

PEACH GENOME STRUCTURAL RE-ANNOTATION RESULTS

Curated BIO BigData Platform

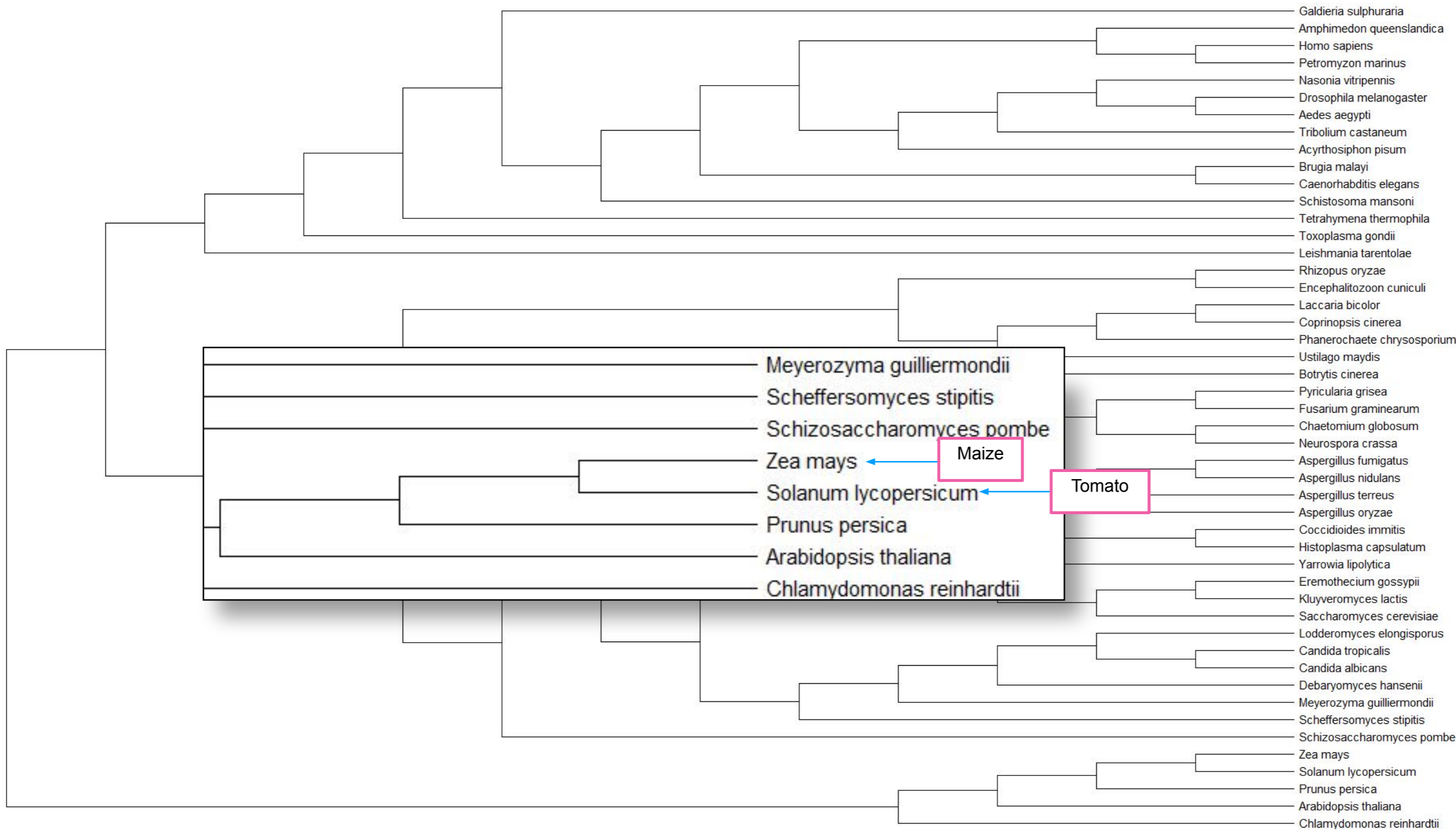
3BIGS

Scientific name: *Prunus persica* (peach)

Files used for analysis:

