

# Package ‘rcolgem’

August 27, 2014

**Type** Package

**Title** statistical inference and modeling of genealogies generated by epidemic and ecological processes

**Version** 0.0.3

**Date** 2014-07-23

**Author** {Erik M Volz, Oliver Ratmann, Ethan Romero Severson}

**Maintainer** Erik M Volz <erik.volz+rcolgem@gmail.com>

**Depends** R (>= 2.15.1), ape, deSolve

**Description** rcolgem is a package for phylodynamic inference using population genetic models. rcolgem implements coalescent models for populations with nonlinear dynamics and potentially many demes, and other population genetic models may be supported in the future. This package is well suited for studying infectious disease epidemics and inference of epidemiological parameters from pathogen phylogenies. The package can also be used for phylogeographic analysis and estimation of demographic histories (population size through time). rcolgem is not a package for conducting phylogenetic inference, although such packages are available in R and such tools may be incorporated in the future. A time-scaled genealogy with known times of sampling is a necessary input for most functions in rcolgem. Currently, rcolgem works best for fitting deterministic demographic models (e.g. systems of ordinary differential equations). Future versions may incorporate particle filters for fitting stochastic models. rcolgem also provides methods for simulating trees conditional on a demographic process.

**License** GPL-3

## R topics documented:

binaryDatedTree . . . . .	2
calculate.cluster.size.moments.from.model . . . . .	2
calculate.cluster.size.moments.from.tree . . . . .	3
coalescent.log.likelihood . . . . .	4
simulate.binary.dated.tree . . . . .	5

---

`binaryDatedTree`      *Create binary dated tree. The `binaryDatedTree` class includes heights for each node and other helper variables*

---

### Description

Create binary dated tree `binaryDatedTree` class, includes heights for each node and other helper variables like `maxSampleTime`.

### Usage

```
binaryDatedTree(phylo, sampleTimes, sampleStates=NULL,
, sampleStatesAnnotations=NULL)
```

### Arguments

`phylo`                    `ape::phylo tree`

`sampleTimes`          Named vector of sample times for each taxon. Names must correspond to `phylo$tip.label`

`sampleStates`          Matrix of discrete character information for each taxon. If `NULL`, this may be inferred from taxon label. Rownames must correspond to `phylo$tip.label`. Column names must correspond to names of demes used in subsequent analysis. Each row is a probability vector summing to 1 (the probability that taxon occupies each deme).

`sampleStatesAnnotations`          Vector of possible discrete character states for taxa. If inferring taxon state from label, this provides the possible matches for taxon annotations. The annotations are extracted from all text following `'_'` at the end of the taxon label.

### Value

A `binaryDatedTree` object. Includes numeric vector heights for each tip and internal node, the time before most recent sample when each node occurs.

---

`calculate.cluster.size.moments.from.model`  
*Calculate cluster size moments from model*

---

### Description

Calculate cluster moments of the distribution of cluster sizes (clade sizes), which change as a function of the threshold TMRCA used for defining clades. Currently, only sampling at a single time point is supported.

**Usage**

```
calculate.cluster.size.moments.from.model(sampleTime,
  sampleStates, maxTime = NA, minTime = NA,
  timeResolution = 50, discretizeRates = FALSE,
  fgyResolution = 100, integrationMethod = "adams")
```

**Arguments**

sampleTime    Time that sample is collected

sampleStates (n X M) matrix of sample states for each taxon

maxTime

minTime

timeResolution  
                    Determines time step for discrete approximations. Larger resolution gives smaller time step.

discretizeRates  
                    if TRUE, a discrete approximation is used to calculate transition rates

fgyResolution  
                    the number of time steps to use if discretizeRates=TRUE

integrationMethod  
                    method passed to deSolve::ode

---

```
calculate.cluster.size.moments.from.tree
```

*Calculate cluster size moments from tree*

---

**Description**

Calculate cluster moments of the distribution of cluster sizes (clade sizes), which change as a function of the threshold TMRCA used for defining clades. Currently, only sampling at a single time point is supported.

**Usage**

```
calculate.cluster.size.moments.from.tree(bdt, heights)
```

**Arguments**

bdt                    a binary dated tree object

heights                vector numeric, heights at which to calculate cluster sizes

---

```
coalescent.log.likelihood
```

*Compute the likelihood of a gene genealogy conditional on a demographic history.*

---

### Description

The likelihood of a dated tree is calculated conditional on a demographic history, which consists of the timeseries of births, migrations, and population sizes in the population.

