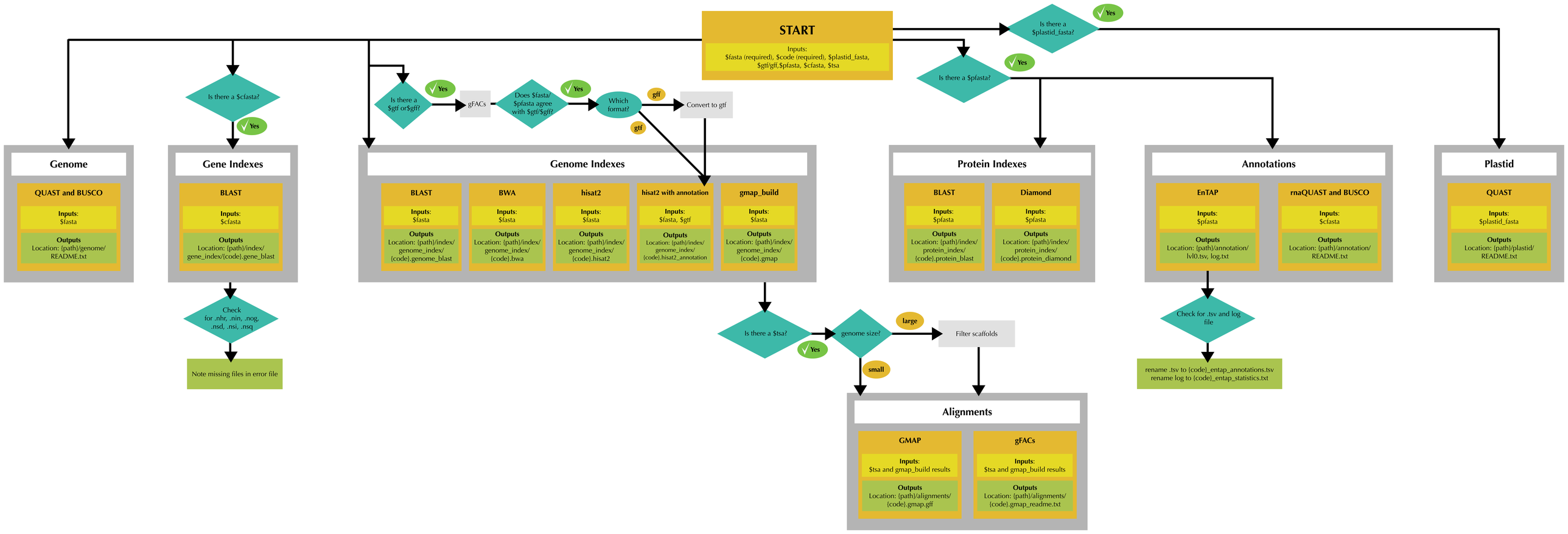
START
Inputs:
$fasta (required), $code (required), $plastid_fasta, $gtf/gff,$pfasta, $cfasta, $tsa
Is there a $plastid_fasta?
Yes
Is there a $pfasta?
Yes
Is there a $cfasta?
Yes
Is there a $gtf or$gff?
Yes
gFACs
Does $fasta/ $pfasta agree with $gtf/$gff?
Yes
Which format?
gff
gtf
Convert to gtf
Genome
QUAST and BUSCO
Inputs:
$fasta
Outputs
Location: {path}/genome/ README.txt
Gene Indexes
BLAST
Inputs:
$cfasta
Outputs
Location: {path}/index/ gene_index/{code}.gene_blast
Check for .nhr, .nin, .nog, .nsd, .nsi, .nsq
Note missing files in error file
Genome Indexes
BLAST
Inputs:
$fasta
Outputs
Location: {path}/index/ genome_index/ {code}.genome_blast
BWA
Inputs:
$fasta
Outputs
Location: {path}/index/ genome_index/ {code}.bwa
hisat2
Inputs:
$fasta
Outputs
Location: {path}/index/ genome_index/ {code}.hisat2
hisat2 with annotation
Inputs:
$fasta, $gtf
Outputs
Location: {path}/index/ genome_index/ {code}.hisat2_annotation
gmap_build
Inputs:
$fasta
Outputs
Location: {path}/index/ genome_index/ {code}.gmap
Is there a $tsa?
Yes
genome size?
large
small
Filter scaffolds
Alignments
GMAP
Inputs:
$tsa and gmap_build results
Outputs
Location: {path}/alignments/ {code}.gmap.gff
gFACs
Inputs:
$tsa and gmap_build results
Outputs
Location: {path}/alignments/ {code}.gmap_readme.txt
Protein Indexes
BLAST
Inputs:
$pfasta
Outputs
Location: {path}/index/ protein_index/ {code}.protein_blast
Diamond
Inputs:
$pfasta
Outputs
Location: {path}/index/ protein_index/ {code}.protein_diamond
Annotations
EnTAP
Inputs:
$pfasta
Outputs
Location: {path}/annotation/ lvl0.tsv, log.txt
rnaQUAST and BUSCO
Inputs:
$cfasta
Outputs
Location: {path}/annotation/ README.txt
Check for .tsv and log file
rename .tsv to {code}_entap_annotations.tsv
rename log to {code}_entap_statistics.txt
Plastid
QUAST
Inputs:
$plastid_fasta
Outputs
Location: {path}/plastid/ README.txt