

## Development of a link extrapolation-based food web model adapted to Korean stream ecosystems

Minyoung Lee, Yongeun Kim<sup>1</sup> and Kijong Cho<sup>2,\*</sup>

Department of Biological Sciences, Ulsan National Institutes of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

<sup>1</sup>Ojeong Resilience Institute, Korea University, Seoul 02841, Republic of Korea

<sup>2</sup>Department of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea

### Contribution to Environmental Biology

- Our findings suggest the potential of a link-extrapolation method for predicting food web structure, even in data-poor environments.
- The developed model can support food web-level understanding in domestic stream ecosystem biomonitoring.
- It can be effectively used by integrating with the National Aquatic Ecological Monitoring Program.

### \*Corresponding author

Kijong Cho

Tel. 02-3290-3064

E-mail. [kjcho@korea.ac.kr](mailto:kjcho@korea.ac.kr)

**Received:** 29 May 2024

**Revised:** 17 June 2024

**Revision accepted:** 20 June 2024

**Abstract:** Food webs have received global attention as next-generation biomonitoring tools; however, it remains challenging because revealing trophic links between species is costly and laborious. Although a link-extrapolation method utilizing published trophic link data can address this difficulty, it has limitations when applied to construct food webs in domestic streams due to the lack of information on endemic species in global literature. Therefore, this study aimed to develop a link extrapolation-based food web model adapted to Korean stream ecosystems. We considered taxonomic similarity of predation and dominance of generalists in aquatic ecosystems, designing taxonomically higher-level matching methods: family matching for all fish (Family), endemic fish (Family-E), endemic fish playing the role of consumers (Family-EC), and resources (Family-ER). By adding the commonly used genus matching method (Genus) to these four matching methods, a total of five matching methods were used to construct 103 domestic food webs. Predictive power of both individual links and food web indices were evaluated by comparing constructed food webs with corresponding empirical food webs. Results showed that, in both evaluations, proposed methods tended to perform better than Genus in a data-poor environment. In particular, Family-E and Family-EC were the most effective matching methods. Our model addressed domestic data scarcity problems when using a link-extrapolation method. It offers opportunities to understand stream ecosystem food webs and may provide novel insights into biomonitoring.

**Keywords:** food web construction, trophic interactions, stream ecosystems, ecological monitoring, National Aquatic Ecological Monitoring Program (NAEMP)

## 1. INTRODUCTION

Understanding food web structure, that is, how species and their trophic links are organized, has been of

longstanding interest in ecology (Cohen *et al.* 2003). It underpins diverse ecosystem mechanisms such as energy and material transfer, primary production, and decomposition, providing system-level insights into

ecosystem functions, resilience, and sustainability (Baird and Ulanowicz 1989; Bellwood *et al.* 2003). In recent years, its importance in applied ecology has been increasing owing to global changes, such as species invasion, pollution, and climate change (Heleno *et al.* 2009; Bartley *et al.* 2019). In particular, Layer *et al.* (2011) emphasized the necessity of biomonitoring. They found that the acidified Broadstone stream (England) was recovered by tracking the food web structure and dynamics, whereas the recovery could not be discovered by a conventional approach focused on the changes in species richness and abundance for aquatic assemblages, ignoring their trophic interactions. Many studies have suggested the potential of food web indices as ecological indicators for detecting environmental changes, stresses, and their impacts, although standardized protocols have not yet been established (Tylianakis *et al.* 2007; Fath *et al.* 2019; Mestre *et al.* 2022). Despite this, food web research is still lacking. This is because the construction of food webs is fundamentally difficult. It is usually conducted through empirical investigations of trophic interactions, such as direct observations, gut content analysis, or stable isotope analysis (Yoshii *et al.* 1999; Grey *et al.* 2002). However, these works are costly, labor-intensive, and require long-term investigation, some of which take years or decades. Moreover, data uniformity is difficult to achieve because it depends on sampling efforts and investigators (Goldwasser and Roughgarden 1997).

Gray *et al.* (2015) developed an automated method for constructing a food web using a link-extrapolation approach. This method estimates the possible trophic links within a food web based on published data, thereby greatly reducing the time and cost of food web construction. The link extrapolation approach is based on the premise that if there is observed data on the interaction between two species, it is valid wherever or whenever two species are found simultaneously. Although, based on this premise, this method cannot reflect the dynamics of a trophic interaction, such as its spatial and temporal dynamics or prey switching of predators, it can effectively describe a summary web that is often spatially or temporally aggregated and provides a broad understanding of the ecosystem trophic structure. Moreover, it has a high performance for predicting food web structure; for instance, it showed higher performance for predicting individual links than previous deterministic food web models including 'Difference,' 'Ratio,'

'Difference/Ratio' (Allesina 2011), and allometric diet breadth model (Petchey *et al.* 2008) and for food web indices in empirical food webs (Gray *et al.* 2015). This approach has been used to compensate for missing links in many empirical studies but it depends on manual work (Goldwasser and Roughgarden 1993; Layer *et al.* 2013; Poisot *et al.* 2016). The development of automated methods allows the construction and analysis of vast amounts of food webs (Heberling *et al.* 2021).

However, its utilization remains challenging because of its data dependency. That is, the accuracy of the constructed food webs depends on the quantity and quality of published trophic interaction data and the researchers' data. For instance, one species observed in a survey cannot have any links unless its interaction has been reported, although it interacts with other species. This ultimately lowers the reliability of constructed food webs. Ecologists who are willing to use this method may face these kinds of issues, although this could be improved through the continued accumulation of data. Gray *et al.* (2015) considered a simple rule based on known ecological knowledge that taxonomically similar species consume similar prey. They showed that genus matching could have a higher performance than species and family matching, which led to underestimated and overestimated links, respectively. Further ecological knowledge is expected to facilitate its utilization and is required to overcome this inherent issue.

In South Korea, the government has intensively conducted stream ecosystem biomonitoring through the National Aquatic Ecological Monitoring Program (NAEMP) since 2008. It has been implemented with streams and tributaries across the whole country, covering 3,035 sites within all five river basins (i.e., the Hangang, Nakdonggang, Youngsangang, Geumgang, and Seomjingang Rivers) (MOE 2018). Targeted aquatic species belonging to the three fauna occupying different positions within a food web (i.e., fish, benthic macroinvertebrates, and epilithic diatoms) were investigated following the NAEMP biological criteria and assessment methods developed for the domestic context, referring to the methods of the US Environmental Protection Agency and European Environment Agency (Kristensen and Bøgestrand 1996; Barbour *et al.* 1999; USEPA 2002).

