**Appendix S1**. Example MAKER2 control files. Instructions for generating these files can be found in Appendix S6.

maker\_bopts.ctl:

#-----BLAST and Exonerate Statistics Thresholds

blast\_type=ncbi+ #set to 'wublast' or 'ncbi'

pcov\_blastn=0.8 #Blastn Percent Coverage Threhold EST-Genome Alignments

pid\_blastn=0.85 #Blastn Percent Identity Threshold EST-Genome Aligments

eval\_blastn=1e-10 #Blastn eval cutoff

bit\_blastn=40 #Blastn bit cutoff

pcov\_blastx=0.5 #Blastx Percent Coverage Threhold Protein-Genome Alignments

pid\_blastx=0.4 #Blastx Percent Identity Threshold Protein-Genome Aligments

eval\_blastx=1e-06 #Blastx eval cutoff

bit\_blastx=30 #Blastx bit cutoff

pcov\_rm\_blastx=0.5 #Blastx Percent Coverage Threhold For Transposable Element Masking

pid\_rm\_blastx=0.4 #Blastx Percent Identity Threshold For Transposbale Element Masking

eval\_rm\_blastx=1e-06 #Blastx eval cutoff for transposable element masking

bit\_rm\_blastx=30 #Blastx bit cutoff for transposable element masking

pcov\_tblastx=0.8 #tBlastx Percent Coverage Threhold alt-EST-Genome Alignments

pid\_tblastx=0.85 #tBlastx Percent Identity Threshold alt-EST-Genome Aligments

eval\_tblastx=1e-10 #tBlastx eval cutoff

bit\_tblastx=40 #tBlastx bit cutoff

eva\_pcov\_blastn=0.8 #EVALUATOR Blastn Percent Coverage Threshold EST-Genome Alignments

eva\_pid\_blastn=0.85 #EVALUATOR Blastn Percent Identity Threshold EST-Genome Alignments

eva\_eval\_blastn=1e-10 #EVALUATOR Blastn eval cutoff

eva\_bit\_blastn=40 #EVALUATOR Blastn bit cutoff

ep\_score\_limit=20 #Exonerate protein percent of maximal score threshold

en\_score\_limit=20 #Exonerate nucleotide percent of maximal score threshold

maker\_opts.ctl:

#-----Genome (Required for De-Novo Annotation)

genome=/home/Wolfelab/Desktop/pen\_cyan\_contigs.fasta.masked #genome sequence file in fasta format

organism\_type=eukaryotic #eukaryotic or prokaryotic. Default is eukaryotic

#-----Re-annotation Using MAKER Derived GFF3

genome\_gff= #re-annotate genome based on this gff3 file

est\_pass=0 #use ests in genome\_gff: 1 = yes, 0 = no

altest\_pass=0 #use alternate organism ests in genome\_gff: 1 = yes, 0 = no

protein\_pass=0 #use proteins in genome\_gff: 1 = yes, 0 = no

rm\_pass=0 #use repeats in genome\_gff: 1 = yes, 0 = no

model\_pass=0 #use gene models in genome\_gff: 1 = yes, 0 = no

pred\_pass=0 #use ab-initio predictions in genome\_gff: 1 = yes, 0 = no

other\_pass=0 #passthrough everything else in genome\_gff: 1 = yes, 0 = no

#-----EST Evidence (for best results provide a file for at least one)

est=/home/Wolfelab/Desktop/MakerEST.fasta #non-redundant set of assembled ESTs in fasta format (classic EST analysis)

est\_reads= #unassembled nextgen mRNASeq in fasta format (not fully implemented)

altest= #EST/cDNA sequence file in fasta format from an alternate organism

est\_gff= #EST evidence from an external gff3 file

altest\_gff= #Alternate organism EST evidence from a separate gff3 file

#-----Protein Homology Evidence (for best results provide a file for at least one)

protein=/home/Wolfelab/Desktop/MakerProtein.fasta #protein sequence file in fasta format

protein\_gff= #protein homology evidence from an external gff3 file

#-----Repeat Masking (leave values blank to skip repeat masking)

model\_org= #select a model organism for RepBase masking in RepeatMasker

rmlib= #provide an organism specific repeat library in fasta format for RepeatMasker

repeat\_protein= #provide a fasta file of transposable element proteins for RepeatRunner

rm\_gff= #repeat elements from an external GFF3 file

prok\_rm=0 #forces MAKER to run repeat masking on prokaryotes (don't change this), 1 = yes, 0 = no

