GenTreat

Pipeline for hybrid assembly of the prokaryotic genomes

User Guide

version 1.0

SUMMARY

PRESENTATION **3**

LIST OF DEPENDENCES AND THEIR INSTALLATIONS **3**

RUNNING GenTreat **5**

THE SOFTWARE **5**

THE PRINCIPAL WINDOW 5

ABOUT WINDOW 6

CREATING A PROJECT 7

RUNNING AGAIN AND ERASING A PROJECT 12

RESULTS **14**

PRESENTATION

Welcome to the user guide of GenTreat - pipeline for hybrid assembly of the prokaryotic genomes. This software was developed in JAVA and its execution must be done in Linux operating systems, the tests performed to validate this pipeline were using the Ubuntu 20.04.4 LTS distribution (GNU/Linux 5.4.0-100-generic x86_64) and Linux Mint 20.2 Cinnamon (Version: 5.0.7). Below is the list of dependencies and how to install them.

LIST OF DEPENDENCES AND THEIR INSTALLATIONS

Many of these dependencies require the user to have superuser privileges, so we advise installing with sudo.

1. To run the tool it is necessary to have the Java virtual machine installed on the operating system, version 17 or higher.
    1.1. Download link: https://www.oracle.com/java/technologies/downloads/.

After installation you can verify that everything is correct by running the test below:

```
(base) allan@allan:~$ java --version
java 17.0.2 2022-01-18 LTS
Java(TM) SE Runtime Environment (build 17.0.2+8-LTS-86)
Java HotSpot(TM) 64-Bit Server VM (build 17.0.2+8-LTS-86, mixed mode, sharing)
(base) allan@allan:~$
```

Figure 1. Checking the Java version.

2. Operating system dependencies:
    2.1. sudo apt-get install python3
    2.2. sudo apt install build-essential
    2.3. sudo apt install cmake
    2.4. sudo apt install python3-distutils
    2.5. sudo chmod 777 /opt/ (Optional, you can choose to create each software folder and give permission only on it)
    2.6. sudo apt install zlib1g
    2.7. sudo apt-get update -y
    2.8. sudo apt-get install -y gzip
    2.9. sudo apt-get install -y bzip2
    2.10. sudo apt-get install libdatetime-perl libxml-simple-perl libdigest-md5-perl git default-jre bioperl

2.11. sudo cpan Bio::Perl

2.12. sudo cpan Bio::SearchIO::hmmer3

2.13. sudo apt install mummer

2.14. sudo apt install python2

2.15. sudo apt install ncbi-blast+

2.16. sudo apt install python-is-python2

3. Installation of SPAdes, Megahit, CISA and Prokka.

3.1. **Installation of SPAdes:**

3.1.1. cd /opt

3.1.2. wget http://cab.spbu.ru/files/release3.14.1/SPAdes-3.14.1-Linux.tar.gz

3.1.3. tar -xzf SPAdes-3.14.1-Linux.tar.gz

3.1.4. mv SPAdes-3.14.1-Linux SPAdes

3.2. **Installation of Megahit:**

3.2.1. cd /opt

3.2.2. wget https://github.com/voutcn/megahit/releases/download/v1.2.9/MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz

3.2.3. tar zvxf MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz

3.2.4. mv MEGAHIT-1.2.9-Linux-x86_64-static megahit

3.3. **Installation of Prokka:**

3.3.1. cd /opt

3.3.2. git clone https://github.com/tseemann/prokka.git

3.3.3. /opt/prokka/bin/prokka --setupdb

Note: The tools must be in the /opt directory and named as follows: SPAdes, megahit, prokka.

RUNNING GenTreat

1. Download the GenTreat zipped file.
2. Save it in the directory of your choice.
3. Unzip the GenTreat package. (Optionally, if you want to perform this step in the terminal, use the command unzip file.zip).
4. Change the permissions on the software directory, it is advisable to have full permissions:

   Exemplo: chmod -R 777 <Diretório>

   4.1. chmod -R 777 /home/gislenne/Downloads/GenTreat.
5. Run GenTreat in its directory:

   5.1. Execution example:

```
java -jar GenTreat.jar
```

Note: As a preference, keep reads and references files in the same directory.

THE SOFTWARE

1. THE PRINCIPAL WINDOW

When starting GenTreat, the window described in Figure 2 will be shown. It is composed of a table containing the list of projects previously created by the user with their respective status, name and creation date. In the same window, there is the Project and About menu.