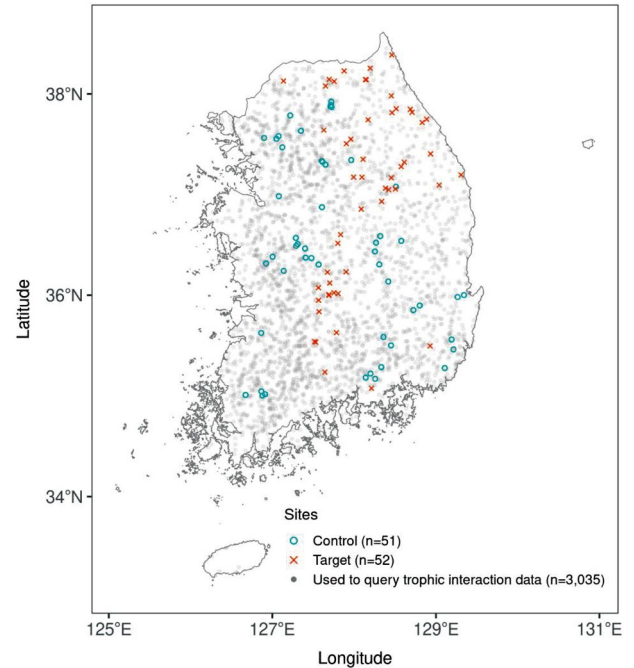
This study aims to develop a link extrapolation-based food web model adapted to South Korean stream ecosystems. However, there is a data dependency problem

caused by the high proportion of endemic species among the species surveyed by the NAEMP. The NAEMP focuses mainly on surveying endemic species that inhabit only certain areas of the world and their interactions are geographically limited. Thus, their trophic interaction data are unlikely to be found unless domestic researchers have reported them; subsequently, their trophic links are not expected to be estimated through a link extrapolation approach because of the lack of data. We regarded this as a data-poor environment. Specifically, the proportion of endemic species was used to define the data environment ( $>0.75$  for data-poor and  $<0.25$  for data non-poor; the target and control groups in this study, respectively). In a data-poor environment, using a taxonomically higher-level matching method can be more effective than using lower-level matching methods (e.g., species or genus matching, as mentioned above), considering that most predators in aquatic systems are generalists who do not consume their diet phylogenetically. Therefore, we designed food web models based on family matching for 1) all fish species, 2) endemic fish species, 3) endemic fish species that play the role of consumers, and 4) endemic fish species that play the role of resources. Food webs were constructed using these four methods and the Genus method of Gray *et al.* (2015) was used for comparison. Model performance was evaluated by comparing the constructed food web from each method with the corresponding empirical food web constructed based on gut content data from other studies. The predictive power of both individual links and food web indices was evaluated.

## 2. MATERIALS AND METHODS

### 2.1. Study site selection and data acquisition

We first collected NAEMP data for all sites (3,035 sites) over 11 years (2008–2018). The NAEMP survey was conducted twice a year in spring (April–May) and autumn (September–October) and almost all taxa were taxonomically resolved at the species level. We obtained survey data for both periods for the three fauna: fish, benthic macroinvertebrates, and epilithic diatoms (the data were retrieved from the website of the Water Environment Information System, <http://www.water.nier.go.kr/>). The obtained data were spatially and temporally aggregated by site name and divided into target



**Fig. 1.** A total of 103 study sites (51 control and 52 target sites).

and control groups based on the proportion of endemic species in the species surveyed at each site ( $>0.75$  and  $<0.25$ , respectively; these proportions were not based on all fish but on fish for which gut content data were available). Gut content data was collected through a literature search using the scientific names and Korean names of all fish species observed at all NAEMP survey sites. We then filtered out the sites with fewer than four fish species for which gut content data were available. This is because the gut content data are necessary to construct an empirical food web and evaluate model performance. Too few fish with embedded gut content data (e.g., only one species per site) may introduce bias in the results. A total of 103 sites (51 targets and 52 controls) representing data-poor and -non-poor environments respectively, were selected as the study sites (Fig. 1).

### 2.2. Food web construction and model development

Food webs were constructed by extrapolating trophic links between species based on trophic interaction data from published literature and biomonitoring data from the NAEMP. Trophic interaction data were collected

using Global Biological Interaction (GloBI) which is a web-based search engine for species interaction data (Poelen *et al.* 2014). We resolved all the surveyed taxa names based on the Global Biodiversity Information Facility (GBIF) backbone taxonomy using the Global Names Resolver, which was conducted using the R package “taxize” (Chamberlain and Szöcs 2013). Then, we queried the predation data (i.e., both “preys on” and “preyed upon by” were used as search options) using the R package “rglobi” (Poelen *et al.* 2017) using the GBIF key identifier of the surveyed taxa. A direct query using taxon names may cause querying errors, such as incorrectly queried data that can result from the existence of the same species names with different upstream taxonomic hierarchies; however, using the identifier can prevent these errors. In this step, we utilized data from all NAEMP sites (3,035 sites) rather than those from our study sites (103 sites) (Fig. 1). This is because the taxonomic matching methods we considered were not limited to accurate species matching, so that can increase data quantity and the probability of link generation. For instance, query data using taxa observed at other sites could be effectively used to generate trophic links for taxa at our study sites through higher-level taxonomic matching. Finally, a registry containing 13,824 trophic interaction data points was used to construct a food web.

Our assumption that taxonomically higher-level matching can perform better than genus matching in a data-poor environment was contradictory to the results of Gray *et al.* (2015). Therefore, before model development, we conducted a preliminary test to identify the feasibility of our assumption. We confirmed the data loss rate during the process of food web construction by comparing the proportion of isolated nodes (i.e., species without links) between two existing methods: genus and family matching. The results showed a higher proportion of isolated nodes using the genus-matching method, especially at the target sites ( $> 87\%$ ), whereas the proportions were low ( $< 6\%$ ) using the family-matching method (Table S1). Higher proportions of isolated nodes indicate a higher rate of node and link data loss, lowering the confidence of the constructed food web, supporting the feasibility of the assumption in this study.

