

## TABLES S6–S15

The *Chao2* estimates and the associated 95% confidence intervals of species richness via a log-normal transformation for five regions based on locality by species incidence matrices (tables S6–S10) and date by species incidence matrices (tables S11–S15). In each of tables S6–S15, the *Chao2* estimate is calculated separately for all species data, described species data (formally described before the end of 2008), and undescribed/new species data (not formally described before the end of 2008). The percentages of described and undescribed in each region are determined by the *Chao2* estimates obtained from the described and undescribed species data.  $Q_1$ : number of species detected in only one sampling unit (uniques);  $Q_2$ : number of species detected in only two sampling units (duplicates); the *Chao2* species richness estimator formula is expressed as  $\hat{S}_{Chao2} = S_{obs} + (1 - 1/T) Q_1^2 / (2Q_2)$ , where  $S_{obs}$  denotes the number of observed species in the data, and  $T$  denotes the total number of sampling units (localities or days). 95% C. I.: 95% confidence interval. All numerical results were obtained from the online software SpadeR (Species-richness Prediction And Diversity Estimation in R) available at <https://chao.shinyapps.io/SpadeR/>. The complete incidence-based frequency counts for each sampling region are provided at the end of the supplementary material.

TABLE S6 (South America) The *Chao2* estimates based on locality by species incidence matrix

South America	Observed species #	Localities #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	242	101	141	44	465.7	(383.8, 594.8)
Described species	62	87	23	17	77.4 (15.10%)	(67.7, 103.8)
Undescribed/new species	180	95	118	27	435.1 (84.90%)	(330.7, 612.0)

TABLE S7 (Africa) The *Chao2* estimates based on locality by species incidence matrix

Africa	Observed species #	Localities #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	105	97	52	16	188.6	(145.1, 279.6)
Described species	30	76	9	5	38.0 (19.39%)	(31.8, 65.1)
Undescribed/new species	75	79	43	11	158.0 (80.61%)	(111.2, 265.1)

TABLE S8 (Asia) The *Chao2* estimates based on locality by species incidence matrix

Asia	Observed species #	Localities #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	225	110	144	42	469.6	(380.0, 611.0)
Described species	44	85	20	7	72.2 (15.35%)	(53.3, 129.7)
Undescribed/new species	181	102	124	35	398.5 (84.65%)	(313.8, 537.2)

TABLE S9 (Subtropical and Mediterranean) The *Chao2* estimates based on locality by  
species incidence matrix

Subtropical & Mediterranean	Observed species #	Localities #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	38	71	20	5	77.4	(50.3, 164.7)
Described species	18	42	6	4	22.4 (16.33%)	(18.8, 42.3)
Undescribed/new species	20	30	14	1	114.7 (83.67%)*	(35.9, 585.7)*

\*Unstable estimate and wide confidence interval due to sparse data (only one duplicate)

TABLE S10 (Caribbean) The *Chao2* estimates based on locality by species incidence matrix

Caribbean	Observed species #	Localities #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	79	62	43	17	132.5	(103.6, 195.5)
Described species	18	51	4	6	19.3 (13.05%)	(18.2, 28.0)
Undescribed/new species	61	47	39	11	128.7 (86.95%)	(89.9, 219.7)

TABLE S11 (South America) The *Chao2* estimates based on date by species incidence matrix

South America	Observed Species #	Days #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	242	132	99	67	314.6	(285.7, 362.5)
Described species	62	117	14	20	66.9 (20.69%)	(63.3, 79.6)
Undescribed/new species	180	124	85	47	256.2 (79.31%)	(224.1, 311.9)

TABLE S12 (Africa) The *Chao2* estimates based on date by species incidence matrix

Africa	Observed Species #	Days #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	105	98	44	22	148.6	(125.5, 197.5)
Described species	30	83	6	6	33.0 (21.61%)	(30.5, 46.6)
Undescribed/new species	75	78	38	16	119.5 (78.39%)	(94.7, 175.9)

TABLE S13 (Asia) The *Chao2* estimates based on date by species incidence matrix

Asia	Observed Species #	Days	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	225	114	133	50	400.3	(336.0, 501.8)
Described species	44	88	20	6	77.0 (19.16%)	(54.6, 146.4)
Undescribed/new species	181	108	113	44	324.8 (80.84%)	(268.9, 416.2)

TABLE S14 (Subtropical and Mediterranean) The *Chao2* estimates based on date by species incidence matrix

Subtropical & Mediterranean	Observed Species #	Days #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	38	47	19	8	60.1	(45.3, 105.2)
Described species	18	30	7	3	25.9 (43.40%)	(19.5, 59.1)
Undescribed/new species	20	23	12	5	33.8 (56.60%)	(23.6, 73.1)

TABLE S15 (Caribbean) The *Chao2* estimates based on date by species incidence matrix.

Caribbean	Observed Species #	Days #	$Q_1$	$Q_2$	The <i>Chao2</i> estimate (percentage of described/undescribed)	95% C. I. range
All species	79	58	36	14	124.5	(98.5, 184.9)
Described species	18	50	4	4	20.0 (15.21%)	(18.3, 32.1)
Undescribed/new species	61	53	32	10	111.2 (84.79%)	(81.1, 186.7)

## APPENDIX

### Complete incidence-based frequency counts in each geographic regions

#### Locality by species data

( $T$ : number of localities,  $Q_i$ : number of species that were detected in  $i$  localities)

Raw data are listed in tables S1–S5.

