Carnation DB

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

**2013/12/17 更新**





Keyword Searchの利用方法

1. キーワードを入力して下さい。
2. 大文字、小文字を区別したい場合は「Case sensitive」にチェックを入れて下さい。デフォルトは区別しません。
3. Includeは文章全体を検索対象にしますがWordは単語単位で検索します。

例えば“L-fuco**kinase**/GDP-L-fucose pyrophosphorylase”というアノテーションがあり、キーワードが“kinase”の場合、Include は“kinase”にヒットしますが、Word はヒットしません。

1. トランスポゾン因子を除いて検索する場合は「Without transposons」にチェックを入れて下さい。
2. １ページに表示する件数を選択して下さい。デフォルトは 50 件表示です。
3. Submit を押して下さい。(検索結果が表示されます)

検索結果画面の下部にはページャーがありますのでページの移動が可能です。直接ページを指定して移動する事も可能です。



Sequence Cut の利用方法

Scaffold 配列から指定したポジションで配列を切り出します。

1. Scaffold ID を入力して下さい。
2. Start と End のポジションを入力して下さい。
3. 相補鎖で切り出す場合は「reverse complement」にチェックを入れて下さい。
4. Submit を押して下さい。



Blast 検索の利用方法

1. **Database:** ご希望のデータベースを選択して下さい。

“without transposons” と記載されているデータベースはトランスポゾン因子を除いたデータセットです。

1. **Search program:** 下記を参考に BLAST のプログラムを選択して下さい。

|  |  |  |  |
| --- | --- | --- | --- |
|  プログラム |  クエリ | データベース |  説明 |
|  BLASTN | 塩基配列 | 塩基配列 | 塩基配列を塩基配列データベースと比較します |
|  BLASTP | アミノ酸配列 | アミノ酸配列 | アミノ酸配列をアミノ酸配列データベースと比較します |
|  TBLASTN | アミノ酸配列 | 塩基配列 | アミノ酸配列に対して、塩基配列データベースを表裏合わせて６通りの読み枠で翻訳しながら比較します |
|  TBLASTX | 塩基配列 | 塩基配列 | 塩基配列を表裏合わせて６通りの読み枠で翻訳しながら、同様に翻訳された塩基配列データベースと比較します |

(DDBJ BLAST ヘルプより一部転載)

1. **Number of search results:** ヒットしたデータを何件表示するかを指定できます。
2. **Query sequence:** 配列を入力して下さい。
3. 更に条件を加えたい場合は Advanced settings にお進み下さい。

6. Submit を押して下さい。（暫くするとBlast の結果が表示されます）

 

Listの利用方法

1. 調べたいID が含まれるカテゴリをクリックします。
2. ID の一覧が表示されますのでご希望のIDをクリックして下さい。

下記のようなアノテーション情報が表示されます。

アノテーション例）



Download データについて

下記ファイルで 'woTE’ が付いているものは、タンパク質をコードするトランスポソン因子を除いてあるものです。

1. DCA\_r1.0\_scaffolds.fa

ゲノムアセンブリ配列のfasta format fileです。

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

PASA, Augustusを用いてEST, cDNA assembly配列を取り込んだ 主にタンパク質をコードする遺伝子のアノテーションファイルです。

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

(2)のアノテーションファイルからmRNA配列を抽出し、予測されるORFの翻訳産物をデータベースに対して blastp検索した出力ファイルです。'Trem' がついているものはUniprotTremble databaseを用いてBLAST検索を行った結果に基づいたものです。

’AtOs’ がついているものは、シロイヌナズナとイネのタンパク質データベース（両方を混ぜ

たもの）を用いてBLAST検索を行った結果に基づいたものです。

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

(3)のデータについて、30%以上の相同性かつevalue 1E-5以下のヒットをまとめたものです。各列の記載事項についてはファイルの行頭に記されています。

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

(3)の結果から最も相同性の高いCDSをfasta fileにしたもので、 ヘッダーには相同タンパク質についてのアノテーションを付してあります。相同タンパク質が見つからなかったものについては、 mRNAの中で最も長いORFを持ったものを抽出しています。

ヘッダーの構成は以下のようになります。

>mRNA ID 遺伝子のエクソン数||ORFのフレーム（e.g., f-3: translation starts at position 3 of forward strand）||相同タンパク質が見つかったものについてはそのアノテーション情報

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

(5)の翻訳産物のfasta fileです。ヘッダーの記述は(5)と同じです。

(7) TE\_annot\_list.txt

’woTE’ で除いたトランスポゾン因子をコードするmRNA ID、アノテーションリストです。

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

(1)のゲノム配列に対して、tRNAscan toolを用いて tRNA遺伝子を予測した結果をリストにしています。

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

ゲノム配列をArabidopsis rRNA 配列に対してblast 検索し、 相同性80%以上、rRNA配列のカバレッジ85%以上のものについてリストにしています。

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

ゲノム配列をSnoScanツールを用いてsnoRNA遺伝子を予測し、スコア28以上の遺伝子の選別および重複する遺伝子の除去を行ったリストです。

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

ゲノム配列をmiRBase databse (植物miRNAのみを抽出したもの) に対して MapMi toolを用いてmiRNAを予測し、スコア35以上のものをリストにしています。

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

既知のリピート配列データベースおよびカーネーションゲノム新規リピート配列ライブラリー(13) に対して、 RepeatMaskerおよびTransposonPSIを用いて見いだしたトランスポゾン配列 および単純繰り返し配列のリストです。

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

RepeatScoutおよびPilerを用いて、カーネーションゲノム特異的に見いだされた新規リピート配列ライブラリーのfasta fileです。