Original version (July 2023)

User's Guide for iNEXT.3D Online: Software for Interpolation and Extrapolation in three dimensions.

Anne Chao, K.-H. Hu

Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Overview

iNEXT.3D (iNterpolation and EXTrapolation in three Dimensions) Online is the R-based interactive online version of iNEXT.3D available via the link https://chao.shinyapps.io/iNEXT_3D/or http://chao.stat.nthu.edu.tw/wordpress/software_download/. Clicking these links, you will be directed to the online interface window. Users do not need to learn/understand R to run iNEXT.3D Online. The interactive web application was built using Shiny (a web application framework).

iNEXT.3D is the extension of shiny iNEXT Online (Chao and Jost 2012, 2015; Chao et al., 2014, 2019). It features two statistical analyses (non-asymptotic and asymptotic) for "Taxonomic diversity", "Phylogenetic diversity" and "Functional diversity" based on Hill numbers:

(1) A non-asymptotic approach based on interpolation and extrapolation

iNEXT.3D computes the estimated three kinds of diversities for standardized samples with a common sample size or sample completeness. This approach aims to compare diversity estimates for equally-large (with a common sample size) or equally-complete (with a common sample coverage) samples; it is based on the seamless rarefaction and extrapolation (R/E) sampling curves of Hill numbers for q = 0, 1 and 2. iNEXT.3D offers three types of R/E sampling curves for three kinds of diversities:

- Sample-size-based (or size-based) R/E sampling curves: This type of sampling curve plots the diversity estimates with respect to sample size.
- Coverage-based R/E sampling curves: This type of sampling curve plots the diversity estimates with respect to sample coverage.
- Sample completeness curve: This curve depicts how sample coverage varies with sample size. The sample completeness curve provides a bridge between the size- and coverage-

based R/E sampling curves.

(2) An asymptotic approach to infer asymptotic diversity

iNEXT.3D computes the estimated asymptotic diversity profiles including q-profile, time-profile, and tau-profile. Note that time profile for only when Phylogenetic diversity. Tau profile for only when Functional diversity and the threshold type is "single threshold".

How to cite

If you publish your work based on results from iNEXT.3D Online, you should make references to the following Online reference:

- Chao, A., Henderson, P. A., Chiu, C.-H., Moyes, F., Hu, K.-H., Dornelas, M and. Magurran, A. E. (2021). Measuring temporal change in alpha diversity: a framework integrating taxonomic, phylogenetic and functional diversity and the iNEXT.3D standardization. *Methods in Ecology and Evolution*, 12, 1926-1940.

Data

Three types of data

iNEXT.3D Online supports three types of data:

- Individual-based abundance data: Data for each assemblage/site include sample species abundances in an empirical sample of individuals (called a reference sample). When there are N assemblages and S species, input a txt file consist of a S by N matrix (in addition to the site labels and the species names). Input a txt file consist of species by assemblages matrix.

- Sampling-unit-based incidence data: There are two kinds of input data.

  (1) Incidence-raw data : For one assemblage, input a txt file for a reference sample consist of a species-by-sampling-unit matrix; when there are N assemblages, input N txt file consist of a species-by-sampling-unit matrix.

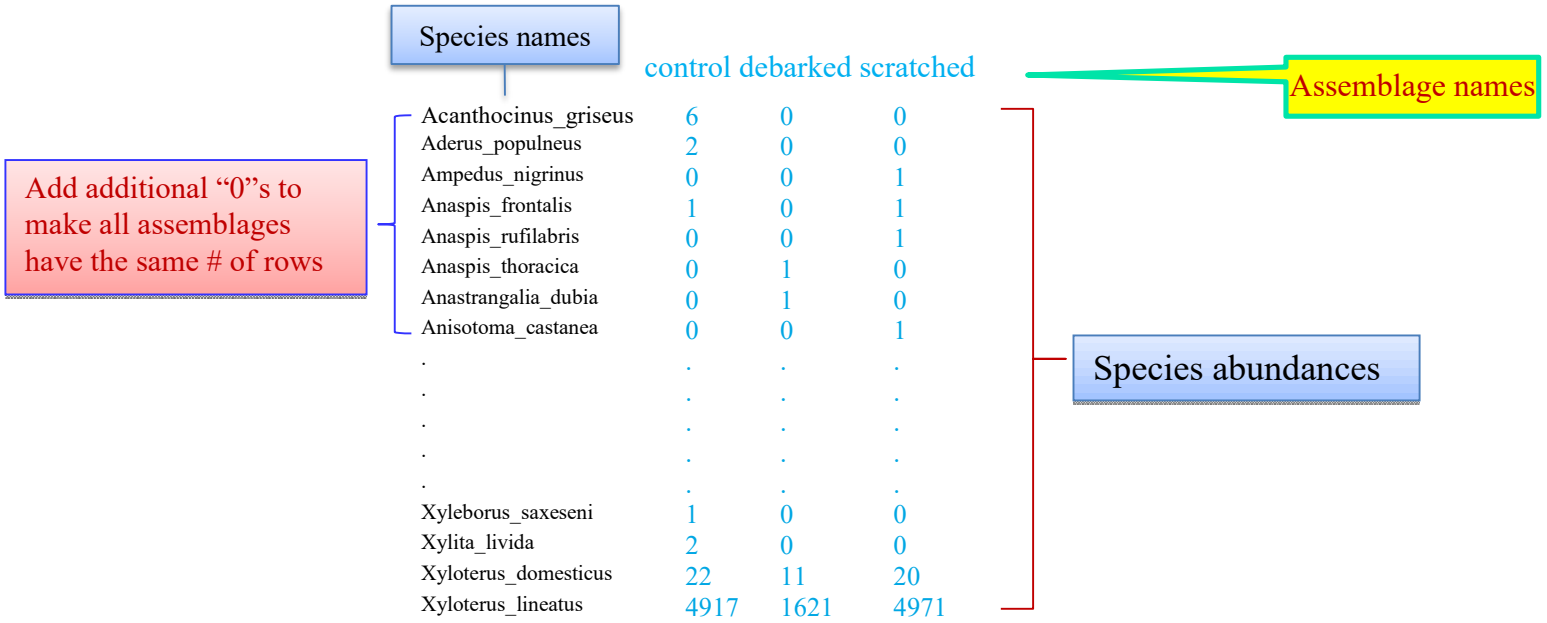
  (2) Incidence-frequency data (for only when Taxonomic and Functional diversity): Input data for each assemblage consist of species sample incidence frequencies (row sums of each incidence raw matrix). When there are N assemblages and S species, input a txt file consist of a (S + 1) by N matrix (in addition to the site labels and the species names). The first row must be the total number of sampling units, followed by the species incidence frequencies.

