

## Article

# Watershed: a key for microbial biogeography

Xiao-Yan Yang<sup>1,2,3†</sup>, Wei Deng<sup>1,2,3†</sup>, Fa Zhang<sup>1,2,3†</sup>, Shuo-Ran Liu<sup>1,2,3</sup>, Davide Fornacca<sup>1,2,3</sup>, Na Li<sup>1,2,3</sup>, Yan-Peng Li<sup>1,2,3</sup>, Xi-Jie Fan<sup>1</sup>, Xin Zhang<sup>1,2,3</sup>, Li-Lei Liu<sup>1,2,3</sup>, Ya-Jiao Zhu<sup>1</sup>, Kun Tan<sup>1,2,3</sup>, Ying Gao<sup>1</sup>, Qing-Hua Cai<sup>5\*</sup> and Wen Xiao<sup>1,2,3,4\*</sup>

<sup>1</sup> Institute of Eastern-Himalaya Biodiversity Research, Dali University, Dali 671003 China; yangxy@eastern-himalaya.cn (YXY); dengweir@hotmail.com (DW); zhangf@eastern-himalaya.cn (ZF); liusr@eastern-himalaya.cn (LSR); fornacca@eastern-himalaya.cn (DF); lin@eastern-himalaya.cn (LN); liyp@eastern-himalaya.cn (LYP); 1119887599@qq.com (FXJ); zhangx@eastern-himalaya.cn (ZX); liulilei7398@dingtalk.com (LLL); 3466185191@qq.com (ZYJ); tank@eastern-himalaya.cn (TK); gaoy@eastern-himalaya.cn (GY).

<sup>2</sup> Collaborative Innovation Center for Biodiversity and Conservation in the Three Parallel Rivers Region of China, Dali 671003 China.

<sup>3</sup> The Provincial Innovation Team of Biodiversity Conservation and Utility of the Three Parallel Rivers Region, Dali University, Dali 671003 China.

<sup>4</sup> International Centre of Biodiversity and Primates Conservation, Dali 671003 China.

<sup>5</sup> State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072 China.

† These authors contributed equally to this work.

\* Correspondence: qhcai@ihb.ac.cn (CQH); xiaow@eastern-himalaya.cn (XW); Tel.: +86-0872-2219912

**Abstract:** Biogeography research is flawed by the poor understanding of microbial distributions due to the lack of a systematic research framework, especially regarding appropriate study units. By combining pure culture and molecular methods, we studied the biogeographic patterns of nematode-trapping fungi by collecting and analysing 2,250 specimens from 228 sites in Yunnan Province, China. We found typical watershed patterns at the species and genetic levels of nematode-trapping fungi. The results showed that microbial biogeography could be better understood by 1) using watersheds as research units, 2) removing the coverup of widespread species, and 3) applying good sampling efforts and strategies. We suggest that watersheds could help unify the understanding of the biogeographic patterns of animals, plants, and microbes and may also help account for the historical and contemporary factors driving species distributions.

**Keywords:** Watershed; biogeographic patterns; microbial biogeography; biodiversity; spatial distribution; research unit

## 1. Introduction

Biogeographic patterns and their underlying mechanisms are essential in biodiversity research. Understanding biogeographic patterns is crucial for predicting how global change affects biodiversity, designing effective conservation strategies, and achieving sustainable development [1]. Despite microbes being the most important biotic driver of global material and energy flows, biogeographic research on microbiota has long lagged behind that of macro-biota. Debates exist regarding whether heterogeneous biogeographic patterns exist in microbes [2, 3], as so far researchers also believe that its biogeographic pattern is weaker than macro-biota. The strong diffusion ability and extreme environmental adaptability (such as extremely hardy resting stages) of microorganisms are considered to have led to their lack of or weak biogeographic pattern, especially in regions with low heterogeneity [3-5]. It is time to carry out microbial biogeography research based on a new perspective. we put forward.

In this regard, a hypothesis we put forward is that the understanding of microbial biogeographic pattern may be helped by watershed perspective. Specifically, microorganisms mainly diffuse through wind, water currents, animal carrying, etc [6]. Watersheds may restrict these processes, thereby shaping the heterogeneity microbial biogeographic

pattern, and this diffusion restriction should be more obvious in mountain and valley areas. For example, the climate of the dry-hot valley in Yunnan, China is caused by the restriction of the distribution of water vapor due to the blocking of the airflow by the mountains [7]. Even the airflow will be blocked by mountains and unevenly distributed. There is no doubt that the microorganisms that travel with the airflow will also be restricted by this. Secondly, the blocking of mountains will create unique habitats and relatively independent spaces in the watersheds, driving the occurrence of endemic species. These endemic species are relatively rare, but they play a key role in shaping the regional heterogeneity patterns. Third, the widespread species does not have biogeographic pattern at the species level or even conducive to the discovery of the pattern, but there should be a watershed pattern at the genetic level.

Consequently, we conducted this research in Yunnan Province, China, which holds six major international rivers belonging to the Pacific and Indian Ocean systems. The watersheds were largely shaped by the collision of the Indian Ocean and Eurasian plates with the formation of the Qinghai-Tibet Plateau. The spatial heterogeneity in Yunnan is very high, given that mountainous areas occupy 88.6% of the province. For the research object, Chase and Martiny (2018) believe that the study of small taxa is conducive to the discovery of microbial patterns [8]. Therefore this study takes Orbiliaceae nematode-trapping fungi (NTF) as research object which is a cultivable small taxa with 104 known species. Using the stratified sampling method, we collected 2,250 specimens from 228 sites in Yunnan Province (Figure S1, S2), combined pure culture and molecular methods to investigate the NTF diversity both at species and genetic level. By doing so, we aim to verify each of our hypotheses.

