mADMIXTURE v.1.0.0

Author: Bryson Sjodin

August 29th, 2018

URL: https://github.com/bsjodin/mADMIXTURE

TABLE OF CONTENTS

DESCRIPTION

mADMIXTURE is a bash program for running **ADMIXTURE** across a range of k and/or over multiple iterations in a single command. It creates a set of organized directories based upon the chosen options (ie, using multiple k or iterations). This still requires **ADMIXTURE** to be installed, as **ADMIXTURE** is still doing all the heavy lifting. All arguments that can be passed to the standalone **ADMIXTURE** program can be supplied through this program as well. Please reference the **ADMIXTURE** manual for detailed descriptions of these arguments.

If you have any questions/concerns/recommendations for improvement, please contact me at bryson.sjodin@ubc.ca.

BEFORE USE

Before you begin using **mADMIXTURE**, you must first install **ADMIXTURE**, which can be downloaded from the following website:

<div align="center">

http://www.genetics.ucla.edu/software/admixture/

</div>

Refer to the **ADMIXTURE** manual for installation instructions.

Please ensure you cite **ADMIXTURE** in all publications:

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

DOWNLOAD & INSTALLATION

DOWNLOAD:
You can download **mADMIXTURE** from the GitHub directly at:

https://github.com/bsjodin/mADMIXTURE

Alternatively, you can clone the git repository by entering the following into a new terminal window:

```
$ git clone https://github.com/bsjodin/mADMIXTURE
```

EASY INSTALLATION:
mADMIXTURE can quickly and easily be installed with a few simple commands. First, navigate into the **mADMIXTURE** directory.

```
$ cd /path/to/directory/mADMIXTURE
```

Next, enter the following command into terminal to install **mADMIXTURE**:

```
$ bash install.sh
```

mADMIXTURE is now ready to use!

ADVANCED INSTALLATION:
Rather than use the "install.sh" script include, you can manually install **mADMIXTURE**.

First, change the file permissions:

```
$ cd mADMIXTURE
$ chmod 755 mADMIXTURE
```

Next, choose one of the following options for manual installation.

Option 1:
Copy this script into a PATH directory (eg. /usr/local/bin). This an be accomplished in terminal with:

```
$ sudo cp mADMIXTURE /path/to/directory/
```

Option 2:
Alternatively, you can add the directory where this script is saved to your PATH variable in terminal with:

```
$ cd ~
```

Then one of:

```
$ nano .bash_profile
```

<div align="center">OR</div>

```
$ nano .bashrc
```

depending on your operating system.

Scroll to the very end of the file and add the line:
```
export PATH=/path/to/directory:$PATH
```

Save and exit using Ctrl+c. Next type into terminal:
```
$ source .bash_profile
```
OR
```
$ source .bashrc
```
depending on which file you edited earlier.

NOTE: If installation fails for any reason, you can still run the program from the command line with:
```
$ bash mADMIXTURE …
```

USAGE

Run **mADMIXTURE** from the command line using the following:

```
$ madmixture -f input_file.ped -k integer [optional arguments]
```

mADMIXTURE requires that you specify both an input file (-f flag) as well as a minimum k value (-k flag). The input file should be a PLINK file with extension .ped, and alleles should be coded in the 1-2 format (see **ADMIXTURE** and PLINK manuals for reference).

In addition to the required arguments, **mADMIXTURE** can pass along arguments to any **ADMIXTURE** setting. See below for a complete list of options. For more details on each option, please reference the **ADMIXTURE** manual.

REQUIRED ARGUMENTS:
-f|--file <filename>   : Input PLINK file with extension (must be in current working directory).
-k|--minK <int>   : Minimum k. If -K is not used, this will be the only k tested.

OPTIONAL ARGUMENTS:
-K|--maxK <int>   : Maximum k. Optional. This must be larger than -k
-i|--iterations <int>   :Number of iterations. If not set, default is 1.
-j|--threads <int>   : Enables multi-threaded mode use n threads.
-cv <float/int>   : Enables cross-validation and sets the threshold. If no value is given with the -cv flag, threshold defaults to 10.
-B|--bootstraps <int>   : Enables bootstrapping and specifies number of replicates.
-S|--supervised   : Enables supervised mode. Must include a ".pop" file in current working directory.
-P|--projection   : Enables projection analysis. Must include a reference ".P.in" file in current working directory.
-s|--seed <int/str>   : Sets the random seed. If no value is given with the -s flag, defaults to shell variable $RANDOM.
-EM|--algorithm   : Use EM algorithm instead of block-relaxation.
-C|--majC <float/int>   : Sets the major termination criterion.
-c|--minC <float/int>   : Sets the minor termination criterion.

-a|--accel \<str\>   : Changes the acceleration method used
-e|--epsilon \<float\>   : Enables penalized estimation and sets epsilon (must be less than 1).
-l|--lambda \<float\>   : Enables penalized estimation and sets lambda.
--haploid   : Use haploid data (default is diploid).
--help   : Displays help message.
--sort-by-k   : Sort outputs by k (default sorts by run). Only valid for multiple k AND
                            iterations.
--version   : Displays version information.

OUTPUTS

Single k; one iteration:
If only a single k and one iteration are performed, **ADMIXTURE** will run normally, with
outputs written to the current working directory and STDOUT written to a ".log" file.

Multiple k:
If a range of k is specified (ie. both the "-k" and "-K" flags are used), outputs will be sorted into
directories based on file extension. Files with the ".Q" extension will be moved to the new
directory "Q_outs", those with the ".P" extension will be moved to "P_outs", and finally, ".log"
files will be moved to "logs".

Multiple iterations:
If more than one iteration for each k is specified (ie. "-i" flag set greater than 1), outputs for
each iteration will be sorted into separate directories labeled "run_1"," run_2", etc.

Multiple k; multiple iterations:
If both a range of k and multiple iterations are specified, outputs from each iteration will be
located within separate directories by iteration (run_1, run_2, etc.), and individual files within a
run directory will be sorted based on extension into "Q_outs", "P_outs", or "logs". Optionally,
if you use the "--sort-by-k" option, outputs will be sorted by k-value (labelled K1, K2, etc.),
with separate subdirectories for each run.

CV threshold:
If the "-cv" flag is specified, cross-validation scores for each iteration will be saved to a file
"CV_scores$n$.txt", where n designates the iteration number. These files will be located in each
run directory. If multiple iterations are specified, all "CV_scores.txt" files will be concatenated
in the current working directory to a single file named "allCV.txt". Additionally, a file labeled
"meanCV.txt" will be created with the mean CV score for each k-value.

CITATION

If you use this software, please cite the original **ADMIXTURE** paper, as it is really doing all the work. This is an open-source software with no publication associated with **mADMIXTURE**, but a citation/acknowledgement in any publication would be appreciated:

Eg). Sjodin, B. (2018). mADMIXTURE (v.1.0.0) [Computer software]. Retrieved from
https://github.com/bsjodin/mADMIXTURE.