Supplementary Material 2: Statistical Analyses for Merged Traits

Based on merged traits described below, we repeated all analyses presented in the main text. The results provide even stronger evidence for a peak in projectile-point diversity in the Southeast.

Merged traits:
Trait 1 = 1, 2
Trait 2 = 1, 2, 3
Trait 3 = 1, 2, 3
Trait 4 = 1, 2, 3
Trait 5 = 1, 2, 3
Trait 6 = 1, 2, 3 (No change)
Trait 7 = 1, 2 (No change)
Trait 8 = 1, 2, 3

There are a total of 2916 possible classes.

The transition from original to merged characters:
Trait 1, (1, 2) = 1; (3, 4) = 2
Trait 2, (1, 2, 3) =1; (4, 5) = 2; 6=3
Trait 3, (1, 2) = 1; (3, 4) = 2; (5, 6) = 3
Trait 4, (1, 2) = 1; (3, 4) = 2; (5, 6, 7) = 3
Trait 5, (1, 2) = 1; (3, 4) = 2; (5, 6) = 3
Trait 6, No change
Trait 7, No change
Trait 8, (1, 2) = 1; (3, 4) = 2; (5, 6) = 3

The definition of each region is duplicated below:

(1) Southeast versus Northeast

> Southeast = Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.

> Northeast = Maine, Vermont, New Hampshire, Massachusetts, Connecticut, New York, New Jersey, Pennsylvania, Ohio, West Virginia, Indiana, Missouri, Nova Scotia, and Ontario.

(2) Upper Southeast versus Lower Southeast versus Northeast

> Upper Southeast = Arkansas, Kentucky, Missouri, North Carolina, Tennessee, Virginia

> Lower Southeast = Alabama, Florida, Georgia, Louisiana, Mississippi, South Carolina.

Northeast = Maine, Vermont, New Hampshire, Massachusetts, Connecticut, New York, New Jersey, Pennsylvania, Ohio, West Virginia, Indiana, Nova Scotia, and Ontario.

Based on the merged traits, the data summary for the Southeast and the Northeast and all the estimated asymptotes of diversities along with s.e. and 95% confidence intervals are given in Table S1. The corresponding results for the Lower Southeast, Upper Southeast and Northeast appear in Table S2. The rarefaction and extrapolation curves for the two sets of data are shown in Figure S1 and Figure S2, respectively.

Comparing Table 1 with Table S1, and also comparing Table 2 with Table S2, we see that, based on the merged traits, the observed class richness, the number of singletons and the number of doubletons are all substantially reduced, although sample sizes are kept the same as those for unmerged traits. Also, the sample completeness for each sample is drastically increased so that more accurate inference can be made and significance can be more easily revealed.

Southeast vs. Northeast (Table S1, Figure S1)

Both sample-size- and coverage-based rarefaction/extrapolation sampling curves for common and dominant classes show that the Southeast is significantly more diverse than the Northeast. (For sample-size-based rarefaction/extrapolation, the significance for the Southeast > Northeast cannot be revealed for the original traits, but it can be revealed for merged traits for common and dominant classes.) For class richness, due to undetected rare classes, this significance is supported for a maximum fraction of the assemblage size (~60% for original traits, and now ~ 90% for merged traits.) The conclusions here are generally consistent with those based on the original traits presented in the main text. The difference in the coverage limit is mainly due to merge of traits.

Upper Southeast vs. Lower Southeast vs. Northeast (Table S2, Figure S2)

The sample-size- and coverage-based rarefaction and extrapolation plots based on the merged traits exhibit clearer three-area diversity patterns. As with the original traits (Figure 2 of the main text), all plots in Figure S2 show the consistent diversity ordering pattern that the Upper Southeast > Lower Southeast > Northeast. The confidence intervals for the three areas are generally not disjoint because the Upper Southeast and the Lower Southeast have very close diversities especially for common and dominant classes of projectile points. However, their confidence intervals are clearly separable from that of the Northeast.

Based on the merged traits, we can perform pairwise comparison to infer whether the Upper Southeast is significantly richer than the Lower Southeast or the Northeast, as conjectured by Broster et al. (2013). As with the original traits, both sample-size- and coverage-based rarefaction/extrapolation sampling curves show that for common and dominant classes (lower right panel), the Upper Southeast is significantly more diverse

than the Northeast. This conclusion can be extended to the asymptotes, as also shown by the non-overlap CIs in Table S2. For class richness, although the two CIs of the asymptotes are disjoint, data support the same conclusion up to a fraction of 90% of the assemblages. (Similar results are also valid for comparing the Lower Southeast and the Northeast, as the Lower Southeast and Upper Southeast have close diversities.)

To test whether the Upper Southeast is significantly richer than the Lower Southeast, we notice that no evidence exists to support this conjecture for common and dominant classes. Although the two CIs of the asymptotes of class richness are disjoint (Table S2), data only reveal that when sample size is between 500 and 1000 (or sample coverage is between 70% and 90%), the Upper Southeast is significantly richer than the Lower Southeast for class richness (the upper rows of Figure S2).

Supplementary Table S1. (Merged traits) Data summary for the Southeast and the Northeast, with statistical inference for estimated asymptotes of diversities

(a) Data summary with Anderson (1990) definition of the Southeast and the Northeast
   (f_k denotes the number of classes represented by exactly k individuals in the sample)

Area	Sample Size n	Observed richness	Sample Coverage (completeness)	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
Southeast	672	190	87.5%	84	36	17	11	2	6	7	0	4	4
Northeast	384	122	84.1%	61	20	11	4	7	5	1	0	4	1

Area	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}
Southeast	6	5	1	0	3	2	2	0	0
Northeast	2	1	0	0	3	0	0	1	1

(b) Observed diversities and the estimated asymptotes of diversities in the Southeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	190	287.9	28.6	245.8	361.6
Shannon diversity (common class richness)	119.3	151.1	7.0	137.5[*]	164.8[*]
Simpson diversity (dominant class richness)	85.7	98.1	5.1	88.1[*]	108.0[*]

 ＊ Interval does not overlap with the interval for the Northeast

(c) Observed diversities and the estimated asymptotes of diversities in the Northeast

	Observed	Estimated	Estimated	95% lower	95% upper

	richness	asymptote	s.e.	confidence interval	confidence interval
Class richness	122	214.8	33.0	169.2	304.4
Shannon diversity (common class richness)	75.5	101.8	7.4	87.3[*]	116.2[*]
Simpson diversity (dominant class richness)	51.2	59.0	4.6	51.2[*]	68.0[*]

* Interval does not overlap with the interval for the Southeast

Supplementary Table S2. (Merged Traits) Data summary for the Lower Southeast, Upper Southeast and Northeast, with statistical inference for estimated asymptotes of diversities

(a) Data summary for the Lower Southeast, Upper Southeast and Northeast
   (f_k denotes the number of classes represented by exactly k individuals in the sample)

Area	Sample Size n	Observed richness	Sample Coverage (completeness)	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
Lower Southeast	265	114	80.8%	51	31	13	9	1	3	3	0	0	1
Upper Southeast	499	168	82.8%	86	29	12	7	9	4	2	3	4	2
Northeast	292	89	86.0%	41	17	4	8	5	4	0	1	2	0

Area	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
Lower Southeast	1	1	0	0	0	0	0	0
Upper Southeast	3	1	2	3	1	0	0	0
Northeast	1	2	1	0	1	0	1	1

(b) Observed diversities and the estimated asymptotes of diversities in the Lower Southeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	114	155.8	15.3	134.8[#]	197.9[#]
Shannon diversity (common class richness)	86.7	121.2	7.1	107.2[$]	135.2[$]
Simpson diversity (dominant class richness)	65.0	85.7	7.6	70.8[$]	100.7[$]

[#] Interval does not overlap with the interval for the Upper Southeast
[$] Interval does not overlap with the interval for the Northeast

(c) Observed diversities and the estimated asymptotes of diversities in the Upper Southeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	168	295.3	37.9	239.8[*#]	393.4[*#]
Shannon diversity (common class richness)	108.1	148.4	8.4	131.9[*]	165.0[*]
Simpson diversity (dominant class richness)	76.3	89.9	6.1	78.0[*]	101.8[*]

* Interval does not overlap with the interval for the Northwest
# Interval does not overlap with the interval for the Lower Southeast

(d) Observed diversities and the estimated asymptotes of diversities in the Northeast

|	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	89	138.3	20.7	111.3[*]	197.6[*]
Shannon diversity (common class richness)	56.3	73.0	5.6	62.0[*$]	84.0[*$]
Simpson diversity (dominant class richness)	39.0	44.9	3.7	39.0[*$]	52.2[*$]

