Ancestral expression estimation

Hang Ruan

2017/3/28

The idea of ancestral expression estimation is very straight forward, if we define $x = (x_1, ..., x_n)$ are known expression levels of species at the tips of a phylogeny. y is the expression level at ancestral node that we are interested to know. According to Bayes' theorem, the posterior density $P(y|x_1, ..., x_n)$ can be calculated as:

$$P(y|x_1, ..., x_n) = \frac{P(x_1, ..., x_n, y)}{P(x_1, ..., x_n)}$$

In here, we will walk through an example of how to perform ancestral expression estimation on primates' expression data.

TreeExp can be loaded the package in the usual way:

```
library('TreeExp')
```

Then we load primates' expression dataset and primates' time tree.

```
data('primatexp')
data('trees')
```

Expression character tree

Before the estimation of ancestral state of expression, we need an expression tree specified with stationary OU model. The branch length of the expression tree is essential in *aee*. This time we will construct the expression tree by mapping the stationary OU distance onto the primates' time tree:

```
dismat <- expdist(primatexp.objects, taxa = "all",
                  subtaxa = "brain", method = "sou")

primate_tree <- primatetimetree
primate_tree$tip.label <- colnames(dismat)
# make sure their names are the same

exp_tree <- map.ls(primate_tree, dismat)
# map the expression distance onto the primate time tree

exp_tree <- root(exp_tree, outgroup = "Macaque_Brain", resolve.root = T)
exp_tree <- no0br(exp_tree)
# make a little tweak to the expression tree,
# and make sure it is rooted and has no zero branch length
```

Creating variance co-variance matrix

```
var_mat <- varMatInv(objects = primatexp.objects,phy = exp_tree,
                     taxa = "all", subtaxa = "Brain")
```

Ancestral expression estimation

Here, we extract the expression values of known primates, locate MAG gene and extract its expression value vector:

```
exp_table <- exptabTE(primatexp.objects,
                      taxa = "all", subtaxa = "Brain")

MAG_expression <- exp_table[which(rownames(exp_table) == "ENSG00000105695"),]
```

Then we infer the expression values at ancestral nodes of the expression tree:

```
MAG_anc <- aee(MAG_expression, exp_tree, var_mat, select = "all")
```

Finally, we map these estimations on the primate time tree to give a direct presentation of these values:

```
primate_tree$node.label <- sprintf("%.4f",MAG_anc$est)
primate_tree$tip.label <- paste0(exp_tree$tip.label, "  ",
                                 sprintf("%.4f", MAG_expression))

plot(primate_tree, edge.color = "grey80", edge.width = 4,
     show.node.label = T, align.tip.label = T)
```