How to prepare your data input files

- Abundance data

When there are *N* assemblages**,** the observed species abundances should be arranged as a species (in rows) by assemblage (in columns) matrix. The first row (including *N* entries) lists the assemblage labels or site names for the *N* assemblages; see an example below. Beginning with the second row, the entry for the first column in each row denotes a species abundance from Assemblage 1, and the entry for the second column denotes a species abundance from Assemblage 2, and so on. If species names or any identification codes are recorded in your original data, they must be removed to conform to the iNEXT.3D Online data format.

For example, the Beetles' data (one of the three sets for demoing abundance data) consist of species sample abundances. The observed species abundances should be arranged in a text data file as follows for three assemblages: “control”, “debarked” and “scratched”



Missing data, which are empty cells without numbers, cannot be read by iNEXT.3D Online. The number of rows must be the same for the *N* assemblages. For example, there are 84 observed species in the control assemblage, 61 species in the debarked assemblage, 86 species in the scratched site, and total 120 species in three assemblages, the input data must consist of 120 rows (in addition to the site labels and the species names). Therefore, thirty six “0” abundances should be added to the control column, fifty nine “0” abundances should be added to the debarked column and thirty four “0” abundances should be added to the scratched column. Adding “0” abundances will not have any effect on our analysis. Additional rows with frequencies all 0 for unobserved species may be included, but that species will not have any effect on the analysis.

For phylogenetic diversity and functional diversity, species names must be included in your data file, and species names should be some information corresponding to phylogenetic tree or distance matrix. When a species was not observed or detected in a community, you must enter the frequency 0 for a "non-detection". In a typical species-check-list abundance data, numbers in the same row refer to the same species, and this is required for iNEXT.3D Online input. iNEXT.3D Online aims to compare within-assemblage diversity in three dimensions. (Note: if the goal is to assess the extent of differentiation in species composition among assemblages, then among-assemblage information is essential. In this case, use the forthcoming iNEXT.beta3D, a generalization of iNEXT.3D to beta diversity.)

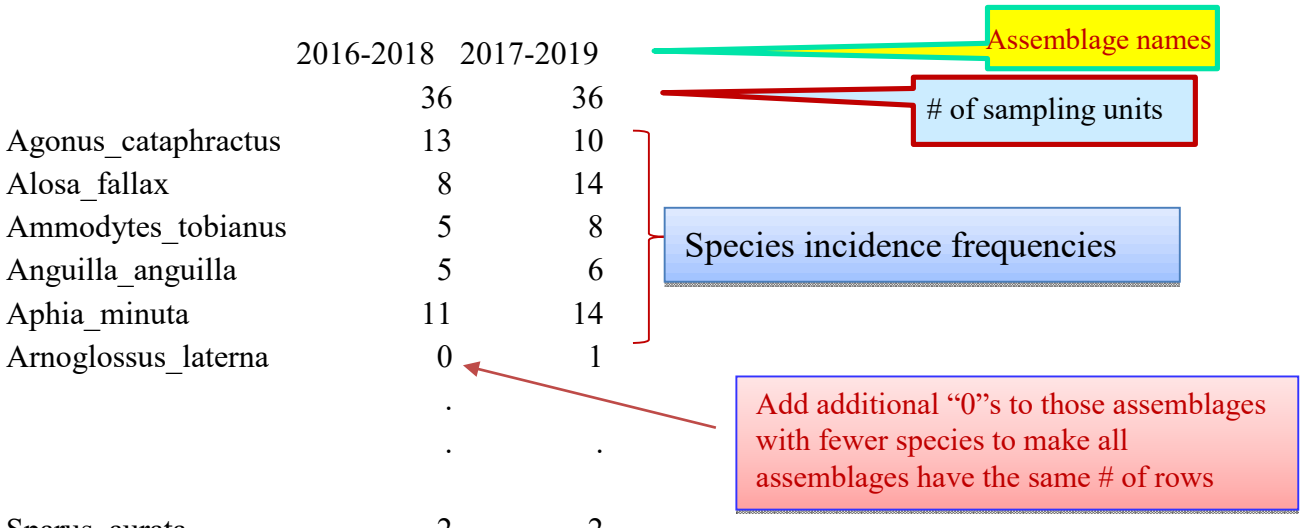
In the special case of $N = 1$ (i.e., there is only one assemblage), all data should be read in one column. That is, the first entry denotes the assemblage label or site name, and the observed species abundances are entered in different rows, with species names specified as row names if you want to compute phylogenetic diversity and functional diversity.

