

```
# environment
TOA_DIR=/home/ranp_u/TOA-master/Package
MINICONDA3_DIR=/home/ranp_u/Miniconda3
MINICONDA3_BIN_DIR=/home/ranp_u/Miniconda3/bin
MINICONDA3_ENVS_DIR=/home/ranp_u/Miniconda3/envs
RESULT_DIR=/home/ranp_u/TOA-results
DB_DIR=/home/ranp_u/TOA-databases
TOA_DB_DIR=/home/ranp_u/TOA-databases/TOA
DATA_DIR=/home/ranp_u/TOA-databases/data
PLAZA_DIR=/home/ranp_u/TOA-databases/PLAZA
NCBI_DIR=/home/ranp_u/TOA-databases/NCBI
INTERPRO_DIR=/home/ranp_u/TOA-databases/InterPro
GO_DIR=/home/ranp_u/TOA-databases/GO
EC_DIR=/home/ranp_u/TOA-databases/EC
KEGG_DIR=/home/ranp_u/TOA-databases/KEGG

# TOA database
TOA_DB=/home/ranp_u/TOA-databases/TOA/toa.db

# basic data
DATASET_FILE=/home/ranp_u/TOA-databases/data/datasets.txt
SPECIES_FILE=/home/ranp_u/TOA-databases/data/species.txt

# other basic data
#EC_IDS_FTP=https://www.enzyme-database.org/downloads/enzyme-data.sql.gz
#EC_IDS_FILE=/home/ranp_u/TOA-databases/EC/enzyme-data.sql.gz
EC_IDS_FTP=ftp://ftp.expasy.org/databases/enzyme/enzyme.dat
EC_IDS_FILE=/home/ranp_u/TOA-databases/EC/enzyme.dat
KEGG_IDS_FTP=ftp://ftp.genome.jp/pub/db/kofam/ko_list.gz
KEGG_IDS_FILE=/home/ranp_u/TOA-databases/KEGG/ko_list.gz

# Gymno PLAZA 1.0
GYMNO_01_CDS_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/Fasta/cds.csv.gz
GYMNO_01_CDS_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-cds.fasta.gz
GYMNO_01_PROTEOME_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/Fasta/proteome.csv.gz
GYMNO_01_PROTEOME_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-proteome.fasta.gz
GYMNO_01_GENEDESC_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/Descriptions
GYMNO_01_GENEDESC_DIR=/home/ranp_u/TOA-databases/PLAZA/gymno_01-descriptions
GYMNO_01_GENEDESC_FILE_PATTERN="gene_description*.csv.gz"
GYMNO_01_INTERPRO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/InterPro/interpro.csv.gz
GYMNO_01_INTERPRO_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-interpro.csv.gz
GYMNO_01_GO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/GO/go.csv.gz
GYMNO_01_GO_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-go.csv.gz
GYMNO_01_MAPMAN_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/MapMan/mapman.csv.gz
GYMNO_01_MAPMAN_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-mapman.csv.gz
GYMNO_01_BLASTPLUS_DB_NAME=gymno_01
GYMNO_01_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/gymno_01-blastplus-db
GYMNO_01_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-blastplus-db/gymno_01
GYMNO_01_DIAMOND_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/gymno_01-diamond-db
GYMNO_01_DIAMOND_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/gymno_01-diamond-db/gymno_01
GYMNO_01_BLAST_XML=$OUTPUT_DIR/gymno_01-alignment.xml
GYMNO_01_ANNOTATION_FILE=$OUTPUT_DIR/gymno_01-annotation.csv
GYMNO_01_NON_ANNOTATED_TRANSCRIPT_FILE=$OUTPUT_DIR/gymno_01-nonann-transcripts.fasta
GYMNO_01_NON_ANNOTATED_PEPTIDE_FILE=$OUTPUT_DIR/gymno_01-nonann-peptides.fasta
```

```

# Dicots PLAZA 4.0
DICOTS_04_CDS_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/Fasta/cds.all_transcripts.fasta.gz
DICOTS_04_CDS_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-cds.fasta.gz
DICOTS_04_PROTEOME_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/Fasta/proteome.all_transcripts.fasta.gz
DICOTS_04_PROTEOME_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-proteome.fasta.gz
DICOTS_04_GENEDESC_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/Descriptions
DICOTS_04_GENEDESC_DIR=/home/ranp_u/TOA-databases/PLAZA/dicots_04-descriptions
DICOTS_04_GENEDESC_FILE_PATTERN="gene_description.*.csv.gz"
DICOTS_04_INTERPRO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/InterPro/interpro.csv.gz
DICOTS_04_INTERPRO_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-interpro.csv.gz
DICOTS_04_GO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/GO/go.csv.gz
DICOTS_04_GO_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-go.csv.gz
DICOTS_04_MAPMAN_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_dicots_04/MapMan/mapman.csv.gz
DICOTS_04_MAPMAN_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-mapman.csv.gz
DICOTS_04_BLASTPLUS_DB_NAME=dicots_04
DICOTS_04_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/dicots_04-blastplus-db
DICOTS_04_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-blastplus-db/dicots_04
DICOTS_04_DIAMOND_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/dicots_04-diamond-db
DICOTS_04_DIAMOND_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/dicots_04-diamond-db/dicots_04
DICOTS_04_BLAST_XML=$OUTPUT_DIR/dicots_04-alignment.xml
DICOTS_04_ANNOTATION_FILE=$OUTPUT_DIR/dicots_04-annotation.csv
DICOTS_04_NON_ANNOTATED_TRANSCRIPT_FILE=$OUTPUT_DIR/dicots_04-nonann-transcripts.fasta
DICOTS_04_NON_ANNOTATED_PEPTIDE_FILE=$OUTPUT_DIR/dicots_04-nonann-peptides.fasta

# Monocots PLAZA 4.0
MONOCOTS_04_CDS_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/Fasta/cds.all_transcripts.fasta.gz
MONOCOTS_04_CDS_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-cds.fasta.gz
MONOCOTS_04_PROTEOME_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/Fasta/proteome.all_transcripts.fasta.gz
MONOCOTS_04_PROTEOME_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-proteome.fasta.gz
MONOCOTS_04_GENEDESC_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/Descriptions
MONOCOTS_04_GENEDESC_DIR=/home/ranp_u/TOA-databases/PLAZA/monocots_04-descriptions
MONOCOTS_04_GENEDESC_FILE_PATTERN="gene_description.*.csv.gz"
MONOCOTS_04_INTERPRO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/InterPro/interpro.csv.gz
MONOCOTS_04_INTERPRO_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-interpro.csv.gz
MONOCOTS_04_GO_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/GO/go.csv.gz
MONOCOTS_04_GO_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-go.csv.gz
MONOCOTS_04_MAPMAN_FTP=ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_monocots_04/MapMan/mapman.csv.gz
MONOCOTS_04_MAPMAN_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-mapman.csv.gz
MONOCOTS_04_BLASTPLUS_DB_NAME=monocots_04
MONOCOTS_04_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/monocots_04-blastplus-db
MONOCOTS_04_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-blastplus-db/monocots_04
MONOCOTS_04_DIAMOND_DB_DIR=/home/ranp_u/TOA-databases/PLAZA/monocots_04-diamond-db
MONOCOTS_04_DIAMOND_DB_FILE=/home/ranp_u/TOA-databases/PLAZA/monocots_04-diamond-db/monocots_04
MONOCOTS_04_BLAST_XML=$OUTPUT_DIR/monocots_04-alignment.xml
MONOCOTS_04_ANNOTATION_FILE=$OUTPUT_DIR/monocots_04-annotation.csv
MONOCOTS_04_NON_ANNOTATED_TRANSCRIPT_FILE=$OUTPUT_DIR/monocots_04-nonann-transcripts.fasta
MONOCOTS_04_NON_ANNOTATED_PEPTIDE_FILE=$OUTPUT_DIR/monocots_04-nonann-peptides.fasta

