

Genetics the R ight way



Hayley Tumas
BiRmington UseR Group
University of Birmingham
27 September 2019

Genetics the R ight way

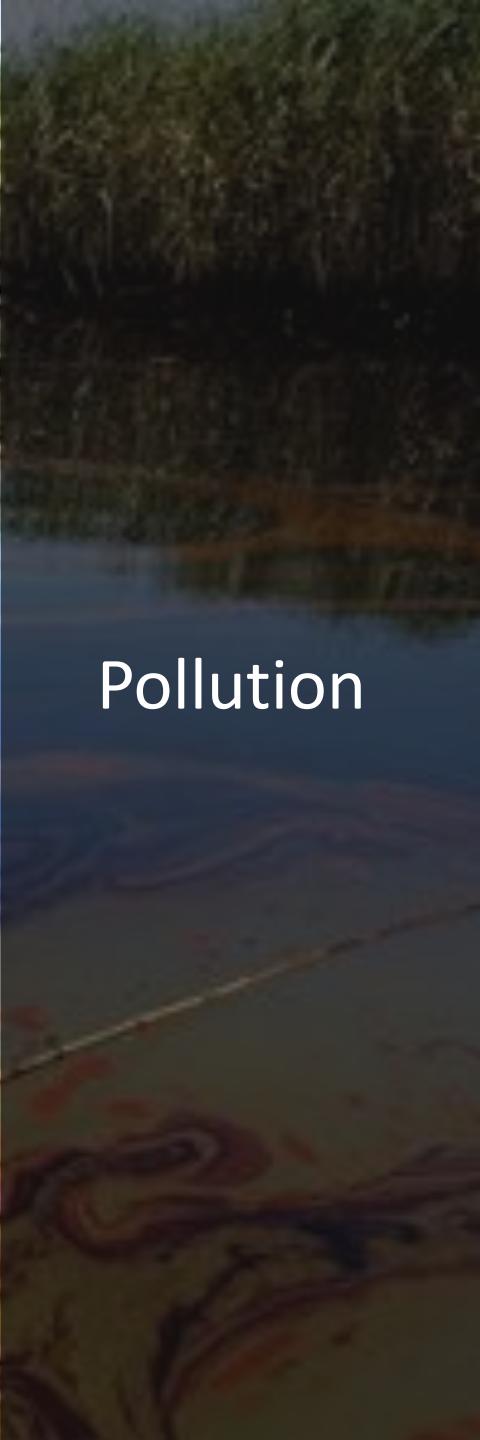
Outline:

- Why conservation genetics?
- How can R improve the genetic workflow?
- How can R help innovate genetic analysis?





