

Webinar Series

Data Use Skills

Featuring Data from Natural History Collections

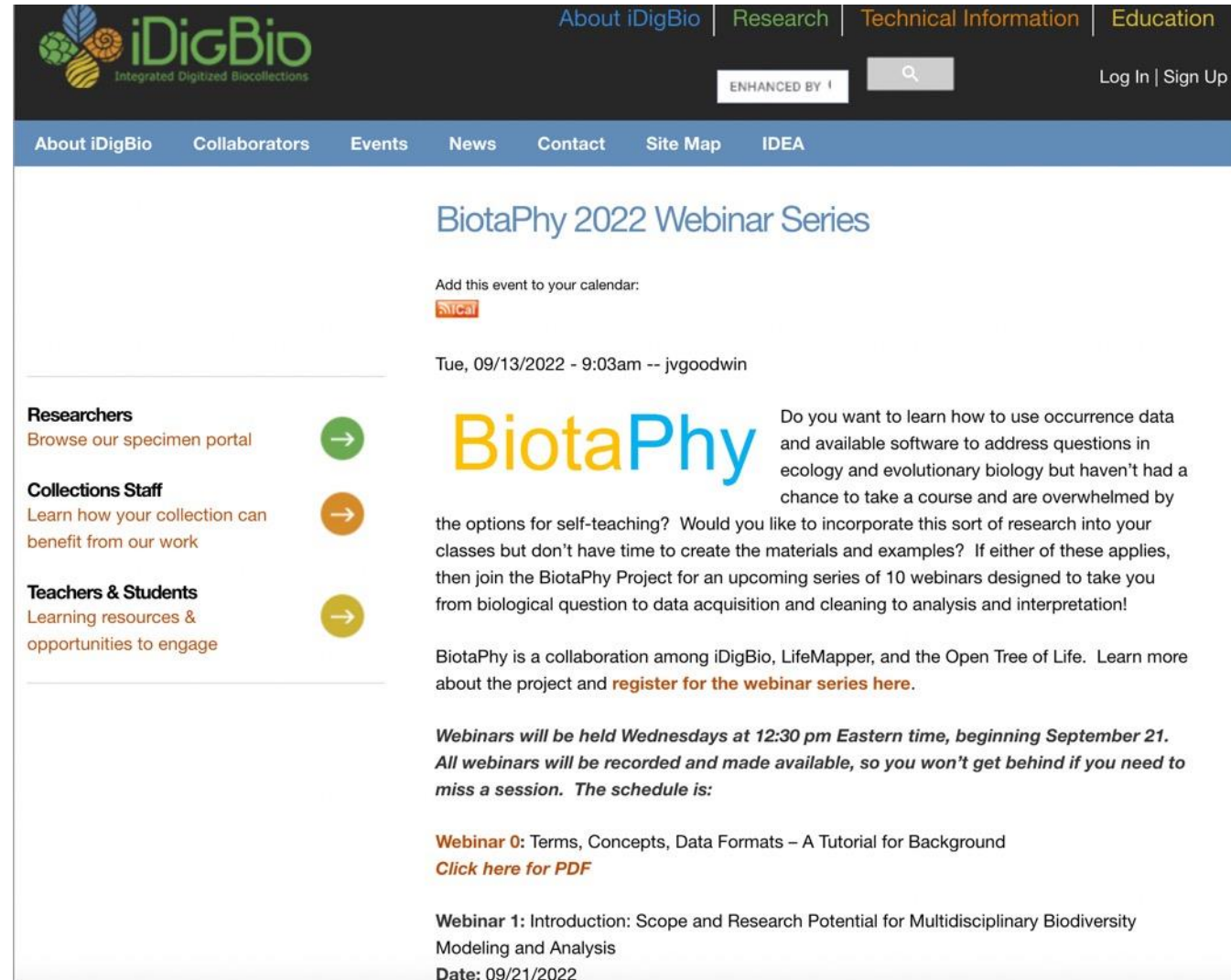
September 21-November 30, 2022

<https://www.idigbio.org/content/biotaphy-2022-webinar-series>

iDigBio:

<https://www.idigbio.org/content/biotaphy-2022-webinar-series>

[iDigBio.org](https://www.idigbio.org)



The screenshot shows the iDigBio website with a navigation bar at the top containing links for About iDigBio, Research, Technical Information, and Education. Below the navigation bar is a search bar and a "Log In | Sign Up" link. The main content area features a sidebar on the left with three sections: "Researchers" (Browse our specimen portal), "Collections Staff" (Learn how your collection can benefit from our work), and "Teachers & Students" (Learning resources & opportunities to engage). The main content area is titled "BiotaPhy 2022 Webinar Series" and includes a calendar link, the date "Tue, 09/13/2022 - 9:03am -- jvgoodwin", and a large "BiotaPhy" logo. The text describes the webinar series as a collaboration among iDigBio, LifeMapper, and the Open Tree of Life, designed to help researchers and students learn how to use occurrence data and available software to address questions in ecology and evolutionary biology. It mentions that the webinars will be held on Wednesdays at 12:30 pm Eastern time, beginning September 21, and that all webinars will be recorded and made available. The schedule for the first two webinars is provided: Webinar 0 (Terms, Concepts, Data Formats – A Tutorial for Background) on 09/21/2022 and Webinar 1 (Introduction: Scope and Research Potential for Multidisciplinary Biodiversity Modeling and Analysis) on 09/21/2022.

About iDigBio Collaborators Events News Contact Site Map IDEA

BiotaPhy 2022 Webinar Series

Add this event to your calendar:

[iCal](#)

Tue, 09/13/2022 - 9:03am -- jvgoodwin

BiotaPhy

Do you want to learn how to use occurrence data and available software to address questions in ecology and evolutionary biology but haven't had a chance to take a course and are overwhelmed by the options for self-teaching? Would you like to incorporate this sort of research into your classes but don't have time to create the materials and examples? If either of these applies, then join the BiotaPhy Project for an upcoming series of 10 webinars designed to take you from biological question to data acquisition and cleaning to analysis and interpretation!

BiotaPhy is a collaboration among iDigBio, LifeMapper, and the Open Tree of Life. Learn more about the project and [register for the webinar series here](#).

Webinars will be held Wednesdays at 12:30 pm Eastern time, beginning September 21. All webinars will be recorded and made available, so you won't get behind if you need to miss a session. The schedule is:

Webinar 0: Terms, Concepts, Data Formats – A Tutorial for Background
[Click here for PDF](#)

Webinar 1: Introduction: Scope and Research Potential for Multidisciplinary Biodiversity Modeling and Analysis
Date: 09/21/2022



Thank You

**Maria Cortez
Aimee Stewart**

**Jill Goodwin
Gil Nelson**

Webinar 8

Introducing Presence-Absence Matrices for Large Scale Analyses

**Learn how to calculate basic diversity statistics and how to utilize a Presence -
Absence Matrix (PAM) to optimize
calculations involving multiple taxa**

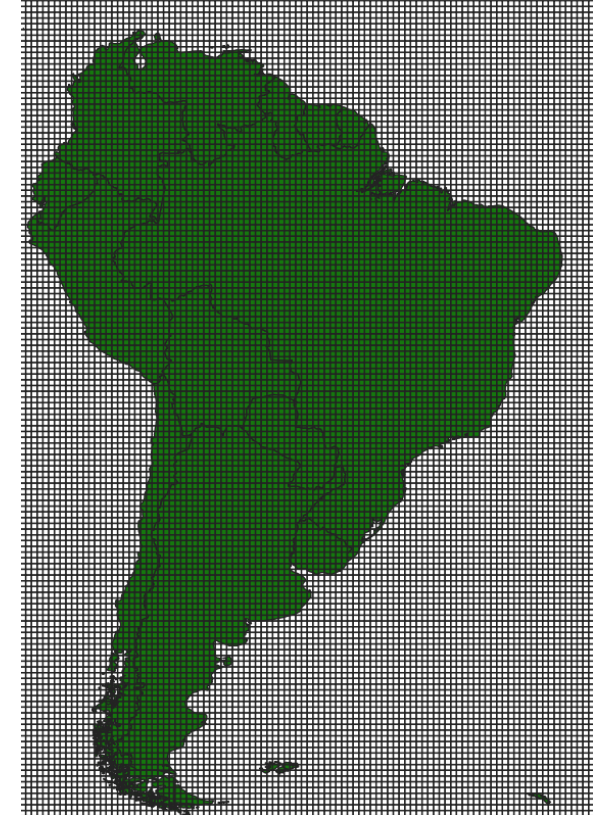
Biological Objectives:

- ✓ **Explore the theory on basic diversity statistics**
 - ✓ **Alpha diversity**
 - ✓ **Beta diversity**
 - ✓ **Gamma diversity**

Technical Objectives:

Build a grid to analyze a multi-species dataset

1. Understand what a grid is
2. Understand the effects of grid scale



Technical Objectives:

Encode Layers into a Presence-Absence matrix
(PAM) for multi-species analysis

1. Learn what a PAM is in a biogeographic context
2. Learn how to create a PAM
3. Understand PAM as a data model

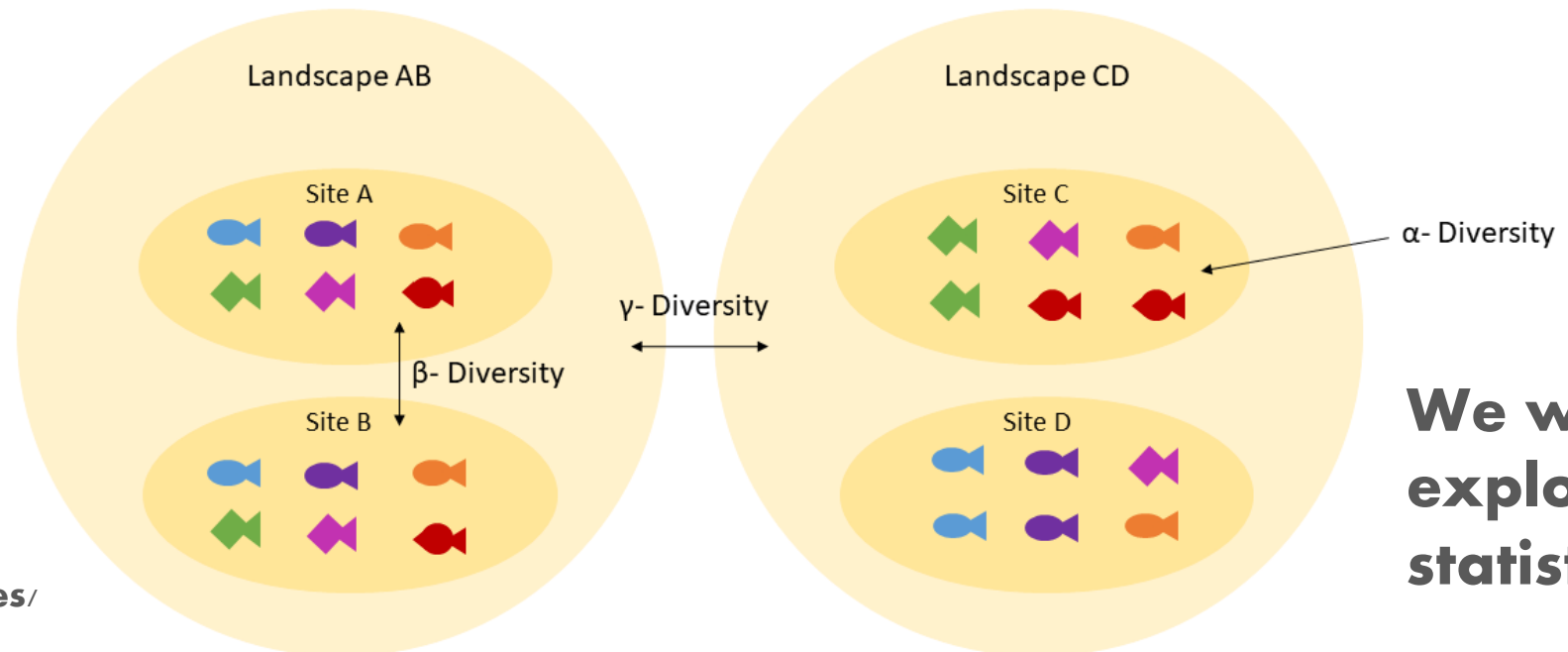
	A	B	C	BNH	BNI
1				Colobanthus apetalus	Colobanthus affinis
3006	3004	144.25	-27.75	0	0
3007	3005	144.75	-27.75	0	0
3008	3006	145.25	-27.75	0	0
3009	3007	145.75	-27.75	0	0
3010	3008	146.25	-27.75	0	0
3011	3009	146.75	-27.75	0	0
3012	3010	147.25	-27.75	0	0
3013	3011	147.75	-27.75	0	0
3014	3012	148.25	-27.75	0	0
3015	3013	148.75	-27.75	0	0
3016	3014	149.25	-27.75	0	0
3017	3015	149.75	-27.75	0	0
3018	3016	150.25	-27.75	0	0
3019	3017	150.75	-27.75	0	0
3020	3018	151.25	-27.75	0	0
3021	3019	151.75	-27.75	0	0
3022	3020	152.25	-27.75	0	0
3023	3021	152.75	-27.75	0	0
3024	3022	153.25	-27.75	0	0

Technical Objectives:

Calculate statistics for a PAM

1. Learn what some of the stats are: alpha, beta and gamma

diversity!



We will explore more statistics later!

1. **Exploring Concepts: Biological background on diversity statistics**
2. **Exploring Concepts: How to ...**
 - a. **build a grid**
 - b. **encode SDM layers**
 - c. **calculate statistics (alpha and beta diversity)**
3. **Exercises: Practicing ...**
 - a. **building a grid**
 - b. **encoding SDM layers**
 - c. **calculating statistics (alpha and beta diversity)**
4. **Session Summary, Q&A and Discussion**

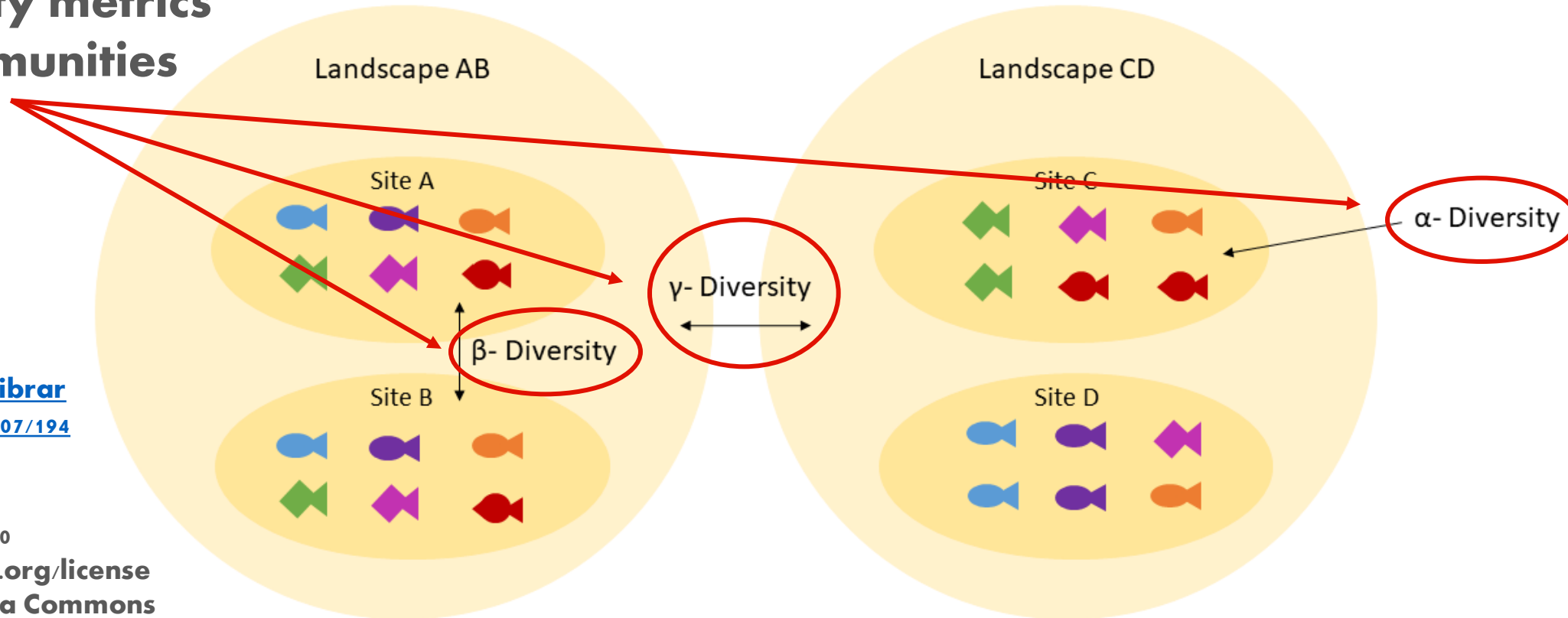
Exploring Concepts: Diversity Statistics

VEGETATION OF THE SISKIYOU MOUNTAINS, OREGON AND CALIFORNIA¹

R. H. WHITTAKER

Biology Department, Brooklyn College, Brooklyn 10, N. Y.

In **1960** Whittaker proposed **3** species-diversity metrics in natural communities