* Interval does not overlap with the interval for the Upper Southeast
$ Interval does not overlap with the interval for the Lower Southeast
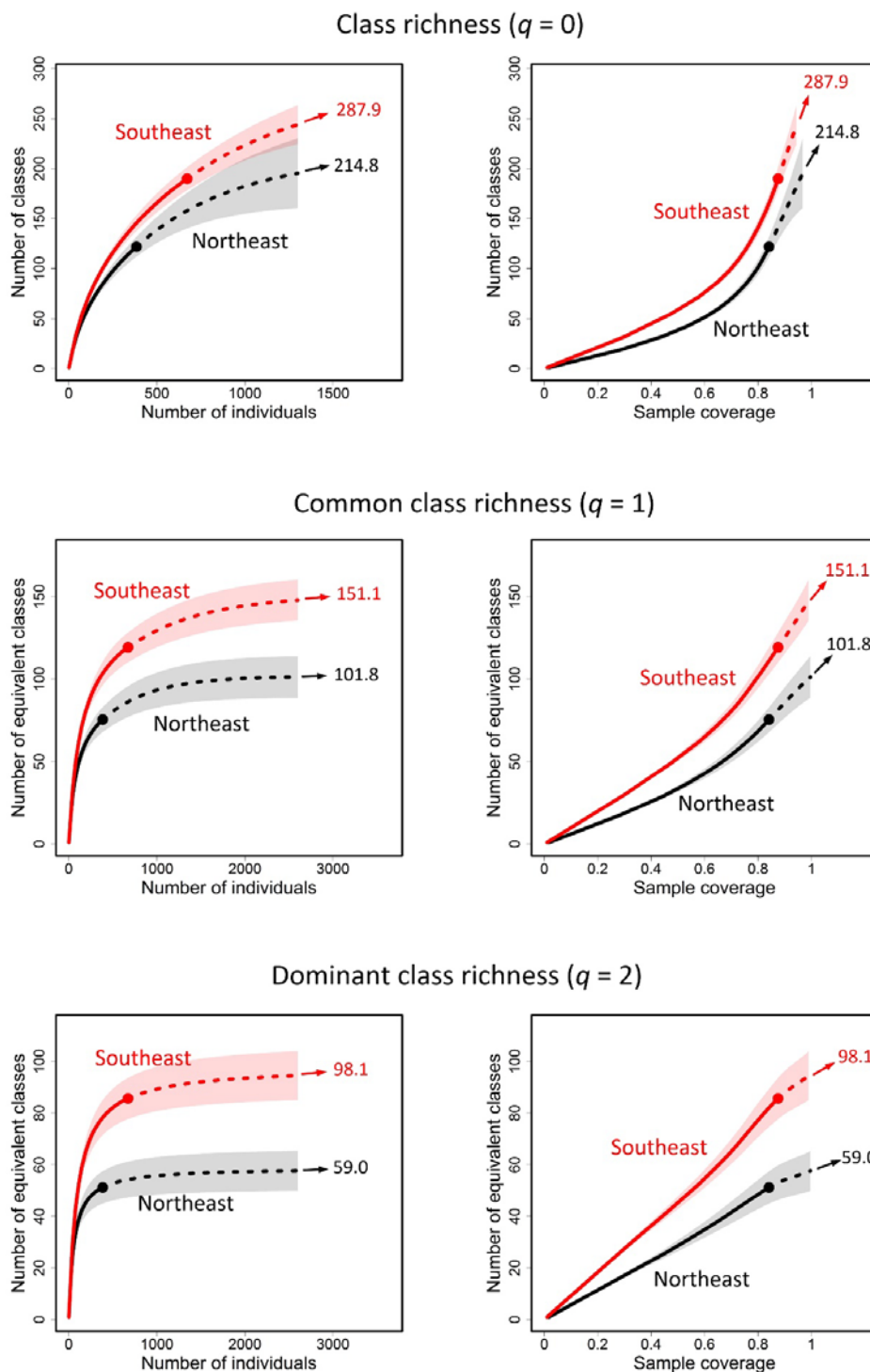
Supplementary Figure S1. (Merged traits) Comparison of sample-size-based (left panels) and sample-coverage-based (right panels) rarefaction and extrapolation for class richness (upper panels), Shannon diversity (middle panels) and Simpson diversity (lower panels) for the Southeast and the Northeast data. Observed samples are denoted by solid dots;