## 2. Materials and Methods

### 2.1 Study area

Yunnan Province is located between 21° 8' 32" and 29° 15' 8" N and between 97° 31' 39" and 106° 11' 47" E, with a total area of 394,100 km<sup>2</sup>, and spans elevations of 76-6740 m. This area is characterized by a sharp elevational gradient and six large rivers - the Mekong, Irrawaddy, Salween, Red, Pearl and Yangtze Rivers - cutting across the province. This complex topography helped shape the region's relatively closed ecological zones, among which biodiversity is extremely high.

### 2.2 Research objects

Nematode-trapping fungi (NTF) consist of eukaryotic microorganisms with terrestrial and aquatic biotypes belonging to the Orbiliaceae family. NTF are predatory microbes and, as their name suggests, prey on nematodes and other small organisms by producing trapping structures with their vegetative hyphae. Globally distributed across various habitats and extreme environments, 104 NTF species have thus far been reported, including in the *Arthrobotrys*, *Drechslerella* and *Dactylella* genera. All NTF species can be pure cultured, and their morphological characteristics are observable under low-magnification microscopes. NTF can be studied qualitatively and semi-quantitatively, making them ideal taxa for studying patterns in the spatial distribution of microbial diversity.

### 2.3 Stratified sampling

We systematically established 282 stratified sampling sites according to Yunnan's hydrographic network (Figure S1). We were unable to sample from sites located in inaccessible terrain, including in deep valleys and high mountains (Figure S2). Sites with strong human disturbances were also avoided. Finally, 228 randomly distributed sites (Ripley's K,  $p = 0.2435$ ) remained (Figure S1a). To ensure the phenological consistency and clear aquatic-terrestrial boundaries of the sites, sampling was conducted from the southernmost to the northernmost region and between February 27, 2014, and June 8, 2014. At each site, 5 sampling points were set parallel to the river in both the aquatic and terrestrial habitats at a separation distance of 10 metres (Figure S1b). In total, 2,250 sampling points

were set (Figure S1, panel a), and 30 sampling points were lost due to the lack of soil at the sampling location. Using the five-point sampling method in 1-m×1-m squares (Figure S1b), samples consisted of ~50 g of mixed soil (or sediment) and were collected at a depth of 0~10 cm below the surface at each sampling point. In total, we collected 2,250 specimens from 228 sites. The samples were then stored in a disposable self-sealing bag at room temperature. We recorded habitat information, including the location, longitude, latitude, and elevation of each sampling site. The samples were sent to the laboratory within 1 week of collection, immediately followed by NTF identification processing.

#### 2.4 NTF classification and identification

**Isolation and purification:** Using the five-point spreading method, the samples were plated on petri dishes with Corn agar medium (CMA) with ~5,000 individuals of *Panagrellus redivivus* to induce NTF germination. We plated three replicates per sample for a total of 6,750 petri dishes plated. The petri dishes were incubated at room temperature (25-28°C). We examined the petri dishes once after 3 weeks and again after 4-5 weeks using a stereomicroscope to isolate all the species. During each examination, we used single-spore isolation to collect and purify any germinated NTF on CMA medium. A sterile toothpick was used during collection. Pure cultures were preserved until we were able to isolate and purify cultures from all specimens.

**Morphological identification:** All preserved pure cultures were revived in CMA medium and placed on a temporary slide using the insertion and sticking method. The morphological characteristics of NTF, including conidia, conidiophores and chlamydospores, were then photographed using a microscope camera (Olympus BX51, Japan). For each NTF strain, the type of trapping device was observed and confirmed using the observation chamber and nematode induction method. The strains were identified according to Li et al., 2014 [9].

**Molecular identification:** Each NTF strain was inoculated on PDA medium to gather the mycelium. We then extracted DNA from the mycelium using a CTAB protocol. Fragments of ITS (internal transcribed spacer), TUB ( $\beta$ -tubulin gene), TEF (translation elongation factor 1-alpha) and RPB2 (RNA polymerase II) were amplified by PCR and sequenced by BioSune Biotechnology Co., Ltd. (Shanghai, China). The homologous ITS and TUB sequences were aligned using the BLAST toolkit from NCBI (National Center for Biotechnology Information Search database).

We were able to identify and classify specific NTF species using morphological and molecular identifications, referring to the taxonomic system used by Li et al., 2014 [9].

#### 2.5 Data analysis

##### 2.5.1 Analysis of NTF species composition and distribution

The occurrence frequency (OF) of each species and sampling site was calculated. The OF (%) was calculated as the number of specimens with NTF occurrence divided by the total number of specimens  $\times 100\%$ . Species with OF  $> 5\%$  were defined as widespread species, while species with OF  $< 1\%$  were defined as rare species.

Using the latitude and longitude of each site and the NTF OFs, we then plotted distribution maps based on the genus, species, and rare species with the raster, sf and ggplot2 packages in R version 4.02.

The mean variance ratio was used to evaluate the NTF species distributions. A mean variance ratio of 1 indicated a random distribution (conforming to Poisson's distribution); a mean variance ratio  $> 1$  indicated a clustered distribution; and a mean variance ratio  $< 1$  indicated an even distribution. The significances of these distributions were verified using a T-test.

We divided the sampling areas into 9 grids using QGIS version 3.10. We selected the six grids with the largest sampling areas (grids A, B, C, D, E and F) for analysis (Figure S3a). The watershed boundaries of Yunnan Province were extracted from a derived digital

elevation model (CGIAR-CSI SRTM v4.1) at a 90-metre resolution using GIS geoprocessing algorithms (Figure S3b). We constructed UpSet plots using grid and watershed units with the UpSetR package in R version 4.02.

