

Introduction to Focus Areas in Bioinformatics – WS21/22 – Week 6

Lecturer: Michael Grünstäudl

Project for week 6

- Deadline: **REPORT**: 04-Dec-2021 at 18:00; **REVIEWS**: 07-Dec-2021 at 18:00
- The entire command-line code employed as well as the project report need to be available through your FU Git repository titled "IFABI-2021" in the directory "Project Week 6" and be visible to the lecturer.

General info on project

This week's project is designed to illustrate a typical bioinformatics workflow in the process of assembling and comparing complete plastid genomes.

Goal: Assemble a plastid genome under two assembly routes and compare best contigs recovered under both routes

Your goal is to assemble the plastid genome of a water lily (*Nymphaea odorata* subsp. *odorata*) starting from raw sequence reads under two different assembly routes: with and without applying sequence depth normalization. Once plastid genome contigs have been assembled, please list the number of contigs generated and visually compare the best contigs of both assembly routes.

Important aspects

- (1.) Please download the raw sequence reads from the total DNA genome skimming of *Nymphaea odorata* subsp. *odorata* (SRR12134661) from NCBI SRA. These are your input reads for the plastid genome assembly. You are welcome to reduce this read set as displayed in the seminar video at your discretion in order to work with smaller files.
- (2.) Please use the following two NCBI RefSeq sequence records as reference genomes for the step of plastome read extraction: *Nymphaea odorata* (NC_057567) and *Nymphaea ampla* (NC_035680)
- (3.) For the actual plastid genome assembly, please employ the software NOVOPlasty (Dierckxsens et al. 2017; <https://doi.org/10.1093/nar/gkw955>).
- (4.) For the sequence depth normalization, please apply a read depth target of 200.
- (5.) **All bioinformatics steps necessary to achieve the project goal are explained and demonstrated in the seminar video of this course week.**

Deliverables

Please collect your entire command-line code for this project in a single file and upload that file together with the project report (in PDF format) to your FU Git repository. Please do not upload any sequence files to your FU Git repository. The report must also be uploaded to the Eduflow system.

- Length, format and structure of the report should be as in course week #5.
- The report should contain the figure that you generate when visually comparing the best contigs of both assembly routes (i.e., visualization of the alignment of the two contigs).

Peer-Review

The peer-review will be conducted as in course week #5.