Supplementary Materials for

Watershed: a key for microbial biogeography

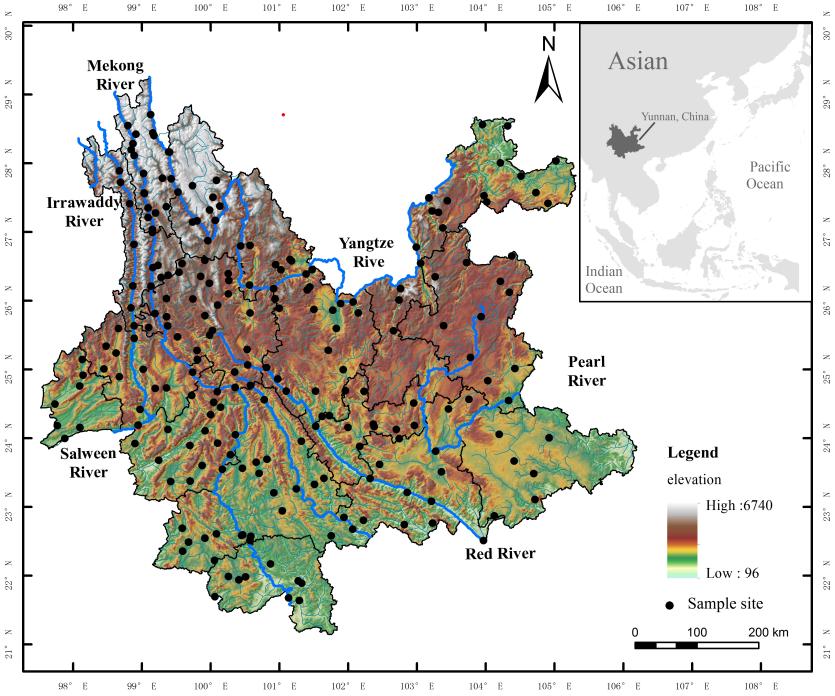
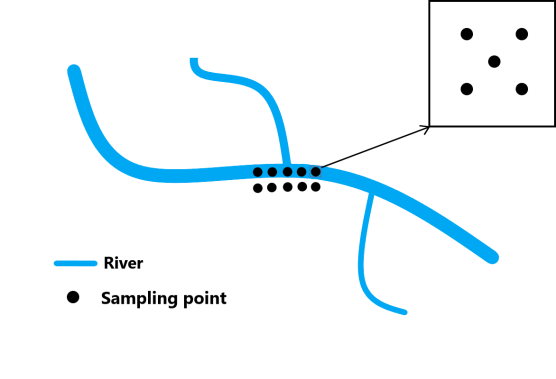
Yang Xiao-Yan(杨晓燕), Deng Wei(邓巍), Zhang Fa(张发), Liu Shuo-Ran(刘硕然), Davide Fornacca, Li Na(李娜), Li Yan-Peng(李延鹏), Fan Xi-Jie(范喜杰), Zhang Xin(张欣), Liu Li-Lei(刘李蕾), Zhu Ya-Jiao(朱娅佼), Tan Kun(谭坤), Gao Ying(高颖), Cai Qing-Hua(蔡庆华)\* , Xiao Wen (肖文)\*

Correspondence to: qhcai@ihb.ac.cn (CQH); xiaow@eastern-himalaya.cn (XW)

**This file includes:**

Figs. S1 to S8

Tables S1 to S4



**Figure S1. a. Map of sampling sites.**

**b. Sampling points at each site.**

b

a

Figure S1.

