B73v4\_protein\_coding\_genes.gff: protein coding genes with AED and QI

\* MAKER\_updated is the latest version of the above file

\*\* Useful to compare also are v3\_models\_mapped\_to\_v4 and canonical\_v4\_models

est2genome\_FLC.gff: Full length cDNAs from genbank

est2genome\_GZT.gff: aligned transcripts from the v3 annotation (not used to generate the v4 annotations)

est2genome\_ISO.gff: Isoseq data has the long read data from 6 tissues. Unfortunately I was given the data in one big file so I don’t know which tissue each sequence is coming from.

est2genome\_MS.gff: Trinity assembled high depth seedling sequencing are polished alignments of assembled mRNA-seq data. Here is a URL for the original publication. [https://www.nature.com/articles/srep04519](https://email.cshl.edu/owa/redir.aspx?C=8LX3uB92wk0wJ1rg5U-dQS4tXIYY2u0CGwFAt5UhNzkyn10Y3-_UCA..&URL=https%3a%2f%2fwww.nature.com%2farticles%2fsrep04519)

est2genome\_TR.gff: 95 trinity assembled RNA-seq experiments (complexity reduced using cdhit) is a little more involved. We started with 95 mRNA-seq experiments that were publicly available in genbank. They were assembled indivisualy using trinity. This is described in this publicaiton [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4280997/](https://email.cshl.edu/owa/redir.aspx?C=5pEhZrelf0HH62IJkCCjtcXptexpgHLSvICyxr12kwcyn10Y3-_UCA..&URL=https%3a%2f%2fwww.ncbi.nlm.nih.gov%2fpmc%2farticles%2fPMC4280997%2f). given the large number of experiments there were a lot of redundant transcripts that took a lot of time to align. I used cdhit to filter out redundant transcripts. This file contains the polished alignments of the non-redundant transcripts.

protein2genome\_AT.gff: Arabidopsis proteins

protein2genome\_BD.gff: Brachypodium proteins

protein2genome\_GZP.gff: v3 proteins (not used to generate the v4 annotations)

protein2genome\_OS.gff: Rice proteins

protein2genome\_SB.gff: Sorghum proteins

protein2genome\_SI.gff: Seteria proteins

augustus\_masked.gff: Gene predictions from augustus

fgenesh\_masked.gff: Gene predictions from fgenesh