Tutorial

Analysing a single sequence file.

Step 1:

Download the tutorial datasets from one of the following links:
Google Drive: http://tinyurl.com/m2xgdpw
Dropbox: http://tinyurl.com/kz77ps5

Step 2:

Open Galaxy and create an account.

- Use the web browser to navigate to: http://10.42.1.212:8080/.
- Click on “you may create one” to open the account creation form.
- Fill in the form and click on submit.

Step 3:

Load the test data into the history.

- Click on the “up” arrow in the upper left corner (next to the “tools” menu).
- Select the “IonXpress_004.fastq” and “primers.csv” files from the test set.
  - Manually set the format for the primer file to “csv”.
- Click on “start” to submit the data.

Step 4:

Demultiplex the sequence file.

- Select the “Split on primer” tool from the tool menu (under “Cluster and Identification Tools”).
- Fill in the form (most settings are already filled in by default).
  - Set the input type to “FASTQ” (matching our uploaded sequence file).
  - Select the sequence file (“IonXpress_004.fastq”).
  - Set the primer type to “csv”.
  - Select the primer file (“primers.csv”)
  - Set both mismatches to “1”.
  - Check the “trim primers box”.
- Click on “Execute”.

Step 5:

Analyse and filter / trim the sequence file.

- Select the “Sequence analyser” tool from the tool menu.
- Fill in the form.
  - Set the file type to “Single”.
  - File format to “FASTQ”.
  - Input file to “IonXpress_004.fastq Demultiplexed (ITS1)”.
- Click on “Execute”.
- Display the “IonXpress_004.fastq Demultiplexed (ITS1) Graphs” result file by clicking on the “eye” symbol right of the file name.

- Check the graphs for indicators of sequence quality such as:
  - Peaks in the sequence lengths.
  - Drop of in quality scores on the 3' end of the sequences.
  - Distribution of quality scores.
- Select the “Sequence trimmer” tool from the tool menu.
- Fill in the form.
  - Set the file type to “Single”.
  - Input file format to “FASTQ”.
  - Output file format to “FASTA”.
  - Input file to “IonXpress_004.fastq Demultiplexed (ITS1)”.
  - The following filter settings:
    - Trim to length to “360”.
    - Minimum read length to “300”.
    - Minimum mean quality to “27”.
- Click on “Execute”.

Step 6:

Cluster and filter the sequences.

- Select the “Cluster and Filter” tool from the tool menu.
- Fill in the form.
  - Set the file type to “Single”.
  - Input file to “IonXpress_004.fastq Demultiplexed (ITS1) Filtered”.
  - Cluster threshold to “0.97”.
  - Minimum cluster size to “2”.
- Click on “Execute”.

Step 7:

Identification of the clusters.

- Select the “BLAST” tool from the tool menu.
- Fill in the form.
  - Set the file type to “Single”.
  - Input file to “IonXpress_004.fastq Demultiplexed (ITS1) Filtered Clustered”.
  - Select “Local BLAST”.
  - Select “GenBank nucleotide database” as the Local BLAST database.
  - Set the maximum number of BLAST hits per sequence to “10”.
  - Minimum BLAST hit percentage to “97”.
  - Minimum BLAST hit coverage to “100”.
  - Maximum e-value to “0.05”.
- Click on “Execute”.

- View the results file ("IonXpress_004.fastq Demultiplexed (ITS1) Filtered Clustered BLAST") by either:
  - Clicking on the "eye" symbol next to the output file.
  - Or downloading the file and opening it in a spreadsheet program. To download the file:
    - Open the file box by clicking on the name.
    - A "floppy" symbols appears under the name which allow the file to be downloaded.

Analysing a ZIP archive.

Download the tutorial file and open Galaxy (see step 1-2 from the first part of the tutorial).

Step 1:

Create a clean history.

- Click on the "Cog" symbol on the upper right corner.
- Select "Create New" to create a new empty history.

Step 2:

Upload the zip archive.

- Click on the "up" arrow in the upper left corner (next to the "tools" menu).
- Select the "IonXpress_Data.zip" and "primers.csv" files from the test set.
  - Manually set the format for the primer file to "csv" and the format for the IonXpress_Data file to "zip".
- Click on "start" to submit the data.
- You can view the data in the ZIP by selecting the "Manage ZIP" tool.
- Fill in the form.
  - Select the "Display ZIP content" option.
  - Select the "IonXpress_Data.zip" file as the input file.
- Click on "Execute".
- Display the contents by clicking on the "eye" symbol for the "Log file" output.

