REAP Software Documentation

Version 1.2

Timothy Thornton[1]

Department of Biostatistics[1]
The University of Washington

REAP

A C program for estimating kinship coefficients and IBD sharing probabilities in samples with admixed ancestry.

Homepage: http://faculty.washington.edu/tathornt/software/REAP/index.html

Release 1.2 April 23, 2013

==========

License

We request that use of this software be cited in publications as follows:

> Thornton T, Tang H, Hoffman TJ, Ochs-Balcom HM, Baan BJ, and Risch N (2012) "Estimating Kinship in Admixed Populations", American Journal of Human Genetics, vol 91 pp. 122-138.

To contact the first author:

Timothy A. Thornton
Department of Biostatistics
University of Washington
Health Sciences Building F-600
Box 357232
Seattle, WA 98195-7232

email: tathornt@u.washington.edu

Contents

1 Overview of REAP

REAP (Relatedness Estimation in Admixed Populations) is a program, written in C, that estimates autosomal kinship coefficients and identity-by-descent (IBD) sharing probabilities using SNP genotype data in samples with admixed ancestry. Some of the features of the program include:

1. supports PLINK's (Purcell et al., 2007) transposed PED (tped) file as the input SNP genotype data file, to allow for the analysis of millions of SNPs without excessive memory allocation, and PLINK's tfam file (which provides family and study ID information of the sample individuals).
2. provides estimation of
   a. pairwise kinship coefficients
   b. pairwise IBD sharing probabilities (probability that a pair of individuals share 0, 1, or 2 alleles IBD).
   c. inbreeding coefficients
3. appropriate handling of missing genotype data
4. user option for specifying a subset of individuals from the sample to include in the relatedness estimation analysis

2 Running REAP

Installation instructions:

1. Download the REAP package. This package contains documentation, source code, example input and output files, and a precompiled executable for Linux 64-bit platforms.
2. Read the file REAP_Documentation.pdf carefully to understand the purpose of this program and how it works.
3. Edit the Makefile as necessary according to the instructions in the Makefile. You should only need to make sure that the correct compiler and compiler options for your machine are chosen.
4. Type "make". This will build an executable program called "REAP". If the message "make: 'REAP' is up to date" appears after typing "make", then to build the executable program you must first delete the precompiled binary REAP program that comes with the software by typing "rm REAP", and then type "make" to build the executable program REAP.

5. REAP is run from the command line via the command 'REAP' with all information, including the type of anlaysis, specified by command line options. To run the executable program REAP:

   First, prepare the input files, e.g., PLINK tped and tfam files, individual ancestry file, and the subpopulation/ancestral population allele frequency file (see Section 3 for more details).

   Then, to run REAP with the default input filenames and settings, one need only type

   ./REAP

   Alternatively, to change input filenames or settings, use flags in the command line. The following flags are available:

   ./REAP -g example.tped -p example.tfam -a admixproportionfile -f allelefreqfile -k 2 -s sampleincludefile -t 0.025 -r 2 -m

   We briefly summarize the meanings of the flags below. More details can be found in section 3:

   -g example.tped Allows the user to specify the name of the SNP genotype data input file that is in the PLINK tped format. Filename defaults to "example.tped" if this flag is not used. To specify a different filename, replace "example.tped" with the appropriate filename.

   -p example.tfam Allows the user to specify the name of the family and study ID information file which is in the PLINK tfam format. This file includes family ID numbers, individual ID numbers, mother ID, father ID, sex, and a phenotype value. Only the first two columns of this file, family ID and individual ID, will be used in the REAP analysis. The filename defaults to "example.tfam". To specify a different filename, replace "example.tfam" with the appropriate filename.

   -a admixproportionfile Allows the user to specify the name of the individual ancestry proportion file, which contains the ancestry admixture proportions for each sample individual. The filename defaults to "admixproportionfile". To specify a different filename, replace "admixproportionfile" with the appropriate filename.

   -f allelefreqfile Allows the user to specify the name of the allele frequency file containing the allele frequencies for each of the subpopulations/ancestral populations. Filename defaults to "allelefreqfile". To specify a different filename, replace "allelefreqfile" with the appropriate filename.

