

# Package ‘breakaway’

August 1, 2019

**Title** Species Richness Estimation and Modeling

**Version** 4.6.10

**Date** 2018-08-01

**Description** Understanding the drivers of microbial diversity is an important frontier of microbial ecology, and investigating the diversity of samples from microbial ecosystems is a common step in any microbiome analysis. ‘breakaway’ is the premier package for statistical analysis of microbial diversity. ‘breakaway’ implements the latest and greatest estimates of species richness, as well as the most commonly used estimates.

**License** GPL-2

**BugReports** <https://github.com/adw96/breakaway/issues>

**LazyData** true

**RoxygenNote** 6.1.1

**Depends** R (>= 3.5.0)

**Imports** MASS, stats, utils, phyloseq, graphics, magrittr, ggplot2, reshape2, R.spc, testthat, tibble, dplyr

**Suggests** covr, DT, knitr, RCurl, rmarkdown, plyr

**VignetteBuilder** knitr

**URL** <https://adw96.github.io/breakaway/>

**NeedsCompilation** no

**Author** Amy Willis [aut, cre],  
Bryan D Martin [aut],  
Pauline Trinh [aut],  
Kathryn Barger [aut],  
John Bunge [aut]

**Maintainer** Amy Willis <adwillis@uw.edu>

## R topics documented:

alpha_estimate . . . . .	3
alpha_estimates . . . . .	4
apples . . . . .	4
betta . . . . .	5
betta_pic . . . . .	7
betta_random . . . . .	8

breakaway . . . . .	9
breakaway_nofl . . . . .	10
build_frequency_count_tables . . . . .	12
chao1 . . . . .	12
chao1_bc . . . . .	13
chao_bunge . . . . .	14
chao_shen . . . . .	14
check_format . . . . .	15
convert . . . . .	15
cutoff_wrap . . . . .	16
good_turing . . . . .	16
hawaii . . . . .	17
kemp . . . . .	17
make_design_matrix . . . . .	19
make_frequency_count_table . . . . .	19
objective_bayes_geometric . . . . .	19
objective_bayes_mixedgeo . . . . .	20
objective_bayes_negbin . . . . .	21
objective_bayes_poisson . . . . .	22
physeq_wrap . . . . .	22
plot.alpha_estimates . . . . .	23
poisson_model . . . . .	24
proportions_instead . . . . .	24
rnbinomtable . . . . .	24
rztnbinomtable . . . . .	25
sample_inverse_simpson . . . . .	25
sample_richness . . . . .	26
sample_shannon . . . . .	26
sample_shannon_e . . . . .	27
sample_simpson . . . . .	27
sample_size_estimate . . . . .	28
sample_size_figure . . . . .	28
toy_metadata . . . . .	29
toy_otu_table . . . . .	29
toy_taxonomy . . . . .	30
true_gini . . . . .	30
true_hill . . . . .	31
true_inverse_simpson . . . . .	31
true_shannon . . . . .	32
true_shannon_e . . . . .	32
true_simpson . . . . .	33
wlrm_transformed . . . . .	33
wlrm_untransformed . . . . .	34

---

alpha_estimate	<i>alpha_estimate</i>
----------------	-----------------------

---

## Description

Build objects of class `alpha_estimate` from their components. `alpha_estimate()` is a constructor method

## Usage

```
alpha_estimate(estimate = NULL, error = NULL, estimand = NULL,
  name = NULL, interval = NULL, interval_type = NULL, type = NULL,
  model = NULL, warnings = NULL, frequentist = NULL,
  parametric = NULL, plot = NULL, reasonable = NULL, other = NULL,
  ...)
```

## Arguments

<code>estimate</code>	The estimate
<code>error</code>	The standard error in the estimate
<code>estimand</code>	What is the estimate trying to estimate? (richness, Shannon...)
<code>name</code>	The name of the method
<code>interval</code>	An interval estimate
<code>interval_type</code>	Type of interval estimate
<code>type</code>	TODO(Amy): Deprecate?
<code>model</code>	What model is fit
<code>warnings</code>	Any warnings?
<code>frequentist</code>	Logical. Frequentist or Bayesian?
<code>parametric</code>	Logical. Parametric or not?
<code>plot</code>	A ggplot associated with the estimate
<code>reasonable</code>	Is the estimate likely to be reasonable?
<code>other</code>	Any other relevant objects
<code>...</code>	Any other objects

## Value

An object of class `alpha_estimate`

---

alpha_estimates	<i>alpha_estimates</i>
-----------------	------------------------

---

### Description

Build objects of class `alpha_estimates` from their components. `alpha_estimates()` is a constructor method

### Usage

```
alpha_estimates(...)
```

### Arguments

... Objects of class `alpha_estimate`, or a list of objects of class `alpha_estimate`

### Value

An object of class `alpha_estimates`

---

apples	<i>(Data) Frequency count table of soil microbes in an apples orchard.</i>
--------	--

---

### Description

(Data) Frequency count table of soil microbes in an apples orchard.

### Usage

```
apples
```

### Format

A data frame with 88 rows and 2 variables:

**index** an index variable

**frequency** number of taxa that were observed with this frequency ...

### Source

<http://onlinelibrary.wiley.com/doi/10.1111/biom.12332/abstract>

### References

Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*, **71**(4), 1042–1049. <http://onlinelibrary.wiley.com/doi/10.1111/biom.12332/abstract>

Walsh, F. et al. (2014). (2014). Restricted streptomycin use in apple orchards did not adversely alter the soil bacteria communities. *Frontiers in Microbiology* **4**, 383.

