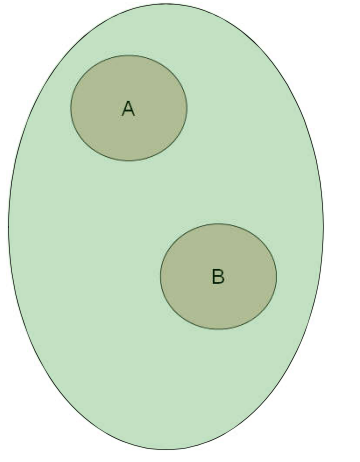


Diversity Metrics

- Alpha diversity (α)
 - 20 species in A
 - 15 species in B
- Beta diversity (β)
 - 8 species in both A and B
- Gamma diversity (γ)
 - 100 species in the region as a whole



- Typically, alpha diversity metrics are summaries by site
 - Species richness
 - Species diversity

sample	Cyprinus	Composita	Ctenophora	Cyprinella	Cyprinella	Notemigona	Notropis	Notropis	Pimephala	Pimephala	Carpodacus	Icthyobry	Minytrema
WIM7	0	1	0	5	0	1	0	0	0	0	0	0	0
WIM8a	0	97	0	3	0	0	13	0	0	0	0	0	0
WIM8b	0	36	0	1	0	0	2	0	0	0	0	0	0
WIM10	0	50	0	0	0	0	75	0	0	0	0	0	0
WIM11	0	5	0	0	0	0	63	0	0	0	0	0	0
WIM12	0	16	0	0	35	0	80	0	0	3	0	0	0
WIM736	0	0	0	27	0	0	0	0	30	0	0	0	0
WIM737	0	48	0	18	0	0	0	0	0	0	0	0	0
WIM738	0	39	0	1	0	2	3	0	7	0	0	0	0
WIM759	0	6	0	3	0	0	31	0	1	3	0	0	0
WIM749	0	59	0	58	6	39	4	0	0	3	0	0	0
WIM750	0	29	0	198	85	2	10	0	0	9	0	0	0
WIM842	0	0	0	9	0	0	0	0	29	0	0	0	0
WIM843	0	48	0	16	0	0	0	0	0	0	0	0	0
WIM844	0	4	0	4	0	0	0	0	0	0	0	0	0
WIM845	0	13	0	5	0	0	68	0	0	1	0	0	0
WIM846	0	70	0	31	3	8	31	0	0	3	0	0	0
WIM847	0	21	0	35	51	0	16	0	0	2	0	0	1
WIM1820	0	0	0	4	0	2	0	0	0	0	0	0	0
WIM1821	0	195	0	1	0	3	20	0	0	0	0	0	0
WIM1823	0	137	0	0	0	1	5	0	2	0	0	0	0
WIM1824	0	9	0	15	0	2	95	0	1	2	1	0	1
WIM85-5	0	243	0	101	12	3	197	2	0	3	0	0	0
WIM1825	0	76	0	8	30	2	107	2	2	2	0	0	2
RCC53	0	0	0	2	0	0	0	0	0	0	0	0	0
RCC954	0	43	0	1	1	1	15	0	0	0	0	0	0

Diversity Indices

- Species richness (S)
- Shannon Index (H')
- Simpson's Index (D)
 - Where n_i = individuals of species i
 - S = number of species
 - N = total number of individuals
 - P_i = relative abundance of each species
- Use raw abundance data

$$H' = - \sum_{i=1}^S p_i \ln p_i$$

$$D = 1 - \sum_{i=1}^S p_i^2$$

Diversity function (vegan package)

- `diversity(community_matrix, index, MARGIN)`
 - Default margin is 1, samples in rows, species in columns
 - Index can be "shannon", "simpson", or "invsimpson"
 - Returns a vector of diversity measures

- `diversity(Brier_Ck, index="shannon", MARGIN=1)`

- Shannon's index by hand:

$$H' = - \sum_{i=1}^S p_i \ln p_i$$

- `Brier_p <- decostand(Brier_Ck, MARGIN=1, method="total")`
- `Brier_p_lnp <- Brier_p * log(Brier_p)`
- `apply(Brier_p_lnp, 1, sum, na.rm=T) * -1`

- `diversity(Brier_Ck,index="simpson",MARGIN=1)`

- Simpson's index by hand:

$$D = 1 - \sum_{i=1}^s p_i^2$$

- `Brier_p<-decostand(Brier_Ck, MARGIN=1, method="total")`
- `Brier_p2<-Brier_p^2`
- `1-apply(Brier_p2,1,sum, na.rm=T)`

- Species richness

- `Brier_pa<-decostand(Brier_Ck,method="pa")`
- `Brier_S<-apply(Brier_pa,1,sum)`

- Function `specnumber(community_matrix)`

- `specnumber(Brier_Ck)`

Measures of evenness

$$J = \frac{H}{H_{\max}} = \frac{-\sum_{i=1}^S p_i \ln p_i}{\ln S}$$

- Measure of the evenness of species abundances within the community.
- Maximum = H_{\max} if all species abundances the same
- Equitability is a percentage of the max