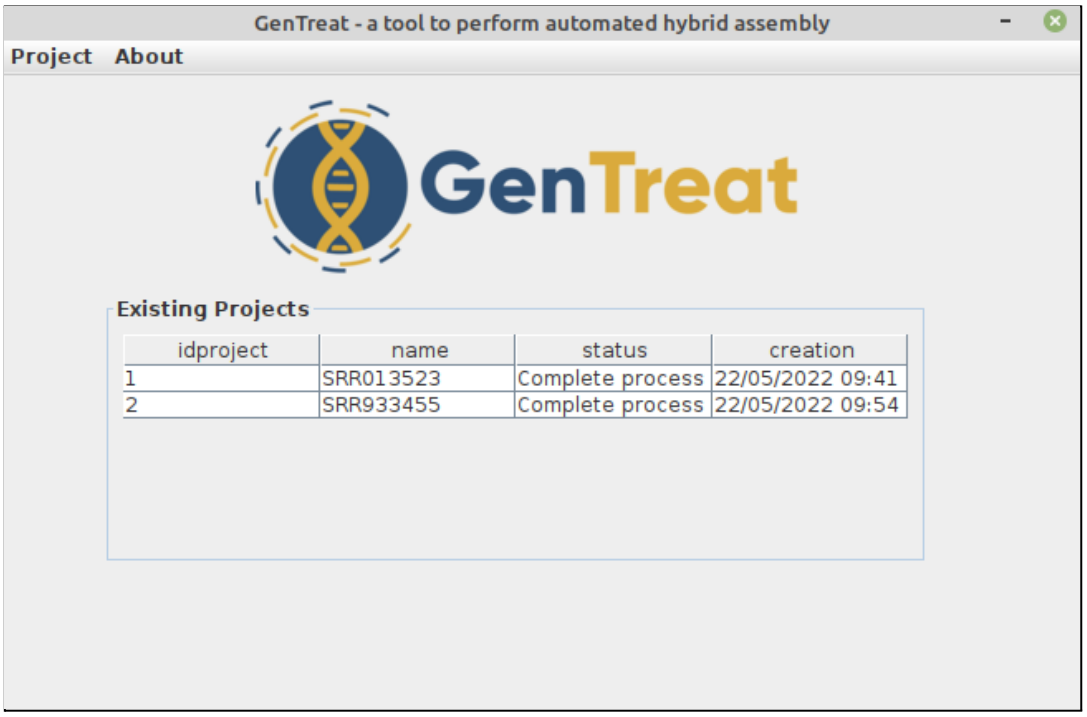

Figure 2. GenTreat main window: This window displays the list of projects previously created by the user.

2. ABOUT WINDOW

Information about the software developers is found in the “About” tab, as shown in Figure 3.

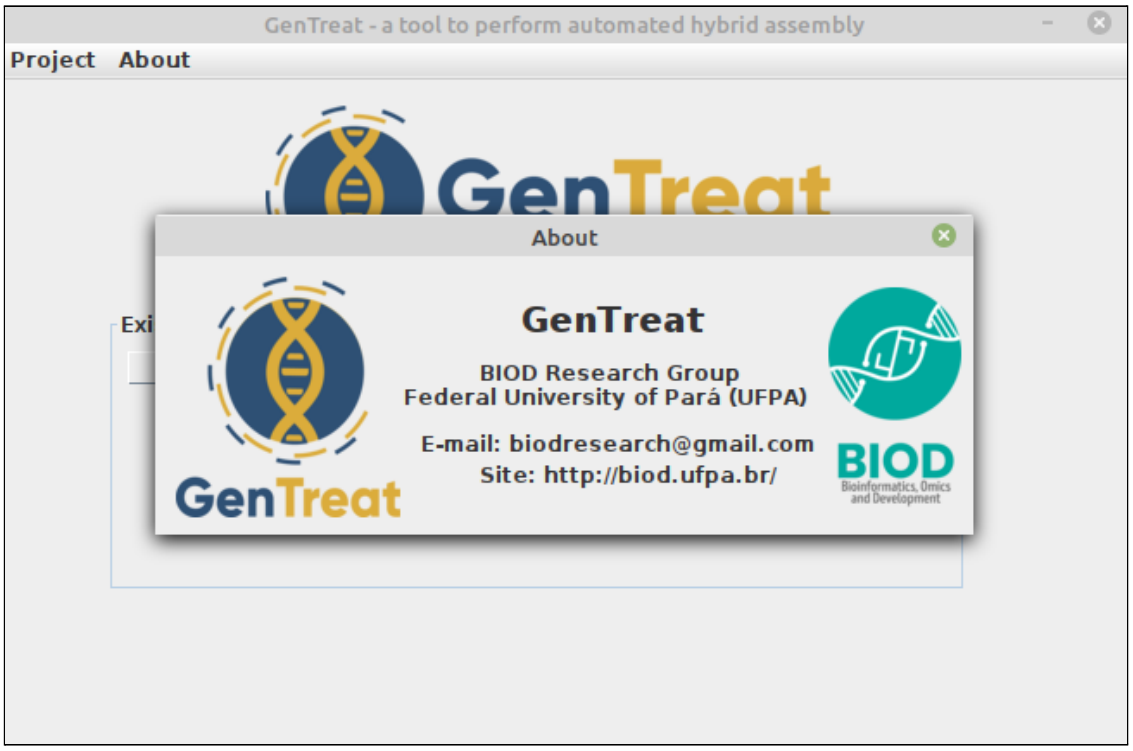

Figure 3. About window.

3. CREATING A PROJECT

To create a project, click on the “Project” menu, then on the “Create" submenu. A window will appear, as shown in Figure 4, to insert the project name (without spaces or accents). After insertion, click on the “Next” button.