   -k 2 Allows the user to specify the number of subpopulations/ancestral populations that are represented in the individual ancestry proportion file and the allele frequency file. The default number is 2. To specify a different number, replace "2" with the appropriate number. For example, "-k 4" in the command line specifies that there are 4 subpopulations/ancestral populations represented in the individual ancestry proportion file and the allele frequency file. If this flag

is not used, then 2, the default number of subpopulations/ancestral populations, will be used.

-r 2 Allows the user to specify the reference allele, either 1 or 2, that was used for the allele frequency file. The default allele is 2. To specify the reference allele to be 1 for the allele frequency file, replace "2" with the "1". For example, "-r 1" in the command line specifies that allele 1 is the reference allele in the allele frequency file. If this flag is not used, then 2, the default reference allele, will be used.

-s sampleincludefile Allows the user to specify a file containing a list of individuals to include in the relatedness estimation analysis. All individuals will be included in the relatedness estimation analysis if this option is not used.

-m Allows the user to bypass the printing of the kinship coefficient matrix and the IBD 0 probability matrix to files. The kinship coefficient and IBD 0 sharing probability matrices will be printed to files named "REAP_Kincoef_matrix.txt" and "REAP_IBD0_matrix.txt", respectively, if this option is not used.

-t 0.025 Allows the user to specify the minimum kinship coefficient threshold to use for printing relatedness estimates of pairs to an output file named "REAP_pairs_relatedness.txt." The default minimum kinship coefficient value is set to .025, and all pairs of individuals who have a kinship coefficient estimate that is above this threshold will have kinship coefficients and IBD sharing probabilities printed to the file. To specify a different minimum kinship coefficient threshold, replace ".025" with the appropriate number. For example, "-t 0.05" in the command line specifies that the minimum kinship coefficient threshold is set to 0.05. If this flag is not used, then 0.025, the default minimum kinship threshold, will be used.

6. You can test the executable program REAP by running it with the sample input files example.tped, example.tfam, admixproportionfile, and allelefreqfile by typing

   ./REAP

   You can then compare the resulting output, which will be printed to the files REAP_pairs_relatedness.txt, REAP_Kincoef_matrix.txt, REAP_IBD0_matrix.txt, REAP_Inbreed.txt, and REAP_Individual_Index.txt with the correct output provided in the sample output files REAP_pairs_relatedness.txt.ex, REAP_Kincoef_matrix.txt.ex, REAP_IBD0_matrix.txt.ex, REAP_Inbreed.txt .ex, and REAP_Individual_Index.txt.ex, respectively.

7. The program stops if any errors are detected in the format of the input files.

3 Input

Required Input Files:

1. **genotype data file**

   The genotype data file is a transposed genotype file containing the SNP names and locations and the genotypes of the sampled individuals. The genotype data file is in the PLINK tped file format. The two alleles of a SNP must be coded as 1 and 2, and missing alleles must be coded as 0.

   To illustrate the format of the genotype data file, consider a study sample with a total of 8 individuals. The first few rows of the genotype data file for this sample could be as follows:

1	rs3094315	0	742429	1	2	2	2	1	1	0	0	1	1	1	2	1	1	1	2
1	rs2286139	0	751595	1	1	1	1	1	1	0	0	1	2	0	0	1	1	1	2
1	rs11240776	0	755132	2	1	2	1	1	2	1	2	1	1	1	1	1	1	1	2
1	rs2980300	0	775852	0	0	2	1	1	1	2	1	2	1	2	2	1	1	1	2
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)

   Column (1) contains the chromosome name (1-22). This information will be ignored by the program (but some string must be present in the column).
   Column (2) contains the rs number or SNP identifier. This information will be ignored by the program (but some string must be present in the column).
   Column (3) contains the genetic distance in Morgans (0 if genetic distance is unknown). This information will be ignored by the program (but some string must be present in the column).
   Column (4) contains the base-pair position in bp units (0 if base-pair position is unknown). This information will be ignored by the program (but some string must be present in the column).
   Columns (5) and (6) contain the marker genotype (one allele in each column) for the 1st individual.
   Columns (7) and (8) contain the marker genotype for the 2nd individual.
   ⋮
   Columns (17) and (18) contain the marker genotype for the 7th individual.
   Columns (19) and (20) contain the marker genotype for the 8th individual.

