

MCRL - Metagenomic Clustering by Reference Library

SOFTWARE INSTALLATION INSTRUCTIONS

Contents

1. Introduction	2
2. Files provided	2
3. Installation & system requirements	3
3.1 Windows distribution	3
3.2 Linux distribution	4
3.3 Running MCRL from MATLAB	5
4. Troubleshooting	6
5. Plotting reference gene networks	8
5.1 Configuration file	8
5.2 Creating plots from MATLAB	9
5.3 Creating plots using the Windows distribution	9
5.4 Creating plots using the Linux distribution	10
5.5 Demo plot	10
6. Support	10

1. Introduction

MCRL is a data mining tool that can be used to probe a metagenome for homologs of a pre-defined reference library. The input to MCRL is an assembled metagenome in nucleotide or amino acid format and a library of reference sequences in amino acid format. MCRL will then perform iterative clustering of the reference library with respect to the given metagenome and provide as output a list of nonredundant reference genes that have homologous counterparts in the metagenome. The MCRL user interface allows users to download the most recent version of the viral RefSeq database and assemble a reference library of viral genes, however, MCRL can be used with any user-defined reference library (see publication for further details).

MCRL can be downloaded from <https://github.com/a-tadmor/MCRL>.

2. Files provided

MCRL_readme_vXXX.pdf	This readme file
MCRL_vXXX_MATLAB.zip	MATLAB sources (zip)
MCRL_vXXX_MATLAB_Linux64.tar.gz	MATLAB sources (tar)
MCRL_vXXX_installer_WIN64.exe	Windows distribution (mrc environment & MCRL)
MCRL_vXXX_Linux64.tar.gz	Linux distribution (MCRL)
Plot_networks_vXXX_WIN64.exe	Windows executable to plot reference gene networks
Plot_networks_vXXX_Linux64.tar.gz	Linux executable to plot reference gene networks
license.txt	MCRL license

3. Installation & system requirements

MCRL can be run on Windows or Linux operating systems as executables or through MATLAB. For demonstration purposes, MCRL comes pre-installed with a demo mini-metagenome and demo mini-reference library. The user can test run the default demo to see that MCRL is properly installed. After MCRL has been installed and the user interface loads, the user can download and install the latest viral RefSeq library from NCBI by pressing the button labeled “Download and assemble the latest viral RefSeq reference library”.

3.1 Windows distribution

Requirements:

1. 64-bit Windows
2. Internet connection and administrator privileges

Instructions:

1. Download and install Visual C++ Redistributable (required for DIAMOND) from: <https://support.microsoft.com/en-us/help/2977003/the-latest-supported-visual-c-downloads>
2. Download `MCRL_vXXX_installer_WIN64.exe` to a **local** directory such as `d:\tmp`.
3. Double click the executable. This executable will automatically install the MATLAB Compiler Runtime (mcr) library and copy the MCRL installation files to a folder designated by the user. Please select here as well a **local** directory such as `d:\MCRL`.
4. To complete the installation, navigate to the `application` folder and double click the executable `MCRL.exe`. Follow the instructions and prompts to install blast and MCRL (accept default entries when prompted). Depending on your system it may take a couple of minutes for the installation to begin. **Note: prompts requiring user response can be minimized in the taskbar.**

Once the main interface of MCRL loads you may start using MCRL.

3.2 Linux distribution

Requirements:

1. 64-bit Linux*
2. Internet connection
3. Root privileges (required for installing the MATLAB Compiler Runtime (mcr) library)

*MCRL has been tested on CentOS Linux 7

Instructions:

1. Download and install the MATLAB Compiler Runtime (mcr) library from:
<https://www.mathworks.com/products/compiler/mcr/index.html>
2. Create a local folder `/home/user/MCRL/bin` and download

`MCRL_vXXX_Linux64.tar.gz`

to the `bin` folder (the parent folder name can be arbitrary).

3. Extract the tarball in the `bin` folder by typing in the shell from within the `bin` folder

```
$ tar -xvf MCRL_vXXX_Linux64.tar.gz
```

4. To complete the installation, type in the in the shell from within the `bin` folder:

```
$ ./run_MCRL.sh <mcr_directory>
```

where `<mcr_directory>` is the directory where the MATLAB Runtime library is installed (or the directory where MATLAB is installed on the machine). For example:

```
$ ./run_MCRL.sh /code/MATLAB/2020a/
```

Follow the instructions and prompts to install blast and MCRL. Depending on your system it may take a couple of minutes for the installation to begin. **Note:**

prompts requiring user response can be minimized in the taskbar. Once the main interface of MCRL loads you may start using MCRL.