#-----Gene Prediction

snaphmm=/home/Wolfelab/Desktop/Snap-training2/run6/snap/Penstemon.hmm #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 pass-through)

est2genome=1 #infer gene predictions directly from ESTs, 1 = yes, 0 = no

protein2genome=0 #gene prediction from protein homology (prokaryotes only), 1 = yes, 0 = no

unmask=0 #Also run ab-initio prediction programs on unmasked sequence, 1 = yes, 0 = no

#-----Other Annotation Feature Types (features MAKER doesn't recognize)

other\_gff= #features to pass-through to final output from an extenal GFF3 file

#-----External Application Behavior Options

alt\_peptide=C #amino acid used to replace non standard amino acids in BLAST databases

cpus=1 #max number of cpus to use in BLAST and RepeatMasker (not for MPI, leave 1 when using MPI)

#-----MAKER Behavior Options

max\_dna\_len=100000 #length for dividing up contigs into chunks (increases/decreases memory usage)

min\_contig=1 #skip genome contigs below this length (under 10kb are often useless)

pred\_flank=200 #flank for extending evidence clusters sent to gene predictors

AED\_threshold=1 #Maximum Annotation Edit Distance allowed (bound by 0 and 1)

min\_protein=0 #require at least this many amino acids in predicted proteins

alt\_splice=0 #Take extra steps to try and find alternative splicing, 1 = yes, 0 = no

always\_complete=0 #force start and stop codon into every gene, 1 = yes, 0 = no

map\_forward=0 #map names and attributes forward from old GFF3 genes, 1 = yes, 0 = no

keep\_preds=0 #Add unsupported gene prediction to final annotation set, 1 = yes, 0 = no

split\_hit=10000 #length for the splitting of hits (expected max intron size for evidence alignments)

softmask=1 #use soft-masked rather than hard-masked (seg filtering for wublast)

single\_exon=0 #consider single exon EST evidence when generating annotations, 1 = yes, 0 = no

single\_length=250 #min length required for single exon ESTs if 'single\_exon is enabled'

retry=1 #number of times to retry a contig if there is a failure for some reason

clean\_try=0 #remove all data from previous run before retrying, 1 = yes, 0 = no

clean\_up=0 #removes theVoid directory with individual analysis files, 1 = yes, 0 = no

TMP= #specify a directory other than the system default temporary directory for temporary files

#-----EVALUATOR Control Options

evaluate=0 #run EVALUATOR on all annotations (very experimental), 1 = yes, 0 = no

side\_thre=5

eva\_window\_size=70

eva\_split\_hit=1

eva\_hspmax=100

eva\_gspmax=100

enable\_fathom=0

maker\_exe.ctl:

#-----Location of Executables Used by MAKER/EVALUATOR

makeblastdb=/home/Wolfelab/Desktop/ncbi-blast-2.2.27+/bin/makeblastdb #location of NCBI+ makeblastdb executable

blastn=/home/Wolfelab/Desktop/ncbi-blast-2.2.27+/bin/blastn #location of NCBI+ blastn executable

blastx=/home/Wolfelab/Desktop/ncbi-blast-2.2.27+/bin/blastx #location of NCBI+ blastx executable

tblastx=/home/Wolfelab/Desktop/ncbi-blast-2.2.27+/bin/tblastx #location of NCBI+ tblastx executable

formatdb=/home/Wolfelab/Desktop/blast-2.2.26/bin/formatdb #location of NCBI formatdb executable

blastall=/home/Wolfelab/Desktop/blast-2.2.26/bin/blastall #location of NCBI blastall executable

xdformat= #location of WUBLAST xdformat executable

blasta= #location of WUBLAST blasta executable

RepeatMasker=/home/Wolfelab/Desktop/RepeatMasker/RepeatMasker #location of RepeatMasker executable

exonerate=/home/Wolfelab/Desktop/exonerate-2.2.0-x86\_64/bin/exonerate #location of exonerate executable

#-----Ab-initio Gene Prediction Algorithms

snap=/home/Wolfelab/Desktop/snap/snap #location of snap executable

gmhmme3= #location of eukaryotic genemark executable

gmhmmp= #location of prokaryotic genemark executable

augustus=/home/Wolfelab/Desktop/maker/bin/../exe/augustus.2.5.5/bin/augustus #location of augustus executable

fgenesh= #location of fgenesh executable

#-----Other Algorithms

fathom=/home/Wolfelab/Desktop/snap/fathom #location of fathom executable (experimental)

probuild= #location of probuild executable (required for genemark)