rarefied segments are denoted by solid lines and extrapolated segments by broken lines. The extrapolation extends up to a maximum sample size of 1300 for class richness, and to a maximum sample size of 2600 for Shannon diversity and Simpson diversity. The sample-coverage-based extrapolation extends to the coverage value of the corresponding maximum sample size, for each sample. The 95% confidence intervals (shaded areas) were obtained by a bootstrap method based on 200 replications. The estimated asymptote of diversity for each curve is shown next to the arrow at the right-hand end of each curve.
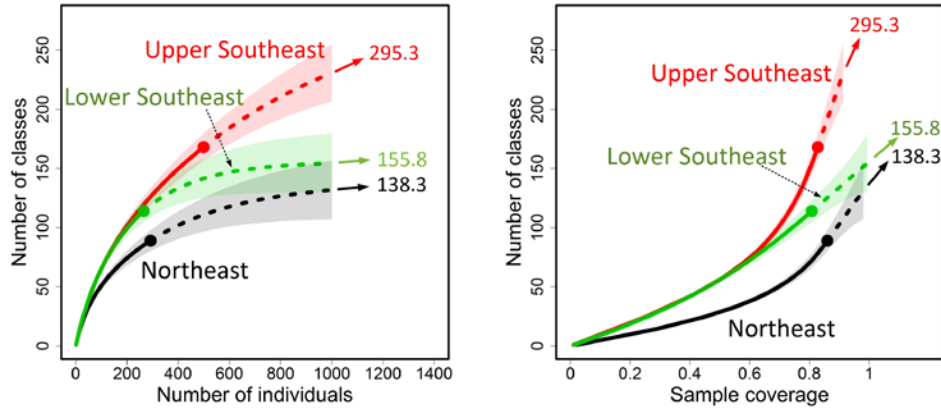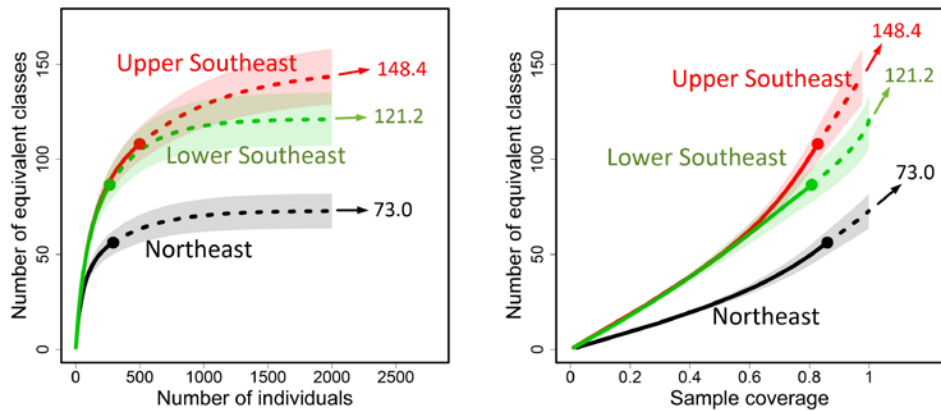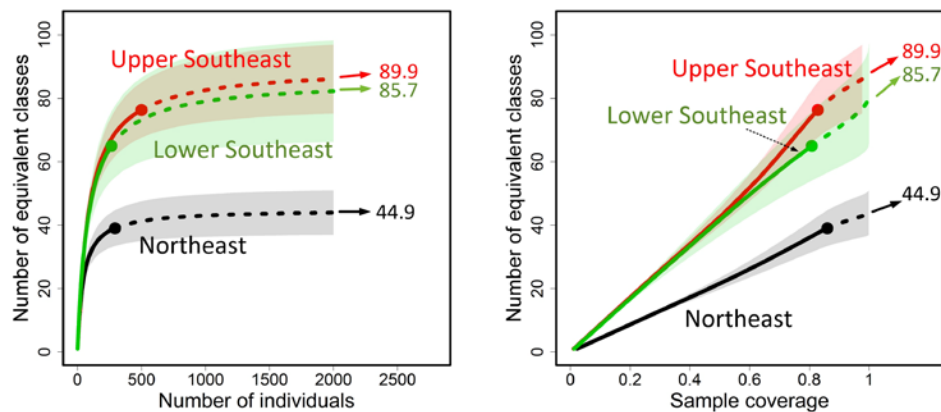
Supplementary Figure S2. (Merged traits) Comparison of sample-size-based (left panels) and sample-coverage-based (right panels) rarefaction and extrapolation for class richness (upper panels), Shannon diversity (middle panels) and Simpson diversity (lower panels)

for the Lower Southeast, Upper Southeast and Northeast data. Observed samples are denoted by solid dots; rarefied segments are denoted by solid lines and extrapolated segments by broken lines. The extrapolation extends up to a maximum sample size of 1000 for class richness, and to a maximum sample size of 2000 for Shannon diversity and Simpson diversity. The sample-coverage-based extrapolation extends to the coverage value of the corresponding maximum sample size, for each sample. The 95% confidence intervals (shaded areas) were obtained by a bootstrap method based on 200 replications. The estimated asymptote of diversity for each curve is shown next to the arrow at the right-hand end of each curve.

