

Tutorial of ViralBottleneck

Bowen

2023-08-20

This package is used for estimating viral transmission bottleneck sizes using different methods.

1. Install package and download test dataset

First step, install the package “ViralBottleneck” and download dataset in `test_dataset` folder

2. Create transmission object

Second step, the transmission object need to be created before bottleneck size estimation. To create transmission object, the working directory need to meet two requirements: transmission pairs table and sample files used for estimation. This package would extract sample files according to the transmission pairs table the users input.

The example of the transmission pairs table is below (in `test_dataset` folder in package):

Donor	Recipient
681_1_H1N1_donor	681_1_H1N1_recipient
684_2_H1N1_donor	684_2_H1N1_recipient
712_3_H1N1_donor	712_3_H1N1_recipient
742_4_H1N1_donor	742_4_H1N1_recipient
751_5_H1N1_donor	751_5_H1N1_recipient
751_6_H1N1_donor	751_6_H1N1_recipient
751_7_H1N1_donor	751_7_H1N1_recipient
779_8_H1N1_donor	779_8_H1N1_recipient
779_9_H1N1_donor	779_9_H1N1_recipient

Note: Do not put the “-” in name of sample.

After making sure the sample files all exist according to the transmission pairs, start to create transmission object. example code:

```
library(ViralBottleneck)
H1N1_t = read.table("H1N1_transmission_pairs.csv",header = TRUE,sep = " ")
H1N1_ob = CreateTransmissionObject(H1N1_t)
```

2.1 Subset transmission object

The transmission object could be used as list.

```
#Get first 3 transmission object
H1N1_ob_subset = H1N1_ob[1:3]
```

3. Summary transmission object

After creating transmission object, the **Summary_ob** function would provide the information of shared sites (the sites belong to shared sites should be sequenced both in donor and recipient.) for users. Example code:

```
Summary_H1N1 = Summary_ob(H1N1_ob)
```

The result:

donors	recipients	number.of.shared.variants
681_1_H1N1_donor	681_1_H1N1_recipient	49
684_2_H1N1_donor	684_2_H1N1_recipient	26
712_3_H1N1_donor	712_3_H1N1_recipient	25
742_4_H1N1_donor	742_4_H1N1_recipient	34
751_5_H1N1_donor	751_5_H1N1_recipient	31
751_6_H1N1_donor	751_6_H1N1_recipient	31
751_7_H1N1_donor	751_7_H1N1_recipient	31
779_8_H1N1_donor	779_8_H1N1_recipient	31
779_9_H1N1_donor	779_9_H1N1_recipient	31

4. Transmission bottleneck size estimation

Finally, start to calculate transmission bottleneck size using transmission object.

4.1 Output of Bottleneck_size_Calculation function

Take calculation using KL as an example:

```
KL_output = Bottleneck_size_Calculation(H1N1_ob, method = "KL")
```

Output like:

donor	recipient	transmission_bottleneck_size	CI_low	CI_high
681_1_H1N1_donor	681_1_H1N1_recipient	20	12	30
684_2_H1N1_donor	684_2_H1N1_recipient	29	15	49
712_3_H1N1_donor	712_3_H1N1_recipient	107	55	183
742_4_H1N1_donor	742_4_H1N1_recipient	24	11	44
751_5_H1N1_donor	751_5_H1N1_recipient	40	22	67
751_6_H1N1_donor	751_6_H1N1_recipient	32	16	57
751_7_H1N1_donor	751_7_H1N1_recipient	41	22	70
779_8_H1N1_donor	779_8_H1N1_recipient	32	18	51
779_9_H1N1_donor	779_9_H1N1_recipient	37	21	60

4.2 Specify transmission pairs during estimation

This package provide a chance that if user need to specify some transmission pairs for estimation

```
subset_transmission_pairs = read.table("H1N1_transmission_pairs_specify.csv",header = TRUE,sep = " ")
KL_specify = Bottleneck_size_Calculation(H1N1_ob, method = "KL", transmission_pairs = subset_transmission_pairs)
```

4.3 Log file

Bottleneck_size_Calculation could create log file containing number of variant sites used in calculation and number of variant sites filtered before calculation in working directory.

Example code:

```
KL_output_log = Bottleneck_size_Calculation(H1N1_ob, method = "KL", log = TRUE)
```

Output of log argument:

donor	recipient	donor_used	donor_unused	recipient_used	recipient_unused
681_1_H1N1_donor	681_1_H1N1_recipient	45	4	45	4
684_2_H1N1_donor	684_2_H1N1_recipient	25	1	25	1
712_3_H1N1_donor	712_3_H1N1_recipient	22	3	22	3
742_4_H1N1_donor	742_4_H1N1_recipient	17	17	17	17
751_5_H1N1_donor	751_5_H1N1_recipient	26	5	26	5
751_6_H1N1_donor	751_6_H1N1_recipient	20	11	20	11
751_7_H1N1_donor	751_7_H1N1_recipient	24	7	24	7
779_8_H1N1_donor	779_8_H1N1_recipient	31	0	31	0
779_9_H1N1_donor	779_9_H1N1_recipient	29	2	29	2

4.4 Likelihood curve

Bottleneck_size_Calculation could create plot of likelihood curve for each transmission pairs in working directory. However, this argument just used for the methods using maximum likelihoods estimation, including KL method, **Presence-Absence** method, **Binomial** method, **Beta_binomial_Approximate** method and **Beta_binomial_Exact** method. The program would create individual folder for each transmission pair to store the plot. Example code for creating plot:

```
KL_output_log = Bottleneck_size_Calculation(H1N1_ob, method = "Beta_binomial_Approximate", plot = TRUE)
```

The plot of likelihood curve for one transmission pairs (681_1_H1N1_donor-681_1_H1N1_recipient) is below:

