

PedFAM User Manual - v0.1

Yiming Zhang and Yufeng Wu

1 Introduction

PedFAM is a probabilistic model-based software tool, which is designed to infer the founders' information of an admixed individual. PedFAM is an extended version of our previous tools PedMix [1] (which can estimate the admixture proportion of recent ancestors from a single child), and parMix [2] (which can infer parental ancestry and call parental genotypes from data of a small number of children). In the current version, PedFAM uses the extant individual's genotypes and the reference allele frequencies to infer the configuration and ratio of founders (we define founders as the first-generation ancestors of the extant individual that are not admixed).

2 Prerequisite

Python version in 3.8.5 has been used to compile PedFAM successfully. The numpy version later than 1.22.1 and the numba version later than 0.56.2 are required for running PedFAM. numba is a decorator for accelerating the PedFAM. For installing the prerequisites, simply run the following command:

```
$ pip install numpy numba
```

3 Download

Source code now available: <https://github.com/YimingZackZhang/PedFAM>.

4 Inputs

4.1 Input Files

PedFAM needs three different input files, and they are shown as follows.

1. Phased genotypes file of the extant individual. Take *example/Geno_C1.dat* for example:

$$Genotypes \begin{cases} first\ haplotype : 0\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0 \\ second\ haplotype : 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1 \end{cases} \quad (1)$$

(2)

2. Position file of SNPs. The physical position of each SNP. Take *example/Position_C1.dat* for example:

position 1: 1765.000353
position 2: 6150.001230
position 3: 9990.001998
position 4: 10480.002096
position 5: 12740.002548
position 6: 13290.002658
position 7: 17030.003406
position 8: 17610.003522
position 9: 22615.004523
position 10: 22850.004570

3. Allele frequency file of reference populations. Take *example/AF_C1.dat* for example:

$$Two\ references \begin{cases} population\ A : 0.990050\ 0.472637\ 0.997512\ 0.781095\ 0.997512\ 0.519900\ 0.997512 \\ population\ B : 0.997512\ 0.982587\ 0.942786\ 0.997512\ 0.594527\ 0.002488\ 0.987562 \end{cases} \quad (3)$$

(4)

5 Usage

For using PedFAM to infer the founders' configuration, first simply type:

```
$ python PedFAM.py -h
```

-g	Number of generations since admixture
-b	Number of blocks per chromosome
-r	Recombination Rate (per base pair per generation)
-c	Number of chromosomes
-p	Number of reference panels
-s	Number of random starting points (Default = 5)

Table 1: the arguments for running PedFAM

For example, the command line for running the example data set looks like this:

```
$ python PedFAM.py -g 5 -b 30 -r 1e-8 -c 1 -p 2 -s 6
```

In this case, PedFAM will infer the founder's configuration in 5 generations ago (-g 5) using 1 chromosome (-c 1). There are 2 reference panels (-p 2), and PedFAM divides each chromosome with 30 blocks (-b 30). The recombination rate is 10^{-8} (-r 1e-8), and PedFAM will run the hill-climbing algorithm 6 times with different random start points (-s 6)

6 Outputs

The PedFAM will output 6 different founder configurations, and pick one with the highest probability then output.

The optimal founders' configuration in the example will be:

```
$ [0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0]
```

with the log probability as -641.0608663128091 .

7 How to cite

The paper, "**Inference of Founding Ancestors of an Admixed Individual**" by Yiming Zhang and Yufeng Wu, is in progress, and we will keep updating the information. Please feel free to contact *Yiming Zhang* via yiming.zhang@cse.uconn.edu or *Yufeng Wu* via yufeng.wu@uconn.edu if you have any questions about PedFAM.

References

- [1] Jingwen Pei, Yiming Zhang, Rasmus Nielsen, and Yufeng Wu. Inferring the ancestry of parents and grandparents from genetic data. *PLoS computational biology*, 16(8):e1008065, 2020.
- [2] Yiming Zhang and Yufeng Wu. Joint inference of ancestry and genotypes of parents from children. *Isience*, 25(8):104768, 2022.