- `Brier_shannon<-diversity(Brier_Ck,index="shannon",MARGIN=1)`
- `Brier_S<-specnumber(Brier_Ck)`
- `Brier_J<-Brier_shannon/log(Brier_S)`

R functions

- Function `diversityresult`

- Calculates diversity indices (richness, abundance, Shannon, Simpson) and evenness indices.
- `diversityresult(matrix,index,method)`
 - Index – type of diversity index to use
 - Method – calculate for all samples pooled, mean of all samples, individual samples, or jackknife
 - Factor – factor dividing samples into groups for comparison
- Default is to pool all samples together
- Will also do bootstrap and jackknife estimates of diversity
- Will also work with factors that divide samples into groups

- Shannon diversity by site

- `diversityresult(Brier_Ck,index="Shannon",method="each site")`

- Shannon diversity pooling sites

- `diversityresult(Brier_Ck,index="Shannon",method="pooled")`

- Mean diversity across sites

- `diversityresult(Brier_Ck,index="Shannon",method="mean")`

- Same as taking the mean from our earlier vector

- `mean(Brier_Shannon)`

- Gamma diversity (chao) estimator

- `diversityresult(Brier_Ck,index="chao",method="pooled")`

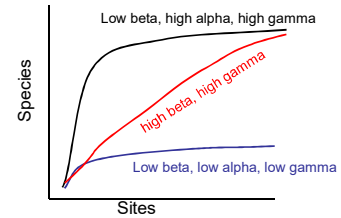
Comparing Diversity with Factors

- Factors may be used to describe sample differences
- Shannons for sites with factor "year" level "a"
 - `diversityresult(Brier_Ck.y=Brier_environmental,index="Shannon",method="eachsite",factor="year",level="a")`
- Function `diversitycomp` will get diversity for all levels of a factor
 - Note – specify factor as factor1=
 - `diversitycomp(Brier_Ck.y=Brier_environmental,index="Shannon",method="pooled",factor1="year")`
 - Specify factor1 and factor2
 - `diversitycomp(Brier_Ck.y=Brier_environmental,index="Shannon",method="pooled",factor1="year",factor2="location")`

Sample	Year	Location
W00001	0	0
W00002	0	0
W00003	0	0
W00004	0	0
W00005	0	0
W00006	0	0
W00007	0	0
W00008	0	0
W00009	0	0
W00010	0	0
W00011	0	0
W00012	0	0
W00013	0	0
W00014	0	0
W00015	0	0
W00016	0	0
W00017	0	0
W00018	0	0
W00019	0	0
W00020	0	0
W00021	0	0
W00022	0	0
W00023	0	0
W00024	0	0
W00025	0	0
W00026	0	0
W00027	0	0
W00028	0	0
W00029	0	0
W00030	0	0
W00031	0	0
W00032	0	0
W00033	0	0
W00034	0	0
W00035	0	0
W00036	0	0
W00037	0	0
W00038	0	0
W00039	0	0
W00040	0	0
W00041	0	0
W00042	0	0
W00043	0	0
W00044	0	0
W00045	0	0
W00046	0	0
W00047	0	0
W00048	0	0
W00049	0	0
W00050	0	0
W00051	0	0
W00052	0	0
W00053	0	0
W00054	0	0
W00055	0	0
W00056	0	0
W00057	0	0
W00058	0	0
W00059	0	0
W00060	0	0
W00061	0	0
W00062	0	0
W00063	0	0
W00064	0	0
W00065	0	0
W00066	0	0
W00067	0	0
W00068	0	0
W00069	0	0
W00070	0	0
W00071	0	0
W00072	0	0
W00073	0	0
W00074	0	0
W00075	0	0
W00076	0	0
W00077	0	0
W00078	0	0
W00079	0	0
W00080	0	0
W00081	0	0
W00082	0	0
W00083	0	0
W00084	0	0
W00085	0	0
W00086	0	0
W00087	0	0
W00088	0	0
W00089	0	0
W00090	0	0
W00091	0	0
W00092	0	0
W00093	0	0
W00094	0	0
W00095	0	0
W00096	0	0
W00097	0	0
W00098	0	0
W00099	0	0
W00100	0	0

