

GATool (Beta) Docs - version 0.2

Matheus Brito de Oliveira

May 19, 2017

Contents

1 Introduction	3
1.1 How does GATOOL Works?	3
1.2 Hardware Requirements	3
2 Installation	3
2.1 Installation of basics requirements on Ubuntu	4
2.1.1 Apache Server	4
2.1.2 PHP5 and libraries	4
2.1.3 JDK - JAVA	4
2.1.4 Python	4
2.1.5 Python-pip	4
2.2.1 FASTQC	4
2.2.2 FASTX-Toolkit and Library	5
2.2.3 SPAdes and Velvet assemblers	6
2.2.4 QUASt and libraries	6
2.2.5 Biopython and dependencies	7
2.2.5 NCBI-Blast+	7
2.3.1 Change the php.ini file	8
2.3.2 Change the /etc/sudoers file	8
2.4 PHP 7.0	8
2.5 Running GATOOL	8

1 Introduction

GATool is a web interface which integrates several tools, with the purpose of to do a genome assembly task and provide to the users a nice and real experience and a very fast way to genome assembly. GATool was created to assist bioinformaticians to assemble bacterial genomes sequenced with Next Generation Sequencing (NGS) platforms quickly, easily and effectively.

1.1 How does GATool Works?

GATool runs on the web and can be executed through any browser on any operational system. However, some minimal hardware requirements are necessary for the correct operational of the tool.

1.2 Hardware Requirements

We performed the analysis with GATool in a 2.50 GHz 64-bit Dual-Core CPU (Intel Core™ i5-7300HQ) and 8GB of RAM, running Ubuntu 16.04 LTS. Some assemblies require more robust hardware.

2 Installation

Some necessary requirements must be installed for the correct functioning of GATool. We can see what are in Table 1 below.

Tabela 1: Requirements	
Requirements	Tools
Linux	Ubuntu 16.04 LTS
Servidor Web	Apache
PHP	Php 5.6 ou superior
JAVA	Oracle JDK 8 ou superior
pip	Python 2.7.10 ou superior

2.1 Installation of basics requirements on Ubuntu

2.1.1 Apache Server

Apache server: Web server required to manage GATOOOL pages, which will be accessed by a browser. In the Linux terminal type the following command:

```
sudo apt-get install apache2  
{adm password}
```

2.1.2 PHP5 and libraries

```
sudo apt-get install php5 libapache2-mod-php5 php5-mcrypt
```

2.1.3 JDK - JAVA

```
sudo add-apt-repository ppa:webupd8team/java  
sudo apt-get update  
sudo apt-get install oracle-java8-installer
```

2.1.4 Python

```
sudo apt-get install python-dev
```

2.1.5 Python-pip

```
sudo apt-get install python-pip
```

2.2 Installation of the tools that make up GATOOOL

GATOOOL is composed of several tools, each with a different purpose, we will show you how to proceed with the installation of each of these tools, as well as their dependencies, on the linux terminal, enter the following commands as shown below.

2.2.1 FASTQC

After downloading and extracting it is necessary to allow the file to execute as a program, since FastQC has two execution modes, with visual interface and without

interface, after this procedure, we move it to the /usr/local/bin folder, so accessible from any location.

```
wget "http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.5.zip"
unzip destinationSource/fastqc_v0.11.5.zip
cd /destinationSource/FastQC/
sudo chmod 755 fastqc
{adm password}
sudo ln -s /home/"user"/destinationSource/FastQC/fastqc /usr/local/bin/fastqc
```

2.2.2 FASTX-Toolkit and Library

FASTX-Toolkit is a powerful toolkit that allows for diverse analysis in reads, from quality to quality filtering and trimmer. Libgtextutils is a library required for the correct operation of the fastx.

i. Libgtextutils

```
wget "http://www.github.com/agordon/libgtextutils/releases/download/0.7/libgtextutils-0.7.tar.gz"
tar -xvf libgtextutils-0.7.tar.gz
cd libgtextutils-0.7/
sudo ./configure
sudo make
sudo make install
```

ii. Fastx-toolkit

```
wget "http://www.github.com/agordon/fastx_toolkit/releases/download/0.0.14/fastx_toolkit-0.0.14.tar.bz2"
tar -jxvf fastx_toolkit-0.0.14.tar.bz2
cd fastx_toolkit-0.0.14/
sudo ./configure
sudo make
sudo make install
```

2.2.3 SPAdes and Velvet assemblers

The genome assemblers used by GATOOL, are SPAdes and Velvet, for the correct installation of Velvet it is necessary to install the zlib library.

a. SPAdes Installation

```
wget http://SPAdes.bioinf.spbau.ru/release3.7.1/SPAdes-3.7.1-Linux.tar.gz
tar -xzf SPAdes-3.7.1-Linux.tar.gz
cd SPAdes-3.7.1-Linux
sudo cp -a bin/* /usr/local/bin
{adm password}
sudo cp -a share/SPAdes/ /usr/local/share/
```

b. Zlib installation

```
sudo apt-get install zlib1g-dev
```

c. Velvet Installation

```
wget http://www.ebi.ac.uk/~zerbino/velvet/velvet_1.2.10.tgz
tar -zxvf velvet_1.2.10.tgz
cd velvet_1.2.10/
sudo make
{adm password}
sudo make 'CATEGORIES=57'
sudo make 'MAXKMERLENGTH=99'
sudo mv velvetg velveth /usr/bin/
```

With these simple commands Velvet will already be installed and configured, but it is necessary to look at the following aspect: the value of k of Velvet is set to 99, if it is necessary to modify it is necessary to recompile the code, changing the parameter MAXKMERLENGTH to the value wanted.

