1   **ARTICLE TITLE** Bioregions are predominantly climatic for fishes of northern lakes

2   **RUNNING HEAD** Bioregionalization of inland fishes

3   **AUTHORS** Charlie J.G. Loewen (charlie.loewen@utoronto.ca)[1], Donald A. Jackson

4   (don.jackson@utoronto.ca)[1], Cindy Chu (cindy.chu@dfo-mpo.gc.ca)[1,2], Karen M. Alofs

5   (kmalofs@umich.edu)[3], Gretchen J.A. Hansen (ghansen@umn.edu)[4], Andrew E. Honsey

6   (ahonsey@usgs.gov)[4†], Charles K. Minns (ken@minns.ca)[1], Kevin E. Wehrly

7   (wehrlyke@michigan.gov)[3,5]

8   [1] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON,

9   Canada

10   [2] Ontario Ministry of Northern Development, Mines, Natural Resources and Forestry,

11   Peterborough, ON, Canada

12   [3] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

13   [4] Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St.

14   Paul, MN, USA

15   [5] Institute for Fisheries Research, Michigan Department of Natural Resources, Ann Arbor, MI,

16   USA

17   [†] Current address: Hammond Bay Biological Station, United States Geological Survey Great

18   Lakes Science Center, Millersburg, MI, USA

19   **CORRESPONDING AUTHOR** Charlie J.G. Loewen, email: charlie.loewen@utoronto.ca,

20   mail: 25 Willcocks Street, Toronto, Ontario, Canada M5S 3B2.

21   **BIOSKETCH** Charlie Loewen is a community and macroecologist. His research takes a

22   quantitative approach to understanding key processes and dynamics of freshwater ecosystems,

23   with a focus on their response to compounding environmental changes.

24   **AUTHOR CONTRIBUTIONS** All authors contributed to data compilation and manuscript

25   revision. CJGL led study conception and design, conducted the analyses, and composed the

26   initial draft manuscript.

1

2   MR. CHARLIE J.G. LOEWEN (Orcid ID : 0000-0002-4389-4134)

3   DR. CINDY CHU (Orcid ID : 0000-0002-1914-3218)

4

5

7

8

9   **ARTICLE TITLE**

10   Bioregions are predominantly climatic for fishes of northern lakes

11   **RUNNING HEAD**

12   Bioregionalization of inland fishes

13   **ABSTRACT**

14   **Aim:** Recurrent species assemblages integrate important biotic interactions and joint responses

15   to environmental and spatial filters that enable local coexistence. Here, we applied a bipartite

16   (site-species) network approach to develop a natural typology of lakes sharing distinct fish

17   faunas and provide a detailed, hierarchical view of their bioregions. We then compared the roles

18   of key biogeographic factors to evaluate alternative hypotheses about how fish communities are

19   assembled from the regional species pool.

20   **Location:** Ontario, Canada and the Upper Midwest, USA.

21   **Time period:** 1957–2017.

22   **Major taxa studied:** Freshwater fishes.

23   **Methods:** Bipartite modularity analysis was performed on 90 taxa from 10,016 inland lakes in

24   the Southwestern Hudson Bay, Mississippi River, and St. Lawrence River drainages, uncovering

25   bioregionalization of North American fishes at a large, subcontinental scale. We then used a

26   latent variable approach, pairing non-metric partial least-squares structural equation modelling

27   with multiple logistic regression, to show differences in the biogeographic templates of each

28   community type. Indicators of contemporary and historical connectivity, climate, and habitat

29   constructs were estimated using a geographic information system.

30   **Results:** Fish assemblages reflected broad, overlapping patterns of postglacial colonization,

31   climate, and geological setting, but community differentiation was most linked to temperature,

32   precipitation, and, for certain groups, lake area and water quality. Bioregions were also marked

33   by non-native species, showing broad-scale impacts of introductions to the Great Lakes and

34   surrounding basins.

35   **Main conclusions:** The dominant effects of climate across broad spatial gradients indicate

36   differing sensitivities of fish communities to rapidly accelerating climate change and

37   opportunities for targeted conservation strategies. By assessing biological variation at the level of

38   recurrent assemblages, we accounted for the non-stationarity of macroecological processes

39   structuring different sets of species on the landscape, and thus offer novel inference on the

40   assembly of inland fish communities.

