Ecological network inferred from field data elucidates invasion of a freshwater ecosystem by the exotic silver carp

Chen Liao[1], Joao B. Xavier[1,*] and Zhenduo Zhu[2,*]

[1]Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, United States of America, [2]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY 14260, United States of America

[*]To whom correspondence may be addressed: J. Xavier (Email: XavierJ@mskcc.org); Z. Zhu (Email: zhenduoz@buffalo.edu)

Keywords:

ecological network; Lotka-Volterra equations; inference; coexistence mechanism; invasive species

Abstract

Networks of interspecific interactions drive community structure, dynamics and stability. The ability to infer interspecies interactions from observational field data would open possibilities to apply network models to manage real world ecosystems. Here, we show this is possible for a freshwater fish community in the Illinois River, United States, using long-term data collected through time and space. We solve the challenge of sparsely sampled field data using latent variable regression and constraints imposed by known trophic structure in the fish community. Network analysis indicates that the most abundant 9 fish coexisted thanks to equalizing mechanisms that reduced fitness differences between strong and weak competitors. Importantly, the network sheds light on the ongoing invasion by the exotic silver carp (*Hypophthalmichthys molitrix*), revealing that the invader outproduces native preys, replacing their contributions to the diets of native predators. Our work shows that field data and constrains imposed by known food webs can improve network inference and produce quantitative insights that could aid in conservation of freshwater ecosystems threatened by invasive species.

Introduction

Ecological network theory seeks to understand how the many interactions between species affect dynamics and stability of the ecosystem. In 1972, Robert May questioned a paradigm established at the time by proving that networks with many, highly connected species tend to be less stable (May 1972). May's pioneering work had immense influence and has led to ecological applications of random matrix theory, which assumes that interaction strengths and types are sampled from random distributions.

Interactions in realistic food webs are far from random in both strength and type. For example, most links of plant-animal mutualistic food webs are weak and asymmetric (Bascompte and Jordano 2007). But to date, the measurement of interactions among species in natural communities has been mostly conducted through ecological field observations in controlled environments (Carrara et al. 2015). If the underlying premise of ecological network theory is correct—that the structure and strength of interactions drive the observed community dynamics—we should be able to infer interaction networks from time-series data of species' abundance. Computational inference would require less labor than directly observing interactions through detailed and laborious fieldwork. It should also be more applicable to large communities with complex dynamic interdependencies between their member species, which are hard to study in isolation. Many tools to infer network structure from field data have been proposed before, ranging from parameter-free correlation-based algorithms to parameter-intensive mechanistic models. Some approaches have been successfully applied to microbial communities (Stein et al. 2013; Berry and Widder 2014; Buffie et al. 2015; Steinway et al. 2015; Bucci et al. 2016; Cao et al. 2017; Venturelli et al. 2018), but inference of macroscopic ecosystems has lagged behind (Milns et al. 2010; Sander et al. 2017; Ushio et al. 2018).

To analyze dynamics of ecological communities, multispecies pairwise models, particularly the Lotka-Volterra equations, have been extensively used in theoretical and computational ecology due to their simplicity, tractability, and transparent logic (Hofbauer and Sigmund 1998; Coyte et al. 2015). In the generalized form of Lotka-Volterra the underlying ecology is phenomenologically summarized with minimal parameterization: the fitness effect of each one-way interaction in a food web is quantified by a

single coefficient whose magnitude and sign represent the interaction strength and type, respectively. The signs of the interaction coefficients can be imposed when we know the relative trophic positions between species in the food web, resulting in a hybrid model (Cohen and Łuczak 1990) that infers population dynamics constrained by known structure.

Here we apply a generalized Lotka-Volterra (GLV) network to study the species-species interactions within the Illinois River fish community. We use field data collected by the Long Term Resource Monitoring Program in the Upper Mississippi River System (Ratcliff et al. 2014), one of the very few long-term monitoring programs in large rivers in the United States (Ward et al. 2017), and we introduce a new latent variable regression approach to allow GLV inference from these noisy and infrequently (annually) sampled data. To improve parameter identifiability, we constrained the inference based on the reconstructed trophic structure of the fish community. The inferred network revealed new insights into the mechanism underlying species coexistence in the ecosystem and the keystone species that help maintain coexistence. Finally, we assessed the impact of the silver carp on the ecosystem, an invasive species with enormous impact that is damaging to the native fish network in the Mississippi and Illinois Rivers (Irons et al. 2007; Solomon et al. 2016), and may invade the Laurentian Great Lakes in the future (Jerde et al. 2013). Network analysis reveals that the silver carp invaded by outgrowing native prey, thus taking their niche in the trophic web.

Results

Fish community varies in space and time

The Illinois River is a major tributary of the Upper Mississippi River, where the long-term monitoring efforts of the fish community spread across six field stations since 1993 (Fig. 1a). To visualize how the fish community structure varied across time and space we first standardized catch-per-unit-effort data to combine fish numbers obtained from the different fishing gears employed (Supplementary Fig. 1 and see Methods). Then we carried out a principle component analysis (PCA) using data from the normalized abundances of 153 fish species for each year and site (Fig. 1b). The data from each site occupied distinct regions of the PCA plot, indicating distinct fish ecologies in space. Still, despite regional differences, the communities were most similar between proximal sites. The first component, which explains the most variance of the data, is strongly determined by variations in the common carp and bluegill, two species highly abundant in the Mississippi River upstream from the confluence with the Illinois River (Pool 4, Pool 8, and Pool 13) but less abundant in the Illinois River (LG) and the Mississippi River downstream from the confluence (Pool 26 and OR).

We conducted a detailed analysis of the fish community in La Grange pool, the only monitoring site along the Illinois River (Fig. 1a). We chose this site to study the ecological impacts of invasive silver carp on native fish population (Fig. 1c). The impact of the silver carp was detected in three sites (OR, Pool 26, and LG) over the course of the invasion (Fig. 1b, inset). The Illinois River is known to have one of the highest silver carp densities worldwide (Sass et al. 2010). The large silver carp density is obvious in the PCA, which shows that the loading vector for the silver carp aligns well with the La Grange community data (Fig. 1b, in brown). In contrast, the Mississippi sites upstream of the confluence with the Illinois River (Pool 4, Pool 8, and Pool 13) where silver carp are barely found (Fig. 1b, inset) are misaligned with the silver carp vector. Fig. 1b and its inset also reveal the invasion path: silver carp entered the Illinois River at the confluence, rather than continuing to migrate up the Mississippi River. There is grave concern that the invader may enter Lake Michigan through the Illinois River, threatening the Great Lakes' ecosystems and multi-billion-dollar fishing industry (Jerde et al. 2013).

97

98 *Latent variable approach coupled with known interactions from the literature improves network*
99 *inference*

100 In order to define the network of species' interactions we seek to parameterize a GLV model. The most
101 common technique is to discretize the model, transform the discretized version to a system of linear
102 equations, and then fit a multilinear regression (see Methods). This method is gaining popularity to infer
103 microbial ecological networks from gut microbiome time series data (Stein et al. 2013) and already served
104 to find a native gut bacterial species that helps stymie invasion by a pathogenic species (Buffie et al. 2015).
105 Compared to microbiome data, field data are often noisier and less frequently sampled (Harwood and
106 Stokes 2003), which amplifies the discretization error in calculating log-derivatives and precludes the
107 inference of a realistic network (Cao et al. 2017). Even a good fit to the transformed linear problem can
108 produce a network that recreates the observed dynamics poorly (Cao et al. 2017).

109 To circumvent this fundamental limitation, we introduce a new method where we optimize log-
110 derivatives rather than using numerical approximations (see Methods and Fig. 2a). This strategy comes
111 from the family of latent variable regressions as the log-derivatives are treated as latent variables (their
112 large uncertainty essentially makes them unobserved variables) and can be learned by minimizing model
113 prediction error. In tests using synthetic data generated by a 3-species GLV model (Supplementary Fig. 2a),
114 the new approach (Fig. 2b) outperformed the most commonly used algorithm based on linear regression
115 (Fig. 2c) in recreating the oscillatory community dynamics. In addition, the inference converged much
116 faster than the derivative-based nonlinear regression (Supplementary Fig. 2b and 3).

117 We further tested the effectiveness of the latent variable regression using experimental data from
118 two separate, independently published predator-prey microbial systems. In these laboratory systems the
119 interspecific relationships are known, which means that finding the correct signs for the interaction network
120 would constitute two independent tests to validate our inference algorithm. As expected, our GLV model
121 inferred the correct network structure (Fig. 2d,f), and reproduced the community dynamics observed
122 experimentally in both tests (Fig. 2e,g). In the first test, the *Paramecium-Didinium* system (Veilleux 1976),

123 the inferred network successfully captured the positive impact of the prey (*Paramecium aurelia*) on the
124 predator (*Didinium nasutum*) and the negative impact in the reverse direction. In the second test, a three
125 species rotifer-algae ecosystem (Kasada et al. 2014), the inferred network revealed that the two algal clones
126 have evolved to possess different survival strategies (Fig. 2f): the second clone (Ch2) grows slower than
127 the first clone (Ch1) but develops resistance to rotifer's predation. The fitness trade-off between prey's
128 defense and growth agreed with experimental observations qualitatively.
129

130 *GLV parameterization from fish population data*

131 To infer a network of species interactions that captured the fish community dynamics in the La Grange pool
132 we focused on the top 9 most abundant native fish species (Fig. 1c, Supplementary Fig. 4, and
133 Supplementary Table 1), which account for 85.5% of the total abundance. We chose not to group the
134 remaining low-abundance species because grouping can produce spurious links between the virtual group
135 and the abundant species (Cao et al. 2017).
136 Then, to constrain GLV inference, we reconstructed a summary food web consisting of all potential
137 interactions among the 9 selected fish species based on Fishbase (http://www.fishbase.org) trophic level
138 indices as well as experimentally observed trophic interactions (see Methods and Supplementary Fig. 5). We
139 next converted these putative interactions to symbolic constraints of the GLV model parameters: a positive,
140 neutral, or negative interaction requires its corresponding parameter to be positive, 0 or negative as well
141 (see Methods and Supplementary Table 2). This approach—which combines latent variable regression with
142 constraints based on known positions on the food web—outperformed the simple linear regression method
143 (Fig. 3a, adjusted R^2 values of 0.80 compared to 0.45). The relative growth rates inferred for each fish
144 species (Supplementary Table 3) ordered the species by their trophic levels, with predators at the top to
145 resource preys at the bottom (Fig. 3b). Hierarchical clustering of the inferred pairwise interactions
146 (Supplementary Table 3) revealed that species occupying the same trophic level share similar interactions
147 with the rest of the community members (Fig. 3c). The inferred network is dominated by weak interactions
148 (Fig. 3d), a pattern thought to promote stability and often observed in natural ecosystems (Wootton and
149 Emmerson 2005; Ushio et al. 2018).

150

151 *Data from another pool validates the interactions inferred in the La Grange pool*

152 Data coming from a single source is often biased, which could make the inferred network unreliable. Here
153 we used data from an additional pool to refine the network and improve model predictions. Pool 26—at the
154 confluence of the Upper Mississippi and Illinois Rivers—is the closest pool and had the most similar
155 community to the one in the La Grange pool (Fig. 1b and Supplementary Fig. 4). Therefore, we assumed
156 that the two communities were similar enough that any pair of species interacting in La Grange pool may
157 interact the same way in Pool 26, if they co-occur.

158 We applied the same latent variable approach to the time series of fish population in Pool 26
159 (Supplementary Fig. 6) to infer an independent network (Supplementary Table 4). For both sites, we
160 quantified the uncertainty associated with each interaction coefficient by assigning its own confidence score,
161 defined as the minimum significance level above which the confidence interval does not contain 0, to
162 indicate how likely the coefficient is significantly different than zero. The confidence scores were generally
163 proportional to the absolute values of their corresponding GLV coefficients (Supplementary Fig. 7). With
164 at least 50% confidence at both sites, we identified 11 negative and 3 positive interspecific interactions (Fig.
165 4), all of which are consistent with empirical observations or ecologically interpretable (Supplementary
166 Information). For example, the two most confident predictions are the positive interactions from emerald
167 shiner and gizzard shad to channel catfish, suggesting that the two prey fish support channel catfish's
168 growth by serving as major food sources.