#### 2.5.2 Genetic divergence and biogeographical distribution analysis of *Arthrobotrys oligospora*

DNA sequencing of ITS, TUB, TEF and RPB2 fragments was conducted on strains of *Arthrobotrys oligospora*, which were found at every sampling site. The four sequences from each strain were summarized in a text file and converted into a fasta file. MAFFT version 7 was used to generate the multi-sequence matrix (homologous sequence searching), and BioEdit was used to manually improve the accuracy of the alignment. We used jmodeltest software to select the optimal calculating alternative model for the piecing sequence. The phylogenetic tree was constructed following the maximum likelihood (ML) method via partition calculation using IQ-Tree version 1.6.5 software. We used FigTree version 1.3.1, Microsoft Word and Photoshop to read the phylogenetic tree.

#### 2.5.3 Biogeographical distribution of NTF in Yunnan

A first prediction was performed using the different watersheds as unique predictors of the clade distribution in the phylogenetic tree of *Arthrobotrys oligospora*. We assigned clades to watersheds using the majority rule and assessed the classification accuracy using error matrix metrics.

From the original dataset, we generated 45 different splits between the training and test samples with training ratios between 30% and 70% (step=5). Each ratio step included five different mixes of training/test samples. Voronoi (Thiessen) polygons were computed around each training set and assigned the clade corresponding to their originating point. The polygons were used to predict the test dataset clades.

#### 1.5.4 Multivariate machine learning model

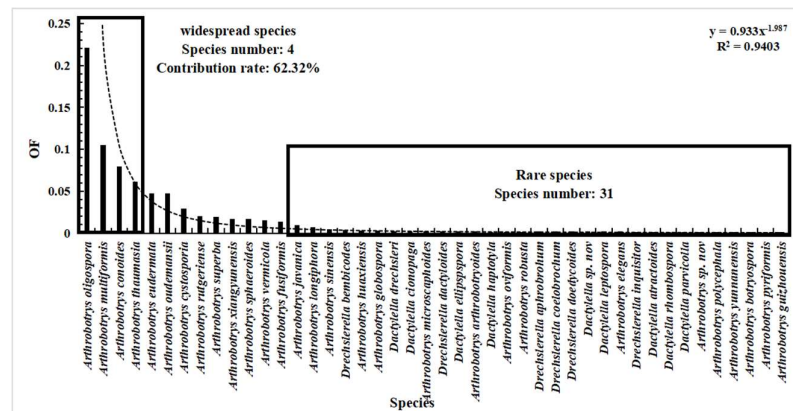
We compiled a dataset of 97 variables divided based on bioclimatic (24), topographic (6), vegetation (12), and soil properties (55) using QGIS (Table S1). In addition to the watershed, 19 predictors were selected among an initial pool of 97 variables. Table S2 shows the selection results after multinomial logistic regression and collinearity screening. In these cases, we always retained the values for the first layer.

To select the most suitable predictor variables to include in our machine learning model, we hierarchically screened each variable in our dataset. We iteratively performed multinomial logistic regressions to assess the effects of each predictor variable on clade. We assessed the relationships among predictor variables using correlation matrices (Pearson's and Kendall's) and assessed multicollinearity using the variance inflation factor.

The selected predictor variables and watershed predictors constituted the final dataset to investigate patterns in clade distribution using an extra-trees classifier (extremely randomized trees). We aimed to improve the prediction ability from the previous analysis (using the watershed subdivision as a unique predictor) and therefore used the same accuracy metrics as those used in that analysis. The hyperparameters were optimized with grid search and cross validation techniques, although we also considered bootstrapping and accuracy estimations on the out-of-bag samples. The relative feature importance was calculated for the best model.

### 3. Results

We identified 44 Orbiliaceae NTF species from 2,250 samples. These species belonged to three genera: *Arthrobotrys* (29 species), *Dactylella* (9 species), and *Drechslerella* (6 species) (Table S3). The ranked order of the occurrence frequency (OF) of NTF species corresponded to a logarithmic distribution model ( $R^2 = 0.940$ ,  $p < 0.05$ ) (Figure 1). Four species were widely distributed (mean OF = 5.22%) and contributed 9.09% of the total species richness and 62.32% of the total OF. Conversely, 31 species were classified as rare (mean OF = 0.18%) and contributed 70.45% of the total species richness and 7.52% of the total OF (Figure 1).

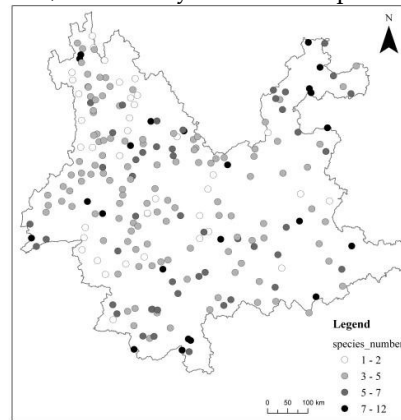


**Figure 1.** Species composition and occurrence frequency ordering of the nematode-trapping fungi collected in this study. Regression lines are calculated by applying logarithmic distribution model.

### 3.1. Widespread species shaped the random distribution pattern of NTF

At the genus level, *Arthrobotrys* was regionally distributed and found in 98.68% of the sampling sites, while *Dactylella* and *Drechslerella* were locally distributed and found in 7.02% and 7.89% of the sampling sites, respectively (Figure S4).

A mean variance ratio of 1.07 ( $t = 0.7029$ ,  $p > 0.05$ ) indicated that the NTF species were randomly distributed (Figure 2). Four widespread species were found in 83.30%, 56.58%, 47.37% and 41.23% of the sampling sites (Figure S5). When the widespread species were removed, the cumulative distribution of the OFs of rare and non-widely distributed species was non-random and clustered (rare species: mean variance ratio = 1.284,  $t = 3.022$ ,  $p = 0.003$ ; non-widely distributed species: mean variance ratio = 1.326,  $t = 3.471$ ,  $p < 0.001$ ).



**Figure 2.** Biogeographic patterns of the species numbers of nematode-trapping fungi. A mean variance ratio of 1.07 ( $t = 0.7029$ ,  $p > 0.05$ ) indicated that the NTF species were randomly distributed.