3.3 Running MCRL from MATLAB

Requirements:

1. MATLAB 2016a or later version
2. Supported operating systems: Windows (32 bit, 64 bit), Linux* (64 bit), MacOS**
3. Optional: Parallel Computing toolbox v4.2 or higher
4. Internet connection and administrator privileges

*MCRL has been tested on CentOS Linux 7

**MCRL is not officially supported for MacOS

Note: To run MCRL from the command line using `command_line_MCRL_EXE.m` or `par_command_line_MCRL_EXE.m` MCRL must first be installed.

For Linux RedHat users: prior to installation of MCRL, please manually download and install blast version 2.2.22+ compatible with your OS from <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.22/>. In the MCRL installation select "Locate sources on computer" to identify the bin folder of the BLAST installation.

For MAC users: prior to installation of MCRL, install conda (<https://docs.conda.io/>) if it is not already installed on your system, in order to enable MCRL to install DIAMOND. Alternatively, follow the instructions in <http://www.diamondsearch.org/> and install DIAMOND manually. For MacOS use this file: `MCRL_vXXX_MATLAB_WIN.zip`.

Instructions:

1. Download and install Visual C++ Redistributable (required for DIAMOND) from:
<https://support.microsoft.com/en-us/help/2977003/the-latest-supported-visual-c-downloads>
2. Download `MCRL_vXXX_MATLAB.zip`
(for Linux you may alternatively download `MCRL_vXXX_MATLAB.tar.gz`)
3. Extract files to a **local** directory such as `d:\MCRL`
for Linux type

```
-xvf MCRL_vXXX_MATLAB.tar.gz
```

4. Start MATLAB, navigate to the `bin` folder in the MCRL installation and run from the MATLAB command prompt `MCRL_EXE`.
5. Follow the instructions and prompts to install blast, clicking "next" and accepting all of the default entries. **Note: prompts requiring user response can be minimized in the taskbar.**

Once the main interface of MCRL loads you may start using MCRL.

4. Troubleshooting

Problems with MCRL installation (Windows)

1. Installation files should be downloaded to a local **offline** folder (e.g., `d:\`) and not a network drive
2. MCRL should also be installed in a local folder (e.g., `d:\MCRL`) and not on a network drive.
3. You require administrator privileges

Problems with MCRL installation (Linux)

The installation tarball should be extracted in a local `bin` folder within the MCRL installation folder (e.g., `/home/user_name/MCRL/bin`).

blast cannot be downloaded via ftp

1. Make sure you have a working internet connection
2. Make sure MCRL is installed in a local drive and not a network drive
3. Make sure the ftp port is not blocked by your firewall
4. Try temporally disabling your firewall
5. Disconnect any active VPNs
6. Contact your IT manager to check the settings on your computer
7. If an ftp connection still fails, manually download and install blast (see instructions below)

blast cannot be executed/installed

1. Make sure there are no antivirus programs blocking executables from running. Contact your IT manager to check the settings on your computer.
2. On Windows you must have administrator privileges to install blast
3. On Windows, make sure you install blast in the default folder (C:\Program Files (x86)\NCBI)
4. On Windows there might be a preexisting installation of blast interfering with MCRL. If Windows is installed on the c:\ drive, check the folder c:\Windows for a file called `ncbi.ini` and temporarily rename it.
5. Install blast manually (see below)

DIAMOND cannot be executed/installed

1. Make sure there are no antivirus programs blocking executables from running. Contact your IT manager to check the settings on your computer.
2. DIAMOND executables are provided for Windows and Linux in the local `diamond` folder in the MCRL installation and require no installation. If MCRL is not able to install DIAMOND for your OS, follow the instructions in <http://www.diamondsearch.org/index.php>. For Linux, the binary files should be copied to the local `diamond` folder in the MCRL installation. For MacOS DIAMOND needs to be on the global path.

Manually installing blast

If there is a problem downloading or installing the blast software, blast v2.2.22+ can be downloaded manually from the NCBI website:

<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.22/>, selecting the software version appropriate for your OS. Once blast is installed, start MCRL, when prompted, click "Locate sources on computer" and locate the `bin` folder of the blast installation.

5. Plotting reference gene networks

5.1 Configuration file

To plot reference gene networks, first configure the setup file `config_file_network.txt` found in the `networks` folder of the MCRL installation. This configuration file has three parts:

1. Select between four options for labeling options nodes:
 - a. `all`: all nodes are labeled with reference gene IDs
 - b. `none`: no labels are plotted
 - c. `auto`: optimize label plotting to image complexity
 - d. `epicenter`: plot only labels of reported reference genes
2. Enter the name of the `MCRL_table_nr.txt` output file using a full path
3. Enter one or more reported reference genes for which you wish to plot the networks. These reference genes must be reported reference genes included in the list provided in the `MCRL_table_nr.txt` output file.

