Annotation pipeline

1. Repeat Masking – *already performed for you!*

2. RNA, transcript, protein evidence

3. RNA-Seq cleanup – *already performed for you!*

4. MAKER, BRAKER

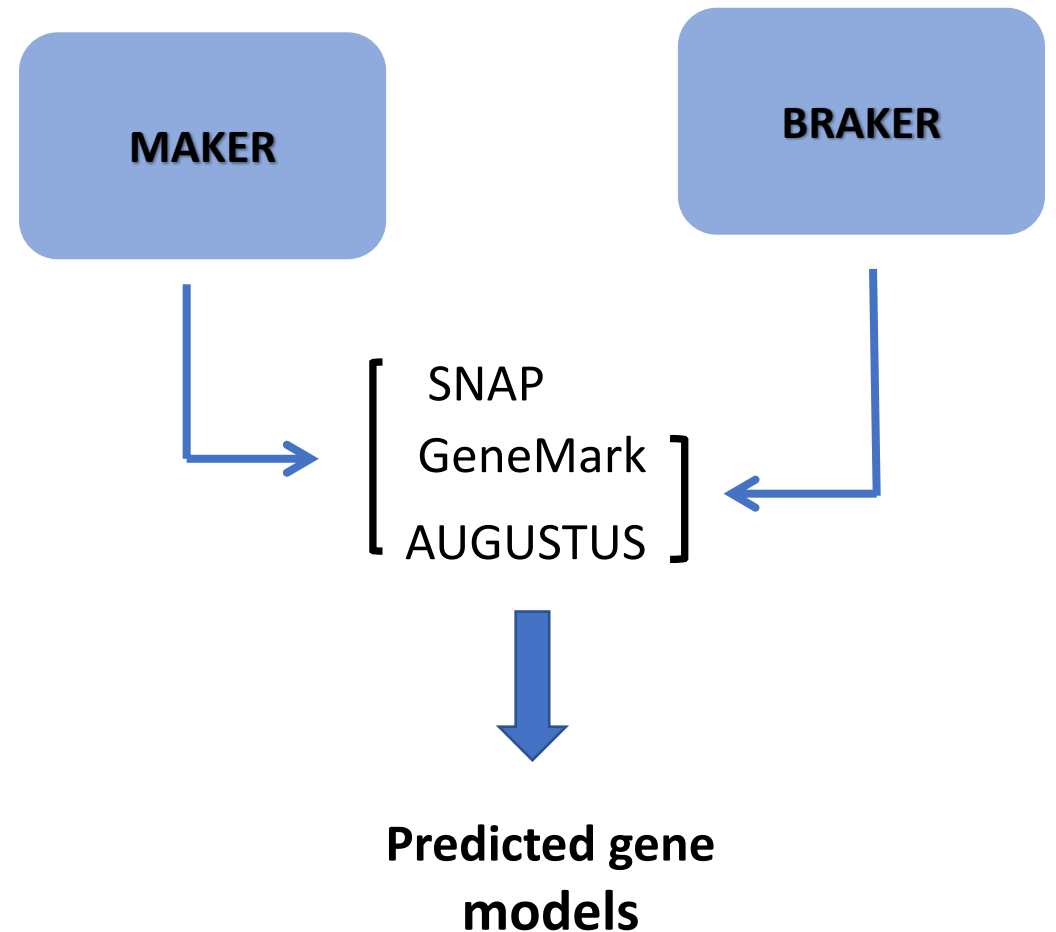5. Quality assessment [BUSCO, gFACs]

6. Training AUGUSTUS

Gene prediction tools

Scripts

1_hisat_pe_annot.sh
2_braker.sh
3_busco.sh
4_trainAugustus.sh

Output files

Braker_output
BUSCO_output
gFACs_output
AUGUSTUS_output

MAKER pipeline

1st Maker Run

(Transcripts, Proteins)

Gene model derived from MAKER run

Train AUGUSTUS

Train SNAP

2nd Maker Run

Control files:

- maker_exe.ctl: path for the underlying executables
- maker_bopt.ctl: stat for BLAST and Exonerate
- maker_opt.ctl: path for input genome files + training parameters

https://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Main_Page
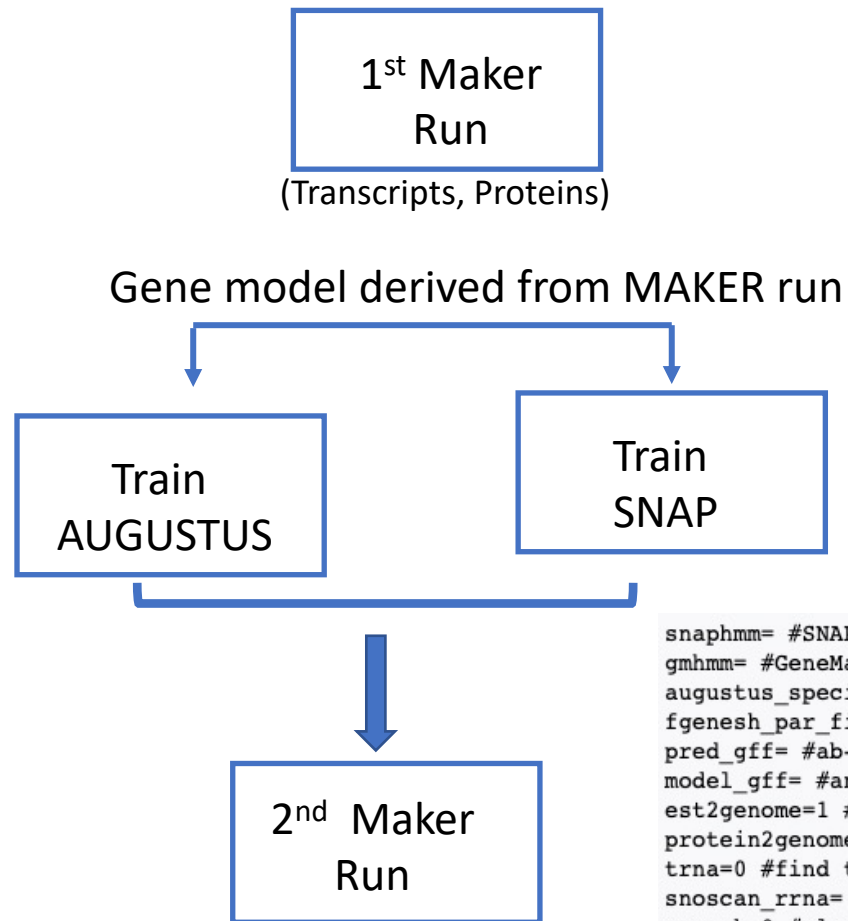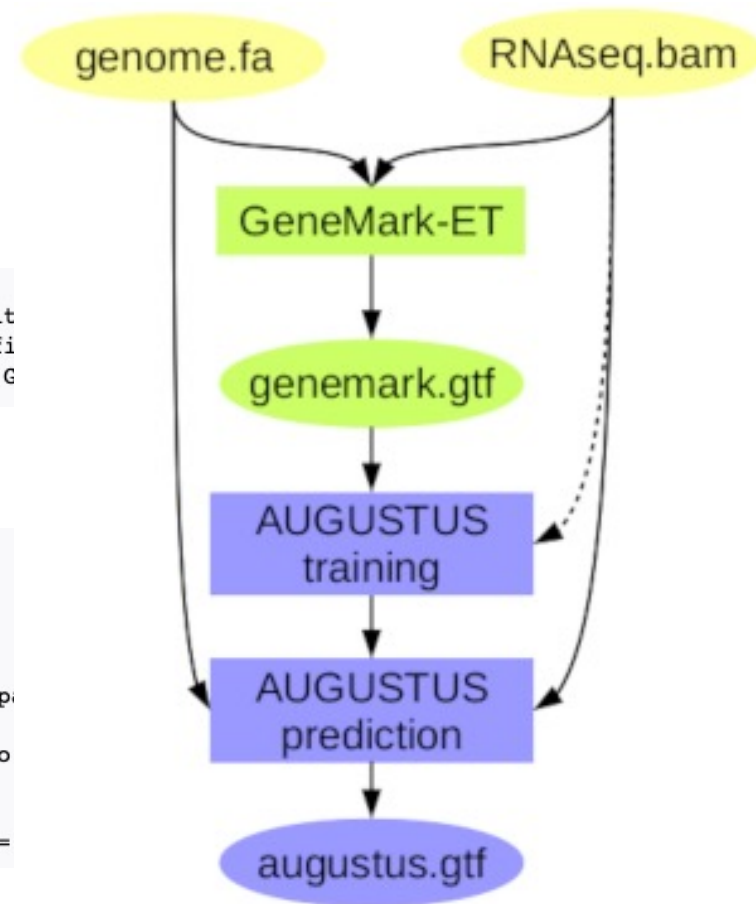
maker_opt.ctl:

```
est= #set of ESTs or assembled mRNA-seq in fasta format
altest= #EST/cDNA sequence file in fasta format from an alt
est_gff= #aligned ESTs or mRNA-seq from an external GFF3 fi
altest_gff= #aligned ESTs from a closly relate species in G
```

```
snaphmm= #SNAP HMM file
gmhmm= #GeneMark HMM file
augustus_species= #Augustus gene prediction species model
fgenesh_par_file= #FGENESH parameter file
pred_gff= #ab-initio predictions from an external GFF3 file
model_gff= #annotated gene models from an external GFF3 file (annotation pa
est2genome=1 #infer gene predictions directly from ESTs, 1 = yes, 0 = no
protein2genome=0 #infer predictions from protein homology, 1 = yes, 0 = no
trna=0 #find tRNAs with tRNAscan, 1 = yes, 0 = no
snoscan_rrna= #rRNA file to have Snoscan find snoRNAs
unmask=0 #also run ab-initio prediction programs on unmasked sequence, 1 =
```

BRAKER with RNA-Seq reads

https://github.com/Gaius-Augustus/BRAKER