Development



Pollution



Climate
Change



Sea Level
Rise



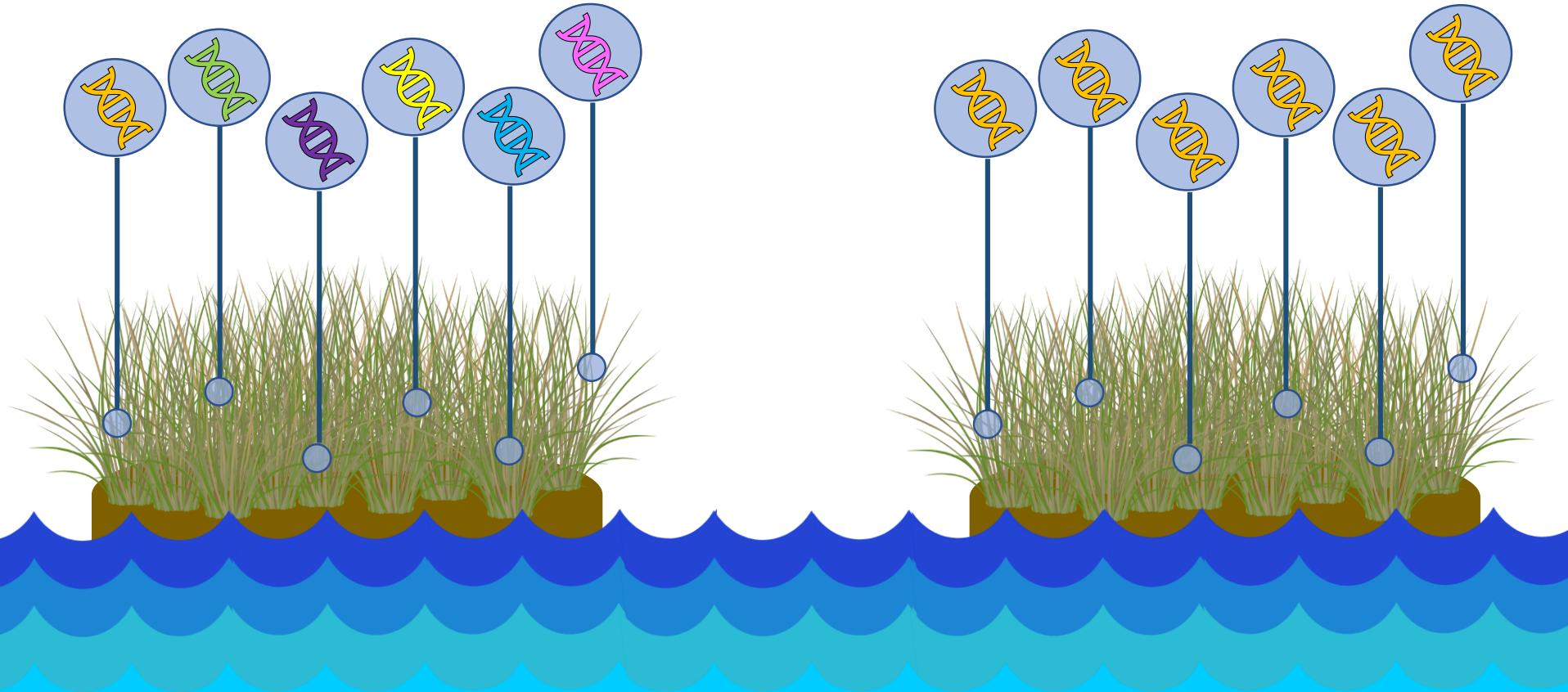
Restoration

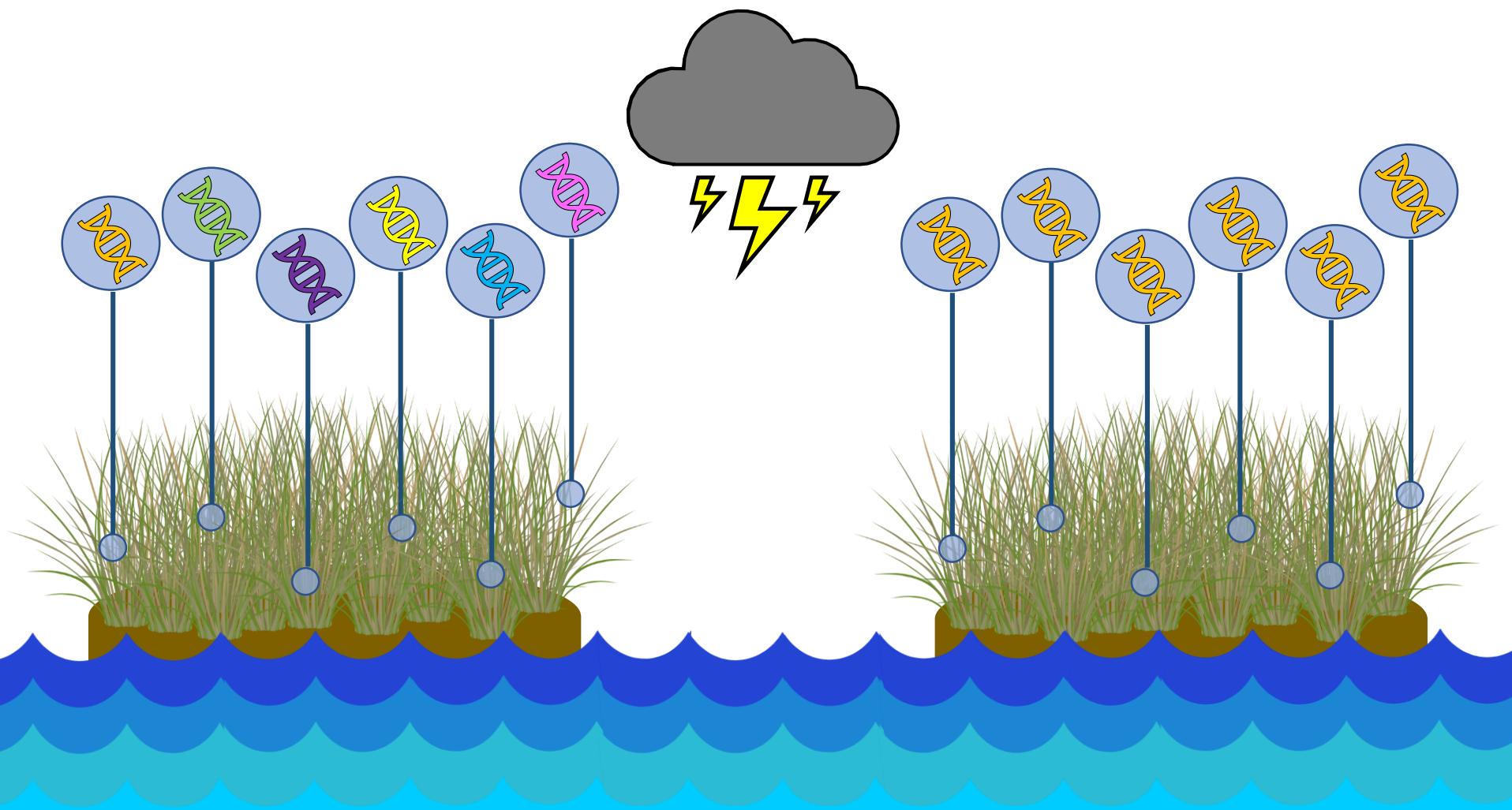


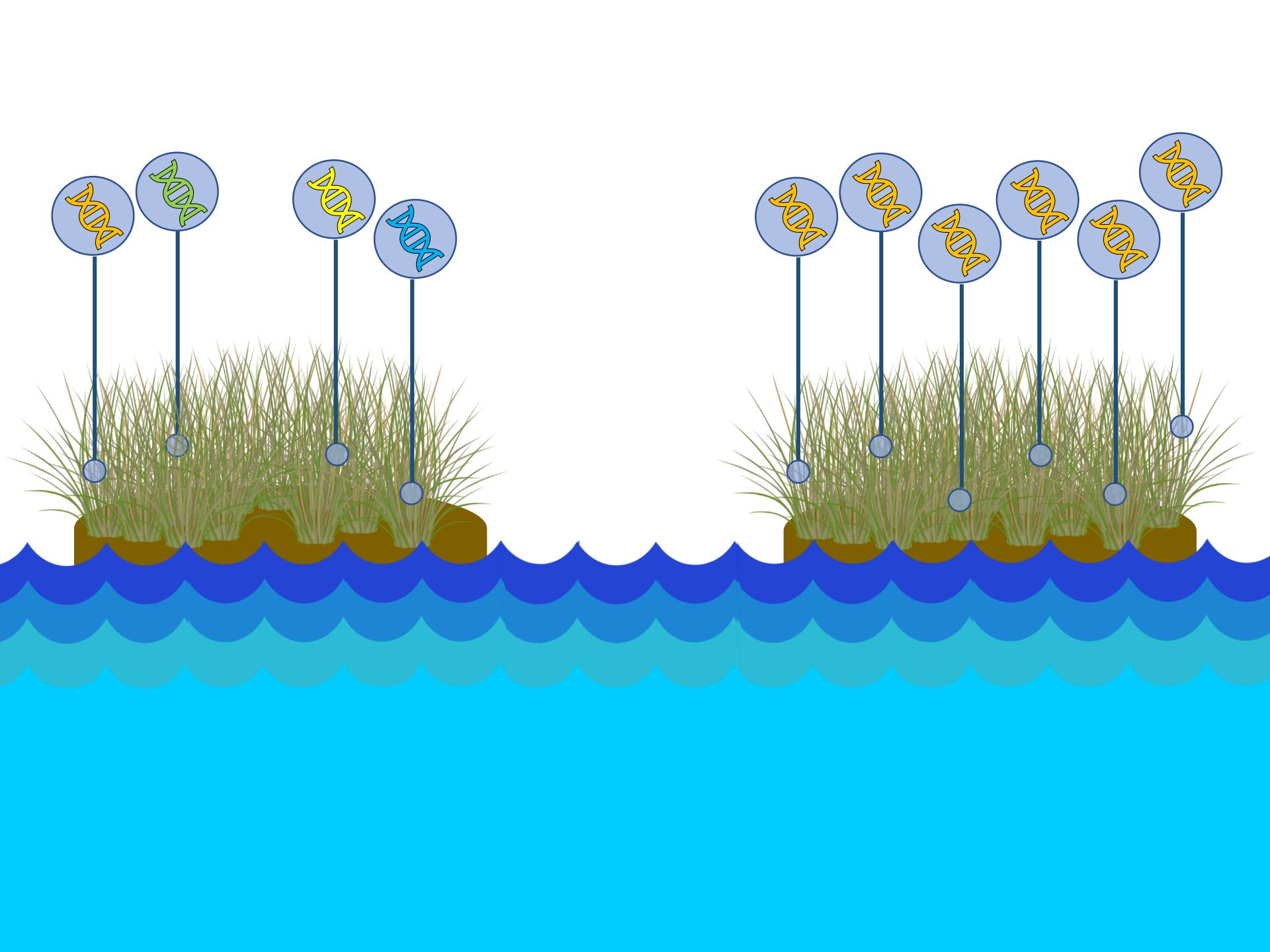


Conservation Genetics

