

PseAAC-Builder Version 2.0

A User Manual

Release Date: Feb. 8th, 2013

By: Pufeng Du, Ph.D.

School of Computer Science and Technology, Tianjin University

Tianjin 300072, China

Department of Computer Science, City University of Hong Kong

Hong Kong

Contents

1 Introduction to the PseAAC algorithm	2
2 What's new in PseAAC-Builder v 2.0.....	2
2.1 Version 2.0 is much faster	2
2.2 Version 2.0 is following POSIX/GNU standard	2
2.3 The new GUI	3
2.4 Portability.....	3
3 Installations.....	3
3.1 Binary distributions.....	3
3.2 Source code compilations.....	4
4 Command line usage	5
5 Useful resources	5
6 Bug report.....	5
7 Forecasting the next version.....	5

Important Notice:

1. Using PseAAC-Builder implies the agreement of the license terms. See the README file for details.
2. It is strongly recommended that users should learn the algorithm basics of pseudo-amino acid composition before using it. A detailed description of pseudo-amino acid composition is provided within this manual. Applying the advanced computational methods, like pseudo-amino acid compositions without understanding its mechanism may be risky and may lead to false discoveries
3. All examples in this manual are only intended to be used to demonstrate the usage of the software. The data and results are scientifically meaningless.

1 Introduction to the PseAAC algorithm

Pseudo amino acid composition (PseAAC) is an algorithm that could convert a protein sequence into a digital vector that could be processed by pattern recognition algorithms. The design of PseAAC incorporated the sequence order information to improve the conventional amino acid compositions. The application of pseudo amino acid composition is very common, including almost every branch of computational proteomics. For the details of the algorithm, the users should refer to the original literature by Prof. Kuo-Chen Chou in the year 2001.

2 What's new in PseAAC-Builder v 2.0

2.1 Version 2.0 is much faster

If you are a user of the PseAAC-Builder v1.0, you should know that, the Mono framework or .net framework was required to execute the version 1.0, as version 1.0 was designed and implemented with C#. Because these framework actually slow down the processing speed of the PseAAC-Builder, we decided to re-design the whole program with GNU C/C++ from scratch in version 2.0. In a comparison test, PseAAC-Builder v2.0 achieved about 100 times speed of version 1.0.

On a desktop PC running Ubuntu 12.04 Linux system, PseAAC-Builder v2.0 can process a little over 17000 sequences per second, while the version 1.0 can only process about 170 sequences per second with the same parameters and the same dataset on the same machine.

2.2 Version 2.0 is following POSIX/GNU standard

As you can see in version 1.0, the command line parameters are generally following the style of BLAST. In version 2.0, with the totally new framework of code, the command line processing system has been redesigned to follow GNU standard, which can be perfectly integrated into UNIX/Linux/Mac system.

Particularly, according to the users experience reports, we decided to remove the default parameter configurations from the program. That means, if you do not enter any parameter on the command line to specify the type of pseudo-amino acid composition, the program will simply wait

input from the STDIN and then do nothing but exit normally. This is a very common behavior of most UNIX/Linux tools, which acquire the input data from the STDIN stream.

2.3 The new GUI

For the GUI part, we reduce some functions of the GUI. The GUI part, which was designed to be an easy-to-use user-friendly interface to the program, has also been redesigned and rewritten with Tcl/Tk. The basic function of specifying input and output file, configuring the parameters are still kept on the GUI. But, the function of showing all available physicochemical properties for chosen and showing the output to the STDOUT in real time has been cut out from the GUI, as the user experiences pointed out these functions are actually useless or significantly slow down the processing.

2.4 Portability

For most system following POSIX/GNU standard, it is easy to compile the source code of version 2.0 with GNU C/C++ compiler and GNU C Library. The only situation, which is a little different, is the Microsoft Windows. For Microsoft Windows system, we will need the Cygwin environment for building the executable. The details of building the Cygwin executable has been stated specifically in the Cygwin version. However, no matter which system you are using, you will need the Tcl/Tk interpreters on your system to execute the GUI program.

3 Installations

3.1 Binary distributions

For the users' convenience, we provide the binary executable for the following system:

- (1) Ubuntu Linux
- (2) Microsoft Windows + Cygwin

If you have already prepared the runtime environment for the above platforms, you may directly use the binary executable from the corresponding package.

3.2 Source code compilations

For any other platform/environments, the users will need to compile the executable from the source code. First, the users should download the source code package and use the following command to unpack the source code package in the users home directory.

```
tar -xzf ~/pseb2-src.tar.gz
```

After the source code has been unpacked, the users should go to the source code directory by typing:

```
cd ~/pseb2-src/src
```

To compile the source code on your platform, you will need the GNU tools-chain, GNU C/C++ compiler and GNU C runtime library. When every dependency is ready, you will need to simply type the following to build the executable.

```
make
```

If there is no errors, you will need to type the following to install the program.

```
make install
```

If everything goes right, you should have the binary executable of the program on the ~/pseb2-src/bin directory and the GUI program in ~/pseb2-src/gui directory. To execute the GUI program, you should type the following.

```
wish ~/pseb2-src/gui/pseb-gui.tcl
```

Of course, the above command will need you to install the Tcl/Tk on your platform.

4 Command line usage

The command line parameters are DIFFERENT from the version 1.0. The detailed usage of the command line can be shown by typing:

```
pseb2 --help
```

Actually, the number of options that can be chosen is the same as the version 1.0. But the meanings and the styles of almost all options are changed. So, if you have scripts using the version 1.0. You may need to modify it to make it applicable with version 2.0.

5 Useful resources

Cygwin: <http://www.cygwin.com/>

Ubuntu Linux: <http://www.ubuntu.com/>

Tcl/Tk: <http://www.tcl.tk/>

Active Tcl: <http://www.activestate.com/activetcl>

6 Bug report

As the release of version is really in a hurry, I believe that there are many problems while using the version 2.0. Every bug report/suggestion/complain/criticism is welcome. You may send all your thoughts about PseAAC-Builder to Dr. Pufeng Du, email: PufengDu@gmail.com.

7 Forecasting the next version

As you were reading this manual, the next version of PseAAC-Builder is being developed. The most useful feature in the next version will be the ability of generating general form PseAAC incorporating many different types of descriptive information of the proteins. If you demand these functions or other similar functions, you should never miss the next version. Furthermore, you are

welcome to take part in the development of the next version. If you have any queries/suggestions regarding the next version, your emails to PufengDu@gmail.com are welcome.