PyRate exercise

Gaurav Kandlikar

October 29, 2015

Step 1

Prep `Ursidae.txt` to be ready for `PyRate` by using `extract.ages()` in `R` utilities:

```
source("pyrate_utilities.r")
extract.ages("example_files/Ursidae.txt")
```

This creates `Ursidae_PyRate.py` in the `example_files/` directory.

Step 2

Run `PyRate` on `Ursidae_PyRate.py`:

From `PyRate` folder in terminal, run the following line-

`python PyRate.py example_files/Ursidae_PyRate.py`

Sample output:

```
                    PyRate 0.600
                    20151024

        Bayesian estimation of speciation and extinction
              rates from fossil occurrence data

            Daniele Silvestro, Jan Schnitzler et al.
                    pyrate.help@gmail.com



        Using Gamma priors on the birth-death rates.

                    started at: Thu Oct 29 16:40:02 2015
                    initializing chain...

0*  post: -2379.83 lik: -2371.41 prior: -8.42 tot length 209.96

1000    post: -1242.94 lik: -1234.47 (-862.71, -371.77) prior: -8.46 tot.l: 318.82 k: 2
    t.frames: [ 23.748   0.   ] (sp.)
    t.frames: [ 23.748   0.   ] (ex.)
    sp.rates: [ 0.244]
    ex.rates: [ 0.382]
    q.rate: 0.697    Gamma.prm: 1.0
    ts: [ 5.494  9.824  6.615  4.088  7.749] ...
    te: [ 3.017  6.266  0.491  3.037  5.598] ...
```

Step 3

After PyRate runs, inspect generated log files.

Ouptut in `Ursidae_1_sum.txt`:

```
PyRate 0.600         build          20151024
input: ['example_files/Ursidae_PyRate.py'] output: example_files/pyrate_mcmc_logs/Ursidae_1
PyRate was called as follows:
Namespace(A=2, B=1, BF=[], M=10, N=None, S=1000, T=1.0, a=0.3, b=0, cite=False, d='', dpp_eK=2.0,
dpp_f=500, dpp_grid=1.5, dpp_hp=2.0, fR=1.0, fS=0.7, fU=[0.02, 0.18, 0], fixSE='', fixShift='',
input_data=['example_files/Ursidae_PyRate.py'], j=1, k=10, logT=2, mC=False, mCov=0, mG=False,
mHPP=False, mL=1, mM=1, mProb='', multiR=1, n=10000000, nT=5, ncat=4, out='', p=1000, pC=1,
pL=[1.1, 1.1], pM=[1.1, 1.1], pP=[1.5, 1.1], pS=2.5, plot='', r=1, s=1000, sw=100, t=0.03,
tC=[0.2, 0.2, 0.15], tQ=[1.2, 1.2], tR=1.2, tS=1.0, tT=1.0, tag='', thread=[0, 0], trait=1,
wd='')
```

`Ursidae_1_mcmc.log` is a table of MCMC posteriors, priors, likelihoods, and `TE` and `TS` for all taxa.

`Ursidae_1_marginal_rates.log`: This file can be opened with `Tracer` (Figure 1). *Note: I am not really sure what to do with this Tracer output- how can we use this information to determine the appropriate number of burnins?*
We can use this information to determine a reasonable burn-in (I will just assume 1000 for now), and then use the burnin estimate to create a plot of the logs.

Here, the list of `l` values are lambda values over time- we can look at these estimates in tracer but we can also use PyRate to generate pretty `R` plots. So we'll do that!

In mcmc file: In this output, `qrate` shows estimates of the preservation rate. In this case it is almost 4- so a single species is expected to leave 4 fossils every million years.
`alpha` = 1, but is different if we use a gamma model
`k birth` and `k death` are discrete values- number of times there were switches in rate shifts. Here `k birth` is a little over 2 which means that there is likely 1 shift. But there is some p(more than 1 shift). `k death` is analogous for extinction rates- likely no shifts.
`root age` is estimated time of origin for the whole data set; `death age` is zero because the lineage here is still extant.

This is followed by a long list of species name, followed by `TS` and `TE` which are estimates of speciation and extinction times.

Step 4

Plotting from the `Ursidae_1_marginal_rates.log`.

We can plot using this command:

`python PyRate.py -plot ~/full/path/to/pyrate_mcmc_logs/Ursidae_1_marginal_rates.log -b 1000`

Note to self- use full path; otherwise program threw an error.

This step creates two documents in the **pyrate_mcmc_logs** folder: the first is an R script (`Ursidae_1_marginal_rates_RTT.r`); the second is a pdf file made using the `R` code (`Ursidae_1_marginal_rates_RTT.pdf`; Figure 2).