# NCBI RefSeq Plant
REFSEQ_PLANT_FTP=ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plant/
REFSEQ_PLANT_LOCAL=/home/ranp_u/TOA-databases/NCBI/ftp.ncbi.nlm.nih.gov/refseq/release/plant/
REFSEQ_PROTEIN_FILE_PATTERN="protein.faa.gz"
REFSEQ_PLANT_PROTEOME_FILE=/home/ranp_u/TOA-databases/NCBI/refseq_plant-proteome.fasta
REFSEQ_PLANT_BLASTPLUS_DB_NAME=refseq_plant
REFSEQ_PLANT_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/NCBI/refseq_plant-blastplus-db
REFSEQ_PLANT_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/NCBI/refseq_plant-blastplus-db/refseq_plant
REFSEQ_PLANT_DIAMOND_DB_DIR=/home/ranp_u/TOA-databases/NCBI/refseq_plant-diamond-db
REFSEQ_PLANT_DIAMOND_DB_FILE=/home/ranp_u/TOA-databases/NCBI/refseq_plant-diamond-db/refseq_plant
REFSEQ_PLANT_FILE_LIST=/home/ranp_u/TOA-databases/NCBI/refseq_plant-file-list.txt
REFSEQ_PLANT_BLAST_XML=$OUTPUT_DIR/refseq_plant-alignment.xml
REFSEQ_PLANT_ANNOTATION_FILE=$OUTPUT_DIR/refseq_plant-annotation.csv
REFSEQ_PLANT_NON_ANNOTATED_TRANSCRIPT_FILE=$OUTPUT_DIR/refseq_plant-nonann-transcripts.fasta
REFSEQ_PLANT_NON_ANNOTATED_PEPTIDE_FILE=$OUTPUT_DIR/refseq_plant-nonann-peptides.fasta

```

```
# NCBI Taxonomy
TAXONOMY TAXDMP FTP=ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdmp.zip
TAXONOMY TAXDMP_FILE=/home/ranp_u/TOA-databases/NCBI/taxdmp.zip
TAXONOMY TAXONNODES_FILE=/home/ranp_u/TOA-databases/NCBI/nodes.dmp
TAXONOMY TAXONNAMES_FILE=/home/ranp_u/TOA-databases/NCBI/names.dmp
TAXONOMY_PROTACCESSION_2_TAXID FTP=ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/accession2taxid/prot.accession2taxid.gz
TAXONOMY_PROTACCESSION_2_TAXID_FILE=/home/ranp_u/TOA-databases/NCBI/prot.accession2taxid.gz
VIRIDIPLANTAE TAXID=33090
VIRIDIPLANTAE_TAXID_LIST_FILE=/home/ranp_u/TOA-databases/NCBI/taxids_${VIRIDIPLANTAE_TAXID}.txt

# NCBI BLAST database
BLAST_DATABASES FTP=ftp://ftp.ncbi.nlm.nih.gov/blast/db/
BLAST_DATABASES_LOCAL=/home/ranp_u/TOA-databases/NCBI/ftp.ncbi.nlm.nih.gov/blast/db/

# NCBI BLAST database NT
NT_FILE_PATTERN="nt.*.tar.gz"
NT_FILE_LIST=/home/ranp_u/TOA-databases/NCBI/nt-file-list.txt
NT_BLASTPLUS_DB_NAME=nt
NT_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/NCBI/nt-blastplus-db
NT_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/NCBI/nt-blastplus-db/nt
NT_BLAST_XML=$OUTPUT_DIR/nt-alignment.xml
NT_VIRIDIPLANTAE_ANNOTATION_FILE=$OUTPUT_DIR/nt-viridiplantae-annotation.csv
NT_CONTAMINATION_ANNOTATION_FILE=$OUTPUT_DIR/nt-contamination-annotation.csv
NT_NON_ANNOTATED_TRANSCRIPT_FILE=$OUTPUT_DIR/nt-nonann-transcripts.fasta

# NCBI Nucleotide GenInfo identifier lists
NUCLEOTIDE_VIRIDIPLANTAE_GI_LIST=/home/ranp_u/TOA-databases/NCBI/nucleotide_viridiplantae.gi

# NCBI BLAST database NR
NR_PROTEOME FTP=ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz
NR_PROTEOME_FILE=/home/ranp_u/TOA-databases/NCBI/nr.gz
NR_FILE_PATTERN="nr.*.tar.gz"
NR_FILE_LIST=/home/ranp_u/TOA-databases/NCBI/nr-file-list.txt
NR_BLASTPLUS_DB_NAME=nr
NR_BLASTPLUS_DB_DIR=/home/ranp_u/TOA-databases/NCBI/nr-blastplus-db
NR_BLASTPLUS_DB_FILE=/home/ranp_u/TOA-databases/NCBI/nr-blastplus-db/nr
NR_DIAMOND_DB_DIR=/home/ranp_u/TOA-databases/NCBI/nr-diamond-db
NR_DIAMOND_DB_FILE=/home/ranp_u/TOA-databases/NCBI/nr-diamond-db/nr
NR_BLAST_XML=$OUTPUT_DIR/nr-alignment.xml
NR_VIRIDIPLANTAE_ANNOTATION_FILE=$OUTPUT_DIR/nr-viridiplantae-annotation.csv
NR_CONTAMINATION_ANNOTATION_FILE=$OUTPUT_DIR/nr-contamination-annotation.csv
