If you encounter any issues, please feel free to reach out to Wang Ma(马旺) (wma19952022@163.com).

# The User Guide of TaphonomeAnalyst 2.0


**Code availability:** The code of TaphonomeAnalyst 2.0 can be downloaded from https://github.com/wma1995/TaphonomeAnalyst and https://zenodo.org/record/8333176.

**Development environments:** Python 3.8.8

**Article:** TaphonomeAnalyst 2.0: integrative analysis software of taphocoenosis co-currence and geochemical data

# Catalogue

# Introduction

In recent years, there has been a significant increase in the number of studies aimed at elucidating the structures and dynamics of ancient communities through taphonomic co-occurrence network (Guo Ma and Tang, 2023; Muscente, 2019; Xu, 2022). As taphonomic co-occurrence data can reflect symbiotic relationships among various groups to some extent, researchers can plot symbiotic networks by aggregating large amounts taphonomic co-occurrence. However, many of these studies have focused on large-scale marine biota, with small-scale lacustrine biota receiving less attention (Muscente et al., 2019; Xu et al., 2022). Marine research typically operates on large scale, allowing researchers to deduce symbiotic relationships from presence/absence data alone (Muscente et al., 2019; Xu et al., 2022). In contrast, lacustrine studies are conducted on a relatively small scale, where the process of remains transport is more complex and the sedimentary environment more variable. Obtaining co-occurrence data for lacustrine researchs necessitates abundance data, which in turn requires the excavation and identification of a large number of specimens, thereby rendering lacustrine co-occurrence research largely unimplemented over the long term.

However, research on lacustrine co-occurrence networks has also seen meaningful progress. Our team has demonstrated that even under less-than-ideal conditions such as time averaging and varied transportation, a specific taphocoenosis can still retain a wealth of community-level information (Guo, Ma and Tang, 2023). We successfully mapped the Daohugou faunal network from the Middle Jurassic of China and divided it into aquatic, edaphic, mudflat, and silvan modules (Guo, Ma and Tang, 2023). This not only provides statistically significant support for traditional networks based on morphological function and taxonomic uniformitarianism but also paves the way for further quantitative studies in paleocommunity ecology.

Our team previously released the original version of TaphonomeAnalyst (Guo, Ma and Tang, 2023), designed for the study of small-scale terrestrial fossil assemblages. The TaphonomeAnalyst software package serves as a comprehensive tool designed for the downstream community analysis of taphocoenosis data (including abundance and taphonomic preservational grade), primarily focusing on community cluster analysis and community network analysis. It integrates functions for taphocoenosis data importation, analysis, and visualization. The design idea of the software is based on accumulating a substantial volume of OTU cooccurrence data from fossil sampling enables researchers to delineate species coexistence networks and discern various environmental zones therein.

Although has core functions such as parsing ancient networks and ancient water environment clustering of lake-buried communities, a principal limitation of original version of TaphonomeAnalyst is its inability to explore the linkage between ecological variables and community structure, which substantially diminishes its research value. To address this constraint, we have developed an advanced iteration, TaphonomeAnalyst 2.0, designed to expand the spectrum of ecological insights that can be derived from taphonomic data. This updated version incorporates several enhancements (Fig. 1; Fig. 2; Table 1):

(1) Integrating OTUs (Operational Taxonomic Units) abundance with geochemical data for joint analysis.

(2) Adding the capacity to deduce synergies between biological differences and multiple geochemical factors.

(3) Adding visualized co-occurrence networks in different environmental types.

| Module | Function | 1.0 | 2.0 |
|---|---|---|---|
| Assessment of sampling effort and estimation of theoretical maximum biodiversity (Module Ⅰ) | Sobs | √ | √ |
| | Chao1 | √ | √ |
| | ACE | √ | √ |
| Relative abundance of OTU analysis (Module Ⅱ) | | √ | √ |
| Proportion of taphonomic preservational grade of species analysis (Module Ⅲ) | | √ | √ |
| Taphonomic environment analysis (Module Ⅳ) | Including the creation of Veen diagrams that compare the diversity found across sedimentary environments or outcrops. | √ | √ |
| Visualization of geochemical data (Module Ⅴ) | | × | √ |
| Assembling dissimilarity-environmental distance test (Module Ⅵ) | | × | √ |
| Mantel Test between species abundance and ecological environmental variables (Module Ⅶ) | | × | √ |
| Species correlation semi-matrix graphics (Module Ⅷ) | | √ | √ |
| Correlational Network Visualization (Module Ⅸ) | SparCC | × | √ |
| | Pearson | √ | √ |
| | Spearman | √ | √ |
| | Kendall | √ | √ |
| Comparison of networks (Module Ⅹ) | The capability to compare networks under different groups of plots. Visualization of total nodes, total linked nodes, total edges, density, modularity, complexity, degree and robustness. | × | √ |

**Comparison of functions of the two TaphonomeAnalyst versions.**

# Operating guide

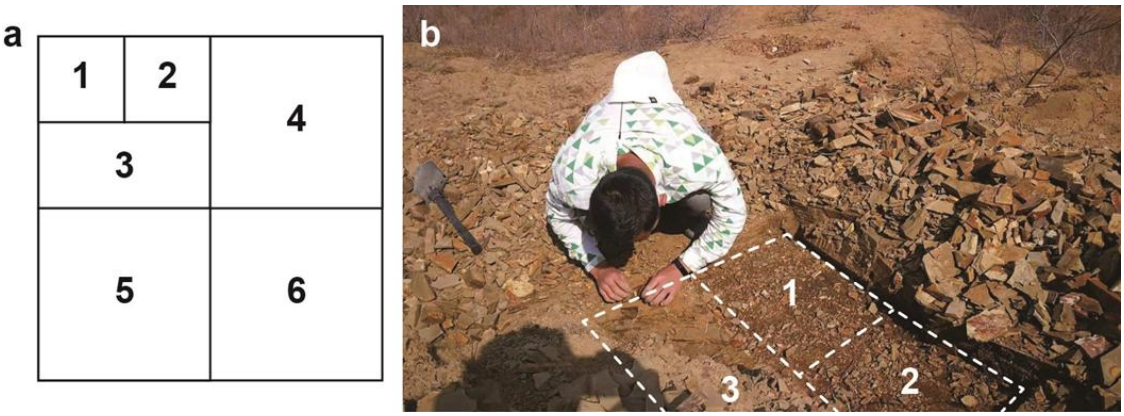
## The basic idea of software

We deem it important to explain the rationale behind naming our software TaphonomeAnalyst and to identify its research focus. Our work has been deeply inspired by the concept of the “microbiome” (Berg et al., 2020; Dhariwal, 2017). Initially, the principal emphasis in microbial ecology research was often placed on microbial communities, which are aggregations of microorganisms coexisting in a shared environment (Berg et al., 2020). Owing to microorganism's small size and distinctive reproductive configurations, microorganisms can easily be conveyed into research settings by various dynamisms. Consequently, microbial communities are prone to hosting a multitude of transient visitors (Berg et al., 2020). These transient visitors typically occur in low abundance and lack ecological functionality. In 1988, Whipps and his colleagues conceptualized the term "microbiome" as a fusion of "micro" and "biome" designating a "characteristic microbial community" within a "reasonably well-defined habitat which has distinct physio-chemical properties" as its "theatre of activity" (Berg et al., 2020; Whipps et al., 1988). The microbiome is an abstract concept derived from the precise statistical analysis of microbial community data at the technical level and the removal of many accidental visitors. In a microbiome, also named ‘abstract characteristic microbial community’, the species co-occurrences and functions within such an assemblage are notably discernible and can exhibit considerable responsiveness to shifts in environmental conditions. Paralleling the definition of a microbial community, the term “taphocoenosis” encompasses an assemblage that includes a mixture of indigenous living in or near water and transient visitors transported from somewhere else. At this juncture, we can successfully define the taphonome in ecology as characterized palaeocommunities in which component function and co-occurrence can be observed and distributed in a certain range of geochemical factors. The taphonome includes a wide variety of species from aquatic settings to nearshore areas, with the maybe exception of the largest predators, whose fossilized remains are typically fragmented and scarce. In ecology, this characteristic community is functional. In stratigraphy, it has stable geochemical data and a consistent sedimentary environment.

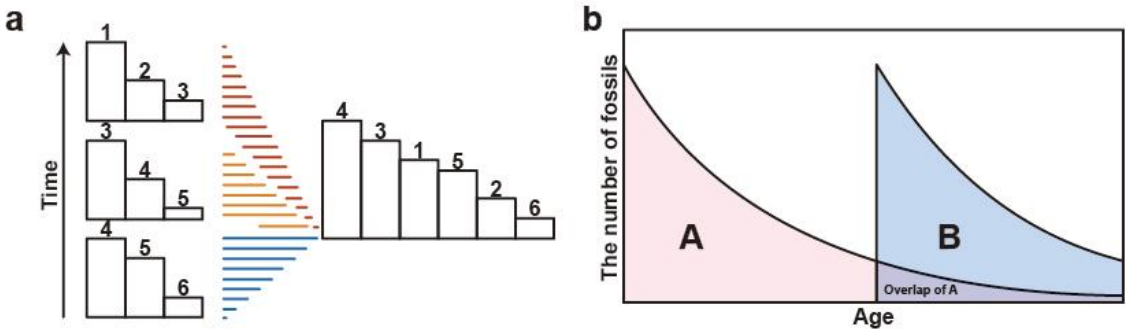
## Data input

Palaeoecological research that involves vertical sampling deviation on the scale of millimeters is uncommon and may lack significance. The majority of palaeoecological datasets are derived from float specimens in several centimeters that means deposition spans thousands of years. For most non-catastrophic depositional environments, samples from strata several centimeters thick might represent a span of thousands of years. In palaeoecological research, it is hard to determine precise assessment time with the burial of organisms. This is understandable given that most fossil assemblages are subject to time-averaging, resulting from the mixing of remains from different communities over time (Dhariwal et al., 2017; Wing et al., 1992; Karr and Clapham, 2015; Wright et al., 2003). As

a result, fossils found within the same stratum might have died and been buried at various points in time. (Fig. 3 a,b) The ecological significance of a taphocoenosis influenced by time-averaging remains a topic of debate (Dhariwal. et al., 2017; Karr and Clapham, 2015; Wing et al., 1992; Wright et al., 2003). However, the prevailing view suggests that a taphocoenosis can either reflect a community under average environmental conditions or certainly constitute the taphonomically altered sum of a community in a particular habitat over some interval of time (Dhariwal et al., 2017; Wing et al., 1992; Karr and Clapham, 2015; Wright et al., 2003). Olszewski (2003) calculated that even if two layers overlap 50% in time, differences between them can differences between them can still be distinguished. The taphonome refers to the characteristic assemblage of organisms ranging from aquatic bodies to nearshore areas. According to theory, as long as the sedimentary environment of the sample layer is consistent, it can be suitable for taphonomic studies. This consistency implies that the characteristic bio-abundance and geochemical data are stable, allowing for meaningful analysis.



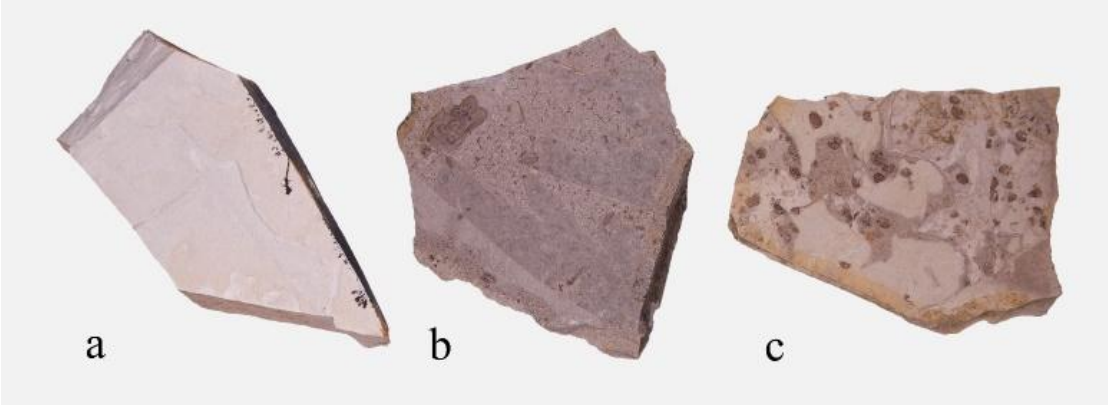

**Nest sampling. a** Schematic diagram of the Nest sampling. The numbers represent the steps of plots excavation. **b** a fieldwork of Nest sampling.