1. See table S6 for the *Chao2* estimate and 95% confidence interval

<b>South America</b>	$T$	$i =$	1	2	3	4	5	6	7	8	9	11	12	13	14	16	20	22
All	101	$Q_i =$	141	44	21	15	3	2	5	1	2	1	1	2	1	1	1	1
Described	87	$Q_i =$	23	17	7	3	2	1	1	1	2	0	1	0	1	1	1	1
Undescribed/new	95	$Q_i =$	118	27	14	12	1	1	4	0	0	1	0	2	0	0	0	0

2. See table S7 for the *Chao2* estimate and 95% confidence interval

<b>Africa</b>	$T$	$i =$	1	2	3	4	5	6	9	10	11	12	13
All	97	$Q_i =$	52	16	9	7	3	8	3	1	1	2	1
Described	76	$Q_i =$	9	5	3	2	1	4	1	1	0	1	1
Undescribed/new	79	$Q_i =$	43	11	6	5	2	4	2	0	1	1	0

3. See table S8 for the *Chao2* estimate and 95% confidence interval

<b>Asia</b>	$T$	$i =$	1	2	3	4	5	6	7	8	10	11	12
All	110	$Q_i =$	144	42	9	8	10	1	5	1	1	2	1
Described	85	$Q_i =$	20	7	2	5	5	0	2	0	0	1	1
Undescribed/new	102	$Q_i =$	124	35	7	3	5	1	3	1	1	1	0

4. See table S9 for the *Chao2* estimate and 95% confidence interval

<b>Subtropical &amp; Mediterranean</b>	<i>T</i>	<i>i</i> =	1	2	3	4	5	6	11
All	71	$Q_i =$	20	5	6	1	4	1	1
Described	42	$Q_i =$	6	4	3	1	3	1	0
Undescribed/new	30	$Q_i =$	14	1	3	0	1	0	1

5. See table S10 for the *Chao2* estimate and 95% confidence interval

<b>Caribbean</b>	<i>T</i>	<i>i</i> =	1	2	3	4	5	6	7	10	12	13	15
All	62	$Q_i =$	43	17	8	2	3	1	1	1	1	1	1
Described	51	$Q_i =$	4	6	2	0	2	0	0	1	1	1	1
Undescribed/new	47	$Q_i =$	39	11	6	2	1	1	1	0	0	0	0



## Date by species data

( $T$ : number of days,  $Q_i$ : number of species that were detected in  $i$  sampling days)

Raw data are listed in tables S1–S5.

1. See table S11 for the *Chao2* estimate and 95% confidence interval

<b>South America</b>	$T$	$i =$	1	2	3	4	5	6	7	8	9	12	13	14	20	21	22	23	24
All	132	$Q_i =$	99	67	26	17	10	4	2	4	3	2	2	1	1	1	1	1	1
Described	117	$Q_i =$	14	20	8	4	2	2	2	2	1	2	0	1	0	1	1	1	1
Undescribed/new	124	$Q_i =$	85	47	18	13	8	2	0	2	2	0	2	0	1	0	0	0	0

2. See table S12 for the *Chao2* estimate and 95% confidence interval

<b>Africa</b>	$T$	$i =$	1	2	3	4	5	6	7	8	9	10	11	12	15	19
All	98	$Q_i =$	44	22	10	5	4	7	1	3	1	2	1	2	2	1
Described	83	$Q_i =$	6	6	5	2	2	2	0	0	1	1	1	1	2	1
Undescribed/new	78	$Q_i =$	38	16	5	3	2	5	1	3	0	1	0	1	0	0

3. See table S13 for the *Chao2* estimate and 95% confidence interval

<b>Asia</b>	$T$	$i =$	1	2	3	4	5	6	7	8	9	10	13	14	23
All	114	$Q_i =$	133	50	13	6	9	3	4	1	1	1	1	2	1
Described	88	$Q_i =$	20	6	4	3	3	2	2	0	1	1	0	1	1
Undescribed/new	108	$Q_i =$	113	44	9	3	6	1	2	1	0	0	1	1	0

4. See table S14 for the *Chao2* estimate and 95% confidence interval

<b>Subtropical &amp; Mediterranean</b>	<i>T</i>	<i>i</i> =	1	2	3	4	5	8
All	47	$Q_i =$	19	8	7	1	2	1
Described	30	$Q_i =$	7	3	5	1	2	0
Undescribed/new	23	$Q_i =$	12	5	2	0	0	1

5. See table S15 for the *Chao2* estimate and 95% confidence interval

<b>Caribbean</b>	<i>T</i>	<i>i</i> =	1	2	3	4	5	6	8	9	11	12	17
All	58	$Q_i =$	36	14	9	6	5	3	2	1	1	1	1
Described	50	$Q_i =$	4	4	1	1	2	0	2	1	1	1	1
Undescribed/new	53	$Q_i =$	32	10	8	5	3	3	0	0	0	0	0