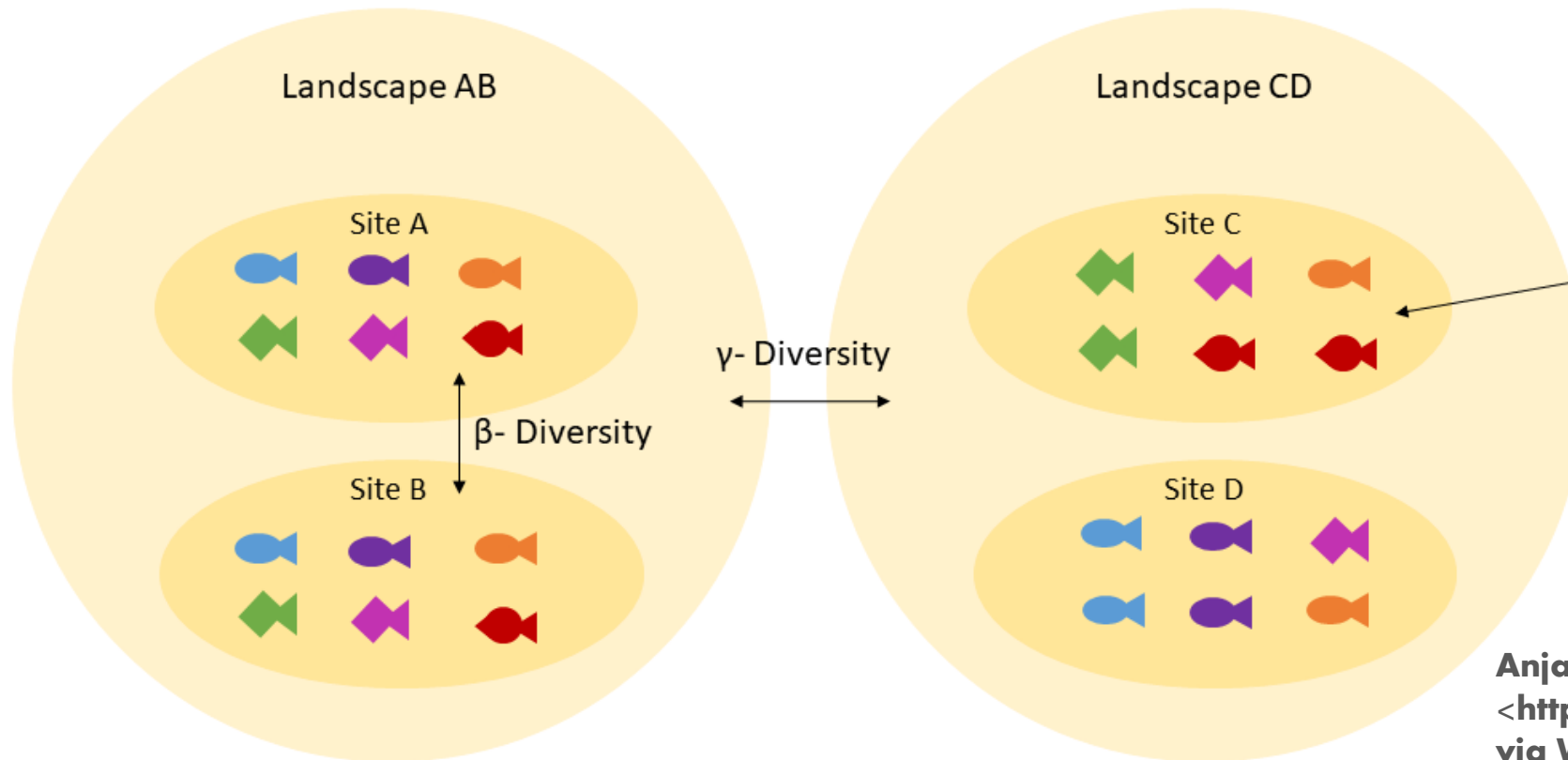
<https://esajournals.onlinelibrary.wiley.com/doi/epdf/10.2307/1943563>

Anja Knaebel, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Exploring Concepts: Diversity Statistics

VEGETATION OF THE SISKIYOU MOUNTAINS, OREGON AND CALIFORNIA¹

R. H. WHITTAKER
Biology Department, Brooklyn College, Brooklyn 10, N. Y.



“The richness in species of a particular stand or community, or a given stratum or group of organisms in a stand.”

α -Diversity

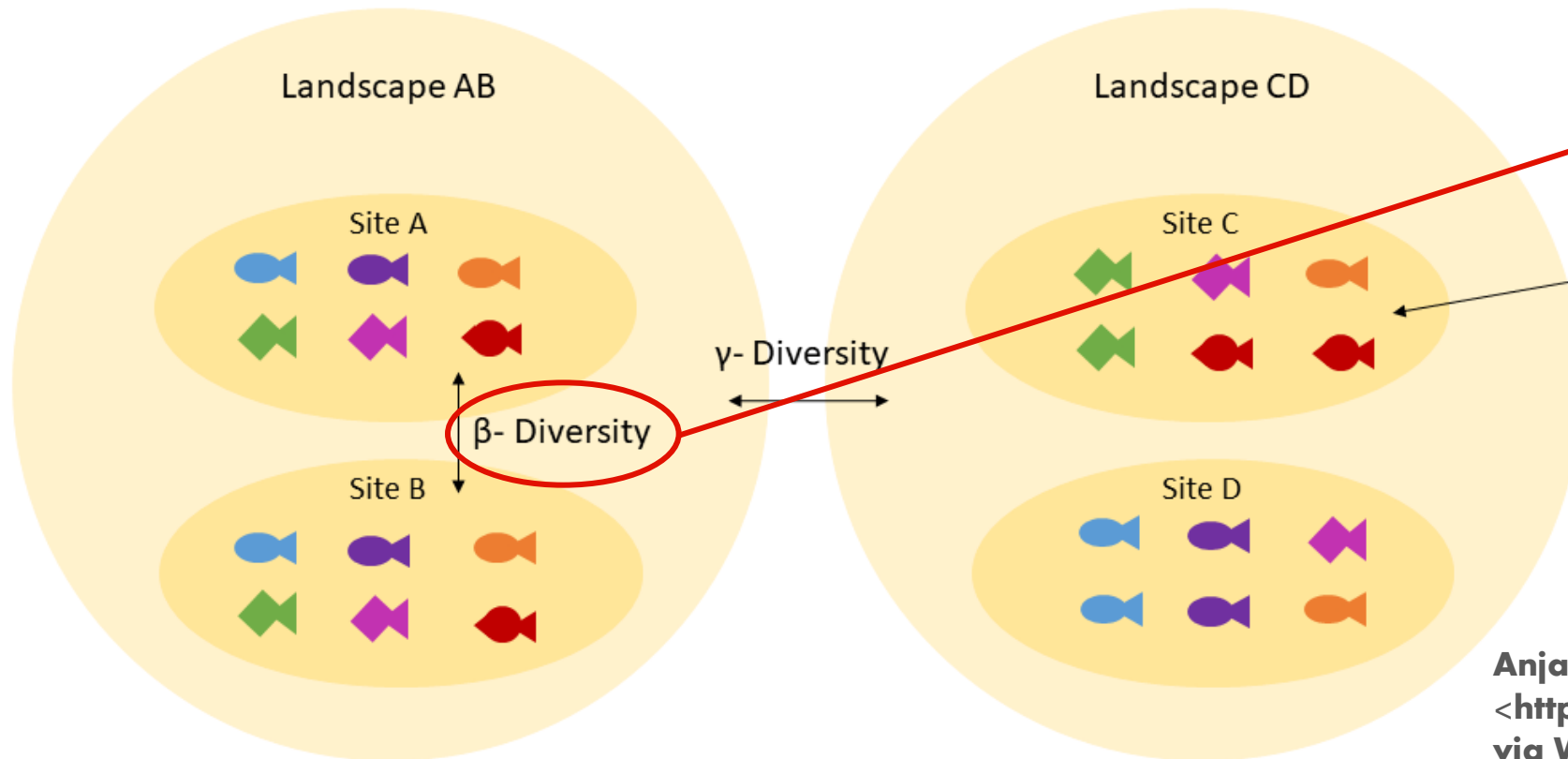
Anja Knaebel, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Exploring Concepts: Diversity Statistics

VEGETATION OF THE SISKIYOU MOUNTAINS, OREGON AND CALIFORNIA¹

R. H. WHITTAKER

Biology Department, Brooklyn College, Brooklyn 10, N. Y.



“The extent of change of community composition, or degree of community differentiation, in relation to a complex gradient of environment, or a pattern of environments”

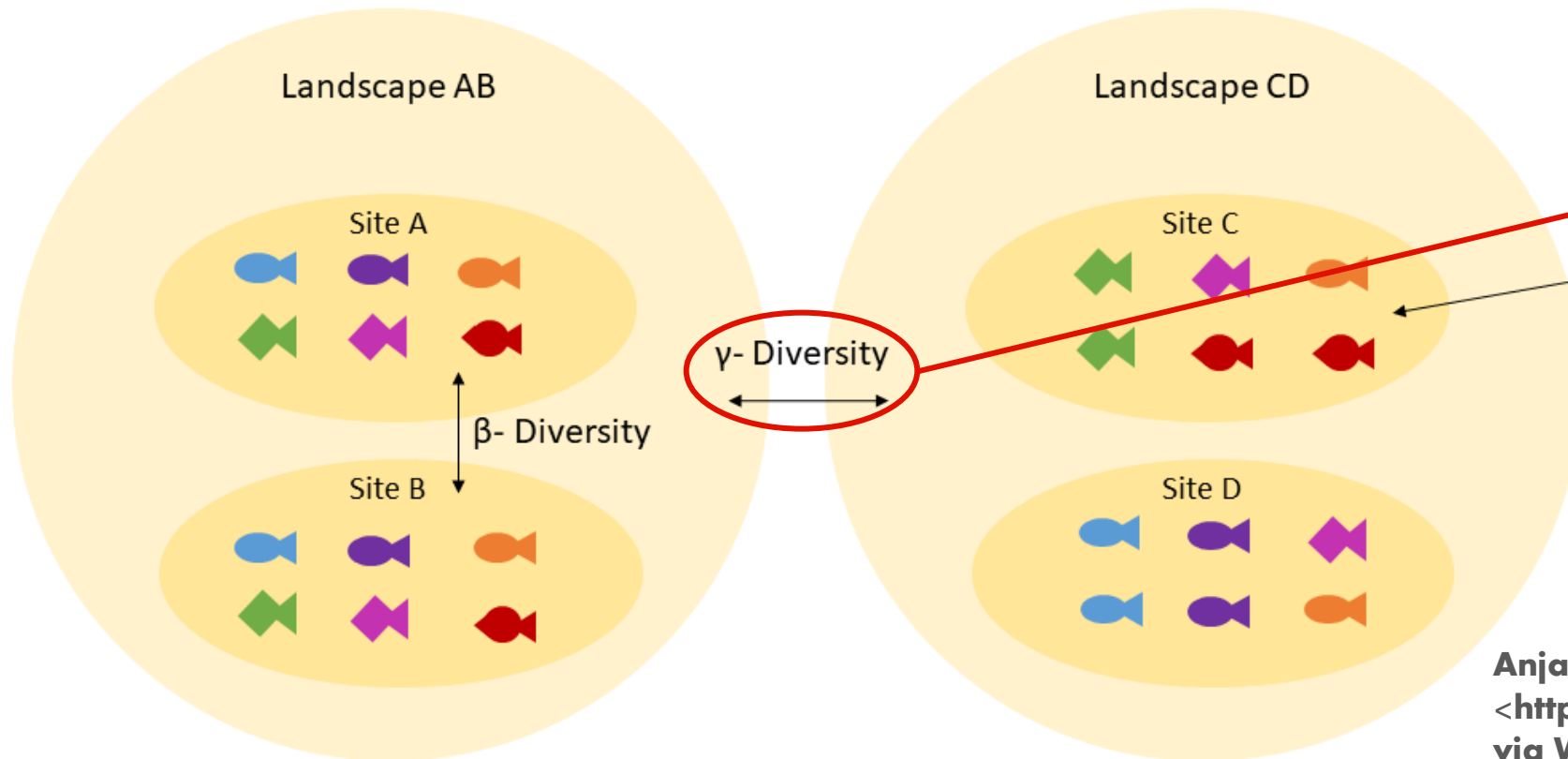
Anja Knaebel, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Exploring Concepts: Diversity Statistics

VEGETATION OF THE SISKIYOU MOUNTAINS, OREGON AND CALIFORNIA¹

R. H. WHITTAKER

Biology Department, Brooklyn College, Brooklyn 10, N. Y.