169

170 *Fitness equalization maintains species coexistence despite competition*

171 The network inferred prompted us to explore the stability of the fish community. The model predicted stable
172 coexistence of all 9 fish species at steady state (Fig. 5a, right). Still, linear stability analysis (see Methods)
173 revealed that 80% of all attainable steady states of the system are stable (Supplementary Fig. 8), suggesting
174 that the community could stabilize in alternative compositions different than those observed so far.

175 To understand the mechanism of species coexistence, we simulated the steady state of two
176 subnetworks consisting of (1) only resource preys and (2) resources preys plus mesopredators. In the

absence of any predators, common carp (a strong competitor) is the most competitive resource prey that drives other preys (weak competitors) to extinction in the long run (Fig. 5a, left). This is expected from the competitive exclusion principle. The common carp was an invasive species first introduced to the United States in 1800s that has caused damages to the fish community due to its rapid growth and high tolerance to poor water quality. However, the presence of mesopredators, particularly freshwater drum, can reverse the outcome of competition by preying on common carp (Fig. 5a, middle). The fitness inequality between the strong and weak competitors persists until the addition of top predators, which suppress population of mesopredators and reduce their predation pressure on resource preys (Fig. 5a, right). Our analysis thus supports the neutral theory (Hubbell 2001): the trophic structure of the inferred network tends to equalize the fitness of otherwise unequal competitors.

When the interactions in the network function simultaneously they give rise to many indirect interactions. This causes trophic cascade effects where some species will have greater impacts than the others on community structure. We therefore quantified the impacts of each species by computing a 'total importance score' based on the inferred network (see Methods). The analysis revealed that only the gizzard shad had a positive impact on the network as a whole (Fig. 5b), suggesting its facilitating role in compensating for the overall competitive environment of the fish community. Its important role in maintaining the community was confirmed by *in silico* experiments where we simulated removing a given species from the network: excluding gizzard shad caused the highest number of secondary extinctions (Fig. 5c). The gizzard shad promotes diversity by reducing competition for less competitive preys and supporting the growth of predator fish that feed on them. In the absence of the gizzard shad, the common carp increased more than 10 folds and the channel catfish collapsed (Supplementary Fig. 9).

Silver carp outgrows native prey and infiltrates predators' diets

To study the impact of silver carp, a present threat to the fisheries in the North America, we included this species into our model as a perturbation to the native food web (Supplementary Fig. 10) while the GLV model inferred for the native community remained unchanged. The silver carp only became detectable recently, and the number of data points in silver carp's time series was insufficient for reliably estimating

new GLV coefficients introduced by silver carp. We solved this problem by using Markov Chain Monte Carlo (MCMC) to approximate the posterior distributions of these parameters (see Methods and Supplementary Table 5), which were later used to predict the long-term effects of silver carp on the native fish.

The simulations indicated that the silver carp integrates stably into the native fish community without causing extinctions (Fig. 6a). Therefore, the invasion impacted the community composition by proportionally reducing the relative abundance of all native fish species. Examining the diet composition of native predators indicated that silver carp replaced the native preys and contributed about 10% of predators' diets (Fig. 6b). This agreement with previous observations of stomach content of native predators in La Grange pool—including channel catfish, black crappie, and white bass—revealed that silver carp had indeed entered their diets (Anderson 2016).

Importantly, our analysis suggested that the silver carp could only invaded when its growth rate exceeded a threshold of 3.33 1/year (Fig. 6c). The need to exceed a threshold growth rate is expected because all native fish species—preys and predators—had a negative impact on the silver carp (Supplementary Fig. 10); therefore, its population growth rate must be sufficiently high to counterbalance the negative pressure from the native fish network. The estimated population growth rate of silver carp is not only above the threshold but even greater than the growth rate of any other native fish, which possibly explains the ongoing invasion.

Discussion

Here we demonstrate the feasibility of inferring ecological networks from field data to produce quantitative insights valuable for the management of real world ecosystems. Field data are invaluable for ecology, but noisy and infrequent sampling can hinder network inference—especially with algorithms such as generalized Lotka-Volterra which require the calculation of time derivatives (Stein et al. 2013). The problem could in principle be solved by measuring better data and at higher rates, but this is often too costly or simply impractical. The inference method we propose here offers a practical solution to obtain realistic networks, based on a latent variable approach combined with constraints from known trophic interactions obtained from the literature.

Applying the inference method to a fluvial fish community revealed the importance of using realistic networks to analyze real world systems. The system shows the coexistence of resource preys despite interspecific competition for shared niche; this is historically intriguing because obtaining stable coexistence of a large community in Lotka-Volterra-type models is extremely improbable (Serván et al. 2018). Using random matrix theory, it has been repeatedly shown that the probability of a feasible solution with strict positive abundances for all species is $1/2^N$ for a N-species community and the chance of stable coexistence is even lower (Serván et al. 2018). However, the low likelihood is inconsistent with the immense diversity observed within natural populations, spurring countless studies to identify possible mechanisms underlying species coexistence (Levine et al. 2017). One such coexistence mechanism, as we revealed here, is the fitness-equalizing effect where predators suppress the population of the most competitive species and thereby dampen the intensity of interspecific competitions. Our study suggests that inferring interaction patterns from time series data may be an effective strategy to advance our mechanistic understanding of species coexistence.

The parameterized network provided intriguing results that help interpret the ongoing invasion by the exotic silver carp. First, silver carp were predicted to be sufficiently integrated into the local fish community, which suggests they may become “native” eventually. This is not surprising, given that common carp, another invasive carp introduced to United States in the 1800s, have become members of the

Illinois River fish community after so many years of establishment and naturalization. Second, the invasion did not seem to cause extinctions and the effects of the prey preference on the population structure of native fish were apparently not drastic. These results could be naïvely interpreted as suggesting that the impact of the silver carp is not detrimental; but the moderate impact may be due to the high productivity and species richness in the Illinois River, which mitigates the effects of interspecific competition for food sources. Indeed, the native predators could even benefit from supplemental prey that they do catch. But caution is still needed when interpreting these results: The overall decrease in native fish abundance caused a measurable shift in the ratio of predator-to-prey which may increase the likelihood of stochastic extinctions with unforeseen—and potentially irreversible—consequences. Our model makes predictions that could be used to reverse the invasion: reducing silver carp's net growth below the critical threshold of 3.33 1/year, for example by targeted harvest, would dramatically curb the invasion.

We have adopted the simple GLV model to increase the generality of our inference approach and to reveal the core interactions driving population dynamics, but our algorithm can be extended in multiple ways. For example, nonlinear functional responses (predation rate as a function of prey density), high-order interactions and environmental factors that encapsulate greater realistic details can be included. Future applications of the algorithm include identification of core interactions from time series collected from different geographical locations with varying ecological and environmental conditions (e.g., other rivers with long-term resource monitoring data). Along with other inference approaches such as dynamic Bayesian network (Sander et al. 2017) and convergent cross mapping (Ushio et al. 2018), such applications will undoubtedly advance our ability to describe and predict ecosystem structure, dynamics and stability, and shed light on disruptive threats posed by invasive species.

Methods

Long term resource monitoring data

The time series data of the Upper Mississippi and Illinois River fish community were collected from the annual reports of the Long Term Resource Monitoring Program (Ratcliff et al. 2014). The program used a multigear and multihabitat sampling design protocol (refer to the program report for details) to collect data from 6 observation sites (Lake City, Minnesota, Pool 4; La Crosse, Wisconsin, Pool 8; Bellevue, Iowa, Pool 13; Alton, Illinois, Pool 26; Havana, Illinois, La Grange Pool; and Cape Girardeau, Missouri, Open River). To standardize the catch per unit effort (CPUE) from multiple gears to the same relative scale, the raw CPUE data during the time period between 1993 and 2015 were converted to relative abundance among species within the same site and summed over all 6 fishing gears (electrofishing, fyke net, mini fyke net, large hoop net, small hoop net, trawling). Parameters of the GLV model were estimated with the standardized CPUE indices since the absolute abundances are not available.

The standardized data was first smoothed using empirical mode decomposition (EMD) and then used as inputs for parameter estimation. The details of the classical EMD algorithm has been extensively reviewed elsewhere (Wang et al. 2014). Briefly, EMD decomposes the input data into several intrinsic mode functions (IMF), which represent the natural oscillation modes of the input time-series data. The trended curve of each original time-series was reconstructed by summed up all IMFs with Hurst exponent no smaller than 0.5.

Reconstruction of summary food web

The trophic structure of the modelled fish community was reconstructed by classifying all species as resource prey, mesopredator, or top predator in a three-tier trophic food web. The classification was determined by requiring (1) the FishBase (https://www.fishbase.de) trophic level index (a floating-point number that equals one plus weighted mean trophic level index of the food items) of any fish species in higher trophic levels is no smaller than that of any fish species in lower levels; (2) the predator of any known predator-prey relationship from prior knowledge occupies a higher trophic level than the level

occupied by the prey. We collected 14 predator-prey pairs from literature (Supplementary Fig. 5b and Supplementary Fig. 10) and assume that each pair observed to interact in other freshwater ecosystems has the potential to interact the same way in the Upper Mississippi and Illinois Rivers.

With the reconstructed trophic structure, all potential interactions besides the known predator-prey relationships can be identified based on three hypothetical rules: (1) fish species on higher trophic levels feed on fish species on the immediate lower level (common prey relationships); (2) the same fish species compete for limited resources within its own population (intraspecific competitions); (3) fish species on the same trophic level compete with each other for limited resources (interspecific competitions). Any pair of fish species whose trophic relationship does not fall into the three categories is assumed to be non-interacting.

GLV model parameterization: Linear Regression

The generalized Lotka-Volterra model is a system of ordinary differential equations with birth-death processes describing how species numbers change in time

$$\frac{dx_i(t)}{dt} = \left[\beta_{i,0} + \sum_{j=1}^{N} \beta_{i,j} x_j(t)\right] x_i(t) \tag{1}$$

where $x_i(t)$ is the abundance of species i at time t and N is the total number of species. $\beta_{i,0}$ is referred to as the net population growth rate (birth minus death) of the species i while $\beta_{i,j}$, known as the pairwise interaction coefficient, represents the population influence of species j on species i.

The most commonly used technique to infer GLV parameters is to discretize Eq. (1) and solve the following linear regression (Stein et al. 2013)

$$\boldsymbol{\beta}^{opt} = \text{argmin}(||\boldsymbol{\beta} \cdot \boldsymbol{X} - \boldsymbol{dL}||_2^2) \tag{2}$$

where $\boldsymbol{\beta}$ $(= [\boldsymbol{\beta}_+ \quad \boldsymbol{\beta_0}])$, $\boldsymbol{X}, \boldsymbol{dL}$ are the matrixes of the GLV parameters, the time-series data, and the log-derivatives of the time-series data respectively ($t_1, t_2, \ldots, t_M$ are discrete time points)

$$\boldsymbol{\beta}_{+} = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,N} \\ \vdots & \ddots & \vdots \\ \beta_{N,1} & \cdots & \beta_{N,1} \end{bmatrix} \tag{3}$$

$$\boldsymbol{\beta_0} = [\beta_{1,0} \quad \cdots \quad \beta_{N,0}]^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{X} = \begin{bmatrix} x_1(t_1) & \cdots & x_1(t_M) \\ \vdots & \ddots & \vdots \\ x_N(t_1) & \cdots & x_N(t_M) \\ 1 & \dots & 1 \end{bmatrix} \tag{5}$$

$$\boldsymbol{dL} = \begin{bmatrix} (ln(x_1))'_{t=t_1} & \cdots & (ln(x_1))'_{t=t_M} \\ \vdots & \ddots & \vdots \\ (ln(x_N))'_{t=t_1} & \cdots & (ln(x_N))'_{t=t_M} \end{bmatrix} \tag{6}$$

The log-derivatives $(d\ln(x_i(t))/dt)$ in Eq. (6) can be numerically approximated from data using the Euler's method

$$\frac{d\ln(x_i(t))}{dt}_{|t=t_k} \approx \frac{\ln(x_i(t_{k+1})) - \ln(x_i(t_k))}{t_{k+1} - t_k} \tag{7}$$

or high-order approaches such as the Runge-Kutta methods or spline interpolation.