45   **INTRODUCTION**

46   Species distributions, and their assembly into communities, reflect an array of contemporary and

47   historical filters operating at multiple spatial scales (Jackson, Peres-Neto, & Olden, 2001; Dias et

48   al., 2014; Mantyka-Pringle, Martin, Moffatt, Linker, & Rhodes, 2014; Ficetola, Mazel, &

49   Thuiller, 2017). Studies seeking to understand sources of variation across ecological

50   communities have often tested their compositional turnover or aggregate properties, such as

51   species richness (e.g. Legendre & Legendre, 1984; Mandrak, 1995; Oikonomou, Leprieur, &

52   Leonardos, 2014). More recently, tools drawn from graph theory have been applied to reveal

53   community typologies and natural boundaries of bioregions (i.e. geographic areas with similar

54   biota) based on multispecies distribution patterns (e.g. Carstensen & Olesen, 2009; Bloomfield,

55   Knerr, & Encinas-Viso, 2018; McGarvey & Veech, 2018). Because co-occurrence implies

56   broadly shared niche characteristics and colonization history, repeated assemblages provide

57   insight into formative ecological processes and a basis for setting habitat management units for

58   conservation planning and targeted interventions (Brooks et al., 2006; Olden et al. 2010; Wehrly,

59   Breck, Wang, & Szabo-Kraft, 2012; Montalvo-Mancheno, Ondei, Brook, & Buettel, 2020).

60   Community differentiation is especially informative for freshwater organisms, such as fishes,

61   which face strong dispersal constraints imposed by the hydrology of the watersheds they inhabit.

62   Freshwater ecosystems exhibit high biodiversity and rates of endemism globally, as well as

63   generally greater threats and species imperilment than terrestrial or marine habitats (Reid et al.,

64   2019). Yet, compared to other realms, inland waters have received little attention paid to

65   delineating recurrent assemblages and their underlying drivers at the broad spatial scales required

66   to develop regional, cross-jurisdictional conservation and management strategies (Brooks et al.,

67   2006; Wehrly et al., 2012; Oikonomou et al., 2014; Heino et al., 2021).

68   Our ability to predict changes in species diversity is challenged by the multitude of contributing

69   spatial and environmental factors (e.g. Dormann et al., 2013; Mantyka-Pringle et al., 2014;

70   Loewen, Strecker, Gilbert, & Jackson, 2020). For instance, spatial patterns of northern fish

71   communities may reflect historical imprints of glaciation (Bailey & Smith, 1981; Legendre &

72   Legendre, 1984; Dias et al., 2014), as well as present-day dispersal networks and regional habitat

73   availability (e.g. Hitt & Angermeier, 2008). As most freshwater organisms are ectothermic, they

74   are also intrinsically sensitive to temperature (Magnuson, Crowder, & Medvick, 1979). Thus,

75   while fish dependence on hydrologic connections to reach new habitats limits their ability to

76   move in response to interannual climate variations, longer-term temperature conditions should

77   constrain the distributional boundaries of species by restricting the fitness of edge populations

78   (Shuter & Post, 1990; Alofs, Jackson, & Lester, 2014). Persistence of species arriving at a new

79   location further depends on habitat quantity and quality. Larger waterbodies may offer more

80   space to more species (MacArthur & Wilson, 1963) and a greater diversity of habitats, including

81   competitive or predatory refugia (Tonn & Magnuson, 1982). Aspects of lake basin morphometry,

82   along with terrestrial (Heino et al., 2021) and geological setting (Conroy & Keller, 1976), also

83   drive gradients in surface water quality and biological production (D'Arcy & Carignan, 1997).

84   Thus, local habitat conditions combine with regional connectivity and climate filters to provide a

85 multilayered template for fish species sorting along geographic gradients (Smith & Powell,

86 1971).

87 Our objectives were two-fold. First, we leveraged an unique dataset of fish records across the

88 Laurentian Shield and surrounding Lowlands of Ontario, Canada and the Upper Midwest, USA,

89 to develop a site-species network graph. We then applied bipartite modularity analysis to detect

90 community typology and provide a hierarchical view of freshwater fish bioregionalization.

