

# Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau

Qian Jin<sup>1</sup>, Huilin Han<sup>2</sup>, XiMin Hu<sup>1</sup>, XinHai Li<sup>3</sup>, ChaoDong Zhu<sup>3</sup>, Simon Y. W. Ho<sup>4</sup>, Robert D. Ward<sup>5</sup>, Aibing Zhang<sup>1\*</sup>

**1** College of Life Sciences, Capital Normal University, Beijing, People's Republic of China, **2** School of Forestry, Experiment Center, Northeast Forestry University, Haerbin, People's Republic of China, **3** Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, People's Republic of China, **4** School of Biological Sciences, University Of Sydney, Sydney, Australia, **5** Wealth from Oceans Flagship, CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia

## Abstract

With the ongoing loss of biodiversity, there is a great need for fast and effective ways to assess species richness and diversity: DNA barcoding provides a powerful new tool for this. We investigated this approach by focusing on the Tibetan plateau, which is one of the world's top biodiversity hotspots. There have been few studies of its invertebrates, although they constitute the vast majority of the region's diversity. Here we investigated species diversity of the lepidopteran family Noctuidae, across different environmental gradients, using measurements based on traditional morphology as well as on DNA barcoding. The *COI* barcode showed an average interspecific K2P distance of  $9.45 \pm 2.08\%$ , which is about four times larger than the mean intraspecific distance ( $1.85 \pm 3.20\%$ ). Using six diversity indices, we did not detect any significant differences in estimated species diversity between measurements based on traditional morphology and on DNA barcoding. Furthermore, we found strong positive correlations between them, indicating that barcode-based measures of species diversity can serve as a good surrogate for morphology-based measures in most situations tested. Eastern communities were found to have significantly higher diversity than Western ones. Among 22 environmental factors tested, we found that three (precipitation of driest month, precipitation of driest quarter, and precipitation of coldest quarter) were significantly correlated with species diversity. Our results indicate that these factors could be the key ecological factors influencing the species diversity of the lepidopteran family Noctuidae on the Tibetan plateau.

**Citation:** Jin Q, Han H, Hu X, Li X, Zhu C, et al. (2013) Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau. PLoS ONE 8(5): e64428. doi:10.1371/journal.pone.0064428

**Editor:** M. Alex Smith, University of Guelph, Canada

**Received:** December 20, 2012; **Accepted:** April 14, 2013; **Published:** May 31, 2013

**Copyright:** © 2013 Jin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by Beijing Municipal Natural Science Foundation Key Projects (KZ201010028028 to ZHANG) and by Natural Science Foundation of China (to ZHANG, grants 31071963 and 31272340, to HAN, grant 30700641), by Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (to ZHANG, grant PHR201107120), by The Research Fund for the Doctoral Program of Higher Education (to ZHANG, grant 20101108120002), and partially by PublicWelfare Project from th403 e Ministry of Agriculture, China (grant 201103024 to ChaoDong). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** CSIRO Marine and Government Research is a part of CSIRO, an Australian government-funded research organisation. It is not a "commercial company" and the employee of CSIRO Marine and Government Research, Dr. Ward, has no competing interests.

\* E-mail: zhangab2008@mail.cnu.edu.cn

## Introduction

The quantification of species diversity, and understanding the processes that drive variation in diversity across space and time, form the basis of some of the most fundamental questions in ecology and evolutionary biology. Measures of species diversity are often regarded as indicators of ecosystem health and function [1]. An extensive range of indices and models for measuring diversity have been devised, and these have been surrounded by considerable debate [1–8]. Traditionally, these have been based on morphology-based species identification for most metazoan groups. However, morphological identification alone is not always feasible, especially for a community with a largely undescribed biodiversity [9–10]. DNA barcoding has gained an important role as part of efforts to develop a global inventory of biodiversity [11–26], but it has also been accompanied by various reservations [27–35]. On 14 March 2013, there were 2,012,391 barcode records

from 172,280 species in the Barcode of Life Data Systems (BOLD) ([www.barcodinglife.org](http://www.barcodinglife.org)). About 74% of the records and 82% of the species are from arthropods. DNA barcoding studies can be generally classified into theoretical and empirical studies. Theoretical studies focus on the methodology of DNA barcoding, such as the precision of species assignment or identification with a variety of methods, including classical phylogenetic approaches [16,17,36,37], pure statistical approaches based on classification algorithms [38], approaches based on artificial intelligence [26,39] and an approach based on fuzzy-set-theory [40]; Empirical studies are typically performed by taxonomists or biologists on their own taxa to provide a reference dataset and to address issues such as phylogenetic relationships [41,42], species assignments of unknowns [41,43], and resolution of cryptic species complexes [44–46]. Until recently, relatively few barcoding studies have been performed to address ecological and evolutionary issues, such as species interactions, host-parasitoid relationships, food-web struc-

tures, DNA barcode accumulation curves [46–52]. In contrast, studies of microbiological community diversity have been more observant of the potential of diversity assessment with DNA barcoding [53–66]. However, the use of DNA barcoding for estimating biodiversity indices has not been widely evaluated for plants or animals [67,68], although some insect groups have been so-studied, including ants [69,70], flies [71–73], and wasps [74,75].

Here we focus on Tibetan moth species (family Noctuidae) to evaluate community species diversity with a DNA barcoding (DB)-based method, and compare it to the traditional morphology (TM)-based method. The Noctuidae constitute one of the largest lepidopteran families, containing some 20,000 species from 20 subfamilies worldwide. A total of 3,751 species have been reported in China [76–80]. The larvae of many of these species are pests of agriculture and forestry, causing serious damage to agricultural production in China [81]. Tibet, known as the roof of the world, has an average altitude of 4,000 meters. Its unique natural environment has resulted in unique assemblages of species and particular morphological adaptations. For example, Tibetan noctuid moths resident above 4,000 meters have smaller body sizes, darker body color, and/or more setae, which are thought to be adaptations to the strong winds and low temperatures at high elevations [81]. The Tibetan Plateau is a biodiversity hotspot and contains many rare species [82–85]. However, owing to the lack of documentation and a largely undescribed species assemblage, biodiversity assessments of the region have generally omitted invertebrates, which probably make up majority of the region's diversity. Since species identification with different barcoding methods has been systematically evaluated elsewhere [39,40,86–90], our study has three aims: (i) to investigate the relationship between TM-based and DB-based measures of species diversity; (ii) to compare species diversities of noctuid moths across different environmental gradients on the Qinghai-Tibetan Plateau; and (iii) to examine relationships between environmental factors and species diversity for this group.

## Results

### Phylogenetic Analysis and DNA Barcode Gaps

Our 615-bp alignment comprised 328 *COI* sequences, sampled from 68 species and 45 genera. After morphological examination, ten species were considered as potential new species based on differences in genitalia. These were included in our calculations of diversity because, upon being identified as ten distinct new species, they would not affect the calculation of diversity. These ten potential new species could be named as spe1, spe2, spe3, ..., spe10, symbolically, and species diversity could be computed as they are considered known species; no mathematical complications arise from the potential new species identified here as distinct species. Twenty-two genera comprised just a single individual of a single species, and these singletons are considered to be rare species in the community. The remaining 23 genera were each represented by more than one individual, with most of these forming monophyletic clades at the species level. There were two exceptions, *Xestia* and *Apamea*; the former was split into two distant clades on the NJ tree, whereas the latter was split into two subclades by *Apamea* and two unassigned sequences which might belong to potential new species. The *COI* barcode showed an average interspecific K2P distance of  $9.45 \pm 2.08\%$ , which is about five times the mean intraspecific distance ( $1.85 \pm 3.20\%$ ). However, there was no clear DNA barcode gap for the *COI* barcode (Fig. 1), indicating

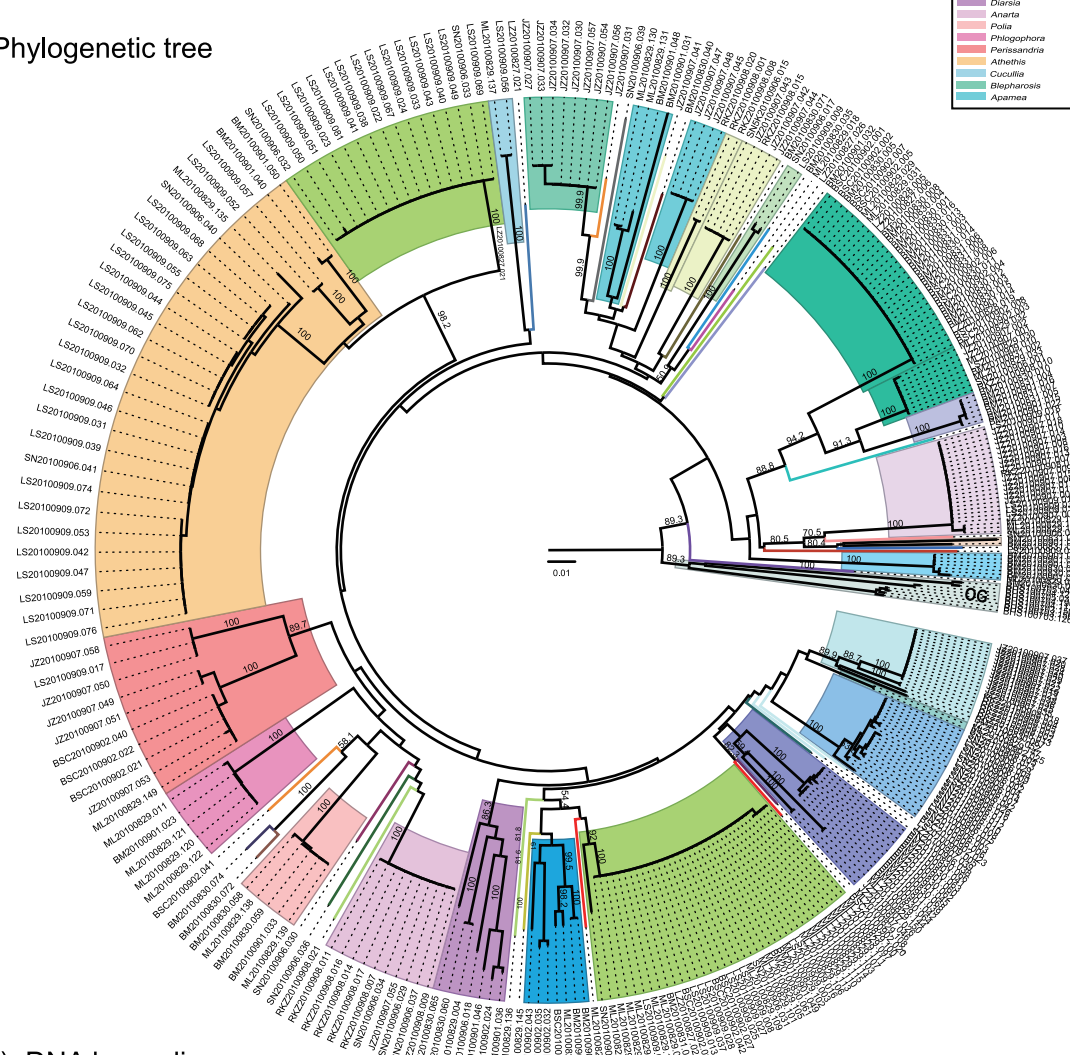
the difficulty of distinguishing sibling species and/or potential cryptic species within this lepidopteran group. These results are generally consistent with those based on the phylogenetic tree. Of the 58 known morphologically identified species, 23 (39.7%) were endemic to the Qinghai-Tibetan Plateau. A comparison of the tree topology between the neighbor-joining tree and the maximum-likelihood tree is presented in Appendix S2. There are some inconsistencies between them, indicating that tree-based barcoding methods are not suitable for determining the phylogeny of our Noctuidae moths. However, it is to be expected that a robust tree topology is not easily achieved with a short DNA marker, like the 648 bp *COI* barcode region.

