



HAL
open science

Addressing the Eltonian shortfall with trait-based interaction models

Dominique Caron, Luigi Maiorano, Wilfried Thuiller, Laura Pollock

► **To cite this version:**

Dominique Caron, Luigi Maiorano, Wilfried Thuiller, Laura Pollock. Addressing the Eltonian shortfall with trait-based interaction models. *Ecology Letters*, 2022, 25 (4), pp.889-899. 10.1111/ele.13966 . hal-04757567

HAL Id: hal-04757567

<https://hal.science/hal-04757567v1>

Submitted on 30 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Addressing the Eltonian shortfall with trait-based interaction models**

2 **Dominique Caron^{1,2}, Luigi Maiorano³, Wilfried Thuiller⁴, Laura J. Pollock^{1,2}**

3 **Affiliations:**

4 ¹ Department of Biology, McGill University, Montreal, QC, Canada

5 ² Quebec Centre for Biodiversity Sciences, Montreal, QC, Canada

6 ³ Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, viale
7 dell'Università 32, 00185 Rome, Italy

8 ⁴ Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Ecologie Alpine, F-
9 38000 Grenoble, France.

10

11 **Statement of authorship:** DC and LJP conceived the study, LM and WT provided comments on the
12 design and analyses. DC collated the data and performed the analyses. DC wrote the first draft of the
13 manuscript, and all authors contributed substantially to revisions.

14 **Data accessibility statement:** All data sources have been previously published. All data used in the
15 analyses will be archived in Dryad and the code used will be hosted on GitHub.

16 **Running title:** Addressing the Eltonian Shortfall

17 **Keywords:** food web, trophic interactions, ecological networks, ecological predictions, trait matching,
18 terrestrial vertebrates, model transferability

19 **Type of article:** Letter

20 **Word counts:** Abstract: 150 words; Main: 5072 words; Text box: NA

21 **Number of references:** 58

22 **Number of figures:** 4

23 **Number of tables:** 1

24 **Number of boxes:** 0

25 **Corresponding author:** Dominique Caron, dominique.caron@mail.mcgill.ca, (+1) 514-433-2947

26

27 LM: luigi.maiorano@uniroma1.it; WT: wilfried.thuiller@univ-grenoble-alpes.fr; LJP:

28 laura.pollock@mcgill.ca

29 **Abstract**

30 We have very limited knowledge of how species interact in most communities and ecosystems despite
31 trophic relationships being fundamental for linking biodiversity to ecosystem functioning. A promising
32 approach to fill this gap is to predict interactions based on functional traits, but many questions remain
33 about how well we can predict interactions for different taxa, ecosystems, and amounts of input data.
34 Here, we built a new traits-based model of trophic interactions for European vertebrates and found that
35 even models calibrated with 0.1% of the interactions (100 out of 71k) estimated the full European
36 vertebrate food web reasonably well. However, predators were easier to predict than prey, especially
37 for some clades (e.g. fowl and storks) and local food web connectance was consistently overestimated.
38 Our results demonstrate the ability to rapidly generate food webs when empirical data are lacking - an
39 important step towards a more complete and spatially-explicit description of food webs.

40 **Introduction**

41 Food webs are collections of trophic interactions describing both the composition and structure of
42 communities. Knowing the trophic structure of a food web is critical to understand how energy flows
43 through ecosystems (Thompson *et al.* 2012), how populations respond to perturbations through time
44 (Zhao *et al.* 2019), and how species are distributed in space (Wiszniewski *et al.* 2013). The food web is also
45 critical for conservation efforts (Harvey *et al.* 2017; Pollock *et al.* 2020) and can be an indicator of the
46 threat of extinction (McDonald-Madden *et al.* 2016). If food web links are lost more quickly than
47 species, then food webs can become simplified and homogenous with fewer trophic levels (Laliberté &
48 Tylianakis 2010; Estes *et al.* 2011; Valiente-Banuet *et al.* 2015). Despite this central importance of food
49 webs in ecology and conservation, we have yet to fully understand predator-prey relationships for many
50 species even in relatively well-studied areas.

51

52 One of the main reasons for this lack of understanding is the Eltonian shortfall – “*the lack of*
53 *knowledge about interactions among species or among groups of species*” (Hortal *et al.* 2015).
54 Available data on food webs are mostly restricted to the United States and Europe similar to other types
55 of species interaction networks (Hortal *et al.* 2015; Cameron *et al.* 2019; Poisot *et al.* 2021).
56 Compounding the problem even further, the sampling effort to detect interactions far exceeds the effort
57 needed to detect species because they require the simultaneous detection of both interacting species and
58 the interaction (Chacoff *et al.* 2012; Jordano 2016). Consequently, rare interactions are often missed
59 while those of dominant species are overestimated. So, even when available, food web datasets are
60 often incomplete and biased. A promising solution is to fill gaps in empirical food webs with expert
61 knowledge and literature review (e.g., Piechnik *et al.* 2008; Maiorano *et al.* 2020), but this approach
62 remains limited to well-studied systems.

63

64 Despite the size of the problem, there is reason to be optimistic about the potential for predicting
65 species interactions to fill gaps in food web data. In neutral food webs, the abundances of coexisting
66 species predict their probability of interacting (Canard *et al.* 2012). In many cases, trophic interactions
67 can be better predicted if they conform to a predictable set of phylogenetic and functional traits
68 (Morales-Castilla *et al.* 2015). Niche theory predicts that two species interact if the foraging traits of
69 the predator match the vulnerability traits of the prey (Williams & Martinez 2000; Gravel *et al.* 2016).
70 This trait-matching framework serves as the basis for most studies aiming to predict interactions (e.g.,
71 Gravel *et al.* 2013; Bartomeus *et al.* 2016; Pichler *et al.* 2020), which have shown promising results.
72 Notably, Eklöf *et al.*, (2013) found that 3-5 traits suffice to predict most trophic interactions, whereas
73 Gravel *et al.*, (2013) showed that predictions of trophic interactions in marine ecosystems are robust to
74 sampling effort. These findings suggest that we can potentially predict missing trophic interactions
75 even when data are sparse.

76

77 Yet, there have been few large-scale, multi-clade tests of predictive models of trophic interactions.
78 Many studies have used trait-matching models in systems where the relationships between the traits of
79 predators and their prey are expected to be strong, such as marine food webs (e.g., Gravel *et al.* 2013;
80 Laigle *et al.* 2018; Albouy *et al.* 2019). Other studies have built more complex models which can
81 outperform simpler models (e.g., Rohr *et al.* 2016), but have many parameters, relying on large datasets
82 that are often unavailable. Therefore, models based on relatively simple trait-matching relationships are
83 likely the best option for a wide range of taxa that have available trait data, but are they realistic enough
84 to make good predictions across a diverse set of taxa and ecosystems? We need a better understanding

85 of the amount of data needed to make reliable predictions and how general the matching rules are
86 across clades and space.

