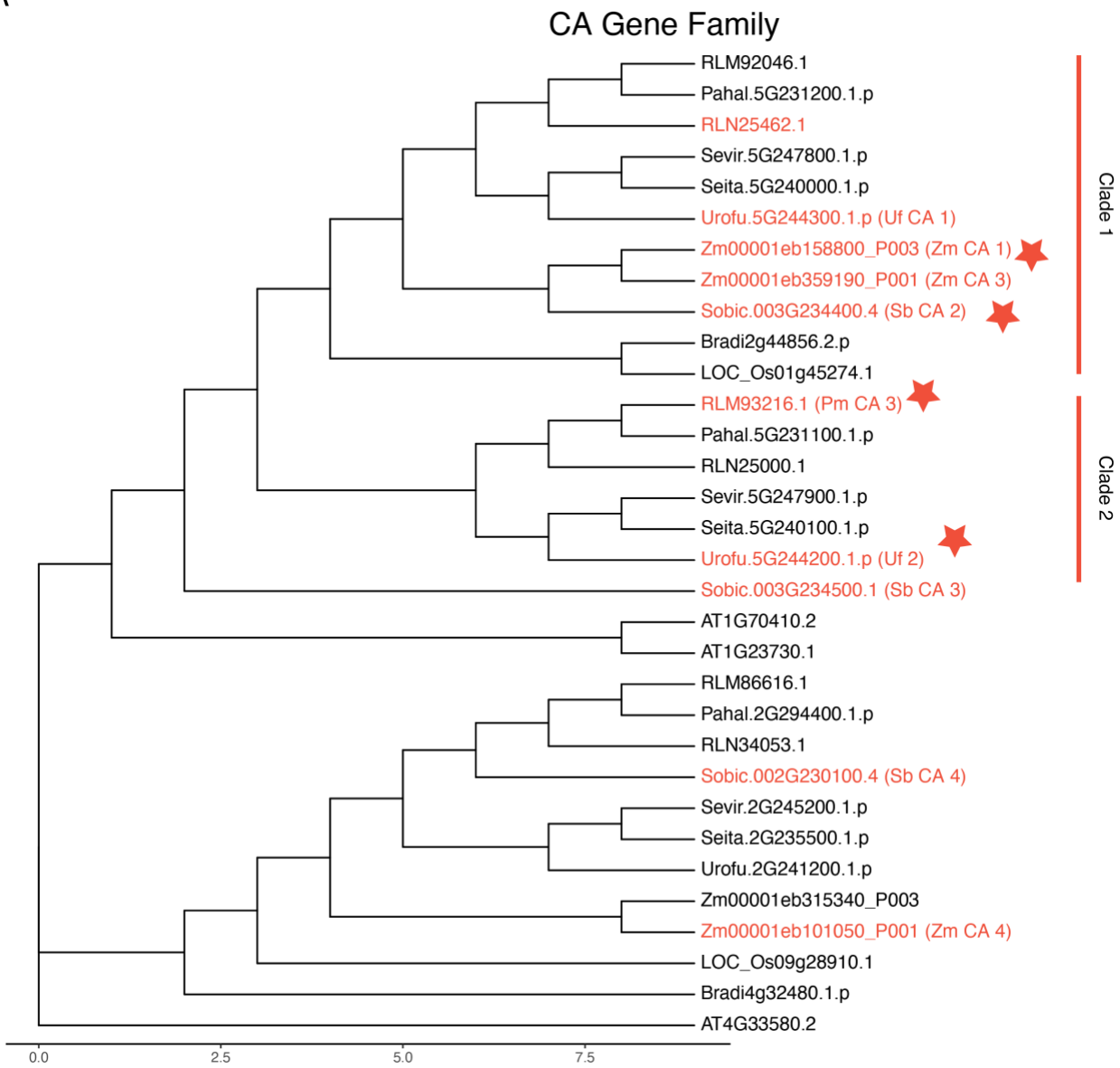


Supplemental Figure 18: *De novo* markers from *Z. mays* used in the annotation of *P. miliaceum.* **A**) UMAP embeddings of the *P. miliaceum* annotation. Each dot represents a cell, and the annotated cell type is labeled by the color. **B**) Examples of marker genes from *Z. mays* and their chromatin accessibility patterns in *P. miliaceum.* Each row corresponds to a cell-type with the *Z. mays* gene ID at the top, and the ortholog from *P. miliaceum* below. Orthologues were assigned by OrthoFinder (see methods). Chromatin accessibility patterns were imputed using only a step size of one, and a KNN value of 25 in order to be conservative in visualization. Cells which are more purple indicate that the gene model in question is more accessible in this cell population.

