

phyloflows: Calculating sources, onward transmissions and flow ratios

Xiaoyue Xi and Oliver Ratmann

2019-09-10

This vignette describes how a number of key summary statistics of inferred transmission flows can be easily calculated with **phyloflows** `source.attribution.mcmc.getKeyQuantities` function. Please work through the vignette *phyloflows: Estimating transmission flows under heterogeneous sampling – a first example* before you go ahead here.

Getting started

We continue our “First_Example”. The following code chunk contains all code needed, up to running **phyloflows** MCMC routine. The only change is that the number of iterations is now 50,000. The MCMC should take about 5 minutes to run.

```
require(data.table)
require(phyloflows)
data(twoGroupFlows1, package="phyloflows")
dobs <- twoGroupFlows1$dobs
dprior <- twoGroupFlows1$dprior
mcmc.file <- file.path(getwd(), 'twoGroupFlows1_mcmc.RData')
control <- list(seed=42, mcmc.n=5e4, verbose=0)
mc <- phyloflows:::source.attribution.mcmc(dobs, dprior, control)
```

Sources of transmission for each group

OK, so now we have samples from posterior distribution of transmission flows within and between the two population groups,

$$\pi = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}).$$

One important summary statistic are the sources of transmissions into each recipient group, defined by

$$\eta = (\eta_{11}, \eta_{21}, \eta_{12}, \eta_{22})$$

where $\eta_{ij} = \pi_{ij} / \sum_s \pi_{sj}$.

Onward transmissions from each group

Another important summary statistic are the proportions of transmissions that originate from each group, defined by

$$\nu = (\nu_{11}, \nu_{21}, \nu_{12}, \nu_{22})$$

where $\nu_{ij} = \pi_{ij} / \sum_s \pi_{is}$.

Transmission flow ratios

Yet another important summary statistic are ratios of transmission flows, defined by

$$\rho_{ij} = \pi_{ij} / \pi_{ji}.$$

Calculating key quantities

phyloflow has a function to calculate the above summary statistics. The basic syntax is as follows:

```
# specify list of user options
#
# burnin.p: proportion of samples to discard as burn-in
#           (only needed when the burn-in was not already removed)
#
# thin: keep every thin-nth iteration
#       (only needed when thinning was not already performed)
#
# quantiles: quantiles of the marginal posterior distributions
#            that will be computed
#
# flowratios: list of vectors of 3 elements. The 3 elements
#             specify the name of the flow ratio (first element),
#             the enumerator of the flow ratio (second element),
#             and the denominator of the flow ratio (third element)
control <- list( burnin.p=0.05,
                 thin=NA_integer_,
                 quantiles= c('CL'=0.025, 'IL'=0.25, 'M'=0.5, 'IU'=0.75, 'CU'=0.975),
                 flowratios= list( c('1/2', '1 2', '2 1'), c('2/1', '2 1', '1 2'))
                 )
ans <- source.attribution.mcmc.getKeyQuantities(mc=mc,
        dobs=dobs, control=control)

#>
#> Removing burnin in set to 5 % of chain, total iterations= 312
#> Computing flows...
#> Computing WAIFM...
#> Computing sources...
#> Computing flow ratios...
ans
```

Note

Note it is also possible to specify a file name to MCMC output or aggregated MCMC output, and it is also possible to input aggregated MCMC output. Please look up the package help for further instructions.

```
?phyloflows::source.attribution.mcmc.getKeyQuantities
```