Output folder structure

Results of searching for protein's homologous with blastp → 🟩 found_homolohous.tsv

Proteins' RefSeq assemblies → 🟩 assemblies_list.tsv

Upstream sequences fasta file → 📄 upstream_sequences.fa

📁 annotated_ORFs

tsv tables with information about annotated ORFs in each locus. → 🟩 locus_id_1.tsv
🟩 locus_id_2.tsv
......

📁 nt/aa_orfs_fasta_files

Fasta files with conserved ORFs sequences → 📄 conserved_ORFs_set_1.fa
📄 conserved_ORFs_set_2.fa
......

📁 nt/aa_msa_fasta_files

Multiple sequence alignments fasta files for conserved ORFs → 📄 conserved_ORFs_set_1.fa
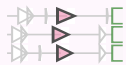📄 conserved_ORFs_set_2.fa
......

Summary statistics for conserved ORFs → 🟩 results_summary.tsv

R script for reproducing MSA visualisation; based on ggmsa R package. → ⟨⟩ msa_plot.R

📁 annotation_visualisation

Visualisation of loci' annotation → conserved_ORFs_set_1.pdf
conserved_ORFs_set_2.pdf
......

📁 nt/aa_msa_visualisation

MSA visualisation obtained with msa_plot.R script. → conserved_ORFs_set_1.pdf
conserved_ORFs_set_2.pdf
......

📁 nt/aa_seqlogo_visualisation

Sequence logo visualisation. → conserved_ORFs_set_1.pdf
conserved_ORFs_set_2.pdf
......