### Comparing Measures of Species Diversity

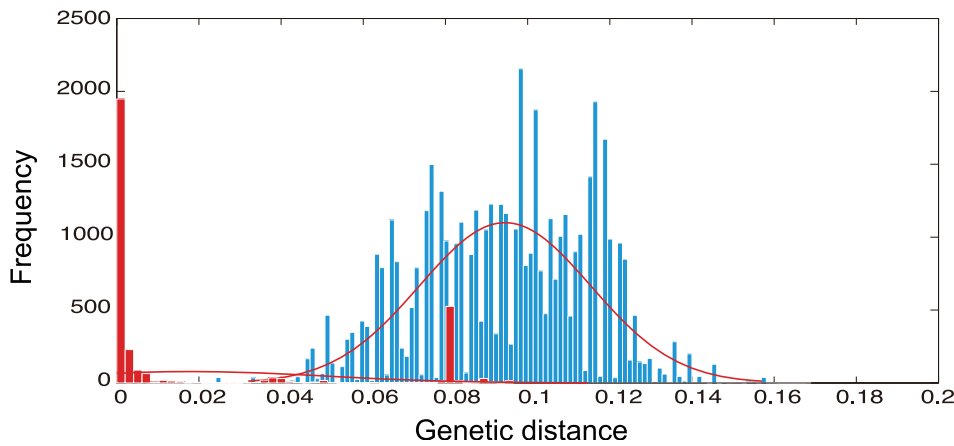
We computed both TM- and DB-based diversities for each community using the six different indices (Fig. 2a–e). For each community, we tested the null hypothesis that DB-based measures of species diversity do not differ from TM-based measures using 1000 replications for each of the six diversity indices. We did not detect any significant differences between TM- and DB-based diversities for the 03 ref and 05 ref data sets, which represent 30% and 50% proportions of database reference sizes respectively (Fig. 2, Appendix S2). For instance, the TM-based Shannon index for the Bome community was 2.7701, while the DB-based values were  $2.1754 \pm 0.1902$  (01 ref),  $2.3558 \pm 0.1502$  (03 ref), and  $2.5537 \pm 0.1091$  (05 ref) (Fig. 2a; Appendix S2;  $P = 0.4454, 0.4404$ , and  $0.4414$  respectively). This indicates that DB-based measures of species diversity can serve as good surrogates for TM-based measures in most of the situations tested in this study. Even for smaller reference sizes (10%), we also failed to reject the null hypothesis, with only a few exceptions such as the Shigatse, Lhasa, and Pagsum Co communities (Fig. 2a–f; Appendix S2). Compared with TM-based species diversities, DB-based methods tended to slightly underestimate diversities. For example, the TM-based Simpson index of the Shigatse community was 0.8199, whereas the DB-based values were  $0.7684 \pm 0.0459$  (01 ref) and  $0.8066 \pm 0.0263$  (03 ref) ( $P = 0.1121 - 0.2553$ ; Fig. 2a; Appendix S2). However, this could be largely corrected mathematically (see below). There were strong correlations between TM and DB-based diversities for these seven communities ( $R = 0.7781 - 0.9822$ ,  $P = 0.0001 - 0.0394$ ; Appendix S3). Simulations with different reference library sizes indicated that the larger the proportion of the reference database for DB-based species diversity, the closer become these estimates of diversity to the TM-based species diversity (Fig. 2a–f; Appendix S4). For example, for a low proportion of reference database (10%; 01 ref), a diversity value of  $1.6700 \pm 0.1445$  was estimated for the Shigatse (RKZ) community using the Shannon index and the DB-based method (Fig. 2a–2; Appendix S4). With an increasing proportion of the reference database (30% of reference database; 03 ref), the Shannon index for the Shigatse community increased ( $1.8491 \pm 0.0965$ ; Fig. 2a–2; Appendix S4), approaching the TM-based Shannon index value for the same community of 1.9408 (Fig. 2a–2; Appendix S4). For the Bome (BM) community, the same tendency was observed ( $2.1754 \pm 0.1902$ ,  $2.3558 \pm 0.1502$ ,  $2.5537 \pm 0.1092$ , for 01 ref, 03 ref and 05 ref, respectively with DB-based diversity measurements; Fig. 2a–2; Appendix S4). The 05 ref value was close to the TM-based species diversity measure (2.7701; Fig. 2a–2; Appendix S4). As mentioned above, the other five diversity indices show the same pattern as the Shannon index (Fig. 2b–f; Appendix S4).

Euxoa	Atrachea	Protexarmis
Unknown1	Helicoverpa	Unknown2
Dichagyris	Ephesia	Amathes
Agrotis	Catocala	Abagrotis
Xestia	Mascomucoughia	Mesia
Hermonassa	Ctenoplysia	Mamestra
Diarsia	Auchmis	Trichestra
Anarta	Cataglyphis	Hadenaria
Polla	Amyra	Lacinipolia
Phlogopthora	Himalaea	Trachea
Persandria	Unknown	Trichoridia
Althelia	Autographa	Unknown3
Cucullia	Orthozona	Oria
Bipharopsis	Olivenebulia	Lophoptera
Apamea	Amphipyra	Leucania

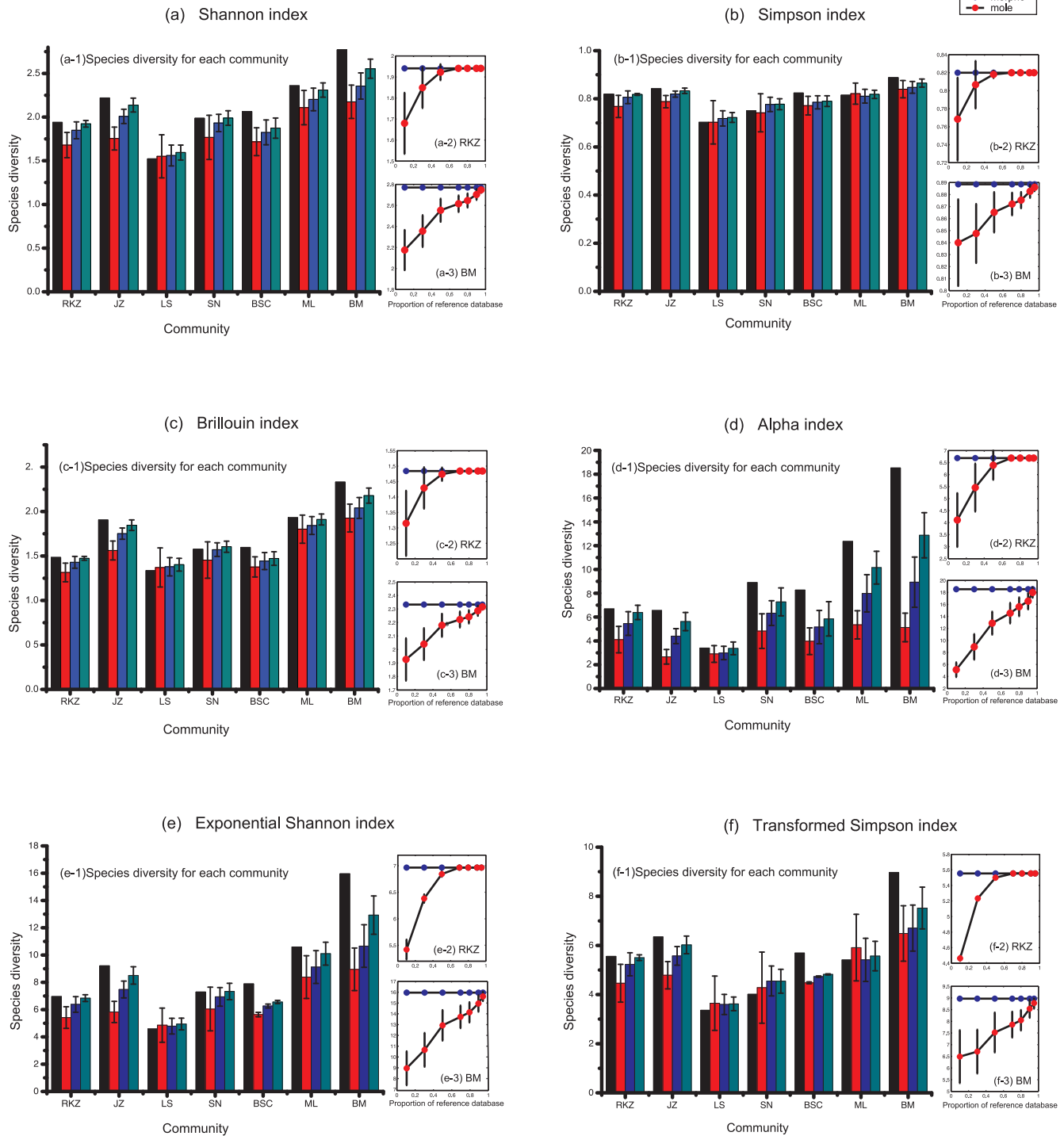
a) Phylogenetic tree



b) DNA barcoding gap



**Figure 1. (a) Phylogenetic tree of Tibetan moth species from the family Noctuidae, inferred using neighbor-joining analysis of 320 COI DNA sequences. Different colours represent different genera. (b) Plot of genetic distances showing the barcoding gap. Intraspecifics in red, interspecifics in blue. Fitted normal distribution curves are shown in red.**  
doi:10.1371/journal.pone.0064428.g001



