Carnation DB

http://carnation.kazusa.or.jp/

**December 17, 2013**





How to use Keyword Search

1. Input your keyword.
2. **Case sensitive:** Check the ‘Case sensitive’ check box if you need to discriminate between upper and lower cases.

1. **Include:** Check the button when searching any words containing the keyword.

**Word:** Check the button when searching words that exactly match the keyword.

For example:

 Annotation “L-fuco**kinase**/GDP-L-fucose pyrophosphorylase”

 Keyword “kinase”

In this case, a search with the “Include” option finds the example annotation word, but a search with the “Word” option does not.

1. **Without transposons:** Check the box when searching databases without transposon-related genes.
2. **Entries per page:** Select the number of hits to display per page. (default: 50)
3. Press the “Submit” button.

There is a pager at the bottom of the result page. You can move to another page by clicking the button or by directly inputting the number.

How to use Sequence Cut

This tool displays a sequence corresponding to a specified range of a scaffold sequence.

1. Input “Scaffold ID”.
2. Input “Start” and “End” position.
3. **reverse complement:** You can get a reverse-complement sequence by checking it.
4. Press the “Submit” button.



How to use Blast Search

1. **Database:** Select the database.

2. **Search program:** Specify the search program from the followings.

|  |  |  |  |
| --- | --- | --- | --- |
| Program | Query | Database | An explanation |
| BLASTN | nucleotide | nucleotide | Aligning your nucleotide sequence with nucleotide sequence database. |
| BLASTP | amino acid | amino acid | Aligning your amino acid sequence with amino acid sequence database. |
| TBLASTN | amino acid | nucleotide | Aligning your amino acid sequence with nucleotide sequence database by translating database sequences taking into account all six possible open reading frames. |
| TBLASTX | nucleotide | nucleotide | Aligning your nucleotide sequence with nucleotide sequence database by translating both sequences taking into account all six possible open reading frames. |

(Quote from DDBJ BLAST.)

1. **Number of search results:** Specify the number of hits to display.

1. **Query sequence:** Paste your sequence in the box.
2. **Advanced settings:** Additional options including Evalue-based filtering are available.

6. Press the “Submit” button.





How to use List

In ‘LIST’, annotations (CDS, translated ORF, gene structure, and BLAST search summary, etc.) of a gene can be displayed.

1. Select a category containing a gene ID to be searched.
2. Click the gene ID to display the annotations.

Annotation example )



About files in Download page

Files with ‘woTE’ in their names contain sequences from which transposable elements are excluded.

1. DCA\_r1.0\_scaffolds.fa

Genome assembly sequence in FASTA format.

1. DCA\_r1.0\_genes.gff3, DCA\_r1.0\_genes\_woTE.gff3

GFF3 annotation files, which were created by combining PASA and cDNA-incorporated Augusts annotation results.

1. DCA\_r1.0\_blastp\_Trem.out, DCA\_r1.0\_blastp\_AtOs.out

Output files of BLASTp searches for translated ORFs that were predicted based on the annotation data (2).

Files with ‘Trem’ in their names were obtained by BLASTp searches against the UniprotTremble database.

Files with ‘AtOs’ were obtained by searches against a database comprised of Arabidopsis and Rice proteins.

1. DCA\_r1.0\_blastp-summary\_Trem.txt, DCA\_r1.0\_blastp-summary\_AtOs.txt

From the BLAST results (3), alignments with identity < 30% and evalue > 1E-5 were filtered out, and their alignment details were summarized in the files.

Explanation of each item is shown at the head of the file.

1. DCA\_r1.0\_cds.fna, DCA\_r1.0\_cds\_woTE.fna

CDS files in FASTA format.

The CDSs were predicted based on the BLAST results (3). In the case of finding no similar protein, CDSs with the longest ORF were given. The header line contains protein annotations given for its similar protein.

The header is described as follows.

>mRNA ID the number of exons contained in a gene||ORF frame（e.g., f-3: translation starts at position 3 of forward strand）|| annotation information of the most similar protein if any

(6) DCA\_r1.0\_pep.faa, DCA\_r1.0\_pep\_woTE.faa

Protein sequence encoded in CDS (5) in FASTA format. The header description is the same as (5).

1. TE\_annot\_list.txt

A file listing mRNA IDs and their annotations about CDSs coding transposable elements.

(8) DCA\_r1.0\_tRNA\_tRNAscan\_list.txt

A list of tRNA gene candidates predicted with tRNAscan.

(9) DCA\_r1.0\_rRNA\_blast\_list.txt

A list of rRNA genes, found by searching with BLASTn against Arabidopsis rRNA sequences.

The blast hits with < 85% coverage and < 80% identity for target were filtered out.

(10) DCA\_r1.0\_snoRNA\_snoscan\_list.txt

A list of snoRNA genes predicted with SnoScan. The results with score < 28 and redundant genes were filtered out.

(11) DCA\_r1.0\_miRNA\_MapMi\_list.txt

A list of miRNA genes, found by searching with MapMi against miRBase library consisting of plant miRNA sequences. The results with score < 35 were filterd out.

(12) DCA\_r1.0\_repeat\_list.txt

A list of transposable elements and simple repeats, found by searching with RepeatMasker and TransposonPSI against a known library and a novel repeat library (13).

(13) DCA\_r1.0\_repeat\_lib.fasta

Carnation-specific repeat sequences, found by de novo repeat finding with RepeatScout and Piper.