Reference

Broster, John B., Mark Norton, D. Shane Miller, Jesse W. Tune, and Jon D. Baker
 2013   Tennessee Paleoindian record: the Cumberland and Lower Tennessee River
 Watersheds. In *In the Eastern Fluted Point Tradition*, edited by Joseph A.M.
 Gingerich, pp. 299-314. University of Utah Press, Salt Lake City.

Supplementary Material 3. Statistical Analysis of Unmerged and Merged Traits Comparing Tennessee River Valley versus Southeast versus Northeast

The definitions of each region are as follows:

> Tennessee River Valley = Arkansas, Missouri, Tennessee, Kentucky, Alabama over 34 latitude, Mississippi over 34 latitude.

> Southeast = Florida, Georgia, Louisiana, North Carolina, South Carolina, Virginia, Alabama under 34 latitude, Mississippi under 34 latitude.

> Northeast = Maine, Vermont, New Hampshire, Massachusetts, Connecticut, New York, New Jersey, Pennsylvania, Ohio, West Virginia, Indiana, Nova Scotia, and Ontario.

Based on the unmerged traits, the data summary for the above three regions, with statistical inference for estimated asymptotes are given in Table S3. The corresponding summary for merged traits is provided in Table S4. The rarefaction and extrapolation curves for unmerged and merged traits are shown in Figure S3 and Figure S4, respectively. The focus here is to test whether we can perform pairwise comparison to infer whether the Tennessee River Valley has greater projectile-point diversity than the Southeast or the Northeast.

Unmerged traits (Table S3, Figure S3)

A general pattern revealed by the plots in Figure S3 is that the diversity curve for the Tennessee River Valley for all plots lies above the curves for the other two areas. Based on the coverage-based curves (right panels in Figure S3) and the summary statistics in Table S3, we obtain that the Tennessee River Valley is significantly greater than that of the Northwest; the conclusion can be extended to entire assemblage and complete coverage for common and dominant classes. For class richness, data support the extrapolation up to sample coverage of 60%. For class richness and common classes, similar conclusions are also valid for comparing the Tennessee River Valley and the Southeast; for dominant classes, the two curves are not statistically distinguishable.

Merged traits (Table S4, Figure S4)

For merged traits, a clearer diversity ordering pattern is shown: Tennessee River Valley > Southeast > Northeast. From the coverage-based curves (right panels in Figure S4) and the summary statistics in Table S4, the diversity of the Tennessee River Valley is significantly greater than that of the Northeast up to the sample coverage of 90% for class richness; for common and dominant classes, the significance difference can be extended to entire assemblage and complete coverage. Similar conclusions are also valid for comparing the Tennessee River Valley and the Southeast.

Supplementary Table S3. (Unmerged traits) Data summary for the Tennessee River Valley, Southeast, and Northeast, with statistical inference for estimated asymptotes of diversities

(a) Data summary for the Tennessee River Valley, Southeast, and Northeast. (f_k denotes the number of classes represented by exactly k individuals in the sample)

Area	Sample size n	Observed richness	Sample coverage (completeness)	f_1	f_2	f_3	f_4	f_5	f_6
Tennessee River Valley	607	451	40.7%	360	55	19	9	4	4
Southeast	157	138	23.7%	120	17	1	0	0	0
Northeast	292	218	42.2%	169	32	11	4	2	0

(b) Observed diversities and the estimated asymptotes of diversities in the Tennessee River Valley

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	451	1627.2	204.2	1290.1[*#]	2099.9[*#]
Shannon diversity (common class richness)	392.9	1227.8	87.9	1055.4[*#]	1400.2[*#]
Simpson diversity (dominant class richness)	323.5	691.4	62.6	568.7[*]	814.2[*]

[*] Interval does not overlap with the interval for the Northeast
[#] Interval does not overlap with the interval for the Southeast