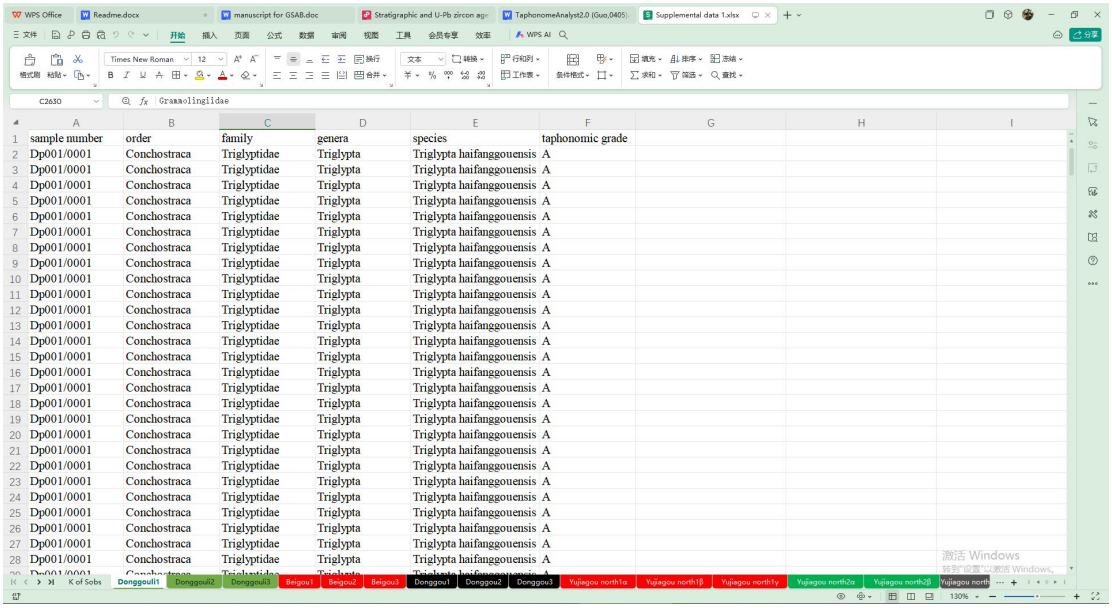

**Impact of time-averaging on the relative abundance of species. a** Time-averaging under variable sedimentary environment (cited in Fürsich and Aberhan, 1990). The bar depicted the numerical abundance of species 1 through 6. The assemblages in three different environments mixed together after forming fossils. Observations indicate that although the three assemblages become intermingled following time averaging, they exhibit a negative quantitative correlation with one another. Statistical methods can be used to separate the three combinations. **b** Overlap in time between two assemblage

distributions (Thomas Olszewski, 1999).

TaphonomeAnalyst 2.0 draws its data from fossil tables and the geochemical abundances of sampling layers. The fossil plots should be excavated using the nest methods. All animal fossils were collected and identified. We suggest using Operational Taxonomic Units (OTUs) at all taxonomic levels, instead of scientific names. An OTU typically represents a species-level taxon that is recognized and morphologically characterized but may not be formally described with a Linnean binomial. This terminology streamlines ecological research by allowing analysis prior to the formal taxonomic description of the taxon in question. Utilizing OTUs is a more annotated and convenient approach, for instance, Orthophlebiidae gen. sp1. Some taxa, such as holometabolic insects, which involve immatures and adults of the same species have large differences in values of indices from their habitats and morphological function. Users should mark immatures with '(I)' after the scientific name to differentiate them from adults. The format of the sample registration form is detailed in Supplementary material 2.




**Fossil samples.** Fossil samples should all be collected from the same stratigraphic layers, without significant changes in the depositional environment within, to ensure stable geochemical data.



| sample number | order | family | genera | species | taphonomic grade |
|---|---|---|---|---|---|
| Dp001/0001 | Conchostraca | Triglyptidae | Triglypta | Triglypta haifanggouensis | A |

**OTU identification form.** Each Sheet represents a plot. Sample number Sample number refers to the unique identification number assigned to a fossil specimen. When multiple animal individuals are present on a flagstone, each individual should be documented separately by entering their details line by line. When identifying specimens, those that cannot be identified should be recorded as "unknown" in the corresponding biological classification category.

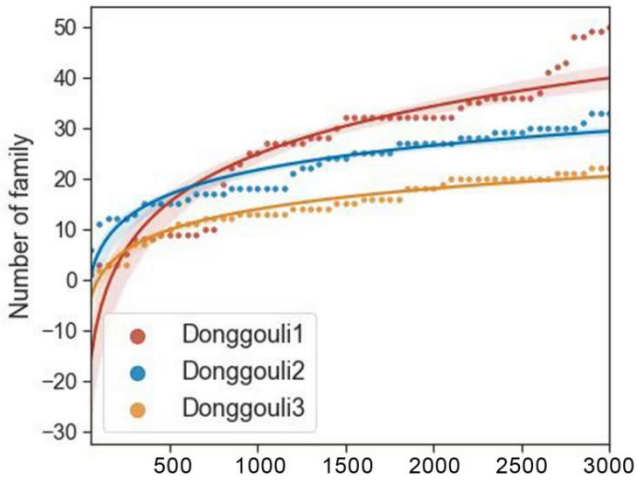
**The recommended output format is PDF, which is more suitable for editing vector elements.**

```
parent_parser = argparse.ArgumentParser(add_help=False)
parent_parser.add_argument('--input', type=str, required=True, help='Absolute or relative path
file.(e.g. "./data.xlsx")')
parent_parser.add_argument('--format', type=str, default='png', choices=['png', 'svg', 'pdf'],
help='Output format.(default: %(default)s)')
parser = argparse.ArgumentParser(description='A comprehensive visual software for study
taphonome.')
parser.add_argument('-v', '--version', action='version', version='TaphonomeAnalyst 2.0')
subparsers = parser.add_subparsers(help='commands')
```

**Assessment of sampling effort and estimation of theoretical maximum biodiversity (Module Ⅰ)**

This module has the capability to employ logarithmic curves for assessing sampling efforts and estimating the theoretical maximum biodiversity. It encompasses the use of Sobs, Chao1, and ACE (Abundance Coverage Estimator) indices (Chao and Yang, 1993; Chao, 1984, 1992, 1993). In the field of lacustrine taphocoenosis research, it is often observed that the abundance of dominant aquatic species far exceeds that of terrestrial species, sometimes by several orders of magnitude. This disparity arises because terrestrial organisms had a much lower probability of fossilization when transported to the water body (Chao and Yang, 1993; Chao, 1984, 1992, 1993). Therefore, we recommend that users use Chao1 or ACE methods whenever possible to evaluate the sample coverage, as these methods are more sensitive to rare species.

$S_{obs}$ is the direct observational diversity, suited to evaluate the coverage of sampling in strata where aquatic species do not have a significant advantage. During the sampling process, some Daohugou samples could contain over 2000 *T. haifanggouensi* among 3000 individuals. In contrast, many terrestrial OTUs only have a few individuals. The $S_{obs}$ curve may tend to flatten out when the number of samples is few, but sample location may still have significant potential diversity. Chao1 is sensitive to OTUs of only one individual, making it more suitable for plots where aquatic species dominate. The abundance coverage estimator considers a wider range of rare species and makes corrections for the coefficient of variation and sample coverage, which is more reasonable. However, due to the difference between the buried community and the present-day community, the definition of the abundance of rare species needs to be considered.




**A presents a visualization of Sampling Coverage curves.** Users have the discretion to determine how many samples should be taken for a change in each steps diversity.

$$Chao1 = S_{obs} + \frac{F_1(F_1-1)}{2(F_2+1)}$$

whereby $S_{obs}$: the direct observational diversity. $F_1$: the number of OTUs whose abundance is one. $F_2$: the number of OTUs whose abundance is two.

$$S_{ace} = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}}\gamma^2_{ace}$$

$S_{abund}$: The count of abundant OTUs (Operational Taxonomic Units) typically includes those exceeding a rarity threshold-often set at 10 individuals. However, in our experience with fieldwork, a threshold of 10 may be excessive. Our software allows the user to set the rare species threshold, when all samples are pooled. $S_{rare}$: the number of rare OTUs (with less than or equal to rare threshold individuals) when all samples are pooled. $C_{ace}$: the sample abundance coverage estimator. $F_1$: hose abundance is one. $\gamma^2_{ace}$: the estimated coefficient of variation for rare OTUs.

$$\gamma^2_{ace} = max\left[\frac{S_{rare}}{C_{ace}}\frac{\sum_{i=1}^{10} i(i-1)F_I}{(N_{rare})(N_{rare-1})} - 1, 0\right]$$

**Users have the flexibility to set the taxonomic level utilized for plotting sampling curves, with the default set at the family rank. Alternative options include order, family, genus, and species levels.**

**$S_{obs}$**

```
samplecurve_parser = subparsers.add_parser(name='samplecurve',
parents=[parent_parser], help='Sampling coverage curve. (Module Ⅰ)\t[regplot]')
samplecurve_parser.add_argument('--level', type=str, default='family', choices=['order',
'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
samplecurve_parser.add_argument('--groups', type=str2dictlist, required=True,
help='Grouping plots (Sheet names) with customized names.\t[e.g.
"plotA:plotA1/plotA2,plotB:plotB1/plotB2"]')
samplecurve_parser.add_argument('--output', type=str, default='./samplecurve',
help='Absolute path or relative path and filename.(default: %(default)s)')
```

```
samplecurve_parser.set_defaults(func=samplecurve)
```

Command line：

python TaphonomeAnalyst2.py samplecurve --input ./Supplementary material2.xlsx --level 'family' --groups 'plot1:plot1-1/plot1-2/plot1-3,plot2:plot2-1/plot2-2/plot2-3,plot3:plot3-1/plot3-2/plot3-3'

**Chao1**

```
chao_parser = subparsers.add_parser(name='chao', parents=[parent_parser], help='Chao1 potential diversity curve. (Module Ⅰ)\t[regplot]')
chao_parser.add_argument('--level', type=str, default='family', choices=['order', 'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
chao_parser.add_argument('--groups', typedissenvtest =str2dictlist, required=True, help='Grouping plots (Sheet names) with customized names.\t[e.g. "plotA:plotA1/plotA2,plotB:plotB1/plotB2"]')
chao_parser.add_argument('--output', type=str, default='./chao', help='Absolute path or relative path and filename.(default: %(default)s)')
chao_parser.set_defaults(func=chao)
```

Command line：

python TaphonomeAnalyst2.py chao --input ./Supplementary material2.xlsx --level 'family' --groups 'plot1:plot1-1/plot1-2/plot1-3,plot2:plot2-1/plot2-2/plot2-3,plot3:plot3-1/plot3-2/plot3-3'

**Ace: In the field of microbial ecology, the default abundance threshold for rare species is set below ten. However, this value may be overestimated. It is recommended that users make adjustments accordingly and provide explanations in their studies.**

