

Results of searching for proteins' RefSeq assemblies and loci

Results of searching for protein's homologous with blastp

Upstream sequences fasta file

Folder with tsv tables containing information about annotated ORFs in each locus.

Folder with fasta files containing results of multiple sequence alignments for conserved ORFs

R script for reproducing MSA visualisation; based on ggmsa R package.
Use ***Rscript msa_plot.R -h*** for more information.

Results summary table containing the main information about found sets of ORFs


Folder with MSA visualisation (pdf files) obtained with msa_plot.R script.

Folder with seqlogo visualisation (pdf files) obtained with python logomaker library.

Output folder structure


 found_homologous.tsv

 assemblies_list.tsv

 upstream_sequences.fa

 annotated_ORFs

 locus_id_1.tsv

 locus_id_2.tsv

.....

 nt/aa_msa_fasta_files


 conserved_ORFs_set_1.fa


 conserved_ORFs_set_2.fa


.....

 results_summary.tsv

 msa_plot.R


 nt/aa_msa_visualisation


 conserved_ORFs_set_1.pdf

 conserved_ORFs_set_2.pdf

.....

 nt/aa_seqlogo_visualisation

 conserved_ORFs_set_1.pdf

 conserved_ORFs_set_2.pdf

.....