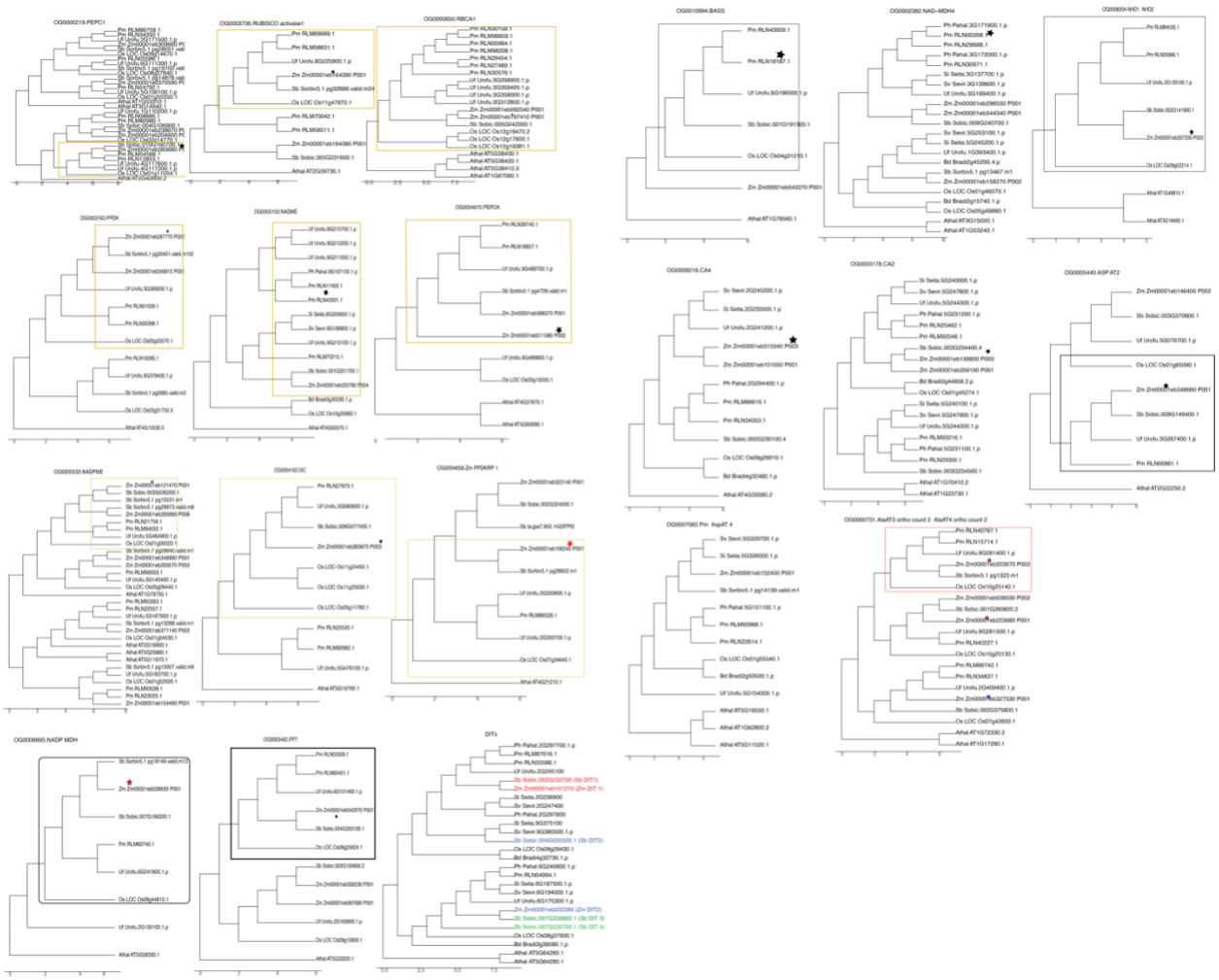


Supplemental Figure 19: *De novo* markers from *Z. mays* used in the annotation of *U. fusca.* A) UMAP embeddings of the *U. fusca* annotation. Each dot represents a cell, and the annotated cell type is labeled by the color. **B**) Examples of marker genes from *Z. mays* and their chromatin accessibility patterns in *U. fusca.* Each row corresponds to a cell type with the *Z. mays* gene ID at the top, and the ortholog from *U. fusca* below. Orthologues were assigned by OrthoFinder (see methods). Chromatin accessibility patterns were imputed using only a step size of one, and a KNN value of 25 in order to be conservative in visualization. Cells which are more purple indicate that the gene model in question is more accessible in this cell population.