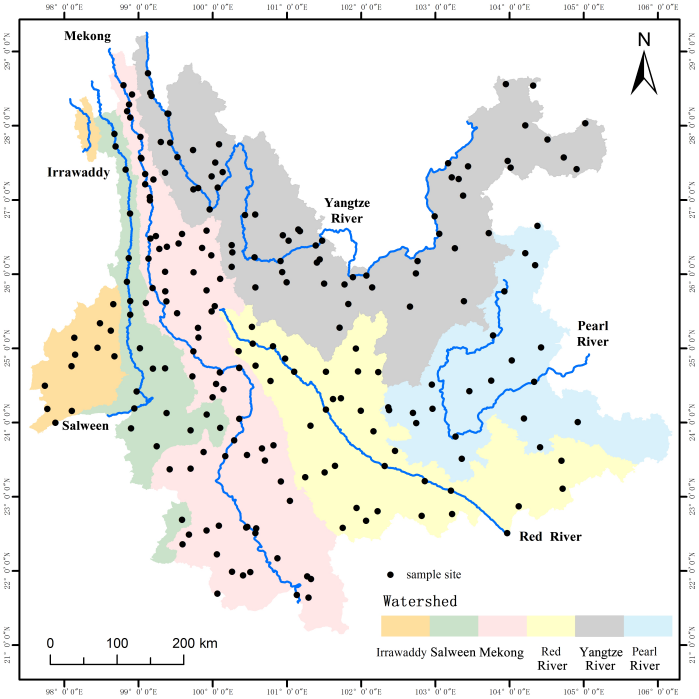
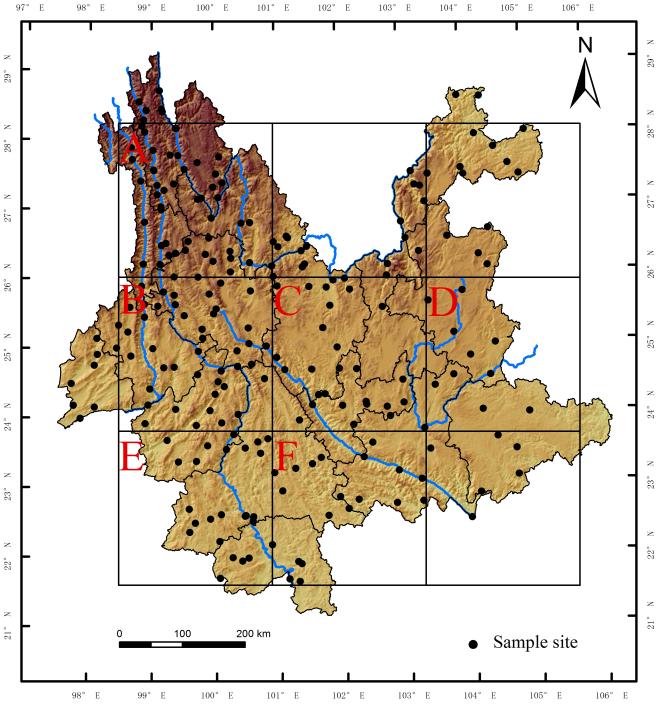
a. Map of sampling sites. b. Sampling points at each site.



Figure S2.

Geographic environments of the sampling sites.

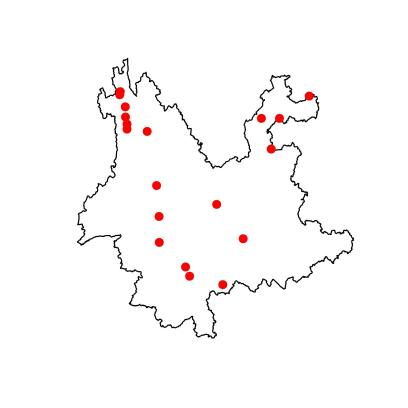
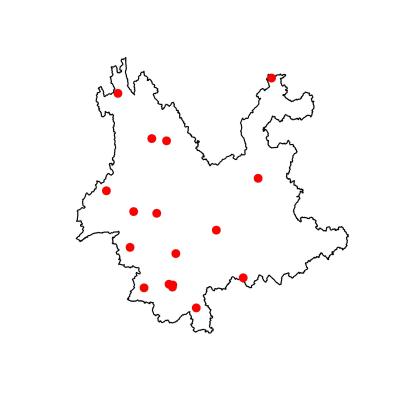
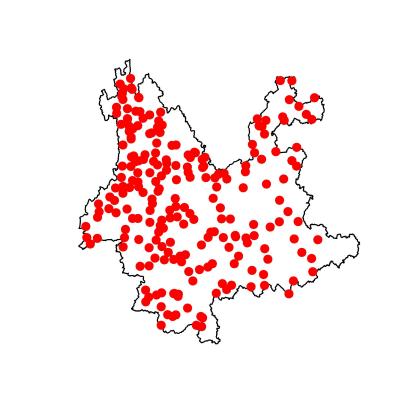
**Figure S3.**



b

a

Divisions of a) grids and b) watershed units.



