

The pyCOGNAT approach for comparative analysis of genomes of chloroplasts and plastids

D.V. Dibrova

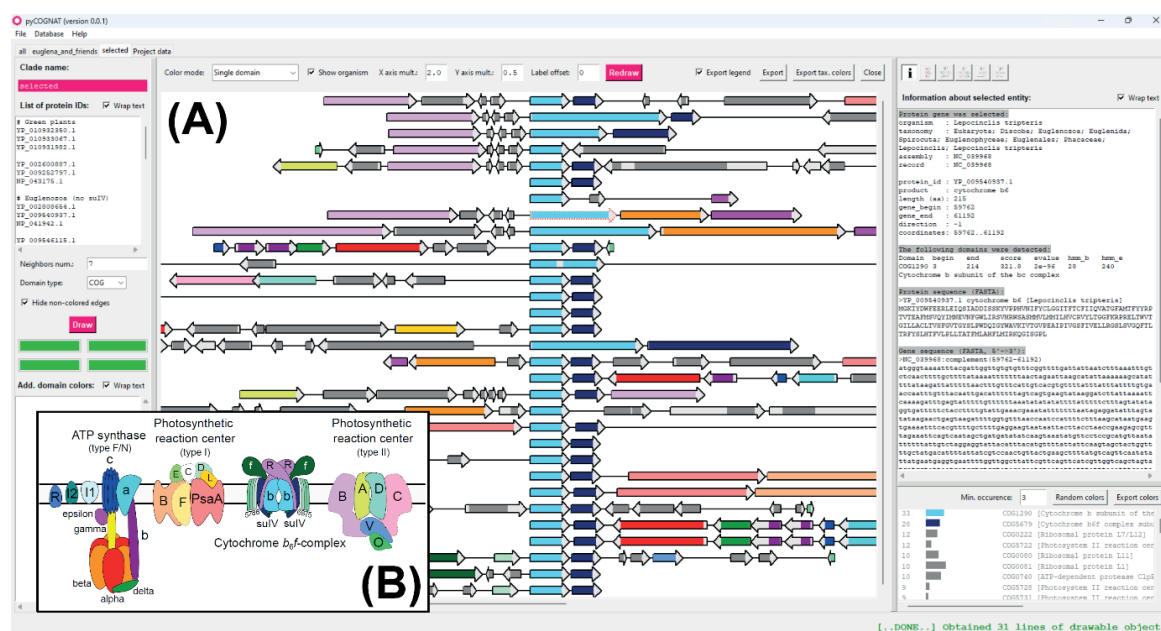
Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

School of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

udavdasha@belozersky.msu.ru; dibrova.daria@yandex.ru

Motivation and aim: Chloroplasts and plastids of modern plants result from endosymbiotic events between eukaryotic host(s) and cyanobacteria. Their genomes are large compared to mitochondria, yet many genes likely undergone a transfer to nucleus. Currently more than 15,000 complete chloroplast and plastid genomes are available, with the large predominance (> 14,000) of vascular plants chloroplasts. We present here a freely available pyCOGNAT software which is suitable for comparative analysis of any set of genomic data and a precalculated database for comparative analysis of plant organelles.

Materials and methods: The pyCOGNAT software is an interactive Tkinter-based cross-platform Python script which shows annotated genomic neighborhoods for any set of genes. The results would be more illustrative however if a set of homologous genes is used. It is based on the approach which we described before (Klimchuk O.I., Konovalov K.A., Perekhvatov V.V., Skulachev K.V., Dibrova D.V., Mulikidjanian A.Y. (2017) COGNAT: a web server for comparative analysis of genomic neighborhoods. *Biol Direct*, 12(1), 26). The pyCOGNAT requires a precalculated database of correspondence between gene products and a set of protein domains (e.g. Pfam domains or COGs) which could be obtained with the additional script from results of the HMMer search (<http://hmmer.org/>). We have calculated this database for a set of complete chloroplast genomes which can be directly used for comparative analysis with our program. The pyCOGNAT shows genomic neighborhood of each gene as arrows and depicts the same protein domains as rectangles of the same color. **Results:** An example of the pyCOGNAT output for a set of cytochromes b_6 of plant b_6f complexes is shown in Figure. The tool is highly versatile and allows to choose coloring scheme both interactively or through text input. It also exports graphical data as an SVG vector file.



(A) The main frame of the pyCOGNAT with a sample of cytochrome b_6 genes. The size of arrows is proportional to the length of genes and colored part on the arrows is proportional to the size of corresponding profile HMM hit on a gene product. Arrow tips are always non-colored. (B) The color code for key subunits of chloroplast photosynthetic complexes and rotary ATP synthase which is used in the panel A

Conclusions: The pyCOGNAT could be useful for comparative genome analysis of plant organelles. The source code of pyCOGNAT is freely available at <https://github.com/udavdasha/pyCOGNAT>. The precalculated database for plant organelles is available at <http://boabio.belozersky.msu.ru/en/tools>.

Funding: The study was conducted under the state assignment of Lomonosov Moscow State University. The original idea behind the pyCOGNAT was developed under support of the Dmitry Zimin “Dynasty” Foundation.