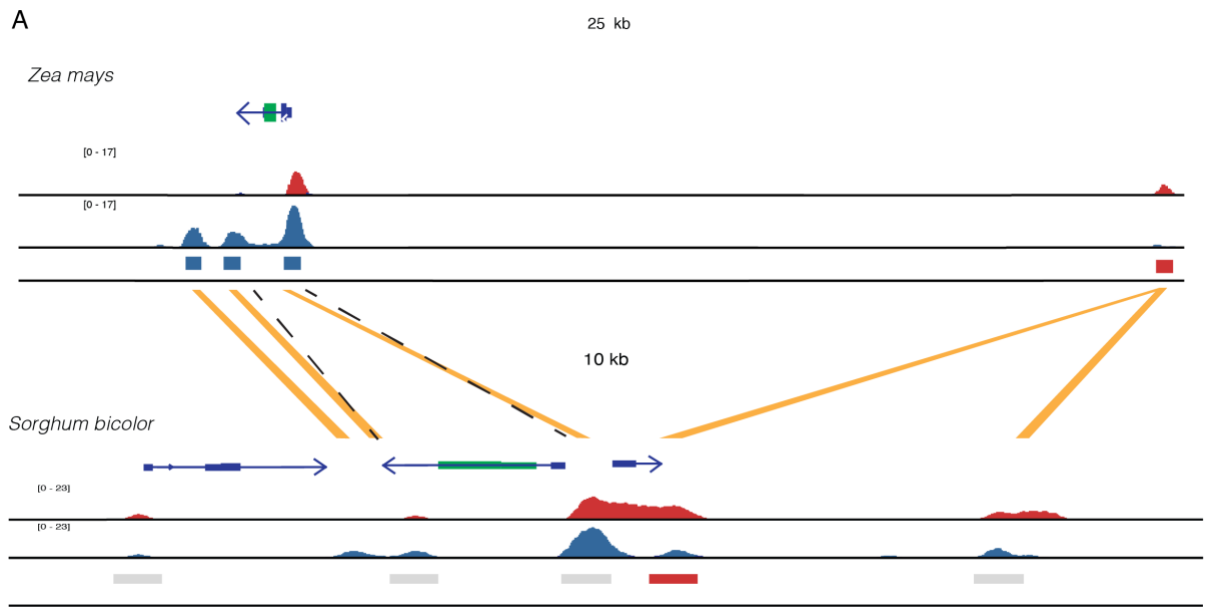
A



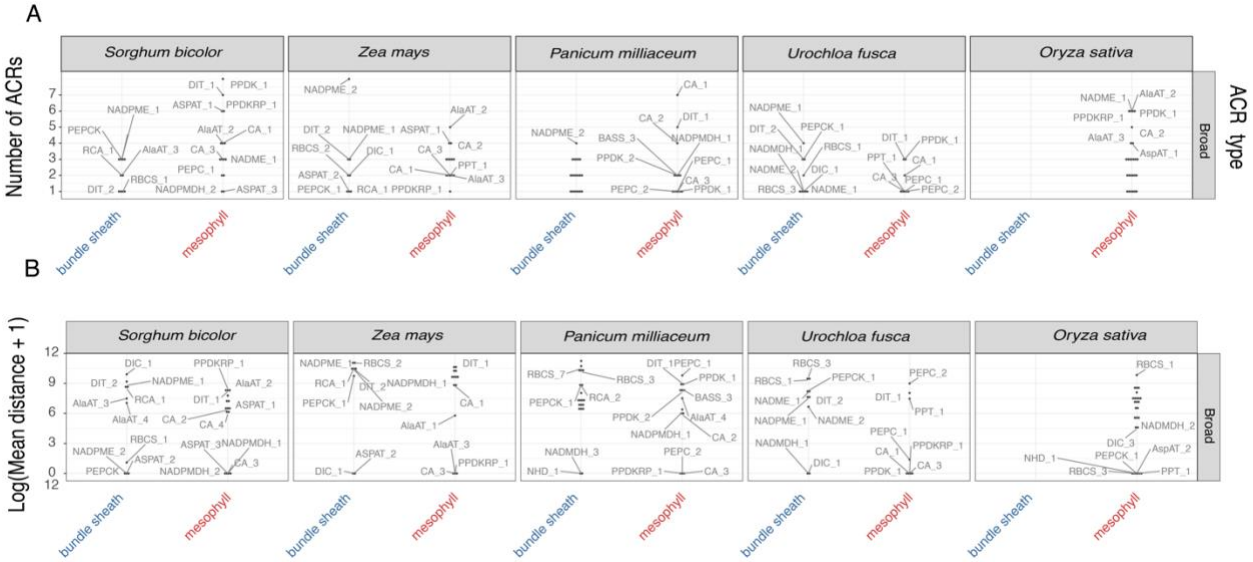
Supplemental Figure 20: **Phylogeny of the *CARBONIC ANHYDRASE* (*CA*) gene family. A**) Phylogeny of the *CA*s that are used in this study are colored in red. Gene models with a red star are the *CA*s for each species which exhibit the most cell-type-specific bias when comparing BS- and MS-specific chromatin accessibility.



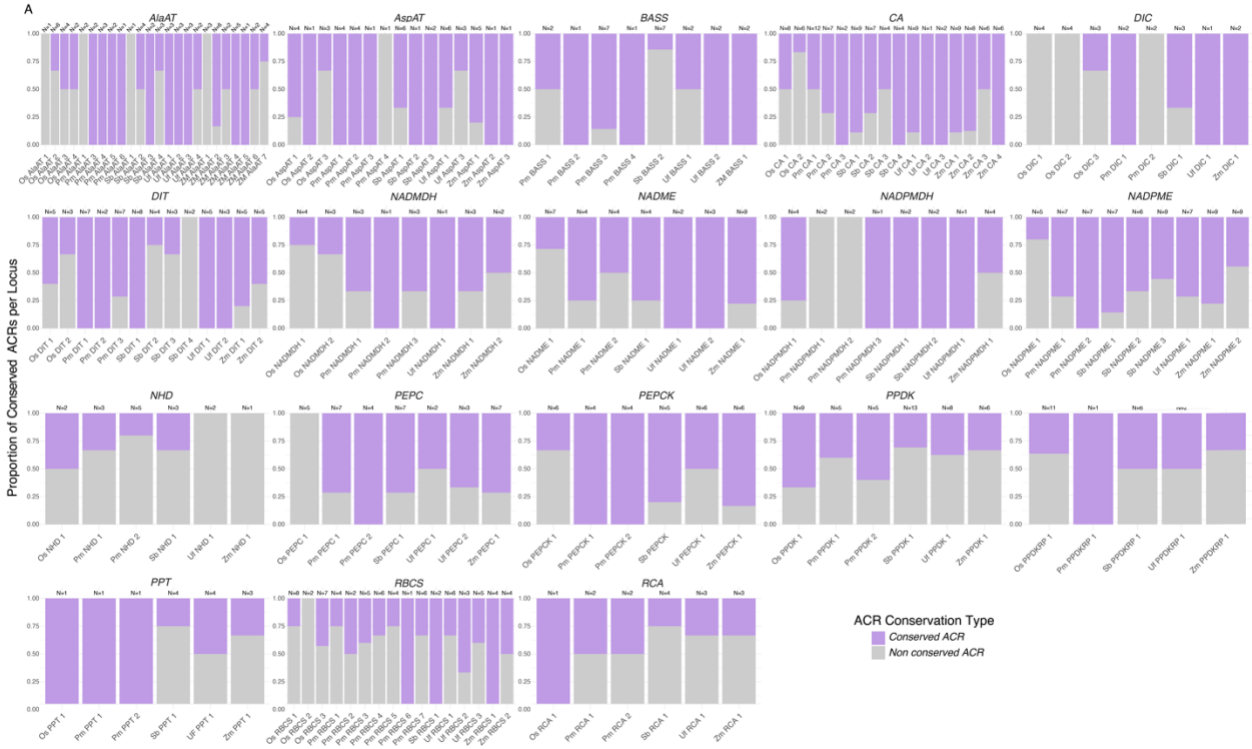
Supplemental Figure 21: **Phylogenies of all C4 genes used in this study. A**) For gene families which are large, gold boxes indicate the gene family members which were considered. Additionally, the gene which was used to identify the gene family is annotated with a star. For example, in the *NADP-ME*'s the *Z. mays* gene Zm00001eb121470 (*ME3*) was used to identify other gene family members in different species.



Supplemental Figure 22: **The *SSU1* locus between *Z. mays* and *S. bicolor shows a complex relationship of ACR usage.* A**) **Top**) The *SSU1* locus in *Z. mays*. With the top track showing gene models, the red track showing MS chromatin accessibility, and the blue track being BS chromatin accessibility. Both tracks are scaled identically to facilitate comparison. Bottom track shows ACRs, with cell-type-specific ACRs either being blue or red to match their cell-type specificity. Gold bars indicate regions with sequence homology. **Bottom**) The *SSU* locus in *S. bicolor*. With the top track showing gene models, here *SSU1* is indicated in green. The red track showing MS chromatin accessibility, and the bottom blue track being BS chromatin accessibility. Both tracks are scaled identically to facilitate comparison. Bottom track shows ACRs, with cell-type-specific ACRs either being blue or red to match their cell-type specificity.



