



**Primary Rapid
Overview of
Metagenomic
Taxonomy**

1. Introduction
2. Installation
 1. Core Deployment Steps
 2. Webserver
 3. Taxonomic Data
3. Running PROMpT
4. Screenshots

Introduction

PROMpT is a pipeline system for rapid metagenomic analysis of NGS amplicon sequencing data with a simplistic web interface allowing non-informatic users access to the benefits from NGS sequencing. While built for NGS, there is also the capacity to load classic analytical techniques (e.g. microscopy) to allow easy comparison between the methods.

It is designed to be implemented in analysis of your chosen taxonomic clade, requiring only reference sequences formatted into a blast database and a taxonomic hierarchy which can both be defined by the user. It also allows for correction factors to be applied to the reference sequences to allow for polyploidy or the effect of size differences observed in the community. The setup steps are documented and scripts to help produce the files are included in this package.

Once deployed, PROMpT will be able to be accessed through the web browser to initiate processing, generate new analyses and visualise the community results at varying taxonomic levels with a number of statistical tools.

Installation

Most dependencies are already installed and configured in the latest version of Bio-Linux (v8) and is the recommended operating system for PROMpT to be deployed upon.

If you wish to install on another linux distribution first ensure the following packages are installed and configured:

cd-hit, blast+, perl, python, R.

Core deployment steps

The latest development version of PROMpT can be directly cloned from github:

```
$ git clone git@github.com:passdan/PROMpT.git
```

To deploy PROMpT the following files must be correctly located on the server system.

```
$ cd PROMpT
$ sudo cp -r scripts /usr/lib/prompt/          ##Or other PATH location
$ sudo cp -r web* /var/www/prompt/            ##Your web location
$ sudo cp web/cgi-bin/* /usr/lib/cgi-bin/      ##cgi-bin on your system
```

The web files must be configured to the web handling 'user':

```
$ sudo chown -r www-data:www-data /var/www/prompt
```

Prompt is now able to run. If you go to a web browser and visit 'localhost/prompt' you should see the home page.

Webserver

A number of steps are required to configure the web-end of the PROMpT package and require administrator access to the system.

Firstly, Ensure that the latest LAMP server is configured. The process for this is described at <https://help.ubuntu.com/community/ApacheMySQLPHP>

Summarily, the following should install all the required packages:

```
$ sudo apt-get update
$ sudo apt-get install lamp-server^
```

The cgi-bin scripts must be present in /usr/lib/cgi-bin and the following lines added to /etc/apache2/sites-available/000-default.conf. This allows you to start analyses from the web interface, and run some of the statistics:

```
ScriptAlias /cgi-bin/ /usr/lib/cgi-bin/
<Directory "/usr/lib/cgi-bin">
    AllowOverride None
    Options +ExecCGI -MultiViews +SymLinksIfOwnerMatch
    Order allow,deny
    Allow from all
</Directory>
```

The following R packages must be installed globally so that they are accessible to the webserver. The simplest method of doing this is:

```
$ sudo su
$ R
>install.packages("gplots")
>install.packages("vegan")
>install.packages("ggplot2")
```

Taxonomic Data

To perform analyses you must supply PROMpT with the reference database which you intend to utilise, and the taxonomic reference file for it. The format for these files can be found in the 'blast_db' and 'taxonomy' directory example files. The codes for the reference sequences can be user defined or corresponding with public databases and the corresponding taxonomic structure file must correspond. Additionally, each sequence which you use for your reference file can be attributed a correction factor to compensate for differences to natural detection. This could relate to multiple copy number of the genome or organelle/plasmid and is discretionary. By default, all correction factors are 1. The first ID in the taxa dictionary is 000 which is reserved for non-matching sequences.

taxonomic_sequences.fasta

```
>001
ACTGCATGCATCGATCGTA...
>002
CATGCATGTCATCATGCATG....
>AJ579565
TGCAGACGTACTGCATCGTCGTC....
...
```

taxa_dictionary.txt

Strain	Species	Genus	Family	Class	correction
000	No_Match	No_Match	No_Match	No_Match	1
002	Surirella_brebissonii	Surirella	Surirellaceae	Bacillariophyceae	1
005	Fragilaria_gracilis	Fragilaria	Fragilariaceae	Bacillariophyceae	1
...					