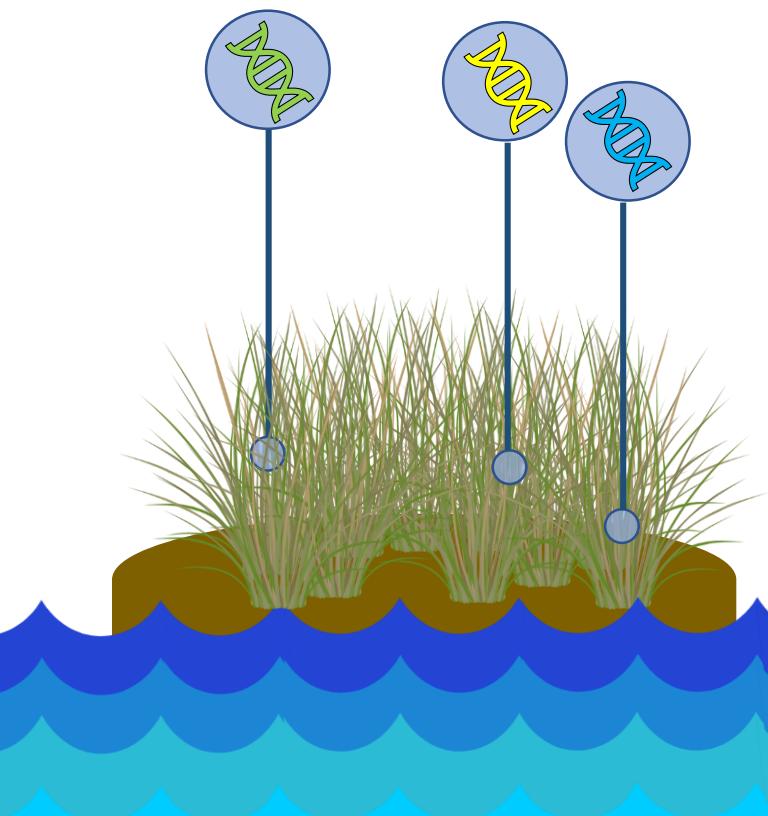


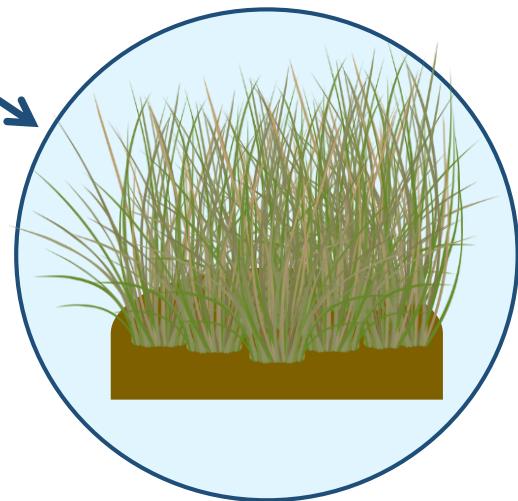
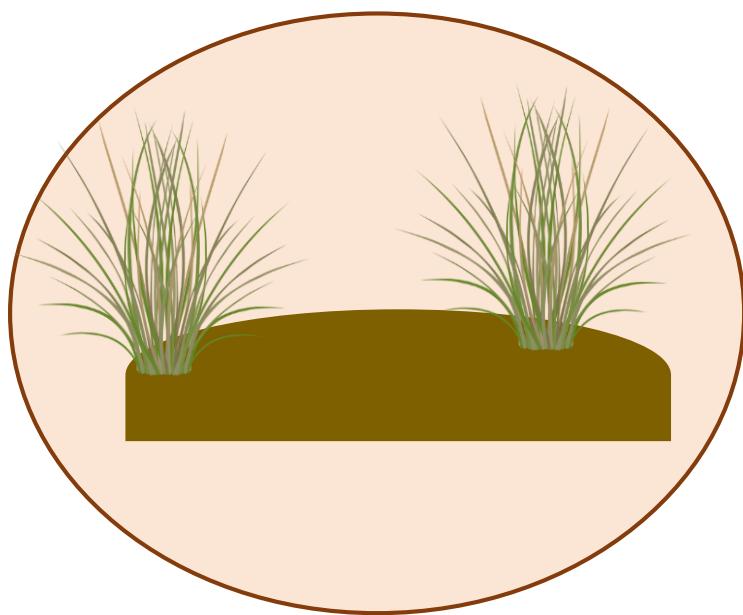
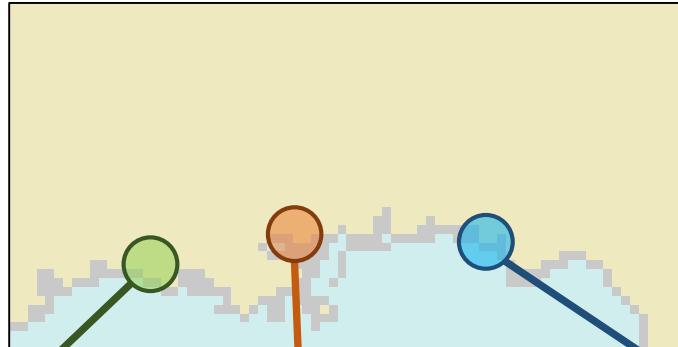
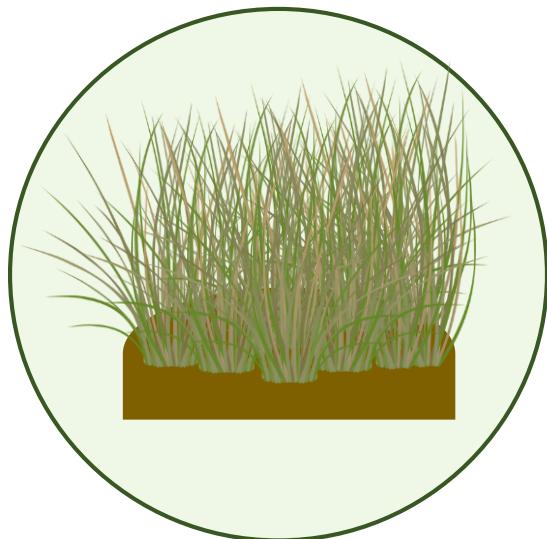