### Usage

```
coalescent.log.likelihood( bdt, births, deaths, nonDemeDynamics
  , t0, x0, migrations=NA, parms=NA, fgyResolution = 2000
  , integrationMethod = 'rk4', censorAtHeight=FALSE
  , forgiveAgtY=.2, returnTree=FALSE)
```

### Arguments

<code>bdt</code>	A <code>binaryDatedTree</code> object which extends <code>ape:phylo</code> and includes dates of tips and internal nodes.
<code>births</code>	A vector or matrix of strings. These are evaluated as equations for the number of births within and between demes. Must have rownames and colnames corresponding to the names of demes.
<code>deaths</code>	A vector of strings. These are evaluated as equations for the rate that lineages in each deme are terminated. Must have rownames corresponding to the names of demes.
<code>nonDemeDynamics</code>	A vector of strings. These are evaluated as equations for the rate of change of state variables that do not correspond to demes. Vector must have names of state variables.
<code>t0</code>	The time of origin of the process. Should be before root of genealogy.
<code>x0</code>	A vector of initial conditions for the demographic process (the state of the system at time <code>t0</code> ). Should include the name and value of all variables mentioned in <code>births</code> and <code>nonDemeDynamics</code> .
<code>migrations</code>	A vector or matrix of strings. These are evaluated as equations for the number of migrations between demes. Must have rownames and colnames corresponding to the names of demes. Should be omitted if there is only one deme.
<code>parms</code>	A list of parameters that will be accessible to differential equations specified in <code>births</code> , <code>migrations</code> and <code>nonDemeDynamics</code> .
<code>fgyResolution</code>	Determines timestep of ODEs (larger values gives smaller time step)
<code>integrationMethod</code>	Passed to <code>lsoda</code> . Higher precision may be obtained with methods such as <i>adams</i> at some computational cost.

sensorAtHeight	Boolean or numeric. If numeric, will ignore all parts of the tree older than this value when calculating the likelihood.
forgiveAgtY	If this is zero, function will return -Inf if at any point the number of lineages exceeds population size. If this value is between zero and one, function will return -Inf if the number of lineages exceeds population size AND number of lineages is less than this value times sample size. This can be useful when fitting models where the number of lineages is close to population size, which often happens during early period of exponential growth.
returnTree	If TRUE, the return value is a list with two elements. The first element is the log likelihood, and the second element is a modified version of the binaryDatedTree that includes the estimated ancestral states along each lineage. Ancestral state information is contained in the matrices lstates and ustates, which has the same number of rows as the tree edge attribute. These respectively describe the state at the beginning and end of each edge (present to past).

### Details

See vignettes for detailed usage.

### Value

The log likelihood (numeric).

### Author(s)

Erik M Volz

### References

E. M. Volz, Complex population dynamics and the coalescent under neutrality, *Genetics*, January, 2012

### See Also

binaryDatedTree

---

simulate.binary.dated.tree

*Simulate a binary dated tree using a coalescent model and given a demographic history*

---

### Usage

```
simulate.binary.dated.tree(births, deaths, nonDemeDynamics, t0, x0, sampleTim
```

**Arguments**

births	A vector or matrix of strings. These are evaluated as equations for the number of births within and between demes. Must have rownames and colnames corresponding to the names of demes.
deaths	A vector of strings. These are evaluated as equations for the rate that lineages in each deme are terminated. Must have rownames corresponding to the names of demes.
nonDemeDynamics	A vector of strings. These are evaluated as equations for the rate of change of state variables that do not correspond to demes. Vector must have names of state variables.
t0	The time of origin of the process. Should be before root of genealogy.
x0	A vector of initial conditions for the demographic process (the state of the system at time t0). Should include the name and value of all variables mentioned in births and nonDemeDynamics.
sampleTimes	A vector of times for each sample.
sampleStates	A vector of states for each sample. Each element gives the index 1..m of the deme from which each sample was taken.
migrations	A vector or matrix of strings. These are evaluated as equations for the number of migrations between demes. Must have rownames and colnames corresponding to the names of demes. Should be omitted if there is only one deme.
parms	A list of parameters that will be accessible to differential equations specified in births, migrations and nonDemeDynamics.
fgyResolution	Determines timestep of ODEs (larger values gives smaller time step)
integrationMethod	Passed to lsoda. Higher precision may be obtained with methods such as <i>adams</i> at some computational cost.

**Value**

A binaryDatedTree, extends ape::phylo and includes heights of all nodes, time before most recent sample when node occurs.

**Author(s)**

Erik M Volz

**References**

E. M. Volz, Complex population dynamics and the coalescent under neutrality, *Genetics*, January, 2012