PIA3, a pipeline for mining assemblies for opsin protein sequences

P-01.3-15

P. Drozdova *[I,II], A. Kizenko *[III], M. Timofeyev[I,II]

[I]Institute of Biology at Irkutsk State University, Irkutsk, Russia, [II]Baikal Research Centre, Irkutsk, Russia, [III]Bioinformatics Institute, St. Petersburg, Russia

Vision is one of the most important senses in animals. The molecular system enabling vision in the absolute majority of animal eyes is based on the visual pigments consisting of opsin proteins and retinal derivatives. Multiple opsins in one species allow distinction of colors. The diversity of opsins in different extant species ranges from one to several dozens. Opsins belong to different subclasses of G-protein coupled receptors, and finding them in newly assembled genomes and transcriptomes by sequence similarity can be challenging due to numerous false positive hits. Approaches to solve this problems have already been proposed, but they are mostly delivered as a bunch of scripts requiring setup for each new user. To address this question, we re-wrote the PIA2 algorithm (https://github.com/xibalbanus/PIA2) as a Python-based pipeline, resulting in PIA3. PIA3 installation and parameter adjustment are straightforward and require minimal knowledge of bash and no knowlegde of Python programming. The pipeline requires a genome or transcriptome assembly and a database of target opsins from other species and outputs putative opsin sequences as amino acids, open reading frames and full opsin-containing contigs, as well as a phylogenetic tree showing the relationship between the database and the newly found sequences. We tested the performance of PIA3 using the endemic Lake Baikal amphipods as an example and found that different species may contain up to five diverse opsin transcripts, but all of them belong to the long wavelength-sensitive opsin subfamily. PIA3 is not restricted to opsins and can be tuned to mine for any gene family. The current version of PIA3 is available from https://github.com/AlenaKizenko/pia3_amphipod_opsins. The work was supported by the Russian Science Foundation (grant #19-74-00045). Expected to be published in: Drozdova et al. (under consideration) BMC Ecology and Evolution.

* The authors marked with an asterisk equally contributed to the work.