

Evolutionary cause of phylogenetic diversity between East Asian and East North American disjunct forest communities

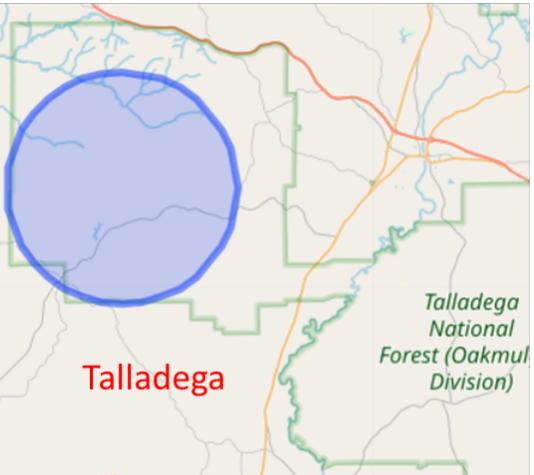
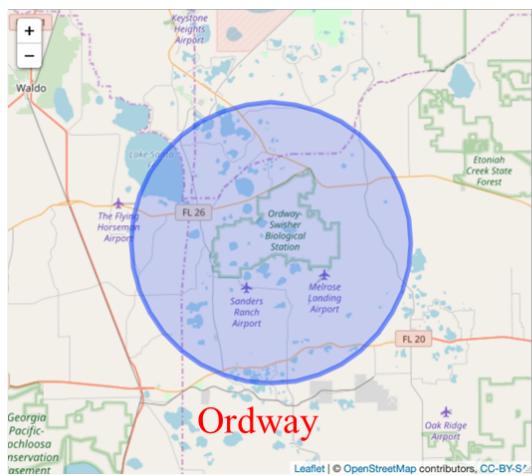
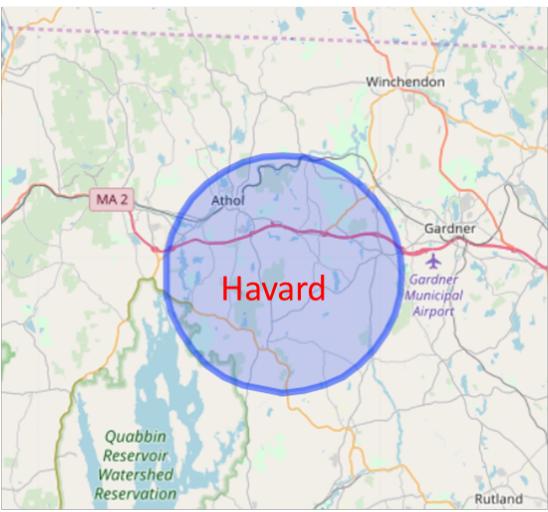
Miao Sun, Hanyang Lin

June 06, 2019

Species list

 Ceweeta floristic list.xlsx	 May 14, 2018 at 09:59	46 KB
 Havard Forest floristic list.xlsx	 Mar 30, 2018 at 12:39	209 KB
 Mountain Lake floristic list.xlsx	 Mar 30, 2018 at 12:42	82 KB
 Ordway floristic list.xlsx	 Mar 30, 2018 at 12:01	190 KB
 Talladega floristic list.xlsx	 Mar 30, 2018 at 10:33	28 KB
 White Mountain floristic list revised.xlsx	 Mar 30, 2018 at 10:41	62 KB

Too less compared with Chinese list

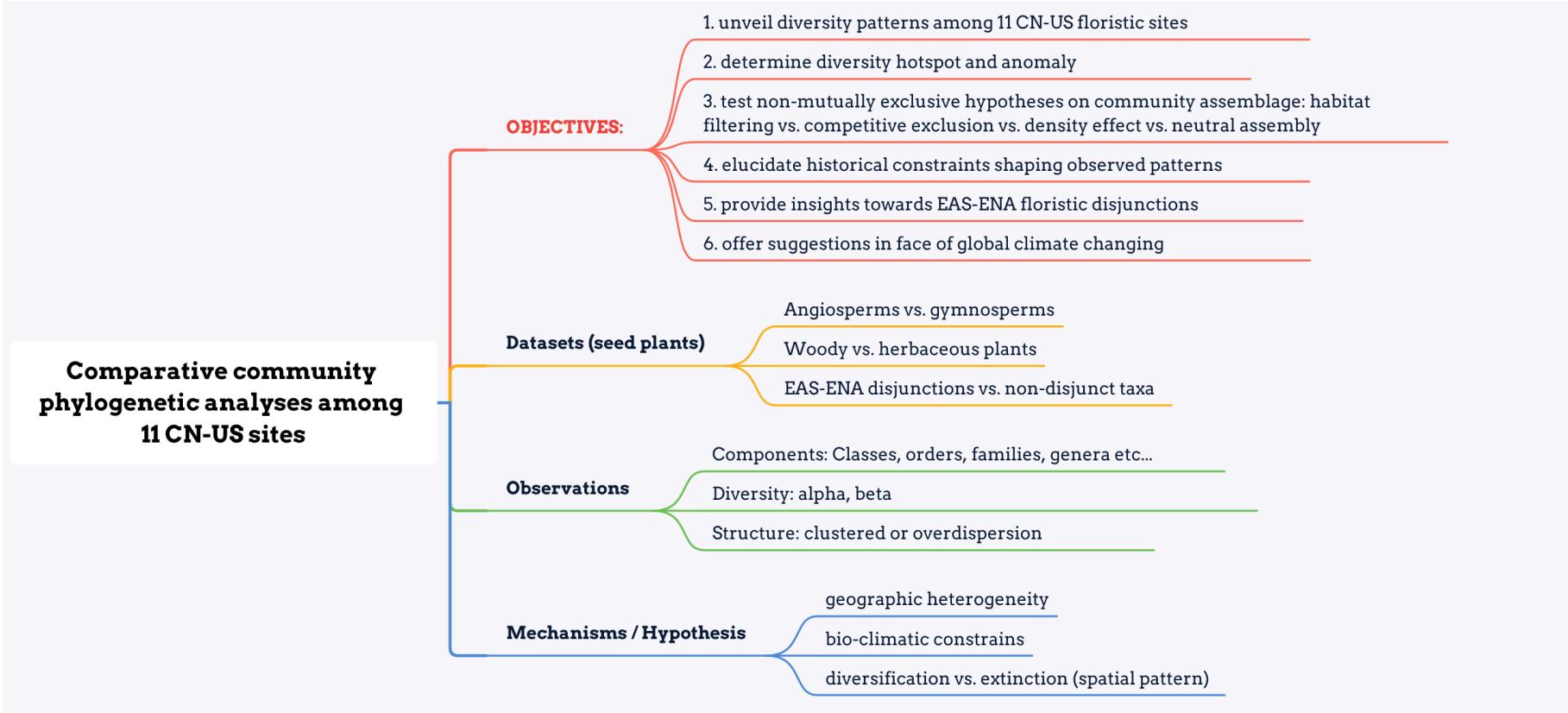


Centered coordinates of each site
Seed plant accepted names
IDigBio + GBIF + old_species_list

Site	Area (km ²)	Elevation	ntaxa	ntaxa_update	Latitude(°)	Longitude(°)
White_Mountain	15.66	232	214	254	44.06388	-71.28731
Havard	49	351	800	897	42.5369	-72.17266
Mountain_Lake	12.6	1160	296	381	37.375654	-80.52214
Coweeta	16.26	914	583	681	35.052555	-83.45086
Talladega	53	135	349	533	32.923284	-87.4203
Ordway	38.44	45	632	749	29.68927	-81.99343
Changbai	1907.81	2744	1141	1140	42.42	128
Dongling	60	2303	773	773	40.06	115.50
Shennong	704.67	3106	1559	1559	31.61	110.31
Tianmu	43	1505	1188	1188	30.42	119.44
Gutian	81.07	1258	1209	1209	29.30	118.13

Adding in more species

Manuscript structure



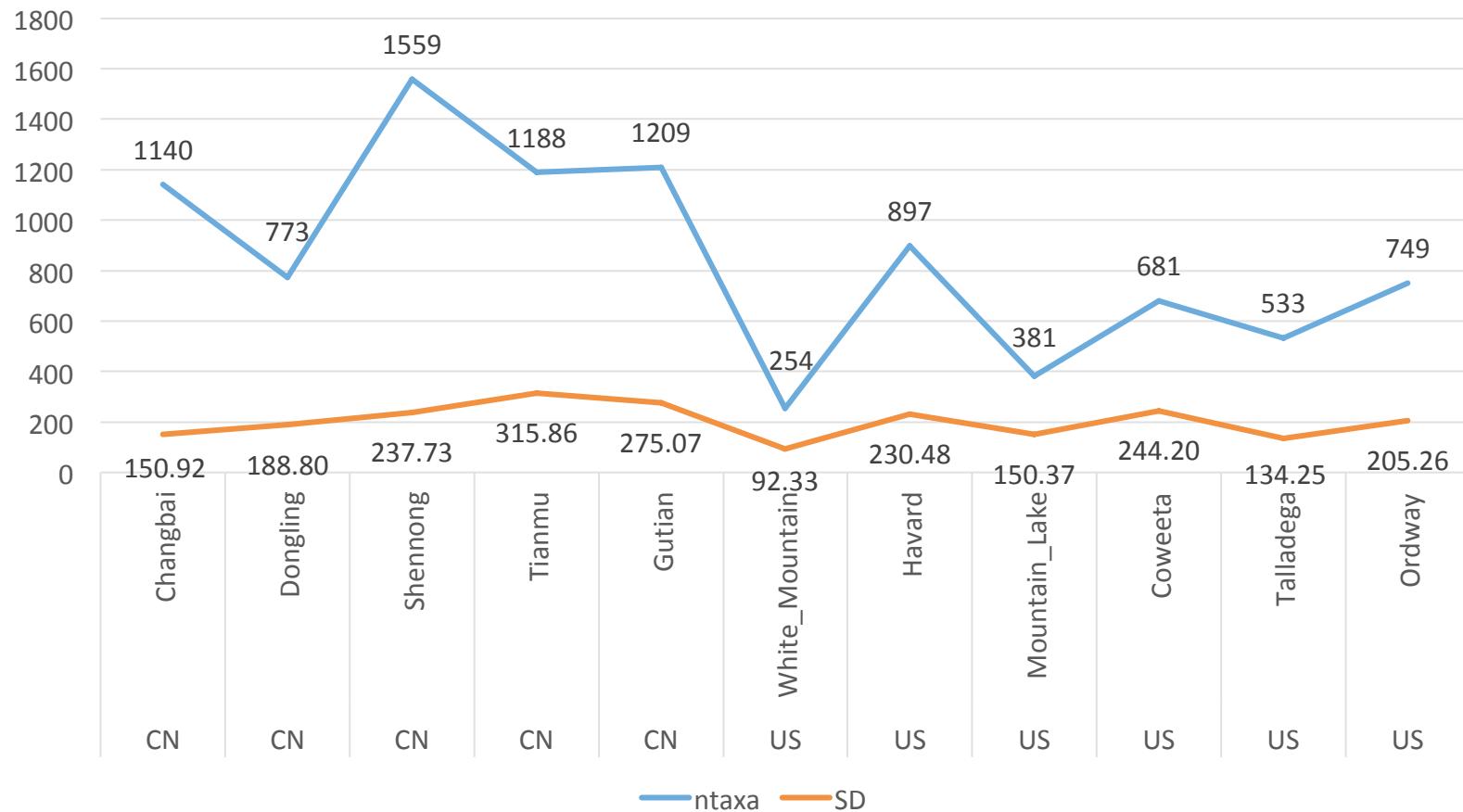
Questions:

- Which floristic region, i.e., EA or ENA, harbors greater species diversity, including species richness (SR) and phylogenetic diversity (PD)?
- Among 11 EAS-ENA sites, where is PD hotspot among US-CN 11 sites? Following any patterns (Latitude, lifeform, disjunct)?
- How angiosperms vs. gymnosperms, woody vs. herbaceous, or EA-ENA disjunction vs Non- disjunction contribute to PD (α & β) of EA and ENA forest communities?
- Insights for community structure and community assemblage process at regional and local scale, respectively?