Output figure files (`.fig` and `.tif` files) will be saved to the local `output` folder.

Example of `config_file_network.txt` file format:

```
# Node labels (all/none/epicenter/auto)
auto

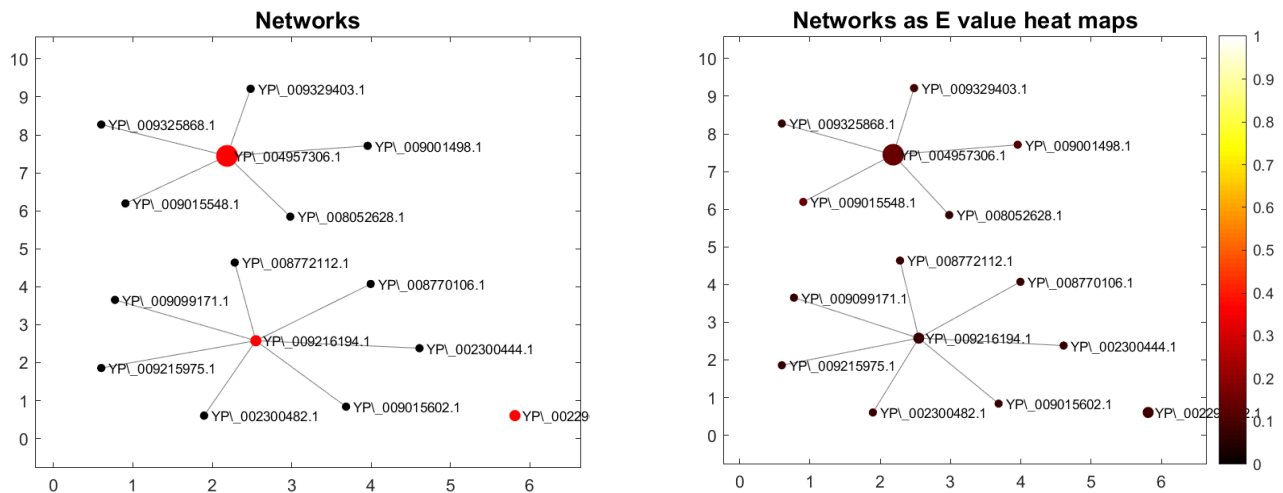
# MCRL_table_nr.txt file
C:\MCRL\FFFFFFFF_BLAST_table_EXAMPLE_MCRL_Table_nr.txt

# Reported reference gene(s)
YP_004957306.1
YP_008770526.1
YP_009238667.1
```

Initial calculation may take a long time but subsequent runs are immediate. It is possible to eliminate the initial computation time if the checkbox “Precompute tracks” is toggled in the user interface before running MCRL, in which case computation of all network tracks

will be performed on the fly (or set to 1 in the command line script). Below is an example of networks plots.

Figure 1. Example of reference gene networks plots computed by MCRL.



5.2 Creating plots from MATLAB

To plot reference gene networks from MATLAB:

1. Edit the `config_file_network.txt` file found in the `networks` folder
2. Execute the source `Plot_networks_EXE.m` from the `bin` folder in MATLAB.

Figures will be saved to the `output` folder in the MCRL installation path with the appropriate `<RUN_ID>`.

5.3 Creating plots using the Windows distribution

To plot reference gene networks using the Windows distribution:

1. Copy the binary file `plot_networks_EXE_WIN64.exe` to the local `bin` folder.
2. Edit the `config_file_network.txt` file found in the `networks` folder
3. Double click `plot_networks_EXE_WIN64.exe`,
or, alternatively, from the command line (from the `bin` folder) type:

```
Plot_networks_WIN64.exe <MCRL_path>\networks\config_file_network.txt
```

where <MCRL_path> is the installation path of MCRL. Figures will be saved to the output folder in the MCRL installation path with the appropriate <RUN_ID>.

5.4 Creating plots using the Linux distribution

To plot reference gene networks using the Linux distribution:

1. Extract the tar file `plot_networks_LINUX64.tar` in the local `bin` folder.
2. Edit the `config_file_network.txt` file found in the `networks` folder
3. From the command line (from the `bin` folder) type:

```
$ run_Plot_networks_LINUX64.sh <mcr_directory>  
  <MCRL_path>\networks\config_file_network.txt
```

5.5 Demo plot

For demonstration purposes, the default file `config_file_network.txt` can be run with the default demo files provided with the MCRL installation, only the path of the *MCRL_table_nr.txt file needs to be updated.

6. Support

In case of bugs or for support please contact Arbel Tadmor (arbel.tadmor@tron-mainz.de).