“The species-diversity of a number of community samples, for some range of environments, which have been combined, so that the diversity value is a resultant of both alpha and beta diversities of these samples.”

Anja Knaebel, CC BY-SA 4.0

<<https://creativecommons.org/licenses/by-sa/4.0/>>,

via Wikimedia Commons



Ecography 33: 2–22, 2010
doi: 10.1111/j.1600-0587.2009.05880.x
© 2010 The Author. Journal compilation © 2010 Ecography
Subject Editor: Robert K. Colwell. Accepted 18 November 2009

A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity

Hanna Tuomisto

H. Tuomisto (hanna.tuomisto@utu.fi), Dept of Biology, FI-20014 Univ. of Turku, Finland.

<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1600-0587.2009.05880.x>

- **Definition of beta diversity can be confusing!!!**
- **Diagram works as a simplified version of species diversity metrics.**

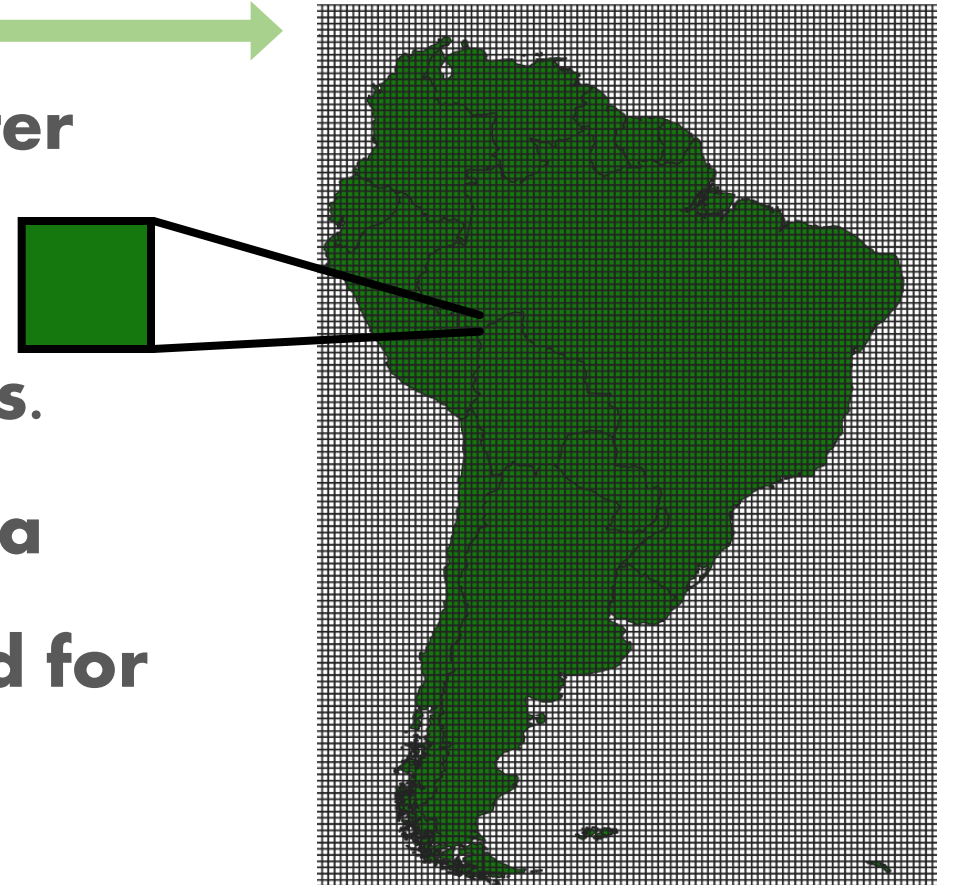
Understand what a grid is ...

Geospatial **grid** built in vector format (later saved as a shapefile).

Grid is used for defining **cells** as polygons.

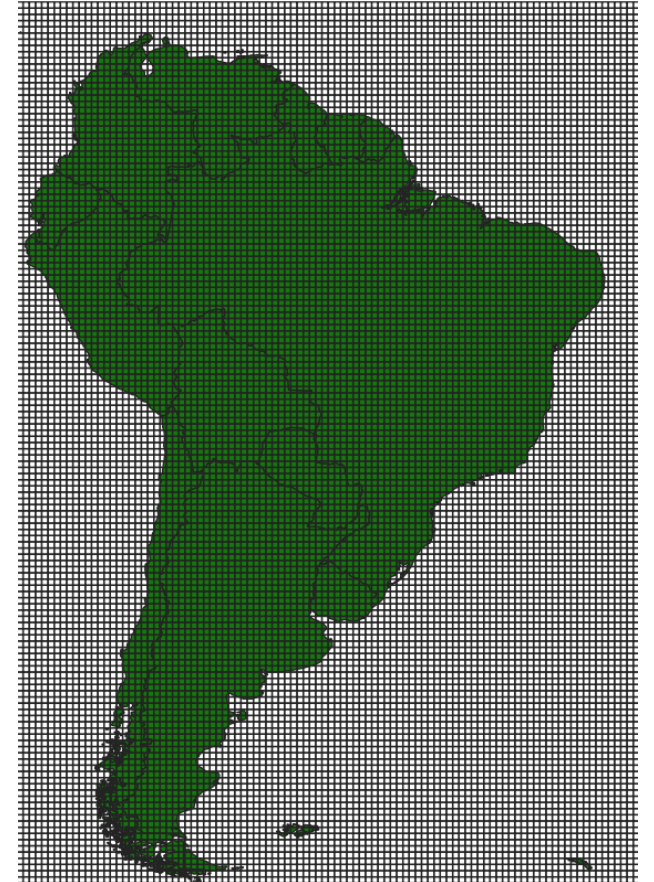
A grid defines the geographic extent for a

PAM or other biogeographic matrix, used for multi-species analyses.



Understand the effects of grid scale ...

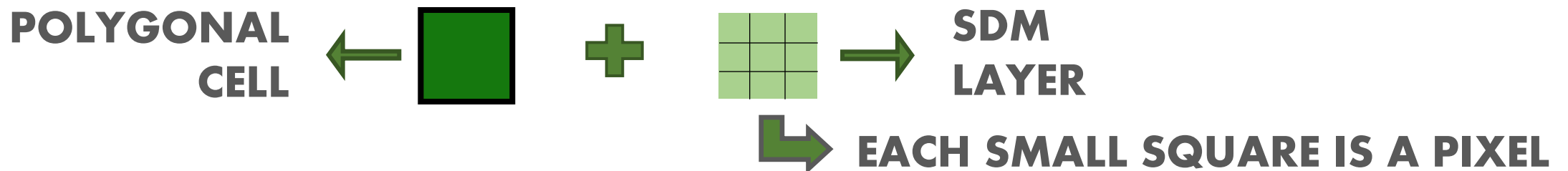
Defining the grid is very important!! Geography based studies rely on *scale* which directly impacts analyses/results. Also, defining grid scale is done considering the resolution of SDMs layers. Usually, the grid resolution is coarser than the SDMs layers!



Learn how to create a PAM matrix ...

Polygonal cells in a grid can be intersected with other geospatial data layers, such as SDMs!

To intersect SDMs with a grid, it is necessary to generalize across different pixel values in the SDM layer contained in a single grid cell.



Learn how to create a PAM matrix ...

Each pixel in an SDM receives a score from 0 – 1 that represents the probability of occurrence of a given species in that pixel.

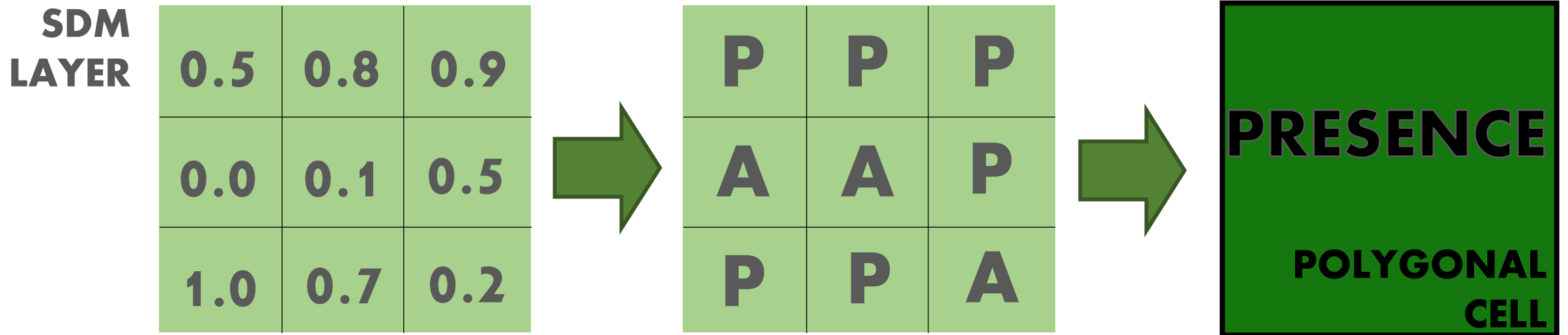
SDM LAYER	LAT Y		
	L	O	N
	0.5	0.8	0.9
	0.0	0.1	0.5
	1.0	0.7	0.2



Each pixel in an SDM is delimited by a unique latitude at the Y axis and longitude at the X axis.