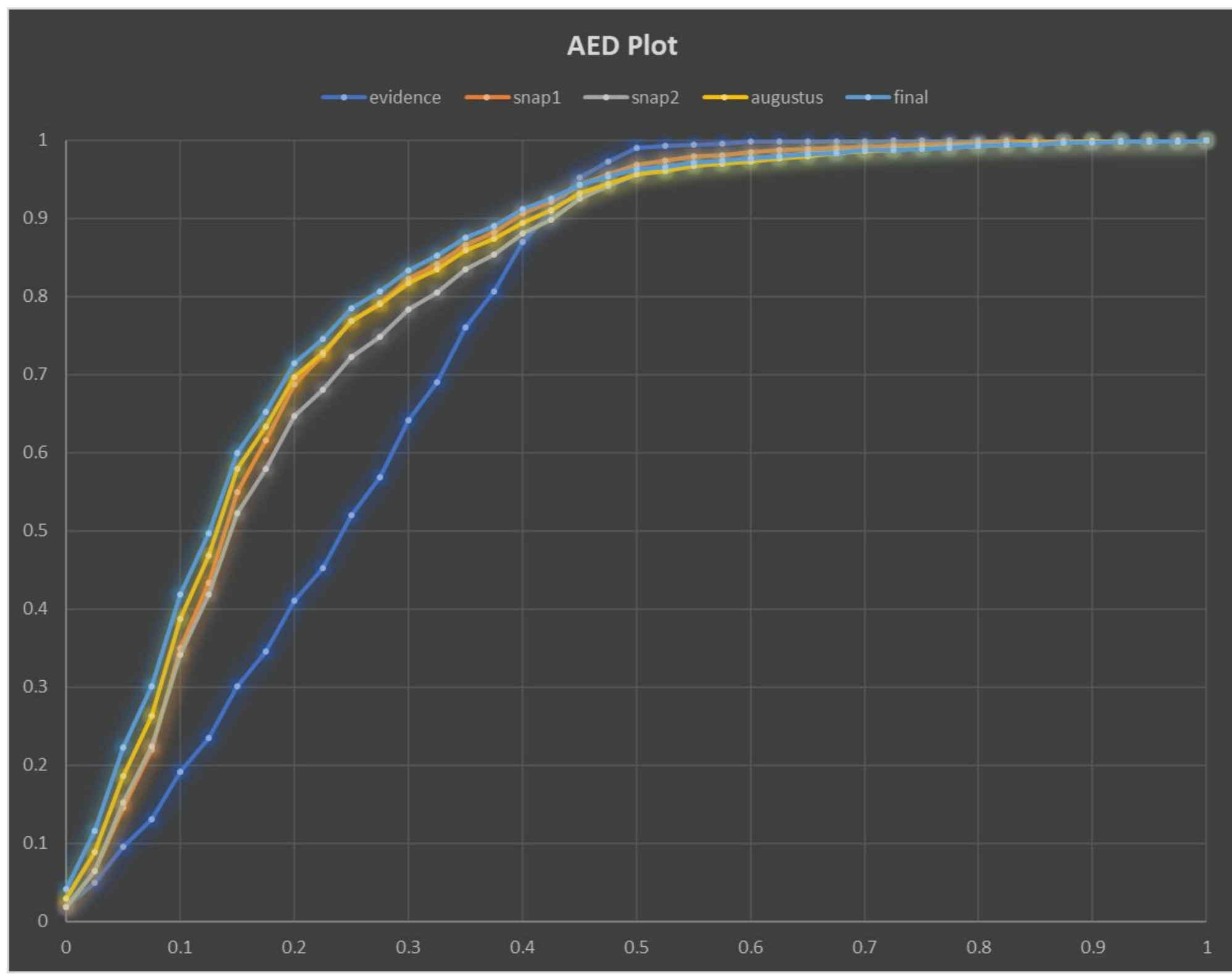
1. Genome file - 1.20181011.Peach.falcon-unzip.primary.errd.hic.fasta.gz
2. RNA evidence - collapse_isoform.collapsed.gff3 (from pacbio) + Public RNAseq data – peach cultivars
3. EST evidence - 3.peach_est_ncbi.fasta
4. Homology evidence - 4.homology.zip (Arabidopsis, Grape, Rice, Soybean, Swissprot_euk_proteins)

Reference used for gene prediction by Augustus - *Solanum Lycopersicon* (Tomato)