These files can be located anywhere on the system but must be accessible to the webuser. Once created, the taxonomic_references must be converted into a blast database with the makeblastdb command (part of the blast+ installed package) and the database identified in the config.txt file.

```
$ makeblastdb -in test.fas -dbtype nucl -out test_dia
```

Running PROMpT

To begin an analysis with PROMpT you must transfer your fasta sequences to the input directory as defined in the config.txt file (one file per sample). Transferring files is not recommended through the web interface because of the large size, so using a file transfer system like winSCP is recommended. Transfer into an available directory on the server accompanied by a mapping file listing samples for analysis, and a modified version of the config.txt file.

Samples.txt

```
Sample10  River_Wye
Sample113  River_Taff
Sample15   River_Sever
...
```

Config.txt

```
## File paths and other settings for running PROMpT ##

[locations]
script_dir: /home/daniel/PROMpT/scripts/
sample_dir: /home/daniel/prompt-DTM/input_fasta/
sample_list: /home/daniel/prompt-DTM/input_fasta/sample_list.txt
tmp_dir: /home/daniel/prompt-DTM/tmp_dir/
blast_db: /home/daniel/prompt-DTM/blast_db/reference

[parameters]
blast_homology: 96
core_no: 3
# Options:both/process_only/analysis_only
run_mode: both
```

Go PROMpT!

PROMpT is now ready to run. Go to localhost/prompt in a web-browser on the system (i.e. chrome/firefox) or identify the system's website/ip address if accessing from a different machine on the network. While designed to be secure, do not assume that the PROMpT system is immune to external attack. It is recommend that you keep access limited to the localised network, or utilise a password protection system (for example .htaccess) when opening to external access.

Bellow are some screen shots of what to expect from PROMpT.

Index

BETA

PROMpT

**Primary Rapid
Overview of
Metagenomic
Taxonomy**

What do you want to do?

[Perform new analysis](#) | [View completed analysis](#)

Introduction

PROMpT is a pipeline system for rapid metagenomic analysis of NGS amplicon sequencing data with a simplistic web interface allowing non-informatic users access to the benefits from NGS sequencing. While built for NGS, there is also the capacity to load classical analytical techniques (e.g. microscopy) to allow easy comparison between the methods.

It is designed to be implemented in analysis of your chosen taxonomic clade, requiring only reference sequences formatted into a blast database and a taxonomic heirarchy which can be defined by the user. Both of these steps are documented and scripts to proiduce the files are included in this package.

Once deployed, PROMpT will be able to be accessed through the web browser to generate new analyses and visualise the community results at varying taxonomic levels with a number of statistical tools.

PROMpT is maintained by [passdan](#)
Theme based upon Tactile by [Jason Long](#). Piecharts were generated by [amCharts](#)

Processing

BETA

PROMpT

**Primary Rapid
Overview of
Metagenomic
Taxonomy**

Navigation

[Home](#) | [Processing](#) | [Results](#)

Processing

On submission, the local fasta file(s) will be submitted to CD-HIT for reduction of duplicate reads. It will then be annotated via BLAST to the reference database, to identify the closest reference in the sequence database. Quality controls remove the matches beyond a cut-off (as defined in the configuration file) and are described taxonomically. On completion, the user is notified and the files are available in the result analysis page for viewing.

Start Processing

[Display/Update Runlog](#)

PROMpT is maintained by [passdan](#)
This page was generated by [GitHub Pages](#). Tactile theme by [Jason Long](#)

Results

BETA

PROMpT

Primary Rapid
Overview of
Metagenomic
Taxonomy

Navigation

Home | Processing | Results

Data Portal

Pie charts abundance by sample

Data Origin:

Barcode Sequencing

Microscopy

Select a Site:

DTM100

DTM113

DTM15

DTM34

DTM42

DTM44

DTM45

DTM47

DTM55

DTM56

DTM69

DTM72

DTM73

DTM74

DTM96

DTM98

DTM99

site: MIC

Choose Taxonomy level...

go

Comparative sample analysis

Choose Taxa Level:

class

family

genus

species

refseq

MIC

☒

DTM100

☐

DTM113

☐

DTM15

☐

DTM34

☒

DTM42

☒

DTM44

☐

DTM45

☐

DTM47

☒

DTM55

☐

DTM56

☐

DTM69

☐

DTM72

☒

DTM73

☒

DTM74

☐

DTM96

☐

DTM98

☒

DTM99

NGS

☒

DTM100

☐

DTM113

☐

DTM15

☐

DTM34

☒

DTM42

☒

DTM44

☐

DTM45

☐

DTM47

☒

DTM55

☐

DTM56

☐

DTM69

☐

DTM72

☒

DTM73

☒

DTM74

☐

DTM96

☐

DTM98

☒

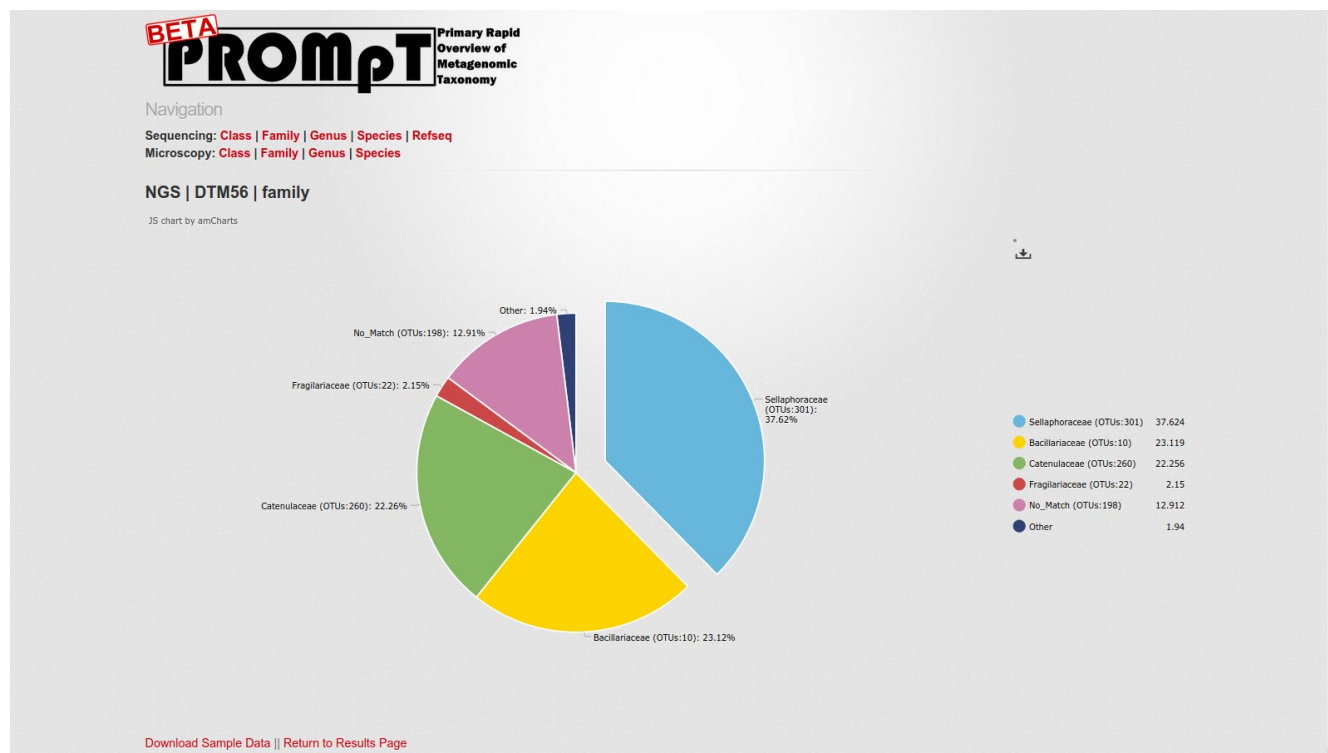
DTM99

Collate Data

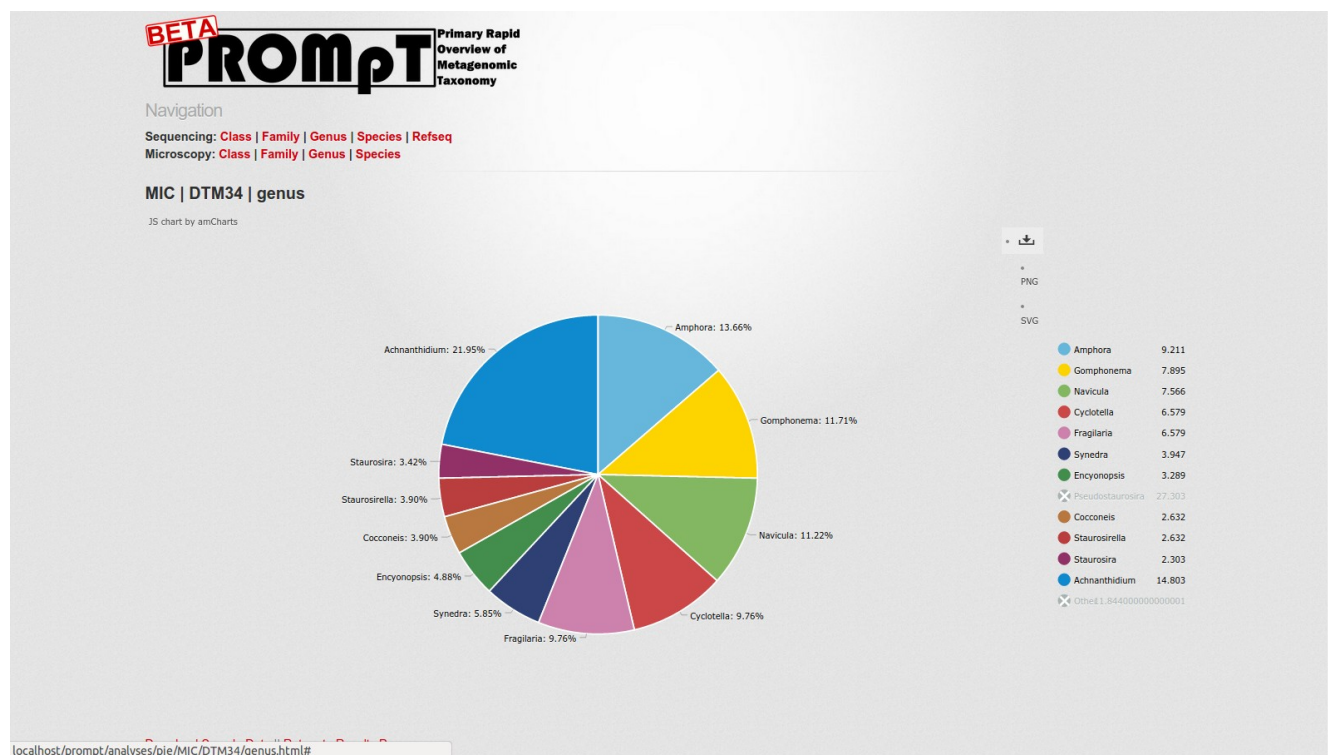
PROMpT is maintained by passdan

This page was generated by GitHub Pages. Tactile theme by Jason Long.

Single Sample Piechart from NGS amplicon data



