**KF-metaweb: A trophic metaweb of freshwater ecosystems of South Korea**

Sagar Adhurya, Da-Yeong Lee, Young-Seuk Park*

*Corresponding author: parkys@khu.ac.kr

Ecology and Ecological Informatics Laboratory, Department of Biology, College of Science, Kyung Hee University, Seoul 02447, Republic of Korea

Running title: Trophic metaweb of freshwater ecosystems

Submitted to Global Ecology and Biogeography on 6 September 2023 with manuscript ID GEB-2023-0495

Abstract

**Motivation:** The metaweb is a dictionary of nodes and their potential interactions developed for a particular region, focusing on a particular type of ecosystem. Based on the local biodiversity information at different spatial and temporal scales, the regional metaweb can be easily decomposed into local webs. The generated local webs are useful for understanding spatiotemporal variations in ecological interactions in a particular region. In this study, an attempt was made to develop a trophic metaweb for freshwater ecosystems in South Korea, called the KF-metaweb. The metaweb contains 23074 interactions between 446 taxa collected from 730 studies. This metaweb can be used to understand the spatiotemporal variability of different local food webs and the effects of the environment on food web properties. Furthermore, this is the first metaweb developed for any Asian ecosystem that contains information about many interactions that are unavailable in any other existing database. In addition, this metaweb study enriches our global understanding of ecological interactions.

**Main types of variables contained:** The data contained trophic interactions between resources (prey) and consumers (predators).

**Spatial location and grain:** The mainland of South Korea and Jeju Island.

**Time period and grain:** 2008–2021

**Major taxa:** Microalgae (belonging to the phyla Cyanobacteria, Bygpa, Cryophyta, Myozoa, Ochrophyta, Charophyta, Euglenozoaa, and Mycetozoa), zooplankton (belonging to the phyla Arthropoda and Rotifera), benthic macroinvertebrates (Platyhelmenthes, Annelida, Arthropoda, and Mollusca), and fish.

**Level of measurement:** Minimum taxonomic resolution was at the genus level for fish and benthic macroinvertebrates and order level for zooplankton and microalgae.
Software format: Excel (*.xlsx)

KEYWORDS: species interaction; aquatic community; river; streams; reservoirs; prey-predator interaction; food web

Introduction

Understanding ecological interactions is crucial for effective conservation planning and for gaining a comprehensive understanding of ecology (Stork, 2010). Unfortunately, the scarcity of ecological interaction data, often referred to as the “Eltonian shortfall,” has hindered research in this field due to the demand for greater effort and expertise (Delmas et al., 2019). Fortunately, as larger databases containing information on interactions, biodiversity, traits, and other essential parameters have gradually emerged (Delmas et al., 2019), the situation is improving. There is growing interest in understanding the spatiotemporal variability of ecological networks (Pilosof et al., 2017; Strydom et al., 2021). The main challenge to the spatiotemporal understanding of ecological networks is their static nature (McCann & Rooney, 2009). To resolve this issue, differential equations that demand more parameters, that is, more data, are sometimes used (Christensen & Walters, 2004; Park et al., 2008). The use of metaweb can be a good solution for mitigating this issue, with comparatively fewer data requirements.

Metaweb is a dictionary of nodes and their potential interactions developed for a particular region, focusing on particular types of ecosystems (Dunne, 2006). Based on local biodiversity information, the metaweb can be easily decomposed into local webs to understand local-level ecological interactions. Thus, it is possible to generate a large number of local webs for different localities and time periods, facilitating an understanding of spatiotemporal variations in the network (Kortsch et al., 2019; Olivier et al., 2019). Additionally, it is possible to relate network properties to environmental variables to understand the influence of climatic factors and habitat properties on network structure and functions (Albouy et al., 2019; Kortsch et al., 2019). Although this concept seems promising and easy to implement, its development in this field is very recent, especially since 2019.

In this study, a metaweb was developed for aquatic organisms in freshwater ecosystems of South Korea. Taxa in the database were gathered from the literature and monitoring datasets. The metaweb was developed through an intensive literature survey of the potential interactions between these taxa. Subsequently, the metaweb was analyzed for its structural properties and the importance of different taxa. This metaweb can be used to understand the spatiotemporal variability of different local food webs and the effects of the environment on food web properties. Furthermore, this is the first metaweb developed for any Asian ecosystem that contains information about many interactions that are
unavailable in any other existing database. This highlights the importance of the metaweb in a global context to better understand ecological interactions, which will ultimately help improve conservation planning and environmental management.