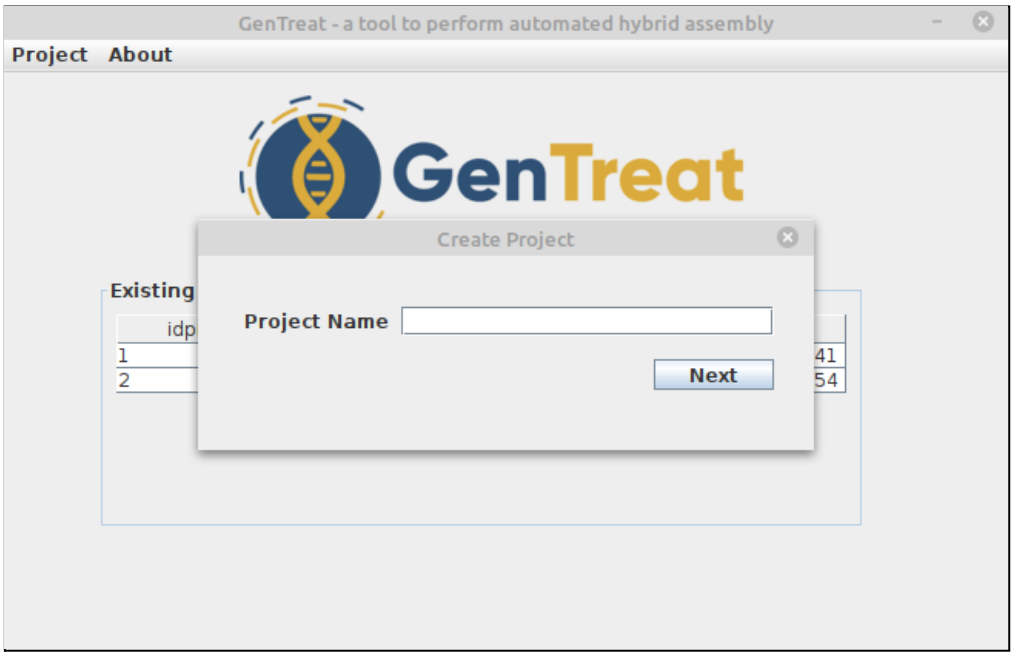

Figure 4. Window to create a project. In this window, the user will inform the name of the project to be executed in GenTreat.

On the next window (Figure 5), it will be possible to inform the parameters of the SPAdes and Megahit tools. Note that the default parameter values for each tool are shown. All can be changed if desired by the user.

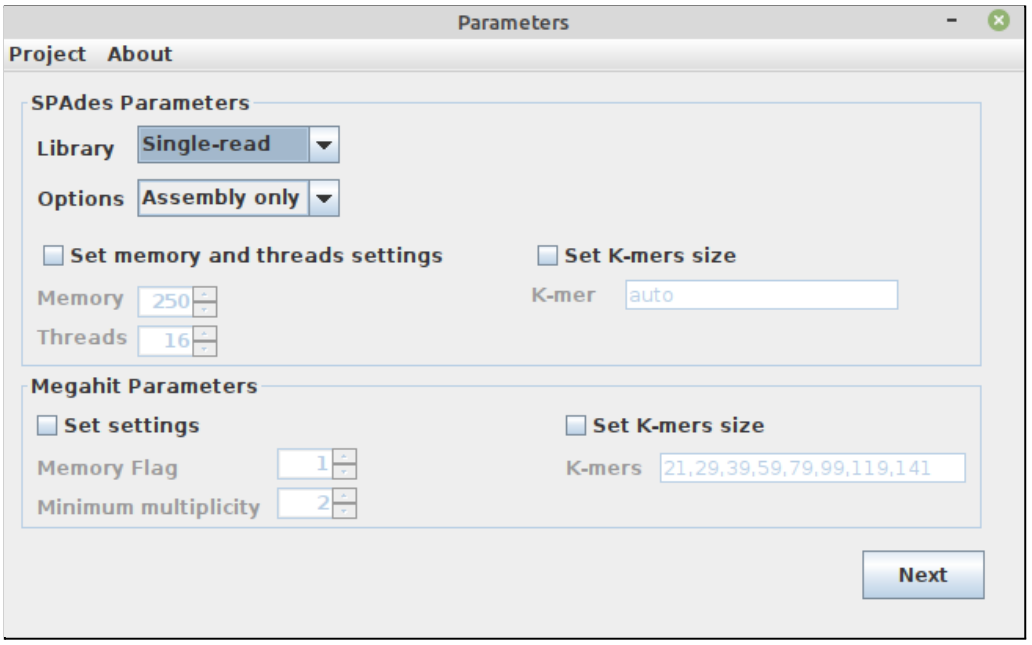

Figure 5. Window to edit parameter values for assemblers.

Here the type of genomic library, Single-end or Paired-end, assembly options, minimum multiplicity for filtering, memory, threads and kmers settings must be informed. After entering the data, click on the “Next” button.

In the next window (Figure 6), there are the parameters of the CISA and Prokka tools, which can be adjusted by the user, the default values of each tool are shown in the window, the option of ordering results using Mauve is also displayed in this window.

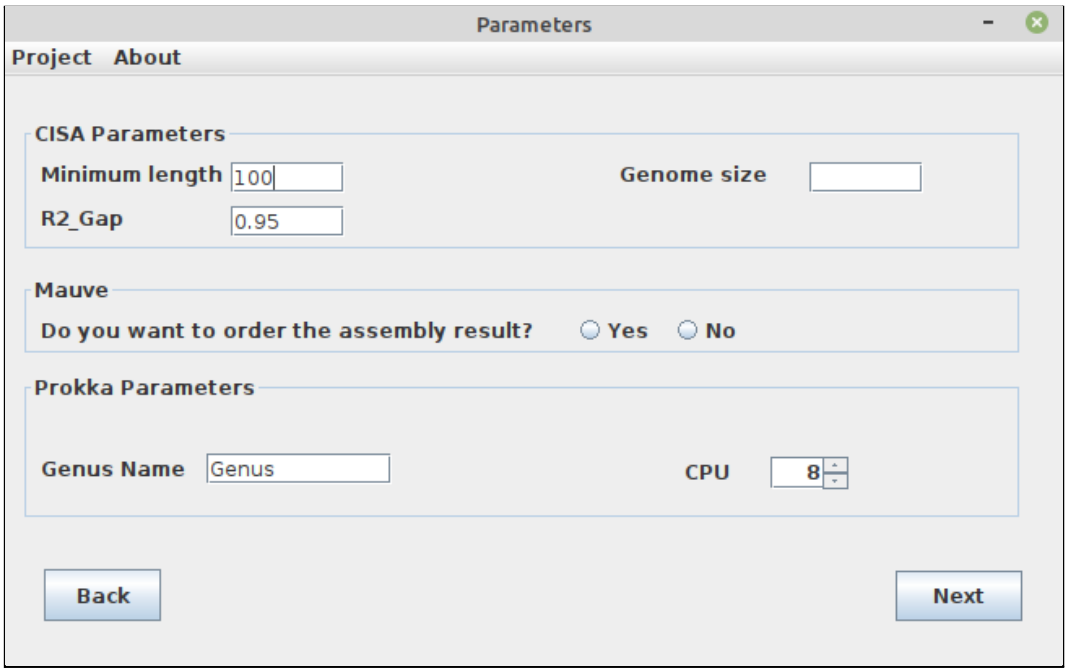

Figure 6. Adding parameter values from CISA, Mauve and Prokka tools.