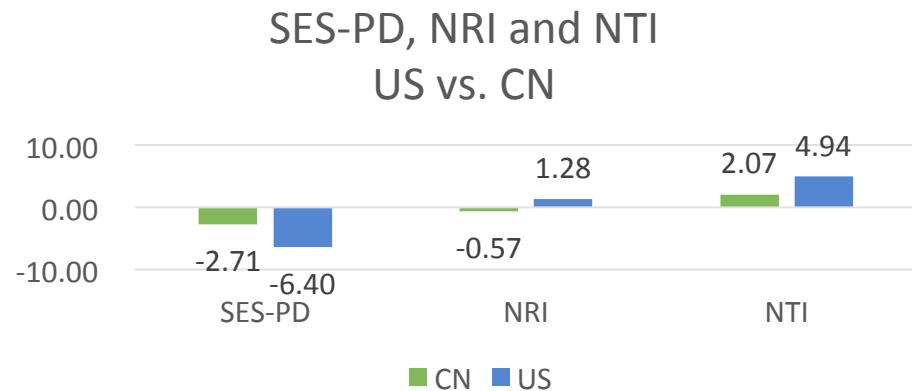
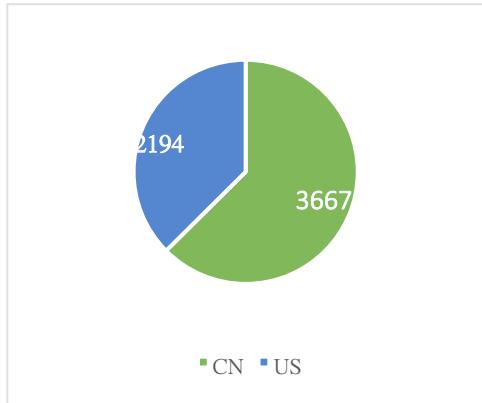
**Q1: Which floristic region harbors greater species diversity,
including species richness (SR) and phylogenetic diversity
(PD)?**

Species richness (SR)

Number of taxa and species density among 11 CN-US sites



CN > US; Shengnong (CN); Harvard (US)

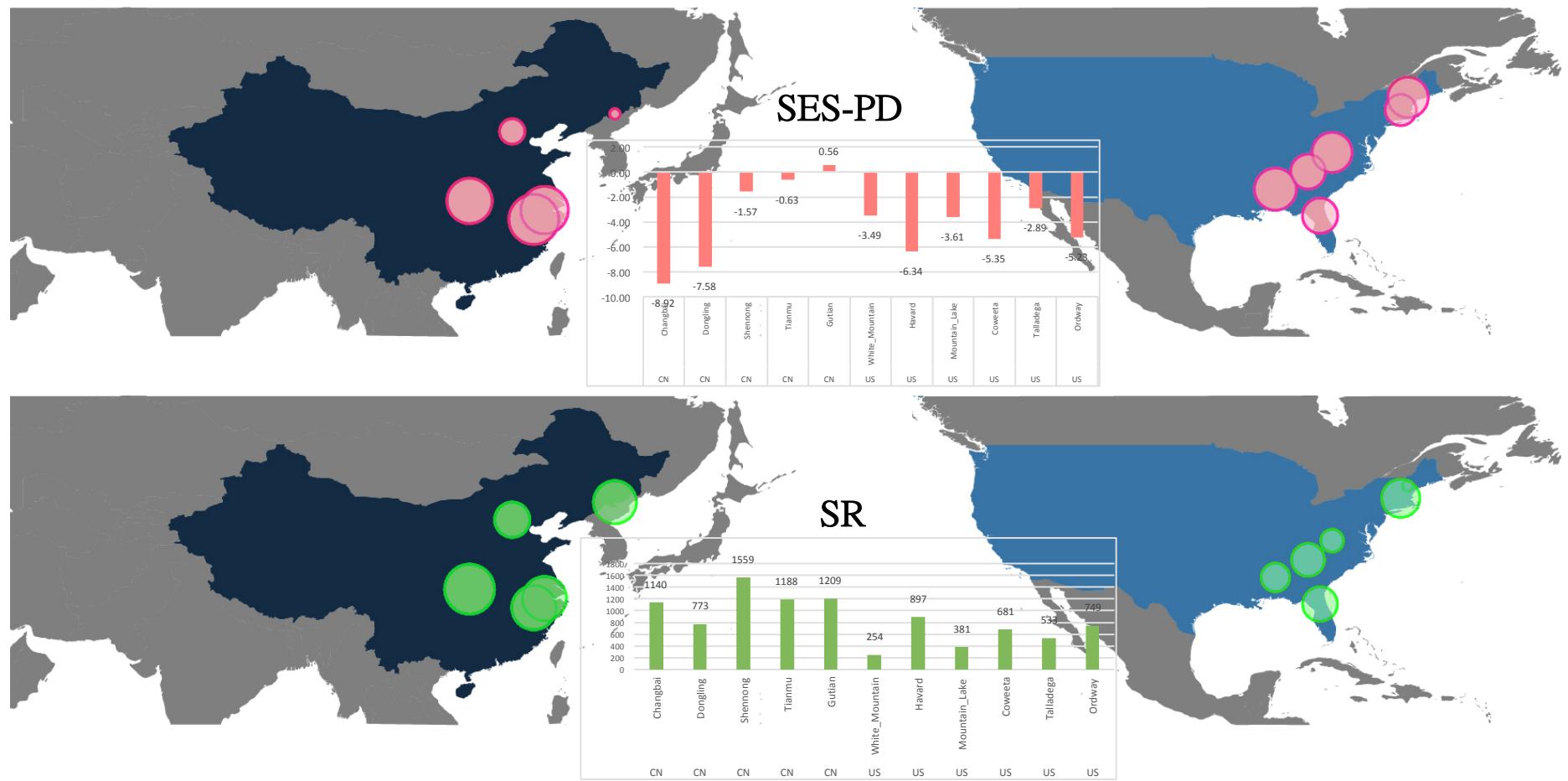


Categories	ntaxa	PD	SES-PD	NRI	NTI
CN	3667	66034.38	-2.71	-0.58	2.08
US	2194	44617.83	-6.40	1.28	4.94

CN > US

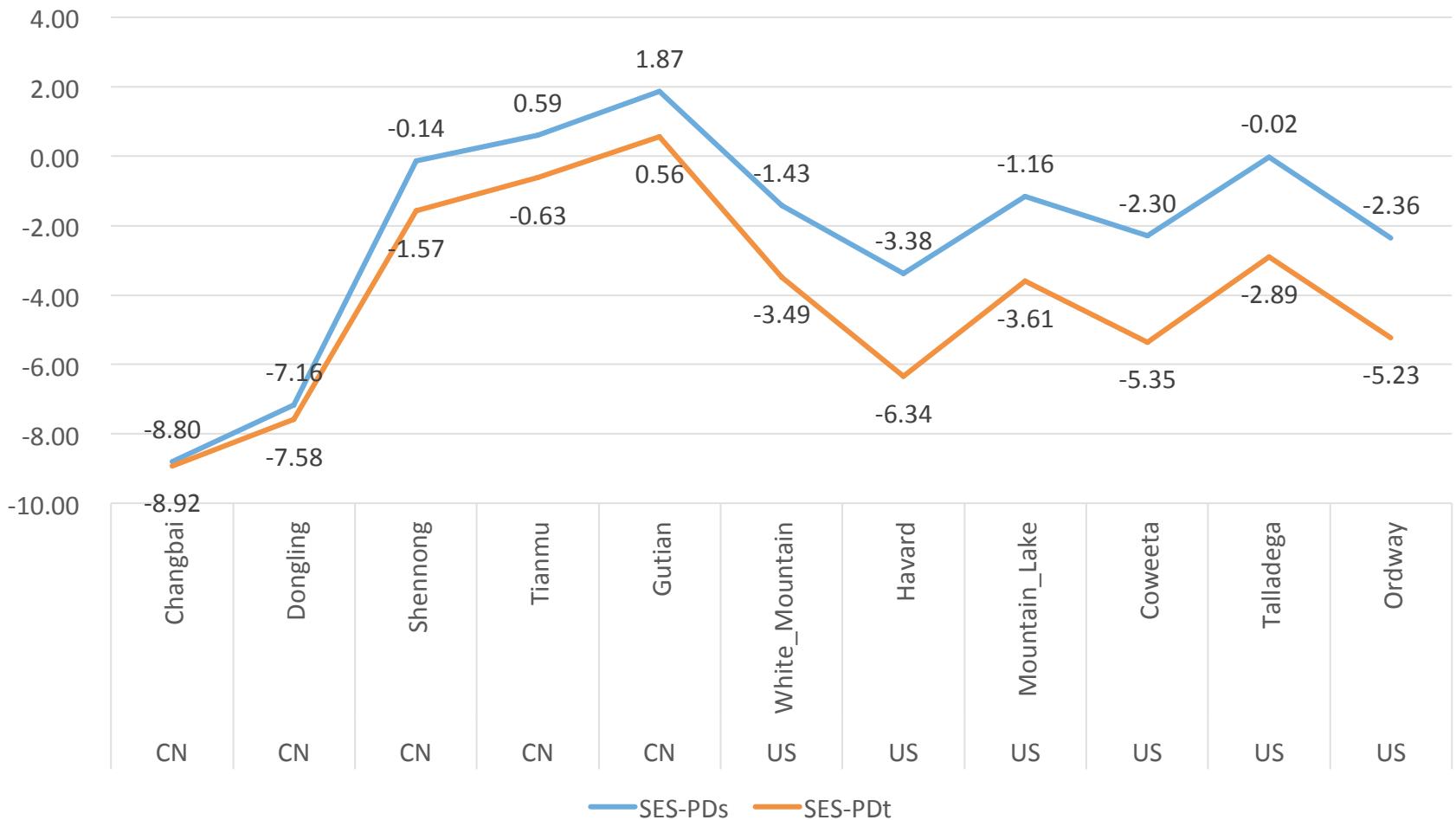
Q2: Among 11 EAS-ENA sites, where is PD hotspot among US-CN 11 sites? Following any patterns (Latitude, lifeform, disjunct)?

α PD --- US-CN 11 sites



CN sites decrease along latitude
Changbai, and White Mountain (SR vs. PD)