### 3.2. The watershed patterns of the widespread species (for genetic level) and rare species

At the genetic level, the phylogenetic tree revealed that the widespread species *A. oligospora* was divided into 5 large clades (Figure 3a). The distribution of these clades was consistent with the natural watershed divisions in Yunnan Province, except for the Irrawaddy River, which had no corresponding clade (Figure 3b). The strains within the other clades were distributed within their corresponding watersheds. A total of 64.70%, 60.71%, 80.00%, 63.89% and 61.11% of the strains from clade 1 to clade 5 were correspondingly distributed in the Yangtze River, Red River, Pearl River, Mekong watershed, and Salween-Irrawaddy watershed, respectively (Figure 3b).

Among the 44 species detected in this study, 31 rare species were found in fewer than 10% of the sampling sites (range: 0.41-9.21%) (Figure S5). Of the 19 rare species with non-

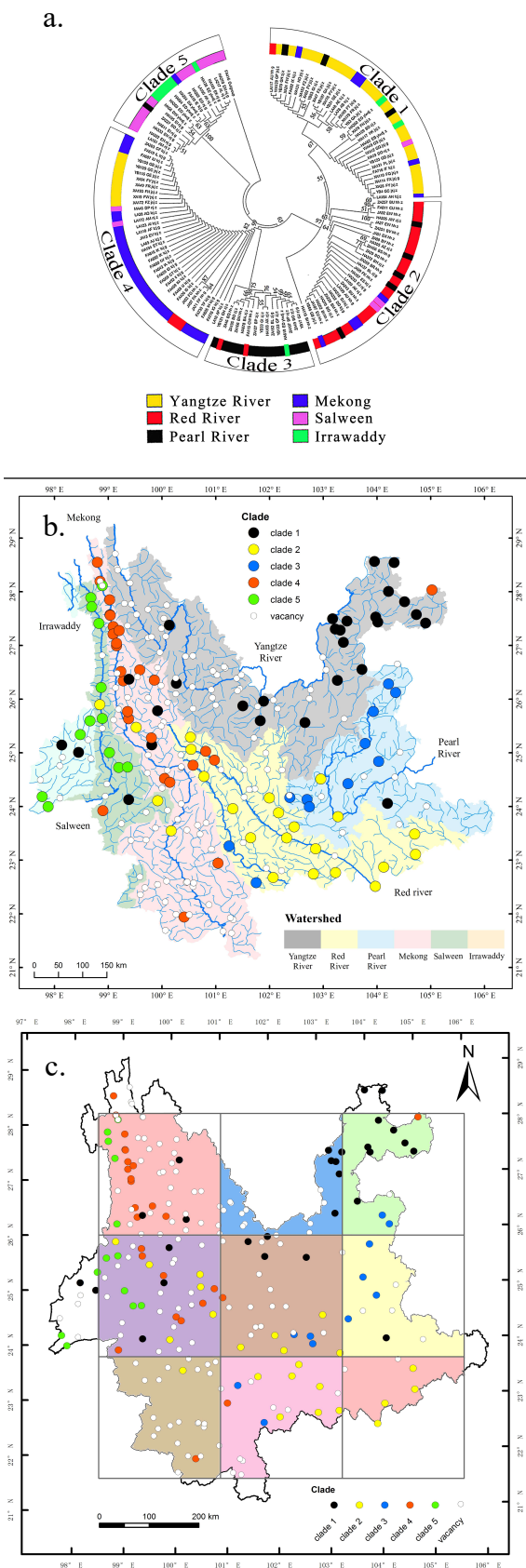


single-site distributions within grids or watersheds, 63.16% were distributed in the same or an adjacent grid and 84.21% were distributed in the same or adjacent watershed (Figure S6). With the exception of *Drechslerella dactyloides*, 8 of the 9 rare species that appeared at only 2 or 3 sites were distributed in the same or adjacent watershed (Figure S6).

### 3.2. Watershed is a key for discover NTF biogeographic pattern

When a grid was used as the unit of analysis, no clear distribution pattern of the five clades in the *A. oligospora* phylogenetic tree was detected (Figure 3c). Based on our phylogenetic tree, the spatial distribution of *A. oligospora* was machine learned using randomly generated Voronoi polygons and watersheds. We found that the 45 maps generated using polygons all had lower accuracy than the maps watersheds generated (Table S4). On average, the accuracy of the polygons was low (mean: 36%, median: 38%), with all but one prediction falling below 50%. By assigning the clades according to the watersheds, 68.8% of clades were classified correctly (Figure S7 A). The watershed unit explained nearly 70% of the *A. oligospora* distribution in Yunnan. None of the additional climate, topographic, soil or vegetation variables significantly improved the model (Figure S8, Table S5).

We were better able to capture the distribution pattern of NTF when using watersheds as the unit of analysis compared with grids. Only 17 intersections were found when we constructed UpSet plots using watershed units (Figure 4). We found 19 species distributed in only one watershed unit, and 76.2% of all species were distributed in the adjacent watersheds. There was a large gap between the datasets (the variance was 7.89, and the mean value was 2.53). Despite finding 22 combinations when constructing the integrative diagrams by grid, only 11 species were distributed in only one grid, and 55.0% of all species were distributed in adjacent grids. There was a small fluctuation between the datasets (the variance was 5.04, and the mean value was 1.77). Ultimately, using watershed units led to higher observations of endemic species, more species distributed within adjacent or similar units, and a more varied species composition compared with those obtained using grid units.



**Figure 3.** Biogeographic patterns of the widespread species *A. oligospora* based on a) genetic clades in with the perspectives of b) watershed units and c) grid units.