- Incidence frequency data

The data input format is generally similar to that for abundance data, except that the total number of sampling units in each assemblage must be provided. That is, the first row lists the assemblage labels or site names, and the second row lists the number of sampling units in the N assemblages, as shown below. Beginning with the third row, the entry in the first column in each row denotes an incidence frequency (total number of detections of a species in all sampling units) from Assemblage 1, and the entry for the second column denotes an incidence frequency from Assemblage 2, and so on.

For example, in the demo Hinkley's fish freq inc. data, species incidence frequencies were collected at 3 years of information (2016-2018 ; 2017-2019).

	2016-2018	2017-2019
	36	36
Agonus_cataphractus	13	10
Alosa_fallax	8	14
Ammodytes_tobianus	5	8
Anguilla_anguilla	5	6
Aphia_minuta	11	14
Arnoglossus_laterna	0	1
	.	
	.	.
Sparus_aurata	2	2



In the above data, for year 2016-2018, there were 36 sampling units; the first species was detected in 13 sampling units, the second one was detected in 8 sampling units, etc. As with the abundance data, the number of rows for each assemblage must be the same, so additional "0" incidences must be added under assemblages with fewer species.

- Incidence raw data:

The data input format is different from abundance data or incidence frequency data. It can upload multiple text files to iNEXT.3D Online. For each file, the file name means the assemblage/site name. In each file, the first row lists the sampling unit labels, and the other rows lists the incidence of each species in each sampling unit, as shown below. So the value in each entry can only contain "0" (not detect) or "1" (detect). The entry in the first column (in addition to species name) in each row denotes a detection from first sampling unit, and the entry for the second column denotes an detection from second sampling unit, and so on. If you have more than one assemblage/site, then you should upload multiple text files.

For example, in the demo data "Hinkley's fish raw inc", it has two assemblages "2016-2018" and "2017-2019". The species incidence detection were collected at 36 sampling units in each assemblage. The observed species number in this assemblage is 59. The data should be read in a text file as follows.

File name: "18/01/2016"

Assemblage names

	18/01/2016	15/02/2016	16/03/2016	…	16/11/2018	14/12/2018
Agonus..	1	1	1	…	0	1
Alosa..	0	0	0	…	1	1
Ammodytes..	0	0	0	…	0	0
Anguilla..	0	0	…	…	0	1
	.	.	.			.
	.	.	.			.
	.	.	.			.
Wood..	0	0	0		0	0
Wood..	0	0	0		0	0

sampling unit name

raw Species incidence

Species names

In the above data, there were 36 sampling units; the first species was detected in sampling unit '18/01/2016'. In this upload data type, the row names (species names) are not essential for taxonomic diversity, but essentail for phylogenetic and functional diversity.

Running procedures for TD, PD, FD analysis

Species identification names are irrelevant in species/taxonomic diversity assessment so they are not essential for upload file (but necessary for phylogenetic and functional diversity). The following Steps 1, 2 and 3 are required procedures (selections are placed in blocks); Step 4 and 5 are optional.

Step 1. Select an analysis method (Taxonomic diversity or Phylogenetic diversity or Functional diversity) from the top menu of iNEXT.3D Online window.

Step 2. "Data Setting" (on the left hand side of the window screen)

(2a) Select Demo data to load demo data or Upload data to load your own data,

(2b) Select data type: For Taxonomic diversity and Functional diversity, we can select abundance data, incidence raw data or incidence frequency data; For Phylogenetic diversity, we can select abundance data or incidence raw data; see above for data input formats.

(2c) For abundance data, there are three demo data sets. Select one: Brazil rainforest, Beetles' data or Hinkley's fish abu. data.

(2d) Then the names/labels for the uploaded assemblages/sites will be automatically shown in the window below; you can select one set or multiple sets for comparison.

(2e) For Phylogenetic diversity and Functional diversity, there are one demo phylogenetic tree and distance matrix, respectively.

Step 3. Press the Run! button to get the output.

Step 4. Optional: "General setting" for statistical procedures. Only for panel "Rarefaction and Extrapolation" and "Asymptotic and Empirical Analysis".

For panel "Rarefaction and Extrapolation":

(4a) Optional: For Phylogenetic diversity, select the diversity type, including PD (effective total branch length) or meanPD (effective number of equally divergent lineages). (Default is meanPD); For Functional diversity, select the threshold type, including single threshold (under specified threshold) and AUC (for an overall FD which integrates all threshold values between zero and one). (Default of single threshold is d_{mean} (quadratic entropy: the mean distance between any two individuals randomly selected from the pooled assemblage)).

(4b) Optional: select a diversity order of q. (Default is q = 0).

(4c) Optional:choose specify endpoint or specify sample sizes. You can specify the endpoint of the extrapolation range or specify the sample sizes for which diversity estimates will be calculated. (Default endpoint = double of the minimum reference

sample size, and the default number of knots = 40). If you choose to specify the sample sizes, then type in those sample sizes in the window space.

(4d) Optional: specify the number of bootstrap replications to compute s.e. and confidence intervals for each estimator. (Default bootstrap replications = 0). We type in "0" to skip all bootstrapping to save running time.