GLV parameters may be constrained during the linear regression. Symbolic constraints (i.e., constraints on the signs) can be derived based on potential interactions between species: (1) $\beta_{i,j} < 0$ and $\beta_{j,i} > 0$ for predator (species j)-prey (species i) relationships; (2) $\beta_{i,i} < 0$ for intraspecific competitions within population of species i; (3) $\beta_{i,j} < 0$ and $\beta_{j,i} < 0$ for interspecific competitions between species j and species i; (4) $\beta_{i,j} = 0$ and $\beta_{j,i} = 0$ for non-interacting species pairs. Population growth rates of species on the lowest trophic level are positive and the rates of species on all higher trophic levels are negative. This makes intuitive sense: predators should decline in the absence of prey.

GLV model parameterization: Latent variable regression

The flowchart of the latent variable regression is shown in Fig. 2a. Conceptually, the method alternates between updating the GLV parameters ($\boldsymbol{\beta}$) and the log-derivatives ($\boldsymbol{dL}$) until convergence is reached. Given $\boldsymbol{dL}^{(k)}$ from a previous iteration k, we optimized $\boldsymbol{\beta}^{(k+1)}$ by solving the linear regression given in Eq. (2) in the current iteration $k+1$. $\boldsymbol{dL}^{(k)}$ is then updated to $\boldsymbol{dL}^{(k+1)}$ according to the Levenberg-Marquardt

optimization algorithm (Moré 1978), which adaptively updates parameter values based on a blend of gradient descent and Gauss-Newton update rules. The objective function to be minimized is the sum of squared errors (SSE) of prediction plus two penalty terms

$$\text{Objective function} = ||\boldsymbol{X} - \boldsymbol{X^s}||_2^2 + \lambda_0||\boldsymbol{\beta_0}||_2^2 + \lambda_+||\boldsymbol{\beta}_+||_2^2 \quad (8)$$

where $\boldsymbol{X}$ is the observed data, $\boldsymbol{X^s}$ is the simulated time series using $\boldsymbol{\beta}^{(k+1)}$, and λ_0 and λ_+ are the penalty coefficients for the growth rate vectors and the interaction matrix respectively. The iterations continue until the convergence criteria for the Levenberg-Marquardt algorithm is met.

To maximize the performance of prediction, λ_0 and λ_+ in Eq. (8) were chosen as the combination that minimizes the averaged prediction error over 10 random runs of 3-fold cross validation. The time series data was first partitioned into 3 subsets of equal size; two training subsets ($\boldsymbol{X^{train}}$) were used to infer the parameters and the remaining test subset ($\boldsymbol{X^{test}}$) was used to evaluate the corresponding prediction error, i.e., $||\boldsymbol{X^{test}} - \boldsymbol{X^s}||_2^2$. The GLV parameters were determined by running the iterative method described above with the optimized penalty coefficients and the complete dataset. The optimal penalty parameters obtained from 3-fold cross validation are $\lambda_0 = 10^{-3}$ and $\lambda_+ = 10^{-2}$.

Coexistence of competing species has been observed in many natural ecosystems (Bode et al. 2011; Hart et al. 2017) and was also assumed for the fish communities we studied. However, the optimal set of GLV parameters obtained from the above procedure does not guarantee this property at steady state. To search for alternative sets that would fit the observed data equally well but predict stable coexistence, we added a perturbation step by randomly sampling the parameter space in the neighborhood of the optimal set and ran nonlinear optimization algorithm using these samples as initial guesses. To sample an initial guess, we modified the optimal set of GLV parameters by only changing the self-regulation coefficients: their values were uniformly sampled between the negative of the maximum absolute value of all interaction coefficients in the optimal set and 0. The perturbation steps were repeatedly applied to the complete dataset until 100 alternative sets of GLV parameters that give rise to stable coexistence of species were found. Among the 100 alternative solutions, the one that has the minimum SSE was chosen to finally parameterize the GLV model.

Linear stability analysis

Linear stability analysis of the GLV model concerns the stability of dynamic system at steady state (i.e., fixed point) under small perturbations. The steady state solutions ($\boldsymbol{x}_{ss} = [x_1^{ss}, x_2^{ss}, \dots, x_N^{ss}]$) of Eq. (1) can be determined by solving $x_i^{ss}\left(\beta_{i,0} + \sum_{j=1}^{N} \beta_{i,j} x_j^{ss}\right) = 0$ where i ranges between 1 and N. In principle, Eq. (1) can have at most 2^N steady states, including no more than one coexistence state. A stable steady state requires that the real part of all eigenvalues of the Jacobian matrix is negative.

Total importance score

The total importance of species i (TI_i) in a directed network is quantified by the sum of effects originated from this species to all other species in the network

$$TI_i = \sum_{j=1, j\neq i}^{N} \sum_{\substack{k \in \text{all paths} \\ \text{from } i \text{ to } j}} \prod_{\substack{p,q \in \text{source,target} \\ \text{nodes of all direct} \\ \text{links of path } k}} \frac{\beta_{p,q}}{\sum_{r \in \text{source nodes of all direct links } r \to q} |\beta_{r,q}|} \quad (8)$$

where the impact of species i on species j is the sum of indirect chain effects over all paths from species i to species j. We assume the indirect chain effect is multiplicative: the effect of indirect interaction along each path is calculated by multiplying the normalized weight of each direct link along the path. For any direct link $p \to q$ (the one-way interaction from species p to specie q), its normalized weight is equal to its corresponding interaction coefficient of the GLV model ($\beta_{p,q}$) divided by the sum of absolute values of interaction coefficients that are associated with all incoming links of node q.

MCMC simulation

MCMC is a class of algorithms that use a sequence of random samples (i.e., Markov chain) to approximate the posterior distribution of model parameters. At each iteration, the algorithm makes a tentative move in the parameter space based on the current parameter values and accepts the move with some probability. Advanced techniques such as adaptive Metropolis samplers and delaying rejection were adopted to enhance

the basic Metropolis-Hastings algorithm (Haario et al. 2006). We simulated 1,000,000 steps of a Markov chain for each parameter in Supplementary Table 5 under the constraints of parameter values between -5 and 0 (for negative interactions) or 0 and 5 (for positive interactions). During MCMC simulation, the population growth rates of native fish species as well as their pairwise interaction coefficients remained unchanged.

Software

The total importance scores for all species in the network were calculated in Python using NetworkX. All other simulations and analysis were performed in MATLAB R2018 (The MathWorks, Inc., Natick, MA, USA). The Levenberg-Marquardt algorithm is implemented in the MATLAB optimization toolbox function lsqnonlin. The parameter confidence intervals of nonlinear regression were calculated using the MATLAB statistics and machine learning toolbox function nlparci. The MATLAB codes of EMD and Hurst exponent estimation can be accessed from https://www.mathworks.com/matlabcentral/fileexchange/52502-denoising-signals-using-empirical-mode-decomposition-and-hurst-analysis. The MCMC toolbox used for this study is available from https://github.com/mjlaine/mcmcstat.

References

Anderson, C. A. 2016. Diet Analysis of Native Predatory Fish to Investigate Predation of Juvenile Asian Carp. PhD thesis, Western Illinois University.

Bascompte, J., and P. Jordano. 2007. Plant-animal mutualistic networks: the architecture of biodiversity. Annual Review of Ecology, Evolution, and Systematics 38:567–593.

Berry, D., and S. Widder. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology 5:219.

Bode, M., L. Bode, and P. R. Armsworth. 2011. Different dispersal abilities allow reef fish to coexist. Proceedings of the National Academy of Sciences of the USA 108:16317–16321.

Bucci, V., B. Tzen, N. Li, M. Simmons, T. Tanoue, E. Bogart, L. Deng, et al. 2016. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome Biology 17:121.

Buffie, C. G., V. Bucci, R. R. Stein, P. T. McKenney, L. Ling, A. Gobourne, D. No, et al. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature 517:205–208.

Cao, H.-T., T. E. Gibson, A. Bashan, and Y.-Y. Liu. 2017. Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. BioEssays 39:1600188.

Carrara, F., A. Giometto, M. Seymour, A. Rinaldo, and F. Altermatt. 2015. Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity. Methods in Ecology and Evolution 6:895-906.

Cohen, J. E., and T. Łuczak. 1990. Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka-Volterra cascade model. Proceedings of the Royal Society of London B 240:607-627.

Cooke, S. L., and W. R. Hill. 2010. Can filter-feeding Asian carp invade the Laurentian Great Lakes? A bioenergetic modelling exercise. Freshwater Biology 55:2138–2152.

Coyte, K. Z., J. Schluter, and K. R. Foster. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663–666.

Haario, H., M. Laine, A. Mira, and E. Saksman. 2006. DRAM: Efficient adaptive MCMC. Statistics and

Computing 16:339–354.

Hart, S. P., J. Usinowicz, and J. M. Levine. 2017. The spatial scales of species coexistence. Nature Ecology and Evolution 1:1066–1073.

Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology and Evolution 18:617–622.

Hofbauer, J., and K. Sigmund. 1998. Evolutionary games and population dynamics. Cambridge University Press, Cambridge.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press, Princeton.

Irons, K. S., G. G. Sass, M. A. McClelland, and J. D. Stafford. 2007. Reduced condition factor of two native fish species coincident with invasion of non-native Asian carps in the Illinois River, U.S.A. Is this evidence for competition and reduced fitness? Journal of Fish Biology 71:258–273.

Jerde, C. L., W. L. Chadderton, A. R. Mahon, M. A. Renshaw, J. Corush, M. L. Budny, S. Mysorekar, and D. M. Lodge. 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. Canadian Journal of Fisheries and Aquatic Sciences 70:522-526.

Kasada, M., M. Yamamichi, and T. Yoshida. 2014. Form of an evolutionary tradeoff affects eco-evolutionary dynamics in a predator-prey system. Proceedings of the National Academy of Sciences of the USA 111:16035–16040.

May, R. M. 1972. Will a large complex system be stable? Nature 238:413–414.

Milns, I., C. M. Beale, and V. A. Smith. 2010. Revealing ecological networks using Bayesian network inference algorithms. Ecology 91:1892–1899.

Moré, J. J. 1978. The Levenberg-Marquardt algorithm: implementation and theory. In: Watson G, editor:Springer Berlin/Heidelberg. 105-116.

Ratcliff, E. N., E. J. Gittinger, T. M. O'Hara, and B. S. Ickes. 2014. Long term resource monitoring program procedures: fish monitoring, 2nd edition. A program report submitted to the U.S. Army Corps of Engineers' Upper Mississippi River Restoration-Environmental Management Program.

Sander, E. L., J. T. Wootton, and S. Allesina. 2017. Ecological network inference from long-term presence-absence data. Scientific Reports 7:7154.

Sass, G. G., T. R. Cook, K. S. Irons, M. A. McClelland, N. N. Michaels, T. Matthew O'Hara, and M. R. Stroub. 2010. A mark-recapture population estimate for invasive silver carp (*Hypophthalmichthys molitrix*) in the La Grange Reach, Illinois River. Biological Invasions 12:433–436.

Serván, C. A., J. A. Capitán, J. Grilli, K. E. Morrison, and S. Allesina. 2018. Coexistence of many species in random ecosystems. Nature Ecology and Evolution 2:1237–1242.

Solomon, L. E., R. M. Pendleton, J. H. Chick, and A. F. Casper. 2016. Long-term changes in fish community structure in relation to the establishment of Asian carps in a large floodplain river. Biological Invasions 18:2883–2895.

Stein, R. R., V. Bucci, N. C. Toussaint, C. G. Buffie, G. Rätsch, E. G. Pamer, C. Sander, et al. 2013. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Computational Biology 9:e1003388.