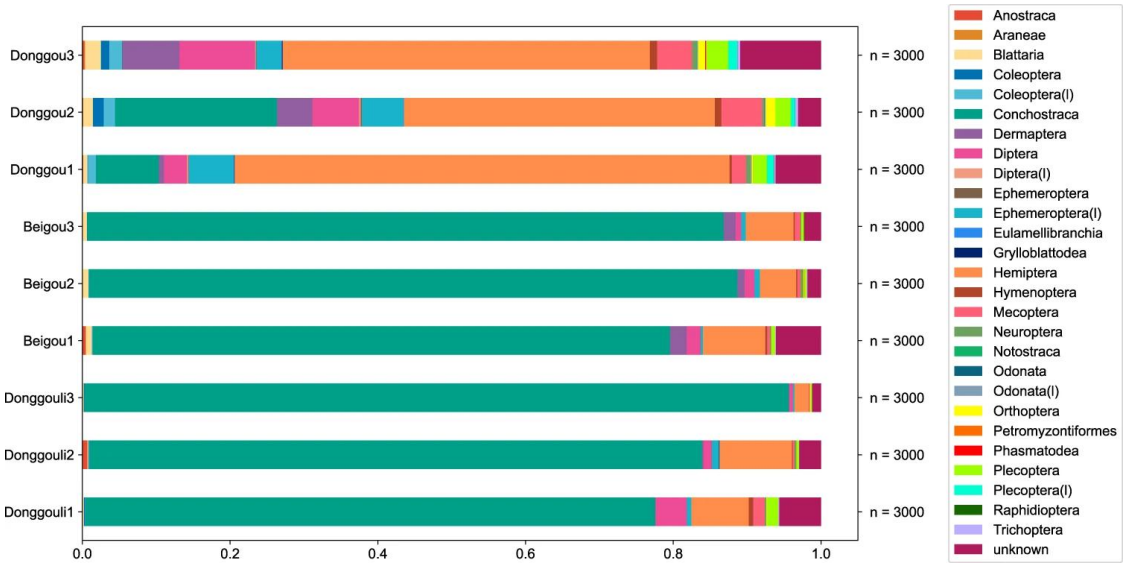
```
ace_parser = subparsers.add_parser(name='ace', parents=[parent_parser], help='ACE potential diversity curve. (Module Ⅰ)\t[regplot]')
ace_parser.add_argument('--level', type=str, default='family', choices=['order', 'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
ace_parser.add_argument('--groups', type=str2dictlist, required=True, help='Grouping plots (Sheet names) with customized names.\t[e.g. "plotA:plotA1/plotA2,plotB:plotB1/plotB2"]')
ace_parser.add_argument('--output', type=str, default='./ace', help='Absolute path or relative path and filename.(default: %(default)s)')
ace_parser.add_argument('--rare', type=int, default=10, help='ACE rare threshold.(default: %(default)s)')
ace_parser.set_defaults (func=ace)
```

Command line：

python TaphonomeAnalyst2.py ace --input ./Supplementary material2.xlsx --level 'family' --groups 'plot1:plot1-1/plot1-2/plot1-3,plot2:plot2-1/plot2-2/plot2-3,plot3:plot3-1/plot3-2/plot3-3' --rare 10

## Relative abundance of OTU analysis (Module Ⅱ)

This module facilitates the generation of bar graphs illustrating species abundances. Given that the abundance of fossils does not reliably indicate the actual diversity of the original community, the relative abundance suggested by an OTU should be regarded as an imperfect measure (McNamara et al., 2012; Smith and Moe-Hoffman, 2007; Wang et al., 2019). The representation of species within a taphocoenosis, or fossil assemblage, is influenced by factors such as taxonomic unit and physical size (Smith and Moe-Hoffman, 2007; McNamara et al., 2012). To some degree, these abundances can offer insights into the varying source distances and trophic levels of the species present in the fossil community.




**Compositional proportion of the Yanliao Fauna by taxa from the sampled plots.** Taxon level is order.

**Users have the flexibility to set the taxonomic level.**

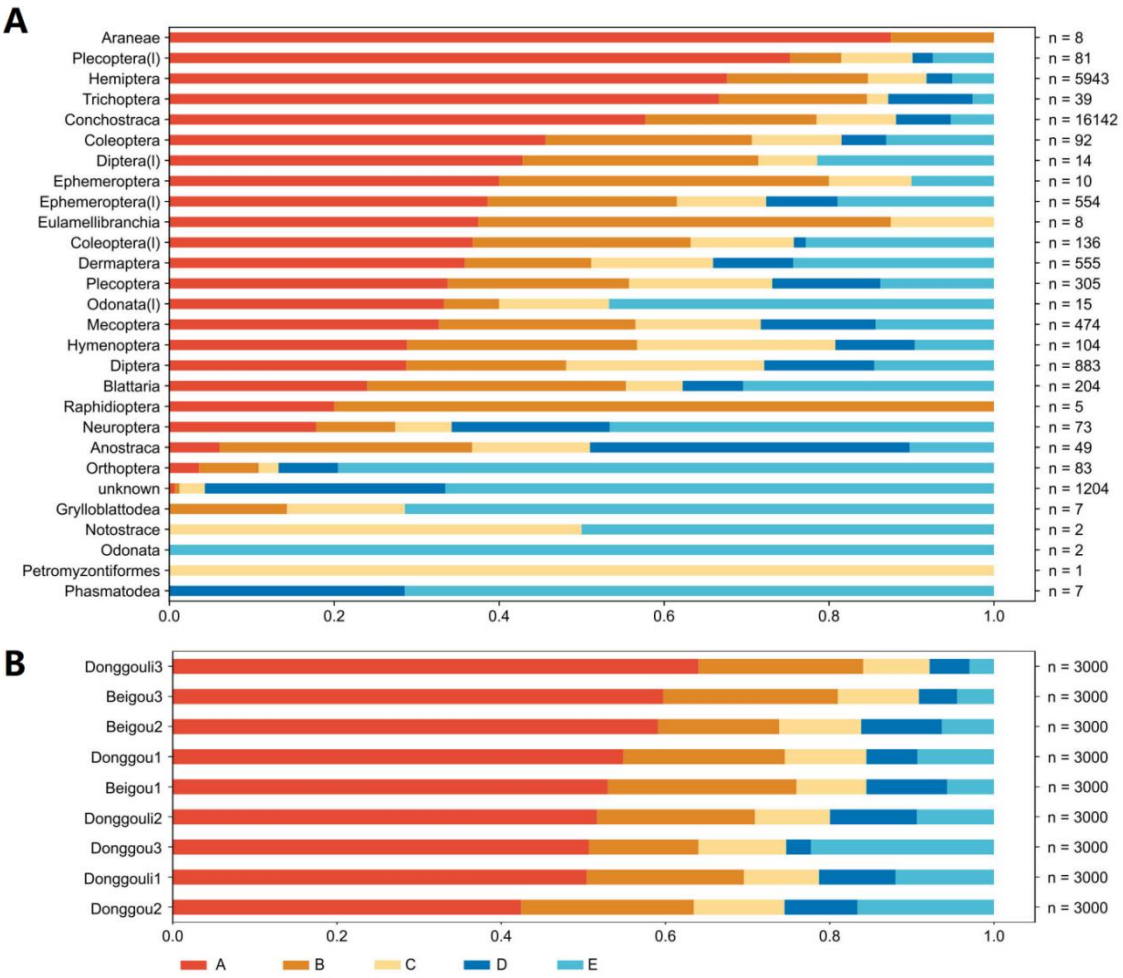
```
abundplots_parser = subparsers.add_parser(name='abundplots',
parents=[parent_parser], help='Abundance-sampling plots. (Module Ⅱ)\t[barh]')
abundplots_parser.add_argument('--level', type=str, default='order', choices=['order',
'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
abundplots_parser.add_argument('--output', type=str, default='./abundplots',
help='Absolute path or relative path and filename.(default: %(default)s)')
abundplots_parser.set_defaults(func=abundplots)
```

Command line：

```
python TaphonomeAnalyst2.py abundplots --input ./Supplementary material2.xlsx --level
'order'
```

## Proportion of taphonomic preservational grade of species analysis (Module Ⅲ)

The taphonomic grade module evaluates the preservation quality of fossils, as reflected by their structural integrity and joint articulations (Guo, Ma and Tang, 2023). The taphonomic grade is categorized into five levels, ranging from A to E, where A represents the best preservation and E indicates the poorest (Guo, Ma and Tang, 2023). This module offers taphonomic grade bar graphs for various taxa. The taphonomic grade module assesses the preservational quality of the fossils based on their structural integrity, such as the extent of intact articulations of joints visible in the fossil. The taphonomic grade can also be utilized to interpret the distance between the original habitat and its eventual deposition into lacustrine sediment. Although influenced by factors like the robustness of body parts, particularly appendages, and body size, this method is extensively employed in taphonomic analyses. Users have the option to choose the level of classification taxa.




**Proportion of taphonomic grades.** Taphonomy Analyst 2.0 offers the capability to selectively output the preservation levels of different OTUs (Operational Taxonomic Units) or the varying degrees of preservation across different sample plots. **a** Proportion of

taphonomic grades (A–E) of the taxa. **b** Proportion of taphonomic grades (A–E) by sampling plot.

**Proportion of taphonomic grades (A–E) of the taxa. Users need to set the taxonomic level themselves.**

```
TGotus_parser = subparsers.add_parser(name='TGotus', parents=[parent_parser],
help='Taphonomic grades-taxa. (Module Ⅲ)\t[barh]')
TGotus_parser.add_argument('--level', type=str, default='order', choices=['order', 'family', 'genera',
'species'], help='Taxonomic level.(default: %(default)s)')
TGotus_parser.add_argument('--output', type=str, default='./TGotus', help='Absolute path or
relative path and filename.(default: %(default)s)')
TGotus_parser.set_defaults(func=TGotus)
```

Command line：

python TaphonomeAnalyst2.py TGotus --input ./Supplementary material2.xlsx --level 'order'

**Proportion of taphonomic grades (A–E) by sampling plot. Users need to set the taxonomic level themselves.**

```
TGplots_parser = subparsers.add_parser(name='TGplots', parents=[parent_parser],
help='Taphonomic grades-sampling plots (in customized order). (Module Ⅲ)\t[barh]')
TGplots_parser.add_argument('--groups', type=str2list, default=None, help='Environment
groups.(Recommend to group the plots by different aquatic and terrestrial environments)\t[e.g.
"plotA1/plotB2,plotB1/plotA2,plotC1/plotC2"]')
TGplots_parser.add_argument('--output', type=str, default='./TGplots', help='Absolute path or
relative path and filename.(default: %(default)s)')
TGplots_parser.set_defaults(func=TGplots)
```

Command line：

Default sort

python TaphonomeAnalyst2.py TGplots --input ./Supplementary material2.xlsx

Custom sort

python TaphonomeAnalyst2.py TGplots --input ./Supplementary material2.xlsx --groups 'plot3-3/plot3-1/plot3-2,plot1-1/plot2-2/plot2-1/plot2-3/plot1-2/plot1-3'

| Grade | Clam shrimps and bivalves | Other arthropods | Vertebrates |
|---|---|---|---|
| A | Shell edge > 90% preserved. growth bands are fully clear. | > 90% preserved. Body articulated, wing veins visible and almost complete. | Body and limbs are complete and articulated. |
| B | Shell edge > 70% preserved. growth bands are almost clear. | 80–90% preserved. Body almost complete, including head, thorax, abdomen and thoracic appendages, details such as antennae or cerci lost. | 70–80% torso and limbs are complete. Partial joint displacement. |
| C | Shell edge > 60% preserved. growth band are partially clear. | 60–80% preserved. Body deformed, at least one of six legs lost. | 60–70% torso preserved. |
| D | Shell edge >50% preserved. | 30–60% preserved. Wings disarticulated, remains of head, thorax and abdomen preserved | Torso with missing tail or head. |
| E | Shell fragments | < 30% preserved. High disarticulated body, isolated structures such as single legs, abdomen and/or wings preserved. | Scattered bones. |


**Definition of taphonomic grades.** The taphonomic grade categorizes the preservational quality of fossils into five levels, ranging from A to E. A signifies the highest quality of preservation, whereas E denotes the poorest quality of preservation.