On this screen, you must inform the minimum length of the contigs, R2_Gap, expected genome size, the option of ordination process, Genus Name and the number of CPUs to be used. After that, click on “Next”.

The next step is to add the input files (Figure 7). If the genomic library chosen, on the screen of figure 5, is single-end and the option “Yes” is selected for ordination on the screen of figure 6, a field will appear to add the dataset containing the raw reads in FASTQ format and another for the reference file in Genbank (GB) format.

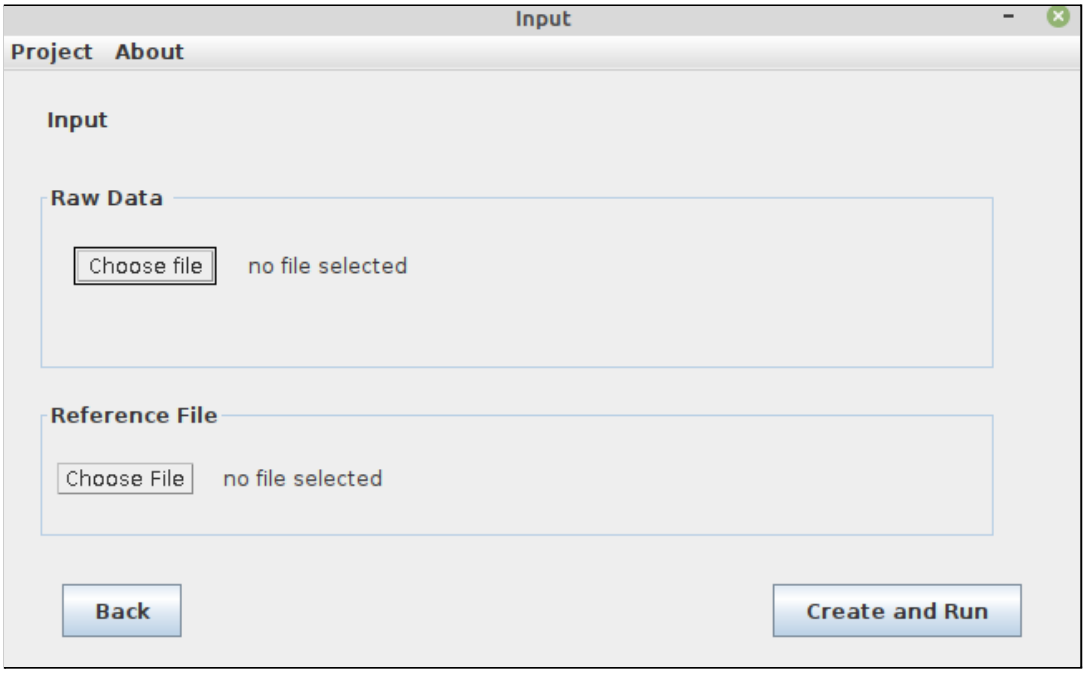

Figure 7. Adding the raw read and reference files.

If the user chooses not to order the assembly result, the option to insert the reference file will not be displayed, as shown in Figure 8.

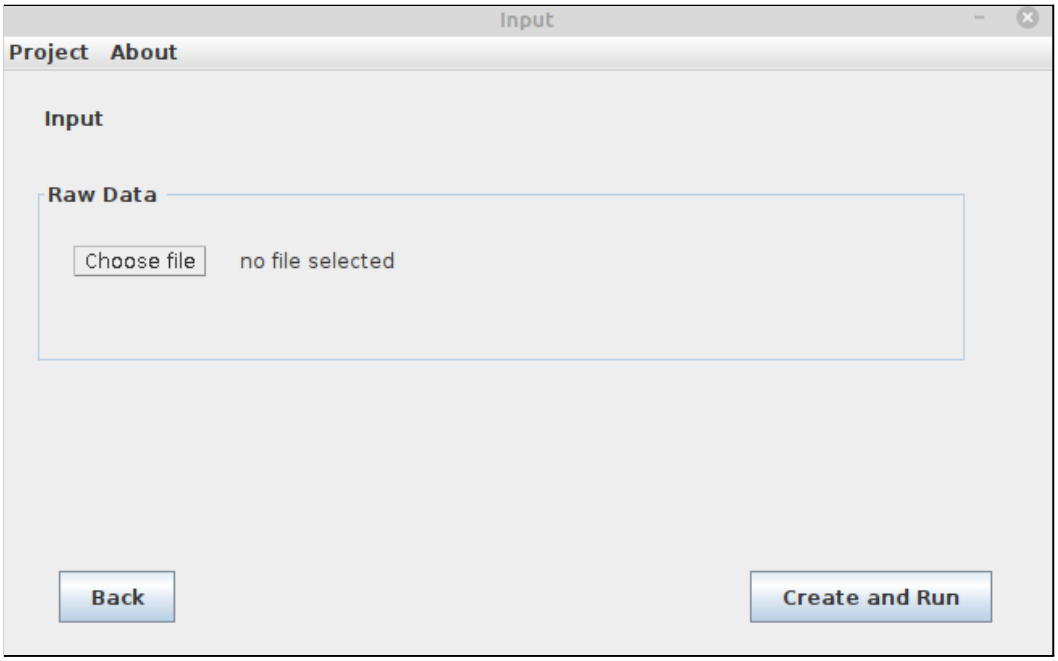

Figure 8. Adding the raw read file.