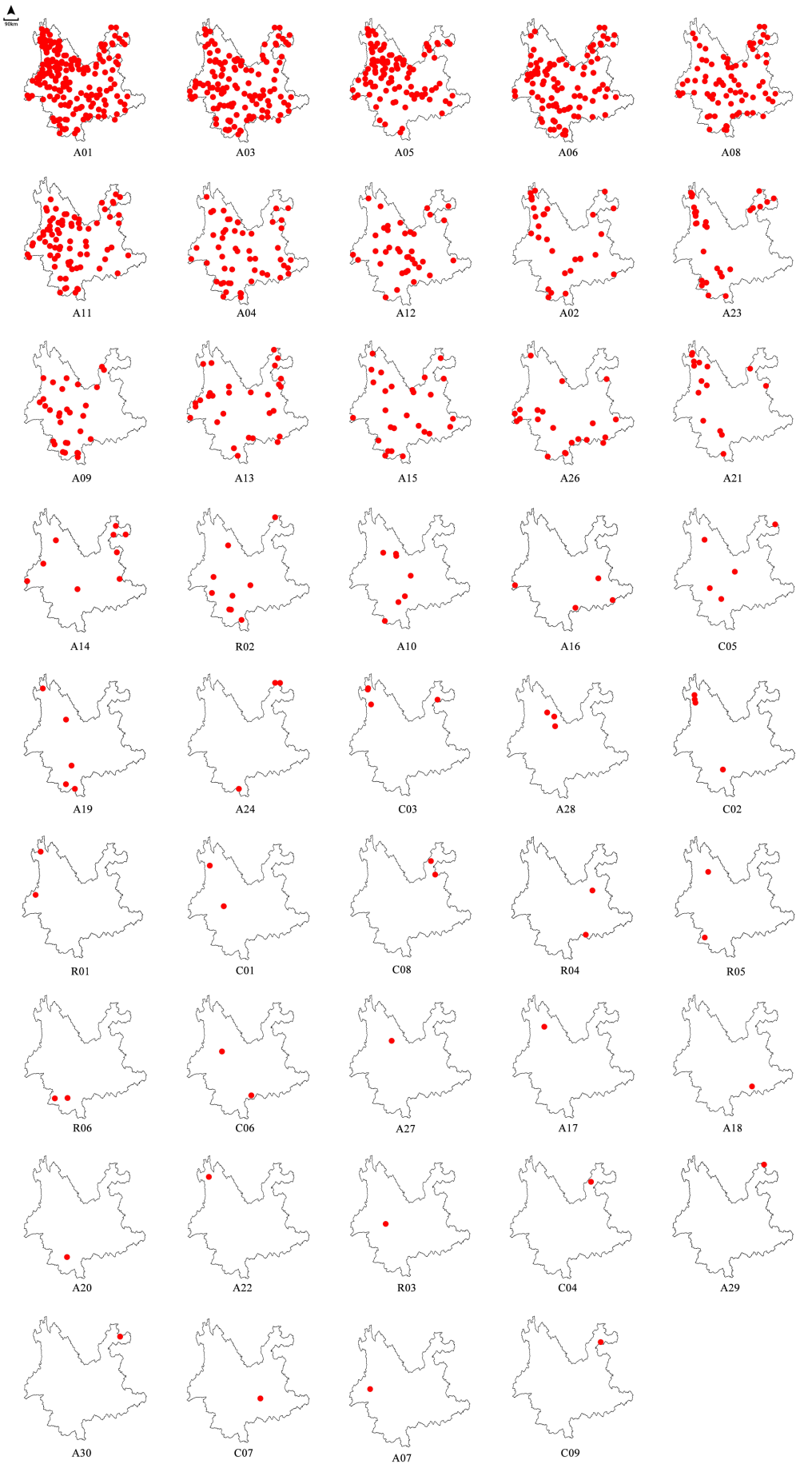
*Arthrobotrys*

*Drechslerella*

*Dactylella*

Figure S4.

Biogeographic patterns of each genus of Nematode-Trapping Fungi.

Figure S5.

Biogeographic patterns of each genus of Nematode-Trapping Fungi.