91 Second, we tested alternative hypotheses about the relative importance of biogeographic factors

92 driving community differentiation, predicting varying roles of connectivity, climate, and habitat

93 filters across groups. For instance, we hypothesized a key role of climate for northern fishes,

94 with contrasting responses to temperature and precipitation (owing to influence on lake thermal,

95 chemical, and hydrological regimes, including timing of ice-out and major run-off events)

96 revealing sensitive regions and community types to target in future climate change adaptation

97 strategies. We expected connectivity to constrain communities both within drainages and in

98 relation to deglaciation patterns, as glacial lakes provided an important means of dispersal and

99 southern sites may have had more time for postglacial colonization. Habitat filters were predicted

100 to be of lesser importance at broad spatial scales, though we anticipated potential differentiation

101 of communities linked to contrasting shield (metamorphic and igneous) and sedimentary

102 lithologies as mediated by their effects on water chemistry (e.g. alkalinity and nutrient levels).

103 **MATERIALS AND METHODS**

104 **Study system**

105 We conducted a large data synthesis to assess fish assemblages and their potential driving factors

106 across a region of inland lakes in northeastern North America spanning over 13.4° of latitude

107 (1,500 km), 21.9° of longitude (1,700 km), and 580 m in elevation (Figure 1a). Fish occurrence

108 records and associated environmental measurements in portions of the Southwestern Hudson

109 Bay (n = 3,637), St. Lawrence (n = 5,757), and Mississippi River (n = 622) drainages were

110 compiled from existing datasets in Ontario (Aquatic Habitat Inventory; Dodge, Goodchild, Tilt,

111 Waldriff, & MacRitchie, 1987), Michigan (Status and Trends; Wehrly, Carter, & Breck, 2021),

112 and Minnesota (Index of Biotic Integrity; Drake & Pereira, 2002). To facilitate visualization of

113 bioregions and certain other geospatial analyses, major drainage areas were further divided into

114   secondary (Figure 1b) and smaller tertiary drainages (equivalent to sub-sub-drainages in Canada

115   and hydrologic unit code 8 drainages in USA; Figure 1c).

116   Fishes were sampled from lakes, ponds, reservoirs, and embayments (herein referred to as lakes

117   or sites) during the open-water season over a 60-year period (1957–2017). Although we were

118   unable to explicitly consider species turnover during this time (e.g., Lynch et al., 2016; Cazelles

119   et al., 2019; Keller, Heneberry, & Edwards, 2019), we found that community differentiation was

120   robust to minor changes in species occurrence (see our sensitivity analysis for non-native species

121   below) and temporal shifts in climate were generally small compared to spatial differences. For

122   example, the mean annual air temperatures of sampled lakes (averaged across the study period)

123   ranged from -4.1 to 9.7 °C, while changes in mean temperatures between the 1950s and 2010s

124   ranged from -0.4 to 1.1 °C across sites (Wang, Hamann, Spittlehouse, & Carroll, 2016). While

125   potentially contributing to shifts in distribution or relative abundance for some populations (e.g.

126   Finigan, Mandrak, & Tufts, 2018), warming is unlikely to have confounded multispecies

127   associations and the delineation of bioregions at our scale of analysis. Samples were collected

128   using a combination of netting, trapping, and electrofishing, and quantitative survey results were

129   collapsed to species presence/absence to reduce biases associated with different methods

130   (Jackson & Harvey, 1997). Records of 132 fish taxa were initially compiled across 11,112 lakes,

131   but several species and lakes were excluded from analysis. We first removed lakes lacking

132   maximum depth or surface area measurements. We further omitted records for fishes with

133   riverine life histories because they were likely transient when captured and therefore infrequently

134   sampled in lake ecosystems. These undersampled species were identified based on expert

135   opinion informed by comparisons to other sampling programs in the region (e.g. Lester,

136   Marshall, Armstrong, Dunlop, & Ritchie, 2003; Drake, & Valley, 2005; Sandstrom, Rawson, &

137   Lester, 2013) and aspects of organismal biology and encounter rates (relating to sampling

138   methods/design; see Table S1.1 in Supporting Information for list of sampled fishes). Finally,

139   because we were interested in patterns of interacting species, we removed sites with only a single

140   species as well as species with only a single occurrence. Our final dataset included 90 taxa across

141   10,016 lakes (mean species richness = 8).