SES-PDs and SES-PDt among 11 CN-US sites

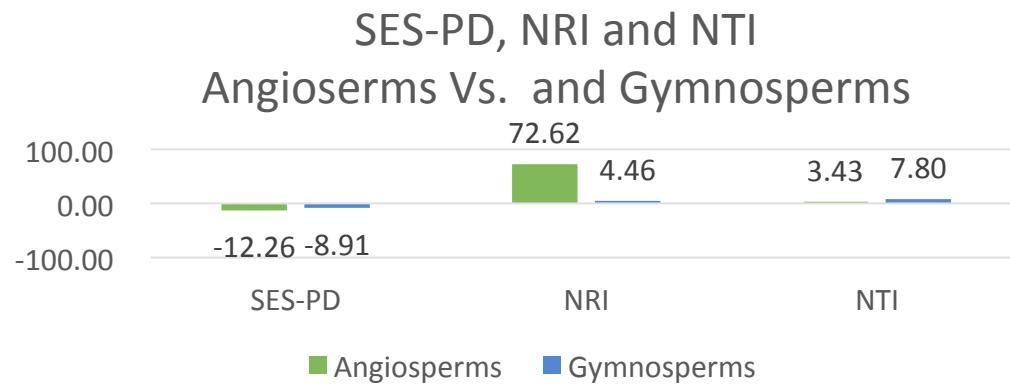
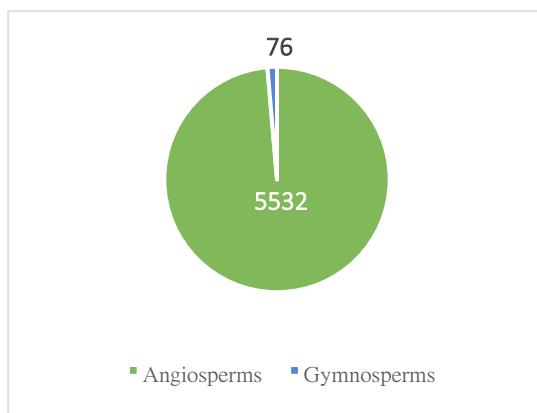
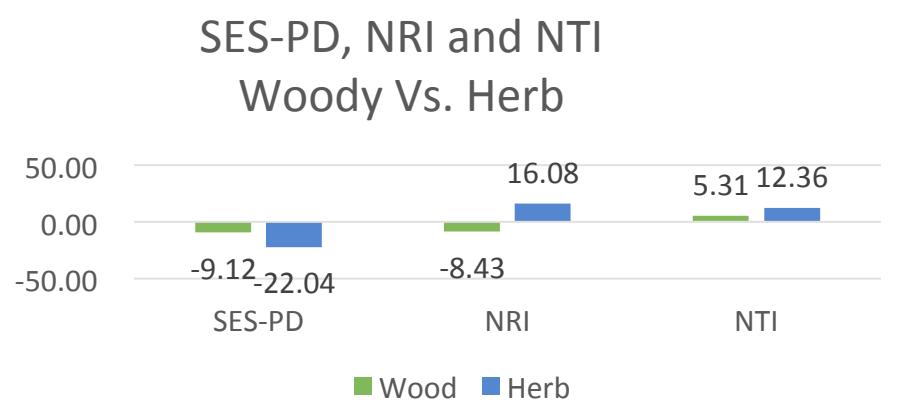
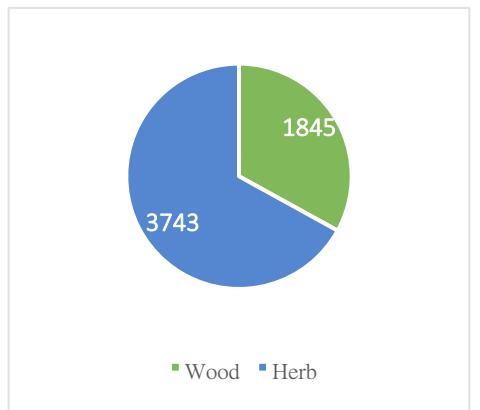
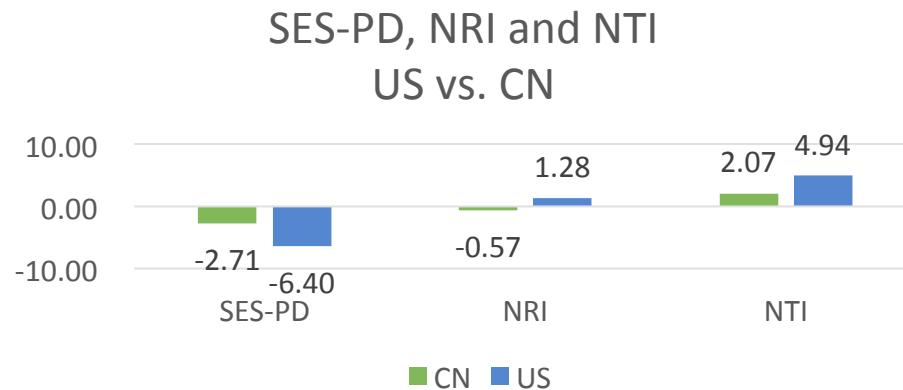
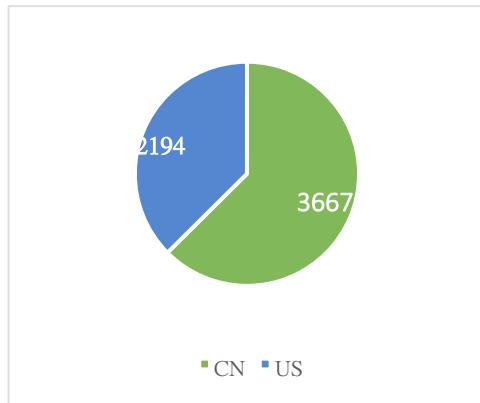


Over all Gutian is hotspot; Talladega is hotspot in US

Site	Area (km ²)	Elevation-max (m)	MAT (°C)	MAP (mm)	ntaxa	SD	SES-PDt	NRIt	NTIt
Changbai	1907.81	2744	2.13	741	1140	150.92	-8.92	1.10	5.50
Dongling	60.00	2303	7.11	493	773	188.80	-7.58	-0.41	5.81
Shennong	704.67	3106	12.94	1156	1559	237.73	-1.57	-0.80	2.44
Tianmu	43	1505	12.92	1487	1188	315.86	-0.63	0.80	1.90
Gutian	81.07	1258	15.91	1963.7	1209	275.07	0.56	-0.89	0.17
White_Mountain	15.66	232	4.8	1300	254	92.33	-3.49	-2.64	2.79
Havard	49.00	351	7.2	1120	897	230.48	-6.34	-0.35	4.82
Mountain_Lake	12.60	1160	9.8	1045	381	150.37	-3.61	-0.69	3.60
Coweeta	16.26	914	11.9	1754	681	244.20	-5.35	2.05	4.18
Talladega	53.00	135	17.1	1425	533	134.25	-2.89	1.17	1.94
Ordway	38.44	45	20.3	1313	749	205.26	-5.23	1.41	4.28

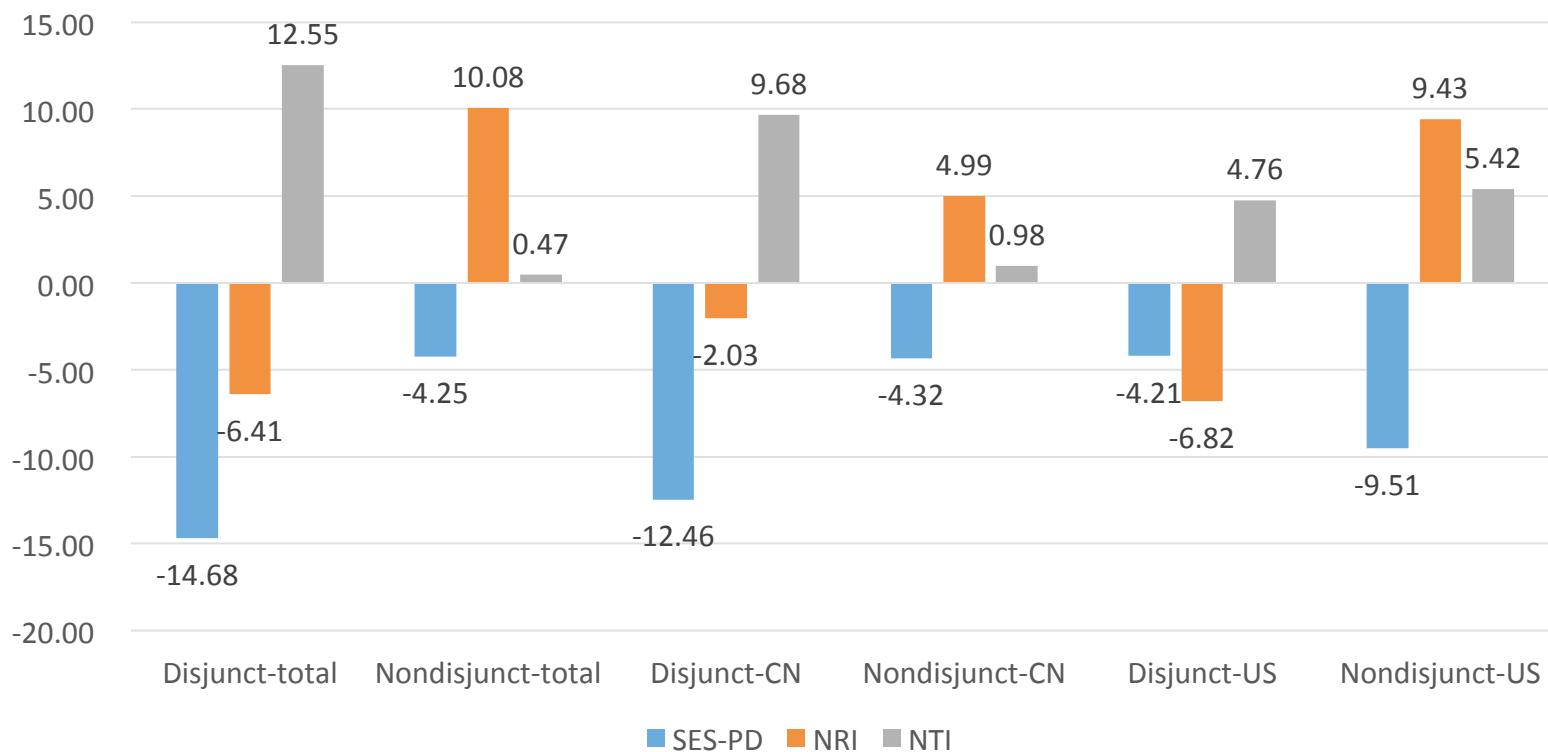
Correlation with env. Variables (small size: 11)

**Q3: How angiosperms vs. gymnosperms, woody vs.
herbaceous, or EA-ENA disjunction vs Non- disjunction
contribute to PD (α & β) of EA and ENA forest communities?**



Latest compilation: 75 genera (25 woody, 50 herbaceous)

Site	Disjunct genera	Disjunct Species	Non-disjunct Species
CN	61	263	3404
US	62	145	2049
Total	70	395	5213



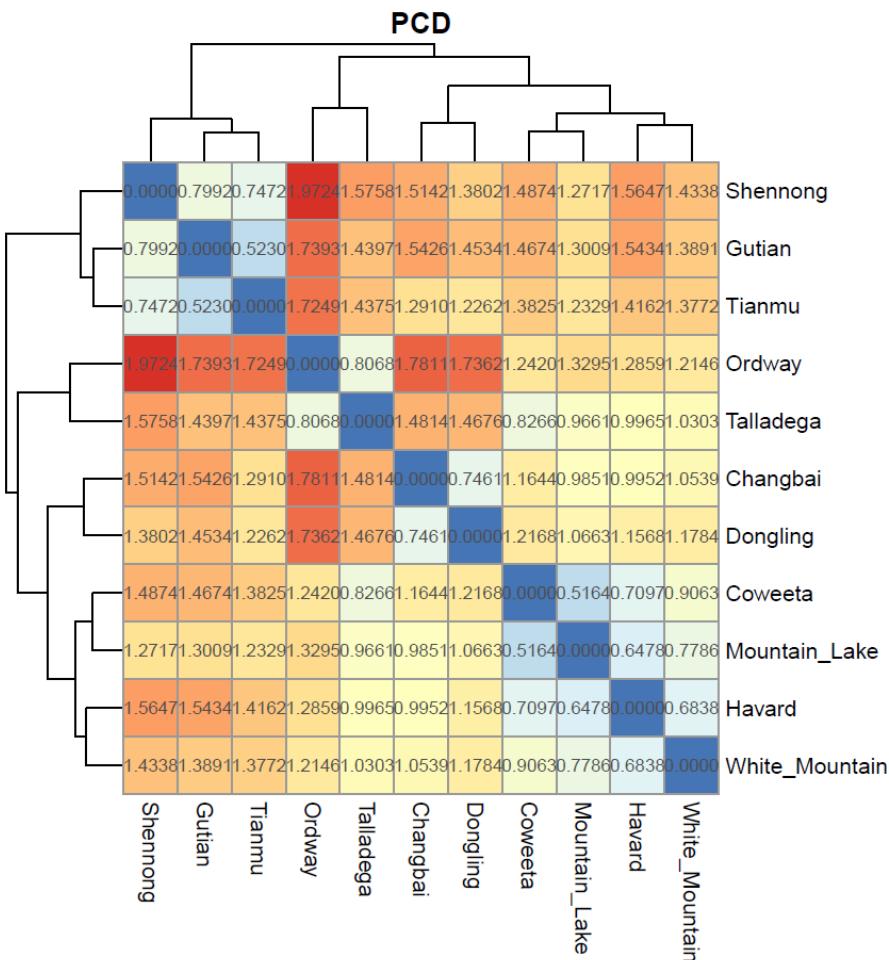
No-disjunct > Disjunct; US opposite pattern

Community phylogeny - β diversity

- Phylogenetic community dissimilarity (**PCD**): pairwise differences between communities
 - PCDc**: a nonphylogenetic component that reflects nonshared species between communities
 - PCDp**: a phylogenetic component that reflects the evolutionary relationships among nonshared species
- **UniFrac**: a phylogenetic beta diversity metric of the unique (non-shared) fraction of total phylogenetic diversity (branch-length) between two communities.