We designed four-link extrapolation methods: taxonomic matching at the family level for fish species (Family), endemic fish species (Family-E), endemic fish spe-

cies that act as consumers (Family-EC), and endemic fish species that act as resources (Family-ER). Food webs were constructed using five matching methods, including four matching methods (Family, Family-E, Family-EC, and Family-ER) and the genus matching method of Gray *et al.* (2015) (Genus) for comparison. Specifically, Family generates links between fish and their prey and between consumers and their fish prey when a trophic link registry includes the data that any fish belonging to the same family eats/is eaten by something. Family-E generates links between endemic fish and their prey and between consumers and their endemic fish prey when a trophic link registry includes data that any fish belonging to the same family eats or is eaten by something. Family-EC generates links between endemic fish and their prey when a trophic link registry includes data indicating that any fish belonging to the same family eats something. Family-ER generates links between consumers and their endemic fish prey when a trophic link registry includes data showing that any fish belonging to the same family is eaten by something.

The constructed food webs were in the form of a summary web, with a temporal range of 2008–2018 for each site. Additionally, counterpart empirical food webs for each site were constructed based on the gut content data examined in other studies. Gut data for the 26 fish species were collected and the data sources are presented in Table S2. If gut content data were not available, the species were excluded from the study. Note that the designed methods focused on fish taxa, whereas the matching method for other taxa was fixed at genus matching. This is because there were insufficient data on the gut contents of benthic macroinvertebrates observed at the study sites, which could reduce the confidence of the model performance evaluation (Williams and Martinez 2008; Petchey *et al.* 2011). Subsequently, we considered only all the generated links concerning fish in the model performance evaluation described later.

### 2.3. Evaluation of model performance

The model performance was evaluated by comparing food web structures between a constructed food web and its counterpart empirical food web. This was done in two ways: one was to evaluate the predictive power for individual links and the other was to evaluate the predictive power for food web indices. Food web indices are commonly used to compare food webs and provide

a summarized explanation of ecosystem trophic structures (Thompson *et al.* 2012; Delmas *et al.* 2019). Their prediction is considered highly important, especially in typical probabilistic food-web modeling studies (Williams and Martinez 2008). We developed a deterministic model and focused more on how to correctly predict individual links; however, the prediction of food web indices was also considered in terms of a broad performance test.

First, the performance evaluation for predicting individual links was done by using True Positives Rate (TPR) and True Skill Statistics (TSS) (Allouche *et al.* 2006; Gray *et al.* 2015) which are calculated by:

$$TPR = \frac{a}{a+c} \quad (\text{Eq. 1})$$

$$TSS = \frac{ad-bc}{(a+c)(b+d)} \quad (\text{Eq. 2})$$

where  $a$  indicates the number of links that are correctly generated,  $b$  is the number of links that are generated but not observed,  $c$  is the number of links that are observed but not generated, and  $d$  is the number of links that are neither generated nor observed. TPR (Eq. 1) indicates the ratio of correctly predicted to generated links and ranges from 0 to 1. TSS (Eq. 2) ranges from  $-1$  (indicating completely inverse link generation) to 1 (identical link structure to the empirical web). For both TPR and TSS, the larger the value (i.e., the closer to 1), the better the model's performance. The Kruskal-Wallis test was conducted to identify if there was a difference among the methods and a post hoc analysis using the paired Wilcoxon signed-rank test was subsequently conducted. In addition, predation matrices were mapped to graphically identify model performance for individual links using two example sites from the control and target groups.

Second, the performance evaluation for food web indices was performed using two indices, generality (Gen) and vulnerability (Vul), which indicate the average prey numbers per predator and average predator numbers per prey, respectively (Bersier *et al.* 2002). Since we only considered fish gut contents in model development and food web construction, Gen and Vul were calculated using the average prey numbers per fish and average fish consumer numbers per prey, respectively. For each index, whether the difference in values between the constructed and the counterpart empirical food

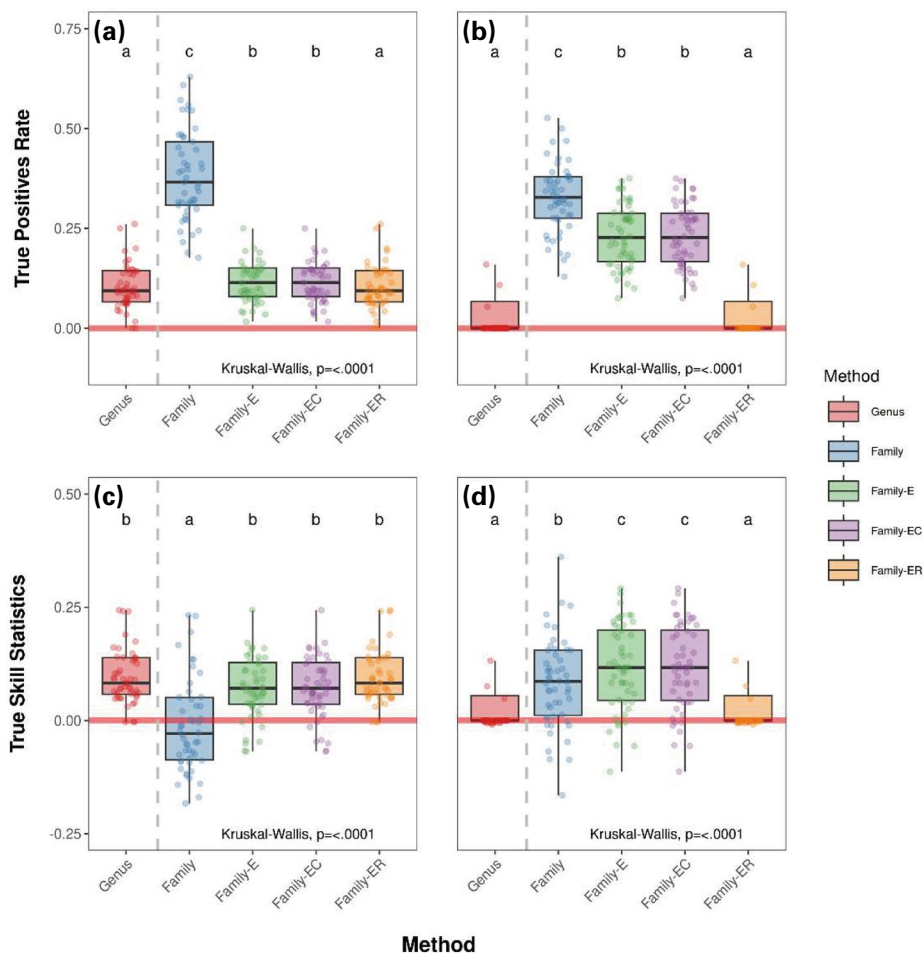
web was equal to zero was analyzed using a one-sample  $t$ -test. Values closer to zero indicate better model performance (i.e., zero is a perfect prediction of the index). Significance indicates that there is a difference in the food web structure between the constructed and empirical food webs. All procedures from data collection to evaluation of model performance and statistical analysis were performed using R version 4.1.0 (R Core Team 2021).