87

88 Here, we built a model to predict trophic interactions based on functional traits that are widely
89 available. We designed our model to be a very general model that can be applied across vertebrate
90 groups using a Bayesian linear model relating trophic interaction (and non-interactions) to a relatively
91 small set of predictor variables. These predictors describe the foraging ability of the predator, the
92 vulnerability of the prey, and the trait-match of interacting species. The relatively simplistic model
93 structure allows for fitting the model with few training data. We test our model using both the recently
94 assembled food web of all terrestrial vertebrates in Europe (Maiorano *et al.* 2020) and the Global Biotic
95 Interactions (GloBI) database (Poelen *et al.* 2014), and use the model to determine: (1) how much data
96 are needed to accurately predict the entire European food web, (2) which traits make the best
97 predictions, (3) how input data (empirical versus multi-sourced data) influence predictions, (4) which
98 taxa are easier or harder to make predictions for, and (5) how well the properties of local food webs can
99 be predicted across space.

100

101 **Materials and methods**

102 *Study area, species, and functional traits*

103 Our study focuses on trophic interactions among all terrestrial vertebrates of Europe: mammals,
104 breeding birds, reptiles, and amphibians. For each species, we extracted five functional traits from
105 Thuiller *et al.* (2015): diet, nesting habitat, activity time, foraging behavior, and body mass (Appendix
106 S1). Because body mass was missing for many amphibians - 52% missing for frogs (*Anura*) and 46%
107 for salamanders (*Caudata*) - we imputed missing body mass from available body length information for

108 these two groups separately. These imputations were justified by the strong relationship between body
109 mass and length in our study ($r^2 = 0.75$ and 0.84 for frogs and salamanders respectively; Appendix S2),
110 and others (Deichmann *et al.* 2008). In all, we gathered these five functional traits for 1055 species:
111 101 amphibians, 507 birds, 267 mammals, and 180 reptiles.

112

113 Using diet, nesting habitat, activity time, foraging behavior, and body mass, we calculated seven
114 predictor variables for each species pair (Table 1). These predictors of predator-prey interactions can be
115 categorized into three types: foraging traits, vulnerability traits, and matching traits (Gravel *et al.* 2016;
116 Rohr *et al.* 2016). Foraging traits (body mass of the predator and its foraging behavior) influence the
117 number of prey of a given predator, whereas vulnerability traits (body mass of the prey) influence the
118 number of predators of a given prey. Matching traits influence the feasibility of the interaction
119 (difference in body mass and diet match) or the encounter rate of species (activity time and habitat
120 match).

121

122 *Interaction data*

123 We extracted interaction data from the trophic metaweb of European terrestrial vertebrates (referred to
124 as the Metaweb; Maiorano *et al.*, 2020). A metaweb documents all potential interactions between all
125 species at the regional scale (Dunne 2006). Maiorano *et al.* (2020) compiled all potential trophic
126 interactions and non-interactions between all terrestrial vertebrates of Europe from guide books,
127 published papers, and completed by expert opinion. As all pairs of species were assessed as to whether
128 they could potentially interact, we assumed all zeros are true non-interactions rather than resulting from
129 missing data. We extracted the interactions and non-interactions between all of the 1 055 species for

130 which we had functional traits. This represents 71 417 potential interactions and 1 041 608 non-
131 interactions.

132

133 Because the Metaweb documents *potential* trophic interactions, we also used a database of empirical
134 observations of species interactions. We extracted all interactions reported in the Global Biotic
135 Interactions (GloBI) platform (Poelen *et al.* 2014). The Metaweb and GloBI are very different in terms
136 of data coverage. While the Metaweb documents only trophic interactions in Europe, GloBI aggregates
137 trophic (and non-trophic) interactions from anywhere in the world. The absences of interactions in the
138 Metaweb can be interpreted as likely to be true absences (i.e., low false negative rate), whereas
139 absences of interactions in GloBI cannot (i.e., high false negative rate). Conversely, the presences of
140 interactions in the Metaweb are more uncertain (i.e., high false positive rate) than in GloBI. We used
141 the package *rglobi* of the R software to extract all trophic interactions between any of the focal 1 055
142 species. In all, we extracted 291 trophic interactions from GloBI involving 194 different species (75
143 predator species and 146 prey species).

144

145 *Predictive model*

146 We modelled the occurrence of a food web interaction for each pair of species as a function of their
147 traits using Bayesian generalized linear models (GLM). We assumed that the occurrence of interaction
148 between species i and j , L_{ij} , is Bernouilli distributed. The corresponding probability of interaction was
149 modelled as the inverse *logit* of a linear function with a common intercept, α , and a set of linear
150 coefficients β associated with the seven predictors T_{ij} (Table 1):

$$151 \quad \text{logit}(P(L = 1)) = \alpha + \sum_{k=1}^7 \beta_k T_{ijk} \quad (\text{Eq. 1})$$

152 We measured the predictive performance of each model on independent validation datasets using the
153 area under the receiver operating characteristic curve (AUC). AUC varies from 0.5 to 1 where 0.5
154 indicates that the model failed to rank interactions higher than non-interactions (i.e., random
155 prediction), and 1 indicates that the model systematically ranked interactions higher than non-
156 interactions (i.e., perfect prediction). We also measured the area under the precision-recall curve, the
157 true positive rate, true negative rate, and true-skill statistics to see if our results were robust to the
158 choice of performance metric (Appendix S8).

159

160 Before fitting the GLMs, we scaled each continuous predictor by subtracting it by its mean and
161 dividing by two times its standard deviation, so that the coefficients of the scaled continuous predictors
162 are directly comparable to coefficients of unscaled binary predictors (Gelman 2008). We used Bayesian
163 inference and Markov chain Monte Carlo algorithm to estimate the model parameters (α and β_k). We
164 used weakly informative priors for the parameters:

$$165 \quad \alpha \sim \text{Normal}(\text{mean} = 0, \text{sd} = 10) \quad (\text{Eq. 2})$$

$$166 \quad \beta_k \sim \text{Normal}(\text{mean} = 0, \text{sd} = 1) \quad (\text{Eq. 3})$$

167 where sd is the standard deviation of the prior distribution. We ran 3 chains, each with 1000 warm-up
168 iterations, followed by 5 000 iterations for inference. We diagnosed convergence visually of a few test
169 runs, and calculated the potential scale reduction factor, \hat{R} , for all runs (Gelman & Rubin 1992;
170 Appendix S3). We conducted the Bayesian analyses using the package *greta* in R (Golding 2019).

