

1 Bayesian implementation

Bayesian implementation is included within the *BayesCode* software, available at <https://github.com/ThibaultLatrille/bayescode>.

1.1 Data formatting

Running the analysis on your dataset and compute posterior probabilities requires three files:

1. A phylogenetic tree in newick format, with branch lengths in number of substitutions per site (neutral markers), from which the values of nucleotide divergence (d) are used.
2. A file containing the mean trait values for each species.
3. A file containing the variation within-species for each trait and the genetic variation within-species (neutral markers).

1.1.1 Phylogenetic tree

The phylogenetic tree must be in newick format, with branch lengths in substitutions per site (neutral markers).

1.1.2 Mean trait for each species

The file containing mean trait values for each species must be in a tab-delimited file with the following format:

TaxonName	Body_mass	Brain_mass
Panthera_tigris	12.26	5.676
Pithecia_pithecia	7.256	3.436
Colobus_angolensis	9.176	4.284
Saimiri_boliviensis	6.845	3.279
⋮	⋮	⋮

The columns are:

- *TaxonName*: the name of the taxon matching the name in the alignment and the tree.
- As many columns as traits, without spaces or special characters in the trait.
- The values can be NaN to indicate that the trait is not available for that taxon.

1.1.3 Trait variation for each species

The file containing trait variation for each species must be in a tab-delimited file with the following format:

TaxonName	Nucleotide_diversity	Body_mass_variance	Body_mass_heritability	Brain_mass_variance	Brain_mass_heritability
Pithecia_pithecia	0.0016	0.22871	0.2	0.00737	0.2
Colobus_angolensis	0.0017	0.00393	0.2	0.00416	0.2
Saimiri_boliviensis	0.0013	0.00022	0.2	0.00045	0.2
Pygathrix_nemausus	0.0016	0.00347	0.2	0.00097	0.2
⋮	⋮	⋮	⋮	⋮	⋮

- *TaxonName*: the name of the taxon matching the name in the alignment and the tree.
- *Nucleotide_diversity*: the nucleotide diversity within-species (neutral markers), cannot be **NaN**.
- As many columns as traits, without spaces or special characters in the trait.
- *TraitName_variance*: the phenotypic variance of the trait within-species, can be **NaN** to indicate that the trait variance is not available for that taxon.
- *TraitName_heritability* (optional): the heritability of the trait within-species, between 0 and 1, cannot be **NaN**.
- The columns with the suffix **_variance** and **_heritability** are repeated for each trait.
- *TraitName_heritability_lower* (optional): the lower bound of the heritability of the trait within-species, between 0 and 1, cannot be **NaN**.
- *TraitName_heritability_upper* (optional): the upper bound of the heritability of the trait within-species, between 0 and 1, cannot be **NaN**.
- If the columns with the suffix **_heritability_lower** and **_heritability_upper** are present, the heritability is randomly drawn from a uniform distribution between the lower and upper bounds.
- If the columns with the suffix **_heritability** is present, it is taken as is.
- If the additive genetic variance (instead of phenotypic variance) is available for a trait, the heritability can be omitted and will automatically be set to 1.0.

1.2 Bayesian estimation

The executable `nodetraits` from *BayesCode* is used to run the Bayesian estimation of the model, and the executable `readnodetraits` is used to read the results.

Assuming that the file `data/body_size/mammals.male.tsv` contains the mean trait values for each species, the file `data/body_size/mammals.male.var_trait.tsv` contains the variation within-species for each trait and the genetic variation within-species (neutral markers), and the file `data/body_size/mammals.male.tree` contains the phylogenetic tree, the following commands are used to run the model and read the results.

1.2.1 Running the model

`nodetraits` is run with the following command:

```
nodetraits --until 2000
           --tree data/body_size/mammals.male.tree
           --traitsfile data/body_size/mammals.male.tsv
           run_mammals_male
```

1.2.2 Reading the results

Once the model has run, the chain `run.mammals.male` is used to compute the posterior distribution of the ratio of between-species variation over within-species variation with `readnodetraits`:

```
readnodetraits --burnin 1000
              --var_within data/body_size/mammals.male.var_trait.tsv
              --output results_mammals_male.tsv
              run_mammals_male
```

The file `data_empirical/chain_name.ratio.tsv` then contains the posterior mean of the ratio of between-species variation over within-species variation, the 95% and 99% credible interval, and the posterior probability that the ratio is greater than 1.