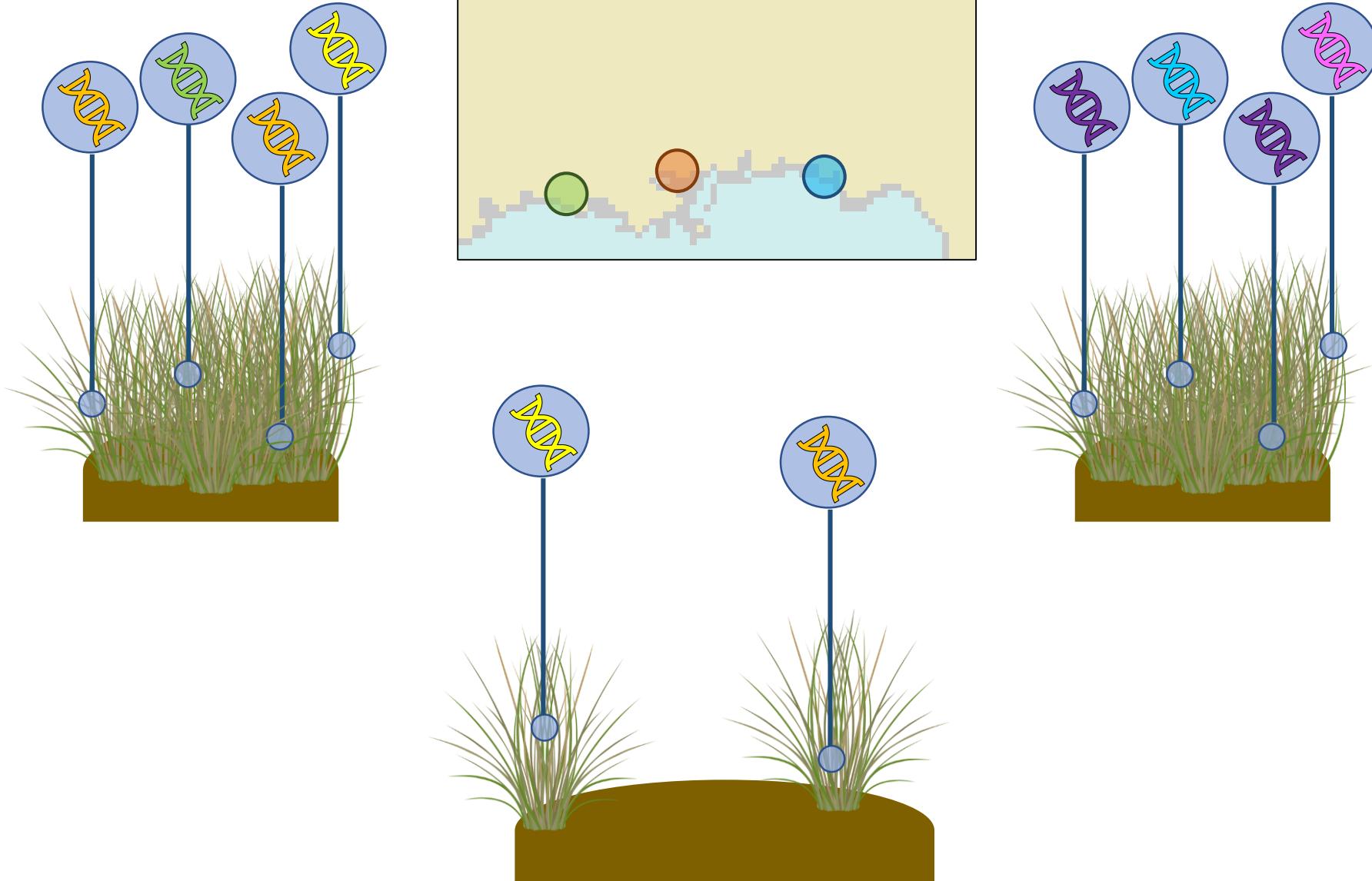




Genetic Diversity

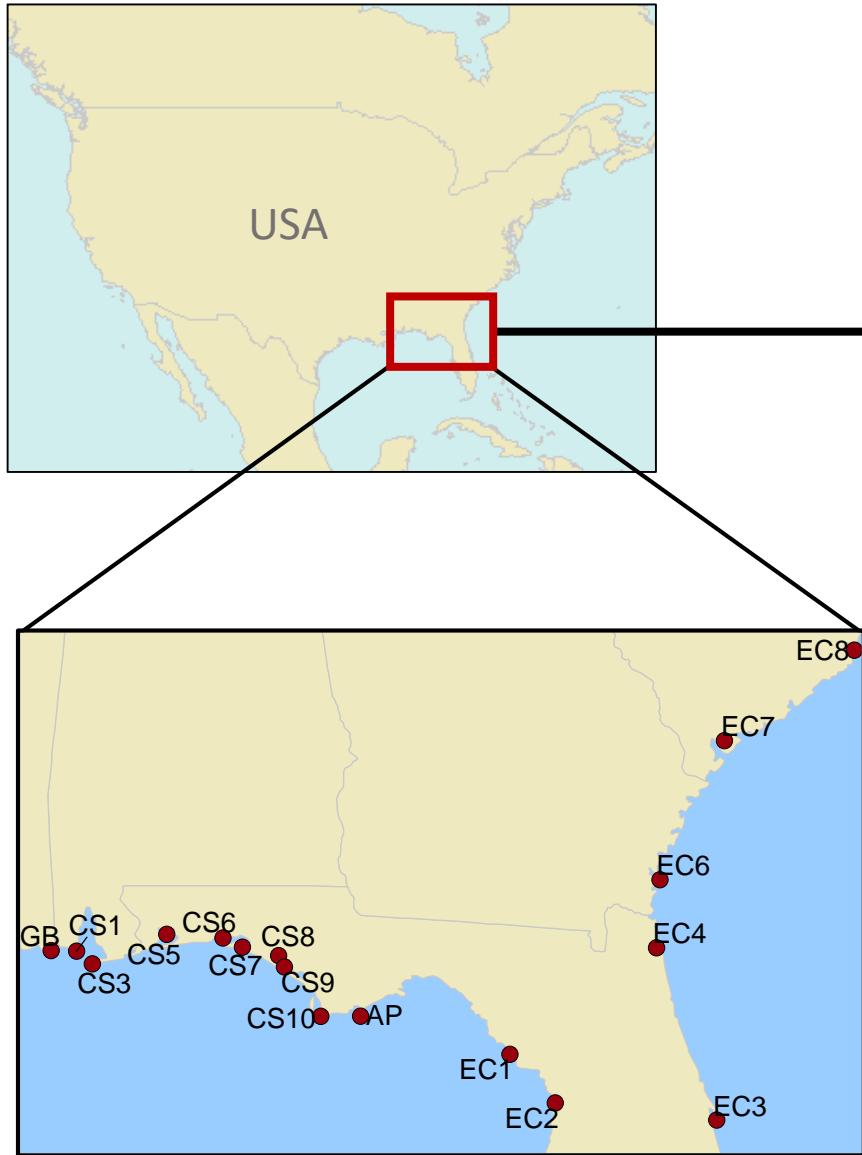




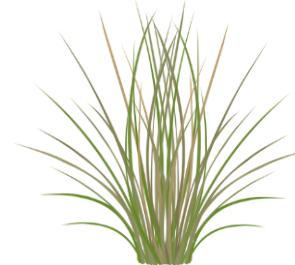


Population Structure

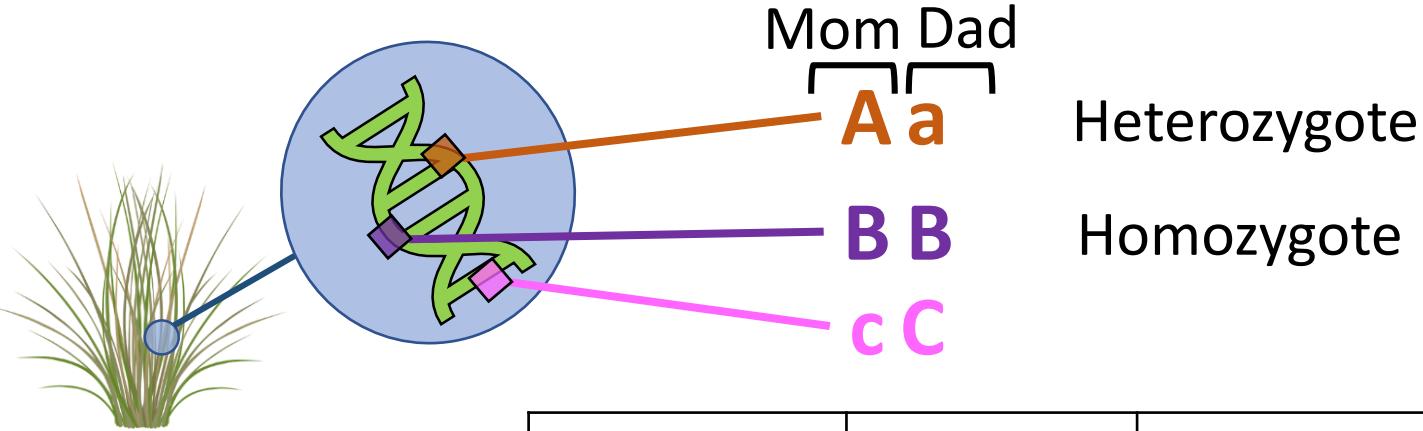
What did I study?



Black Needlerush
Juncus roemerianus
“Juncus”

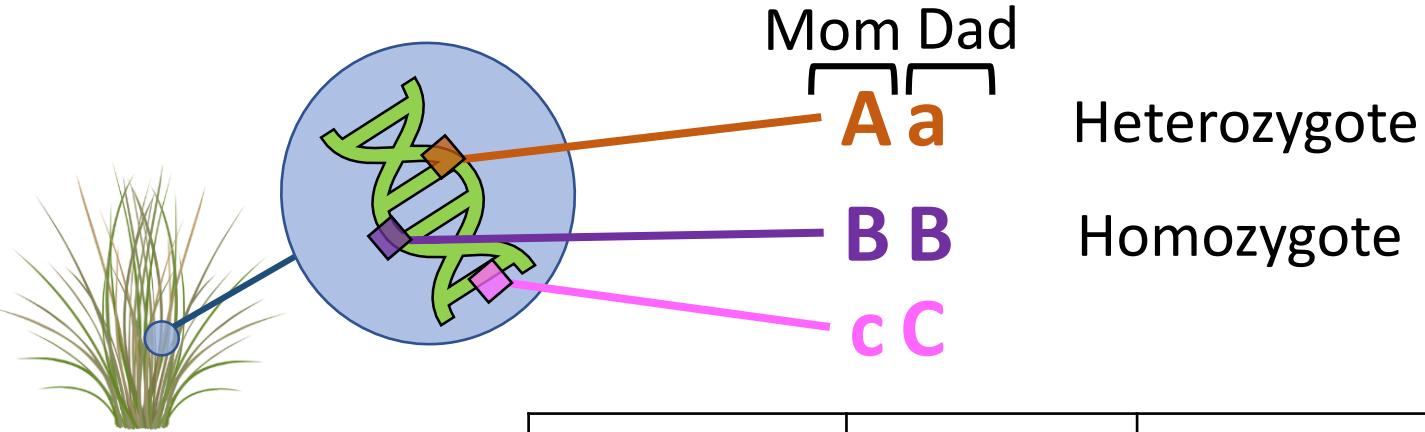


How do we look at genetics?



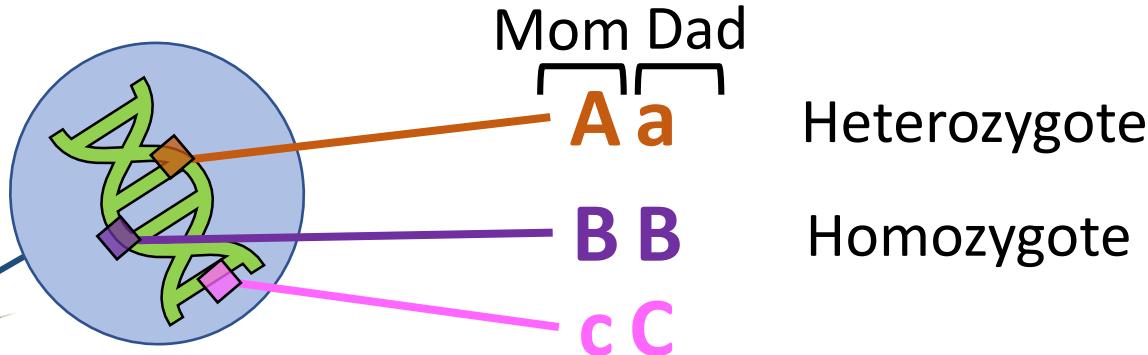
	Locus 1	Locus 2	Locus 3
Sample 1	112:124	100:100	222:226
Sample 2	112:112	130:135	230:230
Sample 3	124:132	130:135	222:238

How do we look at genetics?



	Locus 1	Locus 2	Locus 3
Sample 1	A:C	T:T	A:A
Sample 2	A:A	T:G	A:T
Sample 3	C:C	T:G	T:T

How do we look at genetics?



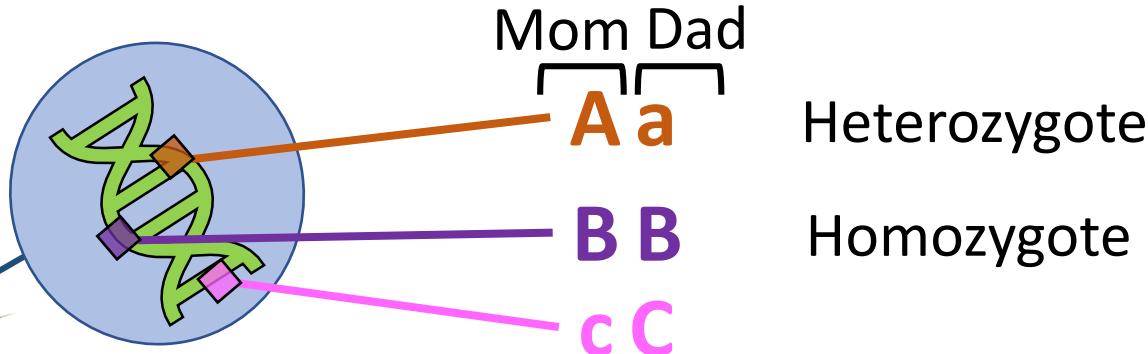
	Locus 1	Locus 2	Locus 3
Sample 1	A:C	T:T	A:A
Sample 2	A:A	T:G	A:T
Sample 3	C:C	T:G	T:T

Genetic Diversity

Observed Heterozygosity
proportion of heterozygotes

Allelic Richness
of alleles/# of loci

How do we look at genetics?



	Locus 1	Locus 2	Locus 3
Sample 1	A:C	T:T	A:A
Sample 2	A:A	T:G	A:T
Sample 3	C:C	T:G	T:T

Population Structure

Principal Components Analysis (PCA)
group samples by similarity in allele frequencies

How do we calculate genetic diversity & population structure?

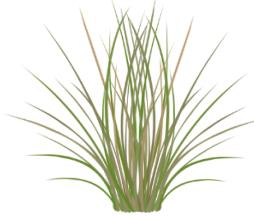
- By hand or in Excel?

How do we calculate genetic diversity & population structure?

- By hand or in Excel?

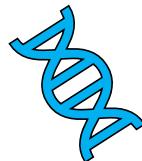
- Numbers game

850 samples



x

19 genetic markers



= 16,150 data points

- Different (and sometimes complicated) equations

How do we calculate genetic diversity & population structure?

- By hand or in Excel?