PCD is partitioned into a nonphylogenetic component that reflects shared species between communities (PCDc) and a phylogenetic component that reflects the evolutionary relationships among nonshared species (PCDp)

US-CN 11 Sites- β diversity



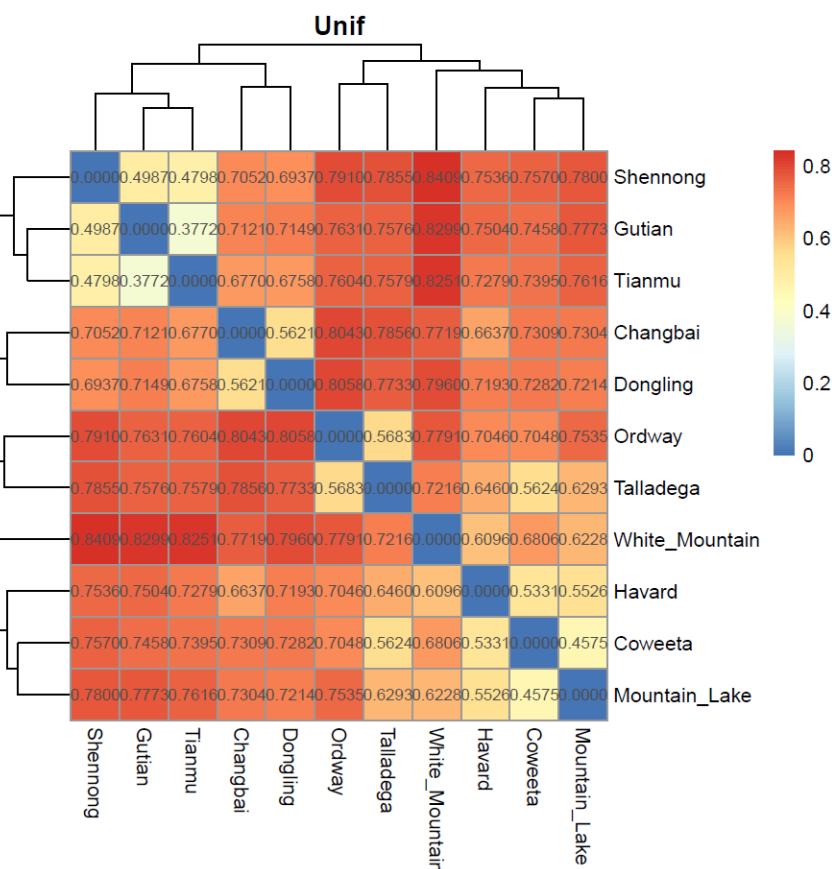
Ordway vs. Shengnong, and most CN sites

Ordway= Talladega

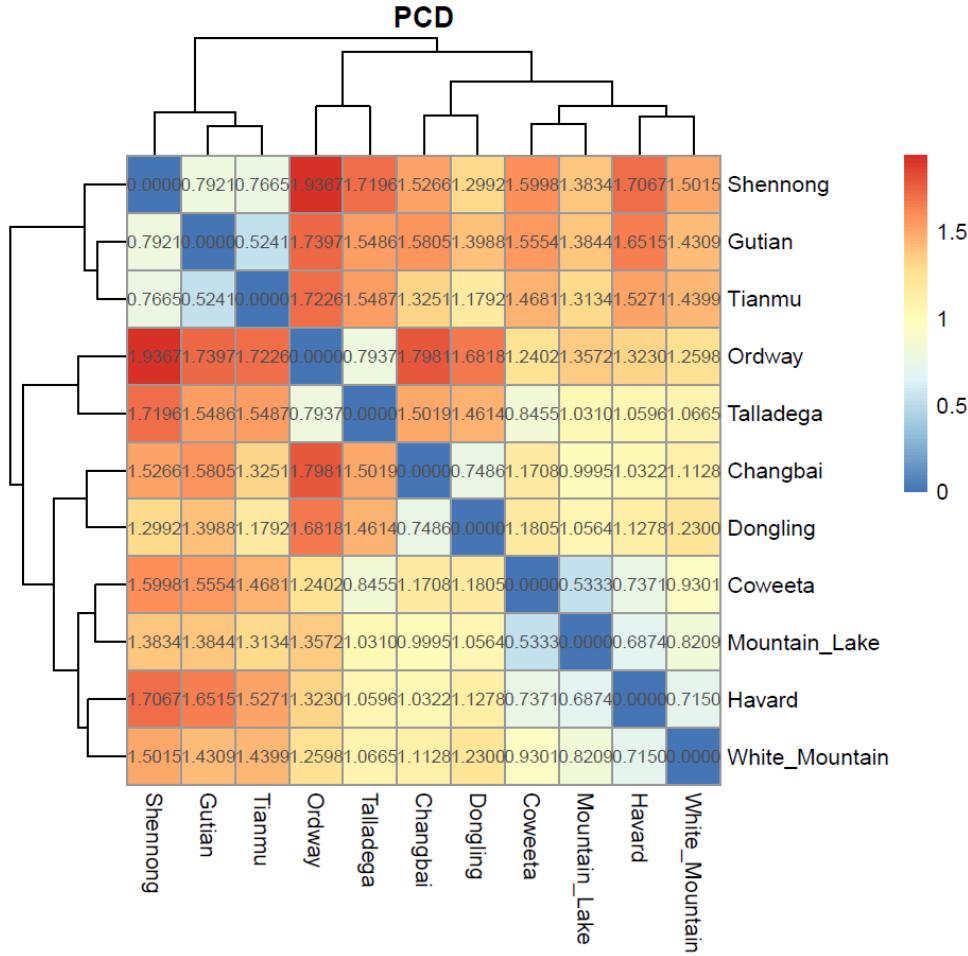
Coweeta=Moutain Lake

Gutian=Tianmu

Changbai vs. most CN south

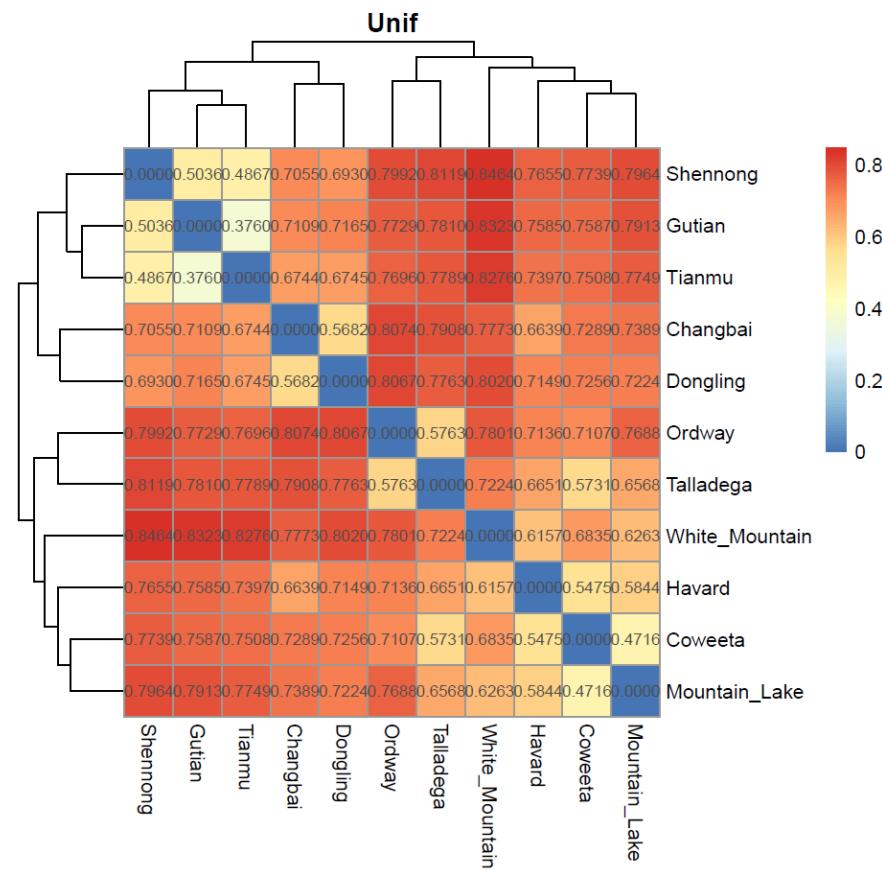


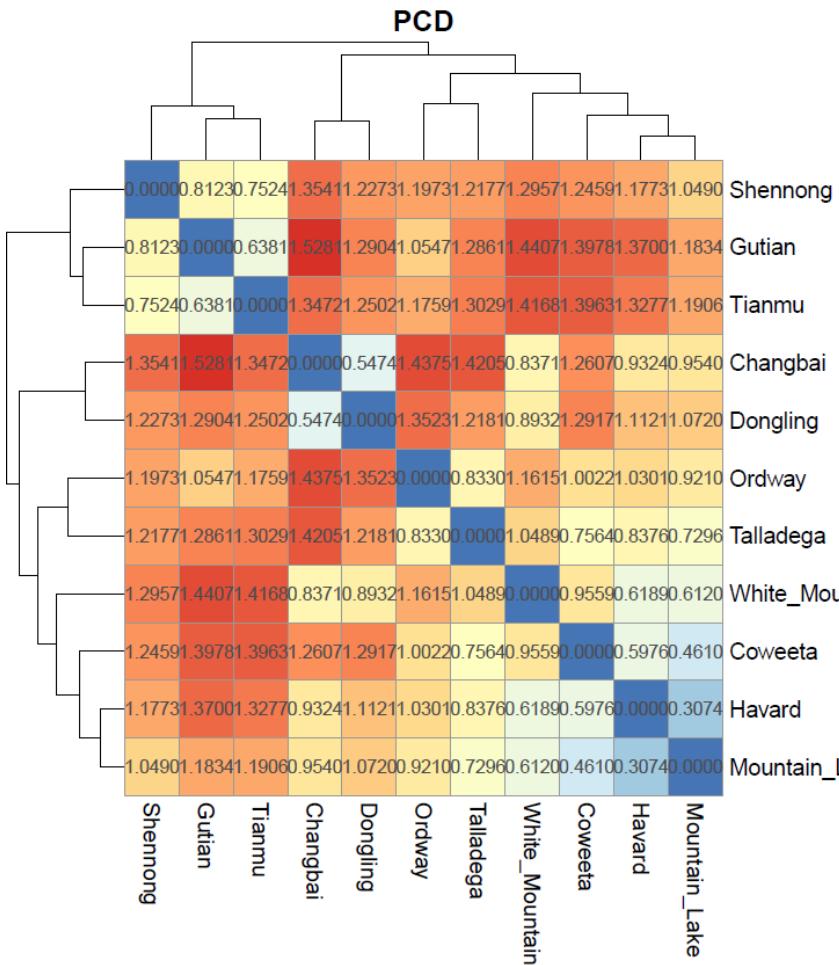
PCD

Ordway vs. Shengnong
Gutian=Tianmu

Disjunct only - β diversity

White Mountain vs. Shengnong
Gutian=Tianmu

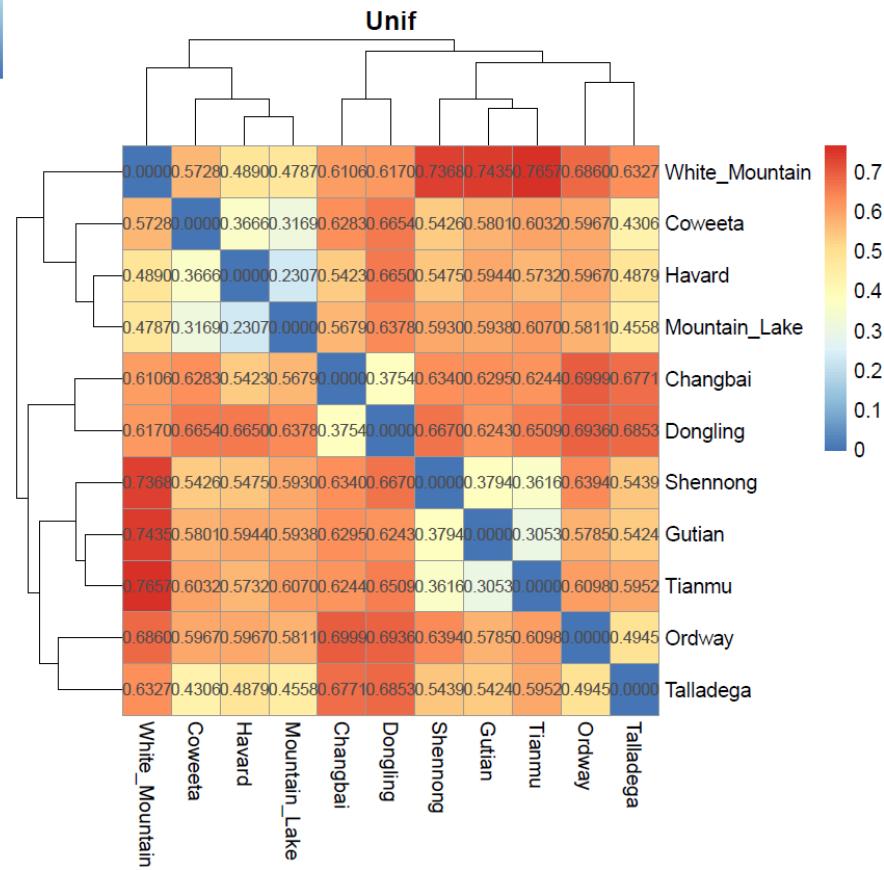
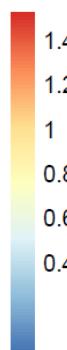