2.2.4 QUAST and libraries

After the pipeline user assembly process, you need to view some assembly statistics such as N50, N75, number of contigs, size of largest contig, and many other important metrics in the assembly process. For this task we used the QUAST (Quality Assessment Tool for Genome Assemblies), it is a script written in python, which generates several statistics about the genome assembly process. For the correct installation of the quast, it was necessary to install a 2D library for graphics generation, called matplotlib.

i. Matplotlib

```
sudo pip install matplotlib
```

ii. QUAST

```
wget https://downloads.sourceforge.net/project/quast/quast-4.2.tar.gz
tar -xzf quast-4.2.tar.gz
cd quast-4.2
sudo ./install.sh
{senha de administrador}
sudo ln -s quast.py /usr/local/bin/quast
```

2.2.5 Biopython and dependencies

For installing biopython you need to install two additional components that are: numpy and reportlab.

a. Numpy

```
sudo pip install numpy
```

b. Reportlab

```
sudo pip install reportlab
```

c. Biopython

```
sudo pip install biopython
```

2.2.5 NCBI-Blast+

This is required to run local alignments between sequences using BLAST. In the Linux terminal, type the following command:

```
sudo apt-get install ncbi-blast+
```

2.3 Changing configurations from Apache Server

GATool needs some additional settings on the Apache server because the default settings do not meet the specifics the tool needs to have.

2.3.1 Change the php.ini file

There are two configuration files called php.ini, which need to be changed. The files are located in the /etc/php/7.0/apache2/php.ini and /etc/php/7.0/cli/php.ini directories. The changes must be made in the two php.ini files.

```
sudo nano /etc/php/7.0/apache2/php.ini
#seek the line: upload_max_filesize e substituir por: upload_max_filesize=1024M
#seek the line: post_max_size e substituir por: post_max_size= 1024M
#seek the line: session.gc_maxlifetime = 7200

sudo nano /etc/php/7.0/cli/php.ini
#seek the line: upload_max_filesize e substituir por: upload_max_filesize=1024M
#seek the line: post_max_size e substituir por: post_max_size= 1024M
#seek the line: session.gc_maxlifetime = 7200
```

These changes allow GATool to load files up to 1 GB in size and also allow the value of the session variables to be kept longer, which helps in assembling genomes that use longer processing times.

2.3.2 Change the /etc/sudoers file

The file containing the user permissions on the linux system is the sudoers file. We need to give permission to the apache user who is known as www-data.

```
#seek the line: # User privilege specification; and add the follow line:
www-data ALL=(ALL) NOPASSWD:ALL
```

2.4 PHP 7.0

We noticed a problem of incompatibility with version 7.0 of php and some routines of the interface GATool. Therefore, users who are running version 7.0 of php should install the following package. You must also install a zip package for PHP 7.0

```
sudo su
{adm password}
sudo add-apt-repository -y ppa:ondrej/php && sudo apt update &&
apt install -y php7.0-mbstring php7.0-zip php7.0-xml
```

```
sudo apt-get install php7.0-zip
```


2.5 Running GATool

After installing the tools shown in the sections above, we can use GATool with all the features, as show in the Fig. 1. We can access GATool through the web with a browser: <http://localhost/gatool/>

The screenshot displays the GATool web interface. A modal window titled "GATool PIPELINE" is open, providing a detailed overview of the tool. The modal includes a "View" section with a description of GATool as a web interface integrating various tools for genome assembly. It lists "Some Features" such as full integration among all tools, support for single end reads, rapid quality analysis, and preprocessing. A "Workflow" section shows a flowchart of the process: Input leads to Analysis, which then leads to Preprocessing, followed by Genome Assembly, and finally Scaffolding, which loops back to Analysis. The "Tools" section lists FATSQC, FASTX-toolkit, SPAdes, and Velvet as integrated tools. The background interface shows a navigation bar with "Entry", "Analysis", "Trimmer", and "Assembly" tabs. The "Entry" tab is active, showing a "Send your Ion torrent reads" section with a file selection button and an "Upload" button. On the right, there is a "History" section with a "Restart process" button and a "Files to download" section with a "download" button. The footer of the page indicates the version is GATool - Genome Assembly Tool (Beta) - v 0.2 - Design by Matheus Brito de Oliveira.

GATool PIPELINE

View

GATool is a web interface which integrates several tools, with the purpose of to do a genome assembly task and provide to the users a nice and real experience. and a very fast way to genomes assembly.

Some Features

- Full integration among all tools
- Single end reads
- Allows a rapid quality analysis of reads
- Rapid preprocessing of the reads

Workflow

Bellow the flow of the GATool pipeline works

```
graph TD; Input --> Analysis; Analysis --> Preprocessing; Preprocessing --> Genome Assembly; Genome Assembly --> Scaffolding; Scaffolding --> Analysis;
```

Tools

GATool pipeline is equiped with several of differents tools. See below a resume of this tools:

- FATSQC - A tool of Quality Control of fragments
- FASTX-toolkit - A rapid kit for preprocessing files
- SPAdes Genome Assembler and Velvet Genome Assembler

© GATool - Genome Assembly Tool (Beta) - v 0.2 - Design by Matheus Brito de Oliveira