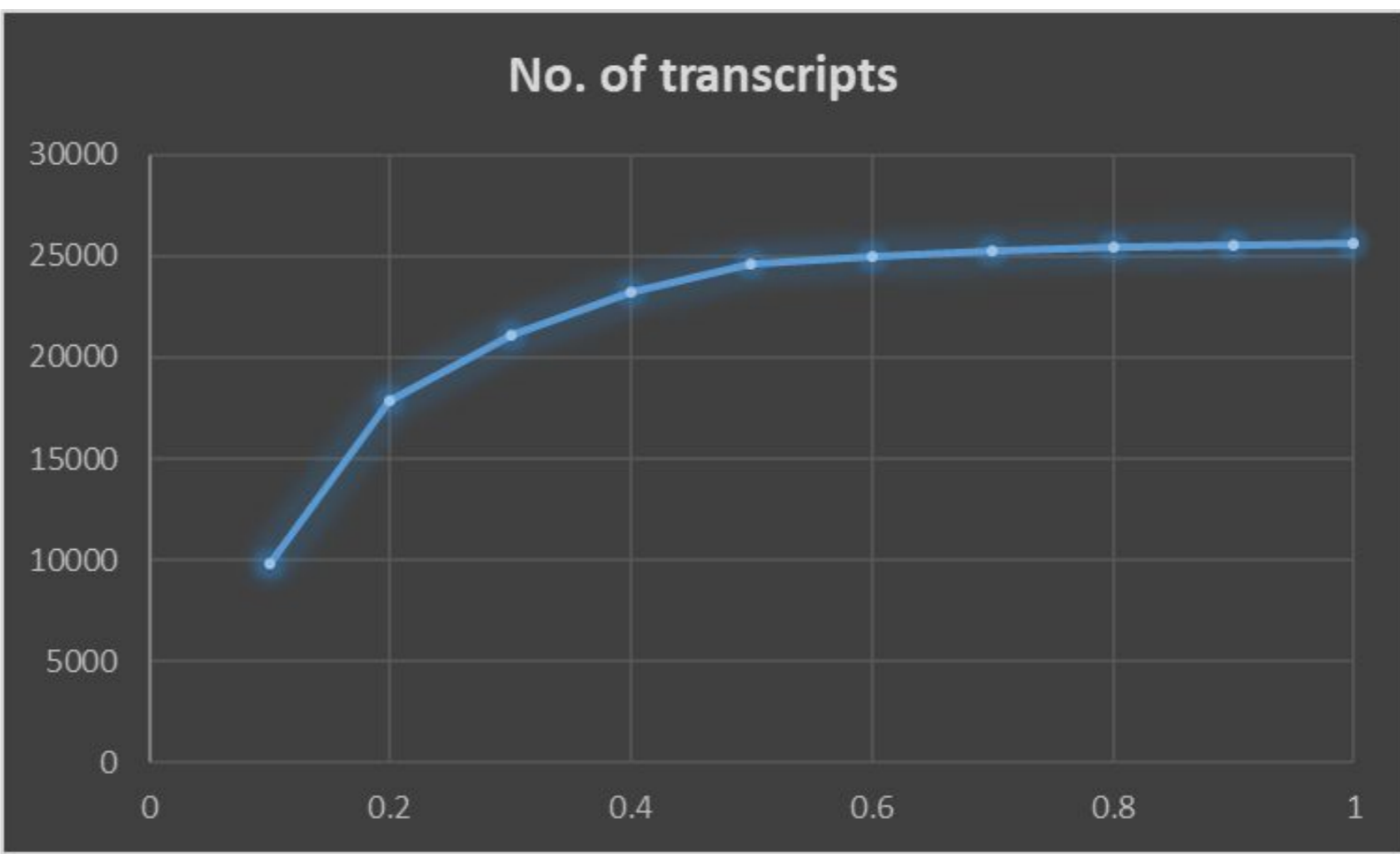
File size	Number of processor (Run type)	Time (hrs)	Status	Remarks
Peach genome (243 M)	36 --cpus 2 (basic_header_corrected_splicing)	6.5 hours	Completed	Genes:20271, Transcripts:58961, Proteins:58961
	36 --cpus 2 (SNAP 1)	6.5 hours	Completed	Genes:18737, Transcripts:19361, Proteins:19361
	36 --cpus 2 (SNAP 2)	6.5 hours	Completed	Genes:19105 , Transcripts:20123, Proteins:20123
	36 --cpus 2 (Augustus - tomato)	7.4 hours	Completed	Genes:19389, Transcripts:20205, Proteins:20205
	36 --cpus 2 (Final Genemark)	7.6 hours	Completed	Genes:21468, Transcripts:25619, Proteins:25619

