14.10.2012

CUO

Codon Usage Optimizer
Beta 0.92

User Guide

Contents

Intro

CUO is written in Java as a multiplatform Graphical User Interface based software. Main function of this program at current development stage is to optimize expression of genes to be transformed into *Chlamydomonas reinhardtii* chloroplast.

CUO is open source and the source code will be published after beta testing. It was written carefully with user/developer friendly structure in mind. When used together with Dropbox, hassle-free sharing of database is easily done in a research group.

The future plan is to develop CUO into a multipurpose synthetic gene design software where sharing, learning, organizing or planning works in biology labs can be as quick and as easy as doing Photoshop without redundant processes.

About author

Kong Khai Jien
Msc Nanotechnology and Regenerative Medicine
University College London
Please email any bug reports or suggestions to:
zcbtfi0@live.ucl.ac.uk

Acknowledgements

UCL Algal Research Group:
Saul Purton, Henry Taunt, Sofie Vonlanthen, Joanna Szaub, Rosanna Young, Janet Waterhouse, Tommaso Barbi, Lamya Al-Haj, Thushyanthi.

iGEME team:
Graham Richard, Andrew Martin, Simcha Jong, Alfred Ho, Ng Kheng Tee, Li Meng, Farhan Sobri, Aakanksha Khandelwal, Anastasia Suprun, Freedanz Ferdinandz, Ali Versi, Lee Mun Ching.

Morph Bioinformatics crew:
Remya Koshy, Balakrishnan, Tulip Sunil Nandu, Nikunj Maheshwari, Rameez Choudhari.

Recipes

Optimizing a gene into *C.reinhardtii* chloroplast genome

1. Get the NA (nucleic acid) or AA (amino acid) sequence of gene of interest.
2. Using File>Sequence Creator, create a FastaNA (if you have the NA) or FastaAA (if you have the AA) file of gene of interest and save it in Library. If NA is used, make sure the gene of interest NA length is divisible by 3 which means the codons are in frame.
3. Open Tools>Moptimizer.
4. Select your target gene by "Select Target" button.
5. Click "Load Weight Table" button to load the preferred weight table. There are a few precompiled choices already. They are:
   - C.reinhardtii Autopick = analysis result of theoretically high expression protein genes in *C.reinhardtii* calculated via mathematical formula.
   - C.reinhardtii Complete = analysis result of all protein genes.
   - C.reinhardtii Handpick = handpicked high expression protein genes analysis. (**Recommended**)
   - C.reinhardtii Low Expression = handpicked low expression protein genes analysis. (Can be used to de-optimize a gene)
   - C.reinhardtii Uncategorized = the left over protein genes apart from handpicked high and low expression protein genes in C.reinhardtii. (theoretically medium expression level)
6. Load Codon Frequency Table and Codon Pair Table as well with each load buttons.
7. Click "Generate" button and choose original in the combobox and click OK. This is to have a preview of the original adapatation of the gene of interest in chosen genome.
8. If it is not satisfactory, click Generate again and choose All the best followed by tick on both avoid codon repeats and avoid bad codon pairs checkboxes then click OK.
9. View the optimized codons. Click on the codons (the one in graphic) to alternate their coding sequence among their siblings to get the best solution that you personally feel most correct.
10. After clicking through the entire gene of interest, use "Detect sites" button on the right to detect the restriction sites occur in the gene of interest.
11. Use Auto Eliminate button to eliminate specific restriction site located in the gene according to demand.
12. Review the gene again to see the minor changes and alternate them as needed.
13. Press "Save" button to save the gene with a name.
14. The sequence is stored in Library folder under the main CUO folder of the program.
15. Use http://mfold.rit.albany.edu/?q=mfold/RNA-Folding-Form to see the folding energy of the mRNA. The less negative is the ΔG the better is the expression. Generally genes optimized for AT-rich *C.reinhardtii* genome has favourable mRNA folding energy.

Outline:

1. Sequence Creator
2. Moptimizer
3. Select Target, Load Weight Table, Load Codon Frequency Table, Load Codon Pair Table
4. Generate (All the best, avoid codon repeats, avoid bad codon pairs)
5. Click through the codons
6. Detect Sites
7. Final review
8. Save
9. mfold web server

Using Browser Panel

The genome is double-clickable. Double click to see what is contained in a genome. Delete key can be used to deleted unwanted sequences.

Using Sequence Creator

- FastaNA = Fasta Nucleic Acid file (.fna)
  - FastaNA>Gene = A gene. Default save location is Library.
  - FastaNA>Restriction = A restriction site. Use > for coding strand cutting position and < for template strand cutting position. Default save location is Toolbox>Restriction
- FastaAA = Fasta Amino Acid file (.faa)
- MultiFasta = Multiple Fasta nucleic acid files compiled into one (.mfa)

Using CSR

Coding Sequence Retriever
Internet connection required. Type in search term and press enter to search the term in NCBI Nucleotide database. Use filter function (the funnel button just beside the search term box, press it to activate it) to filter results. Use Retrieve button the retrieve genome interested. Note that, most of the time the genome are not annotated and the result is that no genes can be seen in the browser panel while entering the downloaded genome.

Using CAI Analyzer

Codon Adaptation Index Analyzer
Reference set refers to genes or genomes that are considered as host organism for the gene of interest. Target sequences refer to the genes that are going to be inserted into the reference set.
Empty target sequences list is allowed.
Pick the reference set and click analyze button to get a pdf report plus .wt, .cft, .cpt files which are useful in gene optimization later.
Weight table file (.wt) can be imported to analyze the target sequences by import button.
Alternatively, use Insert button to insert the weights manually.

Using Moptimizer

Manual Optimizer
Select target sequence with select target button.
Load weight table file (.wt) with load weight table button.
Load codon frequency table file (.cft) with load codon frequency button.
Load codon pair table file (.cpt) with load codon pair table button.
Generate sequence from generate button.
Save the optimized gene with save button.
Detect and eliminate restriction sites occurring in the gene using detect sites button.
Use snapshot button to take photo of the entire gene. Toggle the bars on/off with “View:” part.
Locate a base directly with Go to textfield.
Hover over the amino acid or nucleic acid and wait for one second to see their positon label.
Click on the four square boxes beside CAI or CPI value to get weight of each target sequence’s codon in text form.
Click on the little white button beside the four square box button to view the weights diagram in graph format.

Using Processor

View Tables – Use this function to view any table generated
Kasuza Table – Parse table copied from kasuza database (http://www.kazusa.or.jp/codon/)
Compare Tables – Rapid comparison between tables of same type
Edit Table – Load, view and edit an existing table

Using Translator

Translate – Translate the given nucleic acid sequence into single letter amino acid sequence.
Transcribe – Convert single letter amino acid sequence into its nucleic acid sequence by selecting the first codon in each list of sibling codons.
Complementary – Return the complementary sequence of the given nucleic acid sequence.

Using Operation Panel

Still under construction

Click on any gene or genome on the browser panel and click toStage button on top to put it to the operation panel.
Drag any sequence from browser panel to the canvas.
Right click on canvas to get a drop down list of operations to perform.
Use ‘W’ ‘A’’S’’D’ or mouse drag to pan. Use mouse scroll to zoom in and zoom centrally.
“Ctrl”+mouse click/drag to select multiple fragments.
“Shift”+mouse click/drag to select additional fragments.
“Spacebar” to toggle Sequence Viewer on/off.
“Z” key to toggle moving mode on/off

Using Updater

Refresh button – refresh the update list
Note: Windows users uses updater differently. CUO is closed before Updater is able to launch.

License