### 3. RESULTS AND DISCUSSION

#### 3.1. Model performance in predicting food web individual links

The TPR results showed that Family had the highest performance in both the control and target groups (Fig. 2a, b; Table S3). In the target group, Family-E and Family-EC showed the next highest performance with no significant difference between them (Fig. 2b). Meanwhile, the evaluation using the TSS showed the opposite tendency (Fig. 2c, d). The TSS in the control group was the lowest in Family with a negative geometric mean value (Fig. 2c). This resulted from overestimated (not observed but generated) links between fish and benthic macroinvertebrates (Fig. 3). In the target group, Family-E and Family-EC showed the highest performance, followed by Family and Genus (Family-ER was not different from Genus; Fig. 2d). In particular, they suppressed overestimated and underestimated links relatively well compared to Family and Genus, respectively (indicated by decreased yellow and orange grids, respectively; Fig. 4). Overall, these results indicate that using taxonomically higher-level matching methods in a data-poor environment increases the predictive power of individual links, whereas the opposite is true in a data non-poor environment.

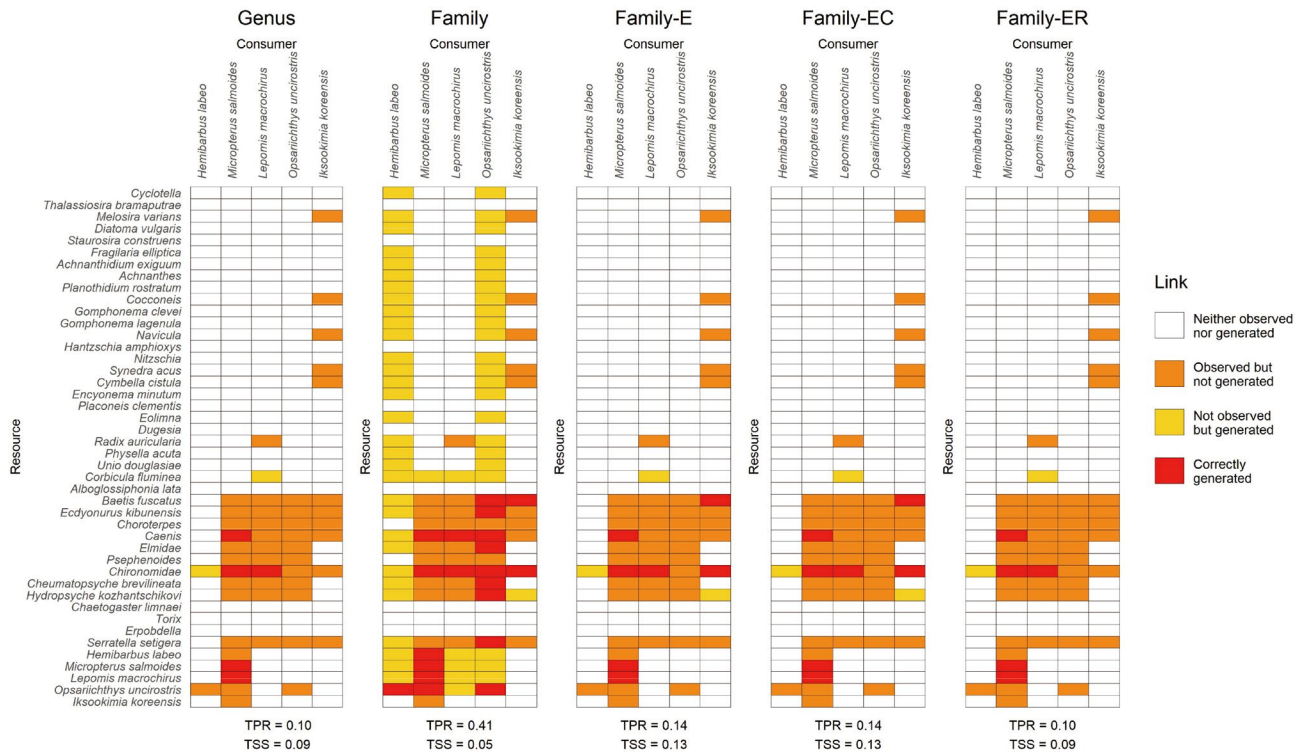
In general, the greater the phylogenetic distance between species, the lower the ecological similarity of feeding characteristics, habitats, and food preferences (Layer *et al.* 2010; Eklöf *et al.* 2012). In freshwater ecosystems, fish consumers tend to be generalists and omnivores (Guenther and Spacie 2006; Clavel *et al.* 2011; Filgueira *et al.* 2016). In insectivorous fish, predation is often determined by the habitat and functional feeding groups (Reis *et al.* 2020; Wang *et al.* 2020). Fish living in surface water feed on small terrestrial insects that



**Fig. 2.** Boxplots for evaluating model performance using True Positives Rate (TPR) and True Skill Statistics (TSS). The Kruskal-Wallis test was conducted to test if there was a difference in performance between methods (Genus, taxonomic matching at the genus level; Family, taxonomic matching at the family level for fish species; Family-E, taxonomic matching at the family level for endemic fish species; Family-EC, taxonomic matching at the family level for endemic fish species who play a role of consumers; Family-ER, taxonomic matching at the family level for endemic fish species who play a role of resources). Post hoc analysis was conducted using a paired Wilcoxon signed-rank test. Results are presented by alphabets (a, b, and c) above boxes. The first column (panels of (a) and (c)) and second column (panels of (b) and (d)) show results of the control and target group, respectively. The red horizontal line indicates TPR or TSS = 0. The grayish dashed line indicates division between Gray's method and proposed methods in this study. Boxes show 25th, 50th, and 75th percentiles and jitters indicate data points for each constructed food web.