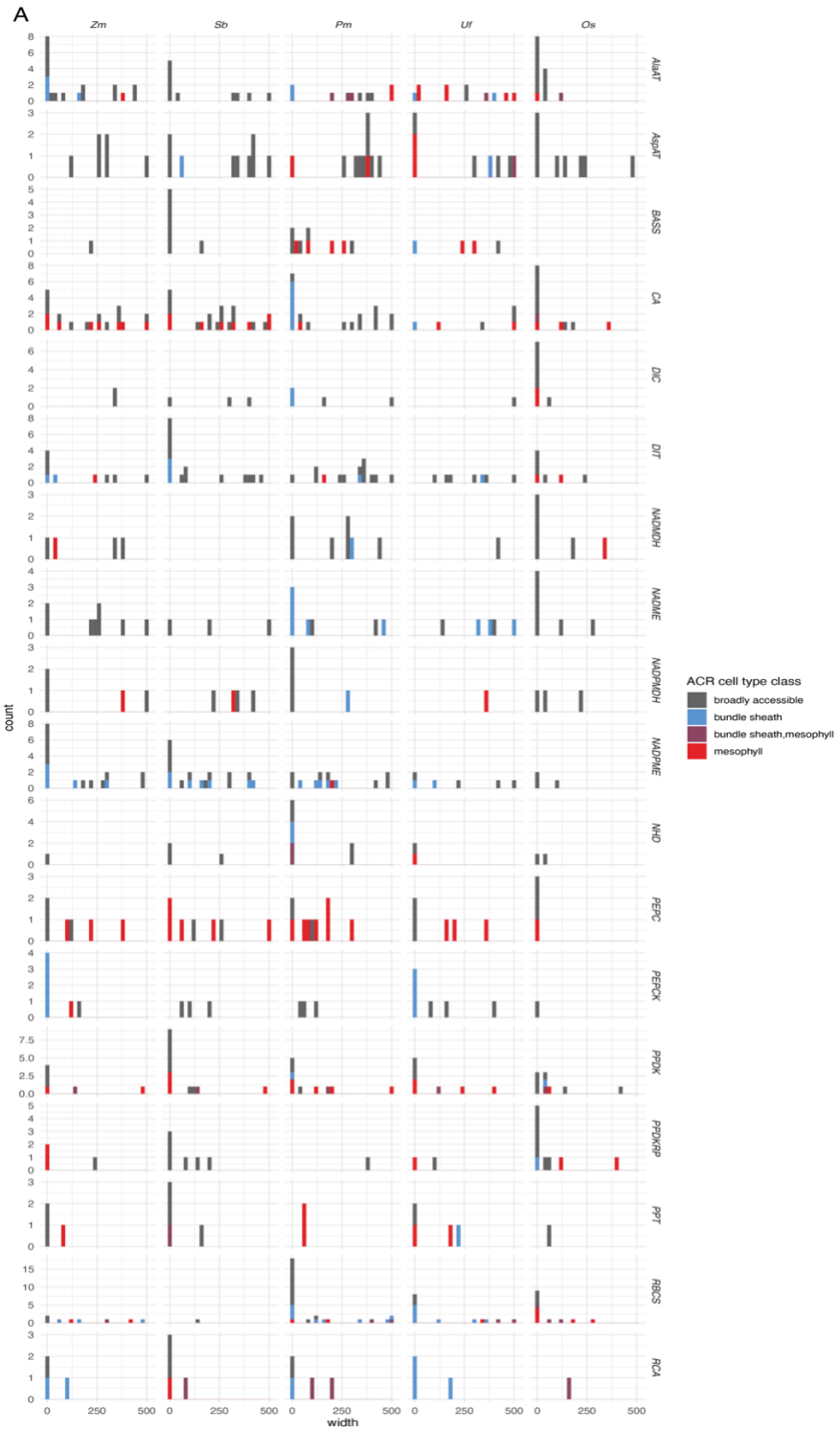
Supplemental Figure 23: **The number of ACRs found around key C4 enzymes and their distance to the corresponding locus. A**) Dot plots showing the number of broadly-accessible ACRs around each enzyme. The x-axis indicates which cell type these enzymes are found in. The y-axis is the number of ACRs. Enzymes are labeled. **B**) Dotplots showing the mean distance of broadly-accessible ACRs to their closest C4 enzyme. The x-axis indicates which cell type these enzymes are found in. The x-axis is the genomic distance to the C4 enzyme in question. If an enzyme had multiple broadly-accessible ACRs, the distance was averaged (mean).



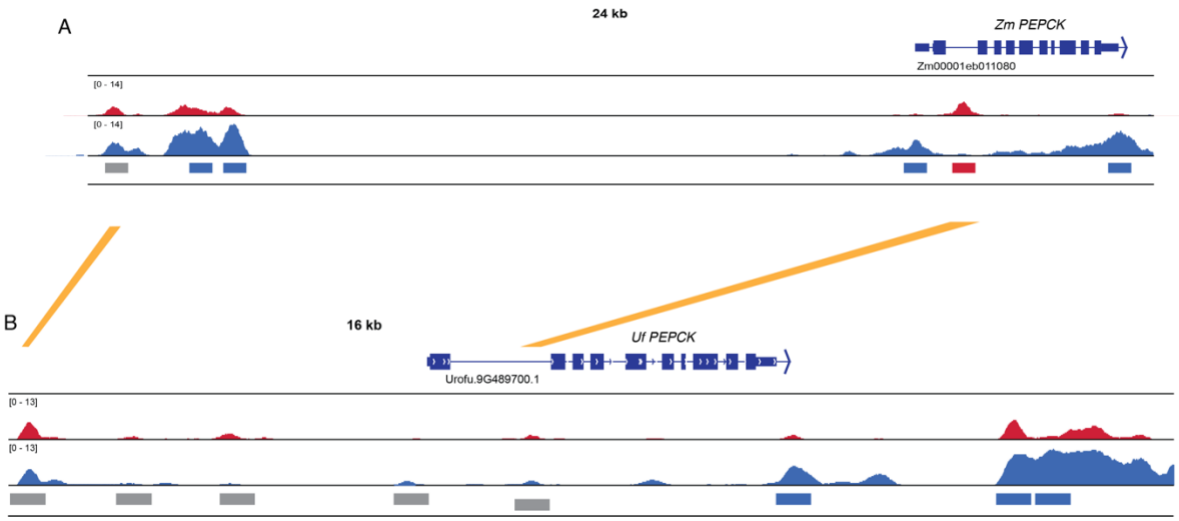
Supplemental Figure 24: **The proportion of all ACRs that are conserved or novel for all gene families. A**) Purple bars represent ACRs that have any sequence aligned to them from a different species, and gray represents ACRs where sequences are unalignable. The number of ACRs in each locus is labeled at the top of each column.