From the plot of Speciation Rate over time, we see that the Ursids had experienced a high speciation rate c. 20 MYA; that speciation rate declined rapidly over 3-5 million years to a basal speciation rate. Over this time period, the extinction rate in the clade has remained at a relatively constant basal rate which is approximately equal to the basal speciation rate. There has been a slight upward trend in extinction rate since ~10 MYA. These two sum up to mean that Ursids experienced a rapid diversification from ~20 to ~17 MYA, after which diversity remained constant for ~10 million years. Since then, there has been a slight decline in diversity, as the upward trend of extinction rates has not been counteracted by any change in speciation rates.

We have a visual interpretation of the graphs above, but we can probabilistically assess the likelihood of models with different numbers of sppn/extn rates (alt, how many rate switches there have been):

`python PyRate.py -mProb ~/full/path/to/Ursidae_1_mcmc.log -b 1000`

The output is as follows:

```
               First 1000 samples excluded as burnin.

       Model              Probability
                      Speciation   Extinction
       1-rate         0.0122       0.6791
       2-rate         0.645        0.2458
       3-rate         0.2686       0.0619
       4-rate         0.0631       0.0117
       5-rate         0.0103       0.0014
       6-rate         0.0008       0.0001
```

This suggests that a 2-rate model (i.e. one rate switch) is the most probable for speciation, and a 1-rate (i.e. no switch) model for extinction.

Step 5

We can correlate diversification rates to a trait associated with the linages. Here are the parameters to use:

`-mCov 3` estimates correlations between trait size and both the extinction and speciation rate- most inclusive.
`-A 0` constant birth death model - no shifts in speciation or extinction rates. Variation only linked to trait variation. On the other hand, average rates might shift thru time and there might be more variation attributed to the rates.
Here it makes sense to assume some shifts- our initial analyses without traits does support a rate shift in speciation. But for the sake of this exercise I am running the model assuming constant rates:

`python PyRate.py ~/full/path/to/Ursidae_PyRate.py -mCov 3 -A 0`

In the output, `cov_sp` and `cov_ex` within the file `Ursidae_1VR_G_COV_marginal_rates.log` are the correlations between speciation and extinction rates and the input trait (here, body size). I ran this analysis for the full 10000000 generations. Neither `cov_sp` nor `cov_ex` were found to be significantly different from zero (Figures 3, 4), although `cov_ex` may be tending towards a negative number (larger lineages have lower extinction rates).

Daniele suggested in class that if we ran this for a more realistic number of steps we would see that diversification rate is positively correlated with body size in Ursideae- which is contrary to the normal prediction that smaller organisms evolve faster than bigger ones. One reason for this difference might be that carnivores benefit from being big, as it opens up a larger source of food.
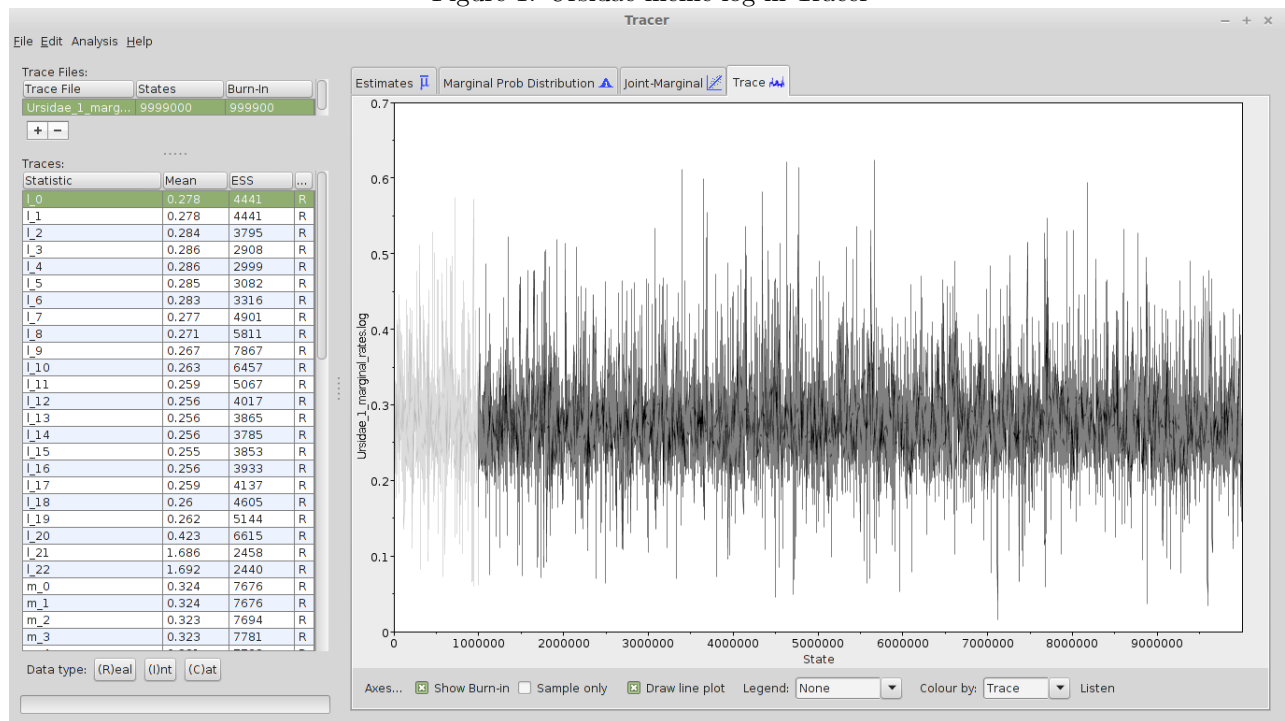
Figure 1: Ursidae mcmc log in Tracer