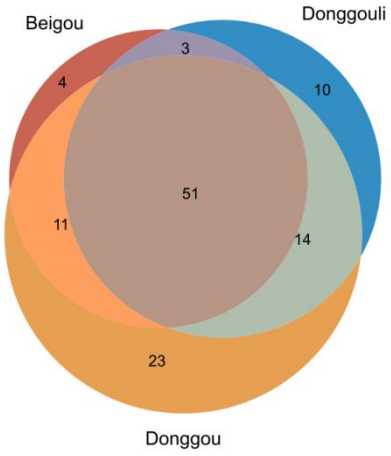
## Taphonomic environment analysis (Module Ⅳ)

The primary aim of this module is to contrast and display the variations in species abundance across different plots, subsequently categorizing these samples into distinct groups. Since aquatic organisms typically undergo minimal transport during fossilization, their populations remain relatively stable, offering a robust reflection of the environmental conditions at the time of fossil deposition. The hierarchical clustering is calculated after a filter threshold set at an individual count greater than 5. Hierarchical clustering is performed using the Average linkage method and Bray-Curtis distance, which are commonly employed in biodiversity studies. We suggest that users conduct a joint analysis by integrating environmental clustering, species distribution, and geochemical heatmap collages. By integrating geochemical heat maps, users can clearly discern the distribution of aquatic OTUs abundance across different clusters, as well as associated differences in environmental factors. Additionally, it generates Venn diagrams that illustrate the differences in diversity within different taphonomic environments.An Average assign is:

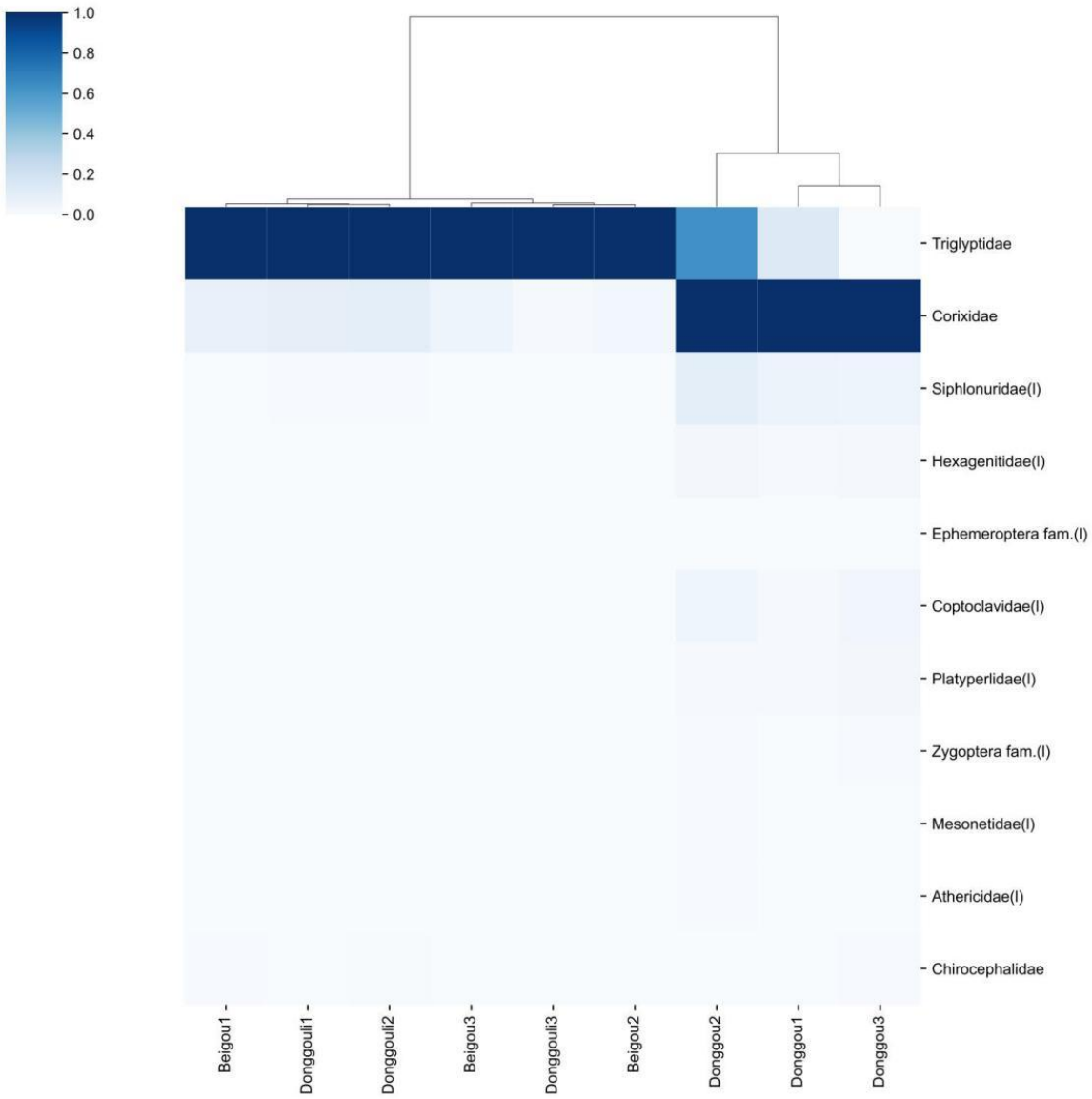
$$d_{(u,v)} = \sum_{ij} \frac{d(u[i], v[j])}{(|u| * |v|)}$$

And the Bray-Curtis distance is:

$$d_{(u,v)} = \frac{\sum_i |u_i - v_i|}{\sum_i |u_i + v_i|}$$




**This module also provides Veen maps comparing biodiversity in different sedimentary environments and outcrops.**




**Diversity and abundance comparisons of different plots and sedimentary environments.** Hierarchical clustering of nine plots of sedimentary environments. Plots were clustered based on aquatic taxonomic abundance (n > 5) using the Average assigns (clustering methods) and Bray-Curtisdistance. The clustering results show that the sedimentary environments can be divided into three types. The color of the heat map indicates the normalized abundance of biological distribution.

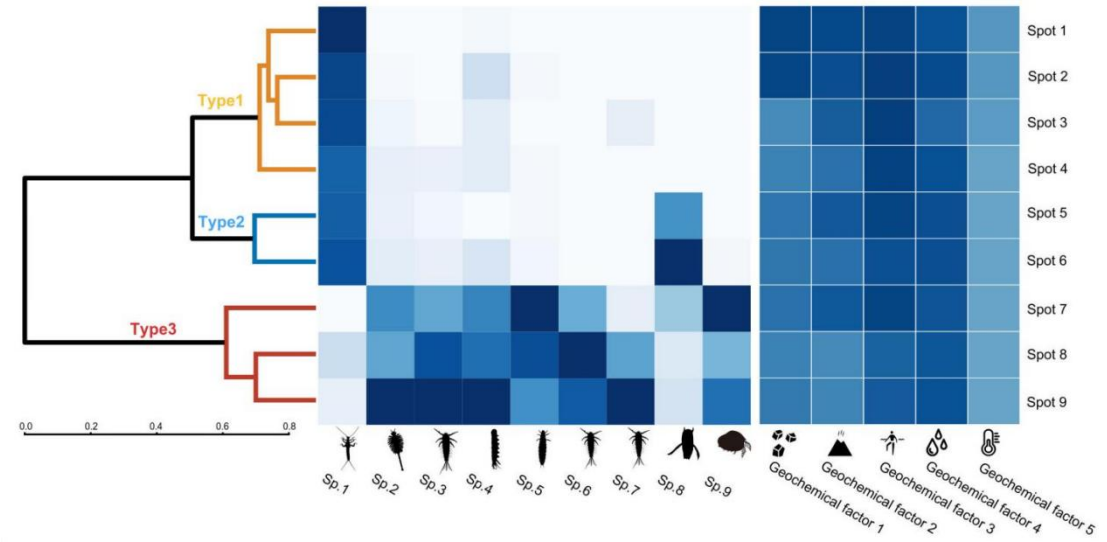
## Visualization of geochemical data (Module Ⅴ)

The environmental variables that shaped ancient ecosystems are not directly measurable and often necessitate the extrapolation of geochemical data for accurate interpretation. The ratios of elements and oxides have been widely used in ancient environmental studies to gain insights into salinity, temperature, water depth, humidity, intensity of volcanic activities and so on (Bai et al., 2020; Chen, Wan, 1999; Chen et

al.,1999; Feng et al., 2003; Hou et al., 2018; Hou et al., 2023; Fu et al., 2018; Harnois,1988; Nesbitt, Young, 1982; McLennan, 1993; Stanistreet et al., 2020; Swain et al., 2022; Wang et al., 2023; Yang et al., 2022; Zhou and Sun, 2023). Beyond serving as final resting grounds for fossilized remains, ancient water bodies act as primary archives for geochemical information, frequently encapsulating environmental contexts crucial to the majority of inhabiting areas for terrestrial organisms from aquatic habitats to littoral zones. This module provides geochemical heat map of different plots. This module provides geochemical heat map of different plots. This figure can be automatically combined with the sedimentary environment cluster tree and the distribution heat map of aquatic organisms. It can simultaneously reflect the distribution of biological and geochemical factors in different environmental groups.

| sample number | Geochemical factor1 | Geochemical factor2 | Geochemical fa | Geochemical factor4 | Geochemical factor5 |
|---|---|---|---|---|---|
| plot1-1 | 0.050277585 | 2.085012087 | 6.381 | 8.215891304 | 12.25805475 |
| plot1-2 | 0.044962642 | 1.88177864 | 6.57 | 8.253242777 | 7.625969722 |
| plot1-3 | 0.120648387 | 6.602077869 | 6.248 | 8.267124956 | 8.960424791 |
| plot2-1 | 0.309145015 | 8.565605513 | 12.865 | 21.71957373 | 8.620356037 |
| plot2-2 | 0.091410129 | 3.985073903 | 7.279 | 7.309533891 | 19.81483572 |
| plot2-3 | 0.06083691 | 5.582370985 | 7.774 | 8.112493614 | 1.450443191 |
| plot3-1 | 0.131808667 | 3.084634553 | 10.252 | 10.18660248 | 12.08372541 |
| plot3-2 | 0.066304692 | 4.216297844 | 6.252 | 7.634473194 | 13.09164963 |
| plot3-3 | 0.10304122 | 2.507430481 | 15.457 | 14.42240219 | 2.46604432 |

**Style of geochemical tables. See** Supplementary material3 for details




**Integrative visualization based on OTUs abundance and geochemical data.** This moudule can be automatically combined with the sedimentary environment cluster tree and the distribution heat map of aquatic organisms. It can simultaneously reflect the distribution of biological and geochemical factors in different environmental groups.

**Module Ⅳ and Ⅴ were integrated together. If geochemical data is not available, only the clustering tree and abundance heatmap will be generated. A list of aquatic OTUs is required for this step. The taxonomic level of aquatic OTUs must correspond your research.**

```
clusterenv_parser = subparsers.add_parser(name='clusterenv', parents=[parent_parser],
help='Hierarchical clustering-sedimentary environment. (Module Ⅳ and Ⅴ)\t[clustermap]')
clusterenv_parser.add_argument('--level', type=str, required=True, choices=['order',
'family', 'genera', 'species'], help='Taxonomic level.(For both statistical and aquatic
OTUs.)')
clusterenv_parser.add_argument('--aquatic', type=str2list, default=None, help='Aquatic
OTUs.(default: all OTUs)\t[e.g. "OTU1,OTU2,OTU3"]')
clusterenv_parser.add_argument('--geochem', type=str, required=False, help='Absolute
or relative path geochemical file.(e.g. "./geochem.xlsx")')
clusterenv_parser.add_argument('--output', type=str, default='./clusterenv', help='Absolute
path or relative path and filename.(default: %(default)s)')
clusterenv_parser.set_defaults(func=clusterenv)
```

Command line：

```
python TaphonomeAnalyst2.py clusterenv --input ./Supplementary material2.xlsx
--aquatic 'Daohugounectes primitinus(I),Triglypta haifanggouensis,Triglypta
haifanggouensis,Yanliaocorixa chinensis,Karataviella popovi,Samarura
gigantea(I),Anisoptera fam. gen. sp1.(I),Platyperla platypoda(I),Ferganoconcha
sibirica,Qiyia jurassica(I),Mesomyzon sp1.,Triops sp1.,Chirocephalidae gen.
sp1.,Eurythoracalis mirabilis(I),Shantous lacustris(I),Foliomimus latus(I),Furvoneta
viriosus(I),Furvoneta raucus(I),Mesobaetis sibirica(I),Clavineta eximia(I)' --geochem
'./Supplementary material3.xlsx' --level 'species'
```

Command line：

```
python TaphonomeAnalyst2.py mantel --input ./Supplementary material2.xlsx --rhome
'C:\Program Files\R\R-4.3.2' --geochem './Supplementary material3.xlsx' --aquatic
'Daohugounectes primitinus(I),Triglypta haifanggouensis,Triglypta
haifanggouensis,Yanliaocorixa chinensis,Karataviella popovi,Samarura
gigantea(I),Anisoptera fam. gen. sp1.(I),Platyperla platypoda(I),Ferganoconcha
sibirica,Qiyia jurassica(I),Mesomyzon sp1.,Triops sp1.,Chirocephalidae gen.
sp1.,Eurythoracalis mirabilis(I),Shantous lacustris(I),Foliomimus latus(I),Furvoneta
viriosus(I),Furvoneta raucus(I),Mesobaetis sibirica(I),Clavineta eximia(I)' --level_aquatic
'species' --level_terrestrial 'family' --corr 'pearson'
```

**Venn diagrams that show differences in biodiversity across various sedimentary environments or outcrops. Users need to define the taxonomic levels and plots groupings.**

```
divvenn_parser = subparsers.add_parser(name='divvenn', parents=[parent_parser],
help='Venn diagram-sampling locations or environments. (Module Ⅳ)\t[venn]')
```

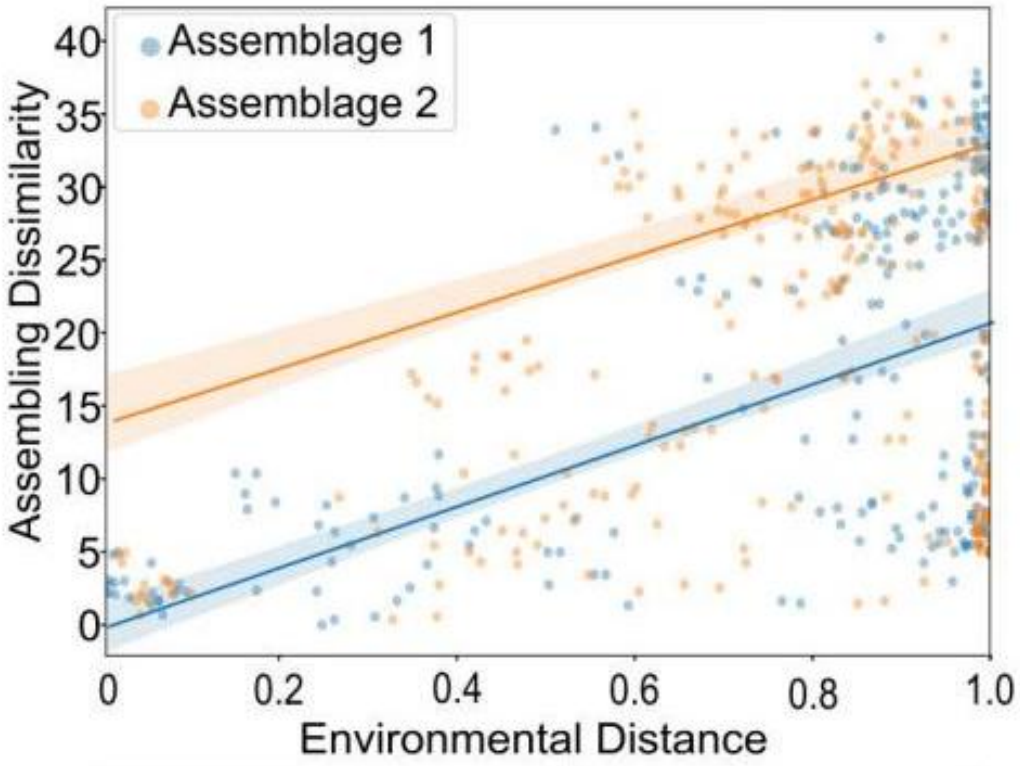
```
divvenn_parser.add_argument('--level', type=str, default='family', choices=['order', 'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
divvenn_parser.add_argument('--groups', type=str2list, required=True, help='Custom Groups.(Recommend to group the plots by environments or locations)\t[e.g. "plotA1/plotB2,plotB1/plotA2,plotC1/plotC2"]')
divvenn_parser.add_argument('--output', type=str, default='./divvenn', help='Absolute path or relative path and filename.(default: %(default)s)')
divvenn_parser.set_defaults(func=divvenn)
```

Command line：

```
python TaphonomeAnalyst2.py divvenn --input ./Supplementary material2.xlsx --groups 'plot3-3/plot3-1/plot3-2,plot1-1/plot2-2/plot2-1/plot2-3/plot1-2/plot1-3' --level 'family'
```

## Assembling dissimilarity – environmental distance test (Module Ⅵ)

Assemblage similarities were quantified using the Bray-Curtis distance derived from various plots, whereas environmental distance was determined by employing a Euclidean distance matrix based on measured geochemical variables. This module quantifies the responses of aquatic and terrestrial community components to changes in environmental factors. In this statistical graph, the greater the slope of pronounced the differentiation among the biological the more pronounced the differentiation among the biological assemblages in response to changes in environmental factors.




**Assembling dissimilarity - environmental distance test.** The slope of the line indicates the degree of abrupt environmental changes corresponding to the assemblage.

**A list of aquatic OTUs is required for this step. The list aquatic OTUs taxonomic level must correspond your research.**

```
dissenvtest_parser = subparsers.add_parser(name='dissenvtest',
parents=[parent_parser], help='Assembling dissimilarity- environmental distance test.
(Module Ⅵ)\t[regplot]')
dissenvtest_parser.add_argument('--geochem', type=str, required=True, help='Absolute
or relative path geochemical file.(e.g. "./geochem.xlsx")')
dissenvtest_parser.add_argument('--aquatic', type=str2list, required=True, help='Aquatic
OTUs.\t[e.g. "OTU1,OTU2,OTU3"]')
dissenvtest_parser.add_argument('--level_aquatic', type=str, required=True,
choices=['order', 'family', 'genera', 'species'], help="Taxonomic level for aquatic OTUs.")
dissenvtest_parser.add_argument('--level_terrestrial', type=str, required=True,
choices=['order', 'family', 'genera', 'species'], help="Taxonomic level for terrestrial OTUs.")
dissenvtest_parser.add_argument('--output', type=str, default='./dissenvtest',
help='Absolute path or relative path and filename.(default: %(default)s)')
dissenvtest_parser.set_defaults(func=dissenvtest)
```

Command line：

```
python TaphonomeAnalyst2.py dissenvtest --input ./Supplementary material2.xlsx
--aquatic 'Daohugounectes primitinus(I),Triglypta haifanggouensis,Triglypta
haifanggouensis,Yanliaocorixa chinensis,Karataviella popovi,Samarura
gigantea(I),Anisoptera fam. gen. sp1.(I),Platyperla platypoda(I),Ferganoconcha
sibirica,Qiyia jurassica(I),Mesomyzon sp1.,Triops sp1.,Chirocephalidae gen.
sp1.,Eurythoracalis mirabilis(I),Shantous lacustris(I),Foliomimus latus(I),Furvoneta
viriosus(I),Furvoneta raucus(I),Mesobaetis sibirica(I),Clavineta eximia(I)' --level_aquatic
'species' --level_terrestrial 'family' --geochem './Supplementary material3.xlsx'
```

## Mantel Test between species abundance and ecological environmental variables (Module Ⅶ)

In 1967, Nathan Mante revolutionized statistical analysis by proposing the Mantel Test. This method advanced beyond the constraints of traditional correlation coefficients, which are only equipped to analyze pairwise relationships among variables within a solitary data matrix. The Mantel Test broke new ground by facilitating the assessment of correlations between two distinct matrices.

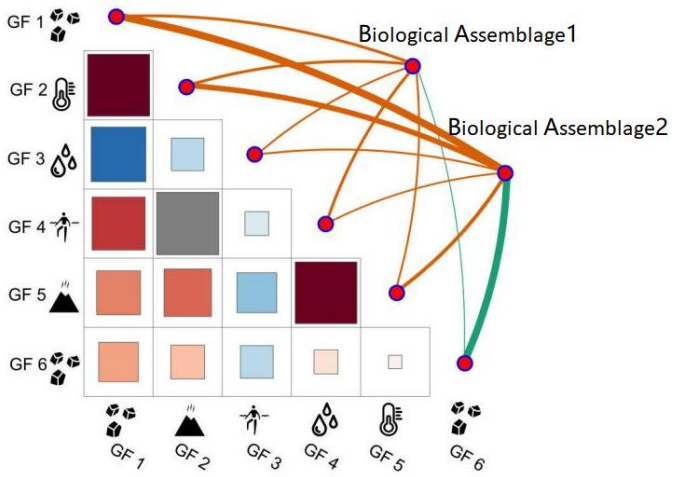
Since its inception, the Mantel Test has been integral to the rapid evolution and application across diverse scientific domains, notably in microbial community ecology. In the realm of paleoecology, the Mantel Test serves as an invaluable tool for probing the connections between geochemical factors and fluctuations in biological abundance. To quantify similarities in species assemblages, the Bray-Curtis dissimilarity metric is commonly employed, derived from comparative data gathered from various plots. Concurrently, the environmental distances are defined using a Euclidean distance matrix predicated on quantified geochemical variables.

The Bray-Curtis distance is:

$$d_{(u,v)} = \frac{\sum_i |u_i - v_i|}{\sum_i |u_i + v_i|}$$

And the Euclidean distance is:

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$




**Mantel test between environmental factors and palaeocommunity composition.** The strength of the correlation is represented by the partial Mantel's r statistic, with line width indicating the magnitude of the correlation and line color denoting statistical significance. Pairwise comparisons of environmental factors were also conducted, with a color gradient representing the strength of the Pearson correlation.

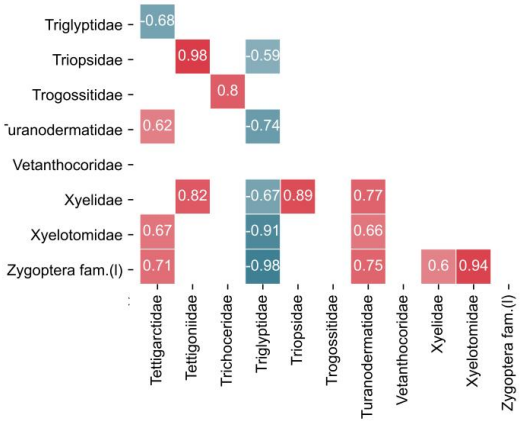
**Here, it is necessary to set the taxonomic level used for quantifying assemblage differences. A list of aquatic OTUs is required for this step. The list of aquatic OTUs taxonomic level must correspond to your research.**

```
mantel_parser = subparsers.add_parser(name='mantel', parents=[parent_parser],
help='Mantel Test between species abundance and ecological environmental variables.
(Module Ⅶ)\t[multiplot]')
mantel_parser.add_argument('--rhome', type=str, required=True, help='Absolute path of
R_HOME.(e.g. "C:\Program Files\R\R-4.3.2")')
mantel_parser.add_argument('--geochem', type=str, required=True, help='Absolute or
relative path geochemical file.(e.g. "./geochem.xlsx")')
mantel_parser.add_argument('--aquatic', type=str2list, required=True, help='Aquatic
OTUs.\t[e.g. "OTU1,OTU2,OTU3"]')
mantel_parser.add_argument('--level_aquatic', type=str, required=True, choices=['order',
'family', 'genera', 'species'], help="Taxonomic level for aquatic OTUs.")
mantel_parser.add_argument('--level_terrestrial', type=str, required=True,
choices=['order', 'family', 'genera', 'species'], help="Taxonomic level for terrestrial OTUs.")
mantel_parser.add_argument('--corr', type=str, default='pearson', choices=['pearson',
'spearman', 'kendall'], help='Correlation algorithm for geochem.(default: %(default)s)')
```

```
mantel_parser.add_argument('--output', type=str, default='./mantel', help='Absolute path or relative path and filename.(default: %(default)s)')
mantel_parser.set_defaults(func=mantel)
```

## Species correlation semi-matrix graphics (Module Ⅷ)

The module for species correlation analysis aims to uncover possible interactions among fossil species, which are indicative of symbiotic relationships. TaphonomeAnalyst 2.0 offers a variety of techniques to compute correlations and to generate semi-matrix graphics, including Pearson's coefficient, as well as Spearman's, Kendall rank correlation or SparCC.




**Semi-matrix of taphonomy correlations among organisms (a part of entire figure).** Red indicates a positive correlation, blue a negative correlation. Missing positions are due to the filtering out of data with insufficient significance levels. Users can adjust the intensity of data filtering as needed.