Supplemental Figure 25: **The proportion of all BS-specific ACRs that are conserved or novel for all gene families. A**) Dark blue bars represent ACRs that have any sequence aligned to them from a different species, and light represents ACRs where sequences are unalignable. The number of ACRs in each locus is labeled at the top of each column.

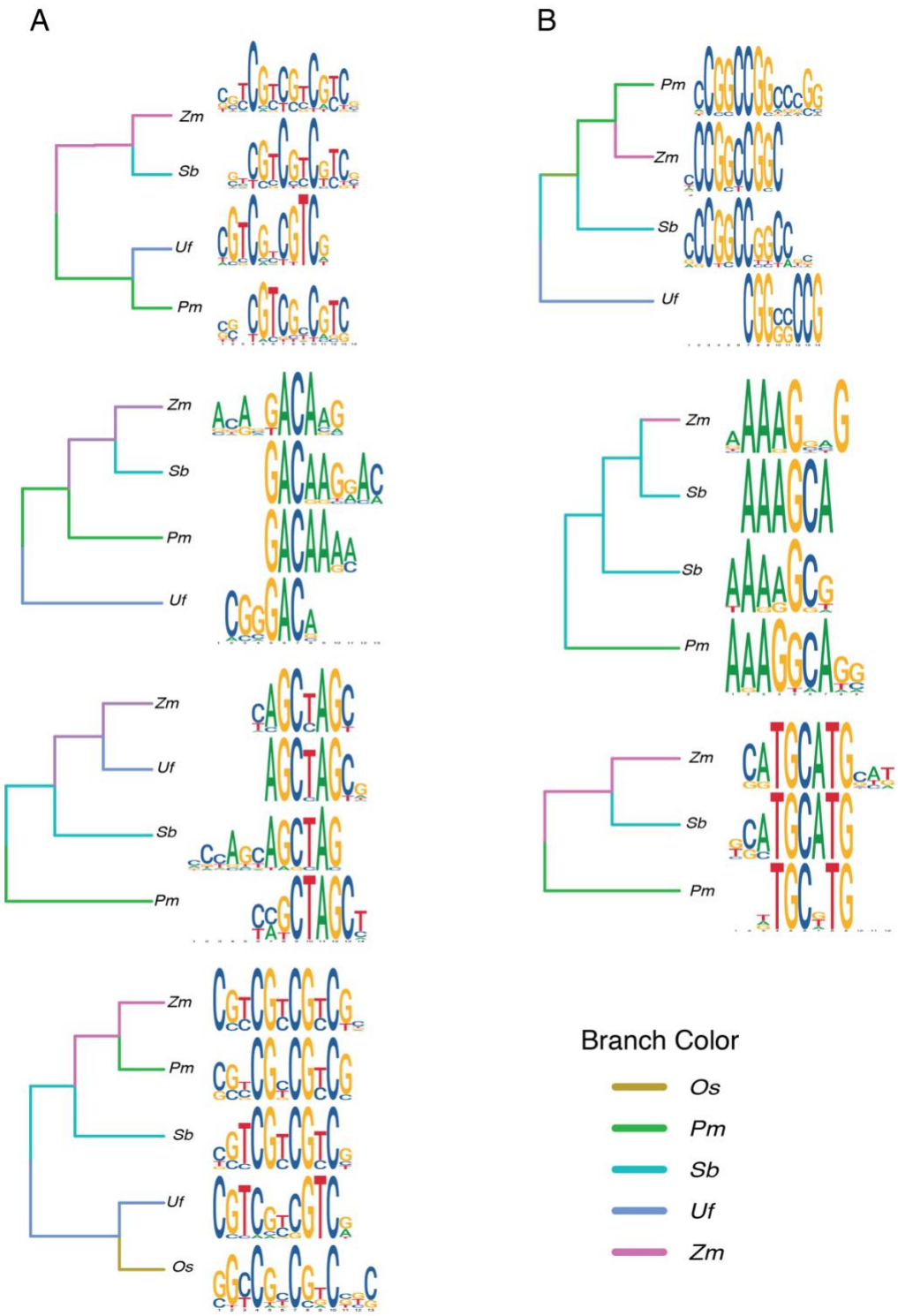
Supplemental Figure 26: **The proportion of all MS-specific ACRs which are conserved or novel for all C4 gene families**. **A**) Dark red bars represent ACRs that have any sequence aligned to them from a different species, and light red represents ACRs where sequences are unalignable. The number of ACRs in each locus is labeled at the top of each column.