142   **Bipartite network modularity analysis**

143   We modelled an unweighted network graph comprised of lakes and fishes as separate sets of

144   nodes, with edges linking species to the sites at which they were observed. While there are

145   several approaches to grouping communities, most fail to fully exploit the bipartite nature of site-

146   species networks to reveal both sites with similar species and species tending to form

147   assemblages. Distance-based methods have conventionally been used to cluster sites based on

148   their species composition and/or spatial configurations (e.g. Legendre & Legendre, 1984;

149   Oikonomou et al., 2014); however, these procedures flatten bipartite networks, reducing species

150   information to one of many site dissimilarity indices. Modularity analysis of site-species

151   networks has proven an effective alternative to uncovering concise bioregions and their

152   transitional zones (see participation coefficients; Guimerà & Amaral 2005) across spatial scales

153   (e.g. Bloomfield et al., 2018; McGarvey & Veech, 2018; Bernardo-Madrid et al. 2019).

154   We used the DIRTLPAb+ algorithm of Beckett (2016), which is a label propagation approach to

155   identifying groups of sites and species that are linked together more densely than others by

156   maximizing Barber's modularity (see Appendix 1 for further details). Barber's index performs

157   well for defining groups in bipartite networks (Thébault 2013) and classifies both sets of nodes

158   (sites and species) simultaneously. We applied the algorithm using the computeModules function

159   in the 'bipartite' package (Dormann, Gruber, & Fründ, 2008) with default settings for random

160   initializations (implemented with R 3.6.1; R Development Core Team). This internally repeating

161   procedure was itself replicated ten times (with different seeds) to check consistency and we

162   recorded site and species assignments of the trial with the highest modularity (Table S1.2).

163   Significance of resultant modularity was evaluated by one-sided randomization tests based on

164   100 null networks obtained from the efficient, sequentially swapping curveball algorithm

165   (Strona, Nappo, Boccacci, Farrorini, & San-Migeul-Ayanz, 2014). The curveball approach

166   produces uniformly distributed null matrices maintaining row and column sums and was

167   implemented with 1,000 burn-in and 1,000 thinning steps across 20 unique chains using the

168   nullmodel and oecosimu functions in the 'vegan' package (Oksanen et al., 2020). Simulations

169   were checked for convergence prior to use. Participation coefficients measuring the among-

170   module connectivity of each site were calculated using the czvalues function (Dormann et al.,

171   2008). To reveal finer-scale structures, we invoked a second round of analysis restricted to each

172   primary module. This multiscale bioregionalization was summarized spatially as the dominant

173   primary (or sub) module across tertiary drainages (with ties adopting the most dominant type

174   among neighbours).

175   Environmental and spatial factors

176   We estimated environmental and spatial variables to capture key dimensions of biogeographic

177   constructs (connectivity, climate, and habitat; Table S1.3) hypothesized to determine natural site

178   typology (see statistical analysis section below for details of latent variable approach). A

179   connectivity construct was formed by regional habitat density, river network dispersal distance,

180   glacial lake/marine distance, and time since glaciation. Variables were estimated from remote-

181   sensing data using a geographic information system (ArcGIS Pro 2.6.2; Esri, Redlands,

182   California, USA). Regional habitat density was calculated by summarizing all lake, pond, and

183   reservoir features within tertiary drainages and dividing the area of this combined habitat by the

184   total area of the watershed (excluding the Great Lakes; 1:50,000 scale or better). River network

185   dispersal distances estimated the lengths of potential contemporary dispersal routes to secondary

186   drainage features in either the Southwestern Hudson Bay drainage area (Lake Winnipeg or

187   Hudson Bay), St. Lawrence drainage area (Lake Superior, Lake Michigan, Lake Huron, Lake

188   Erie, Lake Ontario, or St. Lawrence River) or Mississippi River (Figure 1d; see Appendix 1 for

189   further details). Historical glacial lake/marine distance was calculated as the geodesic distance

190   from each lake to the nearest glacial lake or marine extent polygon (Figure S1.1). Time since

191   glaciation (i.e. age of lake) was calculated by spatially joining sampling points near the center of

192   each lake with a time-series of ice coverage polygons and determining the most recent overlap.