171

172 *Predicting trophic interactions with models trained on the Metaweb*

173 We used the model described above to make predictions for predator-prey pairs in the Metaweb, and
174 determine the information needed to make reliable predictions. More specifically, we determined: (1)
175 the number of trophic interactions needed to calibrate models that could recover most of the Metaweb,
176 and (2) which of the predictors made the most important contribution. To determine the information
177 needed to make predictions, we trained a set of 75 models with progressively more trophic interactions
178 from the Metaweb going from 2 to 5000 pairwise interactions, with 10 000 non-interactions. Each of
179 the 75 models used the same set of seven predictors (Table 1). We then measured how well these
180 models predicted the Metaweb by comparing the predictions generated by the model to an independent
181 validation dataset, which was a sample of 1% of the Metaweb (Metaweb validation dataset; random
182 sampling had little effect on predictive performance; Appendix S4). To compare the importance of trait-
183 based predictor variables, we compared the mean coefficient values of all scaled predictors (Gelman
184 2008) from a model calibrated on the entire Metaweb.

185

186 We also measured how well each of the models described above predicts the empirical interactions in
187 GloBi. To do so, we compared predictions generated by each model to the 291 trophic interactions we
188 extracted from GloBi. Because GloBI only includes interactions, we added 3 845 pairs of species that
189 do not interact in the Metaweb (non-interactions) to the GloBI validation dataset. We added 3 845 non-
190 interactions to make the prevalence of interactions in the GloBI validation dataset comparable to the
191 prevalence in the Metaweb validation dataset.

192

193 *Predicting trophic interactions with models trained on GloBI*

194 We were also interested in how well we could predict the entire Metaweb without using any
195 information from the Metaweb itself. To do this, we combined information on species traits with the

196 observed interactions included in GloBI. We fitted 40 predictive models using a progressively
197 increasing number of trophic interactions extracted from GloBI: from 2 to 291 trophic interactions
198 combined with 10 000 pseudo-absences. To extract pseudo-absences, we randomly drew pairs of
199 species for which no interaction is documented in GloBI. We used pseudo-absences to avoid using any
200 information from the Metaweb. We measured performance of these models with respect to the
201 Metaweb validation dataset, and to the GloBI validation dataset.

202

203 *Predictions of interactions for different taxa*

204 In addition to the overall predictive performance of the models, we wanted to identify whether
205 interactions between certain clades and different types of species (e.g., specialists versus generalists)
206 were more or less predictable with our model. To do this, we first trained a model (*master* model) with
207 10% of the European Metaweb (7 157 trophic interactions and 111 287 non-interactions) and predicted
208 all trophic interactions and non-interactions in the entire food web. We measured the performance of
209 the model to predict all interactions (i.e., the prey and predators) of every species. We further compared
210 these performances for individual species to explore whether interactions of specialists (i.e., species
211 interacting with few species) were harder to predict than interactions of generalists (i.e., species
212 interacting with many species). Specifically, we modelled the performance for each species (the AUC
213 value) as a function of the species generality using three bayesian generalized linear mixed models
214 ($GLMM_{all}$, $GLMM_{prey}$, $GLMM_{predator}$) with random intercepts and slopes for the species group (i.e.,
215 order), and a logit-link function. All species were included in each model, but different representations
216 of generality of each species ($GLMM_{all}$: the number of total interactions; $GLMM_{prey}$: the number of
217 prey; $GLMM_{predator}$: the number of predators) were included as fixed effects in the three separate models.

218 The fixed effects were log-transformed and scaled before running the GLMMs. We fitted the GLMMs
219 using the package *greta* in R (details in appendix S9).

220

221 Finally, we performed a miscalibration analysis to investigate the ecological differences making the
222 trophic interactions of some groups less predictable than others. We first trained group-specific models
223 for each order of predators. For example, considering frogs (*Anura*), we trained a predictive model only
224 using interactions (and non-interactions) involving a frog species as predator. For many orders, most
225 species shared very similar foraging behaviors. This caused some group-specific models not to
226 converge or to over-fit the data. For this reason, we excluded foraging behaviors from the predictors in
227 the miscalibration analysis. We compared the group-specific models to a *general* model trained on the
228 entire Metaweb (71 417 trophic interactions and 1 113 025 non-interactions). Specifically, we
229 calculated the differences between the parameters inferred for the *general* model to the parameters
230 inferred for the group-specific models to measure the miscalibration of the *general* model for each
231 group. By doing so, we identified the miscalibrated coefficients causing the general model to
232 incorrectly predict the prey of specific predator groups. In contrast to the *master* model, the *general*
233 model did not include foraging behaviors as predictors to make it comparable to group-specific models.

234

235 *Predictions of interactions across space*

236 In addition to model performance for different types of species, we also tested how well the model
237 performed in different regions and for local assemblages. This tests the generality of how traits can
238 predict potential trophic interactions with a range of ecological constraints and species pools. First, we
239 used the *master* model described above to make predictions for ‘local’ food webs based on distribution
240 data for each 10km pixel across Europe. Second, we addressed model transferability by training models

241 in one region and using that model to predict to another region. If a model trained in the arctic, for
242 example, can predict interactions in the Mediterranean with regional specialist species, then this is
243 evidence for the generality of trait-interaction relationships and the ability of these models to
244 extrapolate.

245

246 We determined the local/regional webs based on species composition data from Maiorano et al. (2013),
247 where species distributions are determined by the presence of primary habitat within the known species
248 distribution range and validated by field data. Species habitat requirements are determined by experts
249 based on land cover, elevation and distance to water. Species distribution ranges are extracted from
250 atlases (see Maiorano *et al.* 2013 for the full list of atlases). The original data are 300-m cell range
251 maps with three levels: unsuitable, secondary and primary habitat. We upscaled the distributions data to
252 a 10km pixel equal-area grid, and considered a species present in a cell if it had at least one occurrence
253 of primary habitat.

254

255 We created ‘local’ food webs for each 10km pixel across Europe by pruning the Metaweb in every
256 pixel based on the species present in that pixel. We assumed that spatial variation in potential food
257 webs are simply due to species turnover. We are not taking into account intraspecific trait variation,
258 change in abundances or interaction turnovers, which are likely important in the local realization of
259 interactions. We predicted the interactions for each local species assemblage by pruning the predictions
260 of the entire Metaweb (using the *master* model) based on the species pool in each pixel and compared
261 these to the ‘true’ interactions according to the pruned metaweb. If environmental gradients influence
262 the ecological constraints driving trophic interactions, we expect to detect a gradient in the
263 predictability of local food webs. We specifically wanted to investigate whether simpler, less connected

264 food webs (lower connectance) were easier to predict (evaluated with AUC, true positive and true
265 negative rates) than more complex, highly connected food webs.