68  **Methods**

69  **Taxonomic and geographical coverage**

70  This database, named KF-metaweb, covers the interactions among different taxa (nodes) sampled from different streams, rivers, and freshwater lakes in South Korea (Figure 1). A list of taxa was compiled from biomonitoring data from the Aquatic Ecosystem Survey and Health Assessment Method: Stream/River ([https://water.nier.go.kr](https://water.nier.go.kr)), which was collected from 2008 to 2021, covering 3032 sampling locations (National Institute of Environmental Research, 2019), and the Survey on the Environment and Ecosystem of Lakes (Wonju Regional Environmental Office, 2019, 2020, 2021). Additionally, a list of taxa was obtained from the National Ecosystem Survey ([https://nie-ecobank.kr](https://nie-ecobank.kr)), and unpublished data were collected from reservoirs following national data collection guidelines (MoE, 2017).

71  **Taxonomy and systematics**

72  Owing to the unavailability of interaction data at higher taxonomic resolutions, the maximum taxonomic resolution of the metaweb was set to the genus level for benthic macroinvertebrates and fish, and in the case of zooplankton and microalgae, it was set to the order level. In some instances, the specimens were identified only up to the subfamily level (e.g., Acentropinæ), family level (e.g., Saldidae), or class level (e.g., Collembola) in the original datasets. All the nodes were classified into five taxonomic hierarchies according to the Global Biodiversity Information Facility (GBIF; [https://www.gbif.org/](https://www.gbif.org/)) backbone taxonomy: family, order, class, phylum, and kingdom. Some names were updated to match those used in GBIF and checked for other inconsistencies, such as the presence of synonyms and spelling mistakes.

73  **Development of the metaweb**

74  To develop the metaweb (KF-metaweb), geographical and ecological boundaries should be defined first. In this case, the geographical boundary was South Korea, whereas the ecological boundary was a freshwater ecosystem, including streams, rivers, and reservoirs (Step 1 in Figure 2). Thereafter, the list of nodes for which the metaweb was to be developed was compiled from the biomonitoring data, as stated earlier. Interactions for the nodes were then searched in 730 different scientific studies, gray literature, and online databases. In this case, interactions documented in the regional literature were
prioritized. Thereafter, the knowledge gap was filled with further data collection from the international literature (Step 2 in Figure 2). Subsequently, the similarity in interactions in the same taxonomic categories was checked because, in many cases, interactions are documented at higher taxonomic levels, resulting in the same set of interactions for all members belonging to that taxonomic category. In this case, these nodes are merged at a higher taxonomic level (Step 3 in Figure 2). To complete the trophic metaweb, additional nodes were added, including macrophytes, detritus, and Spongillidae. Some nodes with no trophic interactions in the metaweb were removed, for example, (1) parasitic species (Torix, Theromyzon, Hirudo, and other species) were removed for not having any trophic interactions, and (2) Nematomorpha were removed to avoid any documented trophic interactions in the metaweb. This resulted in a total of 446 final nodes in our metaweb, consisting of 49 nodes of microalgae, six nodes of zooplankton, 281 nodes of macroinvertebrates, and 107 nodes of fish, macrophytes, Spongillidae, and detritus. Following this, the list format of the initial metaweb was transformed into a wide matrix format to check for inconsistencies in interactions. As depicted in Figure 2 in Step 4, the rows and columns of the matrix represent resources and consumers, respectively. If the sum of a column is zero, the consumer has no resources, which indicates that it is a basal node (D in the matrix of Figure 2). Similarly, if the sum of a row is zero, this indicates that the resource has no consumers, which means that it is a top node (or apex predator) (F in the matrix shown in Figure 2). If the sum over both the rows and columns is zero for a particular node, it represents a node that does not have any interactions (A in the matrix of Figure 2). Here, inconsistencies result if (i) any node does not have any interactions or (ii) any intermediate or top nodes are represented as basal nodes. In these cases, the first intensive search for additional trophic interactions was conducted. If no interaction is found, then the node is merged with a taxonomically similar node; for example, in Step 5 of Figure 2, nodes A and C are merged so as not to have any interactions with node A. Finally, the matrix is transformed into a list format to represent the final version of the metaweb as a list.