**Figure 2. Species diversities of Tibetan moth species (Noctuidae) for each community, calculated using different diversity indices.** (a) Shannon index; (b) Simpson index; (c) Brillouin index; (d)  $\alpha$  index; (e) Exponential Shannon index; (f) Transformed Simpson index. doi:10.1371/journal.pone.0064428.g002

**Rank-abundance Plots and  $\beta$  Diversity**

We obtained both TM-based and DB-based rank-abundance (RA) curves for each community in Tibet (Fig. 3a-d). For the latter, we further examined the effect of different reference library sizes on the estimation of RA curves. All DB-based RA curves gave

similar and clear patterns for all seven communities when compared to the TM-based RA curve. For example, all DB-based RA curves showed that the Bome community in the east of Tibet has the highest species richness and species evenness (16,24, and 24 for 01 ref, 03 ref and 05 ref respectively), whereas the Shigatse

community in the west of Tibet has the lowest species richness and species evenness (6, 8, and 8 for 01 ref, 03 ref and 05 ref respectively); the TM-based RA curves gave values of 30 for Bome and 9 for Shigatse (Fig. 3a–d). The DB-based RA curves appear to underestimate species richness for each community to some degree, depending on the reference database. Therefore, caution should be still exercised when using DB-based RA curves in an ecological investigation, especially if the reference library is small.

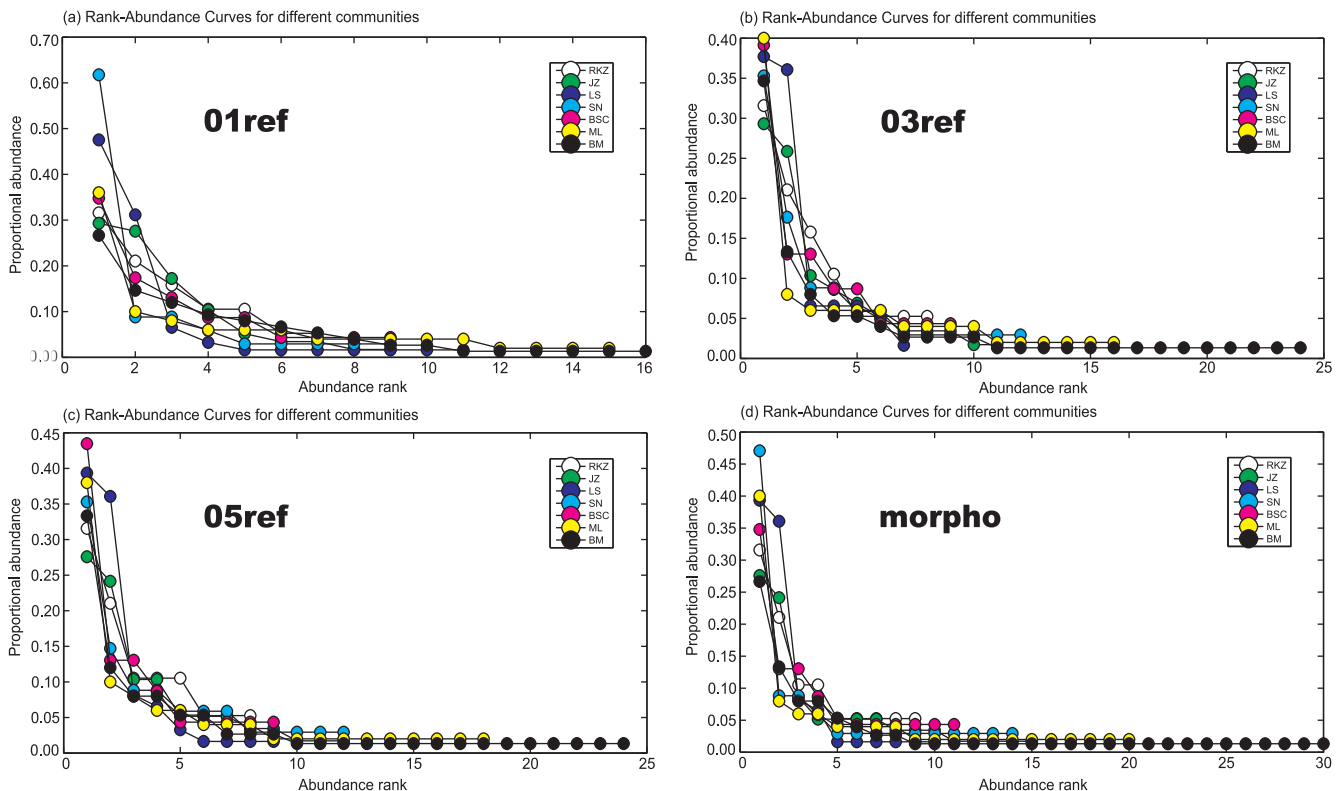
A Jaccard similarity index of 1 means that all species are shared across all areas, while an index of 0 means that no species are shared among the areas compared [8]. The average TM-based and DB-based Jaccard similarity indices were  $0.0949 \pm 0.0455$  and  $0.2225 \pm 0.0327$ , respectively, reflecting the low level of species sharing among the sampling sites. Mainling and Lhasa, Bome and Pagsum Co, had the highest values for the Jaccard similarity index (0.1724 and 0.1724, 5 species respectively), whereas Shigatse and Lhasa had the lowest Jaccard similarity index of 0.0217 (one species, Appendix S5).

The DB-based Jaccard similarity index gave much higher values than the TM-based one, suggesting that Jaccard similarity indices were overestimated by the DB-based method (Appendix S5). However, additional analyses with Mantel tests yielded significant correlations between TM- and DB-based Jaccard similarity indices ( $P < 0.0001$ , 1000 replications), indicating that TM-based and DB-based Jaccard similarity indices are in fact highly correlated.

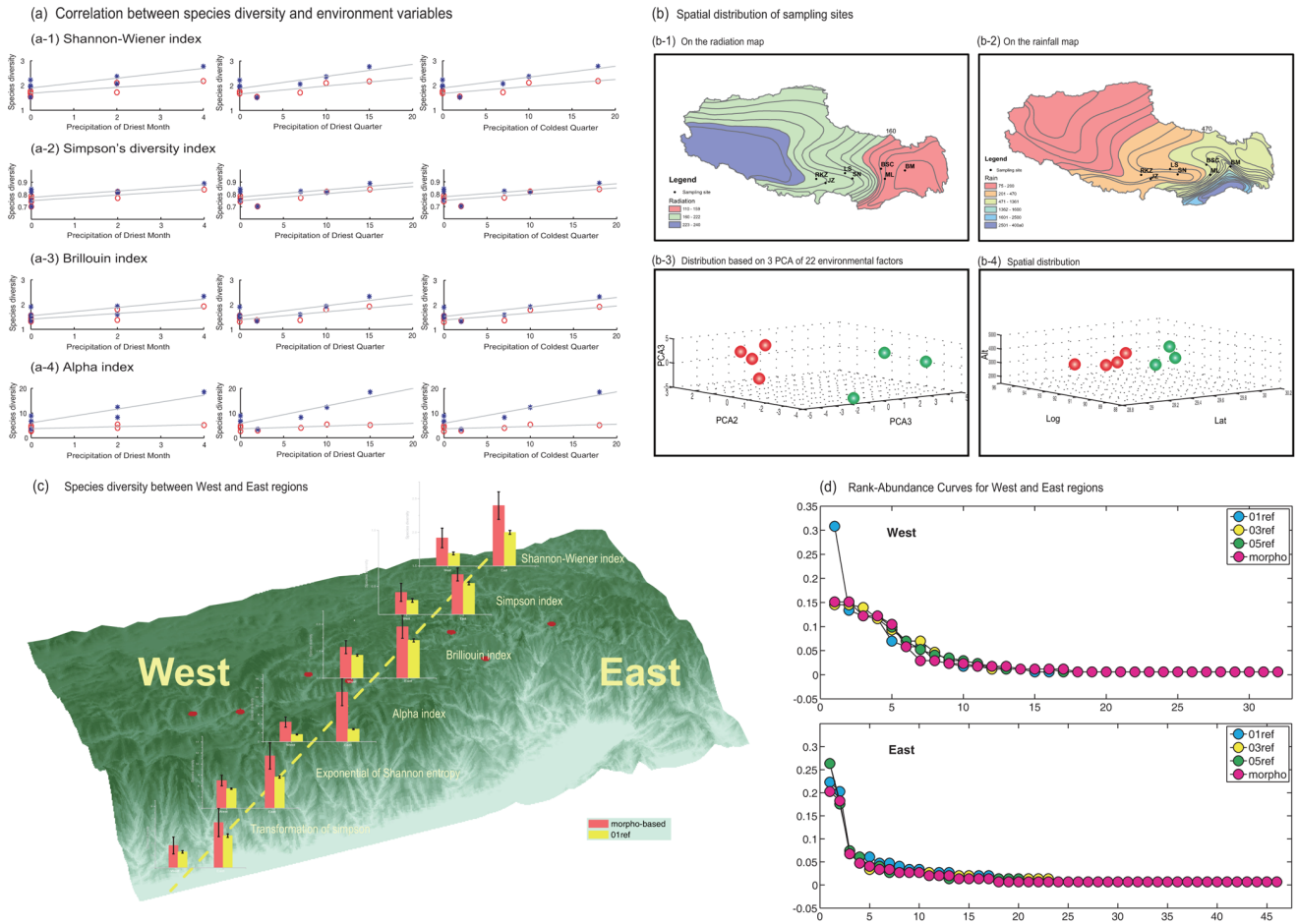
### Variation in Diversity across Different Environmental Gradients on the Qinghai-Tibetan Plateau

