

The IHP-PING Package Manual

Gaston K. Mazandu^{1,2,3*} *et al.*

Type Package

Title: An integrated human protein-protein interaction network generator

Version 2.4.1

Contributors: Christopher Hooper¹, Kenneth Opa², Funmilayo L. Makinda^{2,3}, Victoria Nembaware¹, Nicholas E. Thomford^{3,4}, Emile R. Chimusa¹, Ambroise Wonkam¹ and Nicola J. Mulder², Gaston K. Mazandu^{1,2,3*}

¹*Division of Computational Biology, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Medical School, Anzio Road, Observatory 7925, Cape Town, South Africa.*

²*African Institute for Mathematical Sciences (AIMS), Melrose Road, Muzenberg 7945, Cape Town, South Africa.*

³*Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town (UCT), Medical School, Anzio Road, Observatory 7925, Cape Town, South Africa.*

⁴*School of Medical Sciences, University of Cape Coast, PMB, Cape Coast, Ghana.*

Maintainer Mazandu GK <gaston.mazandu@uct.ac.za, gmazandu@gmail.com, kuzamunu@aims.ac.za>

General description

In the current ‘big data’ driven ‘post-genomic’ era, in which all available sources of information can be explored at the systems level, protein-protein interactions (PPIs) are contributing to elucidation of the complex genetic architecture of organisms. These PPIs are experimentally or computationally predicted, stored in different online resources and updated regularly. As with many biological datasets, this continuously renders older protein-protein interaction (PPI) datasets potentially outdated. The integrated human protein-protein interaction network generator (IHP-PING) tool is a flexible python package, which generates a human PPI network from freely available online resources. This tool extracts and integrates heterogeneous PPI datasets to generate a unified PPI network, which is stored locally for further user applications.

Depends Python (≥ 2.7)

requires [local ncbi-blast, python-selenium, chromium-browser and chromium-chromedriver]

License GLP (<https://www.gnu.org/licenses/gpl-3.0.en.html>)

URL <http://web.cbio.uct.ac.za/ITGOM/post-analysis-tools/ihp-ping-dev/> and <https://github.com/gkm-software-dev/post-analysis-tools>

Release date Friday 19th June, 2020–12:27

*To whom correspondence should be addressed. Tel: +27 21 650 3463; Email: gaston.mazandu@uct.ac.za

Contents

The IHP-PING Package documented	3
1.1 The IHP-PING Package	3
1.2 IHP-PING Environment Management	3
Appendix-1 IHP-PING Administration and Usage	4
2.1 IHP-PING administration	4
2.2 IHP-PING usage	4
2.3 IHP-PING licence and version	4
2.4 Running IHP-PING	5
2.5 Illustrating IHP-PING usage	5
2.6 Running IHP-PING as a Python Package	6
2.7 Important notes	6
2.8 Contributors	6
2.9 Main references	6
2.10 Questions, Comments and Report Bugs	7
2.11 IHP-PING copyright and License	7
2.12 Citing IHP-PING	7
2.13 Other information about IHP-PING development	7

IHP-PING Package	An integrated human protein-protein interaction (PPI) network generator tool, a user-friendly and accessible tool, easing integration of PPI datasets from multiple sources into a unified PPI network on-the-fly, which is stored locally for further user applications.
-------------------------	---

Description

A repository of python modules for easing integration of existing human PPI datasets from multiple sources into a unified PPI network on-the-fly (in real time), which is stored locally for further user applications. This provides a platform that enables the retrieval of human PPI datasets stored in different online resources and produced a real time up-to-date integrated human PPI network with increased accuracy, confidence and coverage.

IHP-PING Environment Management

The IHP-PING system is composed of one main high level folder: `PyPING` and one main python module, `ihppinbuilder.py`, which serves as an interface processing the human PPI interaction network and storing this network locally for further user applications. The `PyPING` folder includes two Python modules: (1) `networkgenerator.py`, which downloads PPI datasets requested by a user and builds the integrated PPI network and (2) `sequenceprocessing.py`, processing and scoring sequence dataset, which includes protein sequences and InterPro domains.

As a package, IHP-PING can be imported in another Python module, however, a user can directly produce a unified human PPI network in three logical steps: PPI Extraction, Mapping Process and Integration Process, via user interface and input processing using a simple single command-line terminal on any computer or any operating system running Python. The only required argument is the list of human PPI datasets to be incorporated into the unified PPI network, with output format parameters according to user preferences. Each user-requested database is downloaded from specific uniform resource locations (URLs). Eight different resources used are shown in **Figure 1** and described in **Table 1** in the main manuscript.

IHP-PING generates an output file in a tabular format, the number of columns depending on the PPI datasets retrieved with each row representing a unique PPI. The first two fields contained the IDs of the two proteins involved in the interaction, remaining columns showing the scores for the interaction from each source. The last value in the row contained the combined score of the interaction (see **Appendix 2**). Once IHP-PING has run successfully, the resultant PPI network can be used in any form of network analysis.

It is worth noting that different sources provided here are freely accessible and, generally, different datasets are free to use for non-profit organizations. However, depending on the use of a unified PPI network being generated from these resources, the user need checks specific conditions required for each datasets. Also the user should bear in mind that there is a need of a credential identifier (ID) for validation and ID protection (VIP) services provision from the DIP Team for free access and use for non-profit organization.

In summary, IHP-PING has been designed, a software for downloading and integrating currently available human PPIs, alongside using protein sequence information to predict further interactions. This software is implemented in Python modules, which can be invoked from the terminal interface on any computer or any operating system running Python, using a single command-line (refer to **Appendix 1**).

Appendix-1 IHP-PING Administration and Usage

2.1 IHP-PING administration

The main website for the IHP-PING package is <http://web.cbio.uct.ac.za/ITGOM/post-analysis-tools/ihp-ping-dev> where users can find essential information about obtaining IHP-PING. It is freely downloadable under GNU General Public License (GPL), pre-compiled for Linux version and protected by copyright laws, a free software and comes with ABSOLUTELY NO WARRANTY. Users are free to copy, modify, merge, publish, distribute and display information contained in the package, provided that it is done with appropriate citation of the library and by including the permission notice in all copies or substantial portions of the module contained in this package.

The whole package itself is relatively small with a total of about 0.2MB, dynamically increasing when retrieving different PPI datasets from different resources. IHP-PING contains one main module and one main folder containing modules required for generating an integrated human PPI network and formatting results to be written into a file. It is currently maintained by one member of the core-development team, Gaston K. Mazandu <gmazandu@gmail.com, gaston.mazandu@uct.ac.za, kuzamunu@aims.ac.za>, who regularly updates the information available in this package and makes every effort to ensure the quality of this information.

2.2 IHP-PING usage

IHP-PING v2.4.1 requires Linux operating system and Python (≥ 2.7), requiring the installation of the NCBI BLAST software locally when retrieving interactions predicted from sequence data. It also requires the selenium Python package, as well as chromedriver and chromium-browser, for retrieving the DIP dataset. These needs to be installed prior to the use of IHP-PING.

To use IHP-PING, the user needs to download the 'tar.gz' file and extract all files as follows:

```
tar xzf ihp-ping-tool.tar.gz
```

or alternatively, it can also be retrieved from the github public platform using git clone command line as follows.

```
git clone https://github.com/gkm-software-dev/post-analysis-tools.git
```

After downloading and/or uncompressing, move to the folder `post-analysis-tools/ihp-ping-dev/`, which should be set as a working directory where IHP-PING and related commands are executed using the following terminal command:

```
cd post-analysis-tools/ihp-ping-dev/
```

2.3 IHP-PING licence and version

As pointed out previously, the IHP-PING package is free to use under GNU General Public License. You are free to copy, distribute and display information contained herein, provided that it is done with appropriate citation of the library. Thus, by using the IHP-PING package, it is assumed that you have read and accepted the agreement provided and that you agreed to be bound to all terms and conditions of this agreement. Please, use the following command line to see the package licence:

```
python setup.py --licence
```

To check the current version of the IHP-PING interface, use the following terminal command:

```
python setup.py --version
```

2.4 Running IHP-PING

As pointed out previously, IHP-PING can be processed through one main python module, `ihppinbuilder.py`, which serves as an interface. Get help on how to run IHP-PING through this interface module using the following command:

```
python ihppinbuilder.py -h
```

The above command should produce the following output:

```
usage: ihppinbuilder.py [-h] [-r list [list ...]] [-o FILE] [-f str]
```

with different tags explained below:

<code>-h, --help</code>	show this help message and exit
<code>-r list [list ...], --resources list [list ...]</code>	Database to be integrated or considered (default: all)
<code>-o FILE, --dir FILE</code>	Folder which will contain the PPI produced (default: current working folder)
<code>-i str, --identifiers str</code>	Identifier outputs: uniprot or genename (default: uniprot)
<code>-f str, --outformat str</code>	Output format tsv, csv or csv2 (default: tsv)

As highlighted by the `help` option, IHP-PING is run using the following one line command:

```
python ihppinbuilder.py -r resources -o outputfolder -i outputProtID -f outputfileformat
```

1. **resources**: represent different online PPI resources to be included in the unified human PPI network to be generated. Different resource arguments are shown in Table ?? and by default all resources are included.
2. **outputfolder**: The path to the folder where the outputs should be written. If not provided, the current working directory is used.
3. **outputProtID**: The identifier (ID) system to be used in the output file. If not provided, the UniProt ID system is used. IHP-PING provides two ID systems: UniProt and genename.
4. **outputfileformat**: The format of the PPI network file produced and this depends on user preferences. Three possible formats are provided: `tsv` (tab separated values), `csv` (comma separated values) and `csv2` (semi-column separated values).