- Numbers game

850 samples



x

19 genetic markers



= 16,150 data points

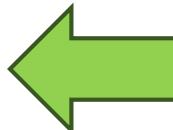
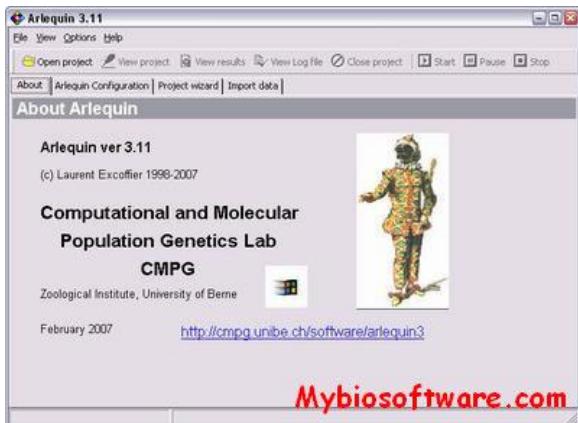
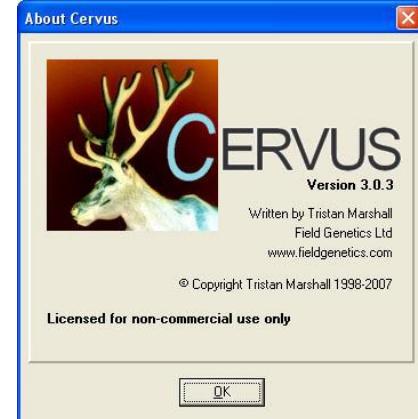
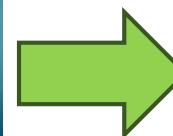
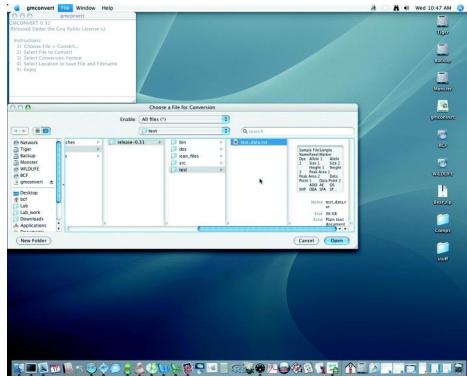
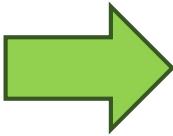
3,000 genetic markers = 2,550,000 data points

- Different (and sometimes complicated) equations

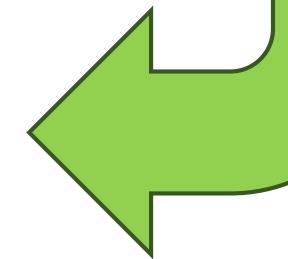
How do we calculate genetic diversity & population structure?

- ~~By hand or in excel?~~
- Using specific software

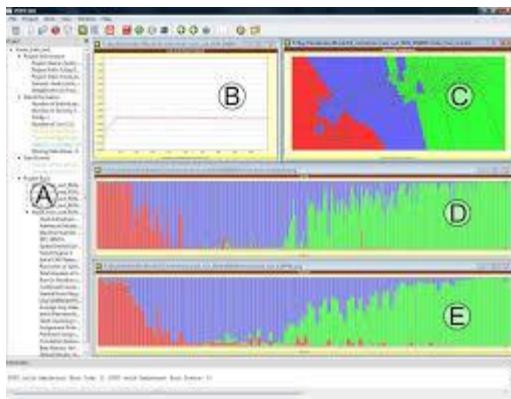
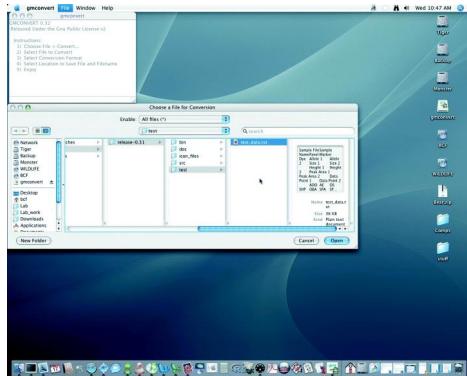
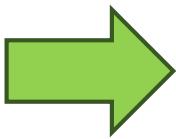
How do we calculate genetic diversity & population structure?



Nodes	Nodes of Web Version	Help
1. Faculty Recruiting Form	Upgraded to Genepop 4.2 compiled binary form accessible provided by François Rousset	Genepop 4.2 Help
2. Lecture Slides/Handouts	Upgraded to Genepop 4.2 compiled binary form accessible provided by François Rousset	Genepop 4.2 Help
3. Resources	Upgraded to Genepop 4.2 compiled binary form accessible provided by François Rousset	Genepop 4.2 Help
4. Site Statistics	Upgraded to Genepop 4.2 compiled binary form accessible provided by François Rousset	Genepop 4.2 Help
5. Basic Instructions for Using Genepop	Upgraded to Genepop 4.2 compiled binary form accessible provided by François Rousset	Genepop 4.2 Help



How do we calculate genetic diversity & population structure?



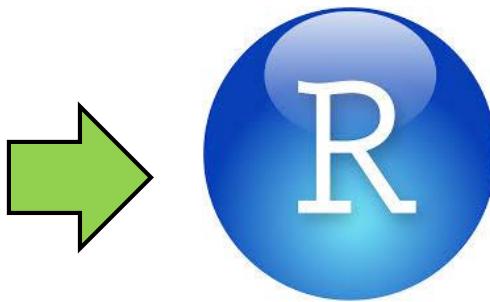
How do we calculate genetic diversity & population structure?

- ~~By hand or in excel?~~
- Using specific software
 - Long process that must be repeated if data updated
 - Each software has a different data format
 - Most GUIs have learning curves
 - Different software uses different equations
 - Output is in different formats
 - Still may not get all of the metrics you want!

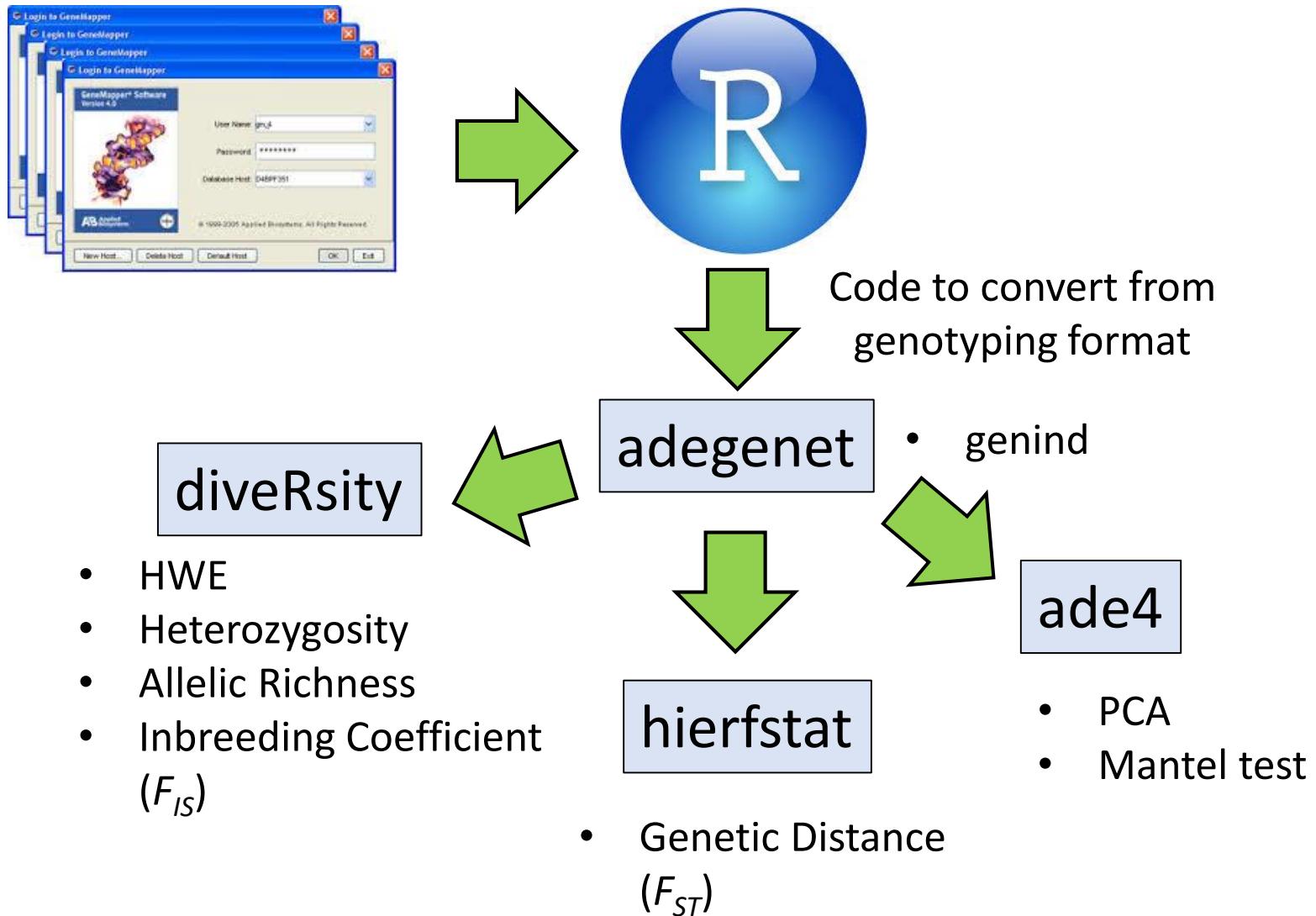
How do we calculate genetic diversity & population structure?

- By hand or in excel?
- Using specific software
- Using specific packages in R

How do we calculate genetic diversity & population structure?