(4e) Optional: specify the level for the confidence interval (Default level is 0.95).

(4f) Optional: Only for Phylogenetic diversity, specify the reference time. (Default is tree height).

For panel "Asymptotic and Empirical Analysis":

(4a) Optional: For Phylogenetic diversity, select the profile type, including q profile or Time profile. (Default is q profile); For Functional diversity only for single threshold, select the profile type, including q profile and Tau profile. (Default is q profile)

(4b) Optional: For Phylogenetic diversity, select the diversity type, including PD (effective total branch length) or meanPD (effective number of equally divergent lineages). (Default is meanPD); For Functional diversity, select the threshold type, including single threshold (under specified threshold) and AUC (for an overall FD which integrates all threshold values between zero and one). (Default of single threshold is d_{mean} (quadratic entropy: the mean distance between any two individuals randomly selected from the pooled assemblage)).

(4c) Optional: specify the number of bootstrap replications to compute s.e. and confidence intervals for each estimator. (Default bootstrap replications = 0). We type in "0" to skip all bootstrapping to save running time.

(4d) Optional: specify the level for the confidence interval (Default level is 0.95).

(4e) Optional: Only for Phylogenetic diversity, specify the reference time. (Default is tree height).

(4f) Optional: Type in diversity order q. (Default is 0, 0.25, 0.5, …, 2.75, 3) for q-profile; type in time when Time values profile; and type in tau values when Tau profile.

Step 5. Press the Run! button to get the output.

Note 1: We use a bootstrap resampling method to compute the s.e. and confidence interval of any estimator involved in the analysis. The default number of bootstrap replications is 0 to save running time. In this case, the computation of s.e. and confidence intervals will be skipped so that the output can be promptly shown.. You may specify a larger number (say, 50) to obtain more accurate results for publication purposes, but it will take a longer time to

get the output.

Note 2: The bootstrap resampling procedures vary with trial, meaning that two different runs for the same data may result in different s.e. estimates and different confidence intervals.

Output for analysis

Along the second row (output) menu, there are five output selection tabs:

- In the "**Data Summary**" tab panel, three kinds of information are shown:
  - For Taxonomic diversity, basic data information (sample size, observed species richness and estimated sample coverage) and the first 10 abundance (or incidence) frequency counts are shown for the reference sample.
  - For Phylogenetic diversity, basic data information (sample size, observed species richness and observed total branch length) and the singletons/doubletons in the sample branch abundance (or incidence) frequency counts are shown for the reference sample.
  - For Functional diversity, basic data information (sample size, observed species richness and estimated sample coverage) are shown for the reference sample. Besides, the first ten speceis abundance frequency counts in the functionally indistinct set at the threshold distinctiveness and the threshold levels are shown when threshold type is "under specified threshold"; or the minimum, mean and maximum distance between two different species are shown when threshold type is "AUC".

- In the "**Rarefaction and Extrapolation**" tab panel, various estimates for interpolated or extrapolated samples and their confidence intervals are displayed. These results can be downloaded by clicking "Download" at the bottom of the displayed figure. Furthermore, three types of sampling curves are shown:
  (1) Sample-size-based rarefaction and extrapolation sampling curve.
  (2) Sample completeness curve.
  (3) Coverage-based rarefaction and extrapolation sampling curve.
  These figures can be downloaded by clicking "Download" at the bottom of the displayed figures.

- In the "**Asymptotic and Empirical Analysis**" tab panel, the estimated asymptotic and empirical (observed) diversity profiles for diversity order are shown. Below the table, all numerical values of diversity are tabulated along with confidence intervals. Furthermore, the sampling curve is shown:
  - ✧ Asymptotic and empirical diversity profiles. (q profile/ time profile (Only for PD)/ tau profile (Only for FD with threshold type is under specified thresholds)).

  These figures can be downloaded by clicking "Download" at the bottom of the displayed figures.

- In the "**Introduction**" panel, users can view a brief introduction to iNEXT.3D Online and a summary of the running procedures.

- In the "**User Guide**" panel, a link will direct users to this user guide.

Example of Taxonomic diversity

Example 1: Run the demo abundance data (Brazil rainforest) for interpolation/extrapolation and Asymptotic analysis of Taxonomic diversity.

Step 1. Select Taxonomic diversity from the top menu.

Step 2. "Data Setting" (on the left hand side of the window)

(2a) Check the Demo data radio button to load demo data.

(2b) Select data type Abundance data.

(2c) There are three demo data sets for abundance data; select Brazil rainforest.

(2d) There are two assemblages (Edge and Interior); you can select one set or multiple sets for comparison. Here we select all.

Step 3. Press the Run! button to get the output.

Step 4. "General Setting": Use all default settings TD except the Number of bootstrap (adjust to ten); no selection is needed.

Step 5. Press the Run! button to get the output.