If the genomic library chosen, on the screen of figure 5, is paired-end and the option "Yes" is selected for ordination on the screen of figure 6, a field will appear to add the two raw reads and another for the reference file (Figure 9).

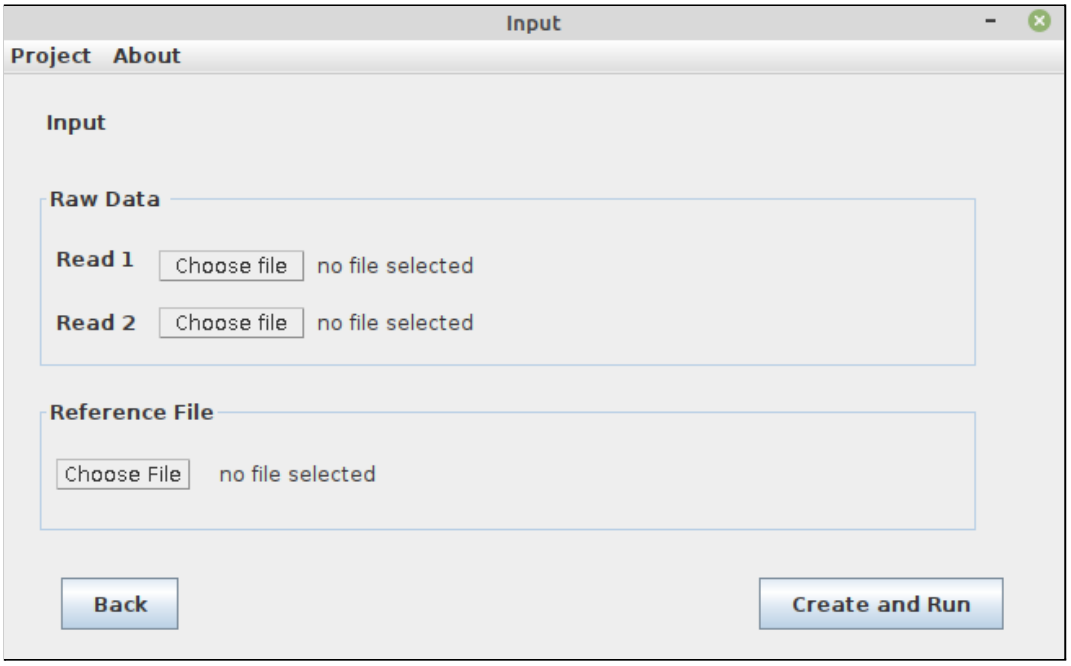

Figure 9. Adding the raw reads and reference files.

If the user chooses not to order the assembly result, it happens in a similar way to the one described for single-end reads, shown in Figure 10.

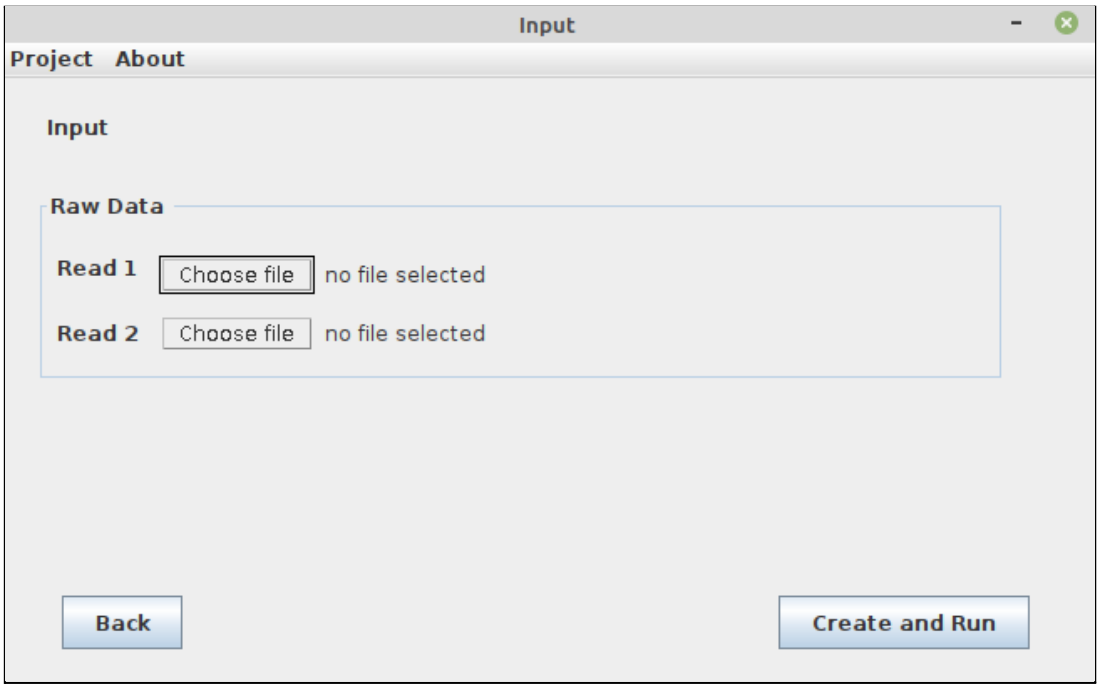

Figure 10. Adding raw read files.

After that, click on the “Create and Run” button to start the process. The Log screen will open, as shown in Figure 11.