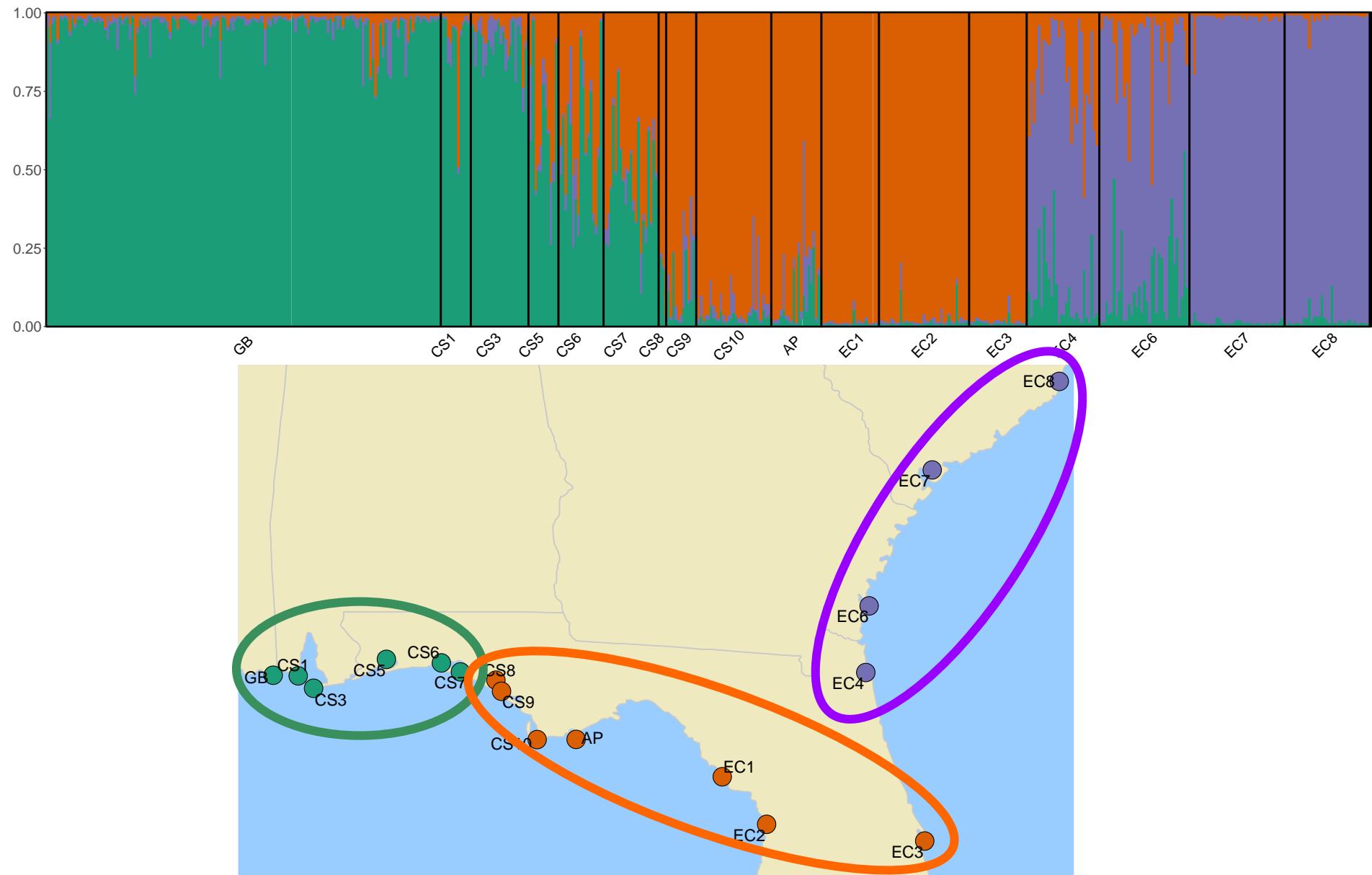
How do we calculate genetic diversity & population structure?



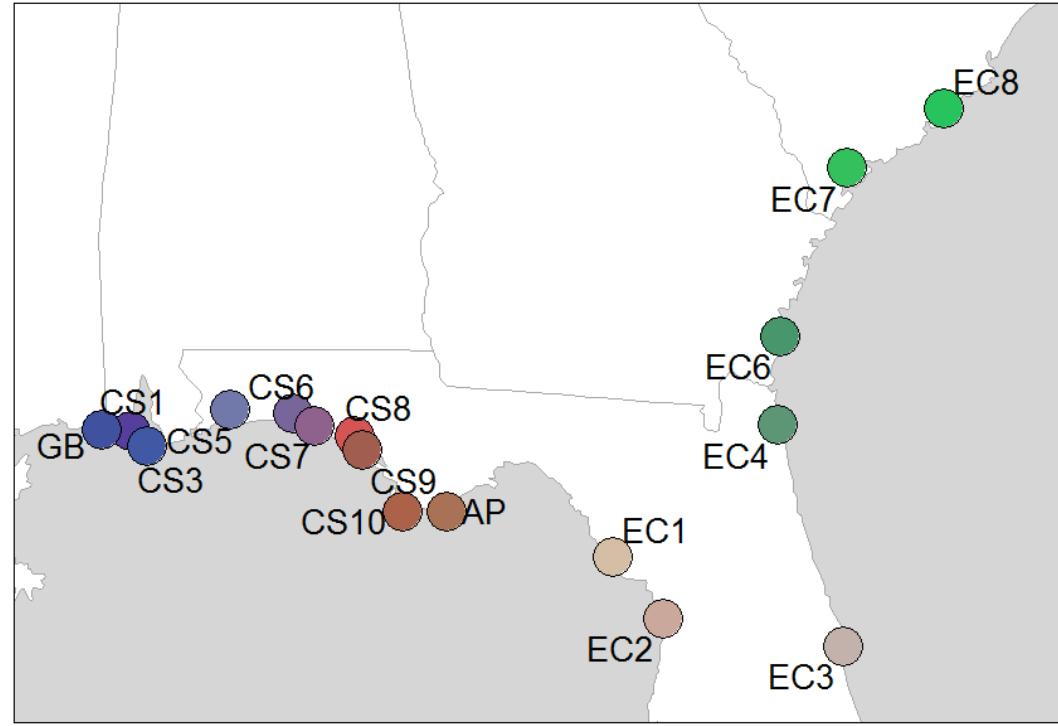
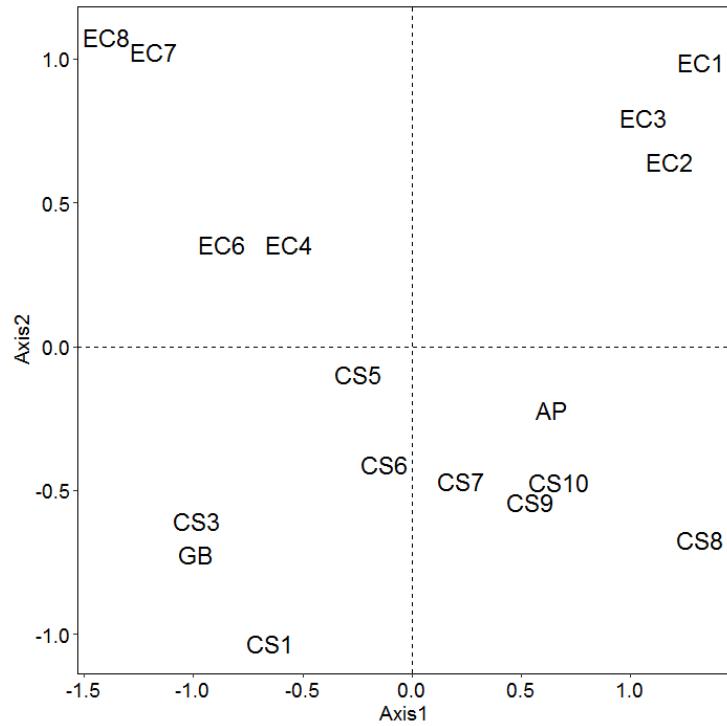
How do we calculate genetic diversity & population structure?

- By hand or in excel?
- Using specific software
- Using specific packages in R
 - Most packages use the same format
 - Multiple metric equations explicitly given
 - Output in csv format
 - Go straight to data visualization
 - Creates easy to follow, ubiquitous, and publishable code