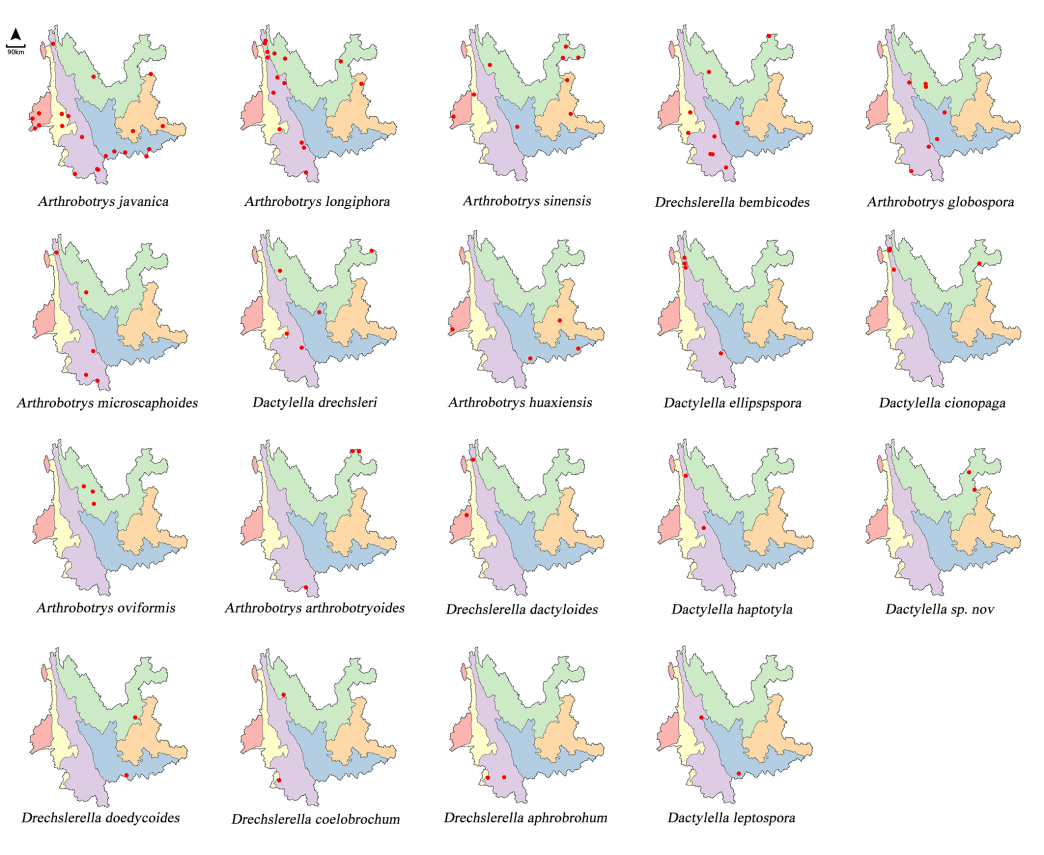
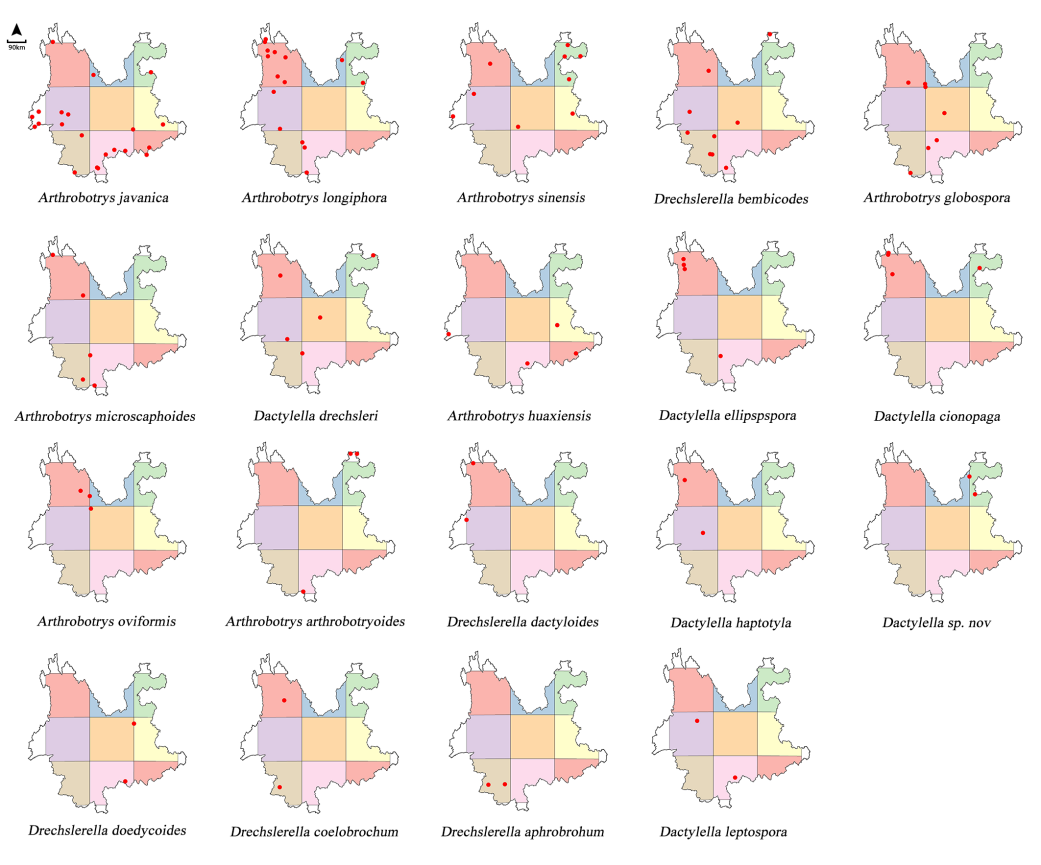