---

beta	<i>Modelling total diversity with beta</i>
------	--

---

## Description

This function tests for heterogeneity of total diversity (observed plus unobserved) across multiple sites. It can account or test for fixed effects that may explain diversity. It returns the significance of the covariates in explaining diversity and a hypothesis test for heterogeneity.

## Usage

```
beta(chats, ses, X = NA)
```

## Arguments

chats	A vector of estimates of total diversity at different sampling locations. ‘breakaway’ estimates are suggested in the high-diversity case but not enforced.
ses	The standard errors in ‘chats’, the diversity estimates.
X	A numeric matrix of covariates. If not supplied, an intercept-only model will be fit.

## Value

table	A coefficient table for the model parameters. The columns give the parameter estimates, standard errors, and p-values, respectively. This model is only as effective as your diversity estimation procedure; for this reason please confirm that your estimates are appropriate and that your model is not misspecified. <code>beta_pic</code> may be useful for this purpose.
cov	Estimated covariance matrix of the parameter estimates.
ssq_u	The estimate of the heterogeneity variance.
homogeneity	The test statistic and p-value for the test of homogeneity.
global	The test statistic and p-value for the test of model explanatory power.
blups	The conditional expected values of the diversity estimates (conditional on the random effects). The authors propose that if the practitioner believes that information from one diversity estimator may inform the others, then using the ‘condfits’ as estimators of total diversity rather than ‘Chats’ may reduce variance of diversity estimates by “sharing strength” across the samples.
blupses	The estimated standard deviation (standard errors) in the blups.

## Note

Ecologists who are interested in the way species richness varies with covariate information often run a regression-type analysis on the observed diversity using their covariate information as predictors. However, in many settings (especially microbial), rare and unobserved taxa play a hugely important role in explaining the subtleties of the ecosystem, however, a regression analysis on the observed diversity level fails to account for these unobserved taxa. By predicting the total level of diversity (for example, via [breakaway](#)) and estimating the standard error in the estimate, one can take account of these unobserved, but important, taxa. In order to account for the estimated nature of the

response, a mixed model approach is taken, whereby the varying levels of confidence in the estimates contributes to a diagonal but heteroscedastic covariance matrix. Given covariates constitute the fixed effects in the mixed model, and significance of the random effect term “sigsq\_u” reflects heterogeneity in the sample, that is, variability that cannot be explained by only the covariates. The authors believe this to be the first attempt at modelling total diversity in a way that accounts for its estimated nature.

### Author(s)

Amy Willis

### References

- Willis, A., Bunge, J., and Whitman, T. (2015). Inference for changes in biodiversity. *arXiv preprint*.
- Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*.

### See Also

[breakaway](#); [breakaway\\_nof1](#); [apples](#)

### Examples

```
beta(c(2000, 3000, 4000, 3000), c(100, 200, 150, 180), cbind(1, c(100, 150, 100,
50)))

## handles missing data
beta(c(2000, 3000, 4000, 3000), c(100, 200, 150, NA))

## A test for heterogeneity of apples diversity estimates vs butterfly estimates
beta(c(1552, 1500, 884), c(305, 675, 205), cbind(1, c(0, 0, 1)))
```

---

betta_pic	<i>function for plotting total diversity</i>
-----------	--

---

## Description

A simple plotting interface for comparing total diversity across samples or a covariate gradient.

## Usage

```
betta_pic(y, se, x = 1:length(y), ylimu = NULL, myy = NULL,
  mymain = NULL, mycol = NULL, labs = NULL, mypch = NULL,
  myxlim = NULL)
```

## Arguments

y	A vector of estimates of total diversity. Other parameter estimates are accessible; this method may be used for plotting any parameter estimates..
se	The standard errors in ‘y’, the diversity (or other parameter’s) estimates.
x	A vector of covariates to form the x-coordinates of the intervals. If no argument is given, defaults to the order.
ylimu	The upper endpoint of the y-axis.
myy	Deprecated, for backwards compatibility
mymain	Deprecated, for backwards compatibility
mycol	Deprecated, for backwards compatibility
labs	Deprecated, for backwards compatibility
mypch	Deprecated, for backwards compatibility
myxlim	Deprecated, for backwards compatibility

## Author(s)

Amy Willis

## References

Willis, A., Bunge, J., and Whitman, T. (2015). Inference for changes in biodiversity. arXiv preprint.

## See Also

[betta](#)

## Examples

```
betta_pic(c(1552, 1500, 884), c(305, 675, 205), mymain = "Example title")
```

---

betta_random	<i>modelling total diversity with random effects</i>
--------------	--

---

## Description

This function extends `betta()` to permit random effects modelling.

## Usage

```
betta_random(chats, ses, X = NA, groups)
```

## Arguments

<code>chats</code>	A vector of estimates of total diversity at different sampling locations.
<code>ses</code>	The standard errors in ‘chats’, the diversity estimates.
<code>X</code>	A numeric matrix of covariates corresponding to fixed effects. If not supplied, an intercept-only model will be fit.
<code>groups</code>	A categorical variable representing the random-effects groups that each of the estimates belong to.