193   Estimated historical ice, lake, and marine extents were obtained at approximately 500-year

194   resolution and 1:7,000,000 scale (Dyke, 2004).

195   A climate construct was formed by mean warmest month temperature, mean coldest month

196   temperature, May to September total precipitation, and total precipitation as snow. Climate

197   variables were elevation-adjusted point estimates of average annual conditions at each lake over

198   the entire study period. We extracted and downscaled estimates from moderate resolution

199   Climate Research Unit data (4 km gridded; Mitchell & Jones, 2005) using ClimateNA v6.40

200   (Wang et al., 2016; Figure S1.2).

201   Finally, a habitat construct was formed by surface area, maximum depth, Secchi depth, and

202   proportion shield lithology. Physical size measurements were retrieved from existing records.

203   Secchi depth, which measures water clarity, provided a proxy for basin productivity given our

204   lack of consistent water chemistry data. Secchi depths were measured at the majority (98%) of

205   sites. For the 191 sites without direct measures, we estimated Secchi depths from their regional

206   relationships to maximum depth. Here, separate linear regression models were constructed for

207   each secondary drainage area stratified into depth categories (0–20 m, 20–40 m, and >40 m) and

208   used to predict missing values. We estimated proportion shield lithology to capture the dominant

209   geological gradient in our study region (i.e. surface exposed portions of igneous and

210   metamorphic Laurentian Shield contrasting the sedimentary rocks of the Interior and Hudson

211   Bay Lowlands). Surficial weathering of rocks provides a natural source of variation in water

212   quality (e.g. ion chemistry) and biological production, especially between major lithological

213   units (Conroy & Keller, 1976). Lake polygons were overlain with detailed lithological maps

214   compiled by Harmann & Moosdoft (2012) at an average scale of 1:3,750,000 and used to

215   calculate the proportions of each basin comprised of either shield or sedimentary rock (Figure

216   S1.3).

217   **Statistical analysis**

218   To evaluate the roles of spatial and environmental constraints on fish communities, we applied

219   non-metric partial least-squares structural equation modelling (NM-PLS-SEM, also called PLS

220   path modelling) paired with multiple logistic regression (Petrarca, Russolillo, & Trinchera,

221   2017). PLS-SEM is a variance-based approach to estimating unobserved latent variables and

222   their relations by a series of simple or multiple ordinary least-squares (OLS) regressions

223   (Sanchez, 2013). A key application of PLS-SEM is for building latent variables as formative

224   constructs using a composite measurement model. Here, latent variables (or constructs) are

225   formed as linear combinations of their manifest variables (or observed indicators) by an iterative

226   process of optimizing weights and paths in the outer measurement and inner structural equation

227   models (Sanchez, 2013). Thus, the algorithm offers an intuitive inferential framework while also

228   applying dimension reduction to develop proxies for multidimensional concepts, such as

229   connectivity, climate, and habitat (e.g. Loewen et al., 2020). PLS methods can also be performed

230   with (non-metric) nominal or ordinal indicators by introducing an optimal scaling step (Petrarca

231   et al., 2017). Here, non-metric variables are assigned values on an interval scale by optimizing

232   additional scaling parameters as part of the iterative PLS-SEM process.

233   We performed NM-PLS-SEM to assess the roles of biogeographic factors using the 'path

234   scheme' method and formative measurement (mode B) with the plspm function in the 'plspm'

235   package (Sanchez et al., 2017). Indicator variables were natural log- or logit-transformed (for

236   proportional data; adjusting by minimum observed values where necessary) and standardized

237   (centered at zero and scaled to unit variance) prior to analysis. Connectivity, climate, and habitat

238   constructs were formed as linear combinations of their respective numeric indicators and used to

239   predict individual modules. The approach was also applied to multilevel categorical outcomes of

240   our modularity analysis. However, because estimating a binomial response by OLS is less than

241   ideal, even with the optimal scaling algorithm, we used a modified approach (Petrarca et al.,

242   2017) first obtaining latent variable scores from NM-PLS-SEM and then re-estimating inner

243   model path coefficients by multiple logistic regression (generalized linear regression with

244   binomial error distribution and a logit-link function; glm function in base R). The explanatory

245   power of logistic models was measured by Tjur's R^2 using the RsqGLM function in the