(c) Observed diversities and the estimated asymptotes of diversities in the Southeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	138	558.8	129.4	371.5[#]	896.5[#]
Shannon diversity (common class richness)	132.3	569.8	90.2	393.0[#]	746.7[#]
Simpson diversity (dominant class richness)	125.1	612.3	77.8	459.8	764.8

[#] Interval does not overlap with the interval for the Tennessee River Valley

(d) Observed diversities and the estimated asymptotes of diversities in the Northeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	218	662.7	106.3	498.2[*]	924.0[*]
Shannon diversity (common class richness)	194.3	553.4	59.2	437.3[*]	669.4[*]
Simpson diversity (dominant class richness)	167.2	389.8	40.0	311.3[*]	468.3[*]

[*] Interval does not overlap with the interval for the Tennessee River Valley

Supplementary Table S4. (Merged traits) Data summary for the Tennessee River Valley, Southeast, and Northeast, with statistical inference for estimated asymptotes of diversities

(a) Data summary for the Tennessee River Valley, Southeast and Northeast
  (f_k denotes the number of classes represented by exactly k individuals in the sample)

Area	Sample size n	Observed richness	Sample coverage (completeness)	f_1	f_2	f_3	f_4	f_5	f_6
Tennessee River Valley	607	191	84.2%	96	31	15	9	6	4
Southeast	157	79	68.3%	50	13	6	1	2	4
Northeast	292	89	86.0%	41	17	4	8	5	4

Area	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
Tennessee River Valley	6	4	3	2	5	2	4	1	1	1	1	0
Southeast	2	0	0	0	1	0	0	0	0	0	0	0
Northeast	0	1	2	0	1	2	1	0	1	0	1	1

(b) Observed diversities and the estimated asymptotes of diversities in the Tennessee River Valley

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	191	339.4	42.1	277.0[*]	447.2[*]
Shannon diversity (common class richness)	120.3	162.0	9.2	143.9[*#]	180.2[*#]
Simpson diversity (dominant class richness)	85.3	99.0	5.7	87.9[*#]	110.2[*#]

[*] Interval does not overlap with the interval for the Northeast
[#] Interval does not overlap with the interval for the Southeast

(c) Observed diversities and the estimated asymptotes of diversities in the Southeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence	95% upper confidence

|				interval	interval
Class richness	79	174.5	39.1	123.2	285.6
Shannon diversity (common class richness)	58.1	102.2	12.6	77.5[#]	126.9[#]
Simpson diversity (dominant class richness)	42.1	57.2	7.6	42.3[#]	72.2[#]

[*] Interval does not overlap with the interval for the Northwest
[#] Interval does not overlap with the interval for the Tennessee River Valley

(d) Observed diversities and the estimated asymptotes of diversities in the Northeast

	Observed richness	Estimated asymptote	Estimated s.e.	95% lower confidence interval	95% upper confidence interval
Class richness	89	138.3	20.7	111.3[*]	197.6[*]
Shannon diversity (common class richness)	56.3	73.0	5.2	62.9[*]	83.2[*]
Simpson diversity (dominant class richness)	39.0	44.9	4.0	39.0[*]	52.7[*]