   For more details on the tped file format, you could consult the PLINK website (http://pngu.mgh.harvard.edu/ purcell/plink/data.shtml#tr). PLINK provides a convenient venue to convert from many different file formats. For example, supposed you had a PLINK ped file named "mydata.ped" which was coded as A/C/G/T or 1/2/3/4 for the four alleles. Then, assuming that the file contains only autosomal SNPs, you could generate the desired REAP genotype input file with the PLINK command:

   ./plink --file mydata --recode12 --output-missing-genotype 0 --transpose --out newfile

   The PLINK software would then create the two files "newfile.tped" and "newfile.tfam". The file "newfile.tped" is a genotype data file that is in the appropriate format for the REAP software. The tfam file "newfile.tfam" is a family and study ID information file

(discussed below in the next subsection) that is also in the appropriate format for the REAP software.

The default filename for the genotype data file is "example.tped". To specify a different filename, use the command-line flag -g followed by the filename. For example, to use the PLINK SNP genotype data file named "newfile.tped", you could type the command

./REAP -g newfile.tped

2. **family and individual ID information file**

This file contains family IDs, individual IDs, parents' IDs, sex, and phenotype values, where individuals must be listed in the same order as in the genotype data file. This file is in the PLINK tfam format and can be created with the software PLINK using the command given in the previous subsection. To illustrate the required format of a PLINK tfam file, consider a study sample with 8 individuals from 3 families. The lines in the PLINK tfam file could be as follows:

F192	ID101	0	0	1	1
F192	ID102	0	0	2	2
F192	ID201	ID101	ID102	2	2
M012	205HG	0	0	1	2
M012	ID309	0	0	2	1
M012	ID2999	205HG	ID309	1	0
M012	ID9203	205HG	ID309	2	1
S230	LF6950	0	0	2	0
(1)	(2)	(3)	(4)	(5)	(6)

Column (1) contains family ID.
Column (2) contain individual ID.
Column (3) contain father's ID. This information will be ignored by the program (but some string must be present in the column).
Column (4) contains mother's ID. This information will be ignored by the program (but some string must be present in the column).
Column (5) contains sex. This information will be ignored by the program (but some string must be present in the column).
Column (6) contains phenotype value. This information will be ignored by the program (but some string must be present in the column).

Each family ID is assumed to be a character string that does not contain a space, newline or tab. The same assumption holds for each individual ID.

The default filename for the phenotype information file used by REAP is "example.tfam". To specify a different filename, use the command-line flag -p followed by the filename. For example, to use the PLINK file "newfile.tfam" file discussed in the previous subsection, you could type the command

./REAP -p newfile.tfam

3. **individual ancestry proportion file**

The file consists of the ancestry proportion estimates for each of the study individuals. Individuals must be listed in the same order as the input PLINK tfam file. To illustrate the required format of the individual ancestry proportion file, consider a study sample with 8 individuals from 3 families. If the the number of subpopulations/ancestral populations is specified to be 3, for example, in the relatedness estimation analysis, then the columns in the individual ancestry proportion file should be organized as follows:

F192	ID101	0.255615	0.0229269	0.721458
F192	ID102	0.367986	0.0619604	0.570053
F192	ID201	0.49122	0.0365117	0.472269
M012	205HG	0.307492	0.0226551	0.669853
M012	ID309	0.717561	0.0143852	0.268054
M012	ID2999	0.220191	0.062522	0.717287
M012	ID9203	0.382879	0.0241328	0.592988
S230	LF6950	0.789815	0.027846	0.182339
(1)	(2)	(3)	(4)	(5)

Column (1) contains family ID.
Column (2) contain individual ID.
Column (3) contains proportional ancestry from population A.
Column (4) contains proportional ancestry from population B.
Column (5) contains proportional ancestry from population C.

The proportional ancestry values should sum to 1 for each of the individuals in this file.

REAP supports as input an individual ancestry proportion file from the FRAPPE software (Tang et al., 2005), which is similar to the format given above but with a colon, ":", appearing between the individual ID and the first proportional ancestry value, i.e., there would be a colon between columns 2 and 3 above for an individual ancestry file obtained from the FRAPPE software. The REAP software automatically detects a FRAPPE individual ancestry proportion file and will appropriately read in the ancestry proportions for file of this type. The FRAPPE software can be downloaded at

http://med.stanford.edu/tanglab/software/frappe.html

The ADMIXTURE software (Alexander et al., 2009) can also be used to obtain individual ancestry proportions for a sample individual. The individual ancestry proportion file from the ADMIXTURE software, however, does not include family ID and individual ID of the sample individuals, so this output file from ADMIXTURE is not in the appropriate format for the REAP software. Please see section 6 for details on how to easily format the individual ancestry proportion file from ADMIXTURE for the REAP

software. The ADMIXTURE software can be downloaded at

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

The default filename for the individual ancestry file used by REAP is "admixproportionfile". To specify a different filename, use the command-line flag -a followed by the filename. For example, to use the file "newancestryfile" as the individual ancestry proportion file in the analysis, you could type the command