NR_NON_ANNOTATED_PEPTIDE_FILE=$OUTPUT_DIR/nr-nonann-peptides.fasta

# NCBI Protein GenInfo identifier lists
PROTEIN_VIRIDIPLANTAE_GI_LIST=/home/ranp_u/TOA-databases/NCBI/protein_viridiplantae.gi

# NCBI Gene
GENE_GENE2GO FTP=ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz
GENE_GENE2GO_FILE=/home/ranp_u/TOA-databases/NCBI/gene-gene2go.gz
GENE_GENE2REFSEQ FTP=ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2refseq.gz
GENE_GENE2REFSEQ_FILE=/home/ranp_u/TOA-databases/NCBI/gene-gene2refseq.gz

# InterPro
INTERPRO_INTERPRO2GO FTP=ftp://ftp.ebi.ac.uk/pub/databases/interpro/interpro2go
INTERPRO_INTERPRO2GO_FILE=/home/ranp_u/TOA-databases/InterPro/interpro2go
```

```
# NCBI Protein GenInfo identifier lists
PROTEIN_VIRIDIPLANTAE_GI_LIST=/home/ranp_u/TOA-databases/NCBI/protein_viridiplantae.gi

# NCBI Gene
GENE_GENE2GO_FTP=ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz
GENE_GENE2GO_FILE=/home/ranp_u/TOA-databases/NCBI/gene-gene2go.gz
GENE_GENE2REFSEQ_FTP=ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2refseq.gz
GENE_GENE2REFSEQ_FILE=/home/ranp_u/TOA-databases/NCBI/gene-gene2refseq.gz

# InterPro
INTERPRO_INTERPRO2GO_FTP=ftp://ftp.ebi.ac.uk/pub/databases/interpro/interpro2go
INTERPRO_INTERPRO2GO_FILE=/home/ranp_u/TOA-databases/InterPro/interpro2go

# Gene Ontology
GO_ONTOLOGY_FTP=http://purl.obolibrary.org/obo/go.obo
GO_ONTOLOGY_FILE=/home/ranp_u/TOA-databases/GO/go.obo
GO_EC2GO_FTP=http://geneontology.org/external2go/ec2go
#GO_EC2GO_FTP=https://build.berkeleybop.org/view/GO/job/Update%20external2go/lastSuccessfulBuild/artifact/external2go/ec2go
GO_EC2GO_FILE=/home/ranp_u/TOA-databases/GO/ec2go.txt
GO_KEGG2GO_FTP=http://geneontology.org/external2go/kegg2go
#GO_KEGG2GO_FTP=https://build.berkeleybop.org/view/GO/job/Update%20external2go/lastSuccessfulBuild/artifact/external2go/kegg2go
GO_KEGG2GO_FILE=/home/ranp_u/TOA-databases/GO/kegg2go.txt
GO_METACYC2GO_FTP=http://geneontology.org/external2go/metacyc2go
#GO_METACYC2GO_FTP=https://build.berkeleybop.org/view/GO/job/Update%20external2go/lastSuccessfulBuild/artifact/external2go/metacyc2go
GO_METACYC2GO_FILE=/home/ranp_u/TOA-databases/GO/metacyc2go.txt
GO_INTERPRO2GO_FTP=http://geneontology.org/external2go/interpro2go
#GO_INTERPRO2GO_FTP=https://build.berkeleybop.org/view/GO/job/Update%20external2go/lastSuccessfulBuild/artifact/external2go/interpro2go
GO_INTERPRO2GO_FILE=/home/ranp_u/TOA-databases/GO/interpro2go.txt

# other transcriptome files
REIDENTIFIED_TRANSCRIPTOME_FILE=$OUTPUT_DIR/reidentified-transcriptome.fasta
TOA_TRANSCRIPTOME_RELATIONSHIP_FILE=$OUTPUT_DIR/toa_transcriptome_relationships.csv
PURGED_TRANSCRIPTOME_FILE=$OUTPUT_DIR/purged-transcriptome.fasta

# peptide sequence files
TRANSDECODER_OUTPUT_DIR=$OUTPUT_DIR/transdecoder
ORF_FILE=$TRANSDECODER_OUTPUT_DIR/longest_orfs.pep
PEPTIDE_FILE=$OUTPUT_DIR/`basename "$REIDENTIFIED_TRANSCRIPTOME_FILE"`.transdecoder.pep
REIDENTIFIED_PEPTIDE_FILE=$OUTPUT_DIR/reidentified-peptides.fasta
TOA_TRANSDECODER_RELATIONSHIP_FILE=$OUTPUT_DIR/toa_transdecoder_relationships.csv

# merger files
MERGED_ANNOTATION_FILE=$OUTPUT_DIR/merged-annotation.csv
PLANT_ANNOTATION_FILE=$OUTPUT_DIR/plant-annotation.csv
MERGED_BLAST_XML=$OUTPUT_DIR/merged-alignment.xml

# statistics
STATS_SUBDIR_NAME=stats
STATS_DIR=$OUTPUT_DIR/stats
STATS_BASE_NAME=stats
ANNOTATION_STATS_FILE=$OUTPUT_DIR/stats/stats.csv
```