## Value

<code>table</code>	A coefficient table for the model parameters. The columns give the parameter estimates, standard errors, and p-values, respectively. This model is only as effective as your diversity estimation procedure; for this reason please confirm that your estimates are appropriate and that your model is not misspecified. <code>betta_pic</code> may be useful for this purpose.
<code>cov</code>	Estimated covariance matrix of the parameter estimates.
<code>ssq_u</code>	The estimate of the heterogeneity variance.
<code>ssq_g</code>	Estimates of within-group variance. The estimate will be zero for groups with only one observation.
<code>homogeneity</code>	The test statistic and p-value for the test of homogeneity.
<code>global</code>	The test statistic and p-value for the test of model explanatory power.
<code>blups</code>	The conditional expected values of the diversity estimates (conditional on the random effects). Estimates of variability for the random effects case are unavailable at this time; please contact the maintainer if needed.



**Author(s)**

Amy Willis

**References**Willis, A., Bunge, J., and Whitman, T. (2015). Inference for changes in biodiversity. *arXiv preprint*.**See Also**[betta](#);**Examples**

```

betta_random(c(2000, 3000, 4000, 3000), c(100, 200, 150, 180), X = cbind(Int = 1,
  Cont_var = c(100, 150, 100, 50)), groups = c("a", "a", "a", "b"))

## handles missing data
betta_random(c(2000, 3000, 4000, 3000), c(100, 200, 150, NA), groups = c("a", NA,
  "b", "b"))

```

---

breakaway	<i>Species richness estimation with breakaway</i>
-----------	---

---

**Description**

breakaway is a wrapper for modern species richness estimation for modern datasets

**Usage**

```

breakaway(input_data, cutoff = NA, output = NULL, plot = NULL,
  answers = NULL, print = NULL, ...)

```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code>
cutoff	The maximum frequency count to use for model fitting
output	Deprecated; only for backwards compatibility
plot	Deprecated; only for backwards compatibility
answers	Deprecated; only for backwards compatibility
print	Deprecated; only for backwards compatibility
...	Additional arguments will be ignored; this is for backward compatibility

**Value**

An object of class `alpha_estimate`

**Note**

‘breakaway’ presents an estimator of species richness that is well-suited to the high-diversity/microbial setting. However, many microbial datasets display more diversity than the Kemp-type models can permit. In this case, the log-transformed WLRM diversity estimator of Rocchetti et. al. (2011) is returned. The authors’ experience suggests that some datasets that require the log-transformed WLRM contain “false” diversity, that is, diversity attributable to sequencing errors (via an inflated singleton count). The authors encourage judicious use of diversity estimators when the dataset may contain these errors, and recommend the use of `breakaway_nof1` as an exploratory tool in this case.

**Author(s)**

Amy Willis

**References**

Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*, **71**(4), 1042–1049.

**See Also**

`breakaway_nof1`; `kemp`; `apples`

**Examples**

```
breakaway(apples)
breakaway(apples, plot = FALSE, output = FALSE, answers = TRUE)
```

---

breakaway_nof1	<i>species richness estimation without singletons</i>
----------------	---

---

**Description**

This function permits estimation of total diversity based on a sample frequency count table. Unlike `breakaway`, it does not require an input for the number of species observed once, making it an excellent exploratory tool for microbial ecologists who believe that their sample may contain spurious singletons. The underlying estimation procedure is similar to that of `breakaway` and is outlined in Willis & Bunge (2014). The diversity estimate, standard error, estimated model coefficients and plot of the fitted model are returned.

**Usage**

```
breakaway_nof1(input_data, output = NULL, plot = NULL,
  answers = NULL, print = NULL)
```

**Arguments**

<code>input_data</code>	An input type that can be processed by <code>convert()</code>
<code>output</code>	Deprecated; only for backwards compatibility
<code>plot</code>	Deprecated; only for backwards compatibility
<code>answers</code>	Deprecated; only for backwards compatibility
<code>print</code>	Deprecated; only for backwards compatibility

**Value**

An object of class `alpha_estimate`

<code>code</code>	A category representing algorithm behaviour. ‘code=1’ indicates no nonlinear models converged and the transformed WLRM diversity estimate of Rocchetti et. al. (2011) is returned. ‘code=2’ indicates that the iteratively reweighted model converged and was returned. ‘code=3’ indicates that iterative reweighting did not converge but a model based on a simplified variance structure was returned (in this case, the variance of the frequency ratios is assumed to be proportional to the denominator frequency index). Please peruse your fitted model before using your diversity estimate.
<code>name</code>	The “name” of the selected model. The first integer represents the numerator polynomial degree and the second integer represents the denominator polynomial degree. See Willis & Bunge (2014) for details.
<code>para</code>	Estimated model parameters and standard errors.
<code>est</code>	The estimate of total (observed plus unobserved) diversity.
<code>seest</code>	The standard error in the diversity estimate.
<code>full</code>	The chosen nonlinear model for frequency ratios.

**Note**

It is common for microbial ecologists to believe that their dataset contains false diversity. This often arises because sequencing errors result in classifying already observed organisms as new organisms. ‘breakaway\_nof1’ was developed as an exploratory tool in this case. Practitioners can run ‘breakaway’ on their dataset including the singletons, and ‘breakaway\_nof1’ on their dataset excluding the singletons, and assess if the estimated levels of diversity are very different. Great disparity may provide evidence of an inflated singleton count, or at the very least, that ‘breakaway’ is especially sensitive to the number of rare species observed. Note that ‘breakaway\_nof1’ may be less stable than ‘breakaway’ due to predicting based on a reduced dataset, and have greater standard errors.

**Author(s)**

Amy Willis

**References**

- Willis, A. (2015). Species richness estimation with high diversity but spurious singletons. *arXiv*.  
 Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*.