./REAP -a newancestryfile

4. **allele frequency file**

The allele frequency file consists of the allele frequencies at the SNPs for each of the subpopulations/ancestral populations. Each row corresponds to a SNP, and if there are "K" ancestral populations specified by the user for the analysis, then each row in the allele frequency file must have K allele frequency values, i.e., an allele frequency for each of the K subpopulations/ancestral populations. Important restrictions to keep in mind are:

a. Each row (i.e., SNP) in the allele frequency file must correspond to the same row (i.e., SNP) in the genotype data file, e.g., row 10 (the 10th SNP) in the allele frequency files correspond to the same SNP given by row 10 in the genotype data file.

b. The allele frequencies given in the allele frequency file must be based on the same reference allele, either 1 or 2, for the SNPs given in the genotype data file. The default reference allele is 2, which specifies that the allele frequencies for the SNPs in the allele frequency file corresponds to allele 2 in the genotype data file. As discussed in section 2 above, "Running REAP", the '-r 1" option in REAP is used to specify that allele 1 is the reference allele for the allele frequency file.

c. The order of the populations in the allele frequency file must be in the same order as the populations in the individual ancestry proportion file. For example, if the columns in the individual ancestry proportion file are ordered as follows: (1) population A ancestry proportion, (2) population B ancestry proportion, and (3) population C ancestry proportion, then the columns in the allele frequency file must ordered as follows: (1) population A allele frequency , (2) population B allele frequency, and (3) population C allele frequency.

To illustrate the required format of the allele frequency file, if the number of ancestral populations is specified to be 3 for the analysis, the first few lines of the allele frequency file could be as follows:

0.420	0.664	0.286
0.701	0.249	0.541
0.581	0.335	0.857
0.784	0.045	0.414
0.300	0.157	0.729
0.601	0.395	0.882
0.285	0.927	0.536

Column (1) contains allele frequencies for population A.
Column (2) contains allele frequencies for population B.
Column (3) contains allele frequencies for population C.

REAP supports as input a subpopulation/ancestral population allele frequency file from the FRAPPE software (Tang et al., 2005) and the ADMIXTURE software (Alexander et al., 2009).

Optional Input:

5. **specify a subset of individuals for the relatedness analysis**

   This optional input file contains the family and individual IDs of the study individuals who will be included in the relatedness analysis. The family and individual IDs of the individuals should be in two columns, where the first column contains the family IDs and the second column contains the individual IDs, and is the same format as the first two columns of a PLINK tfam file previously discussed. If this option is not used, all individuals in the sample will be included in the relatedness analysis. To specify a file that contains a list of individuals to include in the relatedness estimation analysis, use the command-line flag -s followed by the filename. For example, to use a file named "includesample" that contains a list of family and individual IDs, you could type the command

   ./REAP -s includesample

6. **bypass printing of the relatedness estimation matrices to files**

   The command-line flag -m would be used when the user would like to bypass printing the kinship coefficient matrix and the IBD 0 probability matrix to files. For example, if you do not want the matrices to be printed to files, you could type the command

   ./REAP -m

5. **set the minimum kinship coefficient threshold**

   The command-line flag -t can be used to specify the minimum kinship coefficient value that will be used for printing pairs of individuals to the output file "REAP_pairs_relatedness.txt"

(discussed in the next section) The default minimum kinship coefficient value is set to .025, and all pairs of individuals that have a kinship coefficient estimate that is at least as large as this value will have their kinship coefficients and IBD sharing probabilities printed to the file REAP_pairs_relatedness.txt. To specify a different minimum threshold, for example, 0.05, you could type the command

. /REAP -t 0.05

4 Output

1. **REAP_pairs_relatedness.txt** provides relatedness estimates for all pairs of individuals who have a REAP estimated kinship coefficient that is greater the the mininum kinship coefficient threshold, where the default value of 0.025 will be used if the threshold is not specified by the user. Each row in this file provides information for a pair of individuals. The columns are

   FID1: Family ID for the first individual of the pair

   ID1: Individual ID for the first individual of the pair

   FID2: Family ID for the second individual of the pair

   ID2: Individual ID for the second individual of the pair

   N_SNP: The number of SNPs that do not have missing genotypes in either of the individuals and for which the individual-specific minor allele frequencies at the SNPS for both individuals (calculated using individual ancestry proportions and the subpopulation/ancestral population allele frequencies) are greater than or equal to .01.

   IBD0_PROB: REAP estimated probability of IBD=0

   IBD1_PROB: REAP estimated probability of IBD=1

   IBD2:_PROB REAP estimated probability of IBD=2

   KINCOEF: REAP estimated kinship coefficient

   The IBD sharing probabilities are truncated at 0 and 1 in this file to ensure that the IBD sharing probabilities sum to 1.

2. **REAP_Kincoef_matrix.txt** contains an $N \times N$ matrix of REAP estimated kinship coefficients, where N is the number of individuals included in the relatedness estimation analysis. The REAP kinship coefficient estimates are not truncated at 0 and 1.

3. **REAP_IBD0_matrix.txt** contains an $N \times N$ matrix of REAP estimated IBD 0 probabilities, where N is the number of individuals included in the relatedness estimation analysis. The REAP IBD 0 sharing probability estimates are are not truncated at 0 and 1.

4. **REAP_Individual_Index.txt** gives the index of each individual for the kinship coefficient and IBD 0 sharing probability matrices that are printed to the files REAP_Kincoef_matrix and REAP_IBD0_matrix.txt, respectively. For example, the kinship coefficient value given in row 2 and column 5 in the REAP_Kincoef_matrix.txt file would correspond to the individuals in the REAP_Individual_Order.txt file who have an indices of 2 and 5, respectively. Each row provides information for an individual. The columns are

   FID: Family ID

   ID: Individual ID

   Matrix_Index: gives the row (and column) position of an individual for the kinship coefficient matrix and the IBD 0 sharing probability matrix.

5. **REAP_Inbreed.txt** gives the REAP estimated inbreeding coefficients for each individual in the sample included in the relatedness analysis. Each row provides information for an individual. The columns are

   FID: Family ID

   ID: Individual ID

   N_SNP: The number of SNPs that do not have missing genotypes for an individual and for which the individual-specific minor allele frequencies at the SNPS for an individuals (calculated using individual ancestry proportions and the subpopulation/ancestral population allele frequencies) are greater than or equal to .01.

   INBREEDCOEF: gives the REAP inbreeding coefficient estimate

5 Tips

1. Input

   The program will stop if errors are detected in the formats of any of the input files. Please read Section 3 carefully and make sure the input files are in the correct format and have concordant information.

2. Set the minimum kinship coefficient value to a negative number that is large in magnitude if it is desired to have relatedness estimates for all possible pairs printed to the file "REAP_pairs_relatedness.txt. For example, using "-t -100" in the command line specifies that the minimum kinship coefficient threshold is -100, which will be an extreme minimum kinship coefficient threshold for which the kinship coefficients for all possible pairs should be larger than.

3. The number of study individuals for the REAP analysis is set to be smaller than 10,000. To increase this limit, just change the value of MAXPEOPLE in the REAP_SOURCE.c file (seven lines from the top) and recompile the program.

6 Example

1. Consider a PLINK ped file named "mydata.ped". Then the PLINK command below can be used to obtain tped and tfam output files:

   ./plink --file mydata --recode12 -- output-missing-genotype 0 --transpose --out newfile

   This command creates the two files "newfile.tped" and "newfile.tfam" that is in the appropriate format for the REAPsoftware package.

2. Both the FRAPPE software and the ADMIXTURE software can be used to obtain individual ancestry proportions and subpopulation/ancestral population allele frequency estimates. Below are instructions for using the output from the FRAPPE software followed by instructions for using output from the ADMIXTURE software with REAP.