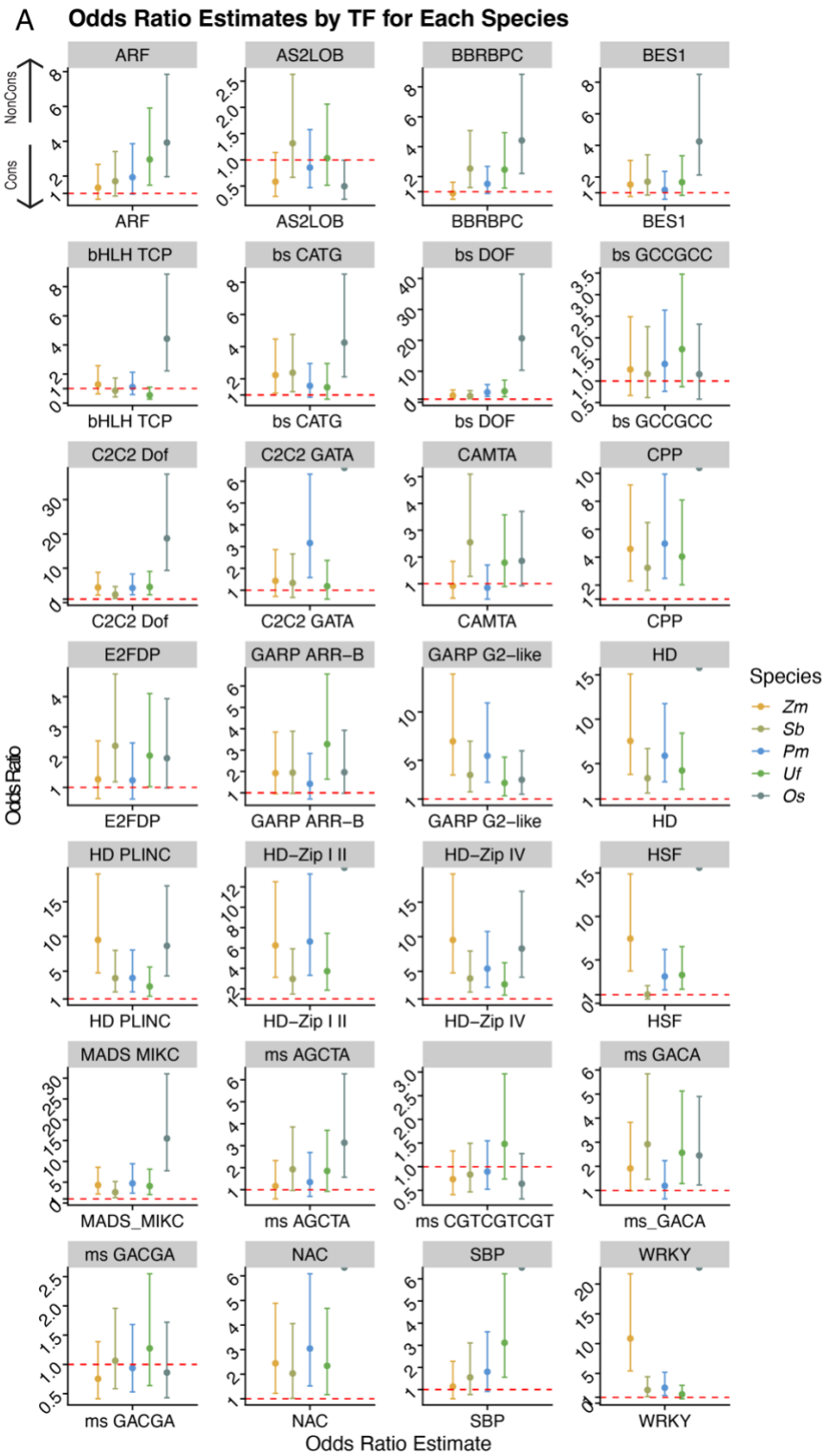
Supplemental Figure 27: **Histogram of sequence conservation of ACRs broken down by gene family, species, and type of ACR. A**) Columns correspond to the species *Zm-Os*, with each row indicating the gene family in question. Within each graph, the x-axis is the length of the given ACR which has homology to another ACR in a different species, and y-axis is the counts. ACRs are colored by their cell-type specificity seen in the corresponding species, with red being MS specific, blue being BS specific, purple being BS and MS specific, and gray being broadly accessible.



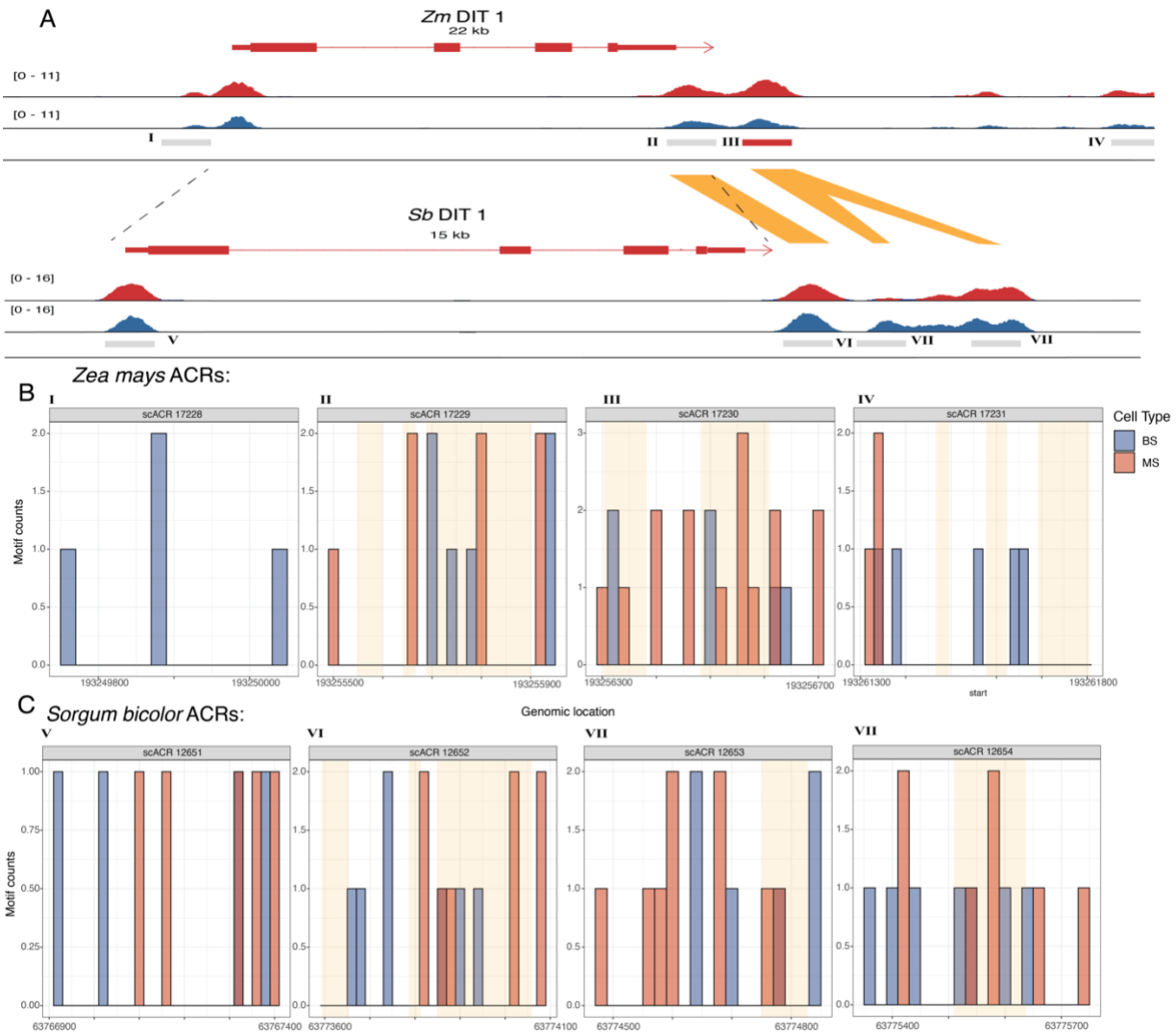
Supplemental Figure 28: **The *PEPCK* locus between *U. fusca* and *Z. mays* highlights a series of novel BS-specific ACRs. A)** Browser shot of the *PEPCK* locus in *Z. mays.* Red tracks correspond to MS chromatin accessibility and blue to BS chromatin accessibility. Tracks are scaled equally to facilitate comparison. Boxes represent ACRs, with cell-type-specific ACRs being colored identically to MS and BS chromatin accessibility tracks. Yellow bars indicate regions of sequence homology between the two loci. **B)** Browser shot of the *PEPCK* locus in *U. fusca.* Red tracks correspond to MS chromatin accessibility and blue to BS chromatin accessibility. Tracks are scaled equally to facilitate comparison. Boxes represent ACRs, with cell-type-specific ACRs being colored identically to MS and BS chromatin accessibility tracks. Yellow bars indicate regions of sequence homology between the two loci.