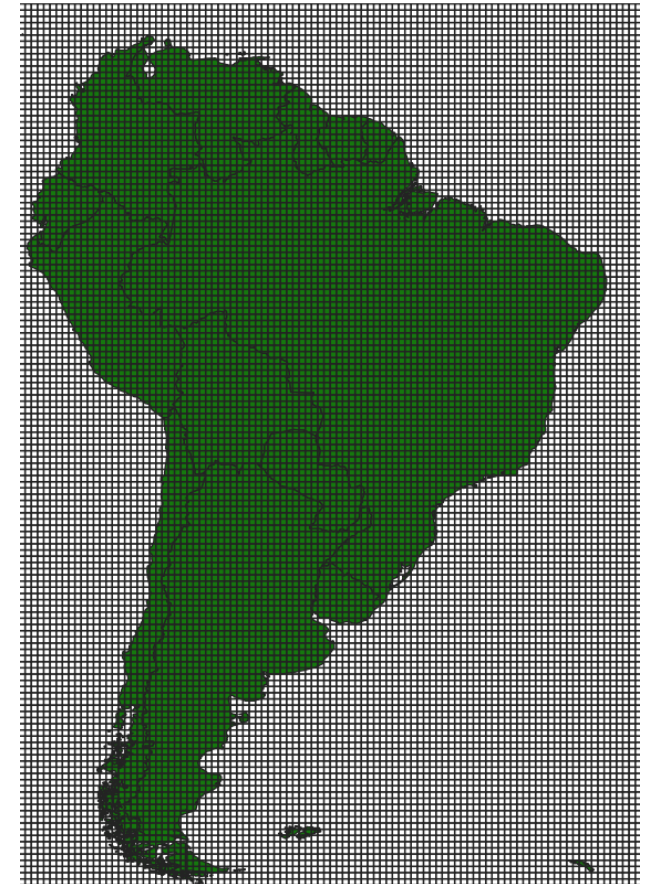
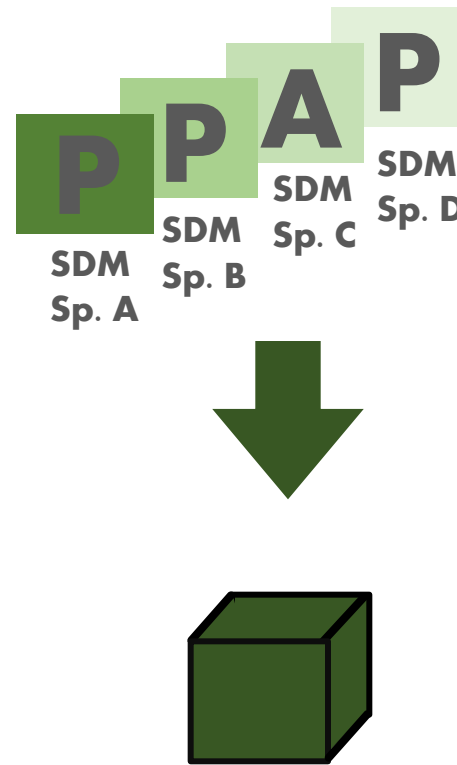
Learn how to create a PAM matrix ...

Values over 0.25 usually account for presence; thus, generalization across these pixels leads to accounting for presence in the grid cell!



Learn how to create a PAM matrix ...

To intersect multiple two-dimensional SDM layers with the same grid cell, it is necessary to stack these layers into a cube.



Learn what a PAM is in a biogeographic context ...

Then, this cube is flattened back into a 2-dimensional multi-layer matrix, by transforming each SDM layer into a matrix of a single column with multiple rows.

L L A T Y

O N X	1	P	2	P	3	A
	4	A	5	P	6	P
	7	P	8	A	9	P
	10	A	11	P	12	P

SDM LAYER
For a single
species

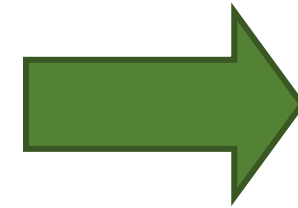
Each small square corresponds to a cell grid

Learn what a PAM is in a biogeographic context ...

Then, this cube is flattened back into a 2-dimensional multi-layer matrix, by transforming each SDM layer into a matrix of a single column with multiple rows.

L L A T Y

O N X	1	P	2	P	3	A
	4	A	5	P	6	P
	7	P	8	A	9	P
	10	A	11	P	12	P



SDM LAYER
For a single
species

Each small square
corresponds to a
cell grid

1	P
2	P
3	A
4	A
5	P
6	P
7	P
8	A
9	P
10	A
11	P
12	P

Understand PAM as a data model

COLUMNS = SPECIES



	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				<i>Colobanthus apetalus</i>	<i>Colobanthus affinis</i>	<i>Rhagodia spinescens</i>	<i>Atriplex limbata</i>	<i>Einadia hastata</i>	<i>Atriplex incrassata</i>	<i>Atriplex nana</i>
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0

Understand PAM as a data model

COLUMNS = SPECIES



	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				Colobanthus apetalus	Colobanthus affinis	Rhagodia spinescens	Atriplex limbata	Einadia hastata	Atriplex incrassata	Atriplex nana
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0

Understand PAM as a data model

**ROWS =
SITES** →

	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				Colobanthus apetalus	Colobanthus affinis	Rhagodia spinescens	Atriplex limbata	Einadia hastata	Atriplex incrassata	Atriplex nana
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0

Understand PAM as a data model

**ROWS =
SITES** →

	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				Colobanthus apetalus	Colobanthus affinis	Rhagodia spinescens	Atriplex limbata	Einadia hastata	Atriplex incrassata	Atriplex nana
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0

Exploring Concepts: Encode SDM Layers

Understand PAM as a data model

COLUMNS = SPECIES



ROWS =
SITES →


LAT/LONG is
represented
by centroid
value of a
given cell

	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				Colobanthus apetalus	Colobanthus affinis	Rhagodia spinescens	Atriplex limbata	Einadia hastata	Atriplex incrassata	Atriplex nana
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0