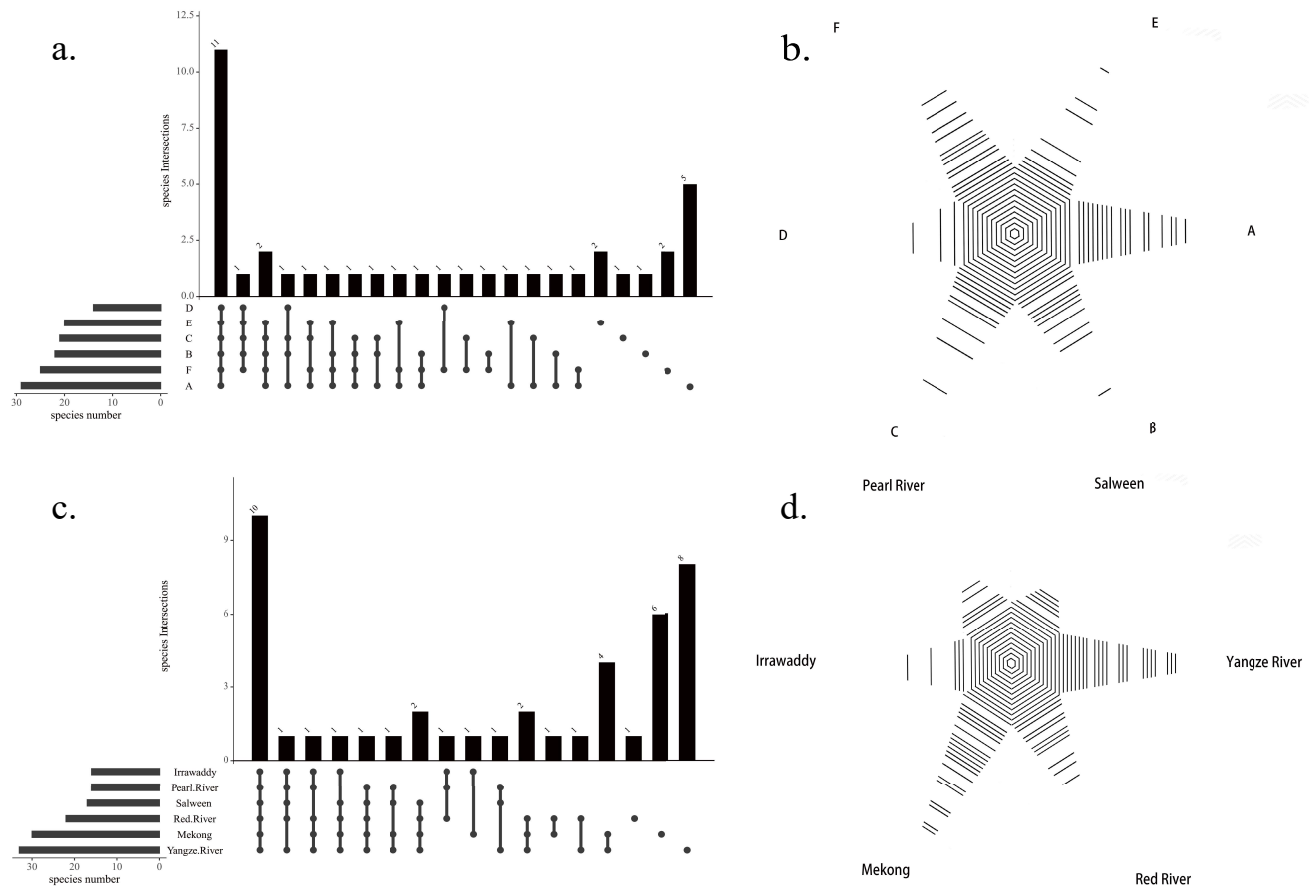


Figure 4. UpSet plot of species from the perspectives of a) grid units and b) watershed units. The distribution pattern was uniform when using grids (c) and heterogeneous when using watersheds (d).

#### 4. Discussion

Nearly half of the reported NTF species ( $N = 44$ ) in three genera were found during this study. At the species level, the distribution pattern was random and had no heterogeneous pattern. However, at the level of widespread species genetic, the level of rare species and genera, NTF has a pattern of local distribution. The heterogeneous distribution of microbial diversity could be better explained by using watersheds as units. Therefore, our three hypotheses have been verified, 1) rare species are indeed important to the microbial biogeographic pattern; 2) widespread species will cover the pattern at the species level, and its geographic pattern needs to be revealed at the genetic level; 3) watersheds are the key to discovering the microbial biogeographic pattern. In the process of gradually removing obstacles to discover the microbial biogeographic pattern, the following points need to be paid attention to.

First, widespread species are widely distributed, covering patterns at the species level, and their biogeographic patterns need to be studied at the genetic level. In this study, the number of widespread species only accounted for less than 5% of the total number of species but contributed 70% of the OF. When excluding the 4 widespread species, the pattern tended to be non-random with a clear heterogeneous pattern. If the distribution of the species with high OFs was gradually superimposed, the patterns of NTF tended to become random. The results indicated that the extremely high OFs and wide distribution of widespread species masked the heterogeneous microbial biogeographic patterns. This may be the major factor in the past that microbial research has been difficult to find or can only find a weak biogeographic pattern. At the genetic level, the 5 clades of *A. oligospora* were consistent with the natural watershed divisions, and these divisions were further supported by machine learning results. Totals of 64.70%, 60.71%, 80.00%, 63.89% and 61.11%



of the strains from clade 1 to clade 5 were correspondingly distributed in the Yangtze River, Red River, Pearl River, Mekong watershed, and Salween-Irrawaddy watershed, respectively. Therefore, in future studies, removing the effects of widespread species at the species level, and paying attention to the lineage differentiation at the genetic level will help microbial biogeography research.