**See Also**

[breakaway](#); [apples](#)

**Examples**

```
breakaway_nof1(apples[-1, ])
breakaway_nof1(apples[-1, ], plot = FALSE, output = FALSE, answers = TRUE)
```

---

```
build_frequency_count_tables
```

*Build frequency count tables from an OTU table*

---

### Description

Build frequency count tables from an OTU table

### Usage

```
build_frequency_count_tables(the_table)
```

### Arguments

<code>the_table</code>	An OTU table as a data frame or a matrix. Columns are the samples and rows give the taxa.
------------------------	---

### Value

A list of frequency count tables corresponding to the columns.

---

```
chao1
```

*Chao1 species richness estimator*

---

### Description

This function implements the Chao1 richness estimate, which is often mistakenly referred to as an index.

### Usage

```
chao1(input_data, output = NULL, answers = NULL)
```

### Arguments

<code>input_data</code>	An input type that can be processed by <code>convert()</code> or a phyloseq object
<code>output</code>	Deprecated; only for backwards compatibility
<code>answers</code>	Deprecated; only for backwards compatibility

### Value

An object of class `alpha_estimate`, or `alpha_estimates` for phyloseq objects

### Note

The authors of this package strongly discourage the use of this estimator. It is only valid when you wish to assume that every taxa has equal probability of being observed. You don't really think that's possible, do you?

**Author(s)**

Amy Willis

**Examples**

```
chao1(apples)
```

---

chao1\_bc

*Bias-corrected Chao1 species richness estimator*


---

**Description**

This function implements the bias-corrected Chao1 richness estimate.

**Usage**

```
chao1_bc(input_data, output = NULL, answers = NULL)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a <code>phyloseq</code> object
output	Deprecated; only for backwards compatibility
answers	Deprecated; only for backwards compatibility

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Note**

The authors of this package strongly discourage the use of this estimator. It is only valid when you wish to assume that every taxa has equal probability of being observed. You don't really think that's possible, do you?

**Author(s)**

Amy Willis

**Examples**

```
chao1_bc(apples)
```

---

chao_bunge	<i>Chao-Bunge species richness estimator</i>
------------	--

---

**Description**

This function implements the species richness estimation procedure outlined in Chao & Bunge (2002).

**Usage**

```
chao_bunge(input_data, cutoff = 10, output = NULL, answers = NULL)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a phyloseq object
cutoff	The maximum frequency to use in fitting.
output	Deprecated; only for backwards compatibility
answers	Deprecated; only for backwards compatibility

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for phyloseq objects

**Author(s)**

Amy Willis

**Examples**

```
chao_bunge(apples)
```

---

chao_shen	<i>The Chao-Shen estimate of Shannon diversity</i>
-----------	--

---

**Description**

The Chao-Shen estimate of Shannon diversity

**Usage**

```
chao_shen(input_data)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code>
------------	---

**Value**

An object of class `alpha_estimate`

---

check_format	<i>Run some basic checks on a possible frequency count table</i>
--------------	--

---

**Description**

Run some basic checks on a possible frequency count table

**Usage**

```
check_format(output_data)
```

**Arguments**

output\_data      A matrix to test

**Value**

A checked frequency count table

---

convert	<i>convert between different inputs for alpha-diversity estimates</i>
---------	---

---

**Description**

Inputs slated for development include phyloseq and otu\_table

**Usage**

```
convert(input_data)
```

**Arguments**

input\_data      Supported types include filenames, data frames, matrices, vectors...

**Value**

Frequency count able

---

cutoff_wrap	<i>Find a cut-off for estimates relying on contiguous counts</i>
-------------	--

---

**Description**

Find a cut-off for estimates relying on contiguous counts

**Usage**

```
cutoff_wrap(my_data, requested = NA)
```

**Arguments**

my_data	Frequency count table
requested	The user-requested cutoff

**Value**

Cutoff value

---

good_turing	<i>The Good-Turing estimate of species richness</i>
-------------	---

---

**Description**

The Good-Turing estimate of species richness

**Usage**

```
good_turing(input_data)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a <code>phyloseq</code> object
------------	---

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects



---

hawaii	<i>(Data) Frequency count table of soil microbes in Hawaii.</i>
--------	---

---

**Description**

(Data) Frequency count table of soil microbes in Hawaii.

**Usage**

```
hawaii
```

**Format**

A data frame with 198 rows and 2 variables:

**index** an index variable

**frequency** number of taxa that were observed with this frequency ...

**Source**

<http://onlinelibrary.wiley.com/doi/10.1111/biom.12332/abstract>

**References**

Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*, **71**(4), 1042–1049. <http://onlinelibrary.wiley.com/doi/10.1111/biom.12332/abstract>

---

kemp	<i>Species richness estimation with Kemp-type models</i>
------	--

---

**Description**

This function implements the species richness estimation procedure outlined in Willis & Bunge (2015). The diversity estimate, standard error, estimated model coefficients, model details and plot of the fitted model are returned.