**Here the user is required to enter the taxonomic rank, correlation type and intensity, and P-value.**

```
corrotus_parser = subparsers.add_parser(name='corrotus', parents=[parent_parser], help='Heatmap-OTUs correlation analysis. (Module Ⅷ)\t[heatmap]')
corrotus_parser.add_argument('--level', type=str, default='family', choices=['order', 'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
corrotus_parser.add_argument('--corr', type=str, default='pearson', choices=['pearson', 'spearman', 'kendall', 'sparcc'], help='Correlation algorithm.(default: %(default)s)')
corrotus_parser.add_argument('--p_value', type=float, default=0.1, help='Maximum threshold of p-value.(default: %(default)s)')
corrotus_parser.add_argument('--output', type=str, default='./corrotus', help='Absolute path or relative path and filename.(default: %(default)s)')
corrotus_parser.set_defaults(func=corrotus)
```

Command line：

python TaphonomeAnalyst2.py corrotus --input ./Supplementary material2.xlsx --level

'family' --corr 'pearson' --p_value 0.1

## Correlational Network Visualization (Module Ⅸ)

### Ⅸ.1 Overview

A network characterized by nodes and links, is a composition of both elements. In such a network, the nodes symbolize taxa found within sampling plots, while the edges represent the taphonomic co-occurrences among these taxa. Animals inhabiting the same microenvironment also demonstrate co-occurrence during the process of fossilization. Therefore, the phenomenon of taphonomic co-occurrence can be partly explained by the extent of habitat overlap between two taxa. In constructing the network, two crucial considerations must be addressed: the methodology for correlation calculation and division of network modules.

### Ⅸ.2 Correlation

Link including various correlation methods, including the Pearson correlation coefficient, Spearman's rank correlation, Kendall rank correlation and SparCC (Sparse Correlations for Compositional data) for network visualization. It also enables users to define their own filters, including correlation strength and P-values.

The most widely employed method in microbial research is SparCC, which full name is Sparse Correlations for Compositional data, is a computational method designed to identify correlations between specifications in a community by analyzing sequential data. SparCC does not use variance as a direct measure of correlation, but instead adopts a more complex method to estimate the correlation between species in sparse data. SparCC improves the estimation of correlation between microbial abundances by log-transforming the observed abundance data and using the variance of log ratios to correct for biases. The method also employs a bootstrap procedure to calculate P-values, allowing for the assessment of the statistical significance of the correlations. However, its performance is constrained when computing interaction networks for high dimensional datasets. Typically, microbial network studies require a minimum of 3000 samples within the dataset, with a OTUs diversity not exceeding 800.

SparCC is based on the log-ratio transformation:

$$y_{ij}=\log\frac{x_i}{x_j}=\log_{x_i}-\log_{x_j}$$

$x_i$: the fraction of OTU i. $x_j$: the fraction of OTU j.

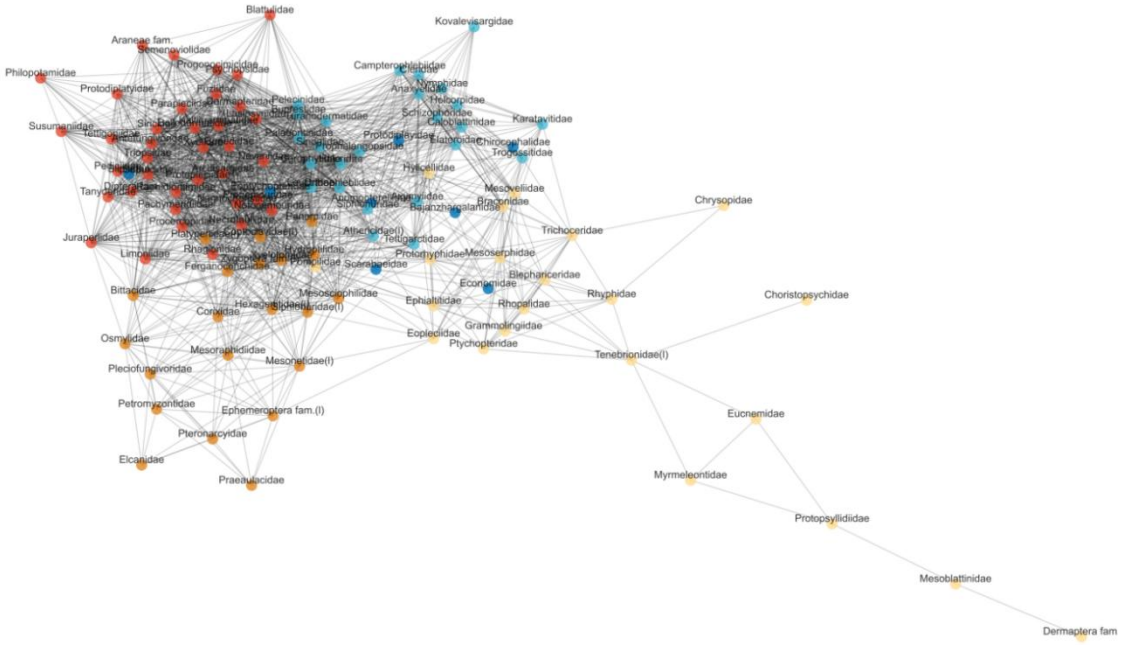
Aitchison proposed using the quantity where the variance is taken across all samples to describe the dependencies in a compositional dataset.

$$t_{ij}=\mathrm{Var}[\frac{x_i}{x_j}]=\mathrm{Var}[y_{ij}]$$

When OUT i and j are absolute correlated, their ratio is constant, therefore $t_{ij}$=0

$$t_{ij}=w_i^2+w_j^2-2\rho_{ij}w_iw_j$$

$w_i^2$, $w_j^2$: The variances of the log-transformed basis abundances OUT i and j. $\rho_{ij}$: the correlation between them OUT i and j.



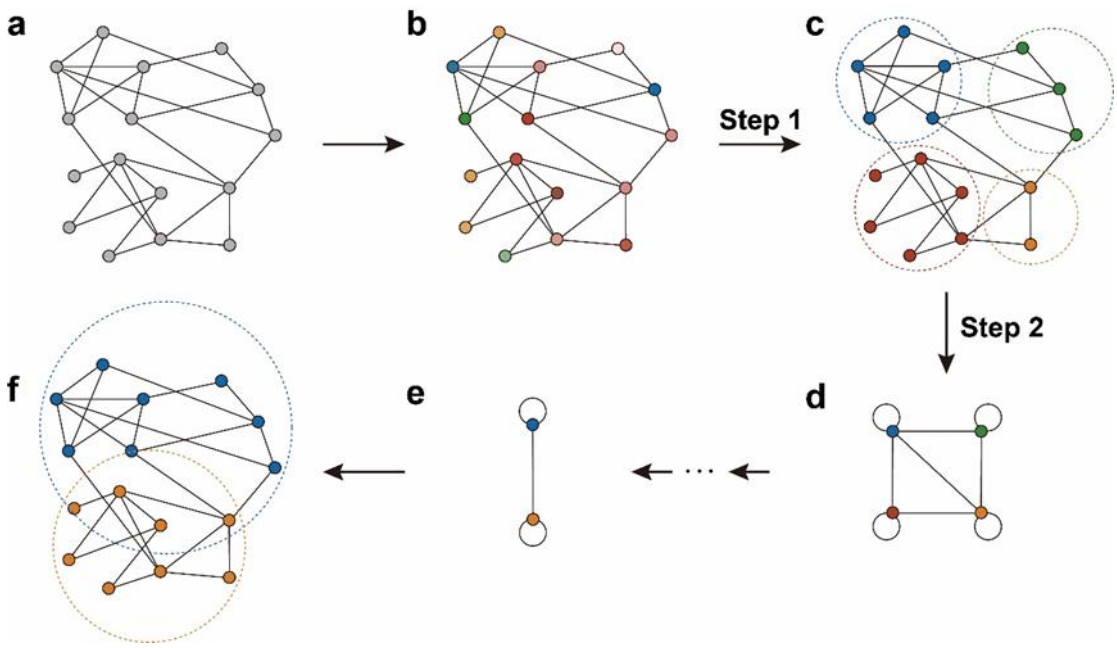

**Network visualization.** Points of the same color represent a module, indicating that these organisms likely inhabit the same environment. The output image is in PDF format, allowing users to adjust the font size in a PDF editor to meet publishing requirements.

## Ⅸ.3 Louvain algorithm

The Louvain algorithm identifies communities based on the concept of modularity. When there is a high density of connections within a module and a low density of connections between modules, the network exhibits high modularity as a result of this partitioning. The iteration process ceases automatically when there is no further increase in modularity. Modularity Q⊆[-0.5,1), and it can also be used to evaluate the effectiveness of network module division. The larger the value, the better the module division effect. When the modularity is between 0.3 and 0.7, it indicates that the clustering effect is very good. The formula for modularity is presented as follows:

$$Q = \frac{1}{2m}\sum_{\upsilon\omega}\left[A_{\upsilon\omega} - \frac{k_\upsilon k_\omega}{2m}\right]\delta(c_\upsilon, c_\omega)$$

m: indicates the total number of connections of all nodes; $\upsilon$ and $\omega$: two nodes in the network. $A_{\upsilon\omega}$=0, if $\upsilon$ and $\omega$ are not linked. $A_{\upsilon\omega}$=1, if $\upsilon$ and $\omega$ are linked. $k_\upsilon$= degree of the node $\upsilon$. $k_\omega$= degree of the node $\omega$. $\delta(c_\upsilon, c_\omega)$=0, if $\upsilon$ and $\omega$ are not linked. $\delta(c_\upsilon, c_\omega)$=1, if $\upsilon$ and $\omega$ are not linked.



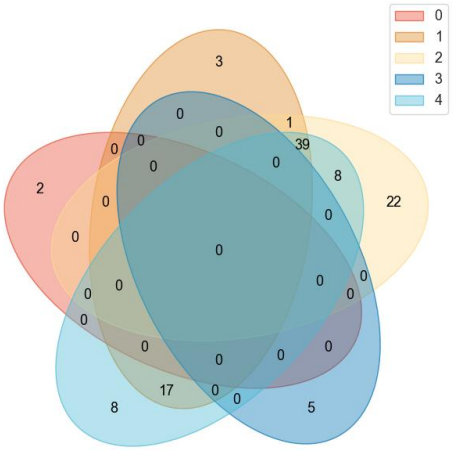

**The operation procedure of Louvain algorithm. a** The network is constructed based on correlation. **b,c** Step one: Each node in the network is assigned to its individual module. Louvain algorithm examine increase in modularity if neighboring nodes of node υ is assigned to the same module. If no increase in modularity when neighboring nodes joined, then node neighboring nodes remains in its original module. This procedure is repeatedly carried out until no further enhancement in modularity can be achieved by altering the community assignments of any nodes, at which point the first phase concludes. **c,d** Step two: After algorithm divided the modules in the first round, each module is merged into a large self-loop node. and its weight is the sum of the link weights of all nodes in the original module. The new merged nodes 's weight is the sum of the link weights of all nodes in the original module. **d,e** The subsequent calculation method is the same as the step one. **f** The final output of the resulting graph.

Louvain algorithm consists of two steps. Step one: each node in the network is assigned to its a individual module. Then, Louvain algorithm examine increase in modularity if neighboring nodes of node υ is assigned to the same module. If no increase in modularity when neighboring nodes joined (i.e., the gain in modularity is zero or negative), then node neighboring nodes remains in its original module. This procedure is repeatedly carried out until no further enhancement in modularity can be achieved by altering the community assignments of any nodes, at which point the first phase concludes. Step two: After algorithm divided the modules in the first round, each module is merged into a large self-loop node. and its weight is the sum of the link weights of all nodes in the original module. The new merged nodes 's weight is the sum of the link weights of all nodes in the original module. The subsequent calculation method is the same as the step one.

