Expression character tree building

Hang Ruan

2017/3/28

Expression divergence of two genomes can be considered as concerted evolution of transcriptome from their common ancestor, which can be measured by among orthologous genes' variance components. And more specifically, the expression levels of same tissue across different species can be treated as taxonomic units, which can be placed at the tips of character tree that represents expression evolution of these species.

In here, we will give an example to build a character tree from expression data (expression phylogeny).

TreeExp can be loaded the package in the usual way:

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

```
## Loading required package: ape

## Loading required package: phytools

## Loading required package: maps

##
##  # maps v3.1: updated 'world': all lakes moved to separate new #
##  # 'lakes' database. Type '?world' or 'news(package="maps")'.  #
```

We load the datasets created from six tissues' expression data of nine tetrapod species

```
data(tetraexp)
```

Distance matrix:

First, we generate an expression distance matrix of these nine tetrapod species:

```
dismat <- expdist(tetraexp.objects, taxa = "all",
                  subtaxa = "Brain",
                  method = "pea")
as.dist(dismat)
```

```
##                  Human_Brain Chimpanzee_Brain Bonobo_Brain Gorilla_Brain
## Chimpanzee_Brain  0.03928547
## Bonobo_Brain      0.05396989       0.02661905
## Gorilla_Brain     0.04541838       0.04346133   0.05570732
## Orangutan_Brain   0.07441185       0.06650002   0.07418112    0.06650577
## Macaque_Brain     0.07096816       0.07026106   0.07568838    0.07142640
## Mouse_Brain       0.14753659       0.14734369   0.14570042    0.15773811
## Opossum_Brain     0.21460036       0.21023433   0.20623812    0.22218695
## Platypus_Brain    0.25206865       0.24840352   0.24224782    0.26264144
## Chicken_Brain     0.27145998       0.27435377   0.26609183    0.27168518
##                  Orangutan_Brain Macaque_Brain Mouse_Brain Opossum_Brain
## Chimpanzee_Brain
## Bonobo_Brain
## Gorilla_Brain
## Orangutan_Brain
## Macaque_Brain         0.07718557
## Mouse_Brain           0.15908033    0.12541961
```

```
## Opossum_Brain          0.22389360      0.19697327  0.17580708
## Platypus_Brain         0.26309704      0.23960607  0.22327274      0.20305429
## Chicken_Brain          0.26891715      0.24476163  0.23916411      0.28310650
##                      Platypus_Brain
## Chimpanzee_Brain
## Bonobo_Brain
## Gorilla_Brain
## Orangutan_Brain
## Macaque_Brain
## Mouse_Brain
## Opossum_Brain
## Platypus_Brain
## Chicken_Brain          0.28657845
```

You can specify **"taxa"** and **"subtaxa"** options in the `expdist` function as well. The default model **"pea"** is to calculate pair-wise distances by Pearson distance, which equals 1-Pearson's coefficient of expression level.

Also, if you already have a data frame with normalized expression values, there are internal functions available for creating expression distance matrix.

For instance,

```
expression_table <- exptabTE(tetraexp.objects, taxa = "all",
                             subtaxa = "Brain")

dismat <- dist.pea(expression_table)
colnames(dismat) <- colnames(expression_table)
rownames(dismat) <- colnames(dismat)
```

if you have your own expression data frame in the format as the "expression_table" here, it will do fine:

```
dismat <- dist.pea(your_own_dataframe)
colnames(dismat) <- colnames(your_own_dataframe)
rownames(dismat) <- colnames(dismat)
```

Expression character tree:

After the expression distance matrix is created, you can construct character tree by Neighbor-Joining, and bootstrap values based on re-sampling orthologous genes with replacements can also be generated by `boot.phylo` function:

```
tr <- NJ(dismat)
tr <- root(tr, "Chicken_Brain", resolve.root = T)

exptable <- exptabTE(tetraexp.objects, taxa = "all",
                     subtaxa = "Brain")

f <- function(xx) {

    mat <- dist.pea(t(xx))
    # the distance metrics here should be the same as you specified
    # when you created the expression distance matrix

    colnames(mat) <- rownames(xx)
    rownames(mat) <- colnames(mat)
```
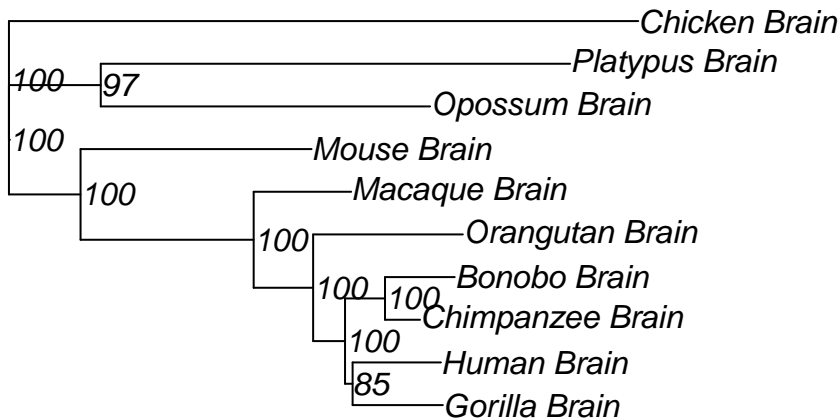
```
    root(NJ(mat), "Chicken_Brain", resolve.root = T)

}

bs <-  boot.phylo(tr, t(exptable), f, B = 100)

# boot.phylo are sampling in columns of a matrix and we want to sample in rows

tr$node.label = bs
plot(tr, show.node.label = TRUE)
```


By now, an expression character tree is successfully constructed. The tree shows expression patterns' similarities in selected genes of designated species. The expression tree is largely in accordance with species tree with minor discrepancy.

Phenomenon of evolutionary history dominates the evolutionary expression pattern can be described as phylogenetic signals. One way to interpret highly consistent expression character tree is that expression levels of transcriptome, representing the regulatory changes, accumulated over time. Though not as concrete as sequence data, expression levels generated from transcriptome data across species show strong phylogenetic signals.