OUTPUT

In the "Data Summary" panel, the following statistics are displayed:

	n	S.obs	SC	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
Edge	1794	319	0.9387	110	48	38	28	13	11	12	5	4	6
Interior	2074	356	0.9407	123	48	41	32	19	17	6	7	6	7

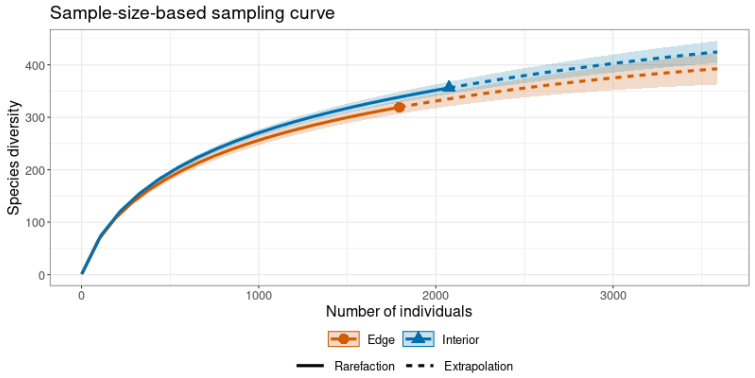
Notes

n = number of observed individuals in the reference sample (sample size).
S.obs = number of observed species in the reference sample.
SC = estimator of the sample coverage of the reference sample.
f1-f10 = the first ten species abundance frequency counts in the sample.

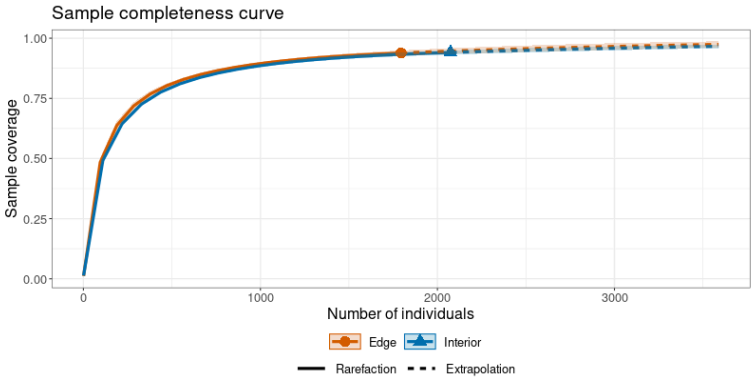
In the “Rarefaction and Extrapolation” tab panel, you can select (1) size-based output, or (2) coverage-based output.

The size-based output is shown below. In the edge assemblage, by default, 40 equally spaced knots (sample sizes) between 1 and 3588 (= 2 x 1794, double the reference sample size) are selected. Diversity estimates and related statistics are computed for these 40 knots (corresponding to sample sizes m = 1, 95, 189,…, 1794…, 3588), and the reference sample is located at the mid-point of the selected knots. The output shows the name of Assemblage, the sample size (m, i.e., each of the 40 knots), the method (Rarefaction, Observed, or Extrapolation, depending on whether the size m is less than, equal to, or greater than the reference sample size), the diversity order (Order.q), the diversity estimate of order q(qD), the 95% lower and upper confidence limits of diversity (qD.LCL, qD.UCL), the sample coverage estimate (SC) and the 95% lower and upper confidence limits of sample coverage (SC.LCL, SC.UCL). These sample coverage estimates with confidence intervals are used for plotting the sample completeness curve.

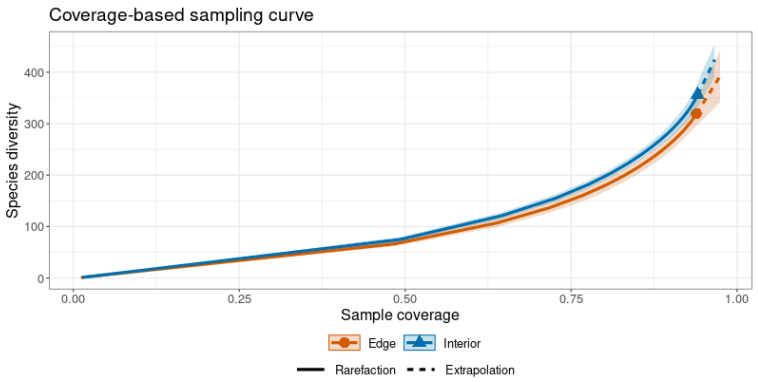
(1) Sample-size-based rarefaction and extrapolation sampling curve



(2) Sample completeness curve



(3) Coverage-based rarefaction and extrapolation sampling curve



- Size-based output:

Assemblage	m	Method	Order.q	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
Edge	1	Rarefaction	0	1.00	1.00	1.00	0.012	0.011	0.013
Edge	95	Rarefaction	0	66.31	65.06	67.55	0.484	0.465	0.503
Edge	189	Rarefaction	0	106.74	103.80	109.68	0.638	0.619	0.656
Edge	284	Rarefaction	0	137.03	132.55	141.51	0.718	0.703	0.734
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Edge	1699	Rarefaction	0	313.04	301.01	325.06	0.936	0.926	0.946
Edge	1793	Rarefaction	0	318.94	306.50	331.50	0.939	0.928	0.949
Interior	1	Rarefaction	0	1.00	1.00	1.00	0.013	0.011	0.016
Interior	110	Rarefaction	0	74.82	73.15	76.48	0.492	0.478	0.506
Interior	219	Rarefaction	0	120.91	117.97	123.85	0.645	0.634	0.656
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

Assemblage	m	Method	Order.q	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
Interior	2074	Observed	0	356.00	344.34	367.66	0.941	0.932	0.950
Interior	2075	Extrapolation	0	356.06	344.40	367.72	0.941	0.932	0.950
Interior	2154	Extrapolation	0	360.67	348.57	372.78	0.942	0.934	0.951

The coverage-based output is shown below. The output includes the Assemblage name, the standardized sample coverage (SC), the corresponding sample size for the standardized coverage (m, i.e., each of the 40 knots), the method (Rarefaction, Observed, or Extrapolation, depending on whether the sample coverage SC is less than, equal to, or greater than the reference sample coverage), the diversity order (Order.q), the diversity estimate of order q (qD), and the 95% lower and upper confidence limits of diversity (qD.LCL, qD.UCL). These diversity estimates and confidence intervals are used for plotting the coverage-based R/E curves.