White Mountain = Harvard
Gutian vs. Changbai

No-Disjunct - β diversity

Harvard=Mountain
Shenong vs. White Moutian



**Q4: Insights for community structure and community
assemblage process at regional and local scale, respectively?**

Categories	ntaxa	PD	SES-PD	NRI	NTI
CN	3667	66034.38	-2.71	-0.58	2.08
US	2194	44617.83	-6.40	1.28	4.94
Woody	1845	38071.57	-9.12	-8.43	5.31
Herb	3743	55541.59	-22.04	16.08	12.36
Angiosperms	5532	86493.48	-12.26	72.62	3.43
Gymnosperms	76	2389.24	-8.91	4.46	7.80

Red: significant over-dispersion

Green: significant clustering

Overall clustering; Woody over-dispersion

Country	Site	nntaxa	SES-PDt	NRIt	NTIt
CN	Changbai	1140	-8.92	1.10	5.50
CN	Dongling	773	-7.58	-0.41	5.81
CN	Shennong	1559	-1.57	-0.80	2.44
CN	Tianmu	1188	-0.63	0.80	1.90
CN	Gutian	1209	0.56	-0.89	0.17
US	White_Mountain	254	-3.49	-2.64	2.79
US	Havard	897	-6.34	-0.35	4.82
US	Mountain_Lake	381	-3.61	-0.69	3.60
US	Coweeta	681	-5.35	2.05	4.18
US	Talladega	533	-2.89	1.17	1.94
US	Ordway	749	-5.23	1.41	4.28

Red: significant over-dispersion

Green: significant clustering

Overall clustering; White Mountain over-dispersion

Category	ntaxa	PD	SES-PD	NRI	NTI
CN	3667	66034.38	-2.71	-0.58	2.08
US	2194	44617.83	-6.40	1.28	4.94
Disjunct-total	395	9451.63	-14.68	-6.41	12.55
Nondisjunct-total	5213	83492.60	-4.25	10.08	0.47
Disjunct-CN	263	7649.42	-12.46	-2.03	9.68
Nondisjunct-CN	3404	62029.77	-4.32	4.99	0.98
Disjunct-US	145	6595.21	-4.21	-6.82	4.76
Nondisjunct-US	2049	40850.76	-9.51	9.43	5.42

Red: significant over-dispersion

Green: significant clustering

Overall clustering; Disjunct taxa over-dispersion

Genus level phylogeny with Target Enrichment data

353 nuclear genes

130 (Soltis) + 150 (ZYP) + 96 (Jenney) + CZD (86)

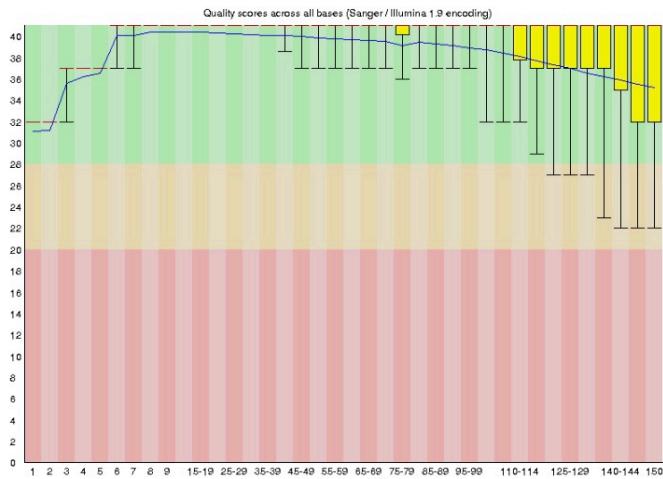
workflow

- FastQC
- Trimmomatic
- HybPiper
- Mafft
- PhyX
- RAxML
- Github (<https://github.com/Cactusolo/IESHTSTE>)

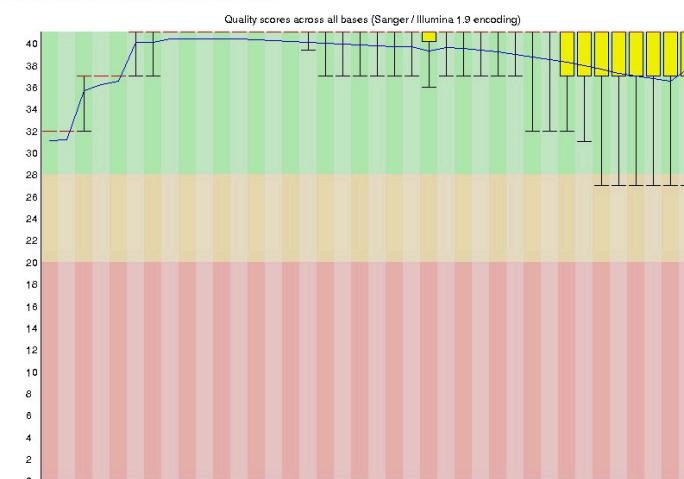
FastQC & Trimmomatic

SLIDINGWINDOW:20:20

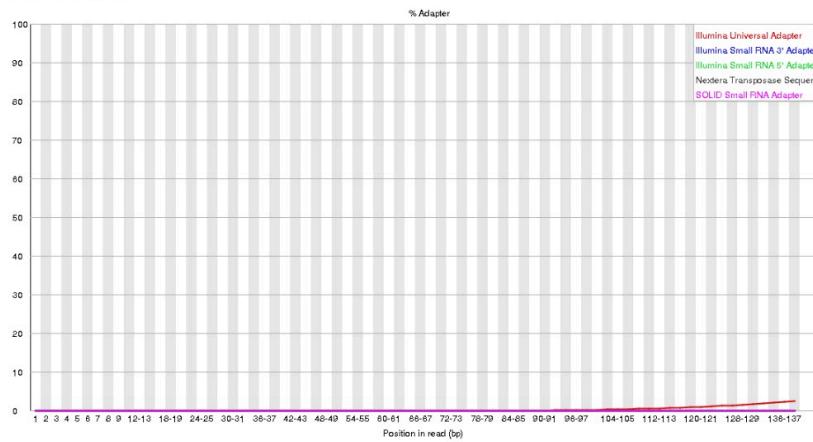
Per base sequence quality



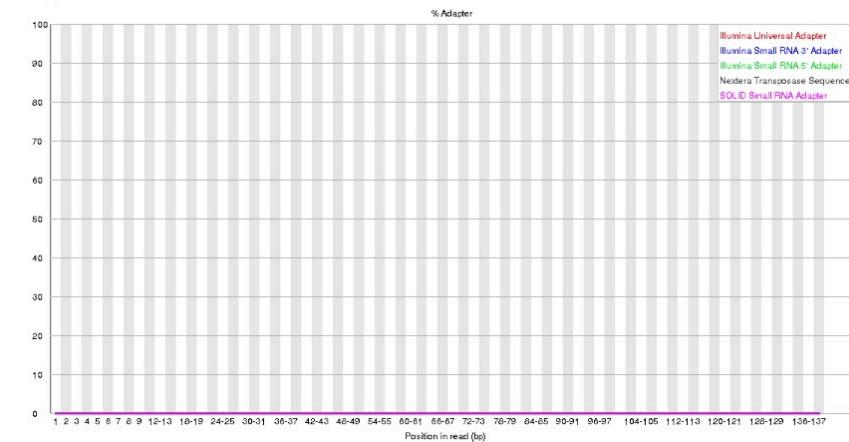
Per base sequence quality



Adapter Content



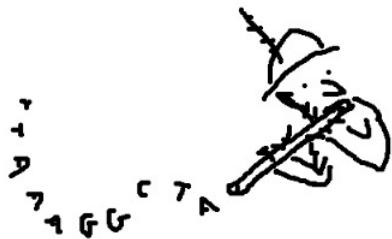
Adapter Content



FastQC

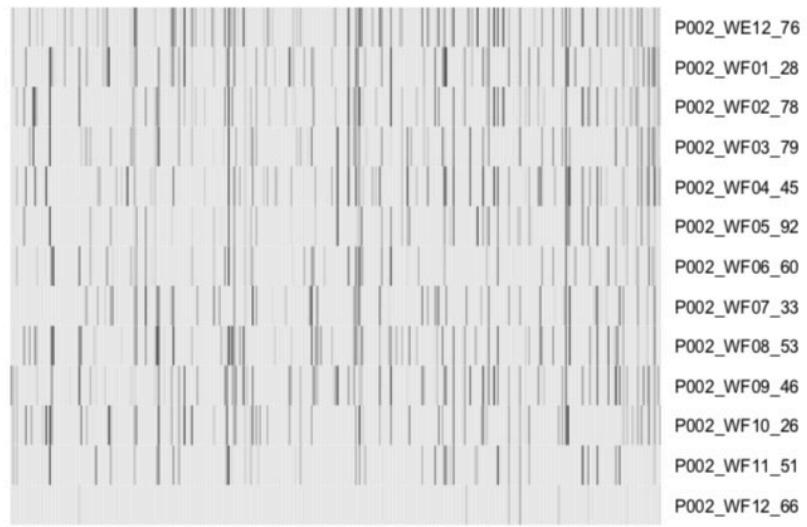
Seq_ID	Seq_len	Total_reads	QC45_mean	T_Deduplicated%
P002_WG01_84_R1_fastqc	150	2292259	35. 8	77. 13
P002_WG01_84_R2_fastqc	150	2292259	29. 49	80. 48
P002_WG02_12_R1_fastqc	150	765520	34. 11	80. 13
P002_WG02_12_R2_fastqc	150	765520	28. 71	82. 94
P002_WG03_22_R1_fastqc	150	983910	34. 74	79. 83
P002_WG03_22_R2_fastqc	150	983910	28. 68	82. 76
P002_WG04_10_R1_fastqc	150	885551	34. 18	79. 21
P002_WG04_10_R2_fastqc	150	885551	28. 2	82. 37
P002_WG05_69_R1_fastqc	150	1396862	34. 75	78. 93
P002_WG05_69_R2_fastqc	150	1396862	29. 05	81. 75
P002_WG06_14_R1_fastqc	150	1170402	35. 26	78. 23
P002_WG06_14_R2_fastqc	150	1170402	28. 93	81. 54
P002_WG07_18_R1_fastqc	150	690660	34. 44	80. 14
P002_WG07_18_R2_fastqc	150	690660	29. 18	82. 72
P002_WH10_34_R1_fastqc	150	971662	33. 54	78. 3
P002_WH10_34_R2_fastqc	150	971662	29. 89	80. 78
P002_WH12_56_R1_fastqc	150	1594675	34. 91	74. 11
P002_WH12_56_R2_fastqc	150	1594675	27. 85	79. 05
P003_WA01_59_R1_fastqc	150	956422	35. 72	72. 89
P003_WA01_59_R2_fastqc	150	956422	33. 39	73. 62
P003_WA02_27_R1_fastqc	150	2667004	36. 59	67. 66
P003_WA02_27_R2_fastqc	150	2667004	30. 94	71. 54
P003_WA06_74_R1_fastqc	150	2343818	36. 39	68. 55
P003_WA06_74_R2_fastqc	150	2343818	32. 43	71
P003_WA07_77_R1_fastqc	150	1389508	35. 96	71. 37
P003_WA07_77_R2_fastqc	150	1389508	29. 77	75. 61
P003_WA08_16_R1_fastqc	150	890044	35. 99	66. 45