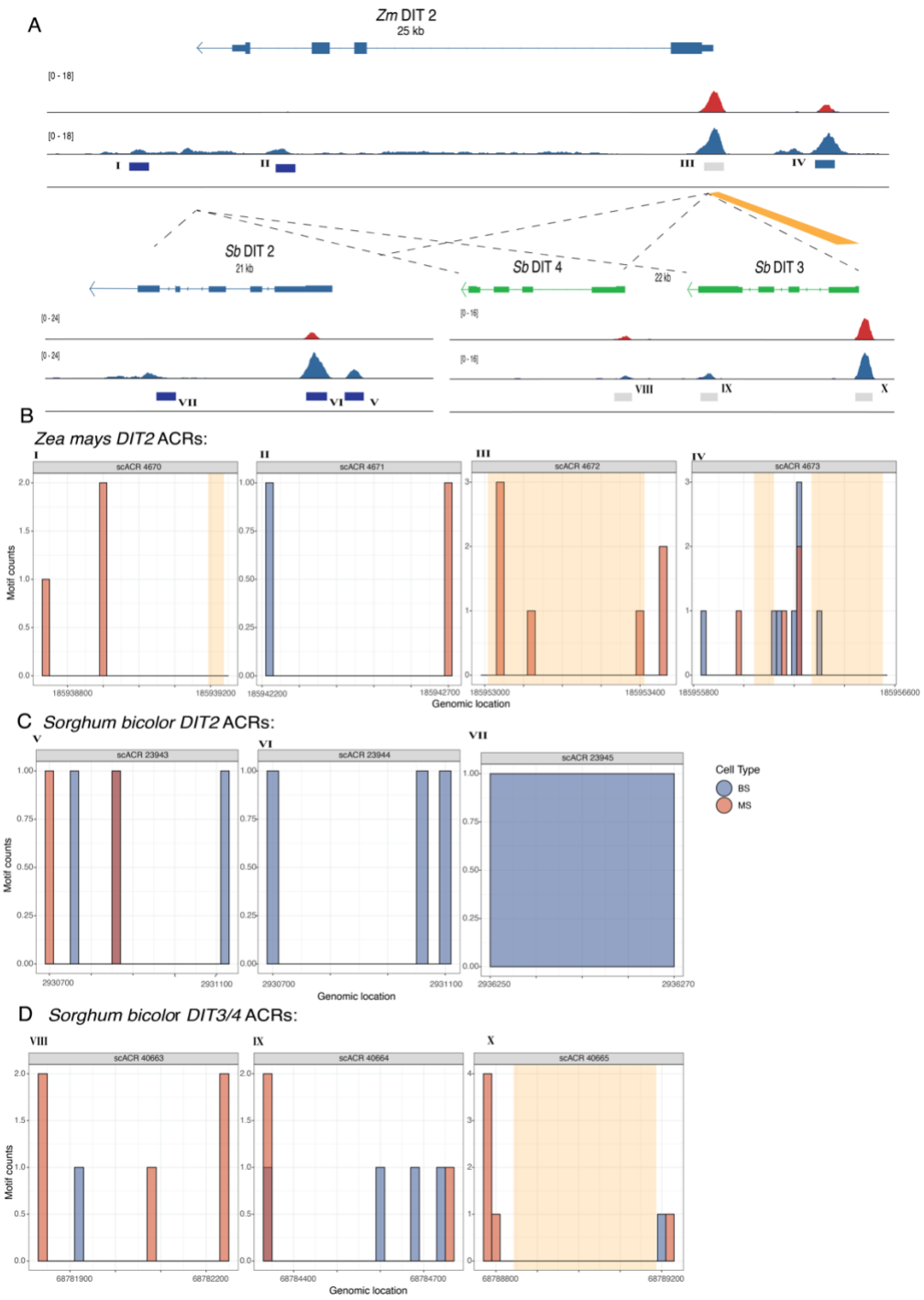
Supplemental Figure 29: **Identification of cell-type-specific *de novo* motifs for BS and MS across all species A**) The four MS specific *de novo* motifs identified in this study. Species are labeled by the tree color labels. **B**) The three BS-specific motifs identified in this study. Species are labeled by the tree color labels. Note, the selected presented in this figure are the motifs used to generate the merged figure, and are not all the motifs used for identification of the motif in question.