Among 22 environmental factors, three (Precipitation of Driest Month, Precipitation of Driest Quarter, and Precipitation of Coldest Quarter) were found to be significantly correlated with species diversity. For example, results for the Shannon index were  $R = 0.8043, P = 0.0292$  for TM-based species diversity and Precipitation of Driest Month,  $R = 0.8165, P = 0.0250$  for DB-based species diversity and Precipitation of Driest Month;  $R = 0.7547, P = 0.0499$  for TM-based species diversity and Precipitation of Driest Quarter,  $R = 0.8364, P = 0.0190$  for DB-based species diversity and Precipitation of Driest Quarter;  $R = 0.7782, P = 0.0393$  for TM-based species diversity and Precipitation of Coldest Quarter,  $R = 0.8393, P = 0.0182$  for DB-based species diversity and Precipitation of Coldest Quarter; for the Brillouin and  $\alpha$  indices (see Fig. 4a). However the Simpson index gave weaker correlations between environmental factors and species diversities ( $R = 0.5619 - 0.7649, P = 0.0451 - 0.1837$  for both TM- and DB-based species diversities; Fig. 4a). It is clear that different diversity indices varied in their power to identify correlations between species diversities and environmental factors. The  $\alpha$  index yielded the strongest correlations, the Simpson index the weakest; the remaining indices showed intermediate values (Fig. 4a).

TM- and DB-based measures of species diversity differ slightly in their ability to detect correlations between ecological factors and species diversities, depending on which species-diversity measure was used. For the Shannon index, DB-based measures detected a greater number of correlations between environmental variables



**Figure 3. Rank-abundance curves of each community based on DNA barcoding (DB) with different reference database (a–c) or traditional morphology (TM) (d).** (a) DB-based rank-abundance curves for each community based on reference size of 10% (01 ref). (b) DB-based rank-abundance curves for each community based on a reference size of 30% (03 ref). (c) DB-based rank-abundance curves for each community based on a reference size of 50% (05 ref). (d) TM-based rank-abundance curves for each community. doi:10.1371/journal.pone.0064428.g003



**Figure 4. Correlations between species diversity and environmental factors.** (a) Significant correlation between species diversity and three environmental factors, including precipitation of driest month, precipitation of driest quarter, and precipitation of coldest quarter. (b) Distribution of sampling sites/communities along different environmental gradients, including radiation and rainfall. Clustering of seven communities based on PCA of 22 environmental factors. (c) Spatial distribution of species diversities between West and East regions. (d) Rank-abundance curves of West and East communities.  
doi:10.1371/journal.pone.0064428.g004

and species diversities than TM-based measures. For example, a significant negative correlation between species diversity and altitude was detected by DB-based but not TM-based measures of species diversity ( $R=0.8571, P=0.0137$  for DB-based;  $R=0.6657, P=0.1023$  for TM-based; Fig. 4a).

Initial investigation of the spatial distribution of the seven communities, based on radiation and rainfall, found that they were distributed in two different ecological regions. In the west, the Shigatse, Gyantse, Lhasa, and Shannan communities are subject to strong radiation, low rainfall, and low humidity. In the east, the Pagsum Co, Mainling, and Bome communities are subject to weak radiation, high rainfall, and high humidity (Fig. 4b). Principal-components analysis based on all 22 environmental factors further confirmed that these seven communities belong to two different types of ecological environments: West and East (Fig. 4c; see also cover photos).

Further comparisons between these two regions indicated a significant, or nearly significant difference in species diversities, depending on which diversity measures were used (DB- or TM-based) (Fig. 4de). On average, five of the DB-based diversity indices (Shannon, Simpson, exponential of Shannon, transformed Simpson, and  $\alpha$ ), yielded higher values in eastern than western communities (Shannon,  $2.0005 \pm 0.0286$  for the East and

$1.6881 \pm 0.0198$  for the West,  $P < 0.0001$ ; Simpson,  $0.8110 \pm 0.0051$  for the East and  $0.7502 \pm 0.0066$  for the West,  $P < 0.0001$ ;  $\alpha$ ,  $4.8188 \pm 0.1377$  for the East and  $3.6243 \pm 0.1229$  for the West,  $P < 0.0001$ ; Fig. 4d). The Brillouin index showed the same trend as the other five diversity indices, but without a significant difference (Brillouin indices of  $1.7018 \pm 0.02917$  and  $1.4254 \pm 0.0173$  for the East and West communities, respectively;  $P = 0.1587$ ; exponential of Shannon,  $7.6581 \pm 0.2094$  for the East and  $5.5375 \pm 0.1123$  for the West,  $P < 0.0001$ ; transformed Simpson,  $5.6244 \pm 0.1439$  for the East and  $4.2948 \pm 0.1000$  for the West,  $P < 0.0001$ ; Fig. 4de). TM-based species-diversity analysis also generated a similar pattern to DB-based species diversity, showing a lower diversity in the West (Shannon,  $1.9179 \pm 0.1452$ ; Simpson,  $0.7789 \pm 0.0321$ ; Brillouin,  $1.5760 \pm 0.1207$ ;  $\alpha$ ,  $6.3875 \pm 1.1318$ ; exponential of Shannon,  $11.4800 \pm 2.3721$ ; transformed Simpson,  $6.6902 \pm 1.1433$ ; Fig. 4c) and higher diversity in the East (Shannon,  $2.3985 \pm 0.2043$ ; Simpson,  $0.8426 \pm 0.0231$ ; Brillouin,  $1.9538 \pm 0.2127$ ;  $\alpha$ ,  $13.0528 \pm 2.9826$ ; exponential of Shannon,  $7.0146 \pm 0.9497$ ; transformed Simpson,  $4.8191 \pm 0.6861$ ; Fig. 4c), with nearly statistical significance ( $P = 0.1076$ ;  $P = 0.1960$ ). The close to significant result was also detected by the  $\alpha$  diversity index ( $P = 0.0653$ ; Fig. 4d), which has been shown to be the most

sensitive diversity index by previous studies [1,8] and in the present study.

## Materials and Methods

### Sampling, DNA Extraction, PCR and Sequencing

We sampled 320 specimens from species of the moth family Noctuidae from seven locations in Tibet (Fig. 5). Samplings were performed with traditional light trap methods collecting moths overnight, typically, in a relatively short time period of some two weeks (Aug. 26–Sept.11) of the year, to avoid the effect of seasons on the diversity assessments. These specimens represented 68 species and 45 genera (Appendix S6; no specific permits were required for the described field studies, the locations are not privately-owned or protected in any way, and the field studies did not involve endangered or protected species). Three species (eight specimens) from the family Pyralididae [91] were included as outgroup taxa when inferring phylogenetic trees. Specimens were identified by one of us, an expert lepidopterist in East Asia (H.L.H.) [81,92–100]. Decisions on species status were based exclusively on morphological evidence, with male genital characters further examined when necessary (Appendix S7). DNA samples were prepared from individual insects by extraction of total DNA from animals either frozen or preserved in 100% ethanol. Genomic DNA was extracted using a BIOMED DNeasy kit. The mitochondrial *COI* gene was amplified via PCR using rTaq (TAKARA) with the primers LCO1490 (GGTCA ACAA TCATAA AGATA TTGG), and HCO2198 (TAAAC TTCAG GGTGA CCAAA AAATCA) [101]. The amplification reaction was performed in a total volume of 25  $\mu$ l, including 2.5  $\mu$ l 10  $\times$  buffer, 2.5  $\mu$ l 2.5 mM *MgCl*<sub>2</sub>, 2.0  $\mu$ l 2.5 mM dNTP, 0.1  $\mu$ l of each primer (10 mM), 1  $\mu$ l of template DNA, and 0.125  $\mu$ l of DNA *Taq* polymerase, and 16.675  $\mu$ l of distilled water. The PCR conditions were: 94°C for 2 minutes, 40 cycles of 94°C for 20 seconds, 54°C for 20 seconds, 72°C for 45 seconds, and a final extension at 72°C for 10 minutes. Sequencing was performed with an ABI3130 sequencer. The specimens were vouchered in a collection, and deposited in Capital Normal University (CNU, Beijing, China). Specimen data, trace files and sequences were deposited in BOLD in project Tibetan Lepidoptera (project code: NOCTU). All DNA sequences were also deposited in GenBank, with accession numbers JX392408 - JX392727.

### Phylogenetic Analysis

The raw DNA sequences were all checked manually and their ends were trimmed with BioEdit 7.0.1 (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>). The resulting alignment of 615 bp contained no gaps and all sequences could be correctly translated into amino acids. To infer phylogenetic relationships among these species, we performed a neighbor-joining analysis [37] using MEGA 4.0 [102] with the K2P model of nucleotide substitution. We also analysed sequence alignments using maximum likelihood with the program PHYML3.0 [103]. A search using nearest-neighbor interchange was conducted to gain a preliminary estimate of the phylogeny. We then conducted a search using subtree pruning and regrafting to estimate the maximum-likelihood tree. The K2P substitution model was used [16,17]. Branch support values were estimated using 1000 bootstrap replicates. All other parameters were set to their default values.

The distance between intraspecific and interspecific variation (the DNA barcode gap) is an important quantity in DNA barcoding practice. A large DNA barcode gap makes it easy to distinguish among species, whereas small or negative barcoding gaps tend to blur species boundaries and hamper species

assignment. To explore intraspecific and interspecific variation in noctuid moths, we performed an analysis of DNA barcode gaps using a custom Perl script [41].

### Species Diversity Analysis

**Traditional morphology-based method.** Given the large number of indices, it is often difficult to decide which is the best method of measuring diversity [1,3–8,67]. In this study, we use the following criteria to choose diversity measures: discriminant ability, sensitivity to sample size, what component of diversity is being measured, and whether or not the index is widely used. Six diversity indices were selected: the Shannon index, the exponential of Shannon index, the Simpson index, the transformed Simpson index, the Brillouin index, and the log series,  $\alpha$  index [1,3–8,67].