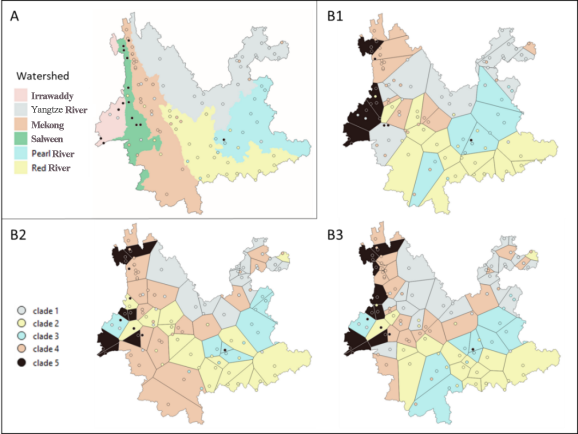
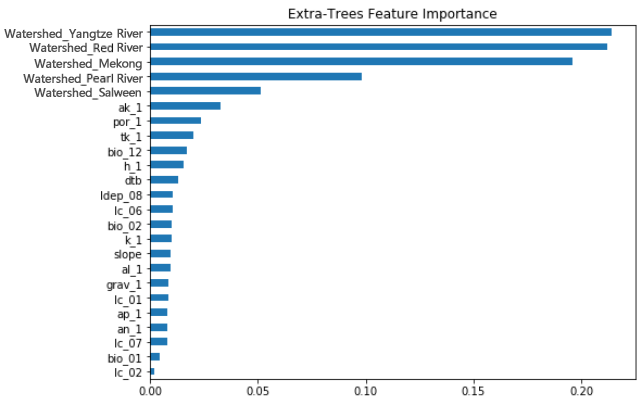