- Coverage-based output:

Assemblage	SC	m	Method	Order.q	qD	qD.LCL	qD.UCL
Edge	0.012	1	Rarefaction	0	1.00	1.00	1.00
Edge	0.484	95	Rarefaction	0	66.31	61.62	70.99
Edge	0638	189	Rarefaction	0	106.74	98.59	114.90
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Edge	0.936	1699	Rarefaction	0	313.04	291.31	334.77
Edge	0.939	1792	Rarefaction	0	318.90	295.33	342.47
Interior	0.013	1	Rarefaction	0	1.00	0.90	1.10
Interior	0.492	110	Rarefaction	0	74.82	69.95	79.68
Interior	0.645	219	Rarefaction	0	120.91	114.46	127.36
Interior	0.941	2074	Observed	0	356.00	332.45	379.55
Interior	0.941	2075	Extrapolation	0	356.06	332.50	379.62

Assemblage	SC	m	Method	Order.q	qD	qD.LCL	qD.UCL
Interior	0.965	3508	Extrapolation	0	421.71	389.89	453.52
Interior	0.966	3588	Extrapolation	0	424.43	392.25	456.61

Notes

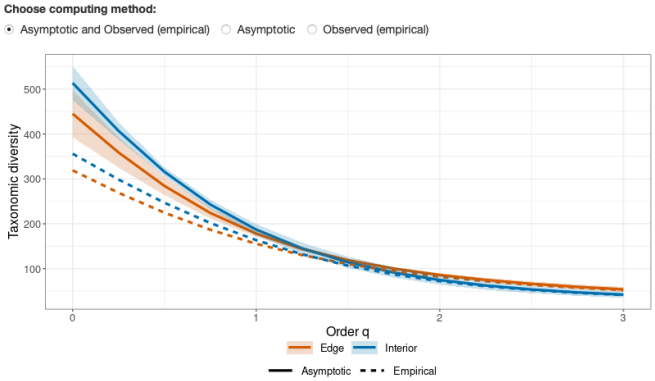
- Assemblage = the datasets you selected in the 'Data setting' on the left hand side of the screen.
- m = sample size for which diversity estimates of order q are computed; by default setting (in the left hand side of the screen), m represents the sample size for each of the 40 knots between 1 and the default endpoint (double the reference sample size). On the "General Setting", you can also either specify the endpoint and knots or specify the sample sizes for which you like to calculate diversity estimates.
- Method = Rarefaction, Observed, or Extrapolation, depending on whether the size m is less than, equal to, or greater than the reference sample size.
- Order.q = the diversity order of q you selected in the "General Setting" on the left hand side of the screen.
- qD = the estimated diversity of order q for a sample of size m.
- qD.LCL, qD.UCL = the bootstrap lower and upper confidence limits for the diversity of order q at the specified level in the setting (with a default value of 0.95).
- SC = the estimated sample coverage for a sample of size m.
- SC.LCL, SC.UCL = the bootstrap lower and upper confidence limits for the expected sample coverage at the specified level in the setting (with a default value of 0.95).

Note that the confidence intervals for diversity (qD) in the size- and coverage-based outputs are different. In the iNEXT.3D method, the confidence intervals of any standardized diversity are obtained by a bootstrap method. In the size-based standardization, the sample size is fixed in each regenerated bootstrap sample. In the coverage-based standardization, for a given standardized coverage value, the corresponding size needed to attain the same level of coverage may vary with regenerated bootstrap samples. Thus, the sampling uncertainty is greater in the coverage-based standardization and the resulting confidence interval is wider than that in the corresponding size-based standardization.

For example, if the size for a future survey will be fixed at a sample size of 1699 in the edge assemblage, we can obtain a 95% confidence interval of (301.01, 325.06) for the expected diversity (q = 0) based on the size-based output. However, if the coverage of a survey is fixed at the level of 0.936, the size needed for the current data is 1699, but the size needed for a regenerated bootstrap sample may be different from 1699; the coverage-based output shows a CI of (291.31, 334.77), which is wider than the former one based on a size of 1794. Because we use a random bootstrapping regeneration process with 50 replications to obtain each CI, the output for qD.LCL and qD.UCL may vary slightly each time you enter the same data.

In the “Asymptotic Analysis” tab panel, you can obtain all numerical values of asymptotic diversity and empirical diversity.

Asymptotic and empirical diversity profiles.



Order.q	qD	s.e.	qD.LCL	qD.UCL	Assemblage	Method
0.00	444.97	26.54	392.96	496.98	Edge	Asymptotic
0.25	358.83	17.23	325.05	392.61	Edge	Asymptotic
.	.	.	.	.	.	.
3.00	53.85	2.87	48.22	59.47	Edge	Asymptotic
0.00	513.52	19.39	475.52	551.52	Interior	Asymptotic
.	.	.	.	.	.	.
3.0	42.17	3.55	35.21	49.12	Interior	Asymptotic
0.00	319.00	8.98	301.40	336.60	Edge	Empirical
0.25	269.06	8.13	253.12	185.00	Edge	Empirical
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
2.75	46.24	2.61	41.14	51.35	Interior	Empirical

3.00	41.48	2.45	36.69	46.28	Interior	Empirical

Notes

- Order.q = the diversity order.
- qD = the asymptotic/empirical diversity estimates of order q.
- s.e. = the bootstrap standard error of the estimated asymptotic/empirical diversity of order q.
- qD.LCL, qD.UCL = the bootstrap lower and upper confidence limits for the diversity of order q at the specified level in the setting (with a default value of 0.95).
- Assemblage = the datasets you selected in the 'Data setting' on the left hand side of the screen.
- Method = computing type. 'Asymptotic' or 'Empirical'.