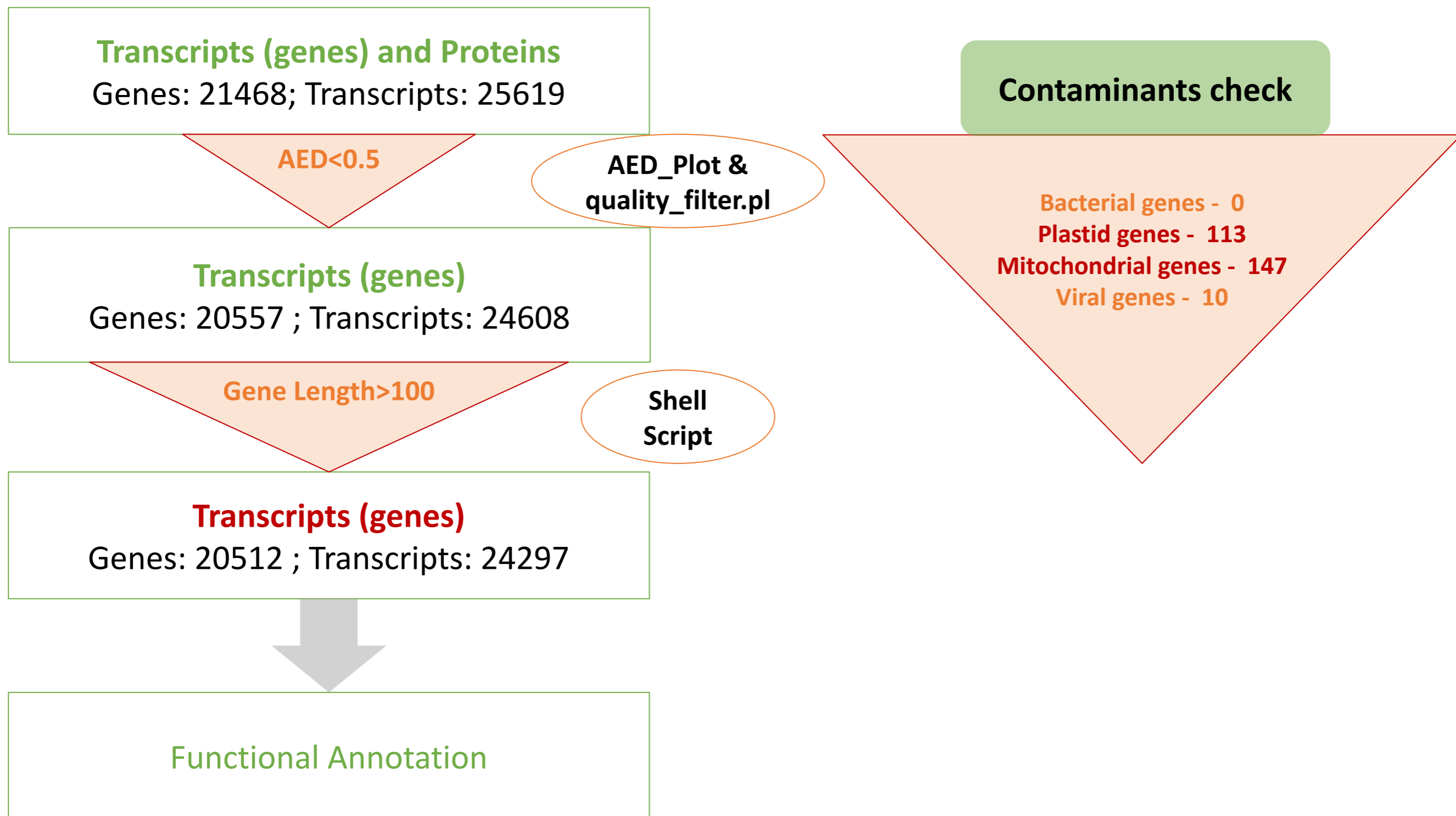
AED Plot



Transcript count and plot at different AED levels

AED (<)	No. of transcripts
0.1	9800
0.2	17853
0.3	21130
0.4	23185
0.5	24608
0.6	25002
0.7	25258
0.8	25427
0.9	25545
1	25609





Feature Type	Count
Total sequence length	241045119
Number of genes	20512
Number of mRNAs	24297
Number of exons	135544
Number of introns	111247
Number of CDS	24297
Overlapping genes	330
Contained genes	87
CDS: complete	0
CDS: start, no stop	0
CDS: stop, no start	0
CDS: no stop, no start	24297
Total gene length	71471553
Total mRNA length	76394457
Total exon length	33157090
Total intron length	43446391
Total CDS length	31510749
Shortest gene	102
Shortest mRNA	102

Feature Type	Count
Shortest exon	2
Shortest intron	4
Shortest CDS	12
Longest gene	81684
Longest mRNA	59400
Longest exon	6583
Longest intron	37689
Longest CDS	16290
mean gene length	3484
mean mRNA length	3144
mean exon length	245
mean intron length	391
mean CDS length	1297
% of genome covered by genes	29.7
% of genome covered by CDS	13.1
mean mRNAs per gene	1
mean exons per mRNA	6
mean introns per mRNA	5