[*] Interval is not overlapped with the one of the Tennessee River Valley

Class richness ($q = 0$)
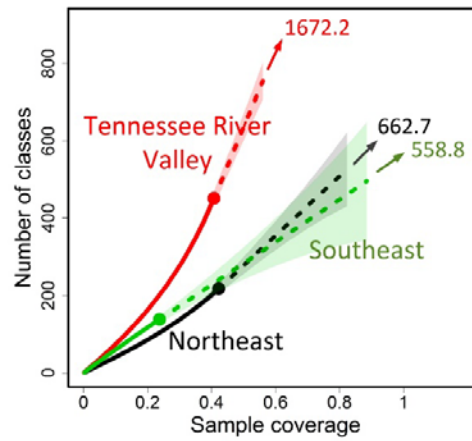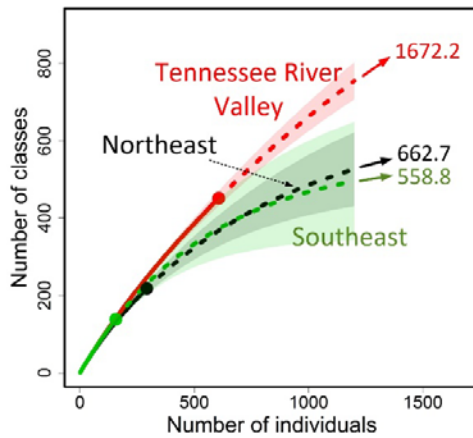
Common class richness ($q = 1$)
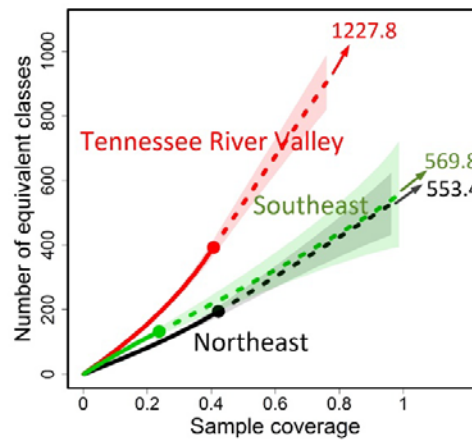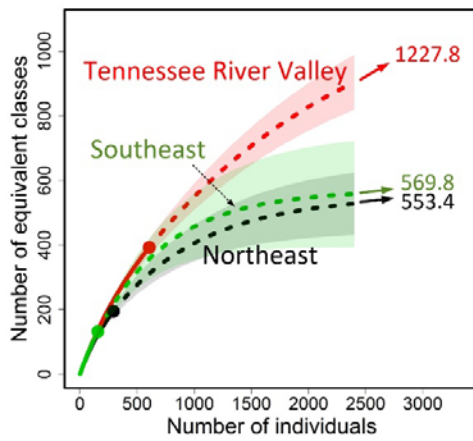
Dominant class richness ($q = 2$)
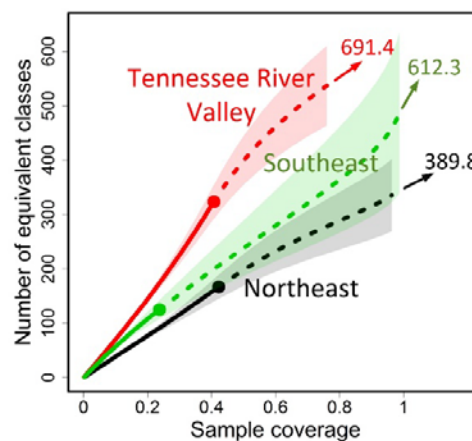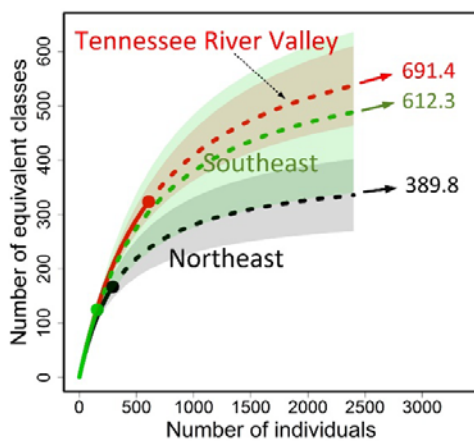
Supplementary Figure 3. (Unmerged traits) Comparison of sample-size-based (left panels) and sample-coverage-based (right panels) rarefaction and extrapolation for class richness (upper panels), Shannon diversity (middle panels) and Simpson diversity (lower panels) for the Tennessee River Valley, Southeast, and Northeast data. Observed samples are denoted by solid dots; rarefied segments are denoted by solid lines and extrapolated segments by broken lines. The extrapolation extends up to a maximum sample size of 1200 for class richness, and to a maximum sample size of 2400 for Shannon diversity and Simpson diversity. The sample-coverage-based extrapolation extends to the coverage value of the corresponding maximum sample size, for each sample. The 95% confidence intervals (shaded areas) were obtained by a bootstrap method based on 200 replications. The estimated asymptote of diversity for each curve is shown next to the arrow at the right-hand end of each curve.