Calculate statistics for a PAM

POLYGONAL CELL

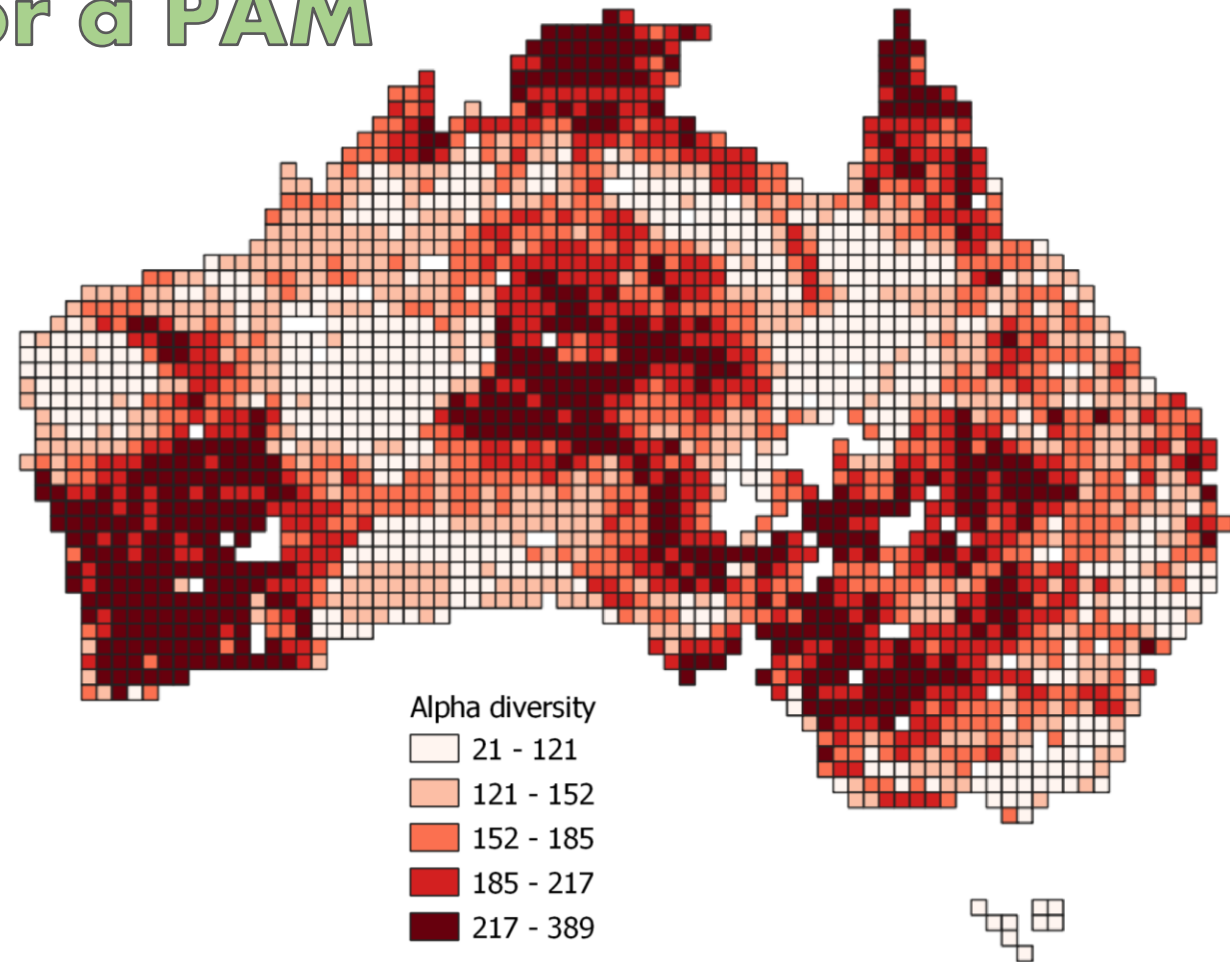


	A	Q	R	S	T	U
1	WKT	alpha	alpha proportional	phi	phi average proportional	Phylogenetic Diversity
1126	POLYGON ((128.0 -16.5,128.5 -16.5,128.5 -17.0,128 -17,128.0 -16.5))	133	0.028106509	71141	0.217791017	5521.499304
1127	POLYGON ((128.5 -16.5,129.0 -16.5,129 -17,128.5 -17.0,128.5 -16.5))	113	0.023879966	56207	0.202527313	4701.04883
1128	POLYGON ((129.0 -16.5,129.5 -16.5,129.5 -17.0,129 -17,129.0 -16.5))	157	0.03317836	72464	0.18792921	5889.895471
1129	POLYGON ((129.5 -16.5,130.0 -16.5,130 -17,129.5 -17.0,129.5 -16.5))	126	0.026627219	64510	0.208462592	5196.282523
1130	POLYGON ((130.0 -16.5,130.5 -16.5,130.5 -17.0,130 -17,130.0 -16.5))	198	0.041842773	84780	0.174341131	6730.982417
1131	POLYGON ((130.5 -16.5,131.0 -16.5,131 -17,130.5 -17.0,130.5 -16.5))	137	0.028951817	67097	0.199413324	5427.129703
1132	POLYGON ((131.0 -16.5,131.5 -16.5,131.5 -17.0,131 -17,131.0 -16.5))	125	0.026415892	60062	0.195641694	5159.118098
1133	POLYGON ((131.5 -16.5,132.0 -16.5,132 -17,131.5 -17.0,131.5 -16.5))	119	0.025147929	61572	0.210672543	5106.485991
1134	POLYGON ((132.0 -16.5,132.5 -16.5,132.5 -17.0,132 -17,132.0 -16.5))	192	0.04057481	94409	0.200209097	6900.66665
1135	POLYGON ((132.5 -16.5,133.0 -16.5,133 -17,132.5 -17.0,132.5 -16.5))	181	0.038250211	89027	0.200269494	6556.122268
1136	POLYGON ((133.0 -16.5,133.5 -16.5,133.5 -17.0,133 -17,133.0 -16.5))	105	0.022189349	68413	0.265290057	4577.727162
1137	POLYGON ((133.5 -16.5,134.0 -16.5,134 -17,133.5 -17.0,133.5 -16.5))	138	0.029163145	73761	0.217630293	5495.218733
1138	POLYGON ((134.0 -16.5,134.5 -16.5,134.5 -17.0,134 -17,134.0 -16.5))	173	0.036559594	88123	0.207402892	6418.042032
1139	POLYGON ((134.5 -16.5,135.0 -16.5,135 -17,134.5 -17.0,134.5 -16.5))	161	0.034023669	81369	0.205780747	6145.732
1140	POLYGON ((135.0 -16.5,135.5 -16.5,135.5 -17.0,135 -17,135.0 -16.5))	176	0.037193576	84444	0.195356455	6605.793222
1141	POLYGON ((135.5 -16.5,136.0 -16.5,136 -17,135.5 -17.0,135.5 -16.5))	189	0.039940828	84509	0.18205927	7239.173472
1142	POLYGON ((136.0 -16.5,136.5 -16.5,136.5 -17.0,136 -17,136.0 -16.5))	203	0.042899408	83290	0.167058455	7733.054866
1143	POLYGON ((136.5 -16.5,137.0 -16.5,137 -17,136.5 -17.0,136.5 -16.5))	205	0.043322063	82496	0.163851593	7674.844733
1144	POLYGON ((137.0 -16.5,137.5 -16.5,137.5 -17.0,137 -17,137.0 -16.5))	216	0.045646661	82580	0.155665641	8027.774269

Calculate statistics for a PAM

**Heat map
created with
QGIS to display
alpha diversity!**

**File needs to be a
GEOJson to be
read in QGIS!**



Time to Exercise!

**What happens when there is a large dataset?
Should we calculate diversity statistics one
taxon at a time?**



**We use BiotaPhy tools to automate
calculating diversity statistics for
multiple taxa at once!**



Before we start ...

Download for the first time OR update the tutorials repository containing test data and configurations.

Initial download:

```
git clone https://github.com/biotaphy/tutorials
```

Update tutorial:

```
cd tutorials
```

```
git pull
```


Let's put the automated framework developed by BiotaPhy to the test!

3 steps:

1. **Build a grid to analyze a multi-species dataset**
2. **Encode layers into a Presence-Absence matrix (PAM) for multi-species analysis**
3. **Calculate statistics for a PAM**

Time to Exercise!

	Steps	Sub-steps		
		Data Preparation	Run Tutorial	Inspect Output
1	Build a grid	1. Script parameter file	✓	✓
2	Encode Layers	1. Encoding layers 2. Script parameter file	✓	✓
3	Calculate statistics	1. Script parameter file	✓	✓

Time to Exercise!

	Steps	Sub-steps		
		Data Preparation	Run Tutorial	Inspect Output
1	Build a grid	1. Script parameter file	✓	✓
2	Encode Layers	1. Encoding layers 2. Script parameter file	✓	✓
3	Calculate statistics	1. Script parameter file	✓	✓

Time to Exercise!

	Steps	Sub-steps		
		Data Preparation	Run Tutorial	Inspect Output
1	Build a grid	1. Script parameter file	✓	✓
2	Encode Layers	1. Encoding layers 2. Script parameter file	✓	✓
3	Calculate statistics	1. Script parameter file	✓	✓

Step 1: Build a grid

Data preparation: Script parameter file

A JSON parameter file is required for this command. The tutorial parameter file is [build_grid.json](#).

These are the required and optional parameters:

- Required:

- **grid_filename**: The relative or absolute path for the output grid.
- **min_x**: The minimum value for X (longitude) coordinate of the grid.
- **min_y**: The minimum value for Y (latitude) coordinate of the grid.
- **max_x**: The maximum value for X (longitude) coordinate of the grid.
- **max_y**: The maximum value for Y (latitude) coordinate of the grid.
- **cell_size**: The size of each cell (in units indicated by EPSG).
- **epsg**: The EPSG code for the new grid.

```
{
  "log_filename": "/volumes/output/build_grid.log",
  "log_console": true,
  "report_filename": "/volumes/output/build_grid.rpt",
  "grid_filename": "/volumes/output/grid_na_5deg.shp",
  "min_x": -164,
  "min_y": 15,
  "max_x": -53,
  "max_y": 64,
  "cell_size": 5.0,
  "epsg": 4326
}
```

Step 1: Build a grid

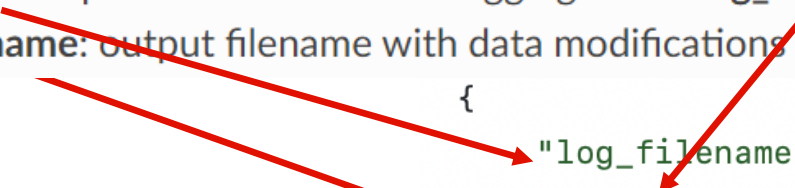
Data preparation: Script parameter file

A JSON parameter file is required for this command. The tutorial parameter file is [build_grid.json](#).

These are the required and optional parameters:

- Optional * **log_filename**: Output filename to write logging data * **log_console**: 'true' to write log to console * **report_filename**: output filename with data modifications made by wranglers

```
{  
  "log_filename": "/volumes/output/build_grid.log",  
  "log_console": true,  
  "report_filename": "/volumes/output/build_grid.rpt",  
  "grid_filename": "/volumes/output/grid_na_5deg.shp",  
  "min_x": -164,  
  "min_y": 15,  
  "max_x": -53,  
  "max_y": 64,  
  "cell_size": 5.0,  
  "epsg": 4326
```

Two red arrows originate from the list items above. One arrow points from the text '* log_filename' to the '"log_filename"' field in the JSON code. The other arrow points from the text '* report_filename' to the '"report_filename"' field in the JSON code.