Example of Phylogenetic diversity

Example 2: Run the demo abundance data (Brazil rainforest) for interpolation/extrapolation and Asymptotic analysis of Phylogenetic diversity.

Step 1. Select Phylogenetic diversity from the top menu.

Step 2. "Data Setting" (on the left hand side of the window)

(2a) Check the Demo data radio button to load demo data.

(2b) Select data type Abundance data.

(2c) There are three demo data sets for abundance data; select Brazil rainforest.

(2d) There are two assemblages (Edge and Interior); you can select one set or multiple sets for comparison. Here we select all.

Step 3. Press the Run! button to get the output.

Step 4. "General Setting": Use all default settings PD except the Number of bootstrap (adjust to ten); no selection is needed.

Step 5. Press the Run! button to get the output.

OUTPUT

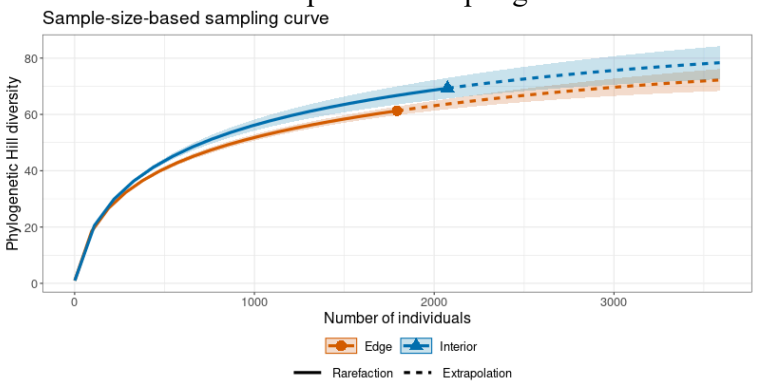
In the “Data Summary” panel, the following statistics are displayed:

	n	S.obs	PD.obs	f1*	f2*	g1	g2	Reftime
Edge	1794	319	24516	110	52	6578	2885	400
Interior	2074	356	27727	123	56	7065	3656	400

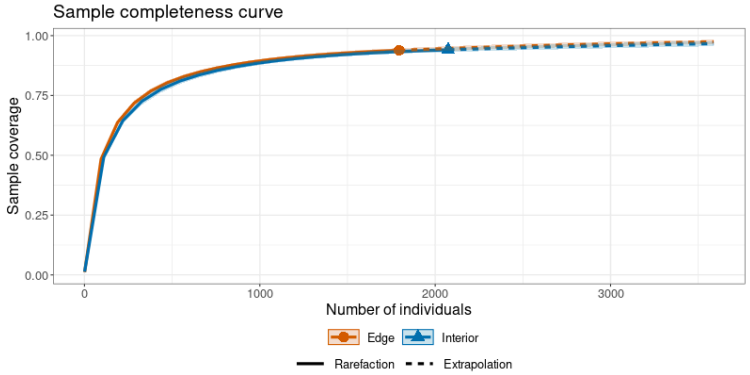
Notes
n = number of observed individuals in the reference sample (sample size).
S.obs = number of observed species in the reference sample.
PD.obs = the observed total branch length in the reference sample.
f1*-f2* = the singletons/doubletons in the sample branch abundance.
g1-g2 = the total branch length of those singletons/doubletons in the sample branch abundance set of the observed tree.
Reftime = reference time for phylogenetic diversity (the age of the root of phylogenetic tree).

In the “Rarefaction and Extrapolation” tab panel, you can select (1) size-based output, or (2) coverage-based output.

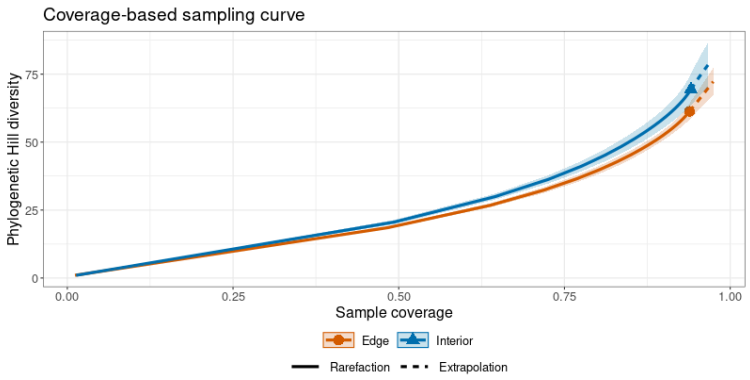
(1) Sample-size-based rarefaction and extrapolation sampling curve



(2) Sample completeness curve



(3) Coverage-based rarefaction and extrapolation sampling curve



- Size-based output:

Assemblage	m	Method	Order.q	qPD	qPD. LCL	qPD. UCL	SC	SC.LCL	SC.UCL	Reftime	Type
Edge	1	Rarefaction	0	1.00	0.99	1.01	0.012	0.011	0.013	400	meanPD
Edge	95	Rarefaction	0	18.55	18.22	18.87	0.484	0.476	0.493	400	meanPD
.	.	.	.	.	.	.	.	.	.		
.	.	.	.	.	.	.	.	.	.		
Interior	1	Rarefaction	0	1.00	0.98	1.02	0.013	0.012	0.015	400	meanPD
Interior	110	Rarefaction	0	20.55	20.12	20.98	0.492	0.478	0.506	400	meanPD

- Coverage-based output:

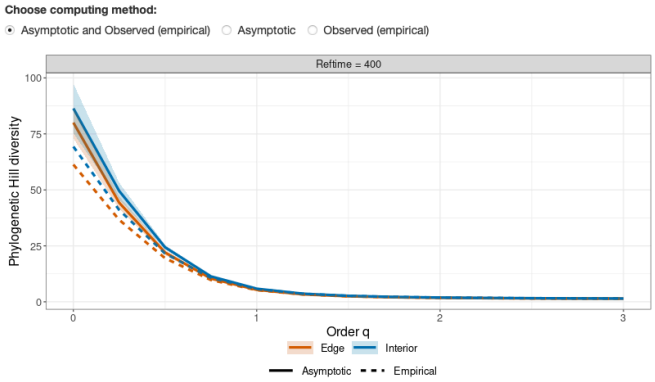
Assemblage	SC	m	Method	Order.q	qPD	qPD.LCL	qPD.UCL	Reftime	Type
Edge	0.012	1	Rarefaction	0	1.00	0.99	1.01	400	meanPD
Edge	0.484	95	Rarefaction	0	18.55	18.06	19.04	400	meanPD
.	.	.	.	.	.	.	.		
.	.	.	.	.	.	.	.		
Interior	0.013	1	Rarefaction	0	1.00	0.98	1.02	400	meanPD
Interior	0.492	110	Rarefaction	0	20.55	19.75	21.35	400	meanPD

Notes

- Assemblage = the datasets you selected in the 'Data setting' on the left hand side of the screen.
- m = sample size for which diversity estimates of order q are computed; by default setting (in the left hand side of the screen), m represents the sample size for each of the 40 knots between 1 and the default endpoint (double the reference sample size). On the “General Setting”, you can also either specify the endpoint and knots or specify the sample sizes for which you like to calculate diversity estimates.
- Method = Rarefaction, Observed, or Extrapolation, depending on whether the size m is less than, equal to, or greater than the reference sample size.
- Order.q = the diversity order of q you selected in the “General Setting” on the left hand side of the screen.
- qPD = the estimated diversity of order q for a sample of size m.
- qPD.LCL, qPD.UCL = the bootstrap lower and upper confidence limits for the diversity of order q at the specified level in the setting (with a default value of 0.95).
- SC = the estimated sample coverage for a sample of size m.
- SC.LCL, SC.UCL = the bootstrap lower and upper confidence limits for the expected sample coverage at the specified level in the setting (with a default value of 0.95).
- Reftime = reference time for phylogenetic diversity (the age of the root of phylogenetic tree).
- Type = the phylogenetic type (‘PD’ or ‘meanPD’).

In the “Asymptotic Analysis” tab panel, you can obtain all numerical values of asymptotic diversity and empirical diversity.

Asymptotic and empirical diversity profiles.



Order.q	qPD	s.e.	qPD.LCL	qPD.UCL	Assemblage	Method	Reftime	Type
0.00	80.03	3.58	73.01	87.04	Edge	Asymptotic	400	meanPD
.	.	.	.	.	.	.	.	.
3.00	1.41	0.01	1.38	1.44	Edge	Asymptotic	400	meanPD

0.00	86.38	5.65	75.31	97.44	Interior	Asymptotic	400	meanPD
.	.	.	.	.	.	.	.	.
3.00	1.48	0.01	1.46	1.50	Interior	Asymptotic	400	meanPD
.	.	.	.	.	.	.	.	.

Notes

- Order.q = the diversity order.
- qPD = the asymptotic/empirical phylogenetic diversity estimates of order q.
- s.e. = the bootstrap standard error of the estimated asymptotic/empirical phylogenetic diversity of order q.
- qPD.LCL, qPD.UCL = the bootstrap lower and upper confidence limits for the phylogenetic diversity of order q at the specified level in the setting (with a default value of 0.95).
- Assemblage = the datasets you selected in the 'Data setting' on the left hand side of the screen.
- Method = computing type. 'Asymptotic' or 'Empirical'.
- Reftime = reference time for phylogenetic diversity (the age of the root of phylogenetic tree).
- Type = the phylogenetic type (‘PD’ or ‘meanPD’).

References

The following papers for pertinent background on rarefaction/extrapolation and related statistical analyses. These papers can be directly downloaded from Anne Chao’s website.

Chao, A., Chiu, C.-H., Villéger, S., Sun, I.-F., Thorn, S., Lin, Y.-C., Chiang, J. M. and Sherwin, W. B. (2019). An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. Ecological Monographs, 89, e01343. 10.1002/ecm.1343.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Chao, A., Henderson, P. A., Chiu, C.-H., Moyes, F., Hu, K.-H., Dornelas, M and Magurran, A. E. (2021). Measuring temporal change in alpha diversity: a framework integrating taxonomic, phylogenetic and functional diversity and the iNEXT.3D standardization. *Methods in Ecology and Evolution*, 12, 1926-1940.

Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, **93**, 2533–2547.

Chao, A. and Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution*, **6**, 873–882.