**Usage**

```
kemp(input_data, cutoff = NA, output = NULL, plot = NULL,
      answers = NULL, print = NULL, ...)
```

**Arguments**

<code>input_data</code>	An input type that can be processed by <code>convert()</code>
<code>cutoff</code>	The maximum frequency count to use for model fitting
<code>output</code>	Deprecated; only for backwards compatibility
<code>plot</code>	Deprecated; only for backwards compatibility
<code>answers</code>	Deprecated; only for backwards compatibility
<code>print</code>	Deprecated; only for backwards compatibility
<code>...</code>	Additional arguments will be ignored; this is for backward compatibility

**Value**

An object of class `alpha_estimate`

<code>code</code>	A category representing algorithm behaviour. ‘code=1’ indicates no nonlinear models converged and the transformed WLRM diversity estimate of Rocchetti et. al. (2011) is returned. ‘code=2’ indicates that the iteratively reweighted model converged and was returned. ‘code=3’ indicates that iterative reweighting did not converge but a model based on a simplified variance structure was returned (in this case, the variance of the frequency ratios is assumed to be proportional to the denominator frequency index). Please peruse your fitted model before using your diversity estimate.
<code>name</code>	The “name” of the selected model. The first integer represents the numerator polynomial degree and the second integer represents the denominator polynomial degree of the model for the frequency ratios. See Willis & Bunge (2015) for details.
<code>para</code>	Estimated model parameters and standard errors.
<code>est</code>	The estimate of total (observed plus unobserved) diversity.
<code>seest</code>	The standard error in the diversity estimate.
<code>full</code>	The chosen nonlinear model for frequency ratios.
<code>ci</code>	An asymmetric 95% confidence interval for diversity.

**Author(s)**

Amy Willis

**References**

- Willis, A. and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*, **71**(4), 1042–1049.
- Rocchetti, I., Bunge, J. and Bohning, D. (2011). Population size estimation based upon ratios of recapture probabilities. *Annals of Applied Statistics*, **5**.

**See Also**

[breakaway](#); [breakaway\\_nof1](#); [apples](#)

**Examples**

```
kemp(apples)
kemp(apples, plot = FALSE, output = FALSE, answers = TRUE)
```

---

make_design_matrix	<i>Make design matrix</i>
--------------------	---------------------------

---

**Description**

Make design matrix

**Usage**

```
make_design_matrix(phyloseq_object, variables)
```

**Arguments**

phyloseq_object	A phyloseq object
variables	variable names

---

make_frequency_count_table	<i>Draw frequency count subtables from an OTU table</i>
----------------------------	---

---

**Description**

Draw frequency count subtables from an OTU table

**Usage**

```
make_frequency_count_table(labels)
```

**Arguments**

labels	A vector of counts of the taxa; i.e. a vector giving the number of times each taxon was observed.
--------	---

---

objective_bayes_geometric	<i>Estimate species richness with an objective Bayes method using a geometric model</i>
---------------------------	---

---

**Description**

Estimate species richness with an objective Bayes method using a geometric model

**Usage**

```
objective_bayes_geometric(data, output = TRUE, plot = TRUE,
  answers = FALSE, tau = 10, burn.in = 100, iterations = 2500,
  Metropolis.stdev.N = 75, Metropolis.start.theta = 1,
  Metropolis.stdev.theta = 0.3)
```

**Arguments**

data	TODO(Kathryn)(Kathryn)
output	TODO(Kathryn)(Kathryn)
plot	TODO(Kathryn)(Kathryn)
answers	TODO(Kathryn)(Kathryn)
tau	TODO(Kathryn)
burn.in	TODO(Kathryn)
iterations	TODO(Kathryn)
Metropolis.stdev.N	TODO(Kathryn)
Metropolis.start.theta	TODO(Kathryn)
Metropolis.stdev.theta	TODO(Kathryn)

---

objective\_bayes\_mixedgeo

*Objective Bayes species richness estimate with the mixed-geometric model*

---

**Description**

Objective Bayes species richness estimate with the mixed-geometric model

**Usage**

```
objective_bayes_mixedgeo(data, output = TRUE, plot = TRUE,
  answers = FALSE, tau = 10, burn.in = 100, iterations = 2500,
  Metropolis.stdev.N = 100, Metropolis.start.T1 = 1,
  Metropolis.stdev.T1 = 2, Metropolis.start.T2 = 3,
  Metropolis.stdev.T2 = 2, bars = 3)
```

**Arguments**

data	TODO(Kathryn)
output	TODO(Kathryn)
plot	TODO(Kathryn)
answers	TODO(Kathryn)
tau	TODO(Kathryn)
burn.in	TODO(Kathryn)
iterations	TODO(Kathryn)
Metropolis.stdev.N	TODO(Kathryn)
Metropolis.start.T1	TODO(Kathryn)

```

Metropolis.stdev.T1
      TODO(Kathryn)
Metropolis.start.T2
      TODO(Kathryn)
Metropolis.stdev.T2
      TODO(Kathryn)
bars      TODO(Kathryn)

```

---

objective\_bayes\_negbin

*Objective Bayes species richness estimate with the Negative Binomial model*

---

## Description

Objective Bayes species richness estimate with the Negative Binomial model

## Usage

```

objective_bayes_negbin(data, output = TRUE, plot = TRUE,
  answers = FALSE, tau = 10, burn.in = 1000, iterations = 5000,
  Metropolis.stdev.N = 100, Metropolis.start.T1 = -0.8,
  Metropolis.stdev.T1 = 0.01, Metropolis.start.T2 = 0.8,
  Metropolis.stdev.T2 = 0.01, bars = 5)

```

## Arguments

```

data      TODO(Kathryn)
output    TODO(Kathryn)
plot      TODO(Kathryn)
answers   TODO(Kathryn)
tau       TODO(Kathryn)
burn.in   TODO(Kathryn)
iterations TODO(Kathryn)
Metropolis.stdev.N
      TODO(Kathryn)
Metropolis.start.T1
      TODO(Kathryn)
Metropolis.stdev.T1
      TODO(Kathryn)
Metropolis.start.T2
      TODO(Kathryn)
Metropolis.stdev.T2
      TODO(Kathryn)
bars      TODO(Kathryn)