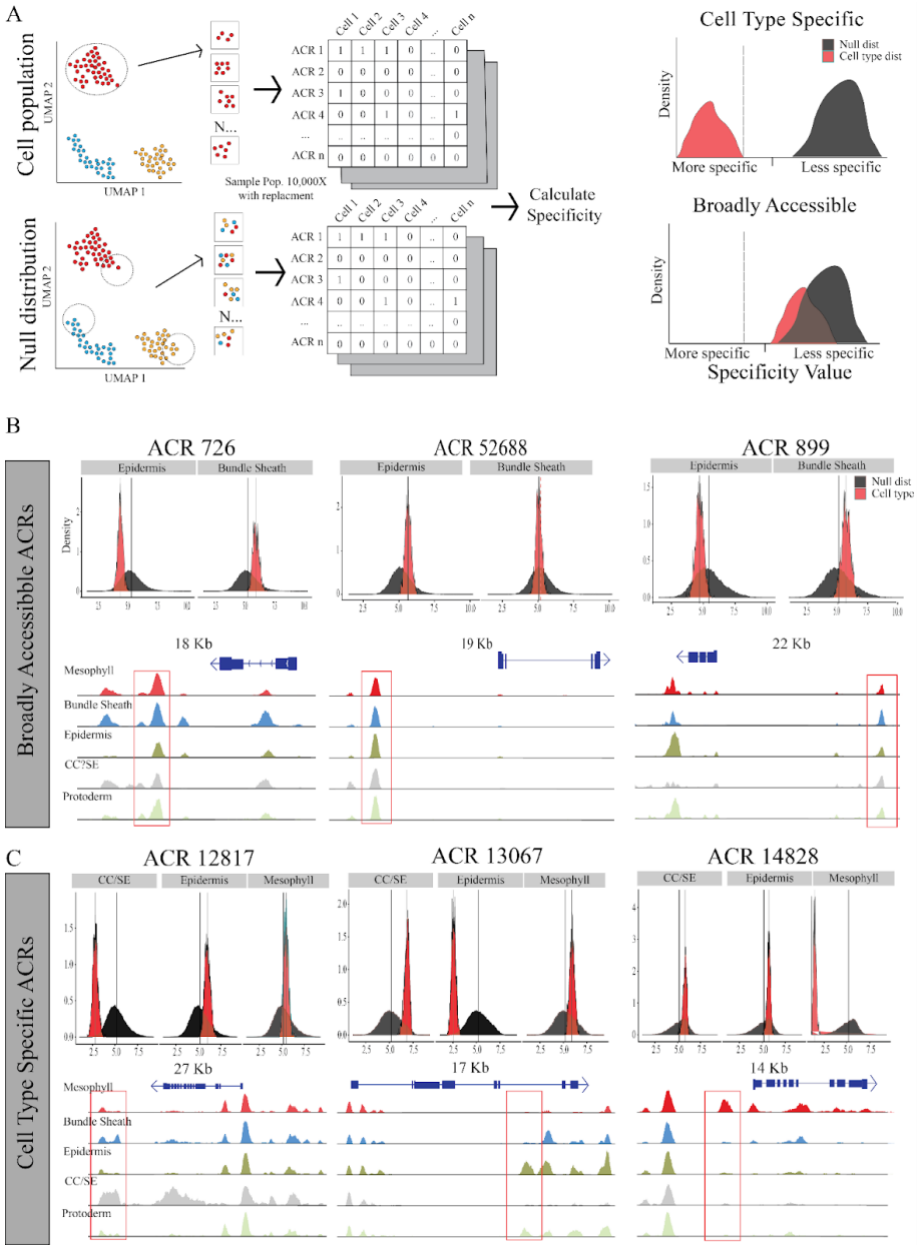
Supplemental Figure 30: **Transcription factor binding motifs show different levels of conservation in ACRs. A**)The Odds ratio of TF motifs when comparing their enrichment in conserved versus non-conserved regions. A higher odds ratio indicates that the motif is more often found in non-conserved regions within ACRs, whereas a lower odds ratio means the motif is in conserved regions. Motifs that are either MS or BS are the BS- and MS-specific motifs identified in this study.



Supplemental Figure 31: **A comparison of the *DIT1* locus ACRs between *Z. mays* and *S. bicolor.* A**) A screenshot of the *DIT1* locus, with MS-specific accessibility tracks being in red, and BS-specific accessibility tracks being in blue. Cell-type-specific ACRs are colored similarly, and gray ACRs are broadly accessible. Note the roman numeral for each ACR correspond to a single ACR graph below **B**) Histograms of TF motif presence in *Z. mays* depicting the number of cell-type-specific motifs found in 20 bp bins in each ACR. X-axis, the genomic location in question, y-axis the count of each cell-type-specific motif. Bars are colored by whether the motif is BS or MS specific. **C**) Histograms of motif presence in *S. bicolor* depicting the number of cell-type-specific motifs found in 20 bp bins in each ACR. X-axis, the genomic location in question, y-axis the count of each cell-type-specific motif. Bars are colored by whether the motif is BS or MS specific.



Supplemental Figure 32: **A comparison of the *DIT2* locus ACRs between *Z. mays* and *S. bicolor.* A**) A screenshot of the *DIT2* locus, with MS-specific accessibility tracks being in red, and BS-specific accessibility tracks being in blue. Cell-type-specific ACRs are colored similarly (red MS and blue BS), with gray ACRs are broadly accessible. Note the roman numeral for each ACR correspond to a single ACR graph below **B**) Histograms of motif presence in *Z. mays* depicting the number of cell-type-specific motifs found in 20 bp bins in each ACR. X-axis, the genomic location in question, y-axis the count of each cell-type-specific motif. Bars are colored by whether the motif is BS or MS specific. **C**) Histograms of motif presence in *S. bicolor DIT2* ACRs. **D**) Histograms of motif presence in *S. bicolor DIT3/4* ACRs.



Supplemental Figure 33: Identification of cell-type-specific ACR's using scATAC-seq data. A) Schematic of how cell-type-specific ACR identification works. In brief, this method uses a bootstrap based approach to generate a null set of cells to generate cell-type-specificity metrics. These metrics and how they're calculated can be seen in equation number 1. This null value is then compared to a real cell population that has also been resampled and the bootstrapped cell-type-specificity metrics are compared between each dataset using a non-parametric based approach. **B**) Examples of three ACRs which are categorized as cell-type specific. Underneath each ACR is two cell types, BS and epidermis, and density plots showing the null distribution as compared to the measured cell-type distribution. Each density plot consists of 10,000 measurements, with the median value of the real cell-type data being shown as a black line. Below the density plots are the screenshots, with cell types labeled, and the ACR in question being highlighted by a red box. **C**) Similar to B, except all ACRs in question have been measured as being cell-type specific. The ACRs in question are highlighted in red boxes. Note the differences in the red and black distributions in the density plots.