Analysis of metaweb

The metaweb was analyzed at two levels: metaweb and node. The metaweb-level analysis focuses on understanding the structural properties of the metaweb. In this case, size, connectance, link density, proportion of top nodes, proportion of basal nodes, proportion of intermediate nodes, proportion of cannibals, omnivory, generality, vulnerability, and clustering coefficient are calculated with ‘multiweb’ package (Saravia et al., 2022) in R (R Core Team, 2023). Modularity is calculated with ‘igraph’ package (Csardi & Nepusz, 2006) in R with an edge-betweenness algorithm to detect the modules in the metaweb as it is a binary network (Leger et al., 2015). Nestedness was calculated with ‘unodf’ package (Cantor et al., 2017) in R. At the node level, the importance of each node in the metaweb was assessed based on centrality, measured with ‘igraph’ package (Csardi & Nepusz, 2006) in R. Further, prey-averaged trophic levels (Levine, 1980) of different nodes were calculated with the
package ‘cheddar’ (Hudson et al., 2013) in R to understand their relative trophic position in this
trophic metaweb.

Results

Metaseb database

This database is available in the latest Excel Workbook (*.xlsx) format and includes four sheets: ‘Metaweb,’ ‘Node taxonomy,’ ‘Reference’ and ‘Trophic level.’ The first sheet contains the interactions among different nodes. It has three columns, i.e., ‘resource,’ ‘consumer’ and ‘reference.’ The first and second columns contain the names of the resource and consumer nodes, respectively, and the third column contains the source-to-source interaction as a reference number. The second sheet contains the names of all the nodes, along with their taxonomy. It has eight columns: ‘Node,’ ‘Genus,’ ‘Family,’ ‘Order,’ ‘Class,’ ‘Phylum,’ ‘Kingdom,’ and ‘Category.’ In the last column, the nodes are categorized into broad groups: macroinvertebrates, fish, microalgae, sponges, zooplankton, detritus, and macrophytes. The third sheet contains two columns: a reference number and a reference. The references used in the first sheet can be found in the third sheet using reference numbers. The last sheet, ‘Trophic Level’ contains the prey-averaged trophic level of each node. It contains two columns: ‘Node’ and ‘Trophic level.’

Structural analysis of the metaweb

This metaweb has 446 nodes and 23,074 interactions, with a link density and connectance of 51.74 and 0.116, respectively. The average path length within our metaweb is 2.36, accompanied by mean and maximum trophic levels of 2.67 and 4.62, respectively. The proportions of the top, basal, intermediate, and cannibalistic nodes are 0, 0.11, 0.89, and 0.23, respectively. Notably, the degree of omnivory in metaweb is 0.58. The value of the clustering coefficient is 0.38, which indicates that the nodes in this metaweb have, on average, well-connected neighbors. The generality and vulnerability of our metaweb are 58.42 ± 1.44 (mean ± standard deviation) and 51.73 ± 0.64, respectively. The modularity of our metaweb was 0.034, which is a lower modularity value. Nestedness for consumers is 0.37 and for resources is 0.46. This indicates a slightly nested structure of the metaweb with generalist resources that share their consumers with specialist resource nodes.

Nodes

Most of the nodes are represented by macroinvertebrates (63%), followed by fish (24%), and microalgae (11%) (Figure 3). Figure 4 shows the detailed composition of the nodes belonging to different groups. Among the macroinvertebrates, most nodes belonged to the phylum Arthropoda
(80%), followed by Mollusca (12.1%). In Arthropoda, most nodes belonged to class Insecta (85.8%), followed by Malacostraca (12%). Most nodes belonging to the class Insecta fell in the order of Trichoptera (26%), Diptera (18%), Plecoptera (13%), Hemiptera (12%), and Coleoptera (12%). Most fish belonged to the orders Cypriniformes (35%) and Perciformes (32%). Classes Bacillariophyceae (42%) and Chlorophyceae (11%) constituted a major portion of microalgal nodes. Microalgae, macrophytes, and detritus are the basal nodes of this metaweb.

**Interactions**

Most of the interactions in this metaweb originated from macroinvertebrates (81.48%), that is, where macroinvertebrates are resources (Figure 3), whereas microalgae and fish contribute 10.03% and 5.03% of the total interactions, respectively. Most of the flows originating from macroinvertebrates go to fish (52.01%), whereas the remaining portion (39.08%) is consumed by macroinvertebrates. The majority of the fish resources are consumed by fish (86.83%), followed by macroinvertebrates (13.07%). In the microalga, 73.42% of the flow ended in macroinvertebrates, followed by fish (23.21%) and zooplankton (4.21%). 78.12% and 19.53% of the detrital flows end up in macroinvertebrates and fishes, respectively. Macrophytes act as resources mostly for the macroinvertebrates (60.58%) followed by fishes (38.46%). Flows originating from zooplankton mostly end up in fish (51.17%), followed by macroinvertebrates (44.63%) and zooplankton (4.2%). It is important to note that the importance of detritus and macrophytes is largely underrepresented because their communities are not detailed.

