Analysing O'Reilly *et al.* matrices

Martin R. Smith martin.smith@durham.ac.uk

2019-01-17

Contents

The files required to reproduce these analyses are included in the R package directory on installation, and can be downloaded from Github.

If you have RStudio, you can open the R Markdown file used to generate this document (`vignettes/Conduct-analyses.Rmd`) to run the R scripts that will copy all necessary files and begin analyses on your behalf.

You will need to specify some paths for automatic downloading:

```r
# Directory in which to install MrBayes
BAYES_DIR <- "C:/Research/MrBayes"

# Directory in which to conduct parsimony analysis
HOME <- "C:/Research/oreilly" # Must not end in a trailing '/'

# GitHub remote
INST_ROOT <- "https://raw.githubusercontent.com/ms609/OReillyEtAl2016/master/inst/"
```

0.1 Bayesian analysis

0.1.1 bayesgen.pl — Generate MrBayes files

A Perl script to execute analysis using Markov models in MrBayes.

The script reads the datasets of O'Reilly *et al.* [1], appends a MrBayes block to the Nexus files, and executes a MrBayes run, saving the consensus trees and preparing them for analysis in R.

Before running the script:

- Install MrBayes

```r
MRBAYES_RELEASE <- "https://github.com/NBISweden/MrBayes/releases/download/v3.2.6/MrBayes-3.2.6_WIN32_x
zipFile <- tempfile(fileext='.zip')
download.file(MRBAYES_RELEASE, destfile=zipFile, method='auto', mode='wb')
unzip(zipFile, c('MrBayes/mrbayes_x64.exe', 'MrBayes/mrbayes_x86.exe'),
      exdir=BAYES_DIR, junkpaths=TRUE)
file.remove(zipFile)
```

- Download the matrices from O'Reilly *et al.* 2016 [2] (doi:%5B10.5061/dryad.10qf3/5%5D(https://dx.doi.org/10.5061/dryad.10qf3)) and unzip its Nexus files to a local directory (default: `C:/Research/oreilly/Matrices`)

```r
tempFile <- tempfile(fileext='.zip')
download.file("https://datadryad.org/bitstream/handle/10255/dryad.112089/oreilly2016matrices.zip", tempF
unzip(tempFile, exdir=paste0(HOME, '/Matrices'))
file.remove(tempFile)
```

- Copy `mrbayesblock.nex` to the `Matrices` directory, and `bayesgen.pl` to the root MrBayes directory. Modify the latter files to specify the path to MrBayes (default: `C:/Research/MrBayes/`) and path to extracted matrices (default: `C:/Research/oreilly/Matrices`)

```r
download.file(paste0(INST_ROOT, "analysis-bayesian/mrbayesblock.nex"),
              paste0(HOME, '/Matrices/mrbayesblock.nex'))

bayesGenPath <- paste0(BAYES_DIR, '/bayesgen.pl')
download.file(paste0(INST_ROOT, "analysis-bayesian/bayesgen.pl"), bayesGenPath)
bayesGen <- readLines(bayesGenPath)
bayesGen[5] <- paste0('$dir = "', HOME, '/Matrices";')
bayesGen[6] <- paste0('$bayes_dir = "', BAYES_DIR, '";')
writeLines(bayesGen, bayesGenPath)

t2nexPath <- paste0(BAYES_DIR, '/t2nex.pl')
download.file(paste0(INST_ROOT, "analysis-bayesian/t2nex.pl"), t2nexPath)
t2nex <- readLines(t2nexPath)
t2nex[2] <- paste0('$dir = "', HOME, '/Matrices";')
writeLines(t2nex, t2nexPath)
```

- Perform the analyses by executing `bayesgen.pl`. (Once Perl is installed, you can just double-click the file.)
- Once the analyses are complete, copy all files ending `.run#.nex` to `C:/Research/oreilly/MrBayes`.

0.2   Parsimony analysis

0.2.1   mptgen.pl — Generate parsimony trees in TNT

A Perl script to generate most parsimonious trees by parsimony search in TNT.