Figure S6.

Biogeographic patterns of rare species in the pespective of a) grids and b) watersheds.

**Figure S7.**

Watershed map (A) and three examples of maps generated using Voronoi polygons from different training sets (Referring to Table S4, B1: v01, B2: v23, B3: v45, respectively). The points indicate species. The legend of the clades refers to all four maps. The same colors are used to highlight correspondence between clades and watershedes, except for Salween and Irrawaddy which host the same clade (clade 5). By training the whole dataset, using the best parameters selected with grid search and cross-validation, we obtained a train accuracy of 70%, while testing on the out-of-bag samples resulted in a score of 68%. The variables best explained the model’s performance were the four basins while the other variables had very little incidence, as shown by the impurity-based feature importance approach (Figure S8) and confirmed by the permutation feature importance approach (Table S5). Feature importance showed that the watersheds are the best predictors.

**Figure S8.**

Impurity-based feature importance from the Extra-Trees classification results.

Table S1.

List of environmental factors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Full Name** | **Units and Processing** | **Time** | **Reference** |
| **bio\_01** | Annual Mean Temperature | degrees °C | 1970-2000  (average) | Fick SE & Hijmans RJ (2017)  WorldClim V2  http://worldclim.org/ |
| **bio\_02** | Mean Diurnal Range | degrees °C  (Mean of monthly(max temp - min temp)) |
| **bio\_03** | Isothermality | (bio\_02/bio\_07)(\*100) |
| **bio\_04** | Temperature Seasonality | (standard deviation \* 100) |
| **bio\_05** | Max Temperature of Warmest Month | degrees °C |
| **bio\_06** | Mean Temperature of Coldest Month | degrees °C |
| **bio\_07** | Temperature Annual Range | degrees °C; (bio\_05 - bio\_06) |
| **bio\_08** | Mean Temperature of Wettest Quarter | degrees °C |
| **bio\_09** | Mean Temperature of Driest Quarter | degrees °C |
| **bio\_10** | Mean Temperature of Warmest Quarter | degrees °C |
| **bio\_11** | Mean Temperature of Coldest Quarter | degrees °C |
| **bio\_12** | Annual Precipitation | millimeters |
| **bio\_13** | Precipitation of Wettest Month | millimeters |
| **bio\_14** | Precipitation of Driest Month | millimeters |
| **bio\_15** | Precipitation Seasonality | Coefficient of Variation |
| **bio\_16** | Precipitation of Wettest Quarter | millimeters |
| **bio\_17** | Precipitation of Driest Quarter | millimeters |
| **bio\_18** | Precipitation of Warmest Quarter | millimeters |
| **bio\_19** | Precipitation of Coldest Quarter | millimeters |
| **srad** | Solar Radiation | kJ m-2 day-1; Mean of monthly averages |
| **vapr** | Water Vapor Pressure | kPa; Mean of monthly averages |
| **wind** | Wind Speed | m s-1; Mean of monthly averages |
| **cloud\_mean** | Mean Annual Cloud Frequency | (%\*100)  Percentage of cloudy days | 2000-2014  (average) | Wilson AM & Jetz W (2016)  EarthEnv  https://www.earthenv.org/ |
| **cloud\_seas** | Seasonal Cloud Concentration Index | (%\*100)  0: all months equally cloudy, 10000: all clouds concentrated in one month |
| **elevation** |  | meters |  | Amatulli G et al. (2018)  EarthEnv  https://www.earthenv.org/ |
| **slope** |  | degrees |  |
| **asp\_cos** |  | (cos(aspect)); -1 to 1 for South to North |  |
| **asp\_sin** |  | (sin(aspect)); -1 to 1 for West to East |  |
| **northness** |  | (sin(slope)\*cos(aspect))  -1 to 1 for South and vertical to North and vertical |  |
| **eastness** |  | (sin(slope)\*sin(aspect))  -1 to 1 for West and vertical to East and vertical |  |
| **watershed** | Location within a major watershed among:  Irrawaddy, Nujiang, Lancang, Jinsha, Red river, Pearl river | Computed from a Digital Elevation Model  (CGIAR-CSI SRTM v4.1, 90 m) |  | Jarvis A et al. (2008)  http://srtm.csi.cgiar.org |
| **lc\_01** | Evergreen/Deciduous Needleleef Trees | % cover  Consensus landcover  (GlobCover, MCD12Q1, GLC2000, DISCover) | 1992-2005  (average) | Tuanmu MN & Jetz W (2014)  EarthEnv  https://www.earthenv.org/ |
| **lc\_02** | Evergreen Broadleaf Trees |
| **lc\_03** | Deciduous Broadleaf Trees |
| **lc\_04** | Mixed/Other Trees |
| **lc\_05** | Shrubs |
| **lc\_06** | Herbaceous Vegetation |
| **lc\_07** | Cultivated and Managed Vegetation |
| **lc\_08** | Regularly Flooded Vegetation |
| **lc\_09** | Urban/Built-up |
| **lc\_12** | Open Water |
| **dtb** | Depth to bedrock | meters  Derived from prediction models using 6382 samples over China and 148 covariates |  | Yan F et al. (2020)  Depth to bedrock map of China 100m  http://globalchange.bnu.edu.cn |
| **pdep** | Soil profile depth | centimeters; Soil depth of 2 layers |  | Shangguan W et al. (2013)  Soil Database of China  http://globalchange.bnu.edu.cn |
| **ldep** | Soil layer depth | centimeters; Soil depth of 11 horizons |  |
| **ph** | pH Value (H2O) | pH units; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |
| **som** | Soil Organic Matter | g/100g; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |
| **tn** | Total N |  |
| **tp** | Total P |  |
| **tk** | Total K |  |
| **sa** | Particle-Size Distribution: Sand |  |
| **si** | Particle-Size Distribution: Silt |  |
| **cl** | Particle-Size Distribution: Clay |  |
| **grav** | Rock fragment |  |
| **an** | Alkali-hydrolysable N | mg/kg; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |
| **ap** | Available P |  |
| **ak** | Available K |  |
| **cec** | Cation Exchange Capacity | me/100g; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |
| **h** | Exchangeable H+ |  |
| **al** | Exchangeable Al3+ |  |
| **ca** | Exchangeable Ca2+ |  |
| **mg** | Exchangeable Mg2+ |  |
| **k** | Exchangeable K+ |  |
| **na** | Exchangeable Na+ |  |
| **bd** | Bulk density | g cm-3; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |
| **por** | Porosity | cm3 100 cm-3; 2 layers (0-4.5 cm, 4.5-9.1 cm) |  |

Table S2.