Step 3:

Demultiplex the sequence file.

- Select the "Split on primer" tool from the tool menu.
- Fill in the form.
  - Set the input type to "ZIP".
  - Select the sequence ZIP ("IonXpress_Data.zip").
  - Set the primer type to "csv".
  - Select the primer file ("primers.csv")
  - Set both mismatches to "1".
  - Check the "trim primers box".
- Click on "Execute".

Step 4:

Separate the primers in the demultiplexed ZIP file.

The output ZIP created in step 3 contains the sequence files from both primers and the unmatched files. These need to be split up prior to filtering and trimming due to the expected length and possibly quality differences of the different primers.

- To view the different output files, select the “Manage ZIP” tool.
- Fill in the form.
  - Select the “Display ZIP content” option.
  - Select the “IonXpress_Data.zip Demultiplexed” file as input.
- Click on “Execute”.
- Display the contents by clicking on the “eye” symbol for the new “Log File”.
- Select the “Manage ZIP” tool again.
- Fill in the form.
  - Select the “Create subset from ZIP” option.
  - Select the “IonXpress_Data.zip Demultiplexed” file as input.
  - In the text field you can either specify the precise names or use regular expressions. To match the ITS1 files type: “*ITS1*”.
- Click on “Execute”.
- A new ZIP file is created in the history that contains the ITS1 sequences.
- In order to check that the new ZIP file indeed contains the ITS1 sequences, select the “Manage ZIP” tool once more.
- Fill in the form.
  - Select the “Display ZIP content” option.
  - Select the “Zip file” file as input.
- Click on “Execute”.
- Display the contents by clicking on the “eye” symbol for the new “Log File”.

Step 5:

Analyse and filter / trim the sequence file.

- Select the “Sequence analyser” tool from the tool menu.
- Fill in the form.
  - Set the file type to “ZIP archive”.
  - File format to “FASTQ”.
  - Input file to “Zip file”.
- Click on “Execute”.
- Display the “Zip file Graphs” result file by clicking on the “eye” symbol right of the file name.
- Check the graphs for indicators of sequence quality such as:
  - Peaks in the sequence lengths.
  - Drop of in quality scores on the 3' end of the sequences.
  - Distribution of quality scores.
- Select the “Sequence trimmer” tool from the tool menu.
- Fill in the form.
  - Set the file type to “ZIP archive”.
  - Input file format to “FASTQ”.
  - Output file format to “FASTA”.
  - Input file to “Zip file”.

    - The following filter settings:
        - Trim to length to “360”.
        - Minimum read length to “300”.
        - Minimum mean quality to “27”.
- Click on “Execute”.

Step 6:

Cluster and filter the sequences.

- Select the “Cluster and Filter” tool from the tool menu.
- Fill in the form.
    - Set the file type to “ZIP archive”.
    - Input file to “Zip file Filtered”.
    - Cluster threshold to “0.97”.
    - Minimum cluster size to “2”.
- Click on “Execute”.

Step 7:

Identification of the clusters.

- Select the “BLAST” tool from the tool menu.
- Fill in the form.
    - Set the file type to “ZIP archive”.
    - Input file to “Zip file Filtered Clustered”.
    - Select “Local BLAST”.
    - Select “GenBank nucleotide database” as the Local BLAST database.
    - Set the maximum number of BLAST hits per sequence to “10”.
    - Minimum BLAST hit percentage to “97”.
    - Minimum BLAST hit coverage to “100”.
    - Maximum e-value to “0.05”.
- Click on “Execute”.
- View the results file (“Zip file Filtered Clustered BLAST”) by either:
    - Unpacking the ZIP file in Galaxy.
        - To unzip the file select the “Manage ZIP” tool.
        - Select the “Unpack items from ZIP” option.
        - Set the “Zip file Filtered Clustered BLAST” file as input.
        - Leave the text form empty (if empty it unpacks everything).
        - Click on “Execute”.
        - The output BLAST files appear in the history.
        - Click on the “eye” icon to display a BLAST file in the browser.
    - Or downloading the ZIP file, unzipping it on your computer and opening the files in a spreadsheet program. To download the file:
        - Open the file box by clicking on the name.