Steinway, S. N., M. B. Biggs, T. P. Loughran, J. A. Papin, and R. Albert. 2015. Inference of network dynamics and metabolic interactions in the gut microbiome. PLoS Computational Biology 11:e1004338.

Ushio, M., C.-H. Hsieh, R. Masuda, E. R. Deyle, H. Ye, C.-W. Chang, G. Sugihara, et al. 2018. Fluctuating interaction network and time-varying stability of a natural fish community. Nature 554:360–363.

Veilleux, B. G. 1976. The Analysis of a Predatory Interaction between *Didinium* and *Paramecium*. Master's thesis, University of Alberta.

Venturelli, O. S., A. C. Carr, G. Fisher, R. H. Hsu, R. Lau, B. P. Bowen, S. Hromada, et al. 2018. Deciphering microbial interactions in synthetic human gut microbiome communities. Molecular Systems Biology 14:e8157.

Wang, Y.-H., C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo. 2014. On the computational complexity of the empirical mode decomposition algorithm. Physica A: Statistical Mechanics and its Applications 400:159–167.

Ward, D. L., A. F. Casper, T. D. Counihan, J. M. Bayer, I. R. Waite, J. J. Kosovich, C. G. Chapman, et al. 2017. Long-term fish monitoring in large rivers: Utility of "benchmarking" across basins. Fisheries 42:100–

475 114.

476 Wootton, J. T., and M. Emmerson. 2005. Measurement of interactions streength in nature. Annual Review

477 of Ecology, Evolution, and Systematics 36:419–444.

478

Figure Legends

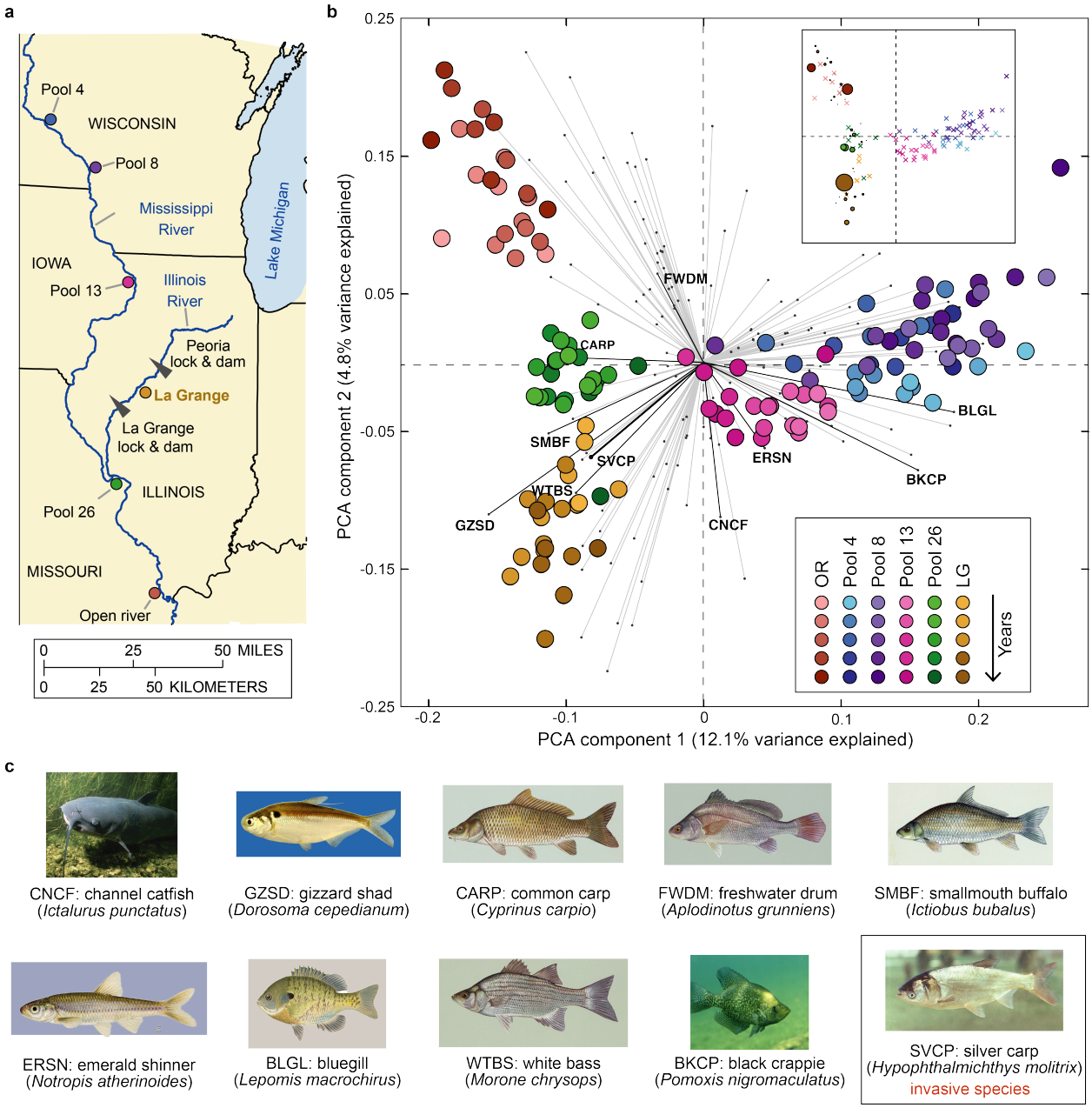


Figure 1: The freshwater fish community varies spatially and temporally in the La Grange pool and nearby water bodies. a, Geographical locations of the six stations monitored by the Long Term Resource Monitoring Program along the Upper Mississippi and Illinois Rivers. b, Biplot of principle component analysis (PCA). Each circle ("score") represents the species abundance distribution of fish community associated with a site and year combination. The color brightness of circles indicates the passage of time:

lighter colors represent earlier data. Each line ("loading vector") represents contribution of an explanatory variable (fish species) to the variance of first two principle components. For all loading vectors, the top 9 dominant fish species in La Grange pool plus silver carp are colored in black while all others are colored in light gray. The inset is the same PCA score plot, but the circle size is scaled to be proportional to the abundance of invasive silver carp (samples missing silver carp are represented with crosses). c, Common names, abbreviations, and species names of the 10 fish species selected to investigate in our study. Fish images were obtained through public domain resources except for silver carp licensed by CC BY 3.0 and gizzard shad provided by Chad Thomas pf Texas State University.

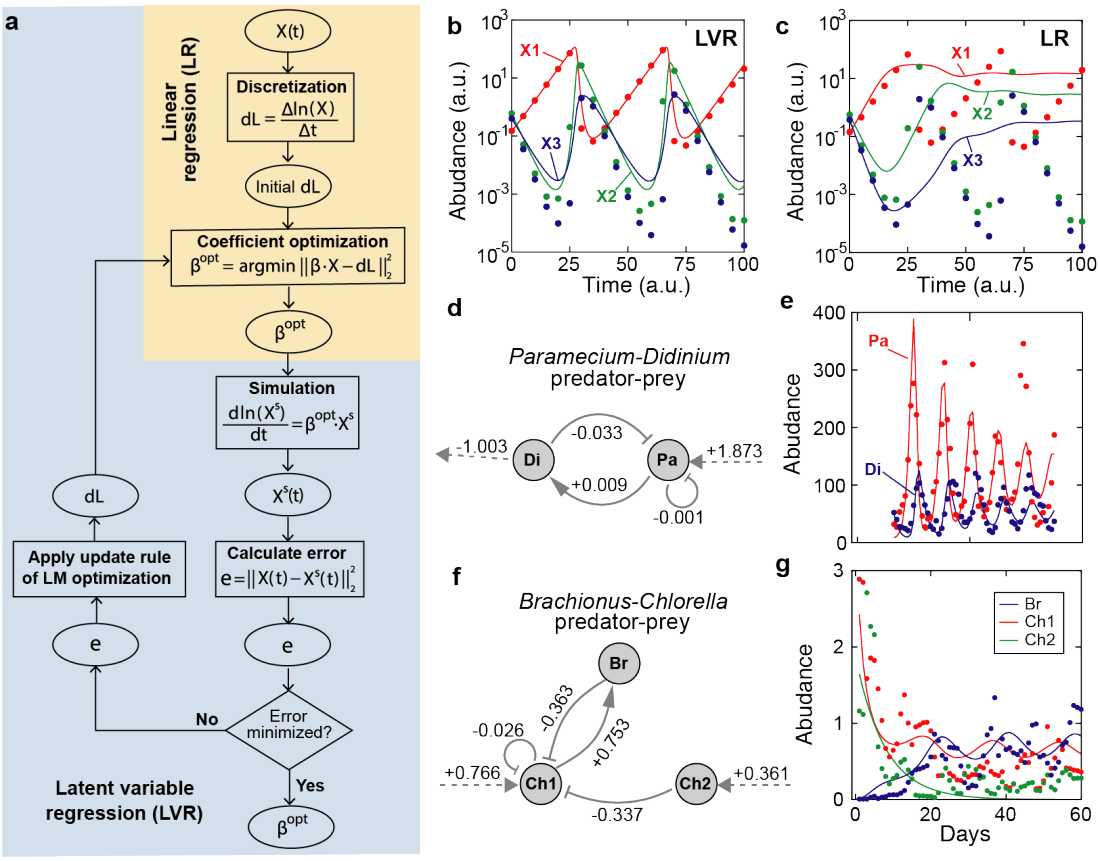
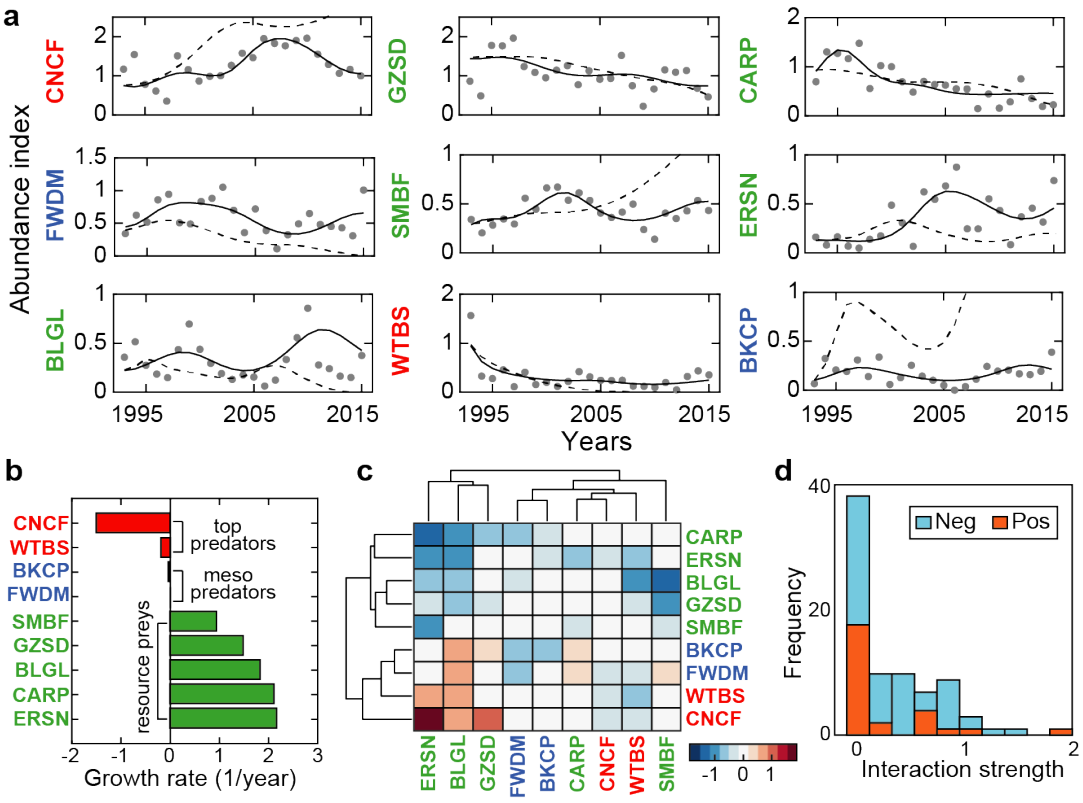


Figure 2: Latent variable regression improves network inference substantially compared to simple linear regression. a, A flowchart showing how linear regression (LR; shaded in light yellow) is expanded by the latent variable regression (LVR; shaded in light blue). $X(t)$: observed time series; $X^s(\mathrm{t})$: simulated time series; β: parameters of the generalized Lotka-Volterra model; dL: time-derivatives of $\ln(X(t))$. The key difference between LR and LVR is that dL in LVR are iteratively learned, while they are numerically approximated and fixed in LR. LM: Levenberg-Marquardt. b,c, Comparison between LVR (b) and LR (c) in fitting synthetic time series generated by a 3-species predator-prey Lotka-Volterra model. d-g, Effectiveness of the LVR method to infer known interactions. d,e, The protozoan predator (*Didinium nasutum*)-prey (*Paramecium aurelia*) ecosystem. f,g, The ecosystem of a rotifer predator (*Brachionus calyciflorus*) and two algae preys (*Chlorella vulgaris*). Lotka-Volterra networks obtained using LVR for the two ecosystems are shown in d and f, where solid links represent interactions (point end for positive effect and blunt end for negative effect) and dashed links represent population growth (incoming links for positive growth rate and outgoing links for negative growth rate). Interaction strengths and population

growth rates are indicated along the network links. e and g show good agreements between observed population dynamics (dots) and model predictions (lines). The first 13 data points of *Didinium nasutum*-*Paramecium Aurelia* dynamics were removed to eliminate the initial transient period. Unit for the y-axis: individuals/mL (e); 10 individual females/mL for the rotifer and 10^6 cells/mL for the algae (g).



512

513 **Figure 3**: Ecological network inferred for the freshwater fish community of the La Grange pool of the
514 Illinois River. a, Comparison of model predictions (lines) with fish population data (gray dots) for the 9
515 dominant native fish species in La Grange pool. Dashed lines: linear regression; solid lines: latent variable
516 regression. b, Inferred population growth rates. c, Clustering of inferred pairwise interaction coefficient
517 matrix. Each coefficient in row i and column j represents the effect of fish species j on species i.
518 Coefficients that are greater, equal to, or smaller than zero stand for positive, neutral, or negative effects,
519 respectively. In a-c, fish species are colored based on their trophic levels (green for resource preys, blue for
520 mesopredators, red for top predators). d, Distribution of interaction strengths shown in c.

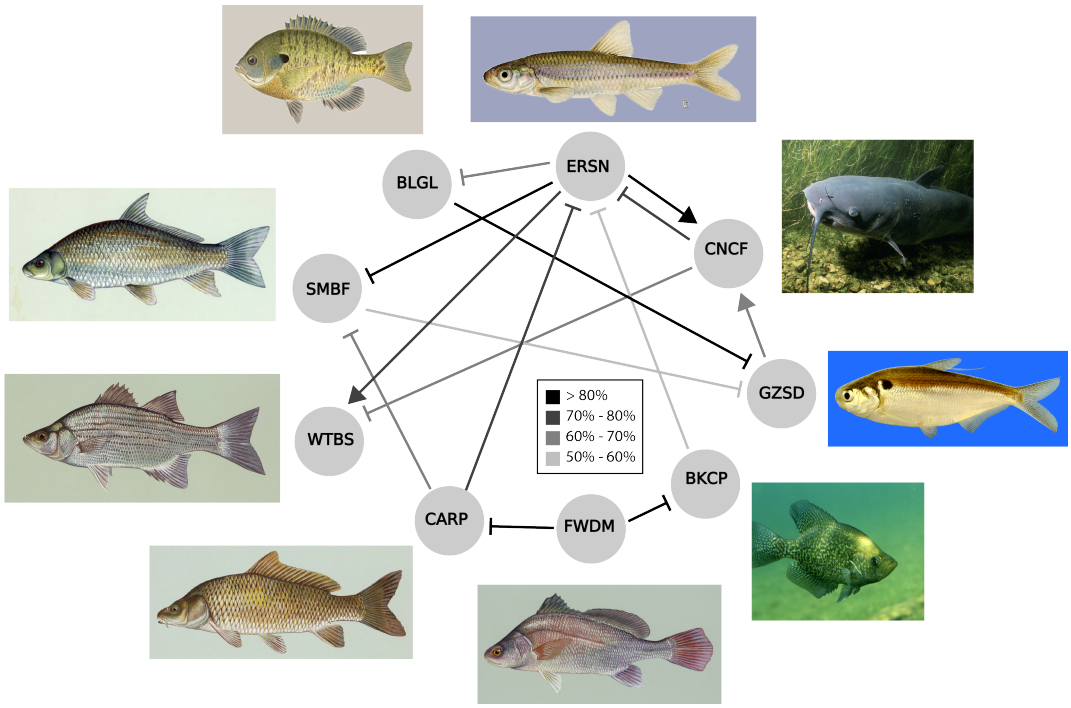


Figure 4: Core subnetwork of interspecific interactions. Point arrows represent positive effects and blunt arrows represent negative effects. The black percentage of each link corresponds to the interval that the smaller confidence score of the link inferred among La Grange pool and Pool 26 falls. The darker the links, the more significant the interactions.

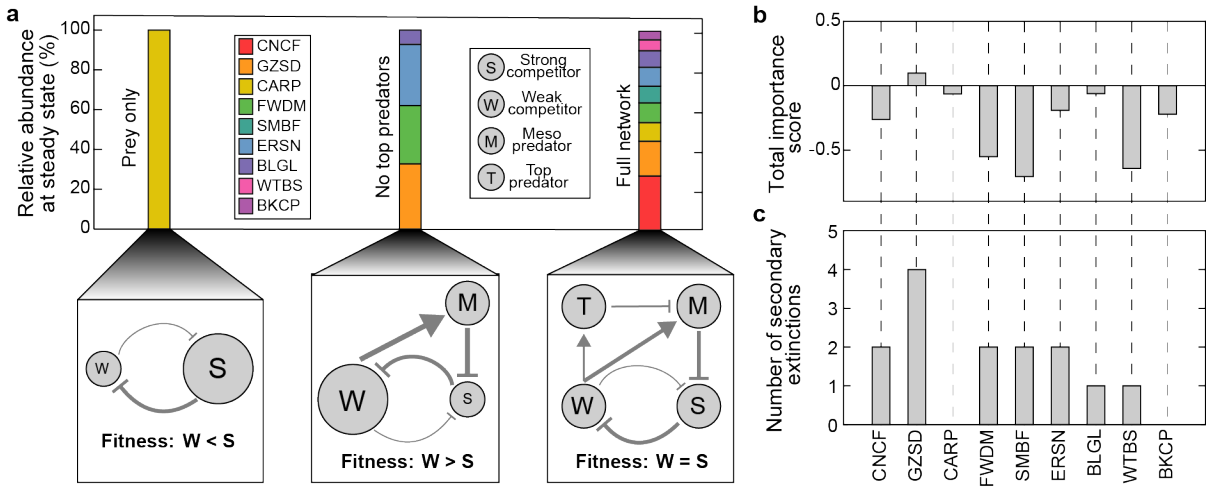


Figure 5: The gizzard shad is a keystone species essential for coexistence. a, Coexistence relies on a fitness-equalizing mechanism. The stacked bars show steady state relative abundance of fish species in subnetworks consisting of only resource preys (upper left) and resource preys plus mesopredators (upper middle) as well as the full network (upper right). The schematic diagrams at the bottom illustrate how fitness between strong and weak competitors among the resource preys are equalized by incorporating predators. b, Topological importance of fish species. The importance score of a species reflects its total impacts on the fitness of all other species in the network. Species with positive scores play overall facilitating roles that stabilize species coexistence. In contrast, species with negative scores tend to exclude other competitors. c, Assessment of effects of single species removal (extinction) on coexistence of others in the community. A species is more important if its removal causes higher number of secondary extinctions.

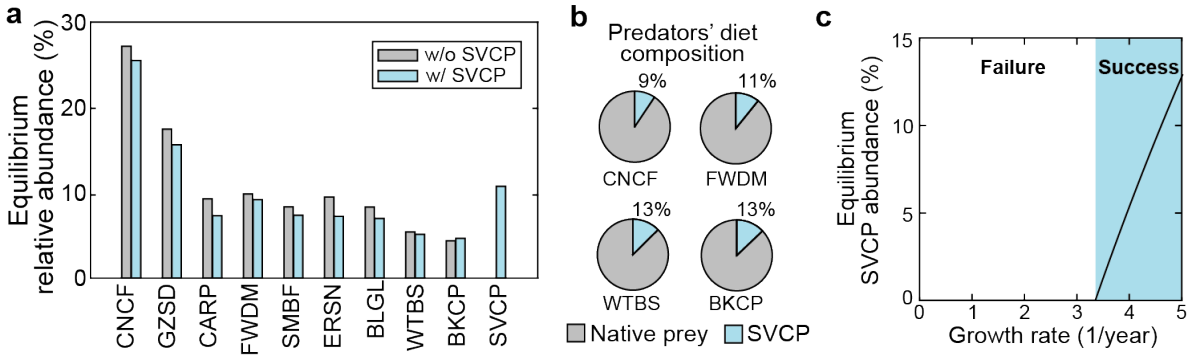


Figure 6: The silver carp invades native fish community by outgrowing other prey. a, Steady state relative abundance of fish species in the absence (gray bars) and presence (blue bars) of silver carp. b, Silver carp as supplemental food sources to native predators. The percentage contribution of a prey to its predator's diet is proportional to the pairwise interaction term describing the positive influence of this prey on the predator. c, Linear-threshold dependence of silver carp's invasion success on its population growth rate.

Supplementary Information for:

Ecological network inferred from field data elucidates invasion of a freshwater ecosystem by the exotic silver carp

Chen Liao[1], Joao B. Xavier[1,*] and Zhenduo Zhu[2,*]

[1]Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, United States of America, [2]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY 14260, United States

[*]To whom correspondence may be addressed: J. Xavier (Email: XavierJ@mskcc.org); Z. Zhu (Email: zhenduoz@buffalo.edu)

Keywords:

ecological network; Lotka-Volterra equations; inference; coexistence mechanism; invasive species

1. Ecological interpretations of inferred fish interaction network

- **The positive interaction from ERSN to CNCF; the negative interaction from CNCF to ERSN; the positive interaction from GZSD to CNCF.** Emerald shiner is a small fish species feeding on a variety of zooplankton, protozoans and diatoms. Gizzard shad is also small, and its diet consist of phytoplankton, zooplankton and detritus[1]. In contrast, channel catfish is an omnivore: although young catfish feed on vegetation and insects, adult channel catfish begin to use other fish as part of the diet[2]. So abundant gizzard shads and emerald shiners may provide a forage base to support growth of channel catfish[3].

- **The positive interaction from ERSN to WTBS; the negative interaction from CNCF to WTBS.** Similar to channel catfish, young white bass feed on zooplankton and small invertebrates but its adults are piscivorous. Since both white bass and channel catfish consume small fish such as emerald shiner and gizzard shad[4,5], the two species may have niche overlap and negative impact each other due to competition.

- **The negative interaction from BLGL to GZSD.** Bluegill is an omnivore and its diet consist of insect larvae, crayfish, leeches, snails and other very small fish. Similar to gizzard shad and emerald shiner, bluegill is prey to many larger predator fish. Gizzard shad is known to compete with bluegill for food resources[6,7].

- **The negative interaction from ERSN to BLGL.** The interaction between emerald shiner and bluegill remains to be found. Given that gizzard shad compete with both fish species for food[6–8], it is very likely that emerald shiner and bluegill have niche overlap and compete for similar resources.

- **The negative interaction from FWDM to CARP.** Freshwater drum generally eat zooplankton when they are young but start to feed on inserts and fish in adult ages[9]. It is very interesting that some people confuse drum with common carp because the drum's underslung mouth makes many people believe it is a bottom feeder, just like the common carp. Although freshwater drum also chase prey in open water, it is likely that both bottom feeders compete for food and space near the bottom of water.

- **The negative interaction from FWDM to BKCP.** Black crappies mainly eat plankton and crustaceans when they are young and larger individuals are basically piscivorous and primarily feed on small fish. Since adult black crappie and freshwater drum share common preys such as gizzard shad[10,11], their competitions for food resources seem to be unavoidable.

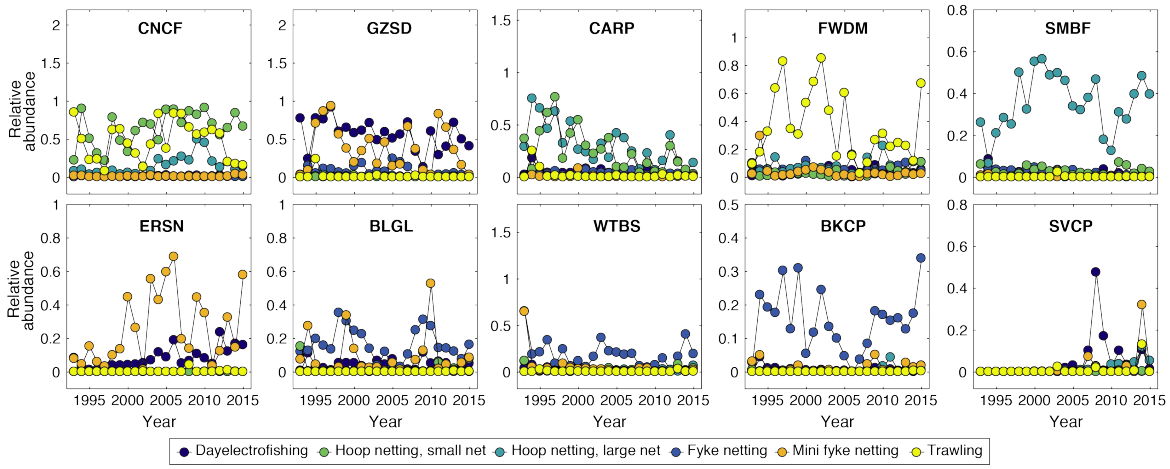
- **The negative interaction from BKCP to ERSN.** Adult black crappies prefer forage fish and minnows such as gizzard shad and emerald shiner. In fact, they can feed on anything that fits into their mouths. Emerald shiner is a small fish with a typical length of 8.6 cm and the small size makes it a bait used by anglers for fishing crappie.

- **The negative interaction from SMBF to GZSD; the negative interaction from ERSN to SMBF.** Smallmouth buffalo is a detritivore and uses its ventral sucker mouth to eat vegetation, insets and other organisms from the bottom of a body of water. The diet of adult smallmouth buffalo contains 55% of zooplankton and 31% phytoplankton[12]. Since gizzard shad and emerald shiner also feed on plankton, they may compete with smallmouth buffalo for diet and habitat due to the diet overlap. It was observed that the diet between smallmouth buffalo and gizzard shad do have certain overlaps, despite the relative proportions of detritus and zooplankton differ[13].

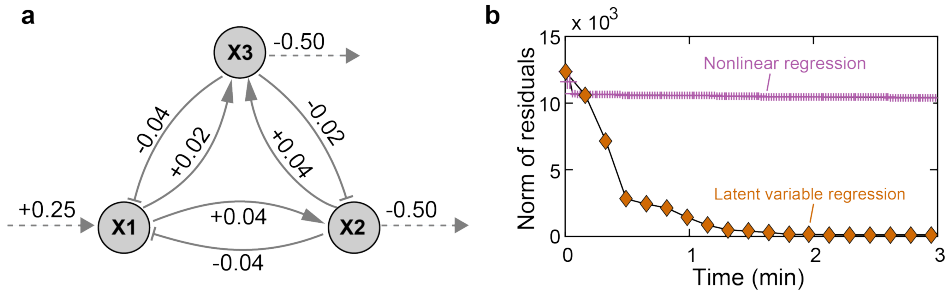
- **The negative interaction from CARP to ERSN; the negative interaction from CARP to SMBF.** Common carp was first introduced to the United States since 1800s and became an invasive species.

The invasion of common carp has caused damages to native fish population due to competition for food and space[14]. Common carp eat nearly anything, including aquatic plants, mollusks, insects and debris[15]. They also reproduce rapidly and can survive in poor water quality[16]. The inhibitions of common carp on emerald shiner and smallmouth buffalo are expected because invasive common carp are more competitive than native consumers.

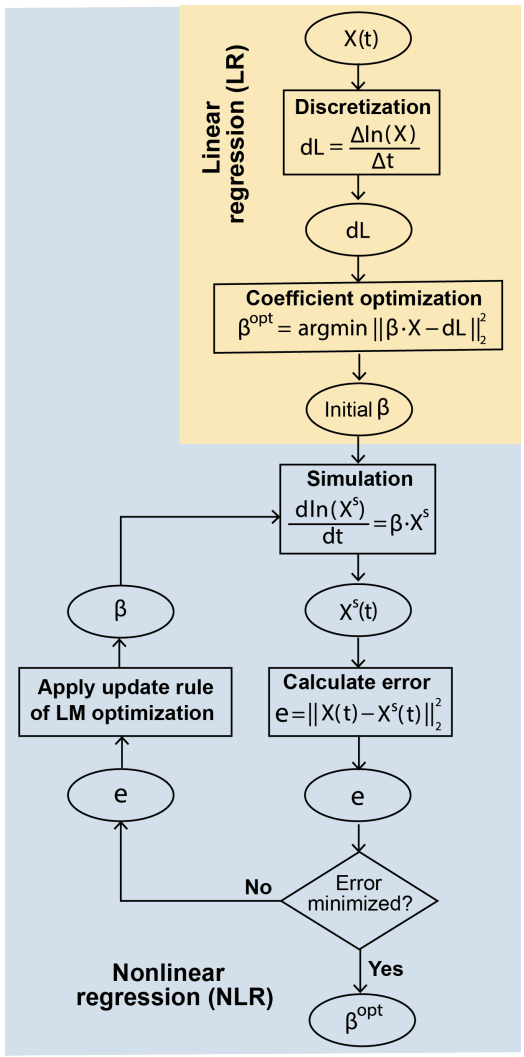
2. Supplementary Figures



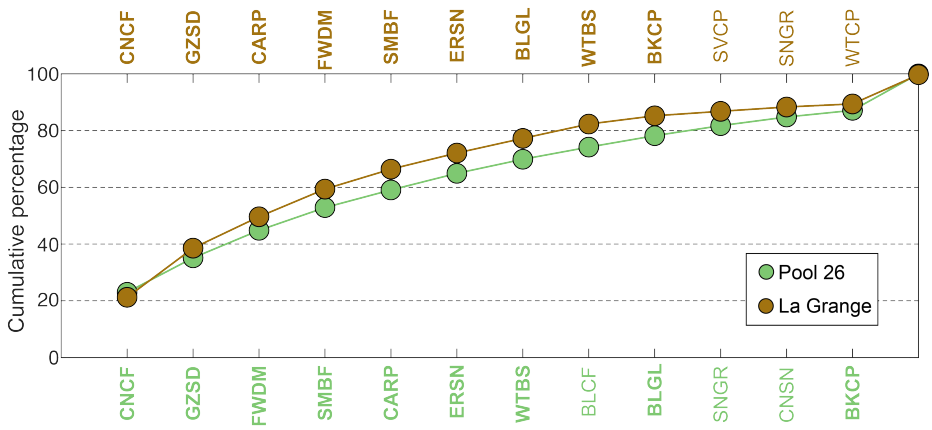
Supplementary Figure 1 Relative abundance (normalized among species) from multiple fishing gears for the dominant fish species in La Grange pool. Different fishing gears are selective for certain types of fish and no gear can catch all of them, suggesting the need to combine fish samples from different gears.



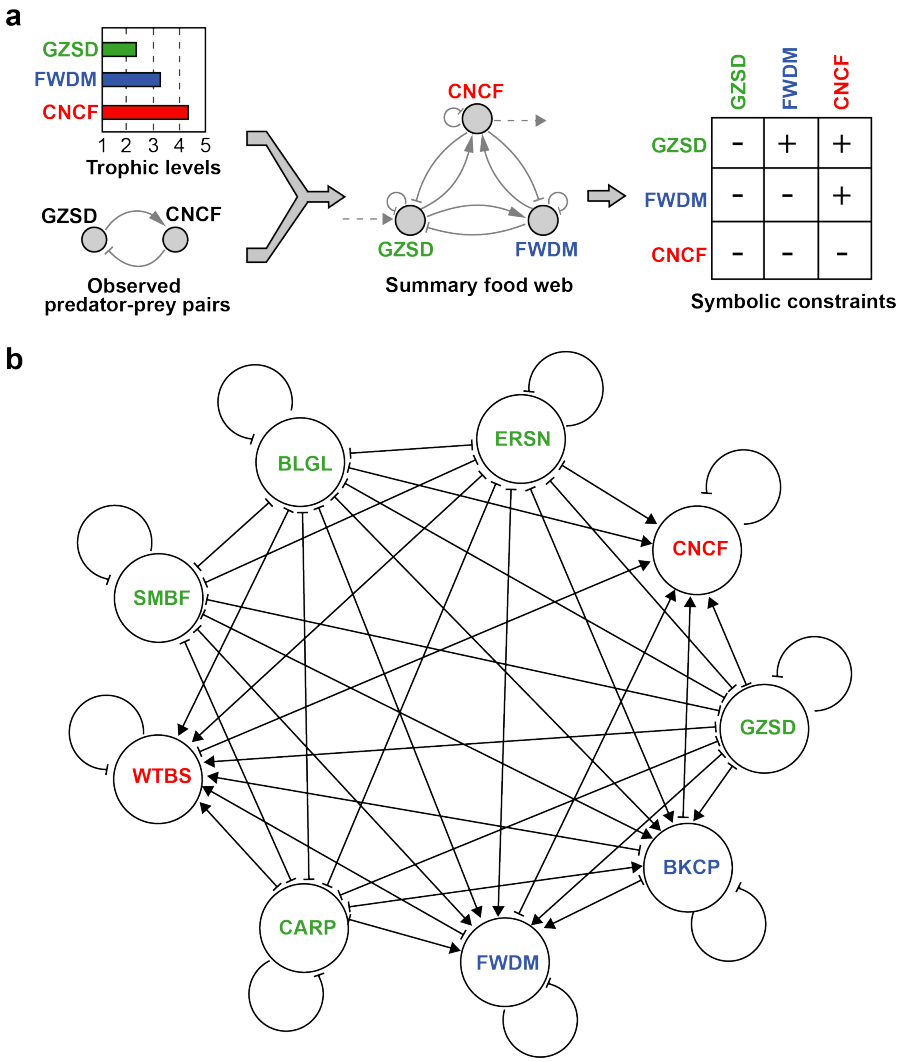
Supplementary Figure 2 a, 3-species predator-prey Lotka-Volterra model used to generate synthetic data in Fig. 2b,c in the main text. Solid links represent interactions (point end for positive effect and blunt end for negative effect) and dashed links represent population growth (incoming links for positive growth rate and outgoing links for negative growth rate). Interaction strengths and population growth rates are indicated along the network links. **b,** Comparison between latent variable regression and nonlinear regression for their convergence rates. Each symbol (diamond and cross) marks the time of completion of one optimization iteration.



Supplementary Figure 3 Flow chart of nonlinear regression. $X(t)$: observed time series; $X^s(\mathrm{t})$: simulated time series; β: parameters of the GLV model; dL: time-derivatives of $\ln(X(t))$. Nonlinear regression takes the output of linear regression as initial guesses of β and adaptively update β until the difference between observed and simulated time series is minimized. LM: Levenberg-Marquardt.

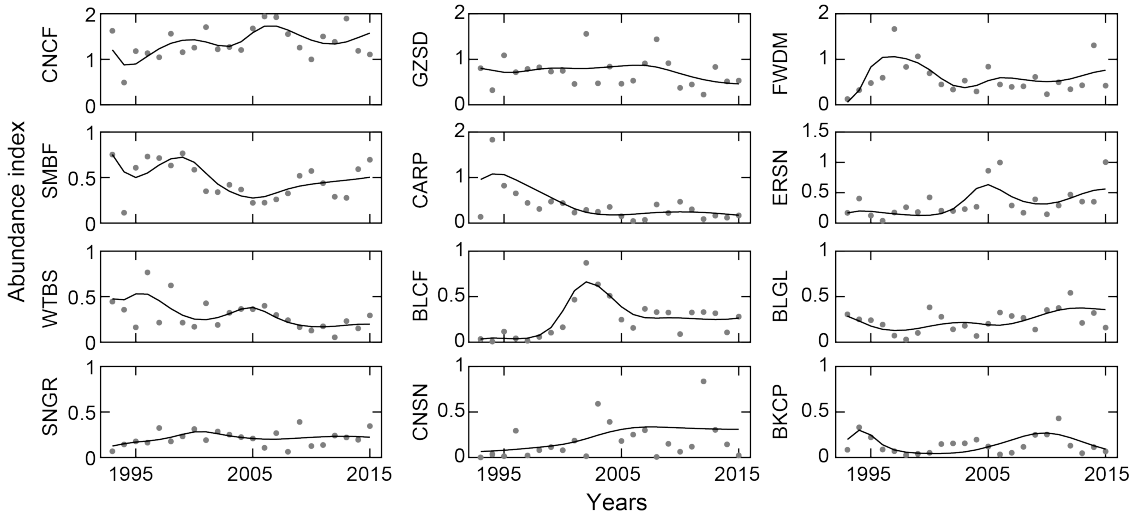


Supplementary Figure 4 Cumulative percentage of the averaged abundance index between 1993 and 2015. The top 12 dominant fish species at both La Grange pool and Pool 26 are shown in order of most to least abundant, where fish species common to both sites are made bold.

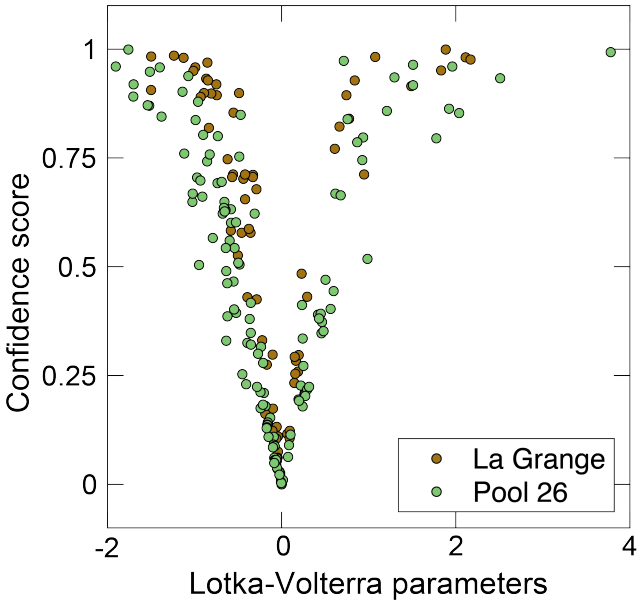


Supplementary Figure 5 a, Schematic diagrams showing steps to identify symbolic constraints of interactions from literature data. First, a summary food web consisting of all potential interactions is reconstructed based on Fishbase (http://www.fishbase.org) trophic level indices as well as experimentally observed trophic interactions. Second, negative, neutral and positive interactions in the food web are converted to symbolic constraints of the interaction coefficients in the generalized Lotka-Volterra (GLV) model. For example, a positive interaction requires its corresponding GLV interaction coefficient to be positive as well. See Methods in the main text for details. **b,** Reconstructed summary food web for the 9 dominant fish species in La Grange pool. Point arrows represent positive effects and blunt arrows represent negative effects. The fish name abbreviations are colored according to their trophic levels: green for resource preys, blue for mesopredators, and red for top predators. The literature reported predator-prey relationships used to identify potential interactions include: BKCP-BLGL[17],

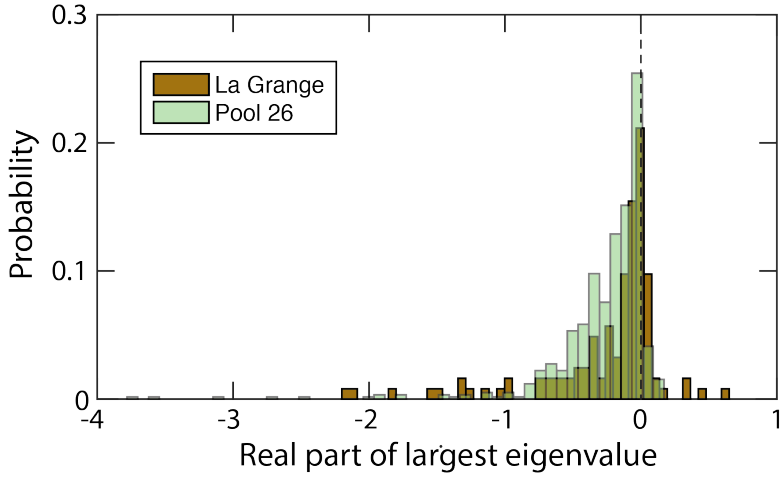
CNCF-BLGL[18], CNCF-ERSN[3], CNCF-GZSD[3], FWDM-GZSD[19], WTBS-BKCP[20], WTBS-BLGL[20], WTBS-FWDM[20], WTBS-ERSN[21], WTBS-GZSD[21], WTBS-CARP[22] (the former species is a predator and the latter species is a prey).



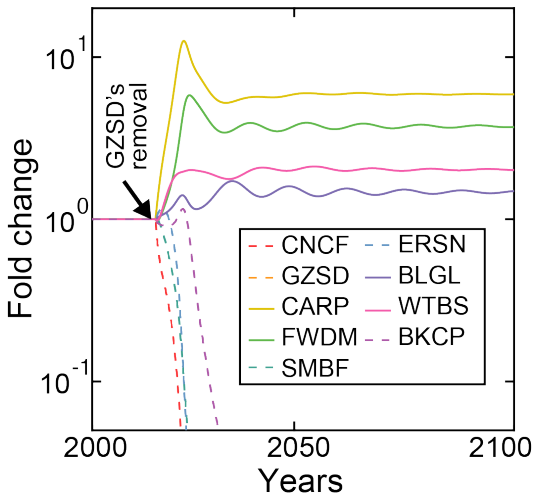
Supplementary Figure 6 Comparison of model predictions (lines) with experimentally observed abundance indices (gray dots) for the 12 dominant fish species in Pool 26.



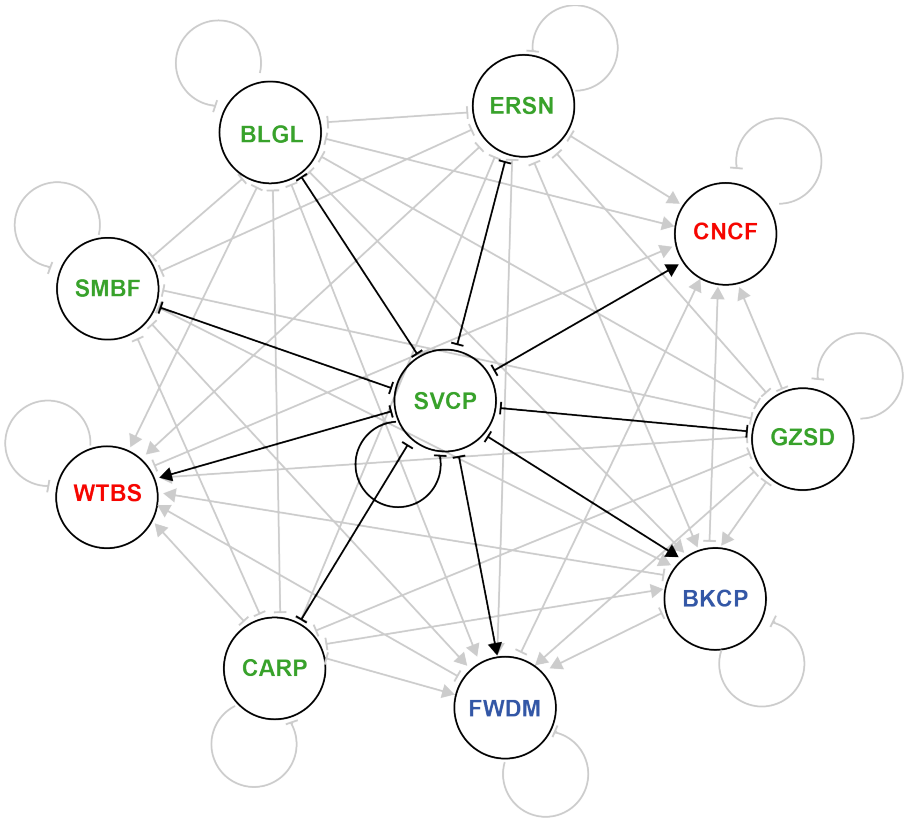
Supplementary Figure 7 General proportionality between the absolute values of the Generalized Lotka-Volterra (GLV) model parameters and their corresponding confidence scores, which were calculated as the minimum significance levels above which the confidence intervals of the GLV parameters do not contain 0. The use of confidence score allows for fair comparisons of significance levels of GLV inferences among different geographical locations.



Supplementary Figure 8 Dynamic stability of alternative steady states for fish communities in La Grange pool and Pool 26. The histograms represent distribution of the largest real eigenvalues of all alternative steady states. A N-species system has at most 2^N (marginally) steady states, which correspond to all possible permutations of species removals, including one state for total extinction, one state for total coexistence, and the other $2^N - 2$ states for partial coexistence with various numbers of extinct species. For a steady state to be stable, its largest real eigenvalue must be negative.



Supplementary Figure 9 Fold change of fish abundance indices in the absence of gizzard shad to those when gizzard shad is present. Black arrow indicates the time point gizzard shad is removed from the network. Dashed lines represent time courses of fish species that go extinction after gizzard shad's removal.



Supplementary Figure 10 Potential interactions between silver carp and native fish species in La Grange pool (black links). The fish name abbreviations are colored according to their trophic levels: green for resource preys, blue for mesopredators, and red for top predators. The literature reported predator-prey relationships used to identify potential interactions include: CNCF-SVCP[23], WTBS-SVCP[21], BKCP-SVCP[23] (the former species is a predator and the latter species is a prey). Gray links represent interactions among/within native fish species.

3. Supplementary Tables

Scientific name	Common name	Abbr.	Fishbase trophic level index	Trophic level	Species in La Grange model?	Species in Pool 26 model?
Hypophthalmichthys molitrix	Silver Carp	SVCP	2.0 ± 0.00	Resource prey	Yes	No
Dorosoma cepedianum	Gizzard Shad	GZSD	2.4 ± 0.21	Resource prey	Yes	Yes
Notropis atherinoides	Emerald Shiner	ERSN	2.8 ± 0.29	Resource prey	Yes	Yes
Notropis wickliffi	Channel Shiner	CNSN	2.9 ± 0.3	Resource prey	Yes	Yes
Ictiobus bubalus	Smallmouth Buffalo	SMBF	3.0 ± 0.39	Resource prey	Yes	Yes
Cyprinus carpio	Common Carp	CARP	3.1 ± 0.0	Resource prey	Yes	Yes
Lepomis macrochirus	Blue Gill	BLGL	3.2 ± 0.2	Resource prey	No	Yes
Ictalurus furcatus	Blue Catfish	BLCF	3.4 ± 0.44	Mesopredator	Yes	Yes
Aplodinotus grunniens	Freshwater Drum	FWDM	3.4 ± 0.43	Mesopredator	Yes	Yes
Pomoxis nigromaculatus	Black Crappie	BKCP	3.8 ± 0.62	Mesopredator	No	Yes
Lepisosteus platostomus	Shortnose Gar	SNGR	3.9 ± 0.7	Mesopredator	No	Yes
Morone chrysops	White Bass	WTBS	4.0 ± 0.68	Top predator	Yes	Yes
Ictalurus punctatus	Channel Catfish	CNCF	4.2 ± 0.3	Top predator	Yes	Yes

Supplementary Table 1 Summary of fish species included in our study. Fishbase is a global database that covers more than 32, 000 fish species (http://www.fishbase.org). Fish species are classified as resource prey, mesopredator, or top predator based on their Fishabse trophic level indices as well as interaction data about how they interact with each other (see Methods in the main text for details).

	CNCF	GZSD	FWDM	SMBF	CARP	ERSN	WTBS	BLCF	BLGL	SNGR	CNSN	BKCP	Growth
CNCF	-1	1	1	0	0	1	-1	1	1	1	0	1	-1
GZSD	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
FWDM	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	-1
SMBF	0	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	1
CARP	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
ERSN	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
WTBS	-1	1	1	0	1	1	-1	1	1	1	0	1	-1
BLCF	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	-1
BLGL	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
SNGR	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	-1
CNSN	0	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	1
BKCP	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	-1

Supplementary Table 2 Symbolic constraints used to parameterize generalized Lotka-Volterra model for fish community in La Grange pool. The shaded matrix was used to constrain pairwise interaction coefficients and the last column was used to constrain population growth rates. For species interaction coefficients, -1, 0, 1 represent negative, neutral, and positive interaction from the column species to the row species respectively. For population growth rates (the last column), -1 means negative growth rate and 1 means positive growth rate.

	CNCF	GZSD	CARP	FWDM	SMBF	ERSN	BLGL	WTBS	BKCP	Growth
CNCF	-0.489	1.075	0.000	0.150	0.000	1.885	0.744	-0.287	0.092	-1.500
GZSD	-0.104	-0.440	-0.003	-0.001	-0.892	-0.358	-0.750	-0.376	0.000	1.489
CARP	0.000	-0.746	0.000	-0.853	-0.001	-1.237	-0.866	0.000	-0.394	2.113
FWDM	-0.290	0.163	0.229	-0.554	0.188	0.157	0.839	-0.224	-0.107	-0.001
SMBF	0.000	-0.043	-0.420	-0.001	-0.460	-0.992	-0.100	0.000	-0.022	0.947
ERSN	-0.329	-0.171	-0.558	-0.001	-0.047	-1.014	-1.127	-0.836	-0.503	2.170
BLGL	-0.043	-0.062	-0.007	-0.420	-1.499	-0.568	-0.847	-0.931	-0.003	1.832
WTBS	-0.331	0.145	0.000	0.065	0.000	0.774	0.664	-0.621	0.009	-0.188
BKCP	-0.057	0.196	0.293	-0.809	0.003	0.094	0.610	0.000	-0.584	-0.003

Supplementary Table 3 Optimized coefficients of the generalized Lotka-Volterra model for fish community in La Grange pool. The shaded area gives pairwise interaction coefficients (each cell represents the interaction from column species to row species) and the last column gives population growth rates.

	CNCF	GZSD	FWDM	SMBF	CARP	ERSN	WTBS	BLCF	BLGL	SNGR	CNSN	BKCP	Growth
CNCF	-0.487	0.617	0.713	0.000	0.000	0.757	-0.244	0.241	0.262	0.294	0.000	0.199	-0.623
GZSD	-0.016	-0.737	-0.311	-0.484	-0.354	-0.213	0.000	-0.524	-1.521	-0.062	-0.546	-0.642	1.923
FWDM	-0.023	0.193	-1.761	0.926	1.296	1.509	-0.065	-0.989	0.457	-0.051	0.271	-0.971	-0.086
SMBF	0.000	-0.677	-0.001	-1.514	-0.581	-0.901	0.000	-1.400	-0.791	-0.243	0.000	-0.933	2.510
CARP	0.000	-0.186	-0.471	-0.856	-0.657	-0.204	-0.005	-1.139	-0.625	-1.025	-1.022	-0.064	2.039
ERSN	-0.960	-0.598	-0.733	-0.552	-0.827	-0.689	-0.165	-0.541	-0.523	-0.085	-0.172	-1.906	3.779
WTBS	-0.583	0.418	0.234	0.000	0.678	0.936	-1.119	0.868	0.597	0.315	0.000	0.074	-0.408
BLCF	-0.091	1.209	-0.056	1.960	0.221	0.250	-1.701	-0.498	0.014	-0.001	0.241	-1.380	-0.949
BLGL	-0.003	-0.355	-0.237	-0.160	-0.645	-0.664	-0.636	-0.654	-1.539	-0.078	-0.004	-0.214	1.777
SNGR	-0.131	0.504	0.000	0.449	0.102	0.084	-0.164	-0.271	0.264	-0.151	0.481	-0.100	-0.451
CNSN	0.000	-0.091	-0.027	-0.394	-0.367	-0.080	0.000	-0.056	-0.168	-0.054	-1.704	-0.026	0.986
BKCP	-0.005	1.509	-1.072	0.463	0.009	0.429	-0.283	-0.353	0.000	-0.008	0.562	-0.910	-0.636

Supplementary Table 4 Optimized coefficients of the generalized Lotka-Volterra model for fish community in Pool 26. Note that Pool 26 model has 12 fish species, including the 9 dominant species in La Grange pool. The shaded area gives pairwise interaction coefficients (each cell represents the interaction from column species to row species) and the last column gives population growth rates.

	CNCF	GZSD	CARP	FWDM	SMBF	ERSN	BLGL	WTBS	BKCP
Impacts of SVCP on native fish species	0.360 ± 0.198	-0.120 ± 0.079	-0.206 ± 0.135	0.142 ± 0.095	-0.201 ± 0.115	-0.210 ± 0.149	-0.147 ± 0.100	0.176 ± 0.130	0.140 ± 0.121
Impacts of native fish species on SVCP	-0.474 ± 0.304	-0.683 ± 0.539	-0.728 ± 0.731	-0.267 ± 0.338	-0.409 ± 0.513	-0.439 ± 0.539	-1.966 ± 0.732	-0.788 ± 0.963	-0.936 ± 1.051
Growth rate of SVCP	4.703 ± 0.377								
Self-interaction within SVCP	-2.507 ± 1.125								

Supplementary Table 5 Estimated median and standard deviation of Lotka-Volterra model parameters related to silver carp using Markov Chain Monte Carlo (MCMC) simulation. These parameters include 9 coefficients for impacts of silver carp on native fish species (second row), 9 coefficients for impacts of native fish species on silver carp (third row), one coefficient corresponding to population growth rate (fourth row), and one coefficient corresponding to intraspecific interaction (last row). However, there are only 16 data points in the time series of silver carp after 2000, when its population started to increase exponentially[24]. Due to insufficient data, the model parameters estimated for the silver carp invasion have high uncertainty: the variances of some parameters are as large as their median values. Nonetheless, we could obtain best guesses by sampling their posterior distributions using MCMC.

4. Supplementary References

1. Michaletz, P. H. Gizzard shad population dynamics in eutrophic Missouri reservoirs with emphasis on environmental influences on their growth. *J. Freshw. Ecol.* **27,** 185–197 (2012).
2. Busbee, R. L. Piscivorous activities of the channel catfish. *Prog. Fish Cult.* **30,** 32-34 (1968).
3. Hilling, C. D., Welsh, S. A. & Smith, D. M. Age, growth, and fall diet of channel catfish in cheat lake, west virginia. *J. Fish Wildl. Manag.* **7,** 304–314 (2016).
4. Bonn, E. W. The food and growth rate of young white bass (*Morone chrysops*) in Lake Texoma. *Trans. Am. Fish Soc.* **82,** 213–221 (1953).
5. Russell, R. W., Gobas, F. A. P. C. & Haffner, G. D. Role of chemical and ecological factors in trophic transfer of organic chemicals in aquatic food webs. *Environ. Toxicol. Chem.* **18,** 1250–1257 (1999).
6. Partridge, D. G. & DeVries, D. R. Regulation of growth and mortality in larval bluegills: implications for juvenile recruitment. *Trans. Am. Fish Soc.* **128,** 625–638 (1999).
7. Aday, D. D., Hoxmeier, R. J. H. & Wahl, D. H. Direct and indirect effects of gizzard shad on bluegill growth and population size structure. *Trans. Am. Fish Soc.* **132,** 47–56 (2003).
8. Hartman, K. J., Vondracek, B. & Parrish, D. L. Diets of emerald and spottail shiners and potential interactions with other western Lake Erie planktivorous fishes. *J. Great Lakes* **18,** 43-50 (1992).
9. Griswold, B. L. & Tubb, R. A. Food of yellow perch, white bass, freshwater drum, and channel catfish in Sandusky Bay, Lake Erie. *Ohio J. Sci.* **77,** 43–47 (1977).
10. Kutkuhn, J. H. Food and feeding habits of some fishes in a dredged Iowa lake. *Proc. Iowa Acad. Sci.* **62**, 576-588 (1955).
11. Davis, S. L. Long-term changes in population statistics of Freshwater Drum (*Aplodinotus grunniens*) in Lake Winnebago, Wisconsin, using otolith growth chronologies and bomb radiocarbon age validation. *PhD diss.*, University of Wisconsin-Milwaukee (2012).
12. Walburg, C. H. & Nelson, W. R. Carp, river carpsucker, smallmouth buffalo, and bigmouth buffalo in Lewis and Clark Lake, Missouri River. Bureau of Sport Fisheries and Wildlife (1966).

13. Gido, K. B. Feeding ecology of three omnivorous fishes in Lake Texoma (Oklahoma-Texas). *Southwest. Nat.* **46,** 23–33 (2001).

14. Weber, M. J. & Brown, M. L. Relationships among invasive common carp, native fishes and physicochemical characteristics in upper Midwest (USA) lakes. *Ecol. Freshw. Fish* **20,** 270–278 (2011).

15. Mangi, A. & Memon, Z. Analysis of gut contents of Common carp (*Cyprinus carpio*) in district Larkana, Sindh, Pakistan. *J. Entomol. Zool. Stud.* **5,** 2631–2634 (2017).

16. Weber, M. J. & Brown, M. L. Effects of Common Carp on Aquatic Ecosystems 80 Years after "Carp as a Dominant": Ecological Insights for Fisheries Management. *Rev. Fish. Sci.* **17,** 524–537 (2009).

17. Liao, H., Pierce, C. L. & Larscheid, J. G. Diet dynamics of the adult piscivorous fish community in Spirit Lake, Iowa, USA 1995-1997. *Ecol. Freshw. Fish* **11,** 178–189 (2002).

18. Doxtater, G. Experimental predator-prey relations in small ponds. *Prog. Fish Cult.* **29,** 102-104 (1967).

19. Wrenn, W. B. & Shoals, M. Life history aspects of smallmouth buffalo and freshwater drum in Wheeler Reservoir, Alabama. *Proc. Southeast. Assoc. Game* **22,** 479-495 (1968).

20. Sigler, W. F. Life history of the white bass, Lepibema chrysops (Rafinesque), of Spirit Lake, Iowa. *Res. Bull. (Iowa Agric. Home Econ. Exp. Stn.)* **29**, 1 (1949).

21. Wolf, M. C. & Phelps, Q. E. Prey selectivity of common predators on Silver carp (*Hypophthalmichthys molitrix*): controlled laboratory experiments support field observations. *Environ. Biol. Fishes* **100,** 1–5 (2017).

22. Olson, N. W., Guy, C. S. & Koupal, K. D. Interactions among three top-level predators in a polymictic Great Plains reservoir. *N. Am. J. Fish Manag.* **27,** 268–278 (2007).

23. Anderson, C. A. *Diet Analysis of Native Predatory Fish to Investigate Predation of Juvenile Asian Carp*. PhD thesis, Western Illinois Univ. (2016).

24. Irons, Kevin S., et al. Reduced condition factor of two native fish species coincident with invasion of non-native Asian carps in the Illinois River, USA. Is this evidence for competition and reduced fitness? *J. Fish Biol.* **71**, 258-273 (2007).