HybPiper

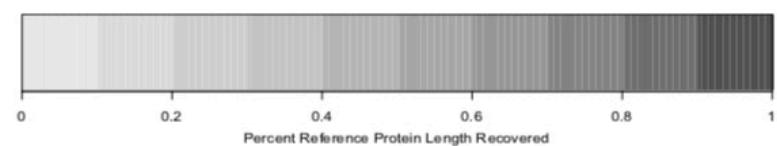
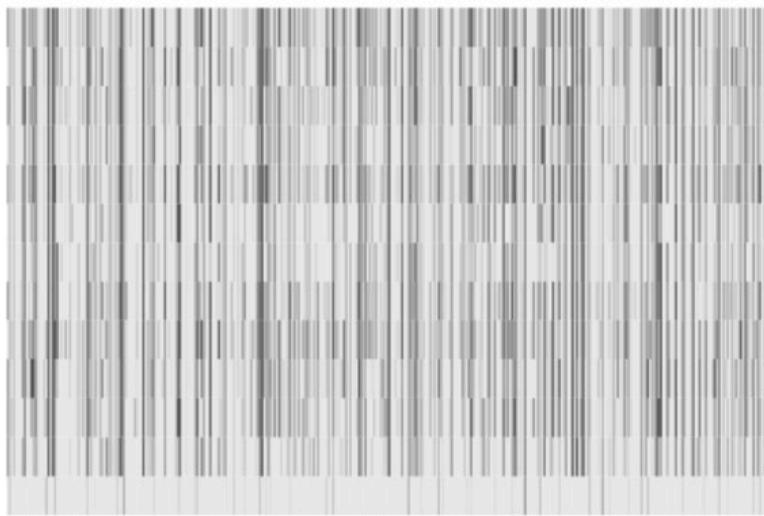
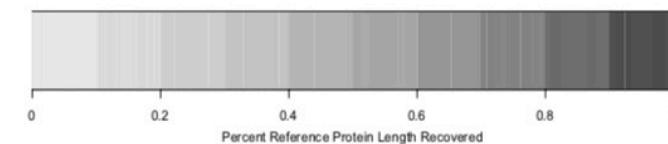


(logo by Elliot Gardner)

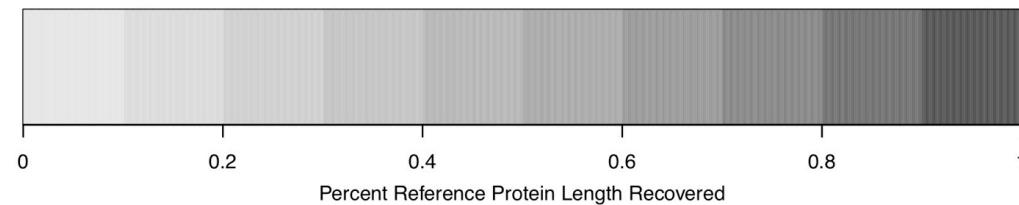
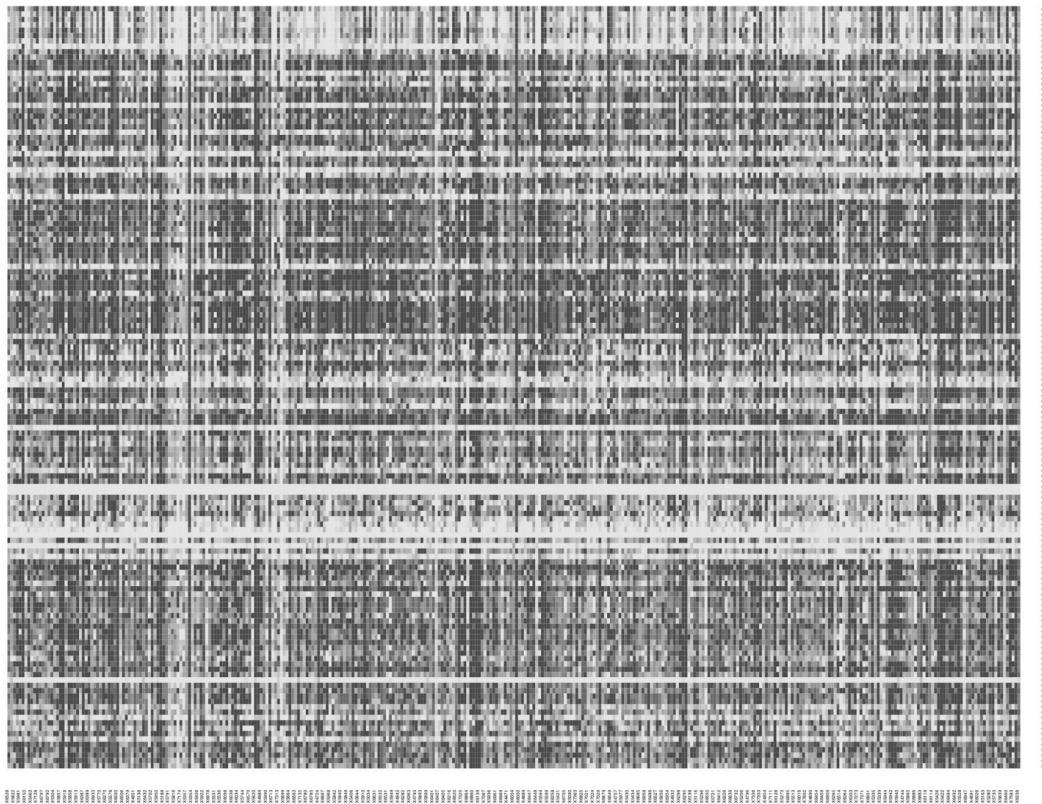
Paired + unpaired reads



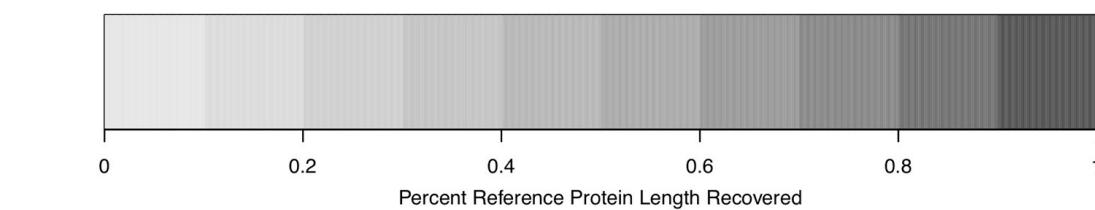
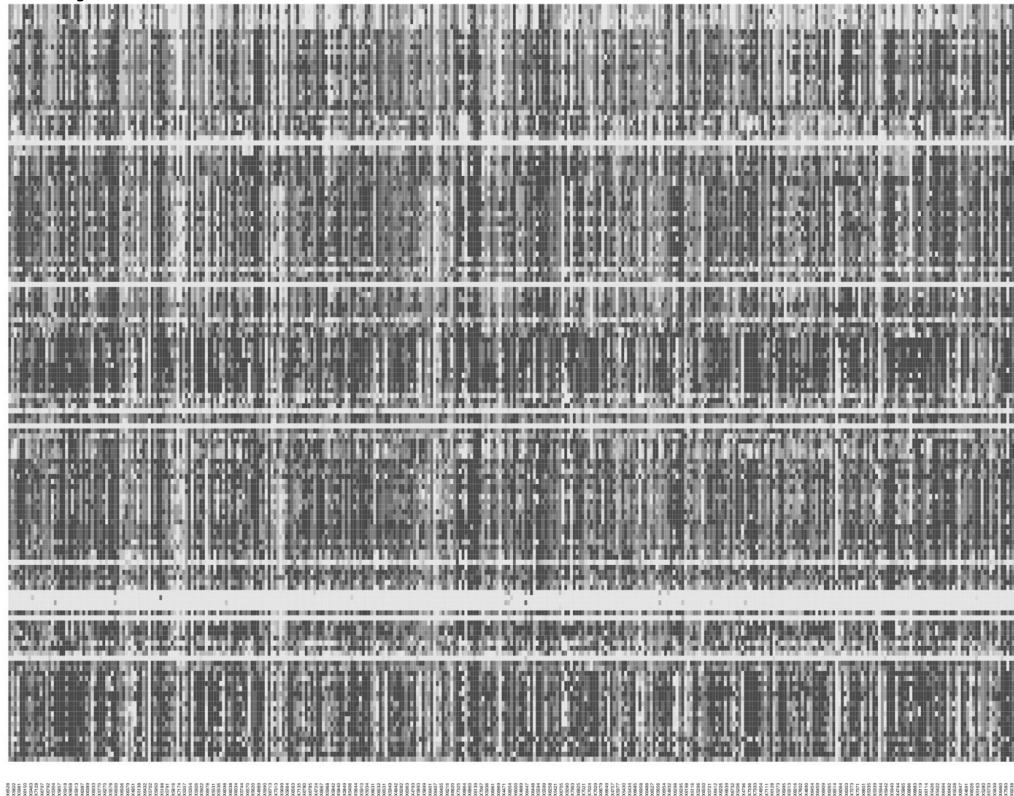
Paired reads only



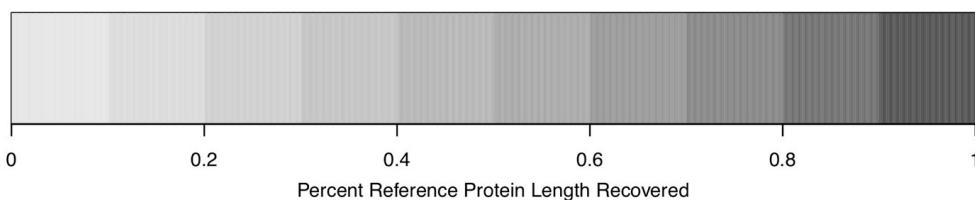
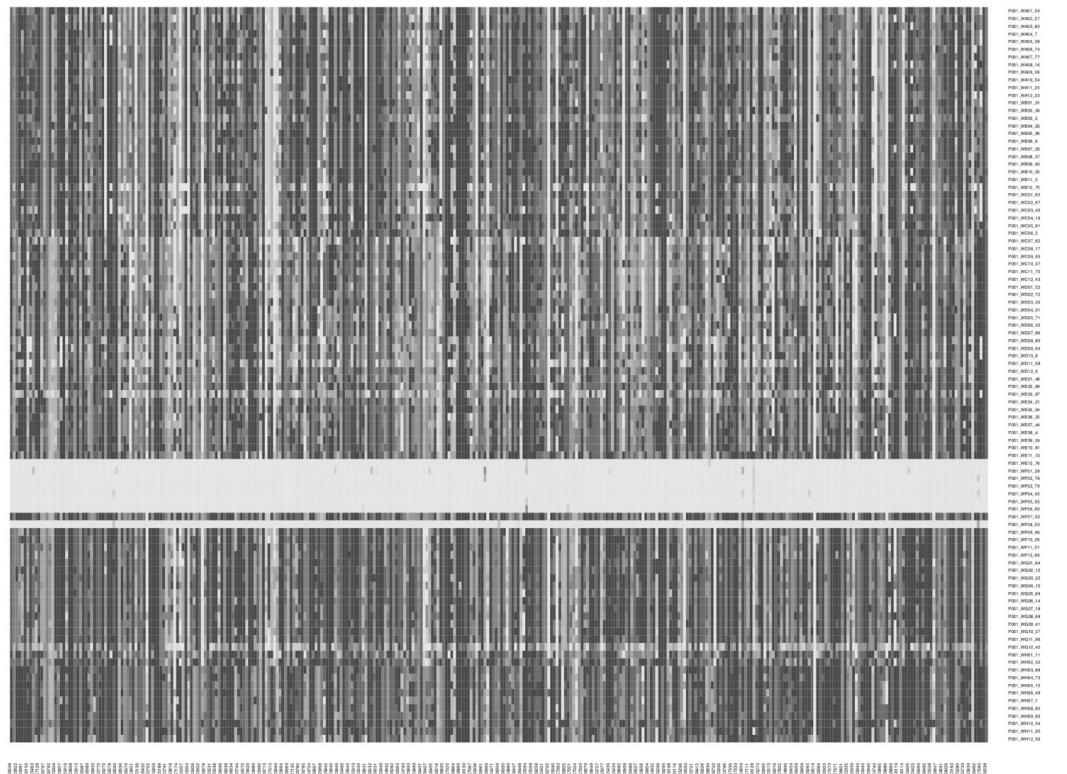
130 (Soltis)