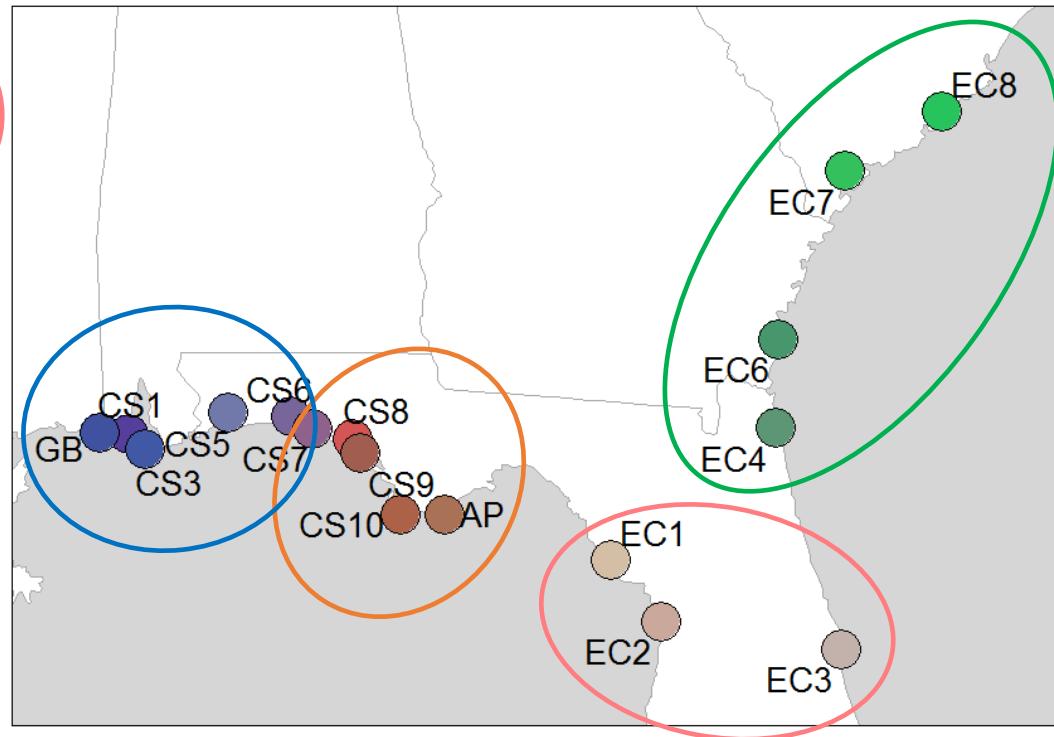
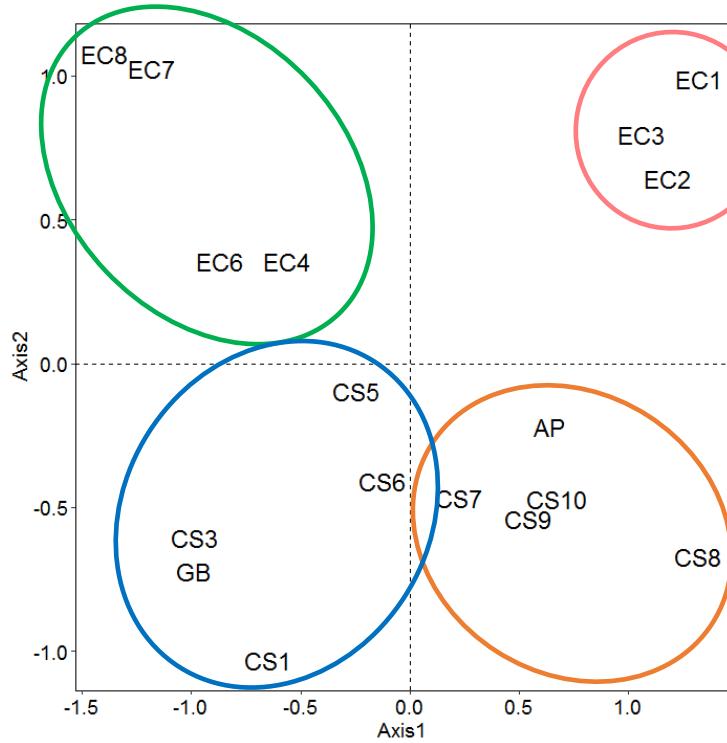
Genetic Structure without ade4



Genetic Structure with ade4



Genetic Structure with ade4



Data Visualization without ggplot2

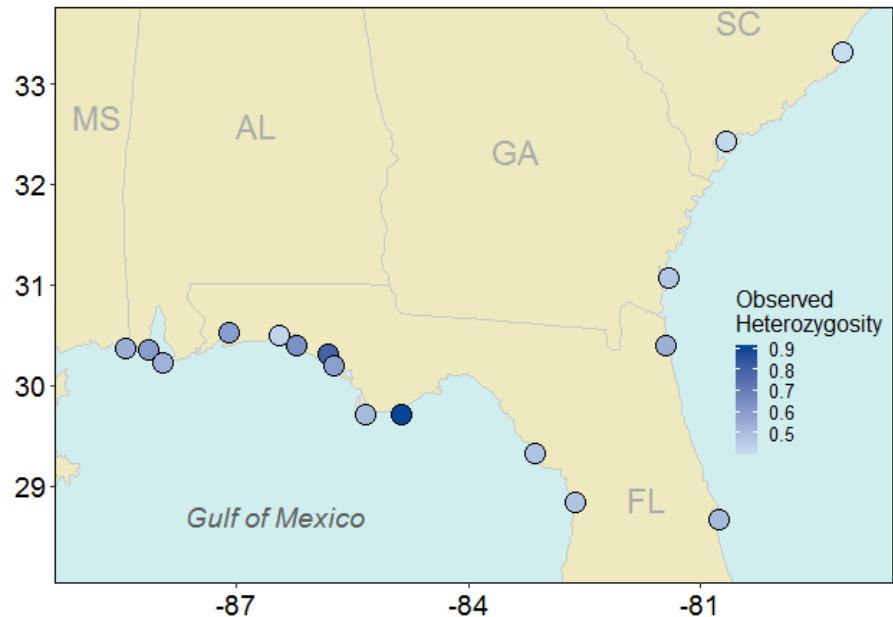
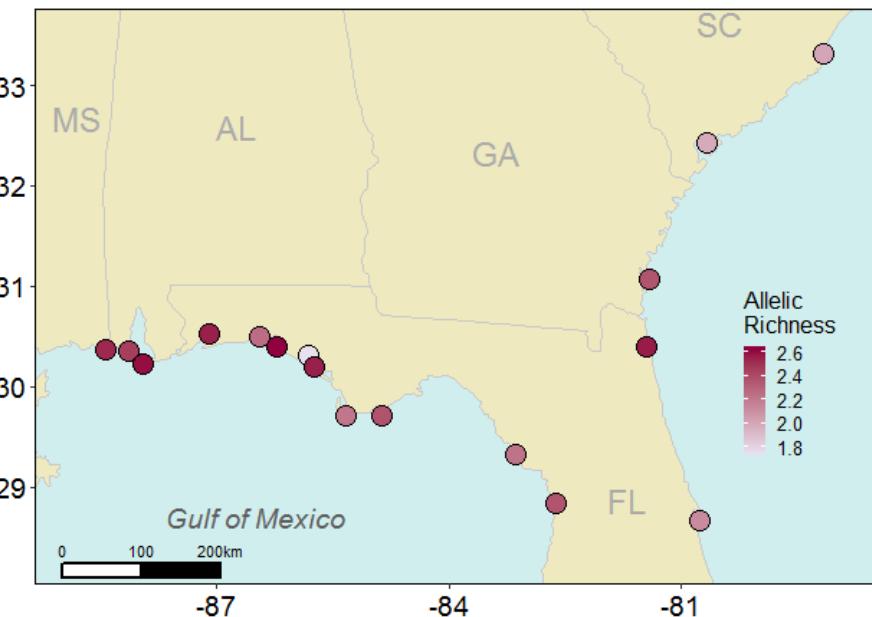
Table 1 Clonal and genetic diversity measures for samples grouped by site and coast

	Clonal diversity						Genetic diversity						
	N	MLG	G _D	H	λ	E.5	N	A _D	A _R	H _O	H _E	PAS	F _{IS}
GB—Moss Point, MS	221	142	0.64	4.788	0.995	0.824	304	6.22	2.53	0.54	0.56	17	0.02
CS1—Heron Bayou, AL	26	10	0.36	1.955	0.843	0.706	30	3.33	2.45	0.6	0.55	0	-0.11
CS3—Gulf Shores, AL	28	22	0.78	3.016	0.981	0.91	30	4.22	2.6	0.53	0.58	1	0.08
CS5—Avalon Beach, FL	26	13	0.48	2.451	0.941	0.902	30	3.77	2.56	0.6	0.61	1	0.01
CS6—Niceville, FL	26	15	0.56	2.611	0.957	0.913	30	3.5	2.26	0.43	0.54	0	0.17
CS7—Santa Rosa Beach, FL	26	22	0.84	2.992	0.979	0.841	30	4.11	2.62	0.64	0.6	2	-0.11
CS8—West Bay, FL	25	3	0.08	0.708	0.440	0.71	30	2.64	1.75	0.79	0.67	0	-0.24
CS9—Panama City Beach, FL	27	12	0.42	2.198	0.892	0.759	30	3.83	2.56	0.59	0.56	1	-0.05
CS10—Cape San Blas, FL	30	30	1.00	3.401	1.000	1	30	3.88	2.22	0.51	0.5	1	0
AP—Apalachicola, FL	27	21	0.77	2.949	0.977	0.882	32	4.24	2.37	0.9	0.53	1	0.07
EC1—Suwanee, FL	24	13	0.52	2.021	0.797	0.494	36	4.06	2.23	0.48	0.48	5	-0.05
EC2—Crystal River, FL	32	30	0.94	3.379	0.996	0.968	40	4.17	2.36	0.48	0.52	3	0.04
EC3—Merritt Island NWR, FL	33	17	0.50	2.643	0.945	0.84	40	3.59	2.12	0.51	0.47	1	-0.14
EC4—Fanning Island, FL	35	26	0.74	3.088	0.970	0.777	40	4.22	2.57	0.54	0.56	1	0.02
EC6—Jekyll Island, GA	36	33	0.91	3.468	0.996	0.961	39	3.83	2.36	0.47	0.5	1	0.05
EC7—Beaufort, SC	34	34	1.00	3.526	1.000	1	38	3.18	1.99	0.42	0.42	0	-0.03
EC8—Awendaw, SC	36	30	0.83	3.323	0.987	0.894	40	3.47	2.01	0.41	0.42	1	0.01
Average/total	76.89	52.5	0.67	2.854	0.923	0.846	849	3.9	2.33	0.56	0.53	2.12	-0.02
SE (\pm)	11.312	7.460	0.062	0.212	0.033	0.031		0.179	0.061	0.031	0.016	0.977	0.023
Gulf coast	518	332	0.64	5.56	0.996	0.671	652	8.06	6.96	0.52	0.65	32	0.17
Atlantic coast	174	140	0.80	4.83	0.996	0.836	197	5.33	5.26	0.45	0.53	4	0.14