```

---

objective\_bayes\_poisson

*Objective Bayes species richness estimate with the Poisson model*

---

### Description

Objective Bayes species richness estimate with the Poisson model

### Usage

```
objective_bayes_poisson(data, output = TRUE, plot = TRUE,
  answers = FALSE, tau = 10, burn.in = 100, iterations = 2500,
  Metropolis.stdev.N = 75, Metropolis.start.lambda = 1,
  Metropolis.stdev.lambda = 0.3, bars = 5)
```

### Arguments

data	TODO(Kathryn)
output	TODO(Kathryn)
plot	TODO(Kathryn)
answers	TODO(Kathryn)
tau	TODO(Kathryn)
burn.in	TODO(Kathryn)
iterations	TODO(Kathryn)
Metropolis.stdev.N	TODO(Kathryn)
Metropolis.start.lambda	TODO(Kathryn)
Metropolis.stdev.lambda	TODO(Kathryn)
bars	TODO(Kathryn)

---

physeq\_wrap

*Wrapper for phyloseq objects*

---

### Description

Wrapper for phyloseq objects

### Usage

```
physeq_wrap(fn, physeq, ...)
```

### Arguments

fn	alpha diversity estimator function with breakaway to be applied to physeq
physeq	A phyloseq object, or an object of class otu_table
...	Additional arguments for fn

**Value**

Object of class alpha\_estimates

---

plot.alpha\_estimates    *Plot function for alpha\_estimates class*

---

**Description**

Plot function for alpha\_estimates class

**Usage**

```
## S3 method for class 'alpha_estimates'
plot(x, physeq = NULL, measure = NULL,
     color = NULL, shape = NULL, title = NULL, trim_plot = FALSE, ...)
```

**Arguments**

x	Object of class alpha_estimates.
physeq	(Optional). Default NULL. Required object of class phyloseq if including a sample_data variable for color or shape.
measure	(Optional). If there are multiple richness measures included in x, this can be set to the the desired measure to be plotted. Defaults to the measure of the first alpha diversity estimate.
color	(Optional). Default NULL. The sample variable to map to different colors. Can be a single character string of the variable name in sample_data or a custom supplied vector with length equal to the number of samples.
shape	(Optional). Default NULL. The sample variable to map to different shapes. Can be a single character string of the variable name in sample_data or a custom supplied vector with length equal to the number of samples.
title	(Optional). Default NULL. Character string. The main title for the graphic.
trim_plot	(Optional). Default FALSE. Boolean indicator for whether you want the plot to include the full confidence intervals.
...	See details

**Details**

... does not currently have any implemented options. Optional arguments currently include "trim\_plot", a Optional

**Examples**

```
## Not run:
library(phyloseq)
data(GlobalPatterns)
alphas <- breakaway(GlobalPatterns)
plot(alphas)

## End(Not run)
```

---

poisson_model	<i>PoissonModel</i>
---------------	---------------------

---

**Description**

A model to estimate the number of missing taxa under a Poisson Model

**Usage**

```
poisson_model(input_data, cutoff = 10)
```

**Arguments**

input_data	A frequency count table
cutoff	The largest frequency to use for predicting f0. Defaults to 10.

---

proportions_instead	<i>OTU table to relative abundances</i>
---------------------	---

---

**Description**

OTU table to relative abundances

**Usage**

```
proportions_instead(the_table)
```

**Arguments**

the_table	An OTU table
-----------	--------------

---

rnbinomtable	<i>Negative binomially distributed frequency count tables.</i>
--------------	--

---

**Description**

Simulate a frequency count table based on a negative binomial model. Zero-truncated, obviously.

**Usage**

```
rnbinomtable(C, size, probability)
```

**Arguments**

C	species richness
size	size parameter for the negative binomial distribution
probability	probability parameter for the negative binomial distribution

**Author(s)**

Amy Willis



---

rztnbinomtable	<i>beta version: Zero-truncated negative binomially distributed frequency count tables.</i>
----------------	---

---

**Description**

Simulate a frequency count table based on a negative binomial model. Zero-truncated, obviously.

**Usage**

```
rztnbinomtable(C, size, probability)
```

**Arguments**

C	species richness
size	size parameter for the negative binomial distribution
probability	probability parameter for the negative binomial distribution

**Author(s)**

Amy Willis

---

sample_inverse_simpson	<i>Plug-in Inverse Simpson diversity</i>
------------------------	--

---

**Description**

This function implements the plug-in Inverse Simpson diversity

**Usage**

```
sample_inverse_simpson(input_data)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a <code>phyloseq</code> object
------------	---

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Examples**

```
sample_inverse_simpson(apples)
```

---

sample_richness	<i>Sample richness estimator</i>
-----------------	----------------------------------

---

**Description**

This function implements the sample richness estimate, which is the number of non-zero taxa per sample.

**Usage**

```
sample_richness(input_data)
```

**Arguments**

input\_data      An input type that can be processed by `convert()` or a phyloseq object

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for phyloseq objects

**Examples**

```
sample_richness(apples)
```

---

sample_shannon	<i>Plug-in Shannon diversity</i>
----------------	----------------------------------

---

**Description**

This function implements the plug-in Shannon diversity

**Usage**

```
sample_shannon(input_data)
```

**Arguments**

input\_data      An input type that can be processed by `convert()` or a phyloseq object

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for phyloseq objects

**Examples**

```
sample_shannon(apples)
```

---

sample_shannon_e	<i>Plug-in Shannon's E ("Equitability")</i>
------------------	---

---

**Description**

This function implements the plug-in Shannon's E

**Usage**