266

267 We created regional food webs by extracting the trophic interactions and non-interactions from the
268 Metaweb between species for which the range intersected in at least 1 % of the European bioregions
269 (see O'Connor et al. 2020) and trained models for each bioregion individually. To investigate the
270 transferability of predictive models across bioregions, we compared the importance of model predictors
271 in each bioregion (e.g. the linear coefficient associated with predator body mass in the arctic versus in
272 the mediterranean bioregion), and how well they could predict the food webs of other bioregions. If
273 environmental gradients influence the ecological constraints driving trophic interactions, we expect
274 poor transferability of models between bioregions.

275

276 **Results**

277 *How much data are needed to accurately predict the entire European food web?*

278 Most pairwise interactions in the Metaweb were predicted reasonably well by models trained on few
279 interactions. Predictive performance increased when more interactions were used to train the model,
280 but performance stabilized around AUC=0.92 at 100 training interactions (0.14% of the total number of
281 interactions in the Metaweb; Fig 1a). Even with as low as 10 interactions (0.014% of all interactions in
282 the Metaweb; Fig 1a), the AUC was over 0.90. Training the model on a lot more presences and
283 absences did not improve substantially model performance. The *master* model, which we trained on
284 10% of the entire Metaweb to predict the interactions of every species, resulted in an AUC of 0.92, well
285 above the 0.5 expected from a null model.

286

287 *Which traits best predict interactions?*

288 All of the traits used as predictor variables in the models were important for predicting interactions: all
289 corresponding linear coefficients were different from 0 (i.e., 95% credible interval did not include 0).

290 The most important predictors were the match between the diet of the predator and the type of prey
291 (median = 2.29, 95%CrI = [2.27, 2.32]), the body mass of the predator (median = 2.08, 95%CrI =
292 [2.04, 2.12]), the body mass of the prey (median = -2.00, 95%CrI = [-2.03, -1.96]), and the match
293 between the body mass of the predator and the body mass of the prey (median = -1.75, 95%CrI = [-
294 1.79, -1.70]). All parameter estimates with their 95% credible interval are available in Appendix S5.

295

296 *How input data (empirical versus multi-sourced data) influences predictions?*

297 We found that results from models trained on the European Metaweb were consistent with models
298 trained on interactions from GloBI. Models fitted using all GloBI interactions (291 interactions)
299 predicted the Metaweb well (AUC=0.91; Fig 1a). Also, all models performed similarly in predicting the
300 realized interactions from GloBI compared to the potential interactions of the Metaweb (Fig 1b).

301

302 *Which taxa are easier or harder to make predictions for?*

303 Overall, the *master* model performed well for all groups (AUC > 0.75 for all groups; Fig 2). The
304 variation between groups was mostly due to variation in the ability of the model to predict prey (Fig 2).
305 For example, prey of carnivorans (*Carnivora*), fowl (*Galliformes* and *Anseriformes*), pelicans
306 (*Pelicaniformes*), storks (*Ciconiformes*), and birds of prey (*Falconiformes*, *Charadriiformes*, and
307 *Strigiformes*) were harder to predict on average (Fig 2). In contrast, predicting predators was similarly
308 easy across all prey groups (AUC > 0.9). One exception to this pattern was amphibians: the prey of

309 *Caudata* and *Anura* were easier to predict (AUC = 0.96 and 0.95 respectively) than their predators
310 (AUC = 0.85 and 0.78 respectively).

311

312 Interactions of specialists tended to be more predictable than interactions of generalists (Fig 3). The
313 order of the species and the number of interactions explained 56% (bayesian $R^2_{\text{GLMM}_{\text{all}}}$) of the variation
314 in the logit-AUC (fixed effect estimate = -0.49; 95%CrI = [-0.54, -0.43]). In general, it is easier to
315 predict species' predators than their prey (Fig 3). Prey of generalist predators tended to be harder to
316 predict (GLMM_{prey} fixed effect estimate = -0.30; 95%CrI = [-0.38, -0.22]). Similarly, predators of high
317 vulnerability prey were harder to predict (GLMM_{predator} fixed effect estimate = -0.24; 95%CrI = [-0.28, -
318 0.20]).

319

320 For most groups in which potential prey were difficult to predict, the group-specific model (i.e. model
321 calibrated on interactions of one predator order) performed much better, with the AUC increasing by
322 0.1 or more (Appendix S6). Our models highlight how some bird groups (e.g., *Galliformes*,
323 *Anseriformes*, *Pelicaniformes*) tend to feed on prey active at different times during the day and nest in
324 different habitats than the focal predator. This contrasts with woodpeckers (*Piciformes*) for which the
325 matches of activity times and nesting habitats of the prey and the predator are especially important. For
326 other groups (e.g., *Rodentia*, *Eulipotyphla*, *Strigiformes*, *Passeriformes*) the relation between the body
327 mass of the predator and the prey is more important to explain interactions than average. We also found
328 that, for carnivorans (*Carnivora*), the predictive performance of the group-specific model remained low
329 (AUC = 0.63). The coefficients of the different predictors were mostly lower (closer to 0) than the
330 *general* model, suggesting that the interactions of carnivorans are difficult to predict from our set of
331 traits. The complete results of the miscalibration analysis are available in the Appendix S6.

332

333 *How well the properties of local food webs can be predicted across space?*

334 Overall, local food webs with more trophic interactions were less predictable than simpler ones (Fig 4).
335 This trend resulted in a slight increase in the predictability of local food webs at higher latitudes (Fig
336 4). True positive rates were higher than true negative rates across the entire continent which caused the
337 predicted connectance (proportion of possible links that are realized) to be systematically overestimated
338 by 2 to 4 times (Fig 4). All bioregional food webs were similarly predictable ($0.89 < \text{AUC} < 0.92$) by
339 the *master* model and models transferred well from one bioregion to another (Appendix S7). Similarly,
340 trait parameters of each bioregional predictive model did not diverge greatly from the *master* model
341 (Appendix S7).

342

343 **Discussion**

344 In this study, we recovered the entire European food web of tetrapods from a fraction of the entire food
345 web and widely available functional traits. Our results indicate that these traits (describing the foraging
346 of predators, vulnerability of prey, and the trait-match between predator and prey) are indeed predictive
347 of the predator-prey pairs that make up large food webs. These trait-interaction relationships appear to
348 be general given: (1) the stability of modeled effects (e.g., the positive effect of body mass differences)
349 across most vertebrate orders and highly contrasting ecosystems, (2) the consistency of predictive
350 power between the complete, yet potential interaction data of the European metaweb and the sparse,
351 empirical data from GloBI, and (3) the efficiency of the model to make good predictions with few input
352 interactions (recovering ~71k interactions with <100 known interactions with an $\text{AUC} > 0.90$). While
353 predictions were mostly reliable, there were exceptions including: poor predictions for the prey of some

354 predator orders, the trophic interactions of generalists and the tendency to overestimate connectance
355 especially in complex food webs, which we discuss further below.