   (a) **Instructions when using the FRAPPE software:**

   Follow the instructions given in the FRAPPE software documentation and using the mydata.ped file discussed in item 1 above, the following command can be used in PLINK to create the appropriate input genotype file for the FRAPPE software:

   ./plink --file mydata –recode12 –out frappedata

   This PLINK command will create the desired FRAPPE input genotype data file name "frappedata.ped". Using the FRAPPE software with the genotype data file frappedata.ped, the output individual ancestry file will be "frappedata_result.txt" and the ancestral/subpopulation allele frequency file will be "cP.txt". (Note, to do the FRAPPE analysis, you must also create additional input files. Please read the FRAPPE documentation carefully for more information on this). Now, to run the REAP software with these two output FRAPPE files as well as the input files "newfile.tped" and "newfile.tfam" files from item 1 above, the following command can be used if the number of ancestral populations is 3:

   ./REAP -g newfile.tped -p newfile.tfam -a frappedata_result.txt -f cP.txt -k 3

   Note that the reference allele is 2 in the FRAPPE output allele frequency file cP.txt, which is also the default reference allele for REAP, so the '-r 2' option is not needed in the above command.

   (b) **Instructions when using the ADMIXTURE software:**

   Follow the instructions given in the ADMIXTURE software documentation and using the mydata.ped file discussed in item 1 above, the following command can be used in PLINK to create the appropriate input genotype file for the ADMIXTURE software with alleles recoded as "1" and "2":

./plink --file mydata –recode12 –out admixturedata

This PLINK command will create the desired ADMIXTURE input genotype data file name "admixturedata.ped". The following ADMIXTURE software command can be used for estimating admixture proportions from 3 ancestral populations:

./admixture admixturedata.ped 3

The output individual ancestry proportion file will be "admixturedata.3.Q" and the output ancestral/subpopulation allele frequency file will be "admixturedata.3.P". The ADMIXTURE output individual ancestry proportion file "admixturedata.3.Q" is not in the proper format for REAP, as previously mentioned in section 3 above. An easy way to obtain an individual ancestry proportion file that is in the appropriate format for REAP when using the ADMIXTURE software is to first create a file containing the family ID and individual ID of the sample individuals, e.g., the first two columns of the mydata.ped file (which corresponds to the first two columns of the "newfile.tfam" file as well as the "admixturedata.ped" file), and then use the UNIX paste command with this file and the ADMIXTURE output file "admixturedata.3.Q". The unix paste command stacks files column-wise. For example if a file named "myID.txt" contains two columns where the first column is family ID and the second column is individual ID for the sample individuals, then the following command in UNIX can be used to create an individual ancestry proportion file named "admixturedata.proportions" that is in the appropriate format for REAP:

./paste myID.txt admixturedata.3.Q > admixturedata.proportions

Now, to run the REAP software with the ADMIXTURE output allele frequency file "admixturedata.3.P" and the individual ancestry proportion file "admixturedata.proportions" created using the paste command, as well as the input files "newfile.tped" and "newfile.tfam" files from item 1 above, the following command can be used if the number of ancestral populations is 3:

./REAP -g newfile.tped -p newfile.tfam -a admixturedata.proportions -f admixturedata.3.P -r 1 -k 3

Note that the the reference allele will be 1 (which is the minor allele) in the ADMIXTURE output allele frequency file "admixturedata.3.P" so the '-r 1' option must be used in the command line for REAP in order to specify that the reference allele is different from the default reference allele of 2.

(Note that one can also perform a supervised analysis with the ADMIXTURE software by using the "–-supervised" option which requires an additional file with a .pop suffix specifying the ancestries of the reference individuals. Please read the ADMIXTURE documentation carefully for more information on this).

3. Alternatively, if one has an individual ancestry proportion file named "myancestryadmixture" that contains ancestry admixture proportions for all of the sample individu-

als listed in "newfile.tfam" as well as an allele frequency file named "populationfreq" containing allele frequencies for the subpopulations/ancestral populations, and if the number of ancestral populations is 3 and the reference allele is 1 in the allele frequency file populationfreq, then the following command can be used:

./REAP -g newfile.tped -p newfile.tfam -a myancestryadmixture -f populationfreq -k 3 -r 1

7 References

1. Alexander DH , Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655-1664.

2. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81, 559-575.

3. Tang H, Peng J., Wang P., and Risch N. (2005) Estimation of Individual Admixture: Analytical and Study Design Considerations. Genet Epidemiol. 28, 289-301.

4. Thornton T, Tang H, Hoffman TJ, Ochs-Balcom HM, Baan BJ, and Risch N (2012) "Estimating Kinship in Admixed Populations", Am. J. Hum. Genet. 91, 122-138.