The most widely used measures of diversity are the information theory indices, which treat the diversity in a natural system in a similar way to the information contained in a message. The Shannon index assumes that individuals/species are randomly sampled from an effectively infinite population or community [104]. It is calculated using the equation:

$$H' = - \sum p_i \ln p_i \quad (1)$$

The quantity  $p_i$  is the proportion of individuals found in the  $i$ th species. The probability that any two individuals, randomly drawn from an infinitely large community, belong to different species is (Simpson 1949):

$$D = \sum p_i^2 \quad (2)$$

where  $p_i$  is the proportion of individuals in the  $i$ th species. As  $D$  increases, diversity decreases and so the Simpson index is usually expressed as:

$$D_{Simpson} = 1 - D \quad (3)$$

The Shannon index and the Simpson index, which have moderate discriminant ability and moderate sensitivity to sample size, are commonly used in ecological studies [105–107]. However, some have recently recommended the use of the exponential of the Shannon index and the transformed Simpson index because they have the “doubling property [3–7]. The exponential of the Shannon index is given by.

$$H_{exp} = e^{(- \sum p_i \ln p_i)} \quad (4)$$

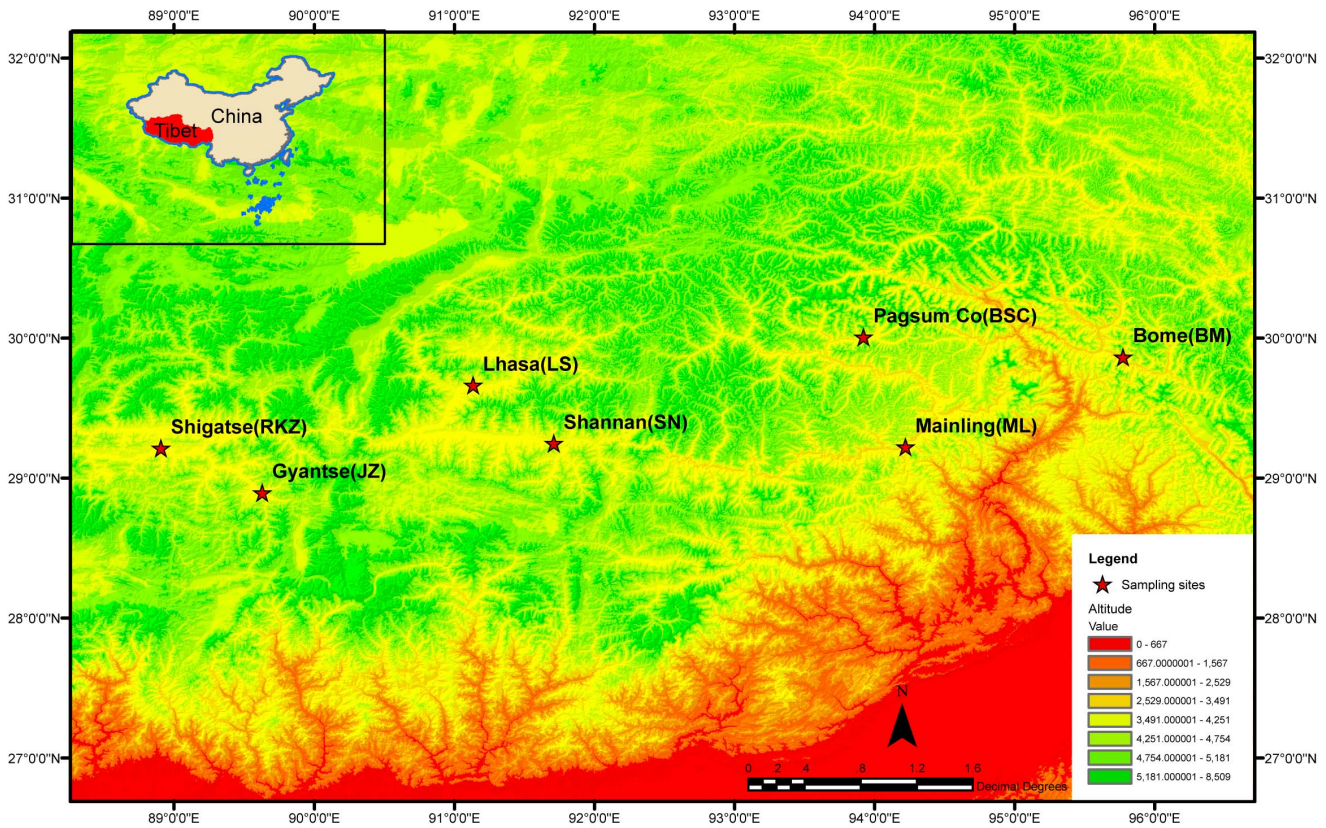
and the transformed Simpson index is given by

$$Simpson_{Transformed} = \frac{1}{(1 - D_{Simpson})} \quad (5)$$

The Brillouin index, which is not widely used, is given by.

$$HB = \frac{\ln N! - \sum \ln n_i!}{N} \quad (6)$$

where  $n_i$  is the number of individuals of species  $i$  and  $N$  is the total number of individuals in the sample. We also used the log series index,  $\alpha$ , owing to its good discriminant ability and because it is not unduly influenced by sample size [1,8,67]. The log series takes



**Figure 5. Sampling sites in the Qinghai-Tibetan Plateau in China.** Samples were collected from seven locations from West to East in Tibet, representing different ecological conditions. RKZ - Shigatse, Gyantse - JZ, Lhasa - LS, Shannan - SN, Pagsum Co - BSC, Mainling - ML, and Bome - BM. doi:10.1371/journal.pone.0064428.g005

the form:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \dots, \frac{\alpha x^n}{n} \tag{7}$$

where  $\alpha x$  and  $\alpha x^2/2$  is the number of species predicted to have one and two individuals respectively, and so on (Fisher et al. 1943; Poole, 1974). The index can be obtained from the equation:

$$\alpha = \frac{N(1-x)}{x} \tag{8}$$

where  $N$  is the total number of individuals,  $x$  is estimated from the iterative solution of

$$S/N = \left(1 - \frac{x}{x[-\ln(1-x)]}\right) \tag{9}$$

**DNA barcoding-based method.** For a community with a very large number of hyperdiverse taxa, such as arthropods, investigations of species diversity are usually hindered by the difficulty of identifying species by traditional morphological means alone. DNA-based species identification [16,17] provides an alternative method for assessment of species diversity. The commonly used barcoding gene, *COI*, was sequenced for all samples. Their species identities were assessed by matching their *COI* barcodes against a reference barcode library. Generally, species in a community are not necessarily closely phylogenetically

related, but they are ecologically related, due to the broad taxon coverage and filtering of environmental factors for a community. Therefore, instead of using phylogeny-based DNA barcoding methods, a non-tree-based Bayesian approach, which takes advantages of both Bayesian theory and bioinformatics, was used to infer species identity via *COI* barcoding (Fig. 6).

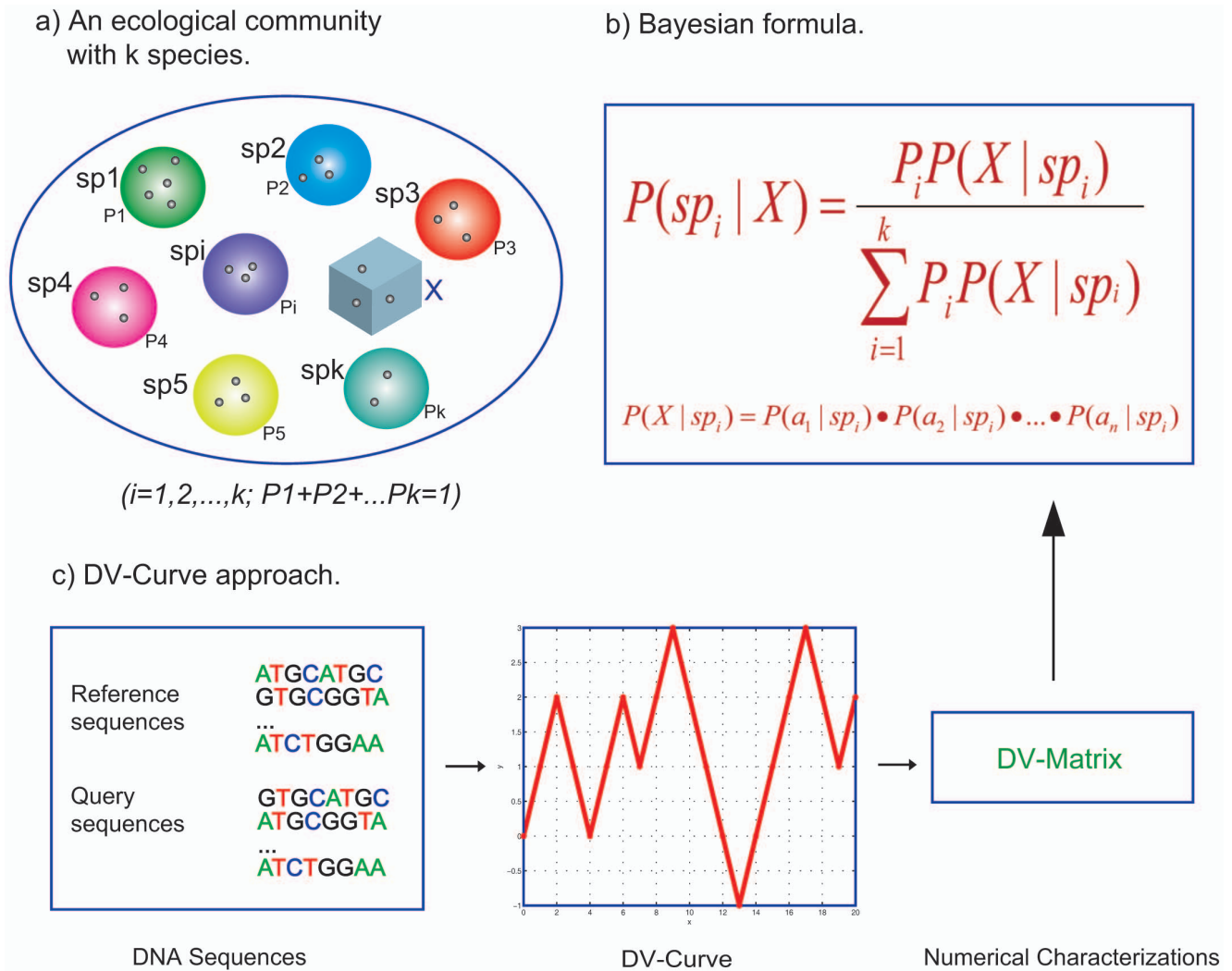
The dual-vector curve (DV-Curve) was proposed by Zhang [108] as a two-dimensional graphical representation for visualizing and analysing DNA sequences (Fig. 6). It is able to represent DNA sequences without degeneracy and loss of information. Let us consider a DNA sequence  $S = s_1 s_2 \dots s_n$  consisting of  $n$  nucleotide sites. Let  $(X_i, Y_i)$  be the point of the DV-Curve, where  $(X_0, Y_0) = (0, 0)$  is the start point. The DV-Curve is uniquely determined by the following formula [108]:

$$Y_{2i-1} = \begin{cases} Y_{2i-2} + 1, & \text{if } s_i = A \text{ or } T \\ Y_{2i-2} - 1, & \text{if } s_i = C \text{ or } G \end{cases} \tag{10}$$

$$Y_{2i} = \begin{cases} Y_{2i-1} + 1, & \text{if } s_i = A \text{ or } C \\ Y_{2i-1} - 1, & \text{if } s_i = T \text{ or } G \end{cases} \tag{11}$$

$$X_{2i-1} = 2i - 1 \tag{12}$$





**Figure 6. Species identity inferred via DNA barcoding for a community with a Bayesian method.** (a) An ecological community with species. (b) Bayesian formula used to infer species membership for unknown samples. (c) DV-Curve approach to construct DV-Matrix for Bayesian inference.  $sp_i$  is the  $i$ th species in the community,  $X$  is an unknown sample with DNA sequenced,  $P_i$  is the probability of a sample belonging to  $sp_i$ .  $P(sp_i|X)$  is the conditional probability of the unknown sample  $X$  belonging to  $sp_i$  given that its DNA is sequenced.  $P(X|sp_i)$  is the probability of having DNA sequence of the unknown sample  $X$  given that DNA sequences of  $sp_i$  are known. doi:10.1371/journal.pone.0064428.g006

$$X_{2i} = 2i \tag{13}$$

where  $i = 1, 2, \dots, n$ .

To numerically characterize a DNA sequence via the DV-Curve, a 24-component vector  $\vec{D}$  as described by Zhang [108] was used:

$$\vec{D}_{DV} = [CM_{1_{xy}}, CM_{2_{xy}}, \dots, CM_{24_{xy}}] \tag{14}$$

The  $CM_{xy}$  value [109] is calculated as follows:

$$(X_c, Y_c) = \left( \frac{1}{2n+1} \sum_{j=0}^{2n} X_j, \frac{1}{2n+1} \sum_{j=0}^{2n} Y_j \right) \tag{15}$$

$$CM_{xy} = \frac{1}{2n+1} \sum_{j=0}^{2n} (X_j - X_c)(Y_j - Y_c) \tag{16}$$

To get equation (14), we need to assign  $A$ ,  $T$ ,  $G$ , and  $C$  to basic Dual-Vectors in  $4!$  different ways to have  $4! = 24$  different DV-Curves for a given DNA sequence. The vector  $\vec{D}_{DV}$  is further used as the input for the Bayes prior attributes.

$$P(sp_i|X) = \frac{P_i P(X|sp_i)}{\sum_{i=1}^k P_i P(X|sp_i)} \tag{17}$$

Where  $sp_i$  is the  $i$ th species in the community,  $X$  is an unknown sample with a DNA sequence,  $P_i$  is the probability of a sample belonging to  $sp_i$ .  $P(sp_i|X)$  is the conditional probability of the

unknown sample  $X$  belonging to  $sp_i$  given that its DNA is sequenced.  $P(X|sp_i)$  is the probability of having the DNA sequence of the unknown sample  $X$  given that the DNA sequences of  $sp_i$  are known (Fig.6).

To explore the effect of reference library size on the assessment of species diversity, we randomly selected as the reference library different scales of the reference database, including 10%, 30%, 50%, 70%, 80%, 90%, and 95% of the whole dataset. These are subsequently referred to as 01 ref, 03 ref, and so on. The full reference library comprises the 322 barcoded and morphologically identified specimens. Once species identities were inferred with barcoding-based methods, species diversity indices were calculated following equations 1–6 and 9. We computed species diversity for each of seven communities in Tibet, and for the whole Tibetan community. All of these calculations were performed with 30 random replications. The differences in species diversity among different communities were statistically examined via a permutation test with 1000 replications. Likewise, the permutation test was also applied to assess the differences in species diversities between morphology-based and barcoding-based methods. To investigate further the relationship between TM- and DB-based species diversities, we additionally performed a correlation analysis between these two types of measurement for all six diversity indices.

### Rank-abundance Curves Based on Morphology or DNA Barcodes

Species abundance distribution can be visualized in different ways, of which the rank-abundance curve (rank/abundance plot) is one of the best known and most informative ones [67,110]. In this species are plotted in sequence from most to least abundant along the horizontal axis, with abundances displayed on the y axis [67]. One of the advantages of a rank-abundance curve is the clearly display of contrasting patterns of species richness and their relative abundances, compared with the inefficient presentation of a histogram [67,111].

By plotting the relative abundance of species against their rank in abundance, we can readily gain information about the diversity of species within a community. We investigated whether DB-based rank-abundance curves could serve as an effective surrogate for TM-based rank-abundance ones. Therefore, we computed both TM- and DB-based rank-abundance plots, and further examined them with the Kolmogorov-Smirnov two-sample test [112].

### Beta Diversity and Correlations between Species Diversities and Environmental Factors

Diversities ( $\beta$ ) between sites/communities were compared using the Jaccard Index of similarity [8],  $J = A/(A + B + C)$ , where A is the number of species shared between the two sites, and B and C are the number of species unique to each site.

To explore the relationship between species diversity and environmental factors, we examined 22 variables, including annual mean temperatures, mean diurnal range, and precipitation of the driest quarter (Appendix S8). Climate data are from the WorldClim dataset [113]. Both TM- and DNA-based indices of species diversity were calculated for the six species-diversity indices. Correlation analyses were performed between each species diversity index and each environmental factor.

The seven sampling sites/communities are distributed across quite different ecological conditions in Tibet, ranging from the dry west to the wetter east. We conducted a principal-component analysis [114] of the 22 environmental variables in order to reduce

the correlation between these variables and to cluster the seven sites/communities.

Diversities of moth species between these two regions (Shigatse, Gyantse, Lhasa, and Shannan in the west, and the remaining three communities in the east) were further compared for the six diversity indices using a t-test. We also constructed rank-abundance plots and evaluated them with the Kolmogorov-Smirnov two-sample test for the two regions [112].

### Discussion and Conclusions

In our study of biodiversity assessment via DNA barcoding for a Tibetan moth community, we found that diversity measurements based on DNA barcoding are able to serve as good surrogates for morphology-based measures. Compared with traditional morphology-based methods, a DNA barcode approach, along with appropriate analytical procedures as outlined here, is fast and convenient and can greatly reduce the time needed for specimen identification. However, we note that our current method is subject to some limitations, for example, requiring a pre-defined reference library.

We have not applied a method based solely on molecular operational taxonomic units, although such approaches have been successfully implemented elsewhere. One of the first ecologically-targeted barcoding studies, of Madagascan ants, applied COI distance thresholds of 2% and 3% to differentiate taxa; it showed that biodiversity richness was not significantly different whether estimated by morphological means or from molecularly-defined taxa [70]. Microbiological community studies have applied a threshold of 3% dissimilarity in 16 S rRNA sequences to identify conspecifics [54,56,60,61]. In fact, barcoding has proven to be an excellent tool for biodiversity surveys where the studied taxa do not have a solid taxonomic foundation [115].

However, an arbitrary threshold may be largely taxon-dependent [24,28,116], and was not applied in current study. Besides the threshold method, the generalized mixed Yule-coalescent (GMYC) model could be an alternative approach for species delimitation based on DNA sequences [21]. The GMYC method models branching events between species with a Yule model [117] and branch events within species using a neutral coalescent model [118]. The method has potential for biodiversity assessments although empirical studies are desirable to examine its consistency with traditional morphospecies since splitting or lumping was often observed compared to traditional taxonomy. An artificial intelligence-based approach, such as BP-based species identification [26], could be another choice for species identification, but was not applied to the current study due to its relatively slow training process during the construction of neural networks, making it impractical for the simulation study. The development of fast algorithms of this method for large datasets is highly desirable. In addition, it is generally recognised that multiple markers, especially the inclusion of nuclear genes, would increase the accuracy of species delimitations, and therefore the accuracy of species diversity assessment. However, we only applied the standard COI barcode in assessments of species diversity because of the widespread generality of the commonly used primers [16,101].

We applied a non-tree based Bayesian method for species identification before the calculation of diversity mainly because species in a community are not necessarily closely phylogenetically related, although they are ecologically related, due to the broad taxon coverage and filtering of environmental factors for a community. Tree-based methods, including Bayesian phylogenetic methods, are not always able to produce reliable assignments of specimens when taxon sampling is incomplete, as is the case here.

In this study, we focus mainly on assessments of species diversity, not barcoding approaches, as the latter have been systematically evaluated with both simulated and empirical data [39,40,86–90] and are outside the scope of the present study.

We found a significant positive correlation between species diversity and precipitation of driest month/coldest quarter for this taxon group. This environmental factor might affect the survival of pupae by changing the moisture of the micro-environments, or by influencing the diversity of their host plants. Our results are supported by the significantly higher diversity in the drier eastern communities compared with the wetter western communities on the Tibetan Plateau, where dramatically different ecological landscapes are present. However, owing to the lack of historical climate records in the sampling sites under investigation, the climate elements considered, such as monthly precipitation and mean, minimum, and maximum temperature, were all derived from the WorldClim dataset [113], which has been commonly used in ecological studies, but is restricted to records from the 1950–2000 period. However, this is unlikely to significantly affect our basic conclusions, because the assemblage of species for a certain community is not a consequence of short time interactions between species and environments, and among species but a long-term adaptation to certain habitats.