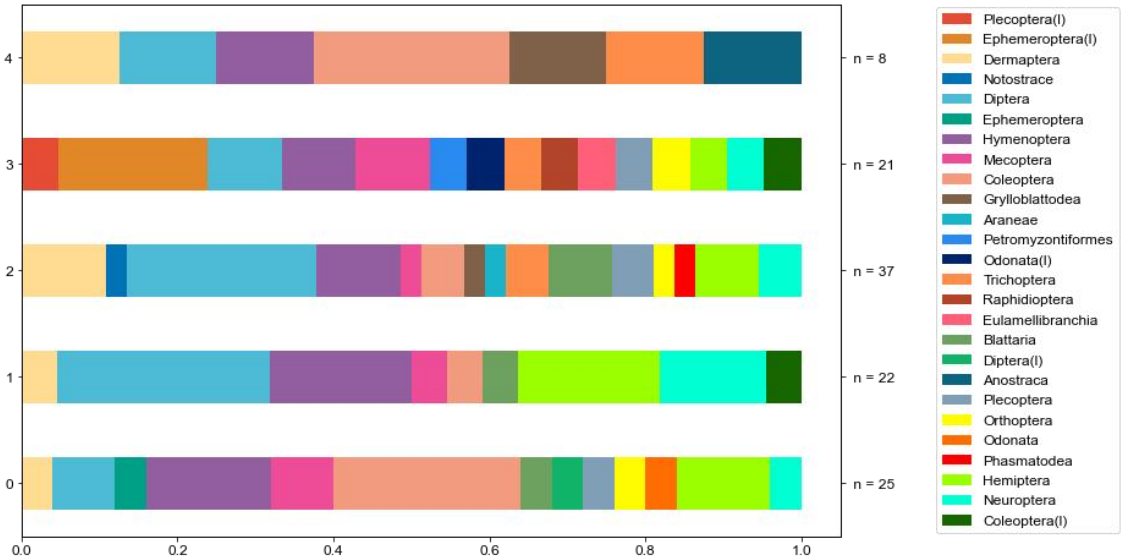
The increase in community modularity can be computed using following formula:

$$\Delta Q = \left[\frac{\sum in + k_{\upsilon,in}}{2m} - \left(\frac{\sum tot + k_{\upsilon}}{2m}\right)^2\right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m}\right)^2 - \left(\frac{k_{\upsilon}}{2m}\right)^2\right]$$

$\sum in$: the sum of the weights of all links in same module. $\sum tot$: the sum of the weights of all links that are external to the module. $k_i$: the sum of weights of node $\upsilon$. $k_{i,in}$:the sum of weights between node $\upsilon$ to the module where node $\upsilon$ are tried moved.




**Veen map of biodiversity between different modules is output with the network.** You can see the diversity and intersections of different modules in the network.




**A histogram of biodiversity for each module is output with the network.**

**Users need to input the taxonomic level, types and strength conditions of correlations, and p-value conditions.**

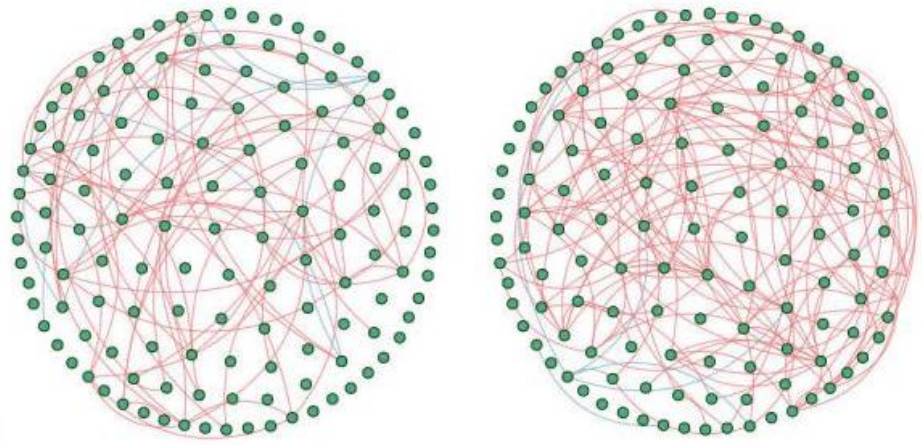
```
cooccurnet_parser = subparsers.add_parser(name='cooccurnet',
parents=[parent_parser], help='Co-occurrence networks. (Module Ⅸ)\t[network]')
cooccurnet_parser.add_argument('--level', type=str, default='family', choices=['order',
'family', 'genera', 'species'], help='Taxonomic level.(default: %(default)s)')
cooccurnet_parser.add_argument('--corr', type=str, default='pearson', choices=['pearson',
'spearman', 'kendall', 'sparcc'], help='Correlation algorithm.(default: %(default)s)')
cooccurnet_parser.add_argument('--corr_coef', type=float, default=0.7, help='Minimum
threshold of correlation coefficient.(default: %(default)s)')
cooccurnet_parser.add_argument('--p_value', type=float, default=0.1, help='Maximum
threshold of p-value.(default: %(default)s)')
cooccurnet_parser.add_argument('--output', type=str, default='./cooccurnet',
help='Absolute path or relative path and filename.(default: %(default)s)')
cooccurnet_parser.set_defaults(func=cooccurnet)
```

Command line：

python TaphonomeAnalyst2.py cooccurnet --input ./Supplementary material2.xlsx --level 'family' --corr 'pearson' --corr_coef 0.7 --p_value 0.1

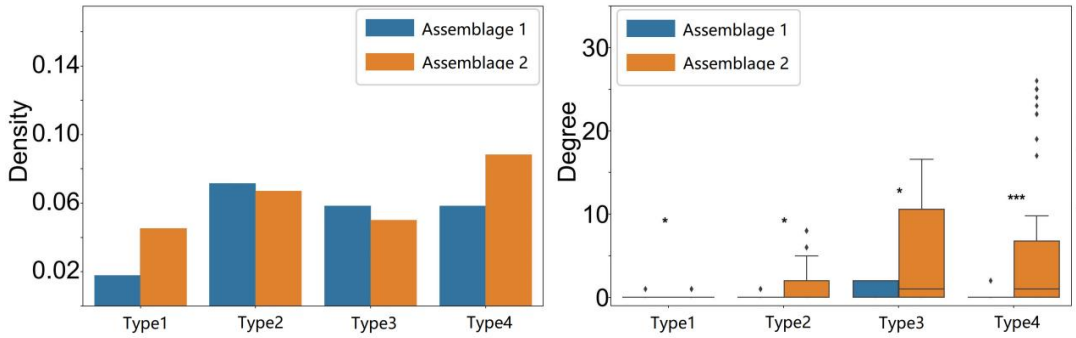
## Comparison of networks (Module Ⅹ)

This module offers the capability to compare networks under different groups of plots. The network set for comparison can be easy to read by using a consistent random layout. Additionally, it presents a bar chart that contrasts key network metrics such as total nodes, total linked nodes, total edges, density, modularity, complexity, degree and robustness.

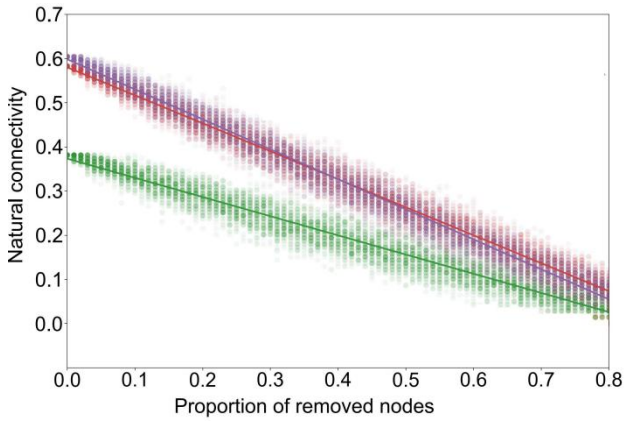


**The network set for comparison can be easy to read by using a consistent random layout.** In this image, points at the same location within the network represent the same species. It enables users to discern differences in community correlations by examining

the number of linked points and the connection density of links.




**The statistical graphs for total nodes, total linked density, modularity, total linked nodes, total edges, density, modularity, and complexity are bar charts.**




**The robustness analysis of the different assemblage.** Two colors represent two different biological assemblages. Robustness is calculated by simulating the "extinction" of organisms within a community. It involves continuously and randomly removing nodes (species) and then observing the decrease in the number of connections within the community (co-ocurrence). Communities with high robustness experience a slower reduction in links when species are removed. Additionally, this function can also output the slope of the linear decline.

**(1) Firstly, input the sample grouping; if not set, the default is the sedimentary environment clustering result.**

**(2) A list of aquatic OTUs is required for this step. The list aquatic OTUs taxonomic level must correspond your research.**

**(3) Users need to input the taxonomic level, types and strength conditions of correlations, and p-value conditions.**

```
netVC_parser = subparsers.add_parser(name='netVC', parents=[parent_parser], help='Generate a unified layout network for comparison. (Module Ⅹ)\t[network/boxplot/barplot]')
netVC_parser.add_argument('--aquatic', type=str2list, required=True, help='Aquatic OTUs.\t[e.g. "OTU1,OTU2,OTU3"]')
netVC_parser.add_argument('--level_aquatic', type=str, required=True, choices=['order', 'family',
```

```
'genera', 'species'], help="Taxonomic level for aquatic OTUs.")
netVC_parser.add_argument('--level_terrestrial', type=str, required=True, choices=['order', 'family',
'genera', 'species'], help="Taxonomic level for terrestrial OTUs.")
netVC_parser.add_argument('--groups', type=str2list, required=True, help='Environment
groups.(Grouping the plots by different aquatic and terrestrial environments)\t[e.g.
"plotA1/plotB2,plotB1/plotA2,plotC1/plotC2"]')
netVC_parser.add_argument('--corr', type=str, default='pearson', choices=['pearson', 'spearman',
'kendall', 'sparcc'], help='Correlation algorithm.(default: %(default)s)')
netVC_parser.add_argument('--corr_coef', type=float, default=0.7, help='Minimum threshold of
correlation coefficient.(default: %(default)s)')
netVC_parser.add_argument('--p_value', type=float, default=0.1, help='Maximum threshold of
p-value.(default: %(default)s)')
netVC_parser.add_argument('--output', type=str, default='./netVC', help='Absolute path or relative
path and filename.(default: %(default)s)')
netVC_parser.set_defaults(func=netVC)
```

Command line：

python TaphonomeAnalyst2.py netVC --input ./Supplementary material2.xlsx --aquatic 'Daohugounectes primitinus(I),Triglypta haifanggouensis,Triglypta haifanggouensis,Yanliaocorixa chinensis,Karataviella popovi,Samarura gigantea(I),Anisoptera fam. gen. sp1.(I),Platyperla platypoda(I),Ferganoconcha sibirica,Qiyia jurassica(I),Mesomyzon sp1.,Triops sp1.,Chirocephalidae gen. sp1.,Eurythoracalis mirabilis(I),Shantous lacustris(I),Foliomimus latus(I),Furvoneta viriosus(I),Furvoneta raucus(I),Mesobaetis sibirica(I),Clavineta eximia(I)' --level_aquatic 'species' --level_terrestrial 'family' --groups 'plot3-3/plot3-1/plot3-2,plot1-1/plot2-2/plot2-1/plot2-3/plot1-2/plot1-3' --corr 'spearman' --corr_coef 0.7 --p_value 0.01

| Evaluation indicators | Significance in netology | Significance in ecology |
|---|---|---|
| Total nodes | All nodes present in the network, both connected and unconnected. | The diversity present in the sampling plots. |
| Total linked nodes | All connected nodes present in the network. | The diversity present clear co-ocurrence with other OTUs. |
| Total edges | All links present in the network. | Taphonomic co-currence. it can also reflect the symbiotic relationship of OTUs to a certain extent. |
| Density | The density of interconnecting edges between nodes in a network. $$d(G) = \frac{2L}{N(N-1)}$$ $N$: Total nodes. $L$: Total edges. | The complexity of symbiotic relationship. |
| Complexity | The complexity of the network. $$C_i = \frac{d_i}{n}$$ $d_i$: Degree of node i. $n$: Total nodes. $C_i$: Complexity of node i. | Community complexity. |
| Degree | The centrality and importance of network. $$\deg(v) = \sum_{u \in V,\ u \neq v} A_{uv}$$ $V$: All the nodes in the network. A: A= 1 if there is an edge from node u to node v, otherwise A=0. | The importance of keystone species in communities |
| Modularity | The strength of a network divided into modules. A network with high modularity has dense connections between nodes within a module, but sparse connections between nodes in different modules. | The degree of differentiation of the community. |
| Robustness | Response of network links after nodes are randomly removed. | The ability of communities to resist environmental change |

**Evaluation indicators of network.**

# FAQ

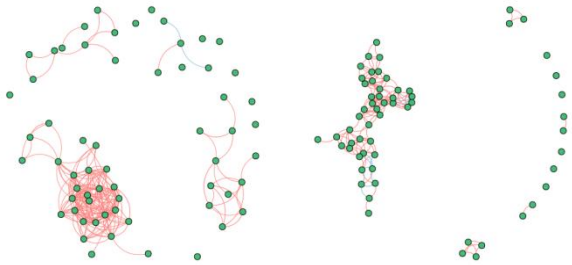
## The error of identical OTUs being identified as two separate ones.