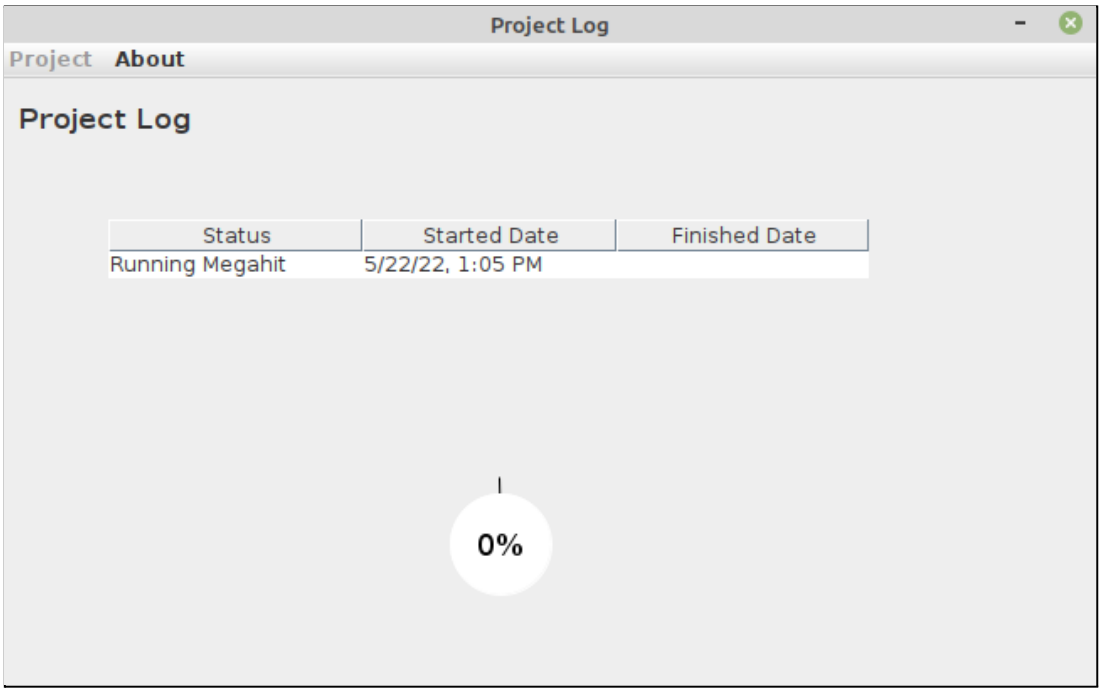

Figure 11. Project Log screen.

When the process is finished, a window will open informing where the results are, as shown in Figure 12.

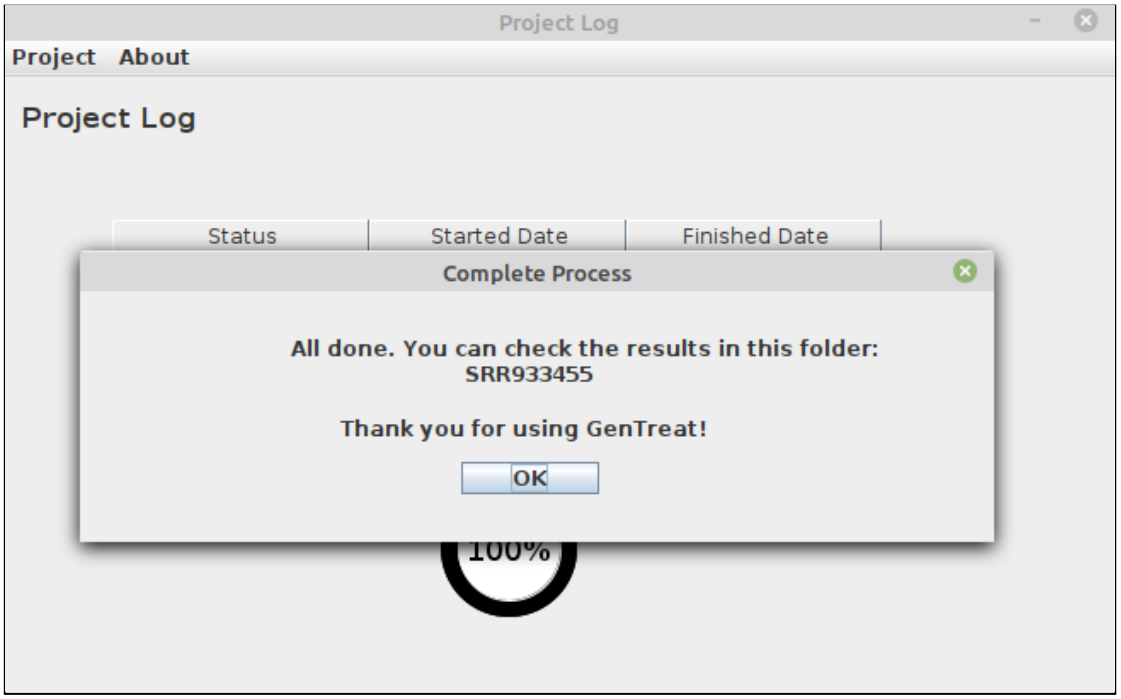

Figure 12. Complete Process screen.

4. RUNNING AGAIN AND ERASING A PROJECT

In GenTreat, there is the possibility to delete a project from the database and also to continue the process if it has not finished for some reason.

On the initial window, in “Existing Projects”, figure 13, all the projects created that are contained in the database will appear, as well as their status. Then, just click on the project you want to continue executing and press the “Continue” button. To delete, just click on the “Delete” button.