246   'modEvA' package (Barbosa, Real, Muñoz, & Brown, 2015). Tjur's coefficient of discrimination

247   is an analogue to the coefficient of determination in linear regression but calculated as the

248   difference in mean fitted values for successful and failed predictions (Tjur, 2009). Finally,

249   because we anticipated potential collinearity among constructs, we used a variation-partitioning

250   framework to reveal their unique, shared, and total explanatory powers. Variation partitioning

251   based on Tjur's R^2 from logistic regression was performed using the varPart function (Barbosa

252   et al., 2015) and results were visualized by approximately area-proportional Euler diagrams

253   using the euler function in the 'eulerr' package (Larsson, 2020). To emphasize broad,

254   biogeographic variation in fish assemblages, we limited our statistical analysis to primary

255   modules. To test the influence of non-native species on bioregionalization patterns (Table S1.1),

256   we conducted sensitivity analyses to determine changes to site and species assignments, as well

257   as interpretations of formative processes, when either non-native species were excluded from the

258   species pool or sites with non-natives were excluded from the site pool.

259   **RESULTS**

260   We identified clear spatial patterns of fish community differentiation, which we summarized by

261   tertiary watershed to identify tractable northern bioregions (Figures 2, S1.4, and S1.5). Sites from

262   module 1 (n = 1,258 lakes; referred to as the southern group) clustered in southern portions of the

263   St. Lawrence and the Mississippi River drainages. Module 2 (n = 3,010; the trout group) and

264   module 3 (n = 1,851; the central Great Lakes group) were situated in more northern and east-

265   central portions of the St. Lawrence drainage, respectively. Module 4 lakes (n = 3,897; the

266   northern group) were dominant in northern areas draining into the Southwestern Hudson Bay.

267   However, none of the assemblages was restricted to any single drainage. Primary modules were

268   internally consistent and had significant non-random structure with a modularity value (Q) of

269   0.332 (P <0.01; Table S1.2). Participation coefficients showed that among-module connections

270   were generally denser in the southern half of the study region, specifying a bioregional transition

271   zone stretching from northern portions of the Mississippi River drainage to the centre of the

272   Great Lakes Basin, as well as adjacent to the St. Lawrence River (Figure 2b). Sub-modularity

273   analysis, where procedures were repeated with restricted site and species pools, revealed further

274   variation within primary modules (Figures 2c–f). Submodule assignments were weaker and more

275   variable across trials but significantly non-random (P <0.01; Table S1.2). Module 1 was shown

276   to possess finer-scale north-south and east-west structure (Figure 2c) while module 4 revealed

277   two central community types (in the north-west and south-east portions of the Southwestern

278   Hudson Bay drainage) with greater heterogeneity to the south-west (Figure 2f).

279   We similarly found groups of associated fishes (Figure 3). Module 1 (n = 44; the southern group)

280   contained the most species, including several warmwater fishes with limited occurrence near

281   their northern range limits in our dataset (especially submodule 1.4). Module 2 (n = 15; the trout

282   group) was characterized by several salmonids, i.e. lake trout (Salvelinus namaycush), brook

283   trout (Salvelinus fontinalis), round whitefish (Prosopium cylindraceum), and rainbow trout

284   (Oncorhynchus mykiss), and a subset of leuciscid minnows. Module 3 (n = 8; the central Great

285   Lakes group) contained species with known associations to the St. Lawrence River and Lake

286   Ontario, such as American eel (Anguilla rostrata) and rainbow smelt (Osmerus mordax), along

287   with certain warm and coolwater sportfishes, including muskellunge (Esox masquinongy),

288   smallmouth bass (Micropterus dolomieu), and rock bass (Ambloplites rupestris). Module 4 (n =

289   23; the northern group) consisted of several species common to northern lakes, including yellow

290   perch (Perca flavescens), white sucker (Catostomus commersonii), and the sportfishes walleye

291   (Sander vitreus), lake whitefish (Coregonus clupeaformis), cisco (C. artedi), and northern pike

292   (Esox lucius).