This is often caused by incorrect spacing after or in the scientific names of OTUs.

## Proportion of taphonomic grades error

The primary reason for the inability to output images is that this module only accepts uppercase letters A through E. Careful proofreading during input is essential, and it is recommended to use Excel's statistical functions to eliminate incorrect data.

## The network comparison error



The network comparison visualization may face challenges due to an overabundance of links. This plethora of connections complicates the software's ability to identify matching layouts. We suggest implementing SparCC or imposing stricter correlation criteria to diminish the link count.

## Geochemical heat map error

In geochemical data, the concentration of certain elements may be below the detection limit, which is generally written as BDL in the main text of the paper. However, in the geochemical tables used for calculations, numerical values must be entered. We have referred to some environmental science papers and suggest using half of the detection limit as the value.

# Future developments

TaphonomeAnalyst 2.0 is an ongoing project that will remain in a state of long-time development. The software is designed to keep pace with the ever-evolving field of

geochemistry and Taphonomy, ensuring that it remains current and relevant as the discipline advances. To comprehensively grasp biological behaviors and interactions, it is essential to gather, integrate, and analyze multiple types of ecological data. Future advancements in TaphonomeAnalyst 2.0 will be geared towards a more holistic analysis that encompasses various forms of palaeoecological data, including functional morphology, types of damages, sporopollen taxa, and body sizes. The objective of TaphonomeAnalyst 2.0 is to facilitate the creation of multilayer ecological networks through the straightforward utilization of a diverse set of fossil community-level data.

## Software Development Environment

The development environment is based on Python 3.8.8 and R 4.3.2, with the following third-party libraries used in Python: matplotlib, numpy, pandas, igraph, networkx, seaborn, scipy, community, skbio, h5py, numba, dask, venn, rpy2. The implementation of the SparCC algorithm originates from the micnet library. In R, the third-party libraries used include dplyr, linkET, and ggplot2.

## Acknowledgments

The original version of TaphonomeAnalyst was published in the Communications Biology. We extend our gratitude to Luke R. Grinham, the Primary Handling Editor at Communications Biology, for his instrumental role in facilitating peer reviews of our original version from two stratigraphy experts and one microbiologist. Additionally, we express our sincere appreciation to the three initial reviewers of the original version of TaphonomeAnalyst manuscript: Roxanne Banker, Alexander Dunhill, and an anonymous reviewer for their valuable insights. This work was supported by grants from the National Natural Science Foundation of China (42288201 and 32020103006) and High-level Teachers in Beijing Municipal Universities (no. BPHR20220114).

## References

Allison, P.A., Briggs, D.E.G., 1991. Taphonomy of nonmineralized tissues in Taphonomy. Taphonomy: releasing the data locked in the fossil record. Springer. Press, pp. 25–70.

Bai, H., Kuang, H., Liu, Y., Peng, N., Chen, X., Wang, Y., 2020. Marinoan-aged red beds at Shennongjia, South China: Evidence against global-scale glaciation during the Cryogenian. Palaeogeogr. Palaeoclimatol. Palaeoecol. 559, 109967.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C., Overbeek, L.V., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M., 2020. Microbiome

definition re-visited: old concepts and new challenges. Microbiome 8, 1–22. https://doi.org/10.1186/s40168-020-00875-0

Blondel, V., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast Unfolding of Communities in Large Networks. J. Stat. Mech.-Theory. E. P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Chao, A., 1984. Non-parametric estimation of the number of classes in a population. Scand. J. Stat. 11, 265–270.

Chao, A., Lee, S.M., 1992. Estimating the number of classes via sample coverage. J. Am. Stat. Assoc. 87 (417), 210–217.

Chao, A., Shen, T. J., 2004. Nonparametric prediction in species sampling. J. Agr. Biol. Envi. St. 9, 253–269.

Chao, A., Yang, M.C. K., 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. Biometrika 80 (1), 193–201.

Chen, J., Wan, G., 1999. Sediment particle size distribution and its environmental significance in Lake Erhai, Yunnan province. Chin. J. Geochem. 18(4), 314–320.

Chen, J., Wan, G., Chen, Z., Huang, R., 1999. Chemical elements in sediments of Lake Erhai and palaeoclimate evoluton. Geochimica 28 (6), 562–570.

Dhariwal, A., Chong, J., Habib, S., King, I.L., Agellon, L.B., Xia, J., 2017. Microbiomeanalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic. Acids. Res. 45 (W1), W180–W188.

Feng, L., Chu, X., Zhang, Q., 2003. CIA (chemical index of alteration) and its applications in the Neoproterozoic clastic rocks. Earth Sci. Front. 10(4), 539–544.

Fu, J., Li, S., Xu, L., Niu, X., 2018. Paleo-sedimentary environmental restoration and its significance of Chang 7 Member of Triassic Yanchang Formation in Ordos Basin, NW China. Pet. Explor. Dev. 45(6), 998–1008.

Fürsich, F. T., Aberhan, M., 1990. Significance of time-averaging for palaeocommunity analysis. Lerhaia 23, 143-152.

Guo, S., Ma, W., Tang, Y., Chen, L., Wang, Y., Cui, Y., Liang, J., Li, L., Zhuang, J., Gu, J., Li, M., Fang, H., Lin, X., Shih, C.K., Labandeira, C.C., Ren, D., 2023. A new method for examining the co-occurrence network of fossil assemblages. Commun. Biol. 6 (1), 1102. https://doi.org/10.1038/s42003-023-05417-6

Harnois, L., 1988. The C.I.W Index: a new chemical index of weathering. Sediment. Geol. 55 (3), 319–322.

Hou, Q., Jin, Q., Niu, C., Zhang, R., Chen, F., Xu, J., Zhang, F. J., 2018. Distribution characteristics and main controlling factors of main hydrocarbon sourcerocks in Liaodong bay area. Earth Sci. 43 (6), 2160–2171.

Karr, J., Clapham, M., 2015. Taphonomic biases in the insect fossil record: Shifts in articulation over geologic time. Paleobiology 41 (1), 16–32.

McLennan, S.M., 1993. Weathering and global denudation. J. Geol. 101 (2), 295–303.

McNamara, M.E., Briggs, D.E., Orr, P.J., 2012. The controls on the preservation of structural color in fossil insects. Palaios 27 (7), 443–454.

Muscente, A.D., Bykova, N., Boag, T.H., Buatois, L.A., Mángano, M.G., Eleish, A., Prabhu, A., Pan, F., Meyer, M.B., Schiffbauer, J.D., Fox, P., Hazen, R.M., Knoll, A.H., 2019.

Ediacaran biozones identified with network analysis provide evidence for pulsed extinctions of early complex life. Nat. Commun. 10 (1), 911.

Nesbitt, H.W., Young, G.M., 1982. Early proterozoic climates and plate motions inferred from major element chemistry of lutites. Nature 299, 715–717. Palaeogeogr. Palaeoclimatol. Palaeoecol. 63 (1–3), 73–89.

Newman, M. E. J., 2006, Modularity and community structure in networks. Proc. Natl. Acad. Sci. 103 (23), 8577-8582.

Olszewski, T., 1999. Taking advantage of time-averaging. Paleobiology, 25(2), 226–238.

Smith, D.M., Moe-Hoffman, A.P., 2007. Taphonomy of Diptera in Lacustrine Environments: A Case Study from Florissant Fossil Beds, Colorado. Palaios 22 (6), 623–629.

Staff, G.M., Powell, E.N, 1998. The palaeoecological significance of diversity: the effect of time averaging and differential preservation on macroinvertebrate species richness in death assemblages.

Stanistreet, I.G., Boyle, J.F., Stollhofen, H., Deocampo, D.M., Deino, A., McHenry, L.J., Toth, N., Schick, K., Njau, J.K., 2020. Palaeosalinity and palaeoclimatic geochemical proxies (elements Ti, Mg, Al) vary with Milankovitch cyclicity (1.3 to 2.0 Ma), OGCP cores, Palaeolake Olduvai, Tanzania. Palaeogeogr. Palaeoclimatol. Palaeoecol. 546, 109656.

Swain, A., Maccracken, S.A., Fagan W.F., Labandeira, C.C., 2022. Understanding the ecology of plant-insect interactions in the fossil record through bipartite networks. Paleobiology 48 (2), 239–260.

Wang, S., Hethke, M., Wang, B., Tian, Q., Yang, Z., Jiang, B., 2019. High-resolution taphonomic and palaeoecological analyses of the Jurassic Yanliao Biota of the Daohugou area, northeastern China. Palaeogeogr. Palaeoclimatol. Palaeoecol. 530, 200–216.

Wang, Y., Xu, Z., Hang, L., Xing, T., Hang, W., 2023. Lycoptera Fossil Elemental Imaging and Paleoenvironmental Research via Laser Ionization Time-of-Flight Mass Spectrometry. At. Spectrosc. 44(2), 55–64.

Whipps, J.M., Lewis, K., Cooke, R.C., 1988. Mycoparasitism and plant disease control. Fungi in biological control systems. Manchester University Press, pp. 161–187.

Wing, S.L., Sues, H.D., Potts, R., DiMichele, W.A., Behrensmeyer, A.K., 1992. Evolutionary paleoecology. Terrestrial Ecosystems through Time: Evolutionary Paleoecology of Terrestrial Plants and Animals. The University of Chicago Press, pp. 1–13.

Wright, P., Cherns, L., Hodges, P., 2003. Missing mollusks: Field testing taphonomic loss in the Mesozoic through early large-scale aragonite dissolution. Geology 31 (3), 211–214.

Xu, H., Zhang, Y., Yuan, D., Shen, S., 2022. Quantitative palaeobiogeography of the Kungurian–Roadian brachiopod faunas in the Tethys: Implications of allometric drifting of Cimmerian blocks and opening of the Meso-Tethys Ocean. Palaeogeogr. Palaeoclimatol. Palaeoecol. 601, 111078.

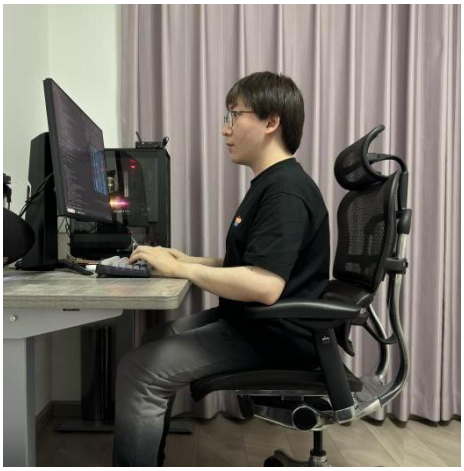
Yang, S., Hu, W., Fan, J., Deng, Y., 2022. New geochemical identification fingerprints of volcanism during the Ordovician-Silurian transition and its implications for biological and environmental evolution, Earth-Sci. Rev. 228, 104016.

Zhou, X., Sun, L., 2023. Factors controlling the formation and evolution of source rocks in the Shahezi Formation, Xujiaweizi fault depression, Songliao Basin. Energy Geoscience 4(2), 100140.

## Developer

**Shilong Guo (郭世龙)**
PhD candidate at Capital Normal University

I am engaged in the quantitative description of succession in Yanliao to Jehol community networks. During the use of the software, if you encounter any ecological issues, please contact me (13820113071@163.com).



**Wang Ma (马旺)**
Bioinformatics engineer, Freshwind Biotechnology (Tianjin) Limited Company

I am currently engaged in bioinformatics research and developed the code of TaphonomeAnalyst 2.0 during my spare time. Please contact me if you have any problems with software operation (wma19952022@163.com).