Class richness (*q* = 0)
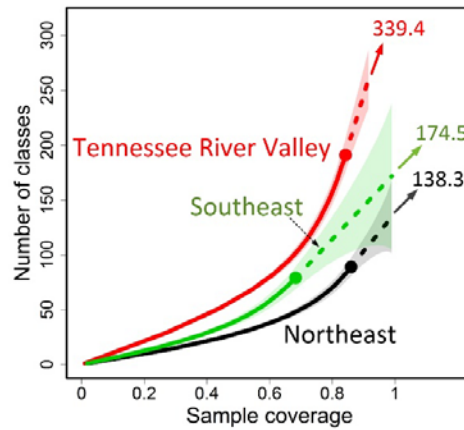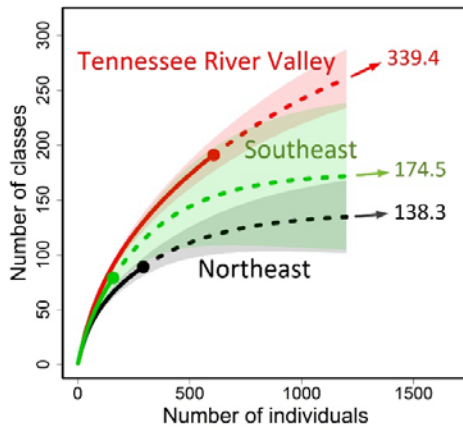
Common class richness (*q* = 1)
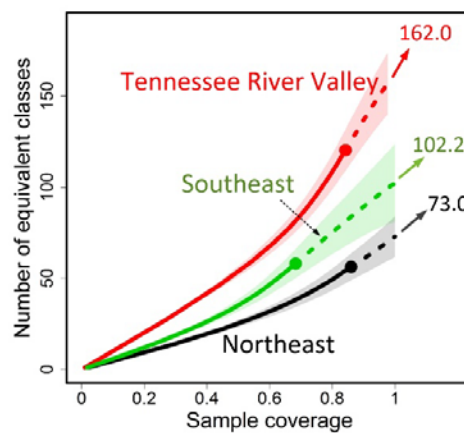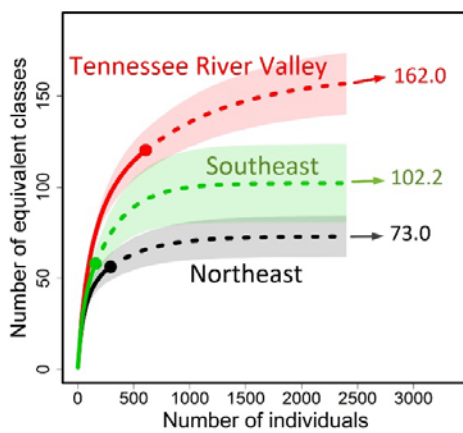
Dominant class richness (*q* = 2)
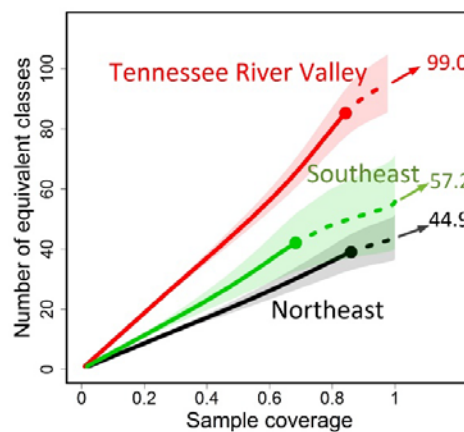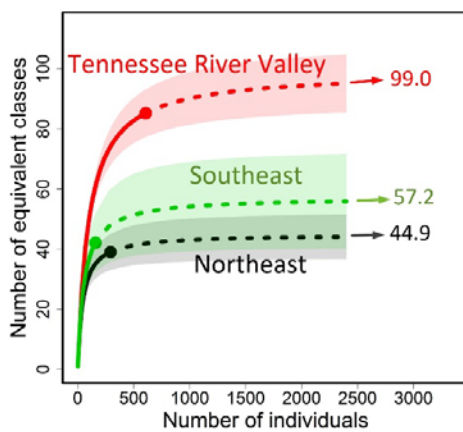
Supplementary Figure 4. (Merged traits) Comparison of sample-size-based (left panels) and sample-coverage-based (right panels) rarefaction and extrapolation for class richness (upper panels), Shannon diversity (middle panels) and Simpson diversity (lower panels) for the Tennessee River Valley, Southeast, and Northeast data. Observed samples are denoted by solid dots; rarefaction parts are denoted by solid lines and extrapolation parts are denoted by dotted curves. The extrapolation is extended up to a maximum sample size of 1200 for class richness, and to a maximum sample size of 2400 for Shannon diversity and Simpson diversity. The sample-coverage-based extrapolation is extended to the coverage value of the corresponding maximum sample size, for each sample. The 95% confidence intervals (shaded areas) are obtained by a bootstrap method based on 200 replications. The estimated asymptote of diversity for each curve is shown after an arrow sign.