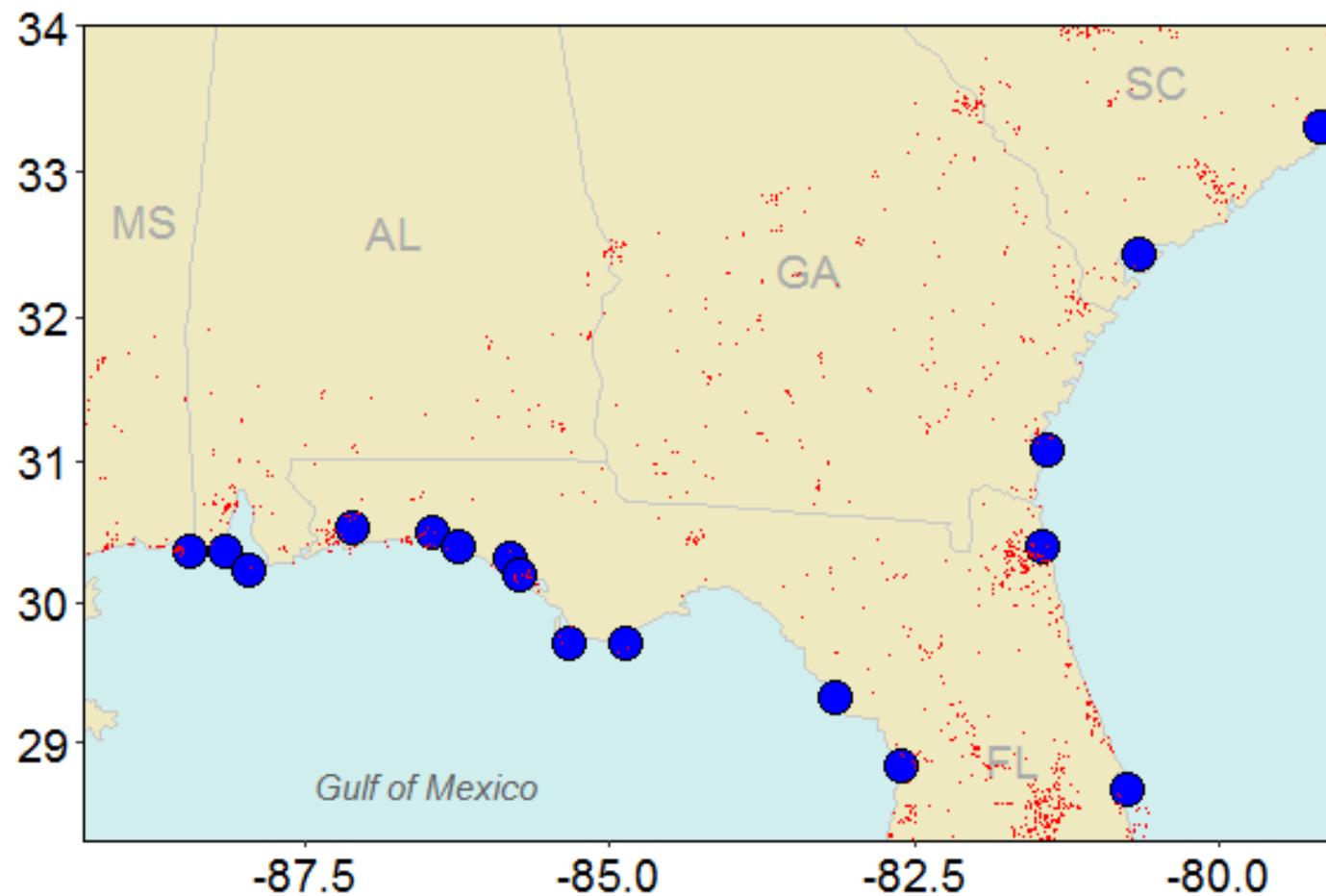
Clonal diversity N reflects samples that had no missing data across a subset of 12 markers, and genetic diversity N reflects all samples collected at each site. Ten sites (GB–AP) were collected as part of another study (Tumas et al. 2018)

NWR National Wildlife Refuge, N total number of samples, MLG number of unique multilocus genotypes, G_D genotypic diversity, H Shannon–Wiener Index of MLG diversity, λ Simpson's Index corrected for sample size, E.5 genotypic evenness, A_D allelic diversity, A_R allelic richness, H_O observed heterozygosity, H_E expected heterozygosity, PAS number of private alleles, F_{IS} inbreeding coefficient

Data Visualization with ggplot2



Effect of developed land on genetic diversity



■ Developed Land

Effect of developed land on genetic diversity



AlabamaToday

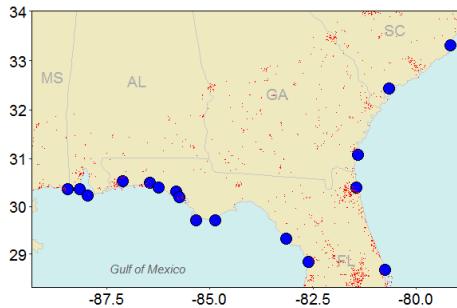


FITSNews



letsroam

Spatial Analysis

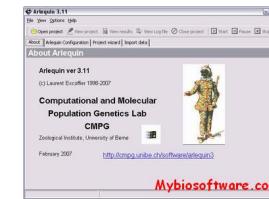
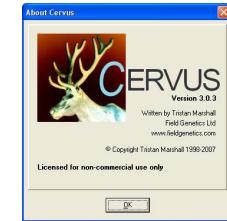
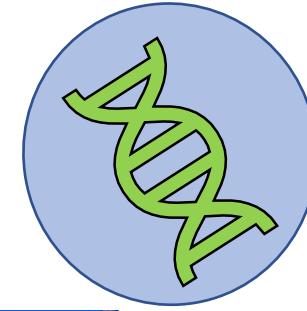


ArcGIS

FRAGSTATS

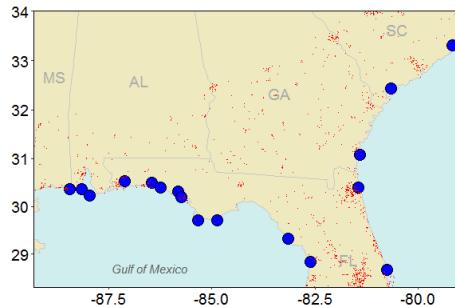
**Traditional Landscape
Ecology**

Genetic Analysis

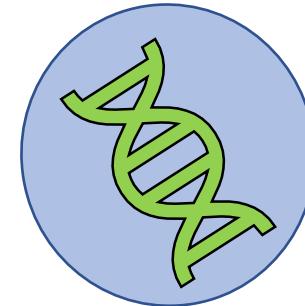
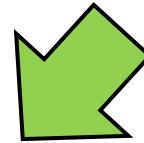
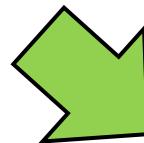


**Traditional Population
Genetics**

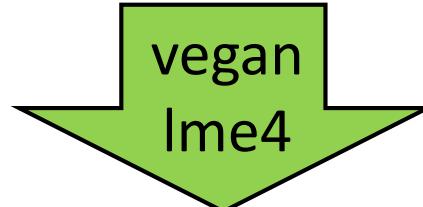
Spatial Analysis + Genetic Analysis



- raster
- rgdal
- sp



- adegenet
- diveRsity



developed land significantly impacts genetic diversity

Summary

- Genetic analyses are key for successful restoration
- Producing a few summary stats requires multiple software programs and data formats
- R provides a streamlined way to analyse genetic data
- Data capabilities in R open up potential for other cross disciplinary analyses
- Genetic analysis R code could be easily shared amongst researchers and practitioners

Questions?



Packages:

- adegenet
- poppr
- hierfstat
- ade4
- diversity
- raster
- rgdal
- sp

Acknowledgements

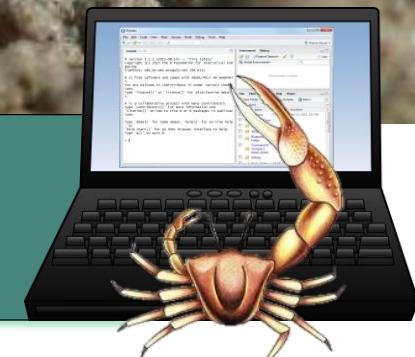


Warnell School of Forestry
& Natural Resources
UNIVERSITY OF GEORGIA



Feel free to contact me:

 @HayleyTumas



abc