The script generates TNT scripts to perform parsimony analysis on each of the O_Reilly *et al.* datasets, under equal and implied weights, with and without suboptimal trees. It then executes these scripts and converts the output into a format suitable for analysis in R.

Before running the script, you'll need an installation of Perl. Strawberry Perl works on MS Windows.

Then:

- Create a local directory (default: `C:/Research/oreilly`) with two subdirectories entitled `Matrices`, and `Trees`. Then, within the new `Trees` directory, create the further subdirectories `eq.100`, `eq.350`, `eq.1000`, `k2.100` (and `.350`, `.1000`), and so forth for `k3`, `k5`, `k10`, `k20` and `k200`.

```r
N_CHAR <- c(100, 350, 1000)
sapply(paste0(HOME, '/', c('', 'Matrices', 'Trees')), dir.create)
sapply(N_CHAR, function (nChar)
  sapply(paste0(HOME, '/Trees/',
               c('eq', 'k2', 'k3', 'k5', 'k10', 'k20', 'k200'),
               '.', nChar), dir.create, showWarnings=FALSE))
```

- Install TNT.

```
zipFile <- paste0(HOME, '/TNT.ZIP')
# This is the Windows path; use the appropriate path for your operating system
download.file("http://www.lillo.org.ar/phylogeny/tnt/ZIPCHTNT.ZIP",
              destfile=zipFile, method='auto', mode='wb')
unzip(zipFile, 'tnt.exe', exdir=HOME)
file.remove(zipFile)
```

- Copy `mptgen.pl` and (optinally) `tnt2nex.pl` into this root directory, updating each file so its variable $dir corresponds to the appropriate path.
  `tnt2nex.pl` translates TNT output into NEXUS format and may be useful if you wish to perform further analysis of TNT output. This will be performed automatically if you uncomment the final line of `mptgen.pl`.

```
tnt2nexPath <-  paste0(HOME, '/tnt2nex.pl')
mptgenPath <-  paste0(HOME, '/mptgen.pl')

download.file(paste0(INST_ROOT, "analysis-parsimony/tnt2nex.pl"), tnt2nexPath)
tnt2nex <- readLines(tnt2nexPath)
tnt2nex[3] <- paste0('$dir = "', HOME, '";')
writeLines(tnt2nex, tnt2nexPath)

download.file(paste0(INST_ROOT, "analysis-parsimony/mptgen.pl"), mptgenPath)
mptgen <- readLines(mptgenPath)
mptgen[3] <- paste0('$dir = "', HOME, '";')
writeLines(mptgen, mptgenPath)
```

- Copy the file `template.run` into the tntscript subdirectory.

```
download.file(paste0(INST_ROOT, "analysis-parsimony/tnt_template.run"),
              paste0(HOME, '/tnt_template.run'))
```

- If you haven't already, download the matrices from O'Reilly *et al.* [2] (doi:%5B10.5061/dryad.10qf3/5%5D(https://dx.doi.org/10.5061/dryad.10qf3)) and unzip its three folders to the `Matrices` directory.

```
tempFile <- tempfile(fileext='.zip')
download.file("https://datadryad.org/bitstream/handle/10255/dryad.112089/oreilly2016matrices.zip", templ
unzip(tempFile, exdir=paste0(HOME, '/Matrices'))
## Optionally:
file.remove(tempFile)
```

- Perform the analyses by executing `mptgen.pl`. (Once Perl is installed, you can just double-click the file.)

0.3   Data analysis

Once the data have ben gneerated, they can be analysed using [https://github.com/ms609/OReillyEtAl2016/blob/master/data-raw/GenerateData.Rmd].

References

1. O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PCJ. 2016 Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters* **12**, 20160081. (doi:10.1098/rsbl.2016.0081)

2. O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PCJ. 2016 Data from: Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Dryad Digital Repository*, doi:10.5061/dryad.10qf3. (doi:10.5061/dryad.10qf3)