

Tutorial for AdmixInfer v1.0.4

Short Description:

AdmixInfer is designed to optimize the parameters of admixture model via maximum likelihood estimation and figure out the model best fit the data. The optimization is under assumption of HI, GA, CGFR and CGFD model.

1.Compile

1.1 Library dependency

AdmixInfer depend on boost library, make sure the boost is installed . Installation of boost can be found at http://www.boost.org/doc/libs/1_58_0/more/getting_started/unix-variants.html

1.2 Compile from source code

It's very easy to compile from the source code by the following commands:

```
bash$ tar -zxvf AdmixInfer.tar.gz
bash$ cd AdmixInfer/src
bash$ make
```

After compiling, you will get the executable AdmixInfer, just typing the command below to get help information:

```
bash$ ./AdmixInfer -h or bash$ ./AdmixInfer --help
```

2. Test with the toy data

2.1 A simple example

```
bash$ ./AdmixInfer --input ../example/sim1.seg
```

Example explanation:

AdmixInfer will read the ancestral tracks from sim1.seg, after a while, the optimal model and corresponding generation will print to screen. The format will explained later .

The following is output of the toy data:

```
//CMD: ./AdmixInfer-fast -i example/sim1.seg
```

Reading ancestral tracks from example/sim1.seg

Start optimizing process...

Results Summary

Parental populations: 1, 2

Admixture proportions: 0.287948, 0.712052

Optimal-model: HI, generation: 51, likelihood: 50290.4

The results show that the toy data. The simulated admixed population has two parental populations with label 1 and 2, corresponding mixture proportions are 0.287948 and 0.712052, the optimal model is HI (Hybrid Isolation), and the generation is 51. The simulated population is under HI model and 30% and 70% mixture proportions, admixture started 50 generations ago.

User can redirect the output to a file, such as:

```
bash$ ./AdmixInfer --input ../example/sim1.seg > sim1_opt.log
```

2.2 A full arguments example

```
bash$ ./AdmixInfer -i ../example/sim1.seg -l 0.01 -u 0.5 -a 0.01 -b 100 -p 0.5 > sim1_fopt.log
```

Example explanation:

Again, AdmixInfer read ancestral tracks from file sim1.seg, discard the tracks shorter than 0.01 Morgan and longer than 0.5 Morgan, the significance level to compute confidence intervals is 0.01 (which will compute 99% confidence intervals), the number of bootstrapping is 100, and 50% tracks will sampled in each bootstrapping. Finally, the outputs will be redirected to sim1_fopt.log.

The results is:

```
//CMD: ./AdmixInfer-fast -i example/sim1.seg -l 0.01 -u 0.5 -a 0.01 -b 100 -p 0.5
```

Reading ancestral tracks from example/sim1.seg

```
Start optimizing process...
Start bootstrapping process...
Bootstrap-1--> model: HI, generation: 51, likelihood: 19003.7
Bootstrap-2--> model: HI, generation: 51, likelihood: 18915.4
.....
Bootstrap-100--> model: HI, generation: 51, likelihood: 18999.4
Finish bootstrapping, processing results...
```

Results Summary

```
-----
Parental populations: 1, 2
Admixture proportions: 0.287948, 0.712052
Optimal-model: HI(100%)
Generation: 51.08, 99% confidence interval: (50.92, 51.24)
-----
```

Again, the optimal model we inferred is HI and generation estimated is 51.08 the 99% confidence intervals are (50.92, 51.24).

3. File format

3.1 Input file format

AdmixInfer is easy to use, only need one file, in which each line represents a ancestral track with the start point, end points, from which ancestry the track originates. The start and end points units are in Morgan.

For example:

```
0.00000000    0.34602058    Yoruba
0.34602058    0.34614778    French
.....
0.40759031    0.41517938    Yoruba
```

4. Arguments

`-i/--input` `<string>`

This argument is required, in which user specify the filename of input ancestral tracks, format described above.

`-l/--lower` `[double]`

This argument is optional, in which user specify the lower bound to discard short tracks. The default is 0, which does not discard any short tracks. However, due to method limitation in local ancestry inference, very short tracks are generally not reliable.

`-u/--upper` `[double]`

This argument is optional, in which user specify the upper bound to discard long tracks. The default is infinity, which does not discard any long tracks.

`-a/--alpha` `[double]`

This argument is optional, in which user specify the significance level to compute confidence intervals, this argument is used with bootstrap. When the number of bootstrap is 1, AdmixInfer does not compute confidence intervals. The default is 0.05, which will compute the 95% confidence intervals.

`-b/--bootstrap` `[integer]`

This argument is optional, in which user specify the number of bootstrapping repeats. The default is 1, which does not perform bootstrapping.

`-p/--proportion` `[double]`

This argument is optional, in which user specify the proportion of tracks in bootstrapping sampled. The default 1.0, which samples equal size of tracks in each bootstrapping repeat.

5. Options:

`-h/--help`

Print help messages of all arguments and options

`-q/--quiet`

Running in quiet mode, default is off.

6. License

GNU GENERAL PUBLIC LICENSE Version 3

<http://www.gnu.org/licenses/gpl-3.0.html>

=====

7. Questions and suggestions

Questions and suggestions are welcomed, feel free to contact

Shawn xyang619@gmail.com