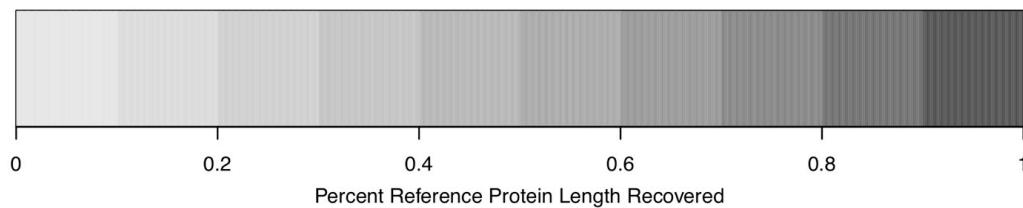
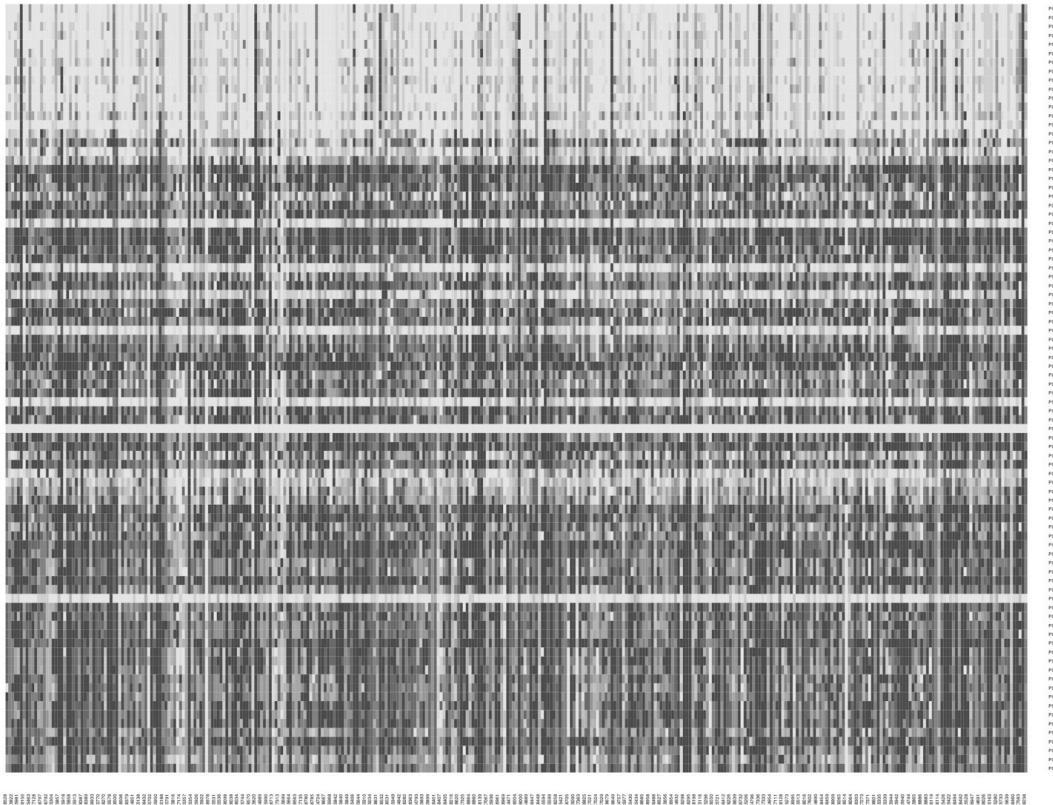
150(ZYP)



96 (Jenney)



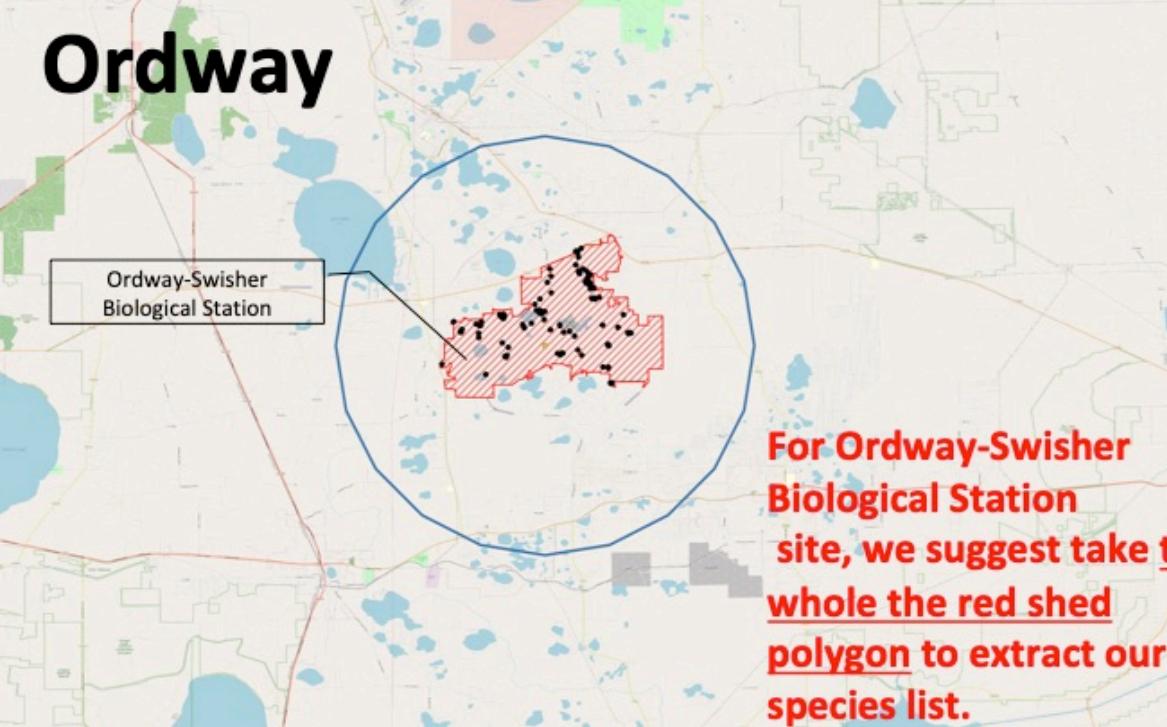
CZD (86)



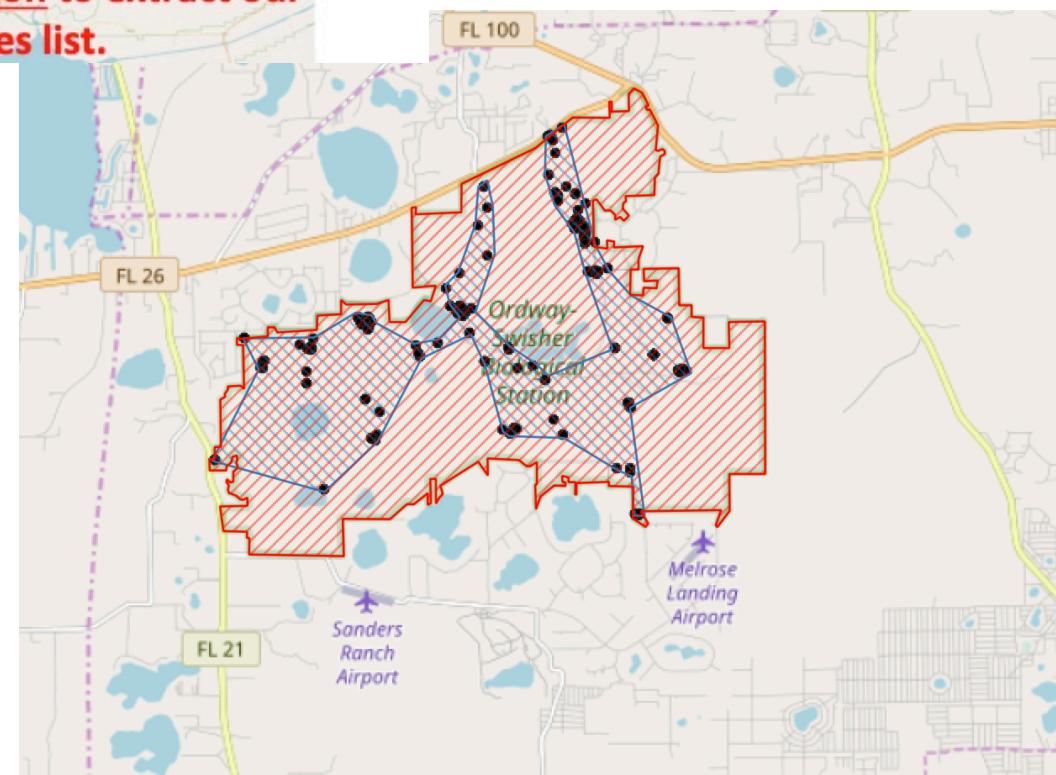
Plans

1. Species list
2. Rerun PD analysis
3. Writing manuscript with Hanyang
4. Generate phylogeny (Exons data)
5. Exons only? Intron and exon flanking region? and paralogs? Other off target genomes (cp and mt), besides nuclear?

Ordway



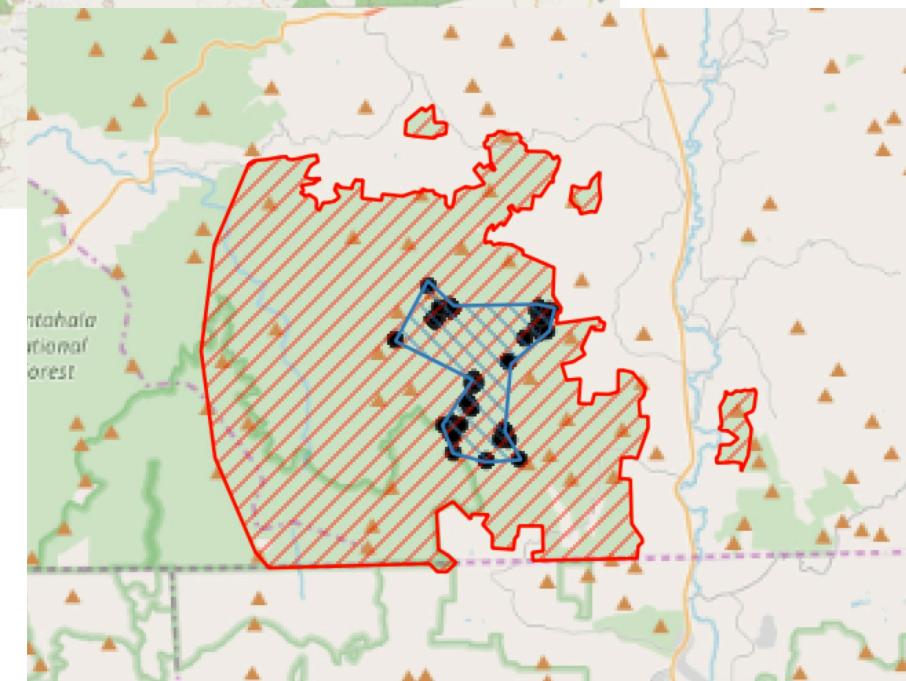
For Ordway-Swisher
Biological Station
site, we suggest take the
whole the red shed
polygon to extract our
species list.



