Analysing Congreve & Lamsdell matrices

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

2019-01-10

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:

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

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

# GitHub remote
INST_ROOT <- "https://raw.githubusercontent.com/ms609/CongreveLamsdell2016/master/inst/"
```

1 Bayesian analysis

`bayesgen.pl` is a Perl script to execute analysis using Markov models in MrBayes.

The script reads the datasets of Congreve and Lamsdell [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

```
MRBAYES_RELEASE <- "https://github.com/NBISweden/MrBayes/releases/download/v3.2.6/MrBayes-3.2.6_WIN32_x
zipFile <- paste0(BAYES_DIR, '/MrBayes.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 Appendix S5 from Congreve and Lamsdell [2] (doi:%5B10.5061/dryad.7dq0j/5%5D(https://dx.doi.org/10.5061/dryad.7dq0j/5)) and unzip its 100 nexus files to a local directory (default: `C:/Research/MrBayes/iw`)

```
tempFile <- tempfile(fileext='.zip')
download.file("https://datadryad.org/bitstream/handle/10255/dryad.108351/S5%20-%20Character%20Weights%20
unzip(tempFile, exdir=paste0(BAYES_DIR, '/iw'), junkpaths=TRUE,
      files = paste0('Weights tests/', formatC(1:100, width=3, flag=0), '.txt.nex'))
file.remove(tempFile)
```

- Copy `mrbayesblock.nex` to the `iw` directory, and `bayesgen.pl` and `t2nex.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/MrBayes/iw`)

```
download.file(paste0(INST_ROOT, "analysis-bayesian/mrbayesblock.nex"),
              paste0(BAYES_DIR, '/iw/mrbayesblock.nex'))

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

2 Parsimony analysis

`mptgen.pl` is 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 Congreve and Lamsdell 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/iw`) with subdirectories entitled `Matrices`, and `Trees`. Then, within the new `Trees` directory, create the further subdirectories `eq`, `k1`, `k2`, `k3`, `k5` and `kX`.

```
sapply(paste0(HOME, '/', c('', 'Matrices', 'Trees')), dir.create)
sapply(paste0(HOME, '/Trees/', c('eq', 'k1', 'k2', 'k3', 'k5', 'kX')), dir.create)
```

- 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 (optionally) `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, '/Trees";')
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 `tnt_template.run` into the root directory.

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

- Download Appendix S1 from Congreve and Lamsdell [2] (doi:10.5061/dryad.7dq0j/1) and unzip its 100 text files to `Matrices`.

```
tempFile <- tempfile(fileext='.zip')
download.file("https://datadryad.org/bitstream/handle/10255/dryad.101095/S1%20-%20TNT%20files.zip", temp
unzip(tempFile, exdir=paste0(HOME, '/Matrices'))
```

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

3 Analysing output data

Once these analyses have generated the necessary data, these can be analysed using the scripts in [https://github.com/ms609/CongreveLamsdell2016/blob/master/data-raw/GenerateData.Rmd]. The results of these analyses are available in the R data objects; to view them, install the package in R and view the help files.

References

1. Congreve CR, Lamsdell JC. 2016 Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices. *Palaeontology* **59**, 447–465. (doi:10.1111/pala.12236)

2. Congreve CR, Lamsdell JC. 2016 Data from: Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices. *Dryad Digital Repository*, doi:10.5061/dryad.7dq0j. (doi:10.5061/dryad.7dq0j)