

## Supporting Text S4

### ***PhyloPythiaS+*: A Self-Training Method for the Rapid Reconstruction of Low-Ranking Taxonomic Bins from Metagenomes**

#### **Benchmarking**

This document describes how to configure the software to reproduce the results.

#### **Software installation**

Follow the installation instructions and go through the tutorial.

#### **Download datasets**

The datasets for the benchmarks can be downloaded from:  
<https://github.com/algbioi/datasets>

Each file is a 7z archive and can be extracted, e.g. using command: `7za x archive.7z`  
Each extracted directory contains a `readme.txt` file describing all the files contained in the directory.

#### **Real datasets**

Follow the tutorial:

- Create the pipeline directory in directory: `/apps/pps/tests`
- Use configuration file:  
`/apps/pps/tools/config_ppsp_vm_refNCBI20121122_example.cfg`  
as a template (i.e. copy this file and modify it appropriately).
- Make sure, you set the following parameters in the configuration file:  
`pipelineDir`  
`inputFastaFile`  
`inputFastaScaffoldsFile`  
`scaffoldsToContigsMapFile`
- Run the pipeline using command:

```
ppsp -c CONFIGURATION_FILE -n -g -o sl6 mg -t -p c s v -r -s
```

- Analyze the results as described in the tutorial.

## Simulated datasets

Follow the tutorial:

- Create the pipeline directory in directory: /apps/pps/tests
- Use configuration file:  
/apps/pps/tools/config\_ppsp\_vm\_refNCBI20121122\_example.cfg  
as a template (i.e. copy this file and modify it appropriately).
- Make sure, you set the following parameters in the configuration file:  
pipelineDir  
inputFastaFile  
referencePlacementFileOut  
excludeRefSeqRank (e.g. excludeRefSeqRank=species)  
excludeRefMgRank (e.g. excludeRefSeqRank=strain)
- Run the pipeline using command:  

```
ppsp -c CONFIGURATION_FILE -n -g -o sl6 mg -t -p c -r -s
```
- Analyze the results as described in the tutorial