Second, although rare species and genera have low abundance, they play a key role in shaping the microbial diversity pattern [10]. It is necessary to be more cautious to conduct research on the biogeographical pattern of microbes based on operational taxonomic units (OTU) alone. At the level of rare species, the 31 rare species were restricted to 0.44% ~ 9.21% of the sampling sites. The genera *Dactylella* and *Drechslerella* were restricted to 7.02% and 7.89% of the sampling sites, respectively. Because these limited distribution and the existence of endemic species were representative of the biogeographic pattern in itself, we concluded that NTF species have heterogeneous biogeographic patterns [2]. Many of the microbial biogeographic pattern studies are OTU based which may ignore rare species and the genetic variation under the species due to the broader resolution of OTUs (usually 97% similarity clustering), and thus hinders the detecting of microbial heterogeneous patterns emerged during the evolution [8, 11, 12]. Therefore, we support the view of Chase and Martiny (2018) and believe that conducting more refined research on a specific taxa and paying more attention to the pattern of the genetic variation will promote the understanding of the microbial biogeographic [8]. The pure culture based study is still indispensable for current microbial biogeography research. The results of the experiments carried out in this study confirm the feasibility of this idea. In the future, we need to combine culture-dependent and culture-independent approaches to carry out more systematic research on microbial biogeography.

Third, microbial biogeography research requires more scientific sampling design, including sample area selection, sampling effort and sampling strategy. Insufficient sampling effort is the main cause of our insufficient understanding of the microbial biogeographic pattern [3]. In accordance with the logic of stratified sampling, we arranged 228 sites in Yunnan Province and collected 2,250 samples intensively, and discovered the NTF biogeographic pattern. Had the sampling effort been reduced by 10% (do not set 10 samples at each site), we estimated that 24 species (77.42% of the rare species) would have been more difficult to observe. If 23 sites were set up only (10% of the current sample site), we estimated that 20 species (64.52% of the rare species) would have been more difficult to observe. What's bad is that most of the current researches on microbial diversity are not high in sampling intensity. The sampling intensity interval between 3~300 samples, with an average of 40 samples collected (summarised according to articles published in the ISME Journal and Microbiome over the past five years). Sampling efforts were subsequently reduced to the same time and resources when using metagenomic methods [13]. With low sampling efforts, many rare species would be ignored. In addition, the selection of research area is also very critical. In this study, a strong biogeographic pattern of NTF was found in Yunnan.

It is precisely because of the rich types of mountains and the strong environmental heterogeneity in Yunnan that we discovered the strong biogeographical pattern of NTF here. It is difficult to find a strong microbial biogeographical pattern if the research is carried out in areas with relatively flat geographical terrain and not strong geographical barriers.

Finally, the watershed unit is the key to understanding the microbial biogeographic pattern. On the one hand, our research results show that a clear watershed pattern of NTF. On the other hand, the heterogeneous distribution of microbial diversity could be better explained by using watersheds as units compared with grids. At the genetic level, the 5 clades of *A. oligospora* were consistent with the natural watershed divisions. That is to say, even if it can be widely distributed, the genetic diversity of *A. oligospora* is differentiated within the watershed. At the species level, more endemic species were observed using watershed units (19 species by using watersheds vs 11 species by using grids). Additionally, 84.21% of rare species were clustered in the same or adjacent watersheds, while

only 63.16% were clustered in the same or adjacent grids. It can be assumed that the spatial compression of the uplift of the Qinghai-Tibet Plateau and the Hengduan Mountains has formed the six major watersheds of Yunnan. Because environments in mountainous regions are geographically isolated by mountains, rivers, and other natural barriers, many new species have evolved in these environments [14]. The results of machine learning found that watershed is the most important factor for this pattern, while other environmental factors have little effect. Therefore, the biogeographical pattern of NTF found in this study is mainly shaped by the geographical isolation caused by the historical changes of the six major river watersheds in Yunnan, and contemporary factors have little influence. Considering that there are natural cascades in the watershed itself, we believe that the future watershed units can also help form a breakthrough in the impact of contemporary and historical factors on the microbial diversity. The proportion of contemporary and historical factors in the watershed units of different scales should be gradual. Historical factors represented by geographical barriers between large watersheds are the main factors shaping the biogeographic patterns. As the scale of the watersheds shrinks, the geographical barriers between sub-small watersheds weaken, and the role of shaping the contemporary environment within the watershed will become more important. In the past, research in this area was often carried out on different spatial, temporal and phylogenetic scales [15]. Even worse, biogeographic patterns are particularly sensitive to the above factors [16]. These confusing results make it impossible for researchers to distinguish variations in microbial biodiversity that are caused by contemporary or historical factors [17]. Natural watershed units can provide a natural scale for the study of the scale effect of the above-mentioned factors driving diversity patterns.

Natural units with clear boundaries should be a fundamental need in studies of biodiversity because the coupling of animal, plant and microbial biodiversity distributions is inevitable in future biogeographic research [18]. While biogeographic units have long been defined for animals and plants (e.g., ecoregions, biota, and biogeographic provinces) [19, 20], these units do not correspond well with each other; great differences exist among different regions and taxa [21]. We propose that natural watersheds are ideal units for biodiversity research because they are relatively independent or “closed” units of material and energy flow, and their boundaries can be natural barriers to transmission [22]. As a result, the watershed provides a unit selection for biogeography research that is suitable for multiple groups.