they encounter in surface water, whereas benthic fish feed on aquatic insects living in benthic areas. In other words, they do not differentiate their prey phylogenetically. For instance, benthic fish indiscriminately eat benthic prey belonging to the Baetidae family, such as *Baetis* sp., *Baetiella* sp., and *Acentrella* sp. The fish feeding on epilithic diatoms were similar. The organisms consumed by fish are mostly fed together during the feeding process of benthic organisms (Son and Byeon 2001). If the data registry for link extrapolation is sufficient, using a lower-level matching method for fish

would not be problematic. However, if not, as in this study containing many endemic species, it probably causes many links to be observed but not generated, thereby reducing model performance. Considering the feeding characteristics of fish, taxonomically higher-level matching methods may be better choices in data-poor environments. In particular, we suggest that Family-E and Family-EC are the most effective methods, according to the results of TSS, which reflect the performance for unobserved as well as observed links, the predictions of which are important for understanding



**Fig. 3.** Predation matrices for Jeongeupcheon stream, an example site in the control group. Matrices for food webs constructed by five methods (Genus, taxonomic matching at the genus level; Family, taxonomic matching at the family level for fish species; Family-E, taxonomic matching at the family level for endemic fish species; Family-EC, taxonomic matching at the family level for endemic fish species who play a role of consumers; Family-ER, taxonomic matching at the family level for endemic fish species who play a role of resources) were mapped. Columns and rows of a matrix indicate fish consumers and their resources, respectively. Characteristics of the matrix element are indicated by color (white, links that are neither observed nor generated; orange, links that are observed but not generated; yellow, links that are not observed but generated; red, links that are correctly generated). Values of TPR (True Positive Rate) and TSS (True Skill Statistic) are presented below the matrices.

the food web structure.

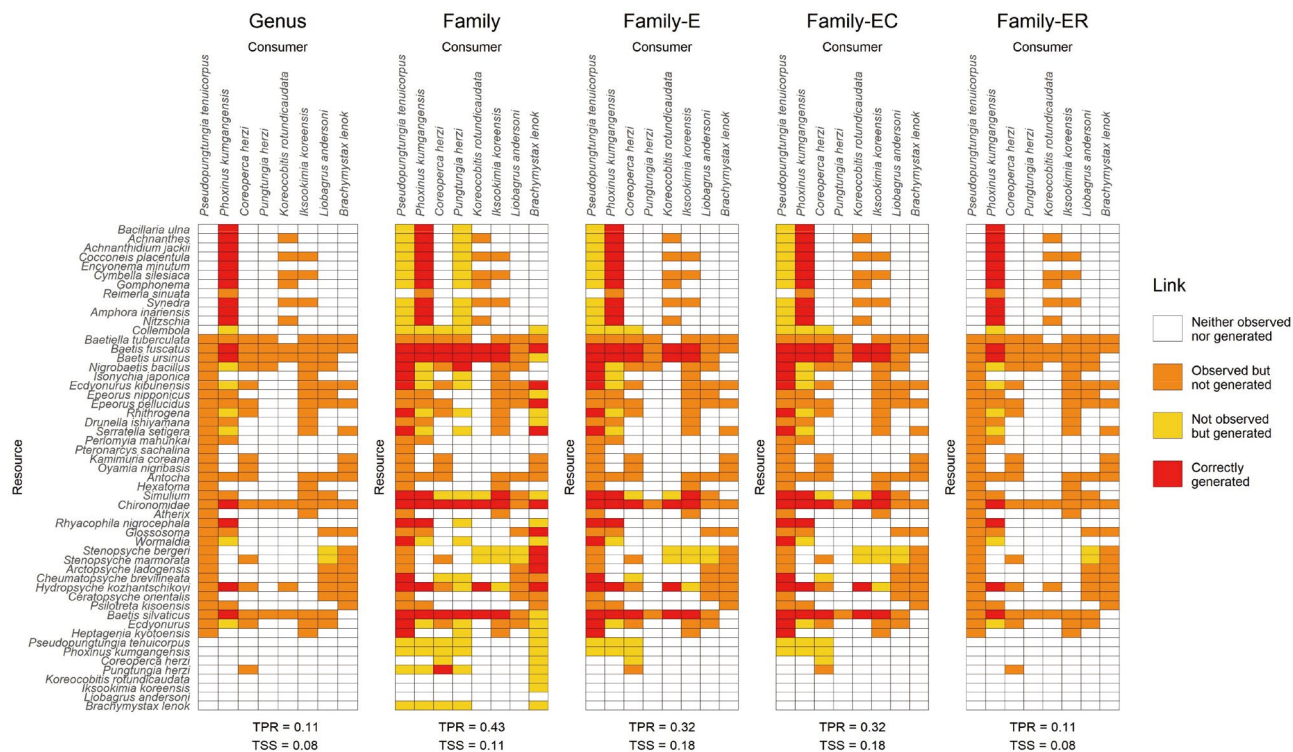
### 3.2. Model performance in predicting food web indices

For both Gen and Vul, the differences between the constructed and empirical food webs differed significantly from zero for all five methods ( $p < 0.0001$ ; Fig. 5). The mean and median values were positive and negative for Family and other methods, respectively (Fig. 5, Table S4). Overall, the results showed similar tendencies regardless of the two indices or groups. However, it is important to compare the results between the target and control groups. In particular, the values in Family in the control group (Fig. 5a, c) were the farthest from zero, whereas those in the target group were the nearest to zero (Fig. 5b, d). This indicates the effectiveness of Family in data-poor environments. This was similar to

the performance evaluation for individual links in that both showed that higher-level matching methods can achieve higher performance than Genus. In addition, family-level matching methods may overestimate or underestimate the generalism of consumers or the vulnerability of resources from consumers. Thus, care must be taken regarding which method to use, and this can be done appropriately depending on the purpose of the research.