Variables selection stages and final data set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Selection stage** | **Climate** | **Topography** | **Landcover** | **Soil** |
| **multinomial**  **logistic**  **regression** | bio\_14,  cloud\_mean, cloud\_seas | asp\_cos, asp\_sin,  northness,  eastness | lc\_03, lc\_04, lc\_05,  lc\_08, lc\_09, lc\_12,  lai, fpar | ldep\_01, ldep\_02, ldep\_04,  ldep\_05, ldep\_06, ldep\_09,  ldep\_10, ldep\_11, tn\_2  som\_1, som\_2, grav\_2  tp\_1, tp\_2, bd\_1, bd\_2  cec\_1, cec\_2, ca\_1, ca\_2  mg\_1, mg\_2, sa\_1, sa\_2  si\_1, si\_2, cl\_1, cl\_2  bd\_1, bd\_2, sa\_1, sa\_2,  si\_1, si\_2, cl\_1, cl\_2 |
| **collinearity**  **and**  **multicollinearity** | bio\_03, bio\_04, bio\_05, bio\_06,  bio\_07, bio\_08, bio\_09, bio\_10,  bio\_11, bio\_13, bio\_15, bio\_16,  bio\_17, bio\_18, bio\_19,  srad, vapr, wind | elevation |  | pdep\_01, pdep\_02, ldep\_03,  ldep\_07, ph\_1, ph\_2, tn\_1,  tk\_2, an\_2, ap\_2, ak\_2, h\_2,  al\_2, k\_2, na\_1, na\_2, por\_2 |
| **Final**  **dataset** | bio\_01, bio\_02, bio\_12 | slope | lc\_01, lc\_02,  lc\_06, lc\_07 | por\_1, ak\_1, ldep\_08, h\_1,  al\_1, dtb, an\_1, ap\_1,  grav\_1, tk\_1, k\_1 |

Table S3.

List of detected species of Nematode-Trapping Fungi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species | OF |  | Species | OF |
| *Arthrobotrys oligospora* | 22.13% |  | *Drechslerella dactyloides* | 0.18% |
| *Arthrobotrys multiformis* | 10.53% |  | *Dactylella ellipspspora* | 0.18% |
| *Arthrobotrys conoides* | 7.96% |  | *Arthrobotrys arthrobotryoides* | 0.18% |
| *Arthrobotrys thaumasia* | 6.13% |  | *Dactylella haptotyla* | 0.13% |
| *Arthrobotrys eudermata* | 4.76% |  | *Arthrobotrys oviformis* | 0.13% |
| *Arthrobotrys oudemansii* | 4.71% |  | *Arthrobotrys robusta* | 0.13% |
| *Arthrobotrys cystosporia* | 2.93% |  | *Drechslerella aphrobrohum* | 0.09% |
| *Arthrobotrys rutgeriense* | 2.00% |  | *Drechslerella coelobrochum* | 0.09% |
| *Arthrobotrys superba* | 1.91% |  | *Drechslerella doedycoides* | 0.09% |
| *Arthrobotrys xiangyunensis* | 1.73% |  | *Dactylella sp. nov* | 0.09% |
| *Arthrobotrys sphaeroides* | 1.69% |  | *Dactylella leptospora* | 0.09% |
| *Arthrobotrys vermicola* | 1.51% |  | *Arthrobotrys elegans* | 0.09% |
| *Arthrobotrys fusiformis* | 1.38% |  | *Drechslerella inquisitor* | 0.04% |
| *Arthrobotrys javanica* | 0.98% |  | *Dactylella atractoides* | 0.04% |
| *Arthrobotrys longiphora* | 0.67% |  | *Dactylella rhombospora* | 0.04% |
| *Arthrobotrys sinensis* | 0.44% |  | *Dactylella parvicolla* | 0.04% |
| *Drechslerella bembicodes* | 0.40% |  | *Arthrobotrys sp. nov* | 0.04% |
| *Arthrobotrys huaxiensis* | 0.31% |  | *Arthrobotrys polycephala* | 0.04% |
| *Arthrobotrys globospora* | 0.27% |  | *Arthrobotrys yunnanensis* | 0.04% |
| *Dactylella drechsleri* | 0.22% |  | *Arthrobotrys botryospora* | 0.04% |
| *Dactylella cionopaga* | 0.22% |  | *Arthrobotrys pyriformis* | 0.04% |
| *Arthrobotrys microscaphoides* | 0.22% |  | *Arthrobotrys guizhouensis* | 0.04% |

Table S4.