Figure 2: Plot of Sppn, Extn, and Diversification Rates from marginal rates file
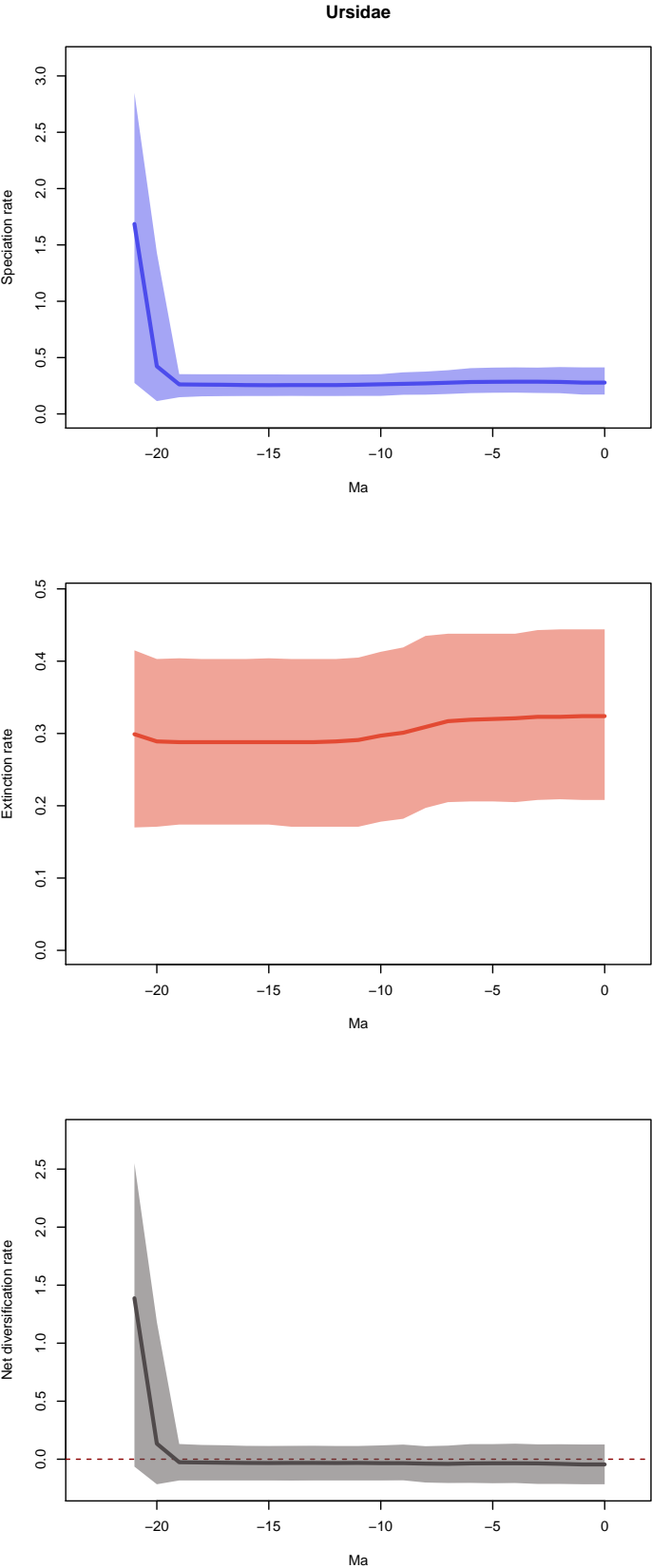
Ursidae

Figure 3: CovEx, which describes any correlation between body size and extinction rate
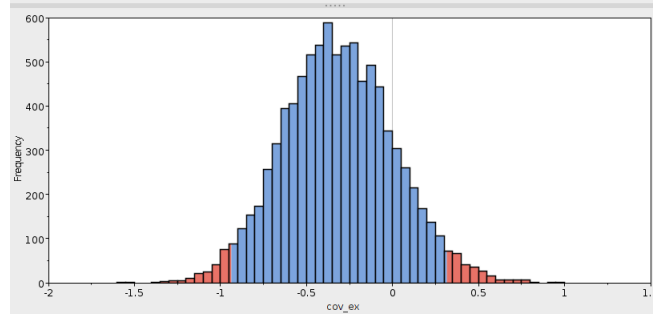

Figure 4: CovSp, which describes any correlation between body size and speciation rate