### 3.3. Limitations, ecological implications, and conclusion

In food web modeling, it is critical to filter forbidden links generated from a model or generate additional links that are not generated from a model but are observed or expected to exist (Morales-Castilla *et al.* 2015; Terry and Lewis 2020). This study suggests that given



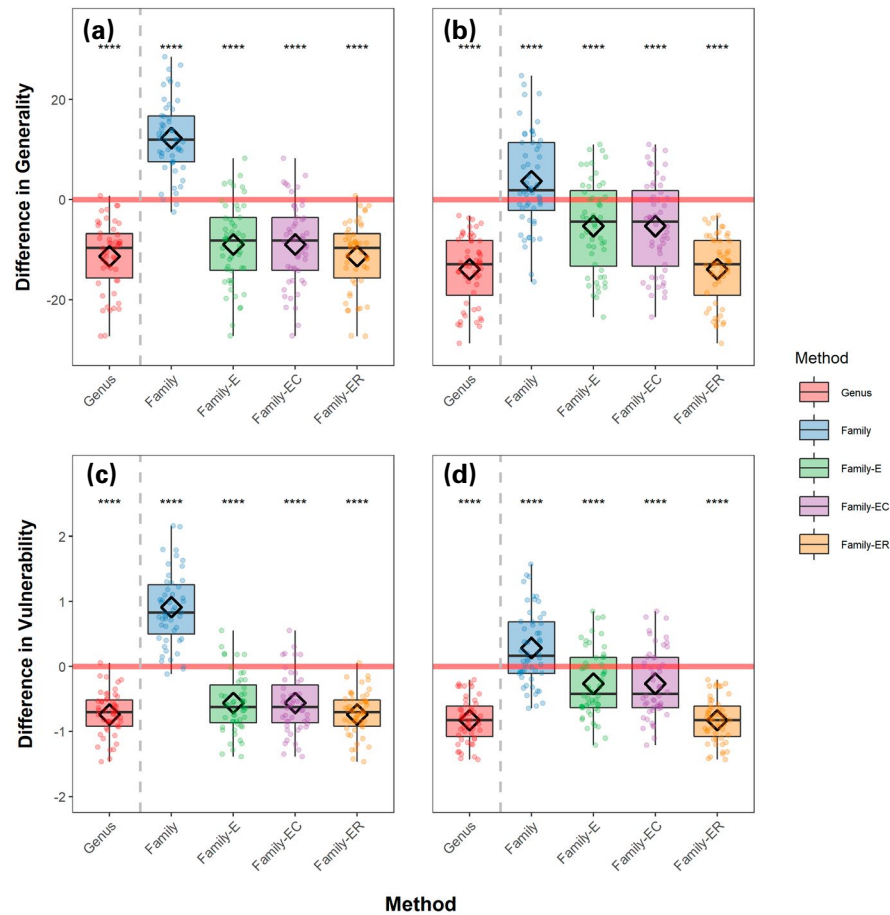
**Fig. 4.** Predation matrices for Naerincheon stream, an example site in the target group. Matrices for food webs constructed by five methods (Genus, taxonomic matching at the genus level; Family, taxonomic matching at the family level for fish species; Family-E, taxonomic matching at the family level for endemic fish species; Family-EC, taxonomic matching at the family level for endemic fish species who play a role of consumers; Family-ER, taxonomic matching at the family level for endemic fish species who play a role of resources). Columns and rows of a matrix indicate fish consumers and their resources, respectively. Characteristics of the matrix element are presented by color (white, links that are neither observed nor generated; orange, links that are observed but not generated; yellow, links that are not observed but generated, red, links that are correctly generated). Values of TPR (True Positive Rate) and TSS (True Skill Statistic) are presented below matrices.

a data-poor environment, taxonomically higher-level matching methods can outperform the lower-level matching method, which correctly generates observed links and does not generate unobserved links. However, our model had some limitations. First, the proposed taxonomically higher-level matching methods are restricted to fish species. Thus, the model depends on biomonitoring data, and its uses are limited; for instance, it may not work well in data conditions where endemic benthic macroinvertebrates are enriched. This can be enhanced by securing the gut content data of benthic organisms or designing additional matching rules. For instance, functional feeding groups (e.g., filter feeders, scrapers, and shredders) or morphological traits (e.g., body size and armoring), which are important determinants of feeding, should be considered in future studies. Secondly, some of the proposed methods showed the same performance (i.e., Genus =

Family-ER and Family-E = Family-EC; Tables S3, S4). Although the underlying ecological concepts of the models differ, distinguished by the roles of fish, their operation ultimately depends on the data. The same performance was observed between Family-E and Family-EC because only consumer fish were present in our data. The Family-ER method did not perform well relative to other family-level matching methods but showed the same performance as that of Genus. For Family-ER to be effective, the two requirements for fish, a carnivore and a resource, should be met simultaneously because the model only reflected the feeding of fish but this could not be achieved from our data. Despite existing limitations, our findings support the potential of a link extrapolation approach for future biomonitoring by showing higher performance for both individual food web links and indices in a data-poor environment.

Current stream ecosystem biomonitoring in South





**Fig. 5.** Boxplots of evaluation of model performance using Generality (Gen) and Vulnerability (Vul). For each Gen and Vul, whether the difference in value between the constructed food web by one of the five methods (Genus, taxonomic matching at the genus level; Family, taxonomic matching at the family level for fish species; Family-E, taxonomic matching at the family level for endemic fish species; Family-EC, taxonomic matching at the family level for endemic fish species who play a role of consumers; Family-ER, taxonomic matching at the family level for endemic fish species who play a role of resources) and the counterpart empirical food web was equal to 0 or not was analyzed using a one-sample  $t$ -test. Significance is presented above the box (all  $p$ -values  $< 0.0001$  are indicated by \*\*\*\*). The first column (panels of (a) and (c)) and second column (panels of (b) and (d)) indicate results of the control and target group, respectively. The red horizontal line indicates that the difference in values between the constructed food web and the counterpart food web is zero. The grayish dashed line indicates the division between Gray's method and proposed methods in this study. Boxes show 25th, 50th, and 75th percentiles and mean values (presented by the diamond symbol). Jitters indicate data points for each difference between the constructed food web and the counterpart food web.