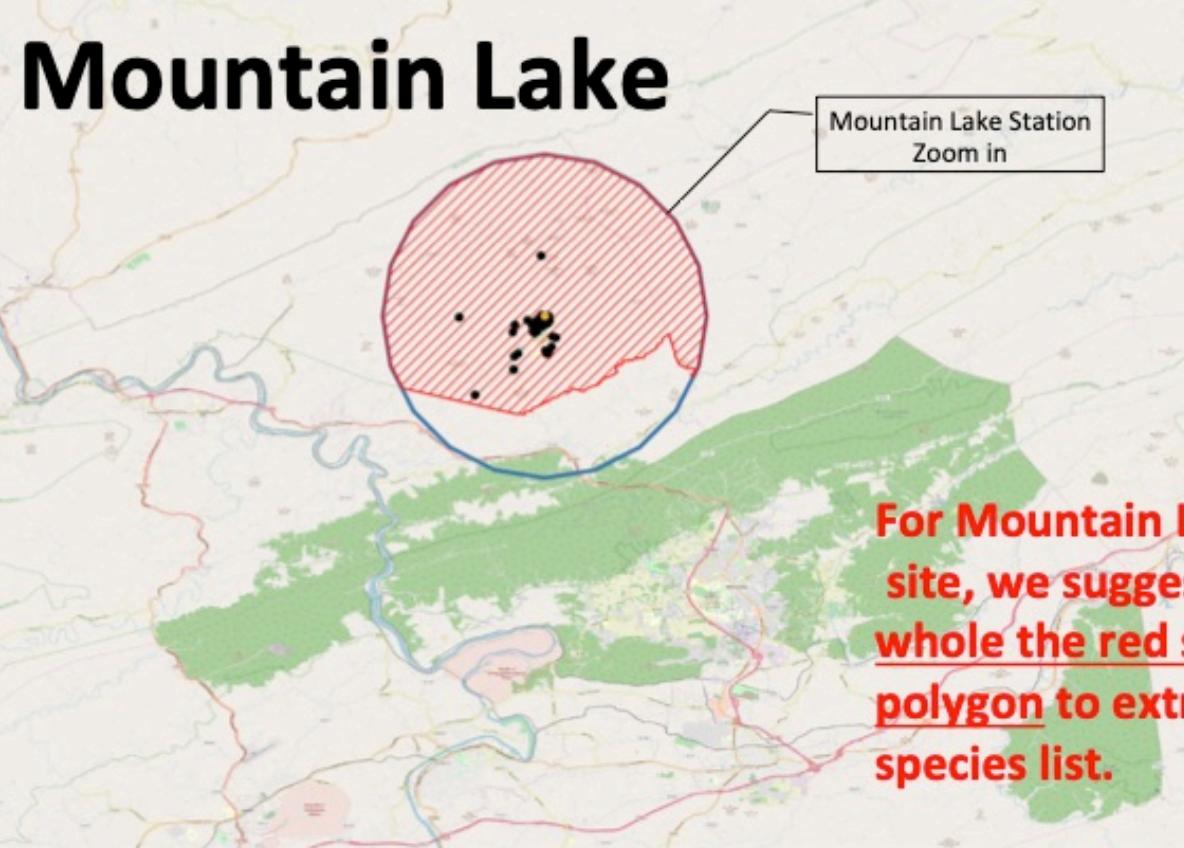
Coweeta

For Ceweeta site, we suggest go with the red shed polygon to extract our species list.

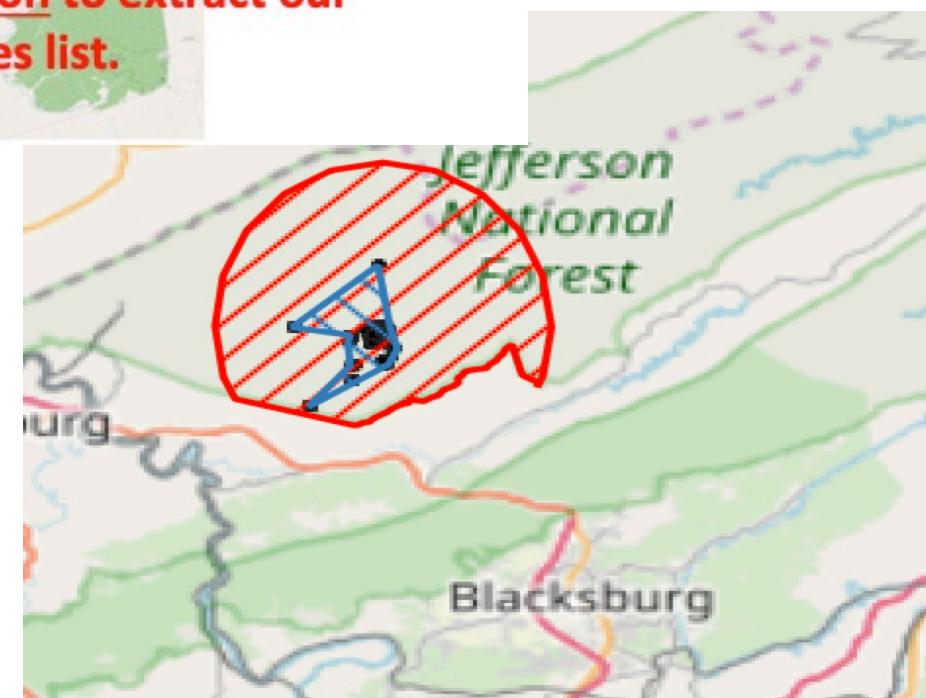
Maybe excluding those tiny small areas (skyblue arrow) on the edge



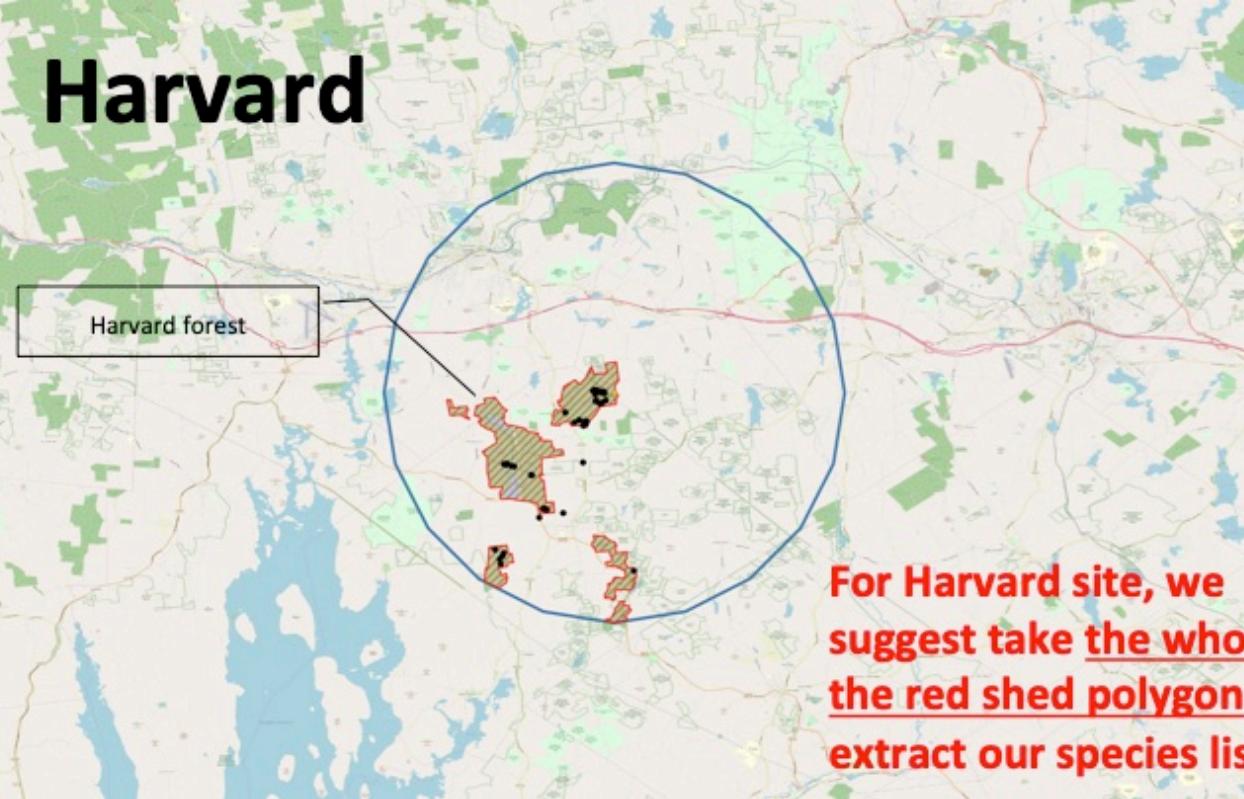
Mountain Lake



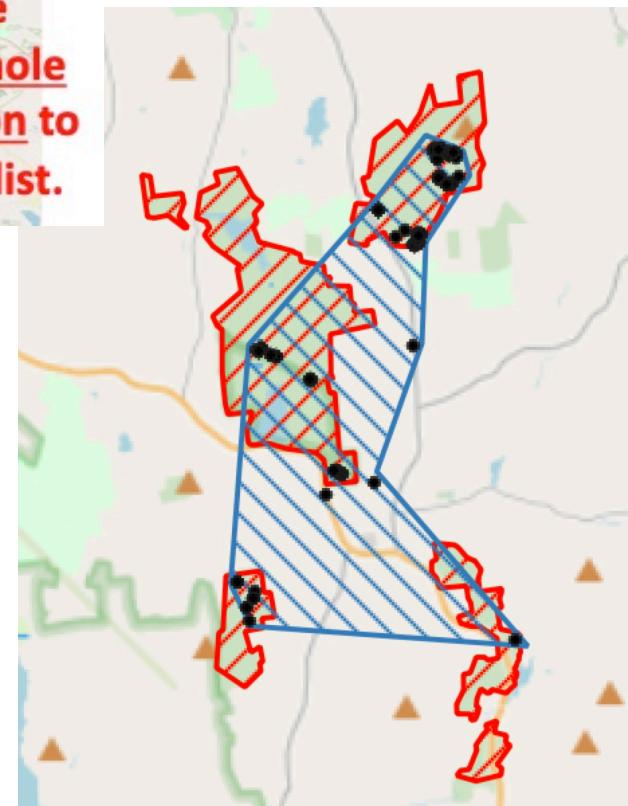
For Mountain Lake
site, we suggest take the
whole the red shed
polygon to extract our
species list.



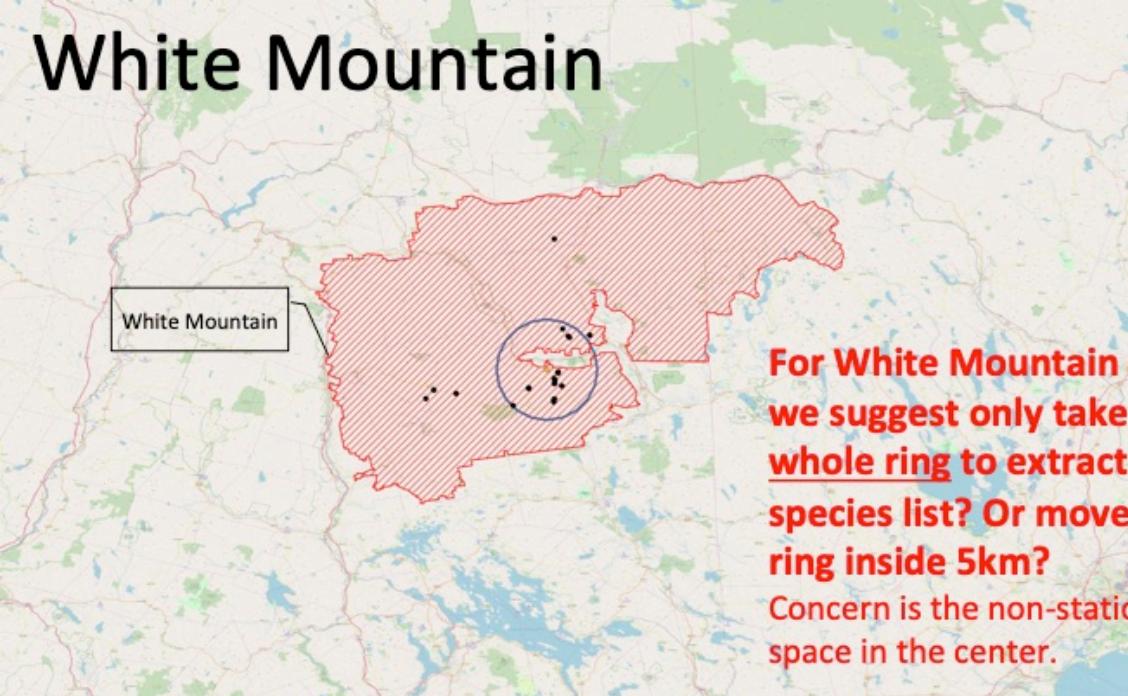
Harvard



For Harvard site, we suggest take the whole the red shed polygon to extract our species list.



White Mountain



For White Mountain site,
we suggest only take the
whole ring to extract our
species list? Or move the
ring inside 5km?

Concern is the non-station
space in the center.