```
sample_shannon_e(input_data)
```

**Arguments**

input\_data      An input type that can be processed by `convert()` or a `phyloseq` object

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Examples**

```
sample_shannon_e(apples)
```

---

sample_simpson	<i>Plug-in Simpson diversity</i>
----------------	----------------------------------

---

**Description**

This function implements the plug-in Simpson diversity

**Usage**

```
sample_simpson(input_data)
```

**Arguments**

input\_data      An input type that can be processed by `convert()` or a `phyloseq` object

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Examples**

```
sample_simpson(apples)
```

---

sample_size_estimate	<i>Estimate the sample size needed to do an unpaired one-way test using beta</i>
----------------------	--

---

**Description**

Estimate the sample size needed to do an unpaired one-way test using beta

**Usage**

```
sample_size_estimate(control_group_est, se_est, diff = 5, alpha = 0.05,
  prop = 0.8, samples = 20, precision = 5)
```

**Arguments**

control_group_est	An estimate of the alpha diversity parameter for the control group
se_est	An estimate of the (common) standard deviation
diff	An estimate of the difference between the control and non-control groups
alpha	Minimum significance level desired
prop	Desired power
samples	Number of bootstrap resamples used to estimate the sample size. Increase for a more accurate estimate.
precision	How much to increment the sample size as we try to increase the power

**Value**

An estimate of the necessary sample size and some details

---

sample_size_figure	<i>Plot the power obtained with sample size</i>
--------------------	---

---

**Description**

Plot the power obtained with sample size

**Usage**

```
sample_size_figure(control_group_est, se_est, diff = 5, samples = 20)
```

**Arguments**

control_group_est	An estimate of the alpha diversity parameter for the control group
se_est	An estimate of the (common) standard deviation
diff	An estimate of the difference between the control and non-control groups
samples	Number of bootstrap resamples used to estimate the sample size. Increase for a more accurate estimate.

**Value**

A plot of the power with the sample size

---

toy_metadata	(Data) Data frame of covariate information about toy_otu_table.
--------------	---

---

**Description**

(Data) Data frame of covariate information about toy\_otu\_table.

**Usage**

```
toy_metadata
```

**Format**

A data frame with 143 rows and 4 variables:

**Years** Year of sampling

**bloom2** Did the sample correspond to a bloom event?

**Period** What season was sampled?

**Site** Where was the sample taken from? ...

---

toy_otu_table	(Data) A toy OTU table.
---------------	-------------------------

---

**Description**

Covariate info available in ‘toy\_metadata’. A data frame with 448 rows and 143 columns. Rows give the abundance of each taxon; columns give the samples

**Usage**

```
toy_otu_table
```

**Format**

An object of class data.frame with 448 rows and 143 columns.

---

toy_taxonomy	(Data) The taxonomy of the OTUs in 'toy_otu_table'.
--------------	---

---

**Description**

(Data) The taxonomy of the OTUs in 'toy\_otu\_table'.

**Usage**

```
toy_taxonomy
```

**Format**

An object of class factor of length 448.

---

true_gini	Calculate the true Gini-Simpson index
-----------	---------------------------------------

---

**Description**

Calculate the true Gini-Simpson index

**Usage**

```
true_gini(input)
```

**Arguments**

input	A vector of proportions.
-------	--------------------------

**Value**

The Gini-Simpson index of the population given by input.

**Note**

This function is intended for population-level data. If you are dealing with a microbial sample, use DivNet instead.

---

true_hill	<i>Calculate the true Hill numbers</i>
-----------	--

---

**Description**

Calculate the true Hill numbers

**Usage**

```
true_hill(input, q)
```

**Arguments**

input	A vector of proportions.
q	The Hill number of interest. $q = 0$ corresponds to species richness, $q = 2$ corresponds to inverse Simpson, etc.

**Value**

The Hill number of the population given by input.

---

true_inverse_simpson	<i>Calculate the true Inverse Simpson index</i>
----------------------	---

---

**Description**

Calculate the true Inverse Simpson index

**Usage**

```
true_inverse_simpson(input)
```

**Arguments**

input	A vector of proportions.
-------	--------------------------

**Value**

The inverse-Simpson index of the population given by input.

**Note**

This function is intended for population-level data. If you are dealing with a microbial sample, use DivNet instead.

---

true_shannon	<i>Calculate the true Shannon index based on proportions</i>
--------------	--

---

**Description**

Calculate the true Shannon index based on proportions

**Usage**

```
true_shannon(input)
```

**Arguments**

input	A vector of proportions.
-------	--------------------------

**Value**

The Shannon index of the population given by input.

**Note**

This function is intended for population-level data. If you are dealing with a microbial sample, use DivNet instead.

---

true_shannon_e	<i>Calculate the true Shannon's equitability index</i>
----------------	--

---

**Description**

Calculate the true Shannon's equitability index

**Usage**

```
true_shannon_e(input)
```

**Arguments**

input	A vector of proportions.
-------	--------------------------

**Value**

The Shannon E's of the population given by input.

**Note**

This function is intended for population-level data. If you are dealing with a microbial sample, use DivNet instead.



---

true_simpson	<i>Calculate the true Simpson index</i>
--------------	---

---

**Description**

Calculate the true Simpson index

**Usage**

```
true_simpson(input)
```

**Arguments**

input            A vector of proportions.

**Value**

The Simpson index of the population given by input.

**Note**

This function is intended for population-level data. If you are dealing with a microbial sample, use DivNet instead.

---

wlrm_transformed	<i>The transformed weighted linear regression estimator for species richness estimation</i>
------------------	---

---

**Description**

This function implements the transformed version of the species richness estimation procedure outlined in Rocchetti, Bunge and Bohning (2011).

**Usage**

```
wlrm_transformed(input_data, cutoff = NA, print = NULL, plot = NULL,
  answers = NULL)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a <code>phyloseq</code> object
cutoff	Maximum frequency count to use
print	Deprecated; only for backwards compatibility
plot	Deprecated; only for backwards compatibility
answers	Deprecated; only for backwards compatibility

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Note**

While robust to many different structures, model is almost always misspecified. The result is usually implausible diversity estimates with artificially low standard errors. Extreme caution is advised.