The 45 maps generated using polygons all had lower accuracy than the maps watersheds generated.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Train** | **Test** | **Correct** | **Accuracy** |  |  | **Train** | **Test** | **Correct** | **Accuracy** |
| v01 | 37 | 88 | 44 | **0.5** |  | v24 | 62 | 63 | 24 | 0.38 |
| v02 | 37 | 88 | 38 | 0.43 |  | v25 | 62 | 63 | 24 | 0.38 |
| v03 | 37 | 88 | 36 | 0.41 |  | v26 | 68 | 57 | 23 | 0.4 |
| v04 | 37 | 88 | 37 | 0.42 |  | v27 | 68 | 57 | 18 | 0.32 |
| v05 | 37 | 88 | 40 | 0.45 |  | v28 | 68 | 57 | 22 | 0.39 |
| v06 | 43 | 82 | 40 | 0.49 |  | v29 | 68 | 57 | 25 | 0.44 |
| v07 | 43 | 82 | 35 | 0.43 |  | v30 | 68 | 57 | 23 | 0.4 |
| v08 | 43 | 82 | 27 | 0.33 |  | v31 | 75 | 50 | 15 | 0.3 |
| v09 | 43 | 82 | 32 | 0.39 |  | v32 | 75 | 50 | 15 | 0.3 |
| v10 | 43 | 82 | 27 | 0.33 |  | v33 | 75 | 50 | 17 | 0.34 |
| v11 | 50 | 75 | 25 | 0.33 |  | v34 | 75 | 50 | 19 | 0.38 |
| v12 | 50 | 75 | 30 | 0.4 |  | v35 | 75 | 50 | 17 | 0.34 |
| v13 | 50 | 75 | 33 | 0.44 |  | v36 | 81 | 44 | 14 | 0.32 |
| v14 | 50 | 75 | 25 | 0.33 |  | v37 | 81 | 44 | 9 | 0.2 |
| v15 | 50 | 75 | 29 | 0.39 |  | v38 | 81 | 44 | 11 | 0.25 |
| v16 | 56 | 69 | 26 | 0.38 |  | v39 | 81 | 44 | 8 | 0.18 |
| v17 | 56 | 69 | 32 | 0.46 |  | v40 | 81 | 44 | 17 | 0.39 |
| v18 | 56 | 69 | 32 | 0.46 |  | v41 | 87 | 38 | 11 | 0.29 |
| v19 | 56 | 69 | 24 | 0.35 |  | v42 | 87 | 38 | 10 | 0.26 |
| v20 | 56 | 69 | 25 | 0.36 |  | v43 | 87 | 38 | 10 | 0.26 |
| v21 | 62 | 63 | 22 | 0.35 |  | v44 | 87 | 38 | 9 | 0.24 |
| v22 | 62 | 63 | 25 | 0.4 |  | v45 | 87 | 38 | 4 | 0.11 |
| v23 | 62 | 63 | 19 | 0.3 |  |  |  |  |  |  |
| **mean** |  |  |  |  |  |  |  |  |  | **0.36** |
| **median** |  |  |  |  |  |  |  |  |  | **0.38** |
| **1st** |  |  |  |  |  |  |  |  |  |  |
| **quart** |  |  |  |  |  |  |  |  |  | **0.32** |
| **3rd quart** |  |  |  |  |  |  |  |  |  | **0.4** |

Table S5.

Permutation feature importance from the Extra-Trees classification results.

|  |  |  |
| --- | --- | --- |
|  | Mean importance | Std |
| Watershed\_Yangtze River | 0.219 | +/- 0.032 |
| Watershed\_Mekong | 0.165 | +/- 0.032 |
| Watershe\_Red River | 0.160 | +/- 0.027 |
| Watershed\_Pearl River | 0.079 | +/- 0.011 |