Minutes of 1st Use Case Workshop (UCWS)

Date: 2022-03-18

Agenda

Goal/ objectives for this workshop

1. Identify needs and requirements of the community in regard to the UCs
2. Prioritize working on corresponding solutions
3. Identify the relation of the UC to the NFDI4Microbiota measures & milestones

On Friday, the 18th of March 2022, a total of 32 participants attended NFDI4Microbiota‘s first Use Case (UC) workshop including NFDI4Microbiota partners as well as external participants interested in the content. The first part of the online workshop consisted of a general introduction by Prof. Dr. Thomas Clavel followed by two UC presentations, which were then separately discussed with the participants in break-out rooms.

The UC GUT was presented by Dr. Thomas Hitch on the topic „Crowd-sourcing high-quality descriptions of novel taxa“. The presentation slides and a video-recording will be made available on NFDI4Microbiota‘s webportal.

The UC MULTI was presented by Dr. Ulisses Nunes da Rocha on "Integrating genomics and transcriptomics for context-specific modeling of microbial metabolism". The presentation slides will be made available on NFDI4Microbiota‘s webportal.

Take home messages and next steps after the workshop

- **GUT**
    - Seq-Code registry & Integration/ Connection of multiple tools to Protologger
    - Chaos of naming species will continue for a couple of years
    - Ideas on how to make protologger's output more long-lasting needed (in cooperation with DSMZ).
- **MULTI**
    - relevance of SOPs for multi-omic analysis and experimental design
    - If multi-omic perspective needed in your project, define it (well)
    - We need to check, if we will be able to interconnect the data. Transformation of data as part of provenance & data quality
- **Use Cases in general**
    - Use Cases were pitched two years ago. Take these scenarios as examples. But Use Cases are current and relevant projects & contributions from our researchers
    - We continue to stay in touch with participants and inform about further developments
    - Input on Use Cases will be needed from Ambassadors. We will provide more information in the meantime
- **Collection of tools**
    - Collection of Tools should be more prominently promoted on the website, which can be implemented on the Website during the re-launch
    - Question whether a repository of tools should be channeled through NFDI4Microbiota or if we should link to https://bio.tools (tools discoveries and promotion). NFDI4Microbiota partners support the inclusion of tools but a filtering through the eye of the NFDI4Microbiota is needed. Rebuttal: Communities exist on bio.tools to provide a curated view of tools (https://bio.tools/communities)

Summary of the questions and discussion points asked during the presentation and break-out sessions

MULTI

- Q: What is the bottleneck of the integration for MULTI?
  A: We need SOPs for computational and wetlab experiments.
- Q: Has the data been submitted already? Can the connection of data be achieved yet?
  A: Everything has been submitted to NCBI or ENA which requires quality data and annotation.
- MULTI needs input from participants on: general needs, what are researchers currently working with (datasets, gaps of pipelines/ tools, struggles of the community). This information will be used to develop easier and accessible solutions for the community.
- MULTI needs the following input from NFDI4Microbiota: An integrated platform and/ or the promotion of tools and pipelines. The possibility to interactively search data sets (interconnectedness).
- Integration of multi OMICS - points of discussion:
  - Ullisses opinion is that experimental design is the most critical part: Very skilled at lab work but don‘t have equipment for multi OMICS analysis i.e. transcriptomics, library preparation, sequencing. Question that preceds the experimental design is „What do you want?/Which part of the data actually needs to be integrated?“
  - There is a need for users to ask the HelpDesk/ Teaching materials for advice on how to frame needs for outsourcing analysis/protocols to companies. The goal would be to train people on how to get their solution and showing different approaches.
  - SOPs for both wet and dry lab are needed and advices on which SOPs to use depending on the scientific questions.
  - Need for statistical advices (e.g., transformations) regarding the integration of omics. Especially because there is not consensus on which tools to use.
  - Potential solution: Review/ benchmark such tools and establish gold standards to help generate recommendations to help researchers (similar to CAMI; https://doi.org/10.1101/2021.07.12.451567).

GUT

- Q: Is protobiome an extension of Protologger?
  A: It is an application of Protologger. We provide the output of taxanomic and functional descriptions and interested experts can interpret the data. The current main foucs will be to engage with the community.
- Q: Protologger is a descriptive tool. Is it possible to combine the tool with redbiom or other tools? How can it relate to the project redbiom?
  A: At this time point it is unclear, how this could be achieved. The primary goals and limitations are to have Protologger and it‘s outputs hosted, searchable, and preserved for the longterm.

Other tools suggested by participants on the topic of taxonomic and functional description of microbiomes is redbiom, which was mentioned during the general discussion. It is possible to use Prokka as an annotation tool and convert the results to KEGG annotation.

How can the community be involved and what would their specific contributions look like?
The aim is to **work with experts from different environments (**e.g. gut, soil, marine expertise) and identify as well as extract MAGs and SAGs that are suitable to be analyzed by Protologger. If suitability is confirmed the analysis can be run to create ‘protologues’ and submit the data. In order to validly name bacteria, the help from the community is needed. The main focus is currently to collect input from ecological experts who know about existing isolates instead of using ANI-values to categorize species.
Specifically, the strategy is to focus on the human gut (based on the personal expertise of Tom Hitch) and continue to expand on taxa descriptions and specifications. From the attending participants of the discussion round, two were experts and able to contribute on the gut isolates, while one is focusing on single organisms. Another specializes in culture and identification without restrictions, while another is knowledgeable on marine bacteria and mice as well as human gut. Additionally, by connecting with NFDI4Biodiversity it might be possible to access the personal stash of retired researchers and deposit their isolates and cultures.
To validly name an isolate, it has to be **deposited in two culture collections** and the sequences published, so that it can be accessed by the the IMNGS tool (https://imngs.org) for further analysis.

The **current cataloging mechanism** has to be improved. It is not discernible if researchers are describing based on MAGs or other. Issues such as data security and cross-referencing with already submitted genomes is currently not possible. The current function is to provide the output of the analysis.

The **seq-code registry** is currently not implemented in Protologger. In the future many problems will arise since isolates published in different registries will have multiple names. The question of which name will be prioritized is unclear, which is why there will be chaos for years to come concerning the names of bacteria. The cultivation community will have a big role to play in navigating this problem.

The question of the **output integration**s has been discussed and the question of being able to use codes interchangeably has arisen, which is problematic when working with both isolates and MAGs.

To increase **Protologger‘s accessibility to the community** a curated database is needed. The idea is to use BACDive and have the overview file of Protologger linked, make the output searchable and connect it to the environment. Additionally there is the need for long-lasting deposition.