1   Supplementary Information
2

3   **Assessment of Insect Communities with Metabarcoding**

4

5   Hendra F. Sihaloho[1,2*], Isham Azhar[1,3], Millawati Gani[4], Juliana Senawi[3,5,6], Lim Shu Yong[7], Qasim
6   Ayub[7], Rahul De[1], Tigga Kingston[1,3], Caleb D. Phillips[1,2*]

7

8   [1]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131
9   [2]Natural Science Research Laboratory, Museum, Texas Tech University, Lubbock, TX 79409-3191
10   [3]Southeast Asian Bat Conservation Research Unit, Lubbock, TX,
11   [4]National Wildlife Forensic Laboratory, Ex-Situ Conservation Division, Department of Wildlife and National Parks
12   Peninsular Malaysia, Km. 10, Jalan Cheras, 56100, Kuala Lumpur, Malaysia
13   [5]Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan
14   Malaysia, 43600 Bangi, Selangor, Malaysia
15   [6]Langkawi Research Centre, Tuanku Abdul Halim Mu'adzam Shah Campus, Universiti Kebangsaan Malaysia, Jalan
16   Teluk Yu, Teluk Burau, 07100 Langkawi, Kedah, Malaysia
17   [7]Monash University Malaysia, Genomics Platform, School of Science, 47500 Bandar Sunway, Selangor Darul Ehsan,
18   Malaysia
19   *hendra.sihaloho@ttu.edu; caleb.phillips@ttu.edu
20

21   *Table S1. Analysis of Variance to test if the subsample combination affected the diversity recovered*
22   *(subsample combination at a given subsampling depth was a nested variable within the light trap).*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Signif
factor(sam)	5	209534481	41906896	72673.782	<2e-16	***
factor(sam):factor(sam_comb)	2964	1141755	385	0.668	1	
Residuals	8910	5137897	577			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aov(formula = jum ~ factor(sam)/factor(sam_comb), data = jum.all)

jum = Number of OTUs

sam = Light trap samples

sam_comb = Light trap subsample combination (12C4)

jum.all = an R dataframe object

23

24 *Supplementary re-analysis of Zizka et al. (2022).*
25 Data published by Zizka et al. (2022) was sub-sampled and re-analyzed using the same pipeline described herein.
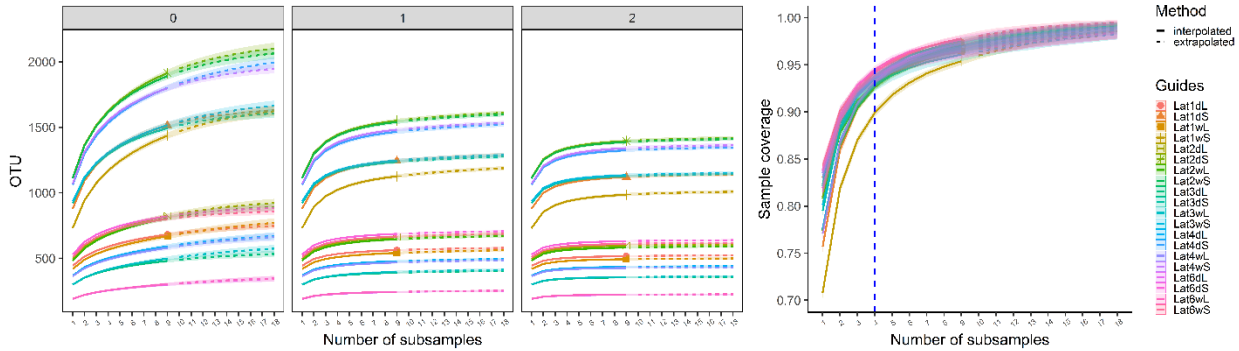26 The results were summarized in Figure S1 and Table S3.

27

28

29 Figure S1. Raw data for insect subsamples from Zizka et al. (2022) was randomly depleted to the average sequencing depth of
30 subsamples of the current study (23,119 read pairs). Rarefaction and extrapolation for different Hill numbers (Hill$_0$: richness, Hill$_1$:
31 exponential Shannon-Wiener, Hill$_2$: Inverse Simpson) and their sample coverage (most right) is reported. Each sample rarefaction
32 curve was extrapolated to twice of subsample depth and includes 95% confidence interval.

33

34 Table S2. Estimated Hill$_0$ with 4 subsamples resulted in 90.30 - 95.10% sample coverage (93.82 ± 0.99, mean ± standard deviation).
35 Site: sample name, t: number of sub-samples, SC: sample coverage, qD: estimate Hill$_0$, qD.LCL: lower confidence limit of Hill$_0$,
36 qD.UCL: upper confidence limit of Hill$_0$.

site	t	method	SC	qD	qD.LCL	qD.UCL
Lat1dL	4	interpolated	0	0.9490	555.68	547.01
Lat1dS	4	interpolated	0	0.9320	1248.50	1234.88
Lat1wL	4	interpolated	0	0.9410	539.86	529.28
Lat1wS	4	interpolated	0	0.9030	1118.44	1103.09
Lat2dL	4	interpolated	0	0.9380	638.44	628.18
Lat2dS	4	interpolated	0	0.9310	1557.50	1542.04
Lat2wL	4	interpolated	0	0.9360	629.95	619.33
Lat2wS	4	interpolated	0	0.9330	1550.76	1535.63
Lat3dL	4	interpolated	0	0.9430	379.75	372.47
Lat3dS	4	interpolated	0	0.9420	1246.03	1232.55
Lat3wL	4	interpolated	0	0.9370	385.33	375.83
Lat3wS	4	interpolated	0	0.9410	1255.87	1241.43
Lat4dL	4	interpolated	0	0.9420	464.11	454.78
Lat4dS	4	interpolated	0	0.9340	1475.30	1460.19
Lat4wL	4	interpolated	0	0.9420	456.82	447.03
Lat4wS	4	interpolated	0	0.9360	1491.66	1477.43
Lat6dL	4	interpolated	0	0.9430	226.29	219.54
Lat6dS	4	interpolated	0	0.9510	679.38	669.62
Lat6wL	4	interpolated	0	0.9430	228.05	221.47
Lat6wS	4	interpolated	0	0.9480	667.87	658.83

References

Zizka VMA, Geiger MF, Hörren T, Kirse A, Noll NW, Schäffler L, Scherges AM, Sorg M (2022) Repeated subsamples during DNA extraction reveal increased diversity estimates in DNA metabarcoding of malaise traps. Ecol Evol 12(11):e9502. https://doi.org/10.1002/ece3.9502