## 5. Conclusions

Future microbial biogeography could be better understood by 1) using watersheds as research units, 2) removing the coverup of widespread species, and 3) applying systematic sampling design. When conducting research, attention should be paid to selecting regions with high heterogeneity, increasing sampling efforts, and introducing stratified sampling strategies. The research methods aimed at small taxa and pure cultivation are a powerful supplement to the current research with metagenomic technology. Research on a single widespread species can focus on its biogeographic distribution at the level of genetic diversity. The watershed unit is helpful to analyze the impact of historical and contemporary factors on biodiversity, and it also provides a natural unit for studying the coupling relationship between microbes and the biogeographic pattern of animal and plant diversity, which is the key to the study of biogeography.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: a. Map of sampling sites. b. Sampling points at each site. Figure S2: Geographic environments of the sampling sites. Figure S3: Divisions of a) grids and b) watershed units. Figure S4: Biogeographic patterns of each genus of Nematode-Trapping Fungi. Figure S5: Biogeographic patterns of each genus of Nematode-Trapping Fungi. Figure S6: Biogeographic patterns of rare species in the perspective of a) grids and b) watersheds. Figure S7: Watershed map (A) and three examples of maps generated using Voronoi polygons from different training sets. Figure S8: Impurity-based feature importance from the Extra-Trees classification results. Table S1: List of environmental factors. Table S2: Variables selection stages and final data set. Table S3: List of detected species of Nematode-Trapping

Fungi. Table S4: The 45 maps generated using polygons all had lower accuracy than the maps watersheds generated. Table S5: Permutation feature importance from the Extra-Trees classification results.

**Author Contributions:** Y.X.Y, X.W and C.Q.H conceived the study; Y.X.Y, X.W, D.W, Z.F, L.S.R and D.F undertook comparative and analytical work and wrote the paper; L.N, T.K and G.Y performed data analyses; Y.X.Y, X.W, D.W, Z.F, L.S.R, L.Y.P, F.X.J, Z.X, L.L.L, Z.Y.J coordinated the laboratory research work; Y.X.Y, X.W, D.W, Z.F, L.S.R, D.F, C.Q.H edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant number U1602262, 31360013, 31760126 and Second Tibetan Plateau Scientific Expedition and Research Program (STEP, SQ2019QZKK2002).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The bioclimatic factors were obtained from WorldClim V2 (<http://worldclim.org/>) and EarthEnv (<https://www.earthenv.org/>). The topographic factors and vegetation factors were obtained from EarthEnv (<https://www.earthenv.org/>). The soil properties factors were obtained from Soil Database of China (<http://globalchange.bnu.edu.cn>). The remaining data are available in the Supplementary Information.

**Acknowledgments:** We thank Chen Jin-Hui (陈锦辉) from Kunming Institute of Zoology, Chinese Academy of Sciences for sample collection. Thanks to Jacob B Kraus, Ren Guo-Peng (任国鹏), She Rong (余容), Zhang Cai-Cai (张彩彩) and Zhu Xiao-Ming (朱晓明) from Institute of Eastern-Himalaya Biodiversity research, Dali University for discussions. Thanks to the followings for their contributions in laboratory research, they are Fang Yi-Hao (房以好), Zhou Han-Yu (周汉宇), Wang Jia-Liang (王家亮), Wang Chuan-Qi (王传琪), Zhou Xin-Juan (周新娟), Wang Hai-Qing (王海清), Ma Rui (马瑞), Li Fei-Teng (李飞腾), Li Jing-Chao (李静超), Jiang Lin (姜林), Gu Rong (顾荣), Wen Hui (温慧), Ni Jun (倪俊), Liao Yong-Hong (廖永红), Liu Sheng-Bin (刘胜斌), Duan Ji-Xia (段继霞), Liu Na (刘娜), Sun Hui-Fang (孙慧芳), Fan Xian-Qin (范先芹), Chen Xiao-Xi (陈晓溪), Wu Tao (吴涛), Mao Xu-Dan (毛旭丹), Wang Ting-Ping (王廷平), Lin Yan-Yun (林燕云), Huang Yuan-Tong (黄圆通), Xie Rong (谢蓉), Wang Yan-Fang (王彦芳), Xia Feng-Yu (夏丰雨), Guo Hao-Miao (郭浩妙), Zhang Ying (张莹), Li Yi-Lun (李宜伦), Yuan Jia-Xing (袁佳兴), Wang Shi-Xiong (王世雄), Zhou Fang-Yuan (周方圆), Xiong Meng-Jia (熊梦佳), Liu Tao (刘韬), Chen Zhi-Hong (陈志鸿), Lu Ya-Xian (陆亚娴), Gao Yu-Ting (高玉婷), Li Zheng-Qiang (李政强), Feng Xing-Xing (冯鑫鑫), Chen Yu-Xiu (陈玉秀), Zhang Xiao-Yu (张晓禹), Huang Zhi-Hao (黄志豪), Fan Ji-Zhuang (范集壮), Huang Fei-Hong (黄飞鸿), Ni Shui-Qing (倪水清), Zhu Zi-Zhong (朱自忠), Liao Guang-Hui (廖光辉), Wang Zhu-Hua (王柱华), Feng Bing (冯冰), Xu Xiao-Yu (徐潇雨), Yan Li-Hong (严丽红), Yue Xiao-Lan (岳晓兰), Teng Man (滕曼), Li Shuang-Mei (李双美), Zhao Jia-Jing (赵家静), Zhang Da-Bo (张大波), Li Zheng-Xiong (李正雄), Zhang Xun-Gang (张勋刚), Zhang Zan (张赞), Huo Bang-Guo (伙帮国), Yao Lu-Jun (姚路军), An Pan (安攀), Yang Zheng-Chao (杨正超), Yang Yao-Quan (杨耀权), Zhou Fa-Ping (周发平), Niu Kai-Yang (牛开阳), Xiong He-Ju (熊和菊), Lin Hui-Biao (林慧彪), Chen Xiao-Jun (陈晓俊), Zhou Jian-Fen (周建芬), Wang Ya-Juan (王娅娟), Qi Yu-Kun (齐宇坤), Huang Shuai-Yi (黄帅义), and Wang Ai-Di (王爱迪).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. K. J. Gaston, Global patterns in biodiversity. *Nature*. 405, 220-7 (2000). <https://doi.org/10.1038/35012228>
2. J. B. Martiny, B. J. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreås, A. L. Reysenbach, V. H. Smith, J. T. Staley, Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol*. 4, 102–112 (2006). <https://doi.org/10.1038/nrmicro1341>
3. K. M. Meyer, H. Memiaghe, L. Korte, D. Kenfack, A. Alonso, B. Bohannan, Why do microbes exhibit weak biogeographic patterns. *ISME J*. 12, 1404–1413 (2018). <https://doi.org/10.1038/s41396-018-0103-3>
4. J. I. Prosser, B. J. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, J. P. Young, The role of ecological theory in microbial ecology. *Nature reviews. Microbiology*, 5(5), 384–392 (2007). <https://doi.org/10.1038/nrmicro1643>
5. S. Liu, H. Wang, P. Tian, X. Yao, H. Sun, Q. Wang, M. Delgado-Baquerizo, Decoupled diversity patterns in bacteria and fungi across continental forest ecosystems. *Soil Biol. Biochem*. 144, 107763 (2020). <https://doi.org/10.1016/j.soilbio.2020.107763>
6. Y. S. Joong, Z. Ge, C. R. Buie, Bioaerosol generation by raindrops on soil. *Nat. Commun*. 8, 14668 (2017). <https://doi.org/10.1038/ncomm.s14668>

