

AVE Allelic Variation Explorer

installation

Below you can find installation instructions with all necessary libraries.

ubuntu server 12.04 LTS

1. install few first prerequisites

```
sudo aptitude install git build-essential python-dev
```

2. install BEDTools

Follow instructions at <http://code.google.com/p/bedtools>. There is a version available in ubuntu repository, but it's usually not the most recent one. I recommend compiling as described at [BEDTools website](#)

To compile it, you need to install zlib1g-dev first

```
sudo aptitude install zlib1g-dev
```

3. install pip (pip installs packages)

```
sudo aptitude install python-pip
```

4. install virtualenv

Before installing all python stuff i recommend installing pythonenv, so that you can install python libraries separately from system ones.

```
sudo pip install virtualenv
virtualenv ave-env
source ave-env/bin/activate
```

5. install needed python libraries

```
pip install cython
pip install pybedtools
pip install numpy
pip install biopython
```

6. install mongod

Follow instructions on [MongoDB website](#).

7. install pymongo -- python mongo driver

```
pip install pymongo
```

8. install [node.js](#) -- server side javascript

Download package from the [node.js website](#).

Unpack and install.

```
tar xvzf node-<version>.tar.gz
cd node-<version>
./configure
make
sudo make install
```

Centos 5.8

1. install few first prerequisites

```
sudo yum install gcc zlib-devel openssl-devel cpio
sudo yum install expat-devel gettext-devel
```

2. install git

Git needs to be compiled from source.

Get the package from [Google Code Website](#). Adjust the url with the latest version number.

```
wget http://git-core.googlecode.com/files/git-1.7.12.1.tar.gz
```

Unpack and install.

```
tar xvzf git-<version>.tar.gz
cd git-<version>
./configure
make
sudo make install
```

3. install BEDTools with dependencies

```
sudo yum install gcc-c++
```

Install BEDTools according to [Installation Instructions](#)

4. install Python

Python in CentOS 5.8 is old. Install newest version from 2.7 branch (not 3), separately from system python, so that it does not mess centos.

Get the source from [Python Website](#).

Unpack and install.

```
tar xvzf Python-<version>.tgz
cd Python-<version>
./configure --prefix=/home/user/alt_python
make
make install
```

Use virtualenv to get whole python environment separate from system python.

Get virtualenv script from github repository and install virtualenv with alternative python version.

```
wget https://raw.githubusercontent.com/pypa/virtualenv/master/virtualenv.py
/home/user/alt_python/bin/python virtualenv.py -p /home/user/ave-env
```

Activate the environment

```
source /home/user/ave-env/bin/activate
```

5. install important python libraries

Remember to install these **after** activating the virtualenv environment.

```
pip install cython
pip install numpy
pip install biopython
pip install pybedtools
```

6. install mongodb

Follow the instructions on [MongoDB website](#). If `mongos start` (as suggested at website howto) does not work, try `/sbin/mongod start`.

7. instal mongodb python drivers

```
pip install pymongo
```

8. install `node.js` -- server side javascript

Dowload package from the [website](#).

Unpack and install.

```
tar xvzf node-<version>.tar.gz
cd node-<version>.tar.gz
./configure
make
sudo python tools/install.py install
```

setting up AVE

These instructions are independent of the operating system. It is important to work in virtualenv ('source ~/ave-env/bin/activate', as explained above).

1. Download the application.
2. Unpack the code and checkout the latest version branch

```
tar xvzf ave.tar.gz
cd ave
```

3. install node packages

```
npm install
```

4. Setup the db

You can setup the db with example mongodump from ave website. To do it download seqdb.tar.gz package. Unpack it and run

```
mongorestore ./seqdb
```

To setup the db with your own data you can use provided script. You will need:

- reference sequence in fasta format

make sure that name of the chromosome (or some other meaningful identifier) is

provided as fasta identifier (the string just after ">"). Like in the example for Chromosome 1 sequence:

```
>Chr1 CHROMOSOME dumped from ADB: Jun/20/09 14:53  
CCCTAAACCCTAAACCCTAAACCCTAAACCTCTGAATCCTTAATCCCTA
```

- gene annotations in [gff3 format](#)
- SNP annotations in [gff3 format](#)

SNPs should be annotated like in this example columns 1-7:

```
Chr1 1001Genomes SNP_adal_3 138 138 3 . .
```

column 8 (key value pairs):

```
Change=T:C;Strain=adal_3;Project=GMINordborg2010;ID=9323.138
```

First column should correspond to seq id from fasta file provided as reference.

In last column:

`Change` follows `reference:variant` order

`Strain` is the name of the strain/accession/ecotype in which this SNP have been called.

`Project` is the sequencing project

`ID` is any unique identifier for this SNP

You can annotate the SNPs in gff file with SNPs location. To do it run

```
python ./ave_tools.py snps_by_location --annot gene_annotation.gff \  
--snps snp_file1.gff --snp_file2.gff
```

The script generates new gff files, one for each snp location, with annotated location in last column:

```
Project=GMINordborg2010;Strain=ale_stenar_44_4;variant_location=CDS;  
ID=992.6992;Change=T:C
```

To import data into the database run:

```
python ./ave_tools.py import --genome TAIR10 --ref \  

```

```
reference.fas --annot gene_annotations.gff snps_annotations.gff
```

after `--genome` provide a name of the genome which was used to map the reads and call variants against

after `--ref` provide a list of fasta files with reference sequence

after `--annot` provide a list of files with gene/trait/snp annotations

starting up AVE

run:

```
node app.js
```

Access app from within web browser (preferably latest chrome). Ip address and port is provided in app.js output.

important ifo

Example SNP annotations have been obtained from [1001 Genomes Project](#). Please read the Data Usage Policy at the project website.