**Node importance**

Four measures of centrality are used to identify the important nodes in the metaweb: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. Degree centrality (Freeman, 1977), a local measure of the number of interactions between a species. This can be further divided into in-degree and out-degree in the case of a directed network, which indicate the number of flows a node receives and contributes, respectively. The in-degree value was highest in fish nodes, with Anguilla (348), Opsarichthys (294), and Oreochromis (286) having the highest values. This indicates a relatively high position of fish at this trophic level. Among macroinvertebrates, Tabanidae (277), Dytiscidae (272), and Anotogaster (268) have the highest in-degree values, although their overall rankings were sixth, ninth, and twelfth, respectively. Regarding the out-degree, detritus (257), Chironomidae (170), and Naviculales (140) have the highest values, indicating their importance as resources in the metaweb.

Closeness centrality (Freeman, 1978; Freeman et al., 1979) is a global measure that indicates the proximity of a species to all other species in a network in terms of the shortest distance between nodes.
A directed web can be further divided into in- and out-, where the former and latter indicate the paths to and from a node, respectively. Approximately 63 nodes have the maximum values of in-closeness centrality, indicating their relative importance as consumers in the metaweb. Among these, one node is fish (Acanthorhodeus), one node is Spongillidae, and all other nodes are macroinvertebrates (dominated by Trichoptera, Gastropoda, and Plecoptera). In contrast, Chironomidae, Copepoda, Diplostraca, Gammaridae, and Baetidae have the highest out-closeness centrality, indicating their relative importance as resources.

Betweenness centrality (Freeman, 1977) describes the number of times a species is between a pair of other species; that is, how many paths pass through it. All the top ten nodes with higher betweenness centrality values belong to invertebrates, with Diplonychus, Chironomidae, Barbronia, Acetes and Macrostromum having the highest values. This indicates the importance of these nodes as intermediate nodes.

In the case of eigenvector centrality (Bonacich, 1987), the node that simultaneously influence all its partners obtains the highest score. In this case, Opsariichthys is the most central node, followed by Anguilla and Pseudorasbora. Except Diplonychus, all other nodes in the top ten nodes having highest eigenvector centrality are fish.

The relative role of each node in the meta-food web is indicated by the trophic level in the fourth sheet of the attached dataset. Dineutus (4.62) has the highest trophic position, followed by Diplonychus (4.48), Ilyocoris (4.45), Misgurnus (4.34), and Lethocerus (4.27).

Usage notes

This database contains information not available in other existing databases, including global databases such as GloBI (Poelen et al., 2014). Furthermore, this is the first metaweb from Asia that indicates its potential to fill critical gaps in understanding ecological interactions. Data are provided in Excel workbook format (*.xlsx), which can be easily converted into other formats such as *.csv and used with any software or programming language. This metaweb can be used to infer local webs for the region in which it is developed, such as the global marine metaweb (Albouy et al., 2019), the German Blight metaweb(Olivier et al., 2019), European Tetrapod metaweb (Braga et al., 2019), and the Swiss Blue and Green metaweb (Ho et al., 2022). From the inferred local webs, it is possible to understand how the network properties vary spatiotemporally. Furthermore, variations in spatiotemporal properties with environmental variables can be used to understand the influence of the environment on the structure of ecological networks. Further, it is important to note that many taxa are
at the high taxonomic level like order or family in the database. Therefore, more focus will be to increase taxonomic resolution of the database in the later versions.

Data availability statement

The database (KF-metaweb) is accessible from Figshare (Adhurya et al., 2023).

References


MoE (2017) Environmental survey guidelines for lakes and reservoirs (호소환경 조사지침), Ministry


R Core Team (2023) R: A Language and Environment for Statistical Computing.


Data availability statement

The database (KF-metaweb) is accessible from Figshare (Adhurya et al., 2023).
Figure legend

Figure 1: Location of the biomonitoring survey sites across South Korea for different reservoirs (indigo) and streams/rivers (red) [at right].

Figure 2: Conceptual diagram for the development of the metaweb. The figure is elaborated in the subsection ‘Development of the metaweb’.

Figure 3: Summary of the metaweb representing flows between different categories of organisms. The percentage value outside the circle indicates the total flow contributed by the nodes belonging to a particular category, whereas the percentage value inside the circle represents amount of flow going from a particular category to other categories. Here, only the values greater than 1% are shown.

Figure 4: Composition of different node groups. From top-left, phylum level composition of macroinvertebrates; order level composition of insects (top-right); order level composition of fishes (bottom-left); and class level composition of microalgae.
Figure 1
Figure 2
Figure 3
Figure 4