293   Climate was the dominant factor underlying community typology, while the importance of

294   connectivity and habitat filters varied across groups. The influence of biogeographic constructs

295   (modelled as latent variables; see meta-model in Figure 4) on the occurrence of each assemblage

296   and the roles of individual environmental and spatial indicators can be interpreted from the

297   model coefficients presented in Figure 5. Here, indicator weights and loadings describe the

298   contributions and correlations of indicators to their respective latent variables in the outer

299   measurement model (Figure 5a), while logistic path coefficients from the inner structural model

300   show the magnitude and directionality (negative/positive) of relationships between latent

301   variables and each of the primary modules (Figure 5b). For example, connectivity constructs

302   were mostly driven by time since glaciation, which contributed positively to module 1 (the

303   southern group), negatively to module 4 (the northern group), and negligibly to modules 2 and 3

304   (the trout and central Great Lakes groups). River network dispersal distance had a stronger

305   negative association to the connectivity construct for module 3, but the small path coefficient for

306   connectivity indicated that its effect was negligible. Similarly, error bars were inflated for

307   indicators of connectivity contributing to module 2 because the construct had a minor effect and

308   coefficients flipped between positive and negative across bootstrapped samples. Climate

309   indicators were generally correlated with one another (and time since glaciation; Table S1.4) and

310   their individual contributions varied across modules. Precipitation as snow had a relatively

311   strong negative correlation with module 1 and positive correlation with module 2, while modules

312   3 and 4 were more associated with higher and lower temperatures, respectively. Habitat

313   indicators also had variable contributions and were most relevant to modules 2 and 4.

314   Specifically, module 2 involved smaller lakes with greater Secchi depth, while module 4 was

315   linked to greater surface area and proportion shield lithology. Categorical analysis indicated that

316   climate was the principal driver of community differentiation overall (Tables S1.5 and S1.6).

317   The predominance of climate in the bioregionalization of northern fishes was further supported

318   by our partitioning of latent variable explanatory power (Figure 6). Here, we found that climate

319   was the most predictive construct for each assemblage except module 2 (the trout group), which

320   was driven by lake area and water clarity. Modules 1 and 4 (the southern and northern groups)

321   had the greatest classification success, with strong explanatory overlap between climate and

322   connectivity (the latter associated with time since glaciation). Variation in connectivity and

323   habitat were of lesser importance to module 3 (the Great Lakes group), contributing almost no

324 additional prediction after accounting for temperature and precipitation effects (see Tables S1.7
325 and S1.8 for detailed results).

326 The most frequent non-natives were common carp (Cyprinus carpio; module 1), rainbow trout
327 (module 2), and rainbow smelt (module 3; Table S1.1). We found that only two native species,
328 spoonhead sculpin (Cottus ricei) and white sucker, changed modules when non-native species (n
329 = 11) were excluded from the species pool (both from module 4 to 2). Assignments changed for
330 646 lakes (6.4%), the majority changing from module 4 to 2 (261) or 3 to 2 (134; Table S1.9).
331 Changes to bioregions were few and mostly in transitional regions of southern Ontario (Figure
332 S1.6b). However, when sites with non-native species were excluded (n = 944; 474 from module
333 1, 243 from module 2, 197 from module 3, and 30 from module 4), assignments changed for nine
334 native species (one from module 1 to 3, one from module 1 to 4, and seven from module 3 to 1)
335 and 1,766 lakes (19.4%). Most site differences involved module 3 (94%), which largely merged
336 with the southern group (Figure S1.6c), but interpretations of how communities were shaped by
337 environmental and spatial processes were generally robust to the inclusion or exclusion of non-
338 native species (Figures S1.7–S1.10).

339 **DISCUSSION**

340 We leveraged exceptional sampling effort across our study region to present a detailed view of
341 freshwater bioregions and their underlying drivers at a broad, subcontinental scale. As
342 bioregionalization is inherently taxa- and scale-dependent (McGarvey & Veech, 2018), we
343 provide a novel classification of northern fish communities and advance understanding of their
344 assembly from the regional species pool. While global maps of fish biogeographical zones have
345 been produced (Abell et al., 2008; Leroy et al., 2019), they have aggregated disparate surveys
346 across larger watersheds and thus have limited use for differentiating local communities or
347 tractable management units. Alternative indices and module detection algorithms also exist (e.g.
348 Guimerà & Amaral, 2005; Rosvall & Bergstrom, 2008), which may reveal different patterns (see
349 Thébault, 2013; Bloomfield et al., 2018). Similar to the flow-based Infomap method