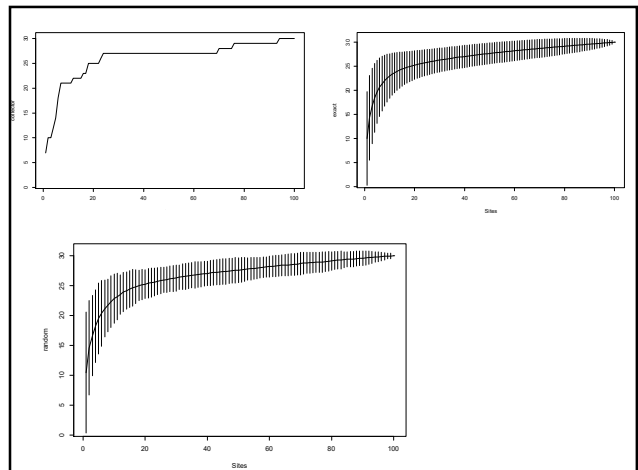
Species Accumulation and Rarefaction

- Curves display the rate of new species addition with additional sampling.
- Alpha, beta and gamma diversity
- Diverse ecosystems (high gamma and high beta diversity) will display rapid increase to higher plateau. Less diverse: slow increase to lower plateau.



Species Accumulation and Rarefaction

- Function: `specaccum`
 - `specaccum(matrix,method)`
- Methods:
 - Accumulate species by:
 - Adding samples in the order collected ("collector")
 - Adding samples in random order ("random")
 - Calculating expected (mean) number for each sample size ("exact")
 - Follow methods of Coleman 1982 ("coleman")
 - Permuting individuals instead of samples ("rarefaction")
- "collector" does not provide error estimates.
- "exact" provides a mean and standard deviation for observed data
- "random" provides mean and standard deviation for permuted data



Randomizations, bootstraps, jackknifes, Monte Carlo, MCMC

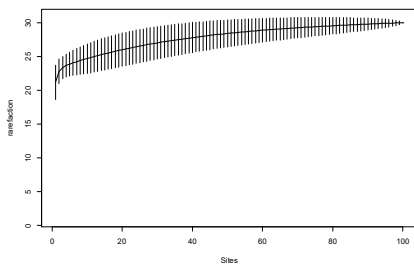
- Function `sample` can be used to randomly sample any data
 - `sample(thing to sample, number of samples, replace)`
 - **Bootstrapping** – estimating some parameter by randomly sampling with replacement
 - See also functions in `permut` package
- In our sample community dataset:
 - `mean(Brier_Ck_matrix) = 14.261`
 - `sample(Brier_Ck_matrix, 100, replace=TRUE) # pick 100 values`
 - `mean(sample(Brier_Ck_matrix, 10000, replace=TRUE)) # 10000`
 - `sample(Brier_Ck_matrix, 3000, replace=FALSE) # sample all without replacement`

Species Accumulation and Rarefaction

- **Rarefaction** – most commonly used technique for estimating diversity
- Pool **individuals** within a dataset and resample individuals and not individual cells within the matrix.
- Use the same `specaccum` function (`method="rarefaction"`).
- Rarefaction requires integer (count) data, untransformed. Other methods will work with log or proportional transformed data.
- **Number of individuals sampled is set as the mean number of individuals per sample in the dataset.**

Species Accumulation and Rarefaction

- Rarefaction curve for Brier Creek community data using `specaccum, method="rarefaction"` which set the number of individuals per sample to 428 (mean number per sample).



Species Accumulation and Rarefaction

- The `specaccum` function uses the function `rrarefy` to do the resampling. You can use `rrarefy` directly if you want more control.
- Function `rrarefy(community, sample)`
 - Provide the community matrix and how many individuals to sample and it will return a permuted sample.
- To duplicate what `specaccum` did:
 - Pool all individuals into one vector
 - `pooled<-apply(Brier_Ck, 2, sum)`
 - Rarefy all samples by randomly selecting 428 individuals
 - `rrarefy(pooled, sample=428)`
 - Repeat...

Species Accumulation and Rarefaction

- One common use for rarefaction is to control for sampling effort (or efficiency) in estimating diversity.
- In the sample community, the number of individuals per sample ranges from 2 to 2554.
- Diversity and the number of individuals is clearly correlated, what is sample diversity if we control for this?

```
rarefied_100<-rarefy(Brier Ck,100, se=T)
```

```
t(rarefied_100)
```

	S	se
WJM7	7	0.00E+00
WJM8a	7.809496	4.17E-01
WJM8b	5	0.00E+00
WJM10	7.303897	7.14E-01
WJM11	9.727267	4.93E-01
WJM12	9.764172	1.25E+00
WJM736	7.665996	5.25E-01
WJM737	8.584473	5.52E-01
WJM738	10.32027	7.08E-01
WJM739	12.07872	8.08E-01

...

Species Accumulation and Rarefaction

- Function `rarefy` will select a constant number of individuals from each sample, return a matrix of the same dimensions
 - `rarefy(community,sample,se)`
 - Provide community data, number to subsample and whether or not you want standard error calculated
- How many individuals should you sample?
 - The number should be less than the number in your least abundant sample
- Note that sampling this number results in no error for the sample with the fewest individuals (all individuals are sampled in each permutation)
- This is where you would consider eliminating samples with fewer individuals