Korea focuses on monitoring changes or the recovery of species richness or abundance of aquatic assemblages, which has been broadly used in stream ecosystem health assessments. There is a growing need for a food web approach in the field of domestic biomonitoring; however, researchers suffer from a lack of trophic interaction data and subsequent difficulties in food web construction. Here, we suggest that the developed link extrapolation-based food web model adapted for Korean stream ecosystems can be effectively utilized in stream ecosystem biomonitoring and health assessment by

integrating it with the present NAEMP, although caution is required. In addition, long-term accumulated ecological data are highly valuable itself as well as in terms of their potential use and there is a continuing need for further interpretations beyond aquatic assemblage-independent investigation and analysis. A link extrapolation approach can provide a food web-level understanding. It helps detect ecological changes that cannot be detected by conventional biomonitoring; therefore, it is expected that our model may provide novel insights into biomonitoring.

In sum, this study aimed to develop a link extrapolation-based food web model adapted to Korean stream ecosystems. We designed family-level matching methods for endemic fish species, with a focus on the generalism of predators in aquatic ecosystems, by improving previous taxonomic matching methods and evaluating their predictive power for food web structure. The results showed that using taxonomically higher-level matching methods can be more effective in a data-poor environment than using the lower-level matching method. A data-poor environment is indeed not a special situation in South Korea; thus, our approach can be utilized in other countries with similar situations. Furthermore, this study suggests that convergence with more ecological knowledge can help improve model performance and overcome the challenges derived from data dependency.

### CRedit authorship contribution statement

**M Lee:** Conceptualization, Formal analysis, Data curation, Writing - Original draft, Software. **Y Kim:** Conceptualization, Validation. **K Cho:** Writing - Reviewing and editing, Supervision.

### Declaration of Competing Interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

### ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (RS-2023-00240819, 2022R1A6A3A01087398, and 2021R1A6A1A10045235) and the Ministry of Science and ICT (2019R1A2C1009812).

### SUPPORTING INFORMATION

Supporting information related to this paper can be found at <https://doi.org/10.11626/KJEB.2024.42.2.XXX>.

### REFERENCES

- Allesina S. 2011. Predicting trophic relations in ecological networks: a test of the allometric diet breadth model. *J. Theor. Biol.* 279:161–168. <https://doi.org/10.1016/j.jtbi.2010.06.040>
- Allouche O, A Tsoar and R Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43:1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Baird D and RE Ulanowicz. 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* 59:329–364. <https://doi.org/10.2307/1943071>
- Barbour MT, J Gerritsen, BD Snyder and JB Stribling. 1999. Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers: Periphyton, Benthic Macroinvertebrates and Fish. 2nd Edition. EPA 841-B-99-002. Office of Water, United States Environmental Protection Agency. Washington D. C.
- Bartley TJ, KS McCann, C Bieg, K Cazelles, M Granados, MM Guzzo, AS MacDougall, TD Tunney and BC McMeans. 2019. Food web rewiring in a changing world. *Nat. Ecol. Evol.* 3:345–354. <https://doi.org/10.1038/s41559-018-0772-3>
- Bellwood DR, AS Hoey and JH Choat. 2003. Limited functional redundancy in high diversity systems: resilience and ecosystem function on coral reefs. *Ecol. Lett.* 6:281–285. <https://doi.org/10.1046/j.1461-0248.2003.00432.x>
- Bersier LF, C Banašek-Richter and MF Cattin. 2002. Quantitative descriptors of food-web matrices. *Ecology* 83:2394–2407. [https://doi.org/10.1890/0012-9658\(2002\)083\[2394:QDOFWM\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2394:QDOFWM]2.0.CO;2)
- Chamberlain SA and E Szöcs. 2013. taxize: Taxonomic search and retrieval in R. *F1000Research*. 2:191. <https://doi.org/10.12688/f1000research.2-191.v2>
- Clavel J, R Julliard and V Devictor. 2011. Worldwide decline of specialist species: toward a global functional homogenization? *Front. Ecol. Environ.* 9:222–228. <https://doi.org/10.1890/080216>
- Cohen JE, T Jonsson and SR Carpenter. 2003. Ecological community description using the food web, species abundance, and body size. *Proc. Natl. Acad. Sci. U. S. A.* 100:1781–1786. <https://doi.org/10.1073/pnas.232715699>
- Delmas E, M Besson, MH Brice, LA Burkle, GV Dalla Riva, MJ Fortin, D Gravel, PR Guimarães, DH Hembry, EA Newman, JM Olesen, MM Pires, JD Yeakel and T Poisot. 2019. Analysing ecological networks of species interactions. *Biol. Rev.* 94:16–36. <https://doi.org/10.1111/brv.12433>
- Eklöf A, MR Helmus, M Moore and S Allesina. 2012. Relevance of evolutionary history for food web structure. *Proc. R. Soc. B-Biol. Sci.* 279:1588–1596. <https://doi.org/10.1098/rspb.2011.2149>