In the future, biodiversity assessments may be further accelerated by meta-barcoding [119]; meta-barcoding of bulked arthropod samples has recently been shown to provide good estimates of within and between community diversity [120]. Whether the preference is for the sequencing of bulked or individual specimens, it is abundantly clear that DNA barcode-based biodiversity assessments are very effective and efficient alternatives to more time-consuming morphological-based measures that require considerable taxonomic expertise.

### Data Accessibility

DNA sequences: Genbank accessions JX392408 - JX392727.

Specimen data, trace files and sequences were deposited at BOLD under a project Tibetan Lepidoptera.

(Project Code: NOCTU; [http://www.barcodinglife.com/index.php/MAS\\_Management\\_ProjectList](http://www.barcodinglife.com/index.php/MAS_Management_ProjectList)).

### References

- Magurran A (1988) *Ecological Diversity and Its Measurement*. Princeton University Press, 1st edition, 179.
- Cabaret J, Schmidt E (2001) Species diversity of nematode communities in the digestive tract of domestic ruminants: multivariate versus univariate estimations. *Parasitol Res* 87: 311–6.
- Jost L (2006) Entropy and diversity. *Oikos* 113: 363–375.
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88: 2427–2439.
- Jost L (2008) Gst and its relatives do not measure differentiation. *Mol Ecol* 17: 4015–4026.
- Jost L (2009) Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecol Econ* 68: 925–928.
- Jost L, DeVries P, Walla T, Greeney H, Chao A, et al. (2010) Partitioning diversity for conservation analyses. *Divers Distrib* 16: 65–76.
- Magurran A (2004) *Measuring Biological Diversity*. Wiley-Blackwell.
- Blaxter M, Floyd R (2003) Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology & Evolution* 18: 268–269.
- May R (1988) How many species are there on earth? *Science* 241: 1441–1449.
- Abdo Z, Golding GB (2007) A step toward barcoding life: a model-based, decision-theoretic method to assign genes to preexisting species groups. *Syst Biol* 56: 44–56.
- Dinca V, Zakharov EV, Hebert PD, Vila R (2011) Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proc Biol Sci* 278: 347–55.
- Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. *Nature* 434: 697.
- Gregory TR (2005) DNA barcoding does not compete with taxonomy. *Nature* 434: 1067.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci U S A* 103: 968–71.
- Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* 270: 313–21.
- Hebert PD, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytoch529 rDNA c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci* 270 Suppl 1: S96–9.
- Marshall E (2005) Taxonomy. Will DNA bar codes breathe life into classification? *Science* 307: 1037.
- Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, et al. (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9: 214.
- Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, et al. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst Biol* 58: 298–311.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, et al. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* 55: 595–609.
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philos Trans R Soc Lond B Biol Sci* 360: 1805–11.

### Supporting Information

**Appendix S1** Comparison of tree topology between NJ and ML trees. The nodes in green on the tree indicate shared nodes between two trees, nodes in dark grey are nodes present on the NJ tree, but not on the ML tree. (PDF)

**Appendix S2** Permutation tests for TM and DB-based species diversities, and diversities for each community. (TXT)

**Appendix S3** Correlation analysis between TM and DB-based species diversities for six different diversity indices. (XLS)

**Appendix S4** The effect of different reference database sizes on the estimation of species diversity. (PDF)

**Appendix S5**  $\beta$  diversity among different communities. (TXT)

**Appendix S6** Sample localities in Tibet, geographical coordinates and sequences used for Noctuidae moth species. (PDF)

**Appendix S7** Species identification based on traditional morphology. a) Some pictures of moths used in this study; b) Female genitalia, example 1; c) Male genitalia, example 2. (JPG)

**Appendix S8** Twenty two environmental variables used in this study. (TXT)

### Acknowledgments

We gratefully acknowledge the constructive comments of Academic Editor, Dr. M. Alex Smith, and two anonymous referees on an earlier version of the manuscript.

### Author Contributions

Conceived and designed the experiments: ABZ. Performed the experiments: QJ. Analyzed the data: QJ ABZ HLH XMH. Wrote the paper: ABZ QJ. Contributed substantially to revisions: CDZ SYWH RDW QJ HLH XMH XHL ABZ.

23. Schindel DE, Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Nature* 435: 17.
24. Ward RD, Hanner R, Hebert PD (2009) The campaign to DNA barcode all fishes, FISH-BOL. *J Fish Biol* 74: 329–56.
25. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005) DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci* 360: 1847–57.
26. Zhang AB, Sikes DS, Muster C, Li SQ (2008) Inferring species membership using DNA sequences with back-propagation neural networks. *Syst Biol* 57: 202–15.
27. Brower A (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperidae). *Syst Biodivers* 4: 127–132.
28. Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst Biol* 55: 729–39.
29. Lou M, Golding GB (2010) Assigning sequences to species in the absence of large interspecific differences. *Mol Phylogenet Evol* 56: 187–94.
30. Meier R, Shiyang K, Vaidya G, Ng PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol* 55: 715–28.
31. Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biol* 2: e354.
32. Silva-Brandao KL, Lyra ML, Freitas AV (2009) Barcoding Lepidoptera: current situation and perspectives on the usefulness of a contentous technique. *Neotrop Entomol* 38: 441–51.
33. Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci U S A* 105: 13486–91.
34. Whitworth TL, Dawson RD, Magalon H, Baudry E (2007) DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proc Biol Sci* 274: 1731–9.
35. Will K, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 4755.
36. Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Mol Phylogenet Evol* 43: 530–42.
37. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–25.
38. Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, et al. (2009) DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics* 10 Suppl 14: S10.
39. Zhang AB, Feng J, Ward RD, Wan P, Gao Q, et al. (2012) A new method for species identification via protein-coding and non-coding DNA barcodes by combining machine learning with bioinformatics methods. *PLoS One* 7: e30986.
40. Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, et al. (2012) A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Mol Ecol* 21: 1848–63.
41. Dai QY, Gao Q, Wu CS, Chesters D, Zhu CD, et al. (2012) Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. *PLoS One* 7: e32544.
42. Tang RW, Yau C, Ng WC (2010) Identification of stomatopod larvae (Crustacea: Stomatopoda) from Hong Kong waters using DNA barcodes. *Mol Ecol Resour* 10: 439–48.
43. Wilson JJ (2011) Assessing the value of DNA barcodes for molecular phylogenetics: effect of increased taxon sampling in Lepidoptera. *PLoS One* 6: e24769.
44. Bradford T, Adams M, Humphreys WF, Austin AD, Cooper SJ (2010) DNA barcoding of stygofauna uncovers cryptic amphipod diversity in a calcrete aquifer in western Australia's arid zone. *Mol Ecol Resour* 10: 41–50.
45. Hubert N, Meyer CP, Bruggemann HJ, Guerin F, Komeno RJ, et al. (2012) Cryptic diversity in Indo-Pacific 608 coral-reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. *PLoS One* 7: e28987.
46. Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, et al. (2011) Barcoding a quantified food web: cryptic concepts, ecology and hypotheses. *PLoS One* 6: e14424.
47. Clare EL, Barber BR, Sweeney BW, Hebert PD, Fenton MB (2011) Eating local: influences of habitat on the diet of little brown bats (*Myotis lucifugus*). *Mol Ecol* 20: 1772–80.
48. Hreck J, Miller SE, Quicke DL, Smith MA (2011) Molecular detection of trophic links in a complex insect host-parasitoid food web. *Mol Ecol Resour* 11: 786–94.
49. Leray M, Boehm J, Mills S, Meyer C (2012) *Moorea* barcode library as a tool for understanding predator-prey interactions: insights into the diet of common predatory coral reef fishes. *Coral Reefs* 31: 383–388.
50. Quicke DL, Alex Smith M, Janzen DH, Hallwachs W, Fernandez-Triana J, et al. (2012) Utility of the DNA barcoding gene fragment for parasitic wasp phylogeny (Hymenoptera: Ichneumonidae): data release and new measure of taxonomic congruence. *Mol Ecol Resour* 12: 676–685.
51. Rougerie R, Smith MA, Fernandez-Triana J, Lopez-Vaamonde C, Ratnasingham S, et al. (2011) Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Mol Ecol* 20: 179–86.
52. Smith MA, Fernandez-Triana J, Roughley R, Hebert PD (2009) DNA barcode accumulation curves for understudied taxa and areas. *Mol Ecol Resour* 9 Suppl s1: 208–16.
53. Fernandez-Gomez MJ, Nogales R, Insam H, Romero E, Goberna M (2011) Role of vermicompost chemical composition, microbial functional diversity, and fungal community structure in their microbial respiratory response to three pesticides. *Bioresour Technol* 102: 9638–45.
54. Gomez-Alvarez V, King GM, Nusslein K (2007) Comparative bacterial diversity in recent hawaiian volcanic deposits of different ages. *FEMS Microbiol Ecol* 60: 60–73.
55. Gudelj I, Weitz JS, Ferenci T, Claire Horner-Devine M, Marx CJ, et al. (2010) An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. *Ecol Lett* 13: 1073–84.
56. Herrera A, Hry M, Stach JEM, Jaffr T, Norman P, et al. (2007) Species richness and phylogenetic diversity comparisons of soil microbial communities affected by nickel-mining and revegetation efforts in New Caledonia. *Eur J Soil Biol* 43: 130–139.
57. Jagersma GC, Meulepas RJ, Heikamp-de Jong I, Gieteling J, Klimiuk A, et al. (2006) Microbial diversity and community structure of a highly active anaerobic methane-oxidizing sulfate-reducing enrichment. *Environ Microbiol* 11: 3223–32.
58. Larsen PE, Gibbons SM, Gilbert JA (2012) Modeling microbial community structure and functional diversity across time and space. *FEMS Microbiol Lett* 332: 91–8.
59. Medihala PG, Lawrence JR, Swerhone GD, Korber DR (2012) Spatial variation in microbial community structure, richness, and diversity in an alluvial aquifer. *Can J Microbiol* 58: 1135–51.
60. Oline DK (2006) Phylogenetic comparisons of bacterial communities from serpentine and nonserpentine soils. *Appl Environ Microbiol* 72: 6965–71.
61. Ovreas L (2000) Population and community and level approaches for analyzing microbial diversity in natural environments. *Ecol Lett* 3: 236–251.
62. Ovreas L, Torsvik VV (1998) Microbial diversity and community structure in two different agricultural soil communities. *Microb Ecol* 36: 303–315.
63. Schloss PD, Handelsman J (2005) Metagenomics for studying unculturable microorganisms: cutting the gordian knot. *Genome Biol* 6: 229.
64. Sirohi SK, Singh N, Dagar SS, Puniya AK (2012) Molecular tools for deciphering the microbial community structure and diversity in rumen ecosystem. *Appl Microbiol Biotechnol* 95: 1135–54.
65. Slabbert E, Kongor RY, Esler KJ, Jacobs K (2010) Microbial diversity and community structure in fynbos soil. *Mol Ecol* 19: 1031–41.
66. Vivas A, Moreno B, Garcia-Rodriguez S, Benitez E (2009) Assessing the impact of composting and vermicomposting on bacterial community size and structure, and microbial functional diversity of an olive-mill waste. *Bioresour Technol* 100: 1319–26.
67. Magurran A, McGill B (2011) *Biological Diversity: Frontiers in Measurement and Assessment*. Oxford New York, USA: Oxford University Press.
68. Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends Ecol Evol* 24: 110–7.
69. Smith MA, Fisher BL (2009) Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. *Front Zool* 6: 31.
70. Smith MA, Fisher BL, Hebert PD (2005) DNA barcoding for effective biodiversity assessment of a hyper-diverse arthropod group: the ants of Madagascar. *Philos Trans R Soc Lond B Biol Sci* 360: 1825–34.
71. Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, et al. (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445: 757–60.
72. Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PD (2007) DNA barcodes affirm that 16 species of apparently generalist parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proc Natl Acad Sci U S A* 104: 4967–72.
73. Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PD (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc Natl Acad Sci U S A* 103: 3657–62.
74. Smith MA, Rodriguez JJ, Whitfield JB, Deans AR, Janzen DH, et al. (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc Natl Acad Sci U S A* 105: 12359–64.
75. Stahlhut JK, Fernandez-Triana J, Adamowicz SJ, Buck M, Goulet H, et al. (2013) DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. *BMC Ecol* 13: 2.
76. Heppner J, Inoue H (1992) *Lepidoptera of Taiwan. Vol 1 Part 2 Checklist* [M]. USA: Scientific Publishers Inc Gainesville.
77. Kristensen NP, Scoble MJ, Karsholt O (2007) *Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity*. *Zootaxa* 1668: 699747.
78. Nye IWB (1975) *The generic Names of Moths of the World* [part 1: Noctuoidea (part) Noctuidae, Agariidae and Nolidae]. London: Unwin Brothers Ltd London.
79. Shen X, Sun H, Zhao H (2007) Chinese noctuid moths and their distribution. *Acta Entomologica Sinica* 50: 709–719.
80. Yu S (1998) A preliminary study on fauna of Noctuidae (Lepidoptera: Ditrysia) insects in China. *Entomotaxonomia* 20: 44–50.

