`epos2plot`: Plot `epos` Results

Bernhard Haubold

Max-Planck-Institute for Evolutionary Biology, Plön, Germany

June 13, 2018

1 Introduction

`Epos2plot` converts `epos` output to a format ready for plotting.

2 Getting Started

- Download `epos2plot` or update an existing copy

  ```
  go get -u  github.com/evolbioinf/epos2plot
  ```

- Install `epos2plot`

  ```
  go install github.com/evolbioinf/epos2plot
  ```

3 Examples

The example data is based on Figure 2a of Liu and Fu (2015): The population size is 10,000 and 10 sequences of length 10Mbp were simulated with 1000 replications using the coalescent simulator `mspms` (Kelleher et al., 2016). Its output was converted to site frequency spectra using `sfs`, and the site frequency spectra in turn were converted to population sizes using `epos` (both also available from `github.com/evolbioinf`):

```
for i in $(seq 1000); do
mspms 10 1 -t 12310 -r 9750 10000000 -eN 0.066 0.3 |
sfs -f |
epos -u 1.2e-1
done > fig2a_10.epos
```

Copy this data to your working directory

```
cp $GOPATH/src/github.com/evolbioinf/epos2plot/data/fig2a_10.epos.bz2 .
```

and uncompress it

```
bunzip fig2a_10.epos.bz2
```

Display the raw data using `pipePlot` (see `github.com/evolbioinf`):

```
epos2plot -r fig2a_10.epos |
pipePlot
```

Instead of plotting raw `epos` results, `epos2plot` by default computes 2.5% and 97.5% quantiles around the median:

```
epos2plot fig2a_10.epos |
head
#Time LowerQ Median UpperQ SampleSize
0 9550 10200 18700 987
463 9550 10200 18700 987
482 9550 10200 18700 987
487 9550 10200 18700 987
491 9550 10200 18700 987
502 9550 10200 18700 987
505 9550 10200 18700 987
506 9550 10200 18700 987
510 9540 10200 18700 987
```

where `Time` contains the time in generations, `LowerQ` the lower quantile of the population size, `Median` its median, and `UpperQ` its upper quantile. Notice how these values are close to the expected 10,000. The column `SampleSize`, finally, lists the number of data points from which the quantiles were computed. As one last example, plot the median:

```
epos2plot fig2a_10.epos |
cut -f 1,3 |
pipePlot
```

References

J. Kelleher, A. Etheridge, and G. McVean. Efficient coalesncent simulation and genealogical analysis for large sample sizes. *PLoS Comput.*, 12:1–2, 2016.

X. Liu and Y.-X. Fu. Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47: 555–562, 2015.