2.5 Illustrating IHP-PING usage

As pointed out previously, move to the `ihp-ping-dev` folder and run following commands for illustration. Please type these commands manually using the computer keyboard, do not use `copy` and `paste`:

```
python ihppinbuilder.py -r sequence mips
```

This produces a tsv format file of human PPI network derived from sequence data and MPPI-MIPS online database saved in the working directory, which is `ihp-ping-dev` folder.

The second and the third commands building some human PPI networks are given below:

```
python ihppinbuilder.py -f csv
python ihppinbuilder.py -r stringdb biogrid dip -i genename -f csv2
```

The second command will generate a unified human PPI network derived from all sources under consideration currently, and save under the working directory (default) in csv format. For the third command, only STRING, BioGrid and DIP databases are used and the network is saved as a csv2 (semi-column separated value) file with the gene name ID system.

2.6 Running IHP-PING as a Python Package

As any python library or package, IHP-PING can be imported and used in another Python models. For accessing and learning about different classes of the two main modules under PyPING, `networkgenerator` providing functions for downloading and integrating the human PPI network and `sequenceprocessing` for processing sequence datasets (protein sequences and InterPro datasets). Please access the python interpreter or the command shell for interactive computing (IPython) and run following commands:

```
>>> from PyPING import *
>>> help(networkgenerator)
>>> help(sequenceprocessing)
```

2.7 Important notes

- To efficiently use the IHP-PING library and to maximally benefit from its use, make sure that you have carefully read this PDF package documentation file, which is provided in the library.
- In some cases, you may need or be required to provide the folder in which interaction file should be. Please make sure that the full path to the folder target is provided.
- Make use of the full screen mode when displaying results on it for a nice and more adapted visualization.

2.8 Contributors

Christopher Hooper, Kenneth Opap, Funmilayo L. Makinda, Victoria Nembaware, Nicholas E. Thomford, Emile R. Chimusa, Ambroise Wonkam and Nicola J. Mulder, Gaston K. Mazandu

Maintainer: Mazandu GK <gmazandu@gmail.com, gaston.mazandu@uct.ac.za, kuzamunu@aims.ac.za>

2.9 Main references

1. Mazandu GK, Chimusa ER, Rutherford K, Zekeng EG, Gebremariam ZZ, Onifade MY, Mulder NJ. Large-scale data-driven integrative framework for extracting essential targets and processes from disease-associated gene data sets. *Brief Bioinform* 2018, 19(6), 1141–1152.

2. Mazandu GK, Mulder NJ. Using the underlying biological organization of the Mycobacterium tuberculosis functional network for protein function prediction. *Infection, Genetics and Evolution* 2012, 12(5), 922–932
3. Mazandu GK, Mulder NJ. Generation and Analysis of Large-Scale Data-Driven Mycobacterium tuberculosis Functional Networks for Drug Target Identification. *Advances in Bioinformatics* 2011, 2011(Article ID 801478), 14 pages. Doi:10.1155/2011/801478.
4. Mazandu GK and Mulder NJ. Scoring protein relationships in functional interaction networks predicted from sequence data. *PLoS One* 2011, 6(4), e18607.

2.10 Questions, Comments and Report Bugs

The IHP-PING team is striving to aggregate knowledge about the human protein-protein interaction online datasets on-the-fly in realistic timeframes, providing a solution to producing a real time or up-to-date unified human PPI network. However, IHP-PING does not guarantee the quality or accuracy of different result outputs. Thus, if it happens that you find errors, please contact the primary source of data set used for more information. If you feel that the errors may be due to some systematic error in the PySML library, please contact the library maintainer at <gmazandu@gmail.com, gaston.mazandu@uct.ac.za, kuzamunu@aims.ac.za>.

2.11 IHP-PING copyright and license

The IHP-PING library is free to use under GNU General Public License (GPL: <https://www.gnu.org/licenses/gpl-3.0.en.html>). You are free to copy, distribute and display information contained herein, provided that it is done with appropriate citation of the tool.

2.12 Citing IHP-PING.

The manuscript is being prepared for publication, before its publication you can cite the preliminary report:

“Hooper C, Opap K, Makinda FL, Nembaware V, Thomford NE, Chimusa ER, Wonkam A, Mulder NJ, Mazandu GK. IHP-PING—A flexible tool for generating integrated human protein-protein interaction networks on-the-fly”. Technical report 2020, H3ABioNet-AIMS node and SADaCC, AIMS & UCT, South Africa. <http://web.cbio.uct.ac.za/ITGOM/post-analysis-tools/ihp-ping-dev/>.

2.13 Other information about IHP-PING development

Please refer to <http://web.cbio.uct.ac.za/ITGOM/post-analysis-tools/ihp-ping-dev/PKG-INFO> (See some other details about the IHP-PING development)

or

go to the local `ihp-ping-dev` folder and type the following command line for the short description:

```
python setup.py --description
```

or alternatively,

```
python setup.py --long-description
```

for the long description.