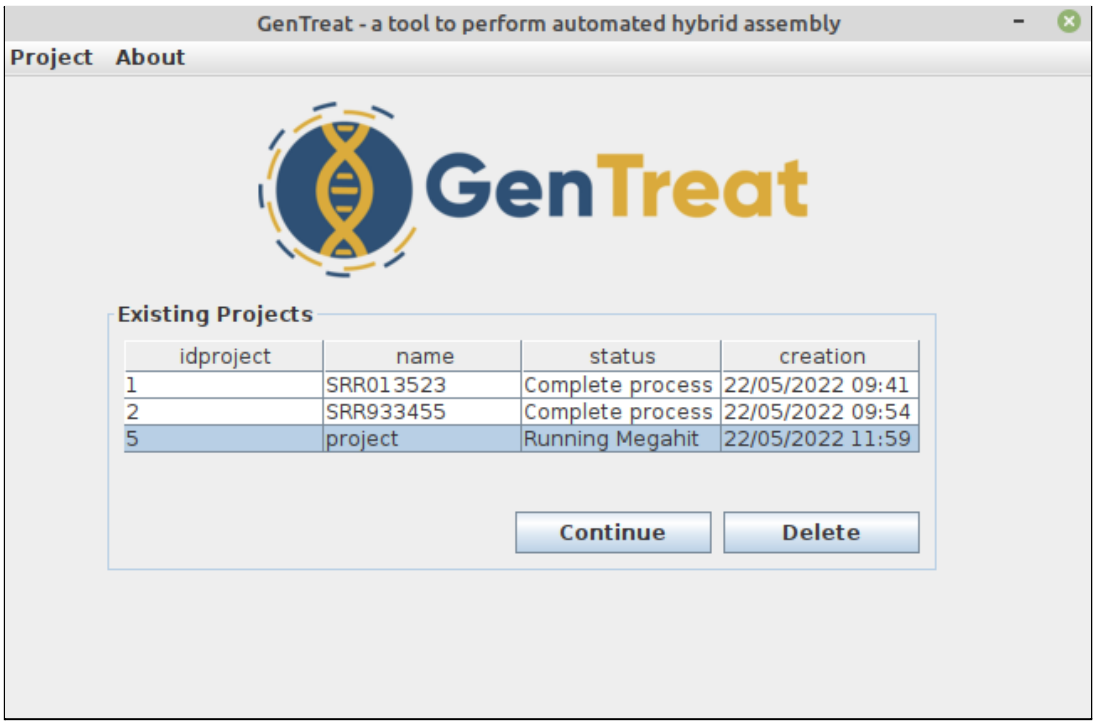

Figure 13. Home screen: selecting an existing project.

Moreover, even if the project is finished, it can be executed again, being necessary only to select the project that has the “Complete Process” status and click on the “Run Again” button, as shown in Figure 14.

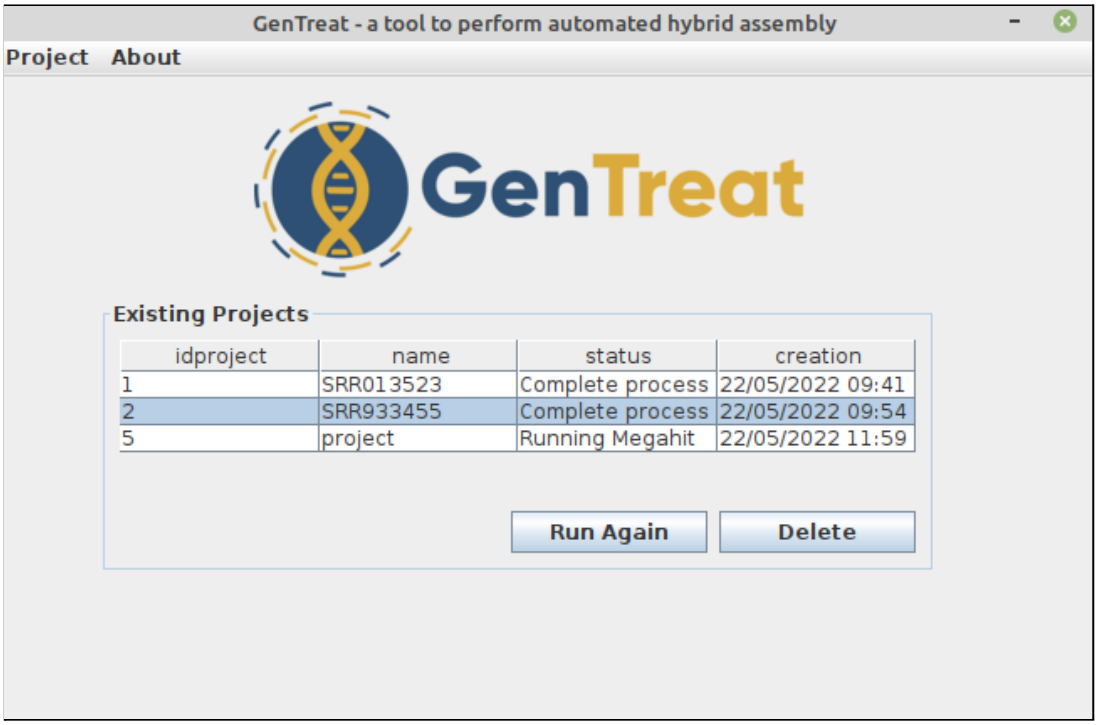

Figure 14. Home screen: selecting an existing project.

When clicking on the "Run Again" button, if the previously completed project folder still exists, a message will appear asking if you want to overwrite it or create a new folder, as shown in Figure 15.