81. Chen YX, Wang BH, Lin D (1991) The Noctuids Fauna of Xizang. Zhengzhou: Henan Scientific and Technical Publishing House.
82. Cun YZ, Wang XQ (2010) Plant recolonization in the Himalaya from the southeastern Qinghai-Tibetan plateau: Geographical isolation contributed to high population differentiation. *Mol Phylogenet Evol* 56: 972–82.
83. Hewitt G (2000) The genetic legacy of the quaternary ice ages. *Nature* 405: 907–13.
84. Liu Z, Ren B, Wu R, Zhao L, Hao Y, et al. (2009) The effect of landscape features on population genetic structure in Yunnan snub-nosed monkeys (*Rhinopithecus bieti*) implies an anthropogenic genetic discontinuity. *Mol Ecol* 18: 3831–46.
85. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853–8.
86. Matz MV, Nielsen R (2005) A likelihood ratio test for species membership based on DNA sequence data. *Philos Trans R Soc Lond B Biol Sci* 360: 1969–74.
87. Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Syst Biol* 57: 750–7.
88. Munch K, Boomsma W, Willerslev E, Nielsen R (2008) Fast phylogenetic DNA barcoding. *Philos Trans R Soc Lond B Biol Sci* 363: 3997–4002.
89. Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. *Syst Biol* 55: 162–9.
90. Ross HA, Murugan S, Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. *Syst Biol* 57: 216–30.
91. Kollar V, Redtenbacher L (1844) Aufzählung und Beschreibung der von Freiherrn Carl v. Hugel auf seiner Reise durch Kaschmir und das Himalayagebirge gesammelten Insecten. In Hugel C F von Kaschmir und das Reich der Siek 4: 393–564.
92. Boursin C (1963) Die Noctuidae-Arten aus Dr. H. C. Hne's China-Ausbeuten. V. Der Amathes-Komplex; VI Die Gattungen: *Raddea* Alph., *Estimata* Kozh., *Erebophasma* n. gen., *Spinipalpa* Alph. und *Perissandra* Warren; VII Die Gattung *Ochropleura* Hb. *ForschBer Landes Nordrhein-Westfalen* 1170: 1–107.
93. Chen Y (1999). *Fauna Sinica* Vol. 16: Insecta, Lepidoptera, Noctuidae.
94. Draudt M (1950) Beiträge zur Kenntnis der Agrotiden-Fauna Chinas. Aus den Ausbeuten Dr. H. Hone's (Beitrag zur Fauna Sinica). *Mitteilungen der Münchner Entomologischen Gesellschaft* 40.
95. Hacker H (1990–2011) Vol. 1(1990), 502 pp, pls. a-f; vol. 3(1993), 587 pp, pls. a-s; vol. 4(1996), 520 pp, pls. a-y; vol. 5(1996), 720 pp, a-y; vol. 6(1998), 843 pp, pls. a-u, *hadula* i-xiii; vol. 7(1999), 377–713, pls. xii-xxvi; vol. 8(2001), 655–700, pls. 31–33; vol. 9(2002), pp, pls. 25; vol. 10(2004), 786 pp, pls. 29; vol. 12(2006), 420 pp, pls. 30; vol. 15(2010), 127–580; vol. 16(2011), 326 pp, pls. 34. Buchreihe zur Entomologie, Schwanfeld, Esperiana.
96. Haruta T (1992–1998) *TINEA*. Vol. 13: Supplement 1(1992), 122 pp, pls. 32, 2(1993), 160 pp, pls. 64; Vol. 14: Supplement 1(1994), 171 pp, pls. 96, 2(1995), 206 pp, pls. 128; Vol. 15: Supplement 1(1998), 314 pp, pls. 160. The Japan Heterocerists' Society, Tokyo. Moths of Nepal.
97. Kononenko V (2010) Noctuidae Sibiricae, VI. 2, Micronoctuidae, Noctuidae: *Ri764* vulinae-Agaristinae. Entomological Press.
98. Kononenko V, Han HL (2007) Atlas Genitalia of Noctuidae in Korea: (Lepidoptera). Center for Insect Systematics, Korea.
99. Kononenko VS (2005) Noctuidae Sibiricae: An annotated check list of the Noctuidae (SL)(Lepidoptera, Noctuoidea: Nolidae, Erebidae, Micronoctuidae, Noctuida) of the Asian part of Russia and the Ural region, volume 1. Entomological Press.
100. Witt T (2008–2011) The Witt Catalogue, a taxonomic atlas of the Eurasian and North African Noctuoidea. Vol. 1(2008) Plusiinae I, 342 pp; Vol. 2(2009) Cuculliinae I, 365 pp; Vol. 3(2009) Apameini I, 393 pp; Vol. 4(2010) Plusiinae II, 280 pp; Vol. 5(2011) Cuculliinae II and Psaphidinae, 380 pp. Heterocera Press, Budapest. Heterocera Press.
101. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294–9.
102. Kumar S, Tamura K, Nei M (2004) Mega3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5: 150–63.
103. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
104. Pielou E (1975) Ecological diversity. Wiley, New York.
105. Muturi EJ, Shillu J, Jacob B, Gu W, Githure J, et al. (2006) Mosquito species diversity and abundance in relation to land use in a riparian agroecosystem in Mwea, Kenya. *J Vector Ecol* 31: 129–37.
106. Ru W, Zhang J, Zhang F, Zhang G, Liu R (2006) Species diversity and community structure of forest communities in Lishan Mountain. *Ying Yong Sheng Tai Xue Bao* 17: 561–6.
107. Xu Y, Chen Y, Li W, Fu A, Ma X, et al. (2011) Distribution pattern of plant species diversity in the mountainous region of Ili River Valley, Xinjiang. *Environ Monit Assess* 177: 681–94.
108. Zhang ZJ (2009) DV-Curve: a novel intuitive tool for visualizing and analyzing DNA sequences. *Bioinformatics* 25: 1112–7.
109. Liao B, Tan M, Ding K (2005) Application of 2-d graphical representation of DNA sequence. *Chem Phys Lett* 414: 296–300.
110. Whitaker RH (1965) Dominance and diversity in land plant communities: Numerical relations of species express the importance of competition in community function and evolution. *Science* 147: 250–60.
111. Wilson J (1991) Methods for fitting dominance/diversity curves. *J Veg Sci* 2: 35–46.
112. Justel A, Pea D, Zamar R (1997) A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters* 35: 251–259.
113. Hijmans R, Cameron S, Parra J, Jones P, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965–1978.
114. Jolliffe I (2002) *Principal Component Analysis*, Series: Springer Series in Statistics. Springer NY XXIX, 2nd edition, 487 pp.
115. Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A (2012) DNA barcoding for community ecology—how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS One* 7: e28832.
116. Rubinoff D (2006) DNA barcoding evolves into the familiar. *Conserv Biol* 20: 1548–9.
117. Nee S, May R, Harvey P, Scholander SG (1994) The reconstructed evolutionary process. *Philos Trans R Soc Lond B Biol Sci* 344: 305–311.
118. Hudson R (1990). Gene genealogies and the coalescent process. Oxford surveys in evolutionary biology.
119. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol* 21: 2045–2050.
120. Yu DW, Ji Y, Emerson BC, Wang X, Ye C, et al. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3: 613–623.