Annotating with DAS

Task Name: das_annotation

Annotate with DAS. Finds similar protein sequence using remote BLAST. Using IDs of sequences found loads annotation for DAS sources. Nucleotide sequences are skipped if any supplied to input.

Parameters:

in - Input amino acid sequence [Url datasets]
out - Output annotated sequence [String]
db - Database against which the search is performed: UniProtKB or clusters of sequences with 100%, 90% or 50% identity. [String]

The following databases are available:

- "uniprotkb";
- "uniprotkb_archaea";
- "uniprotkb_bacteria";
- "uniprotkb_eukaryota";
- "uniprotkb_arthropoda";
- "uniprotkb_fungi";
- "uniprotkb_human";
- "uniprotkb_mammals";
- "uniprotkb_nematoda";
- "uniprotkb_plants";
- "uniprotkb_rodents";
- "uniprotkb_vertebrates";
- "uniprotkb_viruses";
- "uniprotkb_pdb";
- "uniprotkb_complete_microbial_proteomes";
- "uniprotkb_swissprot";
- "UniRef100";
- "UniRef90";
- "UniRef50";
- "uniparc";

f - Low-complexity regions (e.g. stretches of cysteine in Q03751, or hydrophobic regions in membrane proteins) tend to produce spurious, insignificant matches with sequences in the data base which have the same kind of low-complexity regions, but are unrelated biologically. [String]
s - The DAS sources to read features from. [String]
g - This will allow gaps to be introduced in the sequences when the comparison is done. [String]
i - Minimum identity of a BLAST result and an input sequence. [Number]
r - Use first IDs of similar sequences to load annotations [Number]
m - The matrix assigns a probability score for each position in an alignment. [String]
h - Limits the number of returned alignments. [String]
t - The expectation value (E) threshold is a statistical measure of the number of expected matches in a random database. The lower the e-value, the more likely the match is to be signif icant. [String]

Example:

```
ugene das_annotation  --in=test.fa --out=test_out.gb --db=uniprotkb_plants
```