Introduction to Focus Areas in Bioinformatics – WS21/22 – Week 5
Lecturer: Michael Grünstäudl

Project for week 5

-- Deadline: For the **REPORT**: 27-Nov-2021 at 18:00; For the **REVIEWS**: 30-Nov-2021 at 18:00
-- All files need to be available through your FU Git repository titled “IFABI-2021” in the directory “Project Week 5” and be visible to the lecturer.

General info on project

This week's project is designed to illustrate a typical bioinformatics task in the area of mining sequence records on DNA sequence databases. It consists of two goals.

Goal 1: Conduct some sequence data mining

Identify all nucleotide sequence records stored on NCBI that fulfill all of the following four criteria:
-- The records represent complete and verified plastid genomes with a sequence length between 50,000 and 250,000 bp.
-- The records have been first released by NCBI within 2020 (i.e., after 2020-01-01 and before 2020-12-31).
-- The records are stored in the nucleotide section of GenBank (or in NCBI RefSeq, respectively).
-- The raw reads of these records are stored in NCBI SRA.

Once identified via their accession numbers, extract their exact NCBI release dates in 2020 (i.e., their NCBI publication dates).

Goal 2: Visualize the results

Visualize the cumulative growth across 2020 of the absolute number of the sequence records that fulfill the listed criteria in the form of a single figure (e.g., bar chart).

Deliverables

You need to upload all source codes and the report (in PDF format) to your FU Git repository AND the report to the Eduflow system.
-- The report should be about 600-1200 words in length.
-- The report must be delivered in PDF format.
-- The report should contain the figure that you generate under goal #2.
-- The report should be structured as described by PD Dr. Tim Conrad in his course lecture on 25-Oct-2021.

Peer-Review

The peer-review will be conducted via Eduflow as specified by Tim Conrad in his course lecture on 25-Oct-2021. For access to Eduflow, please see the URL-link to Eduflow specified by Tim Conrad.