Single Sample Piechart from Microscopy data (download image button highlighted). Also note the ability to remove taxa from selection and dynamically redraw the chart



Multi-sample comparative data analysis

BETA

PROMPT

Primary Rapid
Overview of
Metagenomic
Taxonomy

Navigation

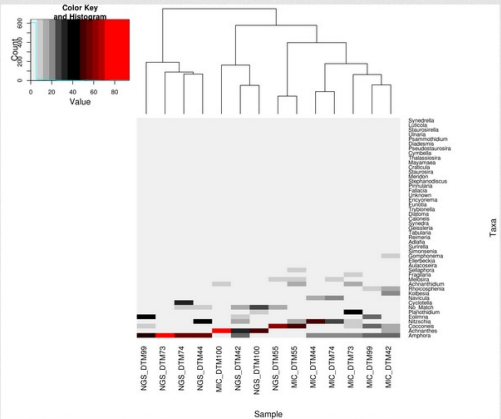
Home | Processing | Results

Analysing:

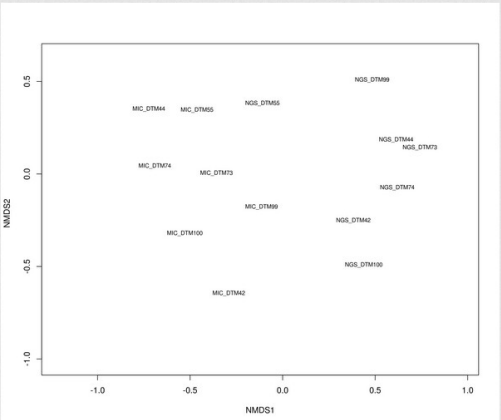
MIC_DTM100
MIC_DTM42
MIC_DTM44
MIC_DTM55
MIC_DTM73
MIC_DTM74
MIC_DTM99
NGS_DTM100
NGS_DTM42
NGS_DTM44
NGS_DTM55
NGS_DTM73
NGS_DTM74
NGS_DTM99
null device 1 null device 1

Download data: [Current_selection.csv](#)

Heatmap



NMDS



Thanks for having a look at PROMpT. If you have any suggestions, ideas or issues please contact me at daniel.antony.pass@gmail.com.