**Author(s)**

Amy Willis

**References**

Rocchetti, I., Bunge, J. and Bohning, D. (2011). Population size estimation based upon ratios of recapture probabilities. *Annals of Applied Statistics*, **5**.

**See Also**

[breakaway](#); [apples](#); [wlrn\\_untransformed](#)

**Examples**

```
wlrn_transformed(apples)
wlrn_transformed(apples, plot = FALSE, print = FALSE, answers = TRUE)
```

---

wlrn_untransformed	<i>The untransformed weighted linear regression estimator for species richness estimation</i>
--------------------	---

---

**Description**

This function implements the untransformed version of the species richness estimation procedure outlined in Rocchetti, Bunge and Bohning (2011).

**Usage**

```
wlrn_untransformed(input_data, cutoff = NA, print = NULL,
  plot = NULL, answers = NULL)
```

**Arguments**

input_data	An input type that can be processed by <code>convert()</code> or a <code>phyloseq</code> object
cutoff	Maximum frequency count to use
print	Deprecated; only for backwards compatibility
plot	Deprecated; only for backwards compatibility
answers	Deprecated; only for backwards compatibility

**Value**

An object of class `alpha_estimate`, or `alpha_estimates` for `phyloseq` objects

**Note**

This estimator is based on the negative binomial model and for that reason generally produces poor fits to microbial data. The result is usually artificially low standard errors. Caution is advised.

**Author(s)**

Amy Willis

**References**

Rocchetti, I., Bunge, J. and Bohning, D. (2011). Population size estimation based upon ratios of recapture probabilities. *Annals of Applied Statistics*, **5**.

**See Also**

[breakaway](#); [apples](#); [wlrm\\_transformed](#)

**Examples**

```
wlrm_untransformed(apples)
```

# Index

## \*Topic **datasets**

- apples, [4](#)
- hawaii, [17](#)
- toy\_metadata, [29](#)
- toy\_otu\_table, [29](#)
- toy\_taxonomy, [30](#)

## \*Topic **diversity**

- betta, [5](#)
- betta\_pic, [7](#)
- betta\_random, [8](#)
- breakaway, [9](#)
- breakaway\_nof1, [10](#)
- kemp, [17](#)
- wlrm\_transformed, [33](#)
- wlrm\_untransformed, [34](#)

## \*Topic **error**

- breakaway\_nof1, [10](#)

## \*Topic **microbial**

- breakaway, [9](#)
- breakaway\_nof1, [10](#)
- kemp, [17](#)

## \*Topic **models**

- breakaway, [9](#)
- breakaway\_nof1, [10](#)
- kemp, [17](#)
- wlrm\_transformed, [33](#)
- wlrm\_untransformed, [34](#)

## \*Topic **nonlinear**

- breakaway, [9](#)
- breakaway\_nof1, [10](#)
- kemp, [17](#)

alpha\_estimate, [3](#)

alpha\_estimates, [4](#)

apples, [4](#), [6](#), [10](#), [11](#), [18](#), [34](#), [35](#)

betta, [5](#), [8](#), [9](#)

betta\_pic, [7](#)

betta\_random, [8](#)

breakaway, [5](#), [6](#), [9](#), [10](#), [11](#), [18](#), [34](#), [35](#)

breakaway\_nof1, [6](#), [10](#), [10](#), [18](#)

build\_frequency\_count\_tables, [12](#)

chao1, [12](#)

chao1\_bc, [13](#)

chao\_bunge, [14](#)

chao\_shen, [14](#)

check\_format, [15](#)

convert, [15](#)

cutoff\_wrap, [16](#)

good\_turing, [16](#)

hawaii, [17](#)

kemp, [10](#), [17](#)

make\_design\_matrix, [19](#)

make\_frequency\_count\_table, [19](#)

objective\_bayes\_geometric, [19](#)

objective\_bayes\_mixedgeo, [20](#)

objective\_bayes\_negbin, [21](#)

objective\_bayes\_poisson, [22](#)

physeq\_wrap, [22](#)

plot.alpha\_estimates, [23](#)

poisson\_model, [24](#)

proportions\_instead, [24](#)

rnbinomtable, [24](#)

rztnbinomtable, [25](#)

sample\_inverse\_simpson, [25](#)

sample\_richness, [26](#)

sample\_shannon, [26](#)

sample\_shannon\_e, [27](#)

sample\_simpson, [27](#)

sample\_size\_estimate, [28](#)

sample\_size\_figure, [28](#)

toy\_metadata, [29](#)

toy\_otu\_table, [29](#)

toy\_taxonomy, [30](#)

true\_gini, [30](#)

true\_hill, [31](#)

true\_inverse\_simpson, [31](#)

true\_shannon, [32](#)

true\_shannon\_e, [32](#)

true\_simpson, [33](#)

wlrm\_transformed, [33](#), [35](#)

wlrm\_untransformed, [34](#), [34](#)