7. M. Huan-cheng, J. A. McConchie, The dry-hot valleys and forestation in southwest China. *J For Res (Harbin)*. 12, 35–39 (2001). <https://doi.org/10.1007/BF02856797>
8. A. B. Chase, J. B. H Martiny, The importance of resolving biogeographic patterns of microbial microdiversity. *Microbiology Australia*. 39, 10.1071/MA18003 (2018). <https://doi.org/10.1071/MA18003>
9. J. Li, K. D. Hyde, K. Q. Zhang, "[Methodology for Studying Nematophagous Fungi]" in *Nematode-Trapping Fungi* (Springer, Netherlands, 2014).
10. M. D. J. Lynch, J. D. Neufeld, Ecology and exploration of the rare biosphere. *Nat Rev Microbiol*. 13, 217–229 (2015).
11. D. Storch, A. L. Izling, The Concept of Taxon Invariance in Ecology: Do Diversity Patterns Vary with Changes in Taxonomic Resolution? *Folia Geobotanica*, 43, 329–344 (2008). <https://doi.org/10.1007/s12224-008-9015-8>
12. A. Gobet, S. I. Böer, S. M. Huse, J. E. van Beusekom, C. Quince, M. L. Sogin, A. Boetius, A. Ramette, Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J*. 6, 542–553 (2012). <https://doi.org/10.1038/ismej.2011.132>
13. B. Karimi, S. Terrat, S. Dequiedt, N. Saby, W. Horrigue, M. Lelièvre, V. Nowak, C. Jolivet, D. Arrouays, P. Wincker, C. Cruaud, A. Bispo, P. A. Maron, N. Bouré, L. Ranjard, Biogeography of soil bacteria and archaea across France. *Sci Adv*. 4, eaat1808 (2018). <https://doi.org/10.1126/sciadv.aat1808>
14. W. N. Ding, R. H. Ree, R. A. Spicer, Y. W. Xing, Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*. 369, 578–581 (2020). <https://doi.org/10.1126/science.abb4484>
15. J. Ladau, E. A. Elie-Fadrosh, Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol*. 27, 662–669 (2019). <https://doi.org/10.1016/j.tim.2019.03.003>
16. S. K Bay, M. A. McGeoch, O. Gillor, N. Wieler, D. J. Palmer, D. J. Baker, S. L. Chown, C. Greening, Soil Bacterial Communities Exhibit Strong Biogeographic Patterns at Fine Taxonomic Resolution. *mSystems*. 5, e00540-20 (2020). <https://doi.org/10.1128/mSystems.00540-20>
17. S. C. Castle, D. A. Samac, M. J. Sadowsky, C. J. Rosen, J. Gutknecht, L. L. Kinkel, Impacts of Sampling Design on Estimates of Microbial Community Diversity and Composition in Agricultural Soils. *Microb Ecol*. 78, 753–763 (2019). <https://doi.org/10.1007/s00248-019-01318-6>
18. M. K. Peters, A. Hemp, T. Appelhans, C. Behler, A. Classen, F. Detsch, A. Ensslin, S. W. Ferger, S. B. Frederiksen, F. Gebert, M. Haas, M. Helbig-Bonitz, C. Hemp, W. J. Kindeketa, E. Mwangomo, C. Ngereza, I. Otte, J. Röder, G. Rutten, D. S. Costa, J. Tardanico, G. Zancolli, J. Deckert, C. D Eardley, R. S. Peters, Mark-Oliver Rödel, M. Schleuning, A. Ssymank, V. Kakengi, J. Zhang, K. Böhning-Gaese, R Brandl, E. K. V. Kalko, M. Kleyer, T. Nauss, M. Tschapka, M. Fischer, I. Steffan-Dewenter, Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nat Commun*. 7, 13736 (2016). <https://doi.org/10.1038/ncomms13736>
19. B. Hausdorf, Units in Biogeography. *Systematic biology*, 4, 648–652 (2002). <https://doi.org/10.1080/10635150290102320>
20. J. R. Smith, A. D. Letten, P. J. Ke, C. B. Anderson, J. N. Hendershot, M. K. Dhami, G. A. Dlott, T. N. Grainger, M. E. Howard, B. Morrison, D. Routh, P. A. San Juan, H. A. Mooney, E. A. Mordecai, T. W. Crowther, G. C. Daily, A global test of ecoregions. *Nat Ecol Evol*. 2, 1889–1896 (2018). <https://doi.org/10.1038/s41559-018-0709-x>
21. J. R. Smith, J. N. Hendershot, N. Nova, G. C. Daily, The biogeography of ecoregions: Descriptive power across regions and taxa. *J Biogeogr*. 47, 1413–1426 (2020). <https://doi.org/10.1111/jbi.13871>
22. O. Timur, "[Use of Watersheds Boundaries in the Landscape Planning]" in *Advances in Landscape Architecture* (IntechOpen, London, 2013).