Type	Count
DNA	72514
LINE	8092
Low_complexity	75002
LTR	97408
RC	8076
rRNA	2486
Satellite	528
Simple_repeat	381140
SINE	4818
snRNA	116
tRNA	806
Unknown	108370

Name	File Location Path
Final GFF3	/home/ec2-user/peach_reannot/genemark_results/peach_genome_corrected_genemark_contrem_final_filtered.gff
Evidence	/home/ec2-user/peach_reannot/evidence_header_corrected/peach_genome_corrected_evidence_final_filtered.gff
SNAP	/home/ec2-user/peach_reannot/snap2_results/peach_genome_corrected_snap2_final_filtered.gff
AUGUSTUS	/home/ec2-user/peach_reannot/augustus_results/peach_genome_header_corrected.all_augustus.gff
AUGUSTUS_SNAP	/home/ec2-user/peach_reannot/augustus_results/peach_genome_corrected_augustus_final_filtered.gff
Protein file	/home/ec2-user/peach_reannot/genemark_results/peach_genome.all_AED_lt_0.5_gene_100_proteins_contrem.fasta
Transcript file	/home/ec2-user/peach_reannot/genemark_results/peach_genome.all_AED_lt_0.5_gene_100_transcripts_contrem.fasta
Repeatmasker file	/home/ec2-user/peach_reannot/genemark_results/peach_genome_header_corrected.all_repeatmasker.gff

S3 path: s3://3bigsgap/peach_reannot_results/peach_final_files/

Peach Genome Post Structural Annotation Results

Type	Count
tRNAscan output file	778

S3 Path:

tRNA sequences fasta: s3://3bigsgap/peach_reannot_results/tRNAscan_results/trna_output.fasta
tRNAscan output file: s3://3bigsgap/ peach_reannot_results/tRNAscan_results/trna_output

S3 File location:

Raw file:

s3://3bigsgap/peach_reannot_results/snoscan_results/peach_genome_header_corrected.fasta.raw

Sorted files:

s3://3bigsgap/peach_reannot_results/snoscan_results/peach_genome_header_corrected.fasta.sortRH
s3://3bigsgap/peach_reannot_results/snoscan_results/peach_genome_header_corrected.fasta.sort-all
s3://3bigsgap/peach_reannot_results/snoscan_results/peach_genome_header_corrected.fasta.sort-by
site

peach_genome_name_corrected.fasta.sort-all - 17419 snoRNA hits over 20 bits

Type	Count
Phase 1 pseudogenes	24778

S3 Path:**Pseudogenes fasta:**

s3://3bigsgap/peach_reannot_results/pseudogene_results/pseudogenes_with_confidence.fasta

Pseudogenes alignments:

s3://3bigsgap/peach_reannot_results/pseudogene_results/pseudogenes_align_details.fasta

Functional Annotation - MAKER Integration

DB	Server1 Location (EC2)	S3 Location
swissprot	/home/ec2-user/peach_reannot/genemark_results/fun_ann/swissprot_Maker.gff	s3://3bigsgap/peach_reannot_results/fn_annot_results/
	/home/ec2-user/peach_reannot/genemark_results/fun_ann/swissprot_Maker_protein.fasta	s3://3bigsgap/peach_reannot_results/fn_annot_results/
	/home/ec2-user/peach_reannot/genemark_results/fun_ann/swissprot_maker_transcript.fasta	s3://3bigsgap/peach_reannot_results/fn_annot_results/
interpro	/home/ec2-user/peach_reannot/genemark_results/fun_ann/Interpro_swissprot_Maker.gff	s3://3bigsgap/peach_reannot_results/fn_annot_results/
	/home/ec2-user/peach_reannot/genemark_results/fun_ann/Interpro2gff3_Maker.gff	s3://3bigsgap/peach_reannot_results/fn_annot_results/
nr	/home/ec2-user/peach_reannot/genemark_results/fun_ann/nr_fasta_fixed.gff	s3://3bigsgap/peach_reannot_results/fn_annot_results/
	/home/ec2-user/peach_reannot/genemark_results/fun_ann/nr_Maker_proteins_final.fasta	s3://3bigsgap/peach_reannot_results/fn_annot_results/
	/home/ec2-user/peach_reannot/genemark_results/fun_ann/nr_Maker_transcripts_final.fasta	s3://3bigsgap/peach_reannot_results/fn_annot_results/

S3 File location: s3://3bigsgap/peach_reannot_results/fn_annot_results/

Curated BIO BigData Platform

3BIGS

core@3bigs.com / 031-8064-1877

3BIGS CO.,LTD. 1403, 156, Gwanggyo-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16506, Republic of Korea