Let's run this tutorial: Build a Grid



Run build_grid command

Initiate the process with the following:

For Linux/Mac systems:

```
./run_tutorial.sh build_grid data/config/build_grid.json
```

For Windows systems:

```
./run_tutorial.bat build_grid data/config/build_grid.json
```

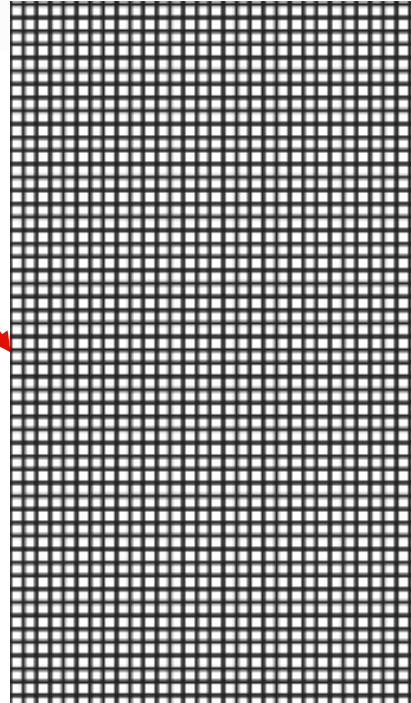
Goal: build a grid to analyze a multi-species dataset

Remember, you will RUN this code in the terminal or in a similar environment (for Windows)!

Let's look at the output: Build a Grid

Output

The build_grid tool outputs a grid in shapefile format, conforming to the arguments specified in the configuration JSON file.



**THE OUTPUT IS SOLELY A GRID.
AT THIS POINT, NO TYPE OF
DATA WILL BE CONTAINED IN IT!**

Step 2: Encode species layers

Input: Layers

Layers are specified in the Script parameter file, described fully in the next section. Each layer of a PAM represents the presence or absence of a taxon in the set of gridcells used in the analysis. Presence or absence is calculated with the `min_coverage`, `min_presence`, and `max_presence` parameters also detailed in the Script parameter file.

If you want to define different values for computing different layers, you can create a matrix from each set of layers that share parameters (each matrix created with a different configuration file containing different parameters and the same grid), then aggregate the matrices in another step.

Data Preparation: Encode Layers

Data preparation: Script parameter file

An example json file for running the `encode_layers` tutorial is at [encode_layers.json](#). These are the required and optional parameters:

Required:

- **grid_filename**: Relative path to shapegrid to use for site geometries
- **out_matrix_filename**: Relative path to write the encoded matrix
- **encode_method**: The only valid option for creating a PAM is "presence_absence". Other options "biogeo", "largest_class", "mean_value", are used for a different types of matrices and operations.

```
{
  "log_filename": "/volumes/output/encode_layers.log",
  "log_console": true,
  "report_filename": "/volumes/output/encode_layers.rpt",
  "encode_method": "presence_absence",
  "min_coverage": 25,
  "min_presence": 0.1,
  "max_presence": 1.0,
  "grid_filename": "/volumes/data/input/grid_na_5deg.shp",
  "out_matrix_filename": "/volumes/output/heuchera_pam.lmm",
  "layer_file_pattern": "/volumes/data/input/heuchera_rfolk_sdm/**/*.asc"
}
```

Data Preparation: Encode Layers

Data preparation: Script parameter file

An example json file for running the `encode_layers` tutorial is at [encode_layers.json](#). These are the required and optional parameters:

Required:

- **grid_filename**: Relative path to shapegrid to use for site geometries
- **out_matrix_filename**: Relative path to write the encoded matrix
- **encode_method**: The only valid option for creating a PAM is “presence_absence”. Other options “biogeo”, “largest_class”, “mean_value”, are used for a different types of matrices and operations.

```
{
  "log_filename": "/volumes/output/encode_layers.log",
  "log_console": true,
  "report_filename": "/volumes/output/encode_layers.rpt",
  "encode_method": "presence_absence",
  "min_coverage": 25,
  "min_presence": 0.1,
  "max_presence": 1.0,
  "grid_filename": "/volumes/data/input/grid_na_5deg.shp",
  "out_matrix_filename": "/volumes/output/heuchera_pam.lmm",
  "layer_file_pattern": "/volumes/data/input/heuchera_rfolk_sdm/**/*.asc"
}
```

Data Preparation: Encode Layers

Data preparation: Script parameter file

An example json file for running the encode_layers tutorial is at [encode_layers.json](#). These are the required and optional parameters:

- Optional
 - **min_coverage**: Minimum percentage of a cell that has to be covered to encode it
 - **min_presence**: Minimum value to be considered present when encoding presence absence
 - **max_presence**: Maximum value to be considered present
 - **layer**: list of a
 - layer filename
 - (optional) label for the encoded layer in the output matrix. If no label is provided for a layer file, the program will first look for another file in the same directory with the same basename, and with the extension ".label". If this Defaults to the the first line of a file in the same directory and with the same basename as lyr_filename and a ".label" extension, OR the basename of the layer file.
 - (optional) attribute. Defaults to None, using the pixel value for raster data.
 - **layer_file_pattern**: File pattern that describes one or more input files.
 - **log_filename**: Output filename to write logging data
 - **log_console**: 'true' to write log to console
 - **report_filename**: output filename with data modifications made by wranglers

```
{
  "log_filename": "/volumes/output/encode_layers.log",
  "log_console": true,
  "report_filename": "/volumes/output/encode_layers.rpt",
  "encode_method": "presence_absence",
  "min_coverage": 25,
  "min_presence": 0.1,
  "max_presence": 1.0,
  "grid_filename": "/volumes/data/input/grid_na_5deg.shp",
  "out_matrix_filename": "/volumes/output/heuchera_pam.lmm",
  "layer_file_pattern": "/volumes/data/input/heuchera_rfolk_sdm/*/*.asc"
}
```

Data preparation: Script parameter file

An example json file for running the encode_layers tutorial is at [encode_layers.json](#). These are the required and optional parameters:

- Optional
 - **min_coverage**: Minimum percentage of a cell that has to be covered to encode it
 - **min_presence**: Minimum value to be considered present when encoding presence absence
 - **max_presence**: Maximum value to be considered present
 - **layer**: list of a
 - layer filename
 - (optional) label for the encoded layer in the output matrix. If no label is provided for a layer file, the program will first look for another file in the same directory with the same basename, and with the extension ".label". If this Defaults to the the first line of a file in the same directory and with the same basename as lyr_filename and a ".label" extension, OR the basename of the layer file.
 - (optional) attribute. Defaults to None, using the pixel value for raster data.
 - **layer_file_pattern**: File pattern that describes one or more input files.
 - **log_filename**: Output filename to write logging data
 - **log_console**: 'true' to write log to console
 - **report_filename**: output filename with data modifications made by wranglers

```
{
  "log_filename": "/volumes/output/encode_layers.log",
  "log_console": true,
  "report_filename": "/volumes/output/encode_layers.rpt",
  "encode_method": "presence_absence",
  "min_coverage": 25,
  "min_presence": 0.1,
  "max_presence": 1.0,
  "grid_filename": "/volumes/data/input/grid_na_5deg.shp",
  "out_matrix_filename": "/volumes/output/heuchera_pam.lmm",
  "layer_file_pattern": "/volumes/data/input/heuchera_rfolk_sdm/**/*.asc"
}
```

Data Preparation: Encode Layers

Data preparation: Script parameter file

An example json file for running the encode_layers tutorial is at [encode_layers.json](#). These are the required and optional parameters:

- Optional
 - **min_coverage**: Minimum percentage of a cell that has to be covered to encode it
 - **min_presence**: Minimum value to be considered present when encoding presence absence
 - **max_presence**: Maximum value to be considered present
 - **layer**: list of a
 - layer filename
 - (optional) label for the encoded layer in the output matrix. If no label is provided for a layer file, the program will first look for another file in the same directory with the same basename, and with the extension ".label". If this Defaults to the the first line of a file in the same directory and with the same basename as lyr_filename and a ".label" extension, OR the basename of the layer file.
 - (optional) attribute. Defaults to None, using the pixel value for raster data.
 - **layer_file_pattern**: File pattern that describes one or more input files.
 - **log_filename**: Output filename to write logging data
 - **log_console**: 'true' to write log to console
 - **report_filename**: output filename with data modifications made by wranglers

```
{  
  "log_filename": "/volumes/output/encode_layers.log",  
  "log_console": true,  
  "report_filename": "/volumes/output/encode_layers.rpt",  
  "encode_method": "presence_absence",  
  "min_coverage": 25,  
  "min_presence": 0.1,  
  "max_presence": 1.0,  
  "grid_filename": "/volumes/data/input/global_2deg_grid.shp",  
  "out_matrix_filename": "/volumes/output/encode_layers.lmm",  
  "layer_file_pattern": "/volumes/data/easy_bake/heuchera_sdm/*/*.asc"  
}
```

Let's run this tutorial: Encode Layers



Run `encode_layers` command

Initiate the process with the following:

For Linux/Mac systems:

```
./run_tutorial.sh encode_layers data/config/encode_layers.json
```

For Windows systems:

```
./run_tutorial.bat encode_layers data/config/encode_layers.json
```

Note: You may get the following warning. This indicates that there is a window with no values, a common occurrence in most datasets, and may be safely ignored.

```
RuntimeWarning: Mean of empty slice  
window_mean = np.nanmean(window[np.where(window != nodata)])
```

Goal: Encode Layers into a Presence-Absence matrix (PAM) for multi-species analysis

Remember, you will RUN this code in the terminal or in a similar environment (for Windows)!