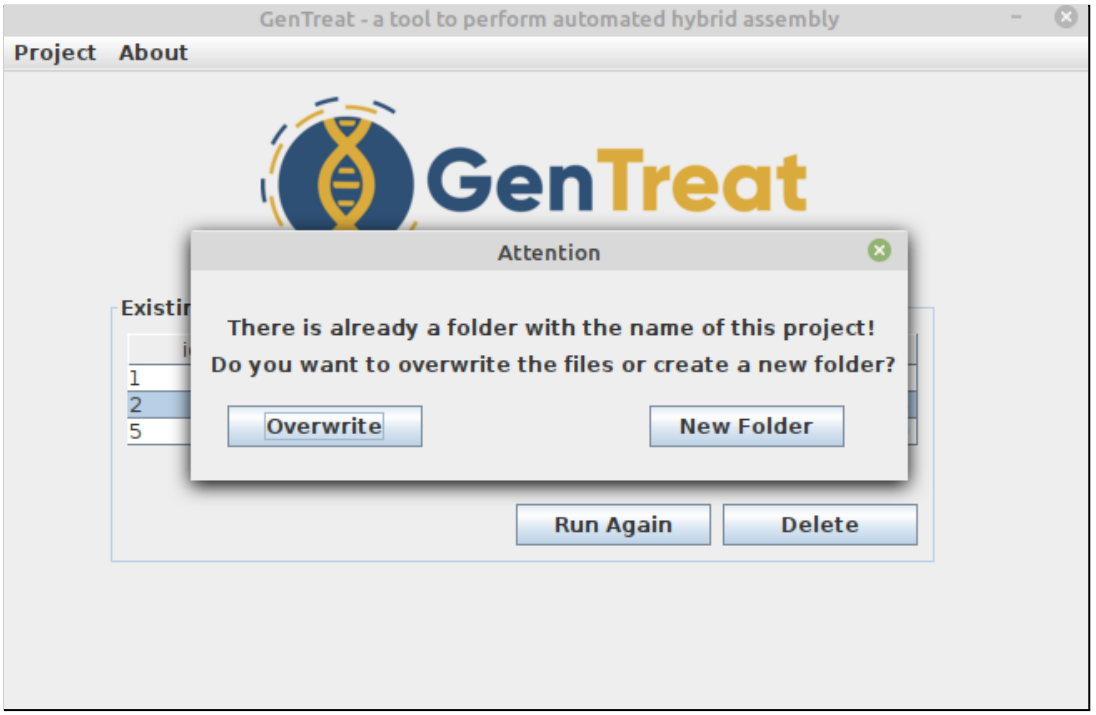

Figure 15. Running a project again.

If the “Overwrite” button is clicked, the folder will be overwritten and the Log screen will be shown, as in figure 11.

However, if the “New Folder” button is clicked, a field will appear to write the name of the folder where you want the results to be saved, as shown in Figure 16. After that, just press “OK” and you will be directed to the Log screen (Figure 11).

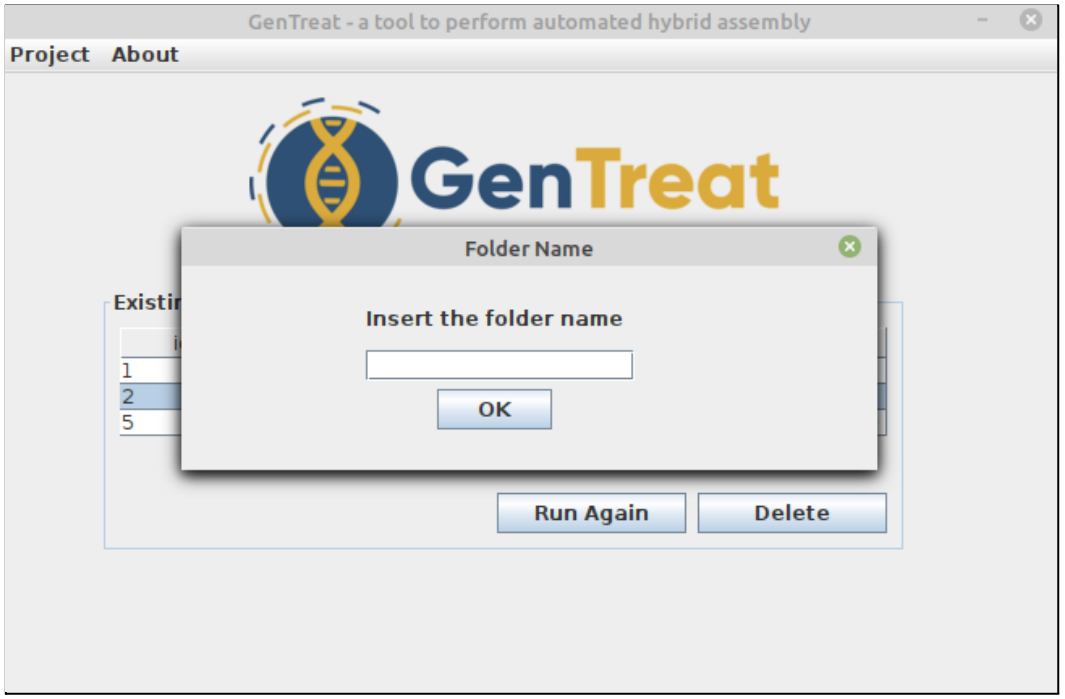

Figure 16. Creating a new folder for project results.

RESULTS

Figure 17 shows the results that are generated after processing: In the spades-assembly folder will be the assembly result of the SPAdes tool, in the megahit-assembly folder will be the assembly result of the Megahit tool, the CISA folder contains the assembly results joined , in the Mauve folder will be the ordered result (if the user has selected the order option) and in the Prokka folder, the annotation result. In addition, there is the GenTreat.fasta result, which is the assembly result that has been joined and ordered (if selected). Finally, the log.txt file, with the log of all tools.

Figure 17. Results.

BIOD

Bioinformatics, Omics
and Development