- Fath BD, H Asmus, R Asmus, D Baird, SR Borrett, VN de Jonge, A Ludovisi, N Niquil, UM Scharler, U Schüchel and M Wolff. 2019. Ecological network analysis metrics: The need for an entire ecosystem approach in management and policy. *Ocean Coastal Manage.* 174:1–14. <https://doi.org/10.1016/j.ocecoaman.2019.03.007>
- Filgueira R, JM Chapman, CD Suski and SJ Cooke. 2016. The influence of watershed land use cover on stream fish diversity and size-at-age of a generalist fish. *Ecol. Indic.* 60:248–257. <https://doi.org/10.1016/j.ecolind.2015.06.006>
- Goldwasser L and J Roughgarden. 1993. Construction and analysis of a large Caribbean food web. *Ecology* 74:1216–1233. <https://doi.org/10.2307/1940492>
- Goldwasser L and J Roughgarden. 1997. Sampling effects and the estimation of food-web properties. *Ecology* 78:41–54. [https://doi.org/10.1890/0012-9658\(1997\)078\[0041:SEATEO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1997)078[0041:SEATEO]2.0.CO;2)
- Gray C, DH Figueroa, LN Hudson, A Ma, D Perkins and G Woodward. 2015. Joining the dots: An automated method for constructing food webs from compendia of published interactions. *Food Webs* 5:11–20. <https://doi.org/10.1016/j.fooweb.2015.09.001>
- Grey J, SJ Thackeray, RI Jones and A Shine. 2002. Ferox trout (*Salmo trutta*) as 'Russian dolls': Complementary gut content and stable isotope analyses of the Loch Ness foodweb. *Freshw. Biol.* 47:1235–1243. <https://doi.org/10.1046/j.1365-2427.2002.00838.x>
- Guenther CB and A Spacie. 2006. Changes in fish assemblage structure upstream of impoundments within the upper Wabash River basin, Indiana. *Trans. Am. Fish. Soc.* 135:570–583. <https://doi.org/10.1577/T05-031.1>
- Heberling JM, JT Miller, D Noesgaard, SB Weingart and D Schigel. 2021. Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2018093118. <https://doi.org/10.1073/pnas.2018093118>
- Heleno RH, RS Ceia, JA Ramos and J Memmott. 2009. Effects of alien plants on insect abundance and biomass: A food-web approach. *Conserv. Biol.* 23:410–419. <https://doi.org/10.1111/j.1523-1739.2008.01129.x>
- Kristensen P and J Bøgestrand. 1996. Biological assessments of river quality. pp. 77–79. In: *Surface Water Quality Monitoring*. The European Environment Agency. Copenhagen, Denmark.
- Layer K, A Hildrew, D Monteith and G Woodward. 2010. Long-term variation in the littoral food web of an acidified mountain lake. *Glob. Change Biol.* 16:3133–3143. <https://doi.org/10.1111/j.1365-2486.2010.02195.x>
- Layer K, AG Hildrew and G Woodward. 2013. Grazing and detritivory in 20 stream food webs across a broad pH gradient. *Oecologia* 171:459–471. <https://doi.org/10.1007/s00442-012-2421-x>
- Layer K, AG Hildrew, GB Jenkins, JO Riede, SJ Rossiter, CR Townsend and G Woodward. 2011. Long-term dynamics of a well-characterised food web: Four decades of acidification and recovery in the Broadstone Stream model system. *Adv. Ecol. Res.* 44:69–117. <https://doi.org/10.1016/B978-0-12-374794-5.00002-X>
- Mestre F, D Gravel, D García-Callejas, C Pinto-Cruz, MG Matias and MB Araújo. 2022. Disentangling food-web environment relationships: A review with guidelines. *Basic Appl. Ecol.* 61:102–115. <https://doi.org/10.1016/j.baae.2022.03.011>
- MOE. 2018. Water Quality Monitoring Program. Ministry of Environment. Sejong, Korea.
- Morales-Castilla I, MG Matias, D Gravel and MB Araújo. 2015. Inferring biotic interactions from proxies. *Trends Ecol. Evol.* 30:347–356. <https://doi.org/10.1016/j.tree.2015.03.014>
- Petchey OL, AP Beckerman, JO Riede and PH Warren. 2008. Size, foraging, and food web structure. *Proc. Natl. Acad. Sci. U. S. A.* 105:4191–4196. <https://doi.org/10.1073/pnas.0710672105>
- Petchey OL, AP Beckerman, JO Riede and PH Warren. 2011. Fit, efficiency, and biology: Some thoughts on judging food web models. *J. Theor. Biology.* 279:169–171. <https://doi.org/10.1016/j.jtbi.2011.03.019>
- Poelen J, S Gosnell and S Slyusarev. 2017. rglobi: R interface to global biotic interactions. R package version 0.2, 11.
- Poelen JH, DS James and CJ Mungall. 2014. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecol. Inform.* 24:148–159. <https://doi.org/10.1016/j.ecoinf.2014.08.005>
- Poisot T, D Gravel, S Leroux, SA Wood, MJ Fortin, B Baiser, AR Cirtwill, MB Araújo and DB Stouffer. 2016. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography* 39:402–408. <https://doi.org/10.1111/ecog.01941>
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>. Accessed June 18, 2024.
- Reis ADS, MP Albrecht and SE Bunn. 2020. Food web pathways for fish communities in small tropical streams. *Freshw. Biol.* 65:893–907. <https://doi.org/10.1111/fwb.13471>
- Son YM and HG Byeon. 2001. Feeding habit of main carnivorous fish (*Erthrocultus erythropterus*, *Opsariichthys uncirostris* and *Micropterus salmoides*) at Lake Paldang. *Inst. Bas. Sci. Seowon Univ.* 15:61–78.
- Terry JCD and OT Lewis. 2020. Finding missing links in interaction networks. *Ecology* 101:e03047. <https://doi.org/10.1002/ecy.3047>
- Thompson RM, U Brose, JA Dunne, RO Hall, S Hladyz, RL Kitching, ND Martinez, H Rantala, TN Romanuk, DB Stouffer and JM Tylianakis. 2012. Food webs: Reconciling the structure and

- function of biodiversity. *Trends Ecol. Evol.* 27:689–697. <https://doi.org/10.1016/j.tree.2012.08.005>
- Tylianakis JM, T Tcharntke and OT Lewis. 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* 445:202–205. <https://doi.org/10.1038/nature05429>
- USEPA. 2002. Biological Assessments and Criteria: Crucial Components of Water Quality Programs. EPA 822-F-02-006. Office of Water, United States Environmental Protection Agency. Washington D. C.
- Wang S, BK Luo, YJ Qin, LH Su, SD Stewart, TT Wang, JP Tang, BD He, JH Zhang, HJ Lin and Y Yang. 2020. Consumer-diet discrimination of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ : Source- and feeding-oriented patterns based on gut content analysis in a large subtropical river of China. *River Res. Appl.* 36:1124–1136. <https://doi.org/10.1002/rra.3644>
- Williams RJ and ND Martinez. 2008. Success and its limits among structural models of complex food webs. *J. Anim. Ecol.* 77:512–519. <https://doi.org/10.1111/j.1365-2656.2008.01362.x>
- Yoshii K, NG Melnik, OA Timoshkin, NA Bondarenko, PN Anoshko, T Yoshioka and E Wada. 1999. Stable isotope analyses of the pelagic food web in Lake Baikal. *Limnol. Oceanogr.* 44:502–511. <https://doi.org/10.4319/lo.1999.44.3.0502>