Let's look at the output: Encode Layers

COLUMNS = SPECIES



**ROWS =
SITES** →

	A	B	C	BNH	BNI	BNJ	BNK	BNL	BNM	BNN
1				Colobanthus apetalus	Colobanthus affinis	Rhagodia spinescens	Atriplex limbata	Einadia hastata	Atriplex incrassata	Atriplex nana
3006	3004	144.25	-27.75	0	0	1	1	0	0	0
3007	3005	144.75	-27.75	0	0	1	1	0	0	0
3008	3006	145.25	-27.75	0	0	1	1	0	0	0
3009	3007	145.75	-27.75	0	0	1	1	0	0	0
3010	3008	146.25	-27.75	0	0	1	0	0	0	0
3011	3009	146.75	-27.75	0	0	1	0	0	0	0
3012	3010	147.25	-27.75	0	0	1	0	0	0	0
3013	3011	147.75	-27.75	0	0	1	0	0	0	0
3014	3012	148.25	-27.75	0	0	1	0	0	0	0
3015	3013	148.75	-27.75	0	0	1	0	0	0	0
3016	3014	149.25	-27.75	0	0	1	0	0	0	0
3017	3015	149.75	-27.75	0	0	1	0	0	0	0
3018	3016	150.25	-27.75	0	0	1	0	0	0	0
3019	3017	150.75	-27.75	0	0	1	0	0	0	0
3020	3018	151.25	-27.75	0	0	0	0	0	0	0
3021	3019	151.75	-27.75	0	0	0	0	0	0	0
3022	3020	152.25	-27.75	0	0	0	0	0	0	0
3023	3021	152.75	-27.75	0	0	0	0	1	0	0
3024	3022	153.25	-27.75	0	0	0	0	0	0	0

Step 3: Calculate statistics for a PAM

Now that a grid has been built, and a PAM has been populated by intersecting species distribution models with the grid, we calculate biogeographic statistics on that PAM.

Data preparation: Script parameter file

An example JSON file for running the `calculate_pam_stats` command is at [calculate_pam_stats.json](#).

These are the required and optional parameters:

- Required:
 - **pam_filename**: The full filename to the input PAM file

```
{
  "log_filename": "/volumes/output/calculate_pam_stats.log",
  "log_console": true,
  "report_filename": "/volumes/output/calculate_pam_stats.rpt",
  "covariance_matrix": "/volumes/output/covariance.lmm",
  "diversity_matrix": "/volumes/output/diversity.lmm",
  "site_stats_matrix": "/volumes/output/site_stats.lmm",
  "species_stats_matrix": "/volumes/output/species_stats.lmm",
  "pam_filename": "/volumes/data/input/heuchera_pam.lmm"
}
```

Data preparation: Script parameter file

An example JSON file for running the `calculate_pam_stats` command is at [calculate_pam_stats.json](#).

These are the required and optional parameters:

- Optional

- **log_filename**: Output filename to write logging data
- **log_console**: 'true' to write log to console
- **report_filename**: output filename with summary
- **covariance_matrix**: The full path and base filename for writing covariance matrices.
- **diversity_matrix**: The full filename for writing the diversity matrix.
- **site_stats_matrix**: The full filename for writing the site statistics matrix.
- **species_stats_matrix**: The full filename for writing the species statistics matrix.
- **tree_filename**: The full filename to an input tree in Nexus format.
- **tree_matrix**: The full filename to an input tree encoded as a matrix.

```
{
  "log_filename": "/volumes/output/calculate_pam_stats.log",
  "log_console": true,
  "report_filename": "/volumes/output/calculate_pam_stats.rpt",
  "covariance_matrix": "/volumes/output/covariance.lmm",
  "diversity_matrix": "/volumes/output/diversity.lmm",
  "site_stats_matrix": "/volumes/output/site_stats.lmm",
  "species_stats_matrix": "/volumes/output/species_stats.lmm",
  "pam_filename": "/volumes/data/input/heuchera_pam.lmm"
}
```

Data preparation: Script parameter file


An example JSON file for running the `calculate_pam_stats` command is at [calculate_pam_stats.json](#).

These are the required and optional parameters:

- Optional

- **log_filename**: Output filename to write logging data
- **log_console**: 'true' to write log to console
- **report_filename**: output filename with summary
- **covariance_matrix**: The full path and base filename for writing covariance matrices.
- **diversity_matrix**: The full filename for writing the diversity matrix.
- **site_stats_matrix**: The full filename for writing the site statistics matrix.
- **species_stats_matrix**: The full filename for writing the species statistics matrix.
- **tree_filename**: The full filename to an input tree in Nexus format.
- **tree_matrix**: The full filename to an input tree encoded as a matrix.

```
{
  "log_filename": "/volumes/output/calculate_pam_stats.log",
  "log_console": true,
  "report_filename": "/volumes/output/calculate_pam_stats.rpt",
  "covariance_matrix": "/volumes/output/covariance.lmm",
  "diversity_matrix": "/volumes/output/diversity.lmm",
  "site_stats_matrix": "/volumes/output/site_stats.lmm",
  "species_stats_matrix": "/volumes/output/species_stats.lmm",
  "pam_filename": "/volumes/data/input/heuchera_pam.lmm"
}
```



Run calculate_pam_stats command

Initiate the calculate_pam_stats process with the following:

For Linux/Mac systems

```
./run_tutorial.sh calculate_pam_stats data/config/calculate_pam_stats.json
```

For Windows systems

```
./run_tutorial.bat calculate_pam_stats data/config/calculate_pam_stats.json
```

**Goal: Calculate Statistics
for a PAM**

**Remember, you will
RUN this code in the
terminal or in a
similar environment
(for Windows)!**

Output

The calculate_pam_stats tool outputs computes various statistics, depending on the output files specified in the command configuration file. Outputs may include:

1. A "report_filename" named in the script parameter file, a summary of point manipulations by each wrangler will be written to this file.
2. A "log_filename" named in the script parameter file, that will be created.
3. A "log_console" named in the script parameter file, logs will be written to the command prompt during execution.
4. One or more "covariance_matrix" files. Each covariance statistic produces a matrix and it is written to the covariance_matrix filename, where the statistic name is appended to the end of the base file name.
5. A "diversity_matrix" containing different diversity statistics.
6. A "site_stats_matrix" containing site statistics.
7. A "species_stats_matrix" containing species statistics.

Let's look at the output: Calculate Statistics

Diversity Matrix!



	A	Q	R	S	T	U
1	WKT	alpha	alpha proportional	phi	phi average proportional	Phylogenetic Diversity
1126	POLYGON ((128.0 -16.5,128.5 -16.5,128.5 -17.0,128 -17,128.0 -16.5))	133	0.028106509	71141	0.217791017	5521.499304
1127	POLYGON ((128.5 -16.5,129.0 -16.5,129 -17,128.5 -17.0,128.5 -16.5))	113	0.023879966	56207	0.202527313	4701.04883
1128	POLYGON ((129.0 -16.5,129.5 -16.5,129.5 -17.0,129 -17,129.0 -16.5))	157	0.03317836	72464	0.18792921	5889.895471
1129	POLYGON ((129.5 -16.5,130.0 -16.5,130 -17,129.5 -17.0,129.5 -16.5))	126	0.026627219	64510	0.208462592	5196.282523
1130	POLYGON ((130.0 -16.5,130.5 -16.5,130.5 -17.0,130 -17,130.0 -16.5))	198	0.041842773	84780	0.174341131	6730.982417
1131	POLYGON ((130.5 -16.5,131.0 -16.5,131 -17,130.5 -17.0,130.5 -16.5))	137	0.028951817	67097	0.199413324	5427.129703
1132	POLYGON ((131.0 -16.5,131.5 -16.5,131.5 -17.0,131 -17,131.0 -16.5))	125	0.026415892	60062	0.195641694	5159.118098
1133	POLYGON ((131.5 -16.5,132.0 -16.5,132 -17,131.5 -17.0,131.5 -16.5))	119	0.025147929	61572	0.210672543	5106.485991
1134	POLYGON ((132.0 -16.5,132.5 -16.5,132.5 -17.0,132 -17,132.0 -16.5))	192	0.04057481	94409	0.200209097	6900.66665
1135	POLYGON ((132.5 -16.5,133.0 -16.5,133 -17,132.5 -17.0,132.5 -16.5))	181	0.038250211	89027	0.200269494	6556.122268
1136	POLYGON ((133.0 -16.5,133.5 -16.5,133.5 -17.0,133 -17,133.0 -16.5))	105	0.022189349	68413	0.265290057	4577.727162
1137	POLYGON ((133.5 -16.5,134.0 -16.5,134 -17,133.5 -17.0,133.5 -16.5))	138	0.029163145	73761	0.217630293	5495.218733
1138	POLYGON ((134.0 -16.5,134.5 -16.5,134.5 -17.0,134 -17,134.0 -16.5))	173	0.036559594	88123	0.207402892	6418.042032
1139	POLYGON ((134.5 -16.5,135.0 -16.5,135 -17,134.5 -17.0,134.5 -16.5))	161	0.034023669	81369	0.205780747	6145.732
1140	POLYGON ((135.0 -16.5,135.5 -16.5,135.5 -17.0,135 -17,135.0 -16.5))	176	0.037193576	84444	0.195356455	6605.793222
1141	POLYGON ((135.5 -16.5,136.0 -16.5,136 -17,135.5 -17.0,135.5 -16.5))	189	0.039940828	84509	0.18205927	7239.173472
1142	POLYGON ((136.0 -16.5,136.5 -16.5,136.5 -17.0,136 -17,136.0 -16.5))	203	0.042899408	83290	0.167058455	7733.054866
1143	POLYGON ((136.5 -16.5,137.0 -16.5,137 -17,136.5 -17.0,136.5 -16.5))	205	0.043322063	82496	0.163851593	7674.844733
1144	POLYGON ((137.0 -16.5,137.5 -16.5,137.5 -17.0,137 -17,137.0 -16.5))	216	0.045646661	82580	0.155665641	8027.774269

1. **Alpha, beta and gamma diversity are commonly used diversity metrics. Be careful with interpretation and rely on recent literature for help!**
2. **Presence-Absence Matrices (PAMs) are a great way to optimize diversity statistics involving multiple species!**
3. **Three steps must be followed to use PAMs:**
 - **Build a grid**
 - **Encode layers**
 - **Calculate statistics**

Any questions??

Please, use the chat to write your question!