356

357 Overall, this generality in trait-interaction relationships and the ability of the model to extrapolate
358 suggests these models could be applied to many ecosystems, even those with very sparse and biased
359 datasets. GloBi only documented 291 trophic interactions compared to the 71 417 trophic interactions
360 documented in the European Metaweb, which illustrate the sparsity of trophic interactions data even in
361 one of the best-studied continents. With very limited data, it is important that models can extrapolate to
362 new conditions (Roberts *et al.* 2017; Santini *et al.* 2021). Our models produced similar results (e.g., had
363 similar coefficients for trait variables) when fitted to different bioregions, and predictive performance
364 of bioregional models did not decrease when one bioregion was used to predict to another. This
365 indicates potential transfer of information from data-rich bioregions to data-poor bioregions. While
366 traits have been shown to promote transferring information from one region to the next for species
367 distributions (Vesk *et al.* 2021), here we show how trait-based models can similarly transfer
368 information on *potential* interactions to entirely new areas. We need more studies on the factors
369 influencing model transferability in food web interactions to fully appreciate the potential of trait-based
370 models to predict interactions in data-poor systems or under future conditions (Yates *et al.* 2018). The
371 prediction and transfer of information on *realized* interactions is likely to be much harder since they are
372 influenced by local abundances, intraspecific trait variation, and indirect interactions from other species
373 (Poisot *et al.* 2015; Pellissier *et al.* 2018).

374

375 While the generality and overall performance of these models are promising, we highlighted some
376 systematic biases. For example, our predictive model systematically overestimated the number of links

377 in local food webs (Fig 4). Interactions were better predicted (true positive rate ≈ 0.9) than non-
378 interactions (true negative rate ≈ 0.8), meaning that predicted interactions includes most realized
379 interactions, but also many incorrectly predicted non-interactions (false discovery rate ≈ 0.7 ; Appendix
380 S8). Thus, our predictive model should be viewed as a first step toward a correct description of a
381 regional and local food webs by reducing the millions of possible interactions to thousands of feasible
382 ones. These predicted interactions could be used to inform targeted sampling or expert elicitation to get
383 a more accurate picture of the true food web.

384

385 In addition, not all trophic interactions were equivalent. The prey of some taxonomic groups were
386 harder to predict than others using a general model (Fig 2). Our miscalibration analysis helped
387 understand where the general model failed for these groups and highlighted some ecological
388 differences in how predator groups choose their prey (Appendix S6). For example, while the match in
389 nesting habitat and activity time of the predator and the prey were not among the important predictors
390 in the general model, these predictors were particularly important for some groups. Woodpeckers
391 (*Piciformes*) tend to feed on prey that share similar nesting habitats and are active at similar times.
392 Conversely, fowl (*Galliformes* and *Anseriformes*), pelicans (*Pelicaniformes*) and storks (*Ciconiformes*)
393 tend to feed on prey that nest in different habitats and active at different times during the day. These are
394 not surprising results: woodpeckers generally both forage and nest in and on trees, while many fowl
395 and storks forage on water but nest in different habitats (Svensson & Grant 2009). We also found that
396 the prey of other predator groups, such as carnivorans (*Carnivora*), falcons (*Falconiformes*), and owls
397 (*Strigiformes*), were harder to predict on average. These groups include many relatively generalist
398 predators that feed on many prey types: small herbivores, but also ungulates, birds, or other carnivores
399 (Svensson & Grant 2009; Hackländer & Zachos 2020). The functional diversity of the diet of generalist

400 predators is larger which make it harder to find general matching rules that explain how these generalist
401 predators choose their prey. Indeed, we found that the interactions of generalist species are harder to
402 predict than that of specialists. Interestingly, predicting specialization has been challenging for other
403 kinds of interactions (Blüthgen *et al.* 2008; Calatayud *et al.* 2016; Olalla-Tárraga *et al.* 2017). Although
404 predicting trophic specialization appeared to not be an issue here for terrestrial vertebrates, further
405 research is needed to understand the role of traits in predicting specialist and generalist interactions.
406

407 The field of species interaction and ecological network prediction has been very active in recent years
408 (Strydom *et al.* 2021b). We presented a predictive model that aims to identify general relationships
409 between traits and food web interactions and be used for multi-clade vertebrate groups with very few
410 trophic interaction data. For systems in which we have more information, other approaches are
411 possible. First, we showed that predators tend to vary in how they choose their prey between taxa,
412 making hierarchical models good candidates to improve predictions. Hierarchical models allow some
413 variation in the regression coefficients between groups (Ovaskainen *et al.* 2017; Gelman *et al.* 2020).
414 Because how species choose their prey tends to be evolutionary conserved, phylogenetic relationships
415 could inform how regression coefficients correlate across clades (Gómez *et al.* 2010). Second,
416 phylogenetic relationships can directly make predictions given enough interaction data (Elmasri *et al.*
417 2020), or to transfer species interaction knowledge between systems (Strydom *et al.* 2021a). Third,
418 machine learning algorithms have been used to predict interactions within networks (i.e. in sample
419 prediction) and oftentimes outperformed linear models (e.g., Desjardins-Proulx *et al.* 2017; Pichler *et*
420 *al.* 2020), but often rely on a larger volume of data, and ecological inference can be less
421 straightforward. There are also advantages to having relatively simplistic linear responses in terms of
422 inference and the potential for better out-of-sample performance (Wenger & Olden 2012). Finally, bio-

423 energetic models can infer energy fluxes between organisms and help quantify food webs using body
424 mass, metabolic demands, and energy loss (Berlow *et al.* 2009; Rall *et al.* 2012). A promising avenue
425 of research is to combine trait-based models to other methods, such as expert elicitation or bio-
426 energetic models, to downscale regional metaweb into local quantitative food webs (Rall *et al.* 2012;
427 Bode *et al.* 2017). This would allow us to investigate how food webs and species interactions vary in
428 space and time, and to forecast the consequences of global changes on the composition and structure of
429 ecosystems.

430

431 **Acknowledgements**

432 We acknowledge that this study was conducted on unceded land which has long served as a site of
433 meeting and exchange amongst Indigenous peoples, including the Haudenosaunee and Anishinabeg
434 nations. We thank DC's PhD supervisory committee members Dominique Gravel, Timothée Poisot and
435 Brian Leung, and members of the FutureWeb working group for useful comments. This research was
436 supported by an NSERC Discovery Grant (NSERC RGPIN-2019-05771), the 2017–2018 Belmont
437 Forum and BiodivERSA joint call for research proposals, under the BiodivScen ERA-Net COFUND
438 program, and with the funding organizations Agence Nationale pour la Recherche (FutureWeb, ANR-
439 18-EBI4–0009). We declare we have no potential sources of conflict of interest.

440 **References**

- Albouy, C., Archambault, P., Appeltans, W., Araújo, M.B., Beauchesne, D., Cazelles, K., *et al.* (2019). The marine fish food web is globally connected. *Nat Ecol Evol*, 3, 1153–1161.
- Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Funct. Ecol.*, 30, 1894–1903.
- Berlow, E.L., Dunne, J.A., Martinez, N.D., Stark, P.B., Williams, R.J. & Brose, U. (2009). Simple prediction of interaction strengths in complex food webs. *PNAS*, 106, 187–19
- Blüthgen, N., Fründ, J., Vázquez, D.P. & Menzel, F. (2008). What Do Interaction Network Metrics Tell Us About Specialization and Biological Traits. *Ecology*, 89, 3387–3399.
- Bode, M., Baker, C.M., Benshemesh, J., Burnard, T., Rumpff, L., Hauser, C.E., *et al.* (2017). Revealing beliefs: using ensemble ecosystem modelling to extrapolate expert beliefs to novel ecological scenarios. *Methods Ecol. Evol.*, 8, 1012–1021.
- Calatayud, J., Hórreo, J.L., Madrigal-González, J., Migeon, A., Rodríguez, M.Á., Magalhães, S., *et al.* (2016). Geography and major host evolutionary transitions shape the resource use of plant parasites. *PNAS*, 113, 9840–9845.
- Cameron, E.K., Sundqvist, M.K., Keith, S.A., CaraDonna, P.J., Mousing, E.A., Nilsson, K.A., *et al.* (2019). Uneven global distribution of food web studies under climate change. *Ecosphere*, 10, e02645.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLoS One*, 7, e38295.

- Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *J. Anim. Ecol.*, 81, 190–200.
- Deichmann, J.L., Duellman, W.E. & Williamson, G.B. (2008). Predicting Biomass from Snout–Vent Length in New World Frogs. *J. Herpetol.*, 42, 238–245.
- Desjardins-Proulx, P., Laigle, I., Poisot, T. & Gravel, D. (2017). Ecological interactions and the Netflix problem. *PeerJ*, 5, e3644.
- Dunne, J. (2006). The network structure of food webs. In: *Ecological Networks: Linking Structure to Dynamics in Food Webs*. pp. 27–86.
- Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro Urgal, R., Chacoff, N.P., *et al.* (2013). The dimensionality of ecological networks. *Ecol. Lett.*, 16, 577–583.
- Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. *Ann. Appl. Stat.*, 14, 221–240.
- Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J., *et al.* (2011). Trophic Downgrading of Planet Earth. *Science*, 333, 301–306.
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Stat Med.*, 27, 2865–2873.
- Gelman, A., Hill, J. & Vehtari, A. (2020). *Regression and Other Stories*. Cambridge University Press.
- Gelman, A. & Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.*, 7, 457–472.
- Golding, N. (2019). greta: simple and scalable statistical modelling in R. *JOSS*, 4, 1601.

- Gómez, J.M., Verdú, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465, 918–921.
- Gravel, D., Albouy, C. & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 371, 20150268.
- Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from predator-prey body size relationships. *Methods Ecol. Evol.*, 4, 1083–1090.
- Hackländer, K., & Zacos, F. E. (2020). *Mammals of Europe: Past, Present, and Future*. Springer.
- Harvey, E., Gounand, I., Ward, C.L. & Altermatt, F. (2017). Bridging ecology and conservation: from ecological networks to ecosystem function. *J Appl Ecol*, 54, 371–379.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu Rev Ecol Evol Syst*, 46, 523–549.
- Jordano, P. (2016). Sampling networks of ecological interactions. *Funct. Ecol.*, 30, 1883–1893.
- Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I. & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127, 316–326.
- Laliberté, E. & Tylianakis, J.M. (2010). Deforestation homogenizes tropical parasitoid–host networks. *Ecology*, 91, 1740–1747.
- Maiorano, L., Amori, G., Capula, M., Falcucci, A., Masi, M., Montemaggiori, A., *et al.* (2013). Threats from Climate Change to Terrestrial Vertebrate Hotspots in Europe. *PLoS One*, 8.
- Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA EU 1.0: A species level trophic metaweb of European tetrapods. *Glob. Ecol. Biogeogr.*, 29, 1452–1457.
- McDonald-Madden, E., Sabbadin, R., Game, E.T., Baxter, P.W.J., Chadès, I. & Possingham, H.P. (2016). Using food-web theory to conserve ecosystems. *Nat Commun*, 7, 1–8.

- Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends Ecol. Evol.*, 30, 347–356.
- O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J. Biogeogr.*, 47, 181–192.
- Olalla-Tárraga, M.Á., González-Suárez, M., Bernardo-Madrid, R., Revilla, E. & Villalobos, F. (2017). Contrasting evidence of phylogenetic trophic niche conservatism in mammals worldwide. *J. Biogeogr.*, 44, 99–110.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., *et al.* (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.*, 20, 561–576.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). Comparing species interaction networks along environmental gradients. *Biol. Rev.*, 93, 785–800.
- Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M. & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods Ecol. Evol.*, 11, 281–293.
- Piechnik, D.A., Lawler, S.P. & Martinez, N.D. (2008). Food-web assembly during a classic biogeographic study: species' "trophic breadth" corresponds to colonization order. *Oikos*, 117, 665–674.
- Poelen, J.H., Simons, J.D. & Mungall, C.J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecol. Inform.*, 24, 148–159.
- Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). Global knowledge gaps in species interaction networks data. *J. Biogeogr.*, 48, 1552–1563.

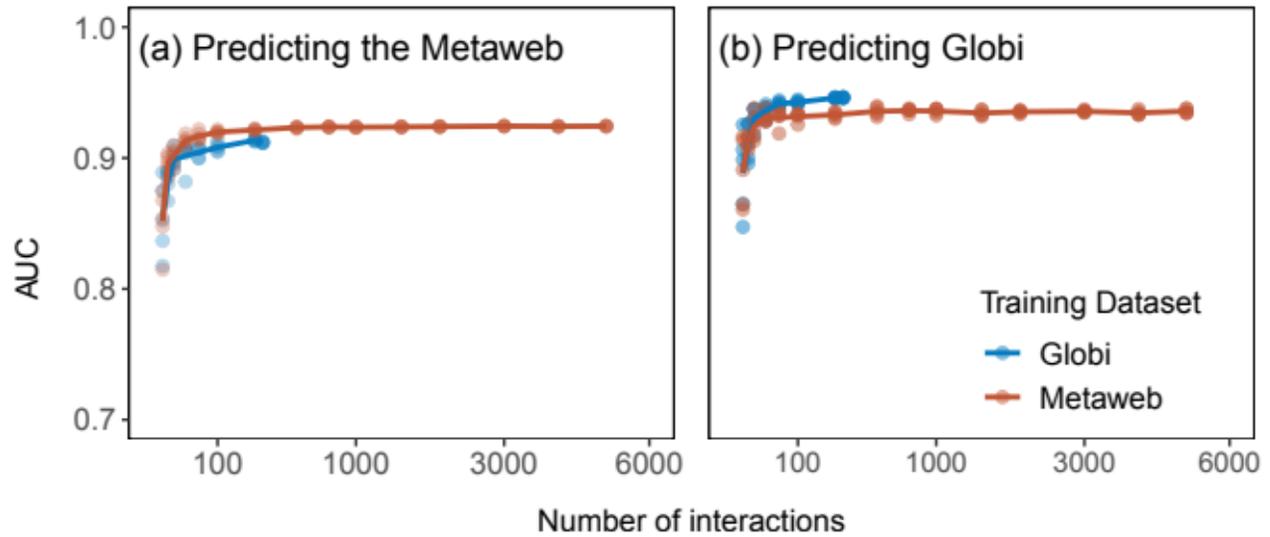
- Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.
- Pollock, L.J., O'Connor, L.M.J., Mokany, K., Rosauer, D.F., Talluto, M.V. & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends Ecol. Evol.*, 35, 1119–1128.
- Rall, B.C., Brose, U., Hartvig, M., Kalinkat, G., Schwarzmüller, F., Vucic-Pestic, O., *et al.* (2012). Universal temperature and body-mass scaling of feeding rates. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 367, 2923–2934.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillerá-Arroita, G., *et al.* (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Rohr, R.P., Naisbit, R.E., Mazza, C. & Bersier, L.-F. (2016). Matching–centrality decomposition and the forecasting of new links in networks. *Proc. Royal Soc. B*, 283, 20152702.
- Santini, L., Benítez-López, A., Maiorano, L., Čengić, M. & Huijbregts, M.A.J. (2021). Assessing the reliability of species distribution projections in climate change research. *Divers. Distrib.*, 27, 1035–1050.
- Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2021a). Food web reconstruction through phylogenetic transfer of low-rank network representation. *EcoEvoRxiv*.
- Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021b). A roadmap towards predicting species interaction networks (across space and time). *Philos Trans R Soc Lond B Biol Sci*, 376, 20210063.
- Svensson, L. & Grant, P.J. (2009). *Birds of Europe*. Princeton field guides. 2nd ed. Princeton University Press, Princeton.

- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladysz, S., Kitching, R.L., *et al.* (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.*, 27, 689–697.
- Thuiller, W., Maiorano, L., Mazel, F., Guilhaumon, F., Ficetola, G.F., Lavergne, S., *et al.* (2015). Conserving the functional and phylogenetic trees of life of European tetrapods. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 370, 20140005.
- Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A., Galetti, M., *et al.* (2015). Beyond species loss: the extinction of ecological interactions in a changing world. *Funct. Ecol.*, 29, 299–307.
- Vesk, P.A., Morris, W.K., Neal, W.C., Mokany, K. & Pollock, L.J. (2021). Transferability of trait-based species distribution models. *Ecography*, 44, 134–147.
- Wenger, S.J. & Olden, J.D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol. Evol.*, 3, 260–267.
- Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.
- Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., *et al.* (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.*, 88, 15–30.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., *et al.* (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends Ecol. Evol.*, 33, 790–802.
- Zhao, Q., Brink, P.J.V. den, Carpentier, C., Wang, Y.X.G., Rodríguez-Sánchez, P., Xu, C., *et al.* (2019). Horizontal and vertical diversity jointly shape food web stability against small and large perturbations. *Ecol. Lett.*, 22, 1152–1162.

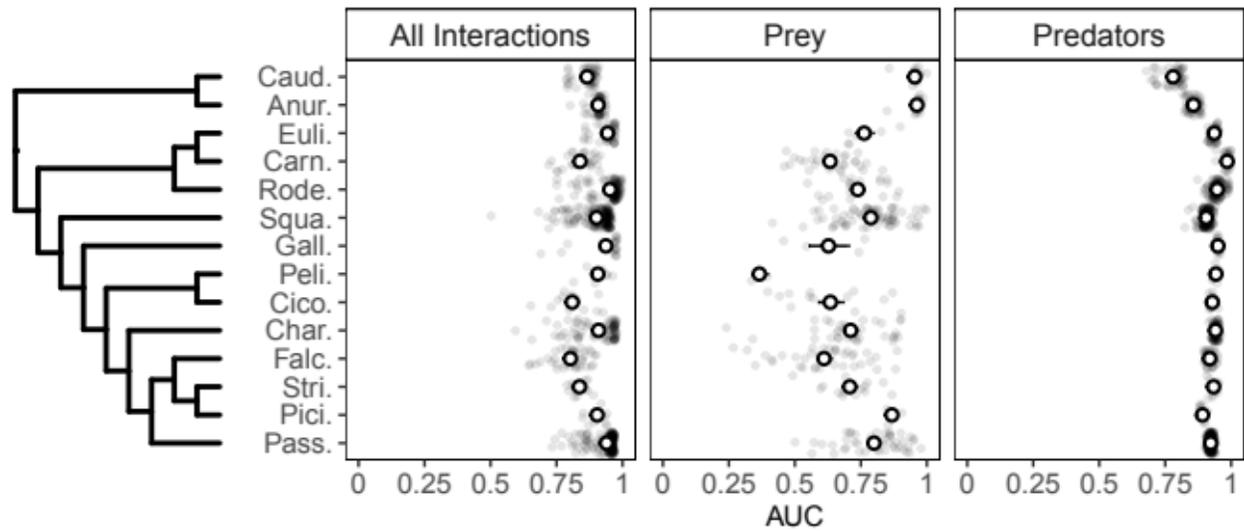
442 *Table 1: Variables used to predict trophic interactions between all species pairs. The variables are*
 443 *grouped as foraging traits, vulnerability traits and matching traits. BM stands for body mass.*
 444

	Foraging		Vulnerability	Matching			
	Predator body mass	Predator foraging behavior	Prey body mass	Diet	Activity time	Nesting habitat	Body mass difference
Functional traits	Body Mass of the predator	Foraging behavior of the predator	Body Mass of the prey	Diet of the predator and the class of the prey	Activity time of the prey and predator	Nesting habitats of the prey and the predator	Body mass of the prey and the predator
Operation	log		log	Does the diet of the predator match the type of prey (1) or not (0)?	Jaccard similarity*	Jaccard similarity*	$(\log(\text{BM}_{\text{predator}}) - \log(\text{BM}_{\text{prey}}))^2$
Variable type	Continuous	4 binary variables	Continuous	Binary	Continuous	Continuous	Continuous

445 *We calculated the Jaccard similarity coefficient for the Activity time and Nesting habitat of the
 446 predator and the prey. Activity time and Nesting habitat are a set of binary variables (Appendix S1).

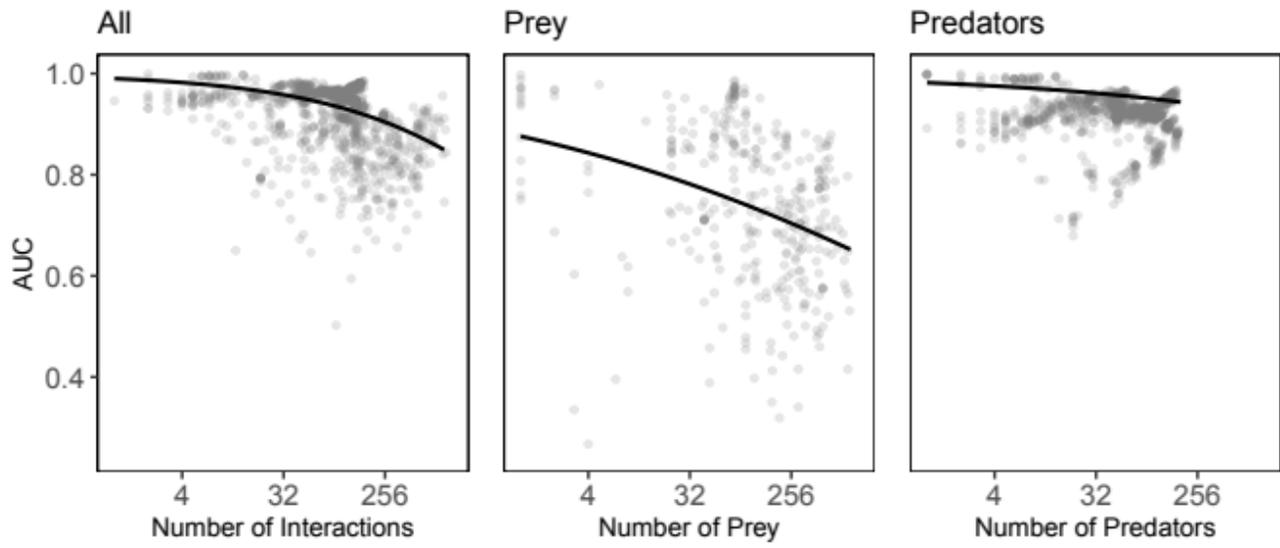


449 *Figure 1: Effect of the number of interactions sampled on the predictive performance of the model. In*
 450 *(a), we measured performance using the European Metaweb validation dataset. In (b), we measured*
 451 *performance with interactions using the GloBI validation dataset. In red and blue are the predictive*
 452 *models trained with interactions sampled from the European metaweb and GloBI, respectively.*



454 *Figure 2: Differences in predictive performance among groups. Predictive performance is measured*
 455 *with the AUC. Each grey point is the predictive performance of the general model for a single species.*
 456 *The white points are the group mean, with the associated standard error. From left to right, the panels*
 457 *represent the performance of the general model to predict all trophic interactions, the prey, and the*
 458 *predators of the focal species. From top to bottom, the groups are Caudata, Anura, Eulipotyphla,*
 459 *Carnivora, Rodentia, Squamata, Galliformes, Pelecaniformes, Ciconiiformes, Charadriiformes,*
 460 *Falconiformes, Strigiformes, Piciformes, and Passeriformes.*

461



463 *Figure 3: Effect of generality on model performance. Each grey point is the predictive performance of*
 464 *the master model to predict all interactions (left; $GLMM_{all}$), the prey (center; $GLMM_{prey}$), and the*
 465 *predators (right; $GLMM_{predator}$) of a single species. We measured generality as the number of*
 466 *interactions (left), the number of prey (center), and the number of predators (right) of a given species.*
 467 *The trend lines are the mean effect of the scaled generality (log transformed) on the logit-AUC.*

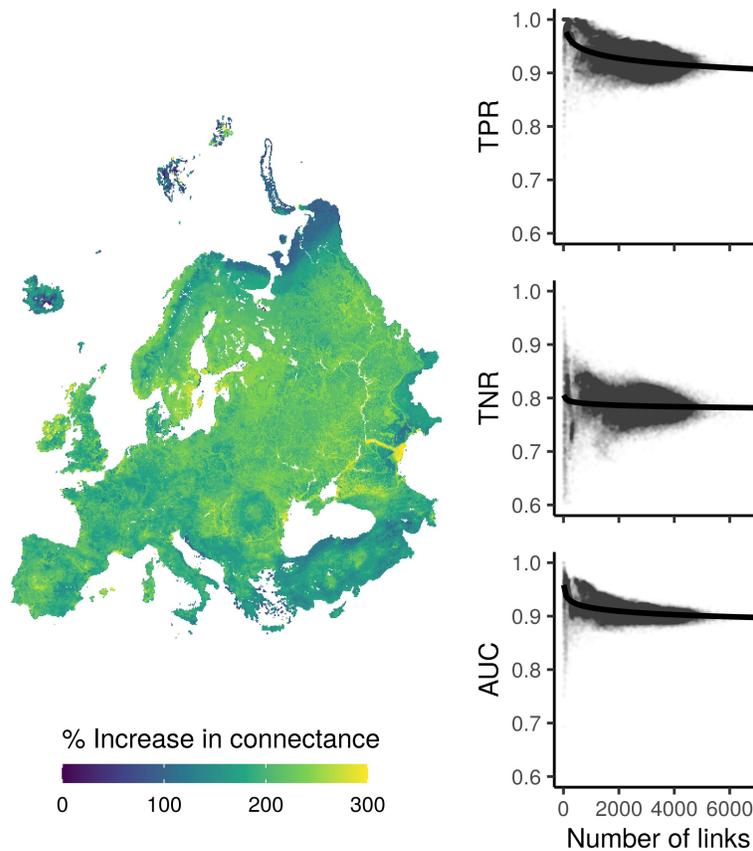


Figure 4: Spatial variation in predictability of local food webs. The left panel shows the overestimation of connectance of each 10km-cell food web. The right panels show the true positive rate (TPR; top), the true negative rate (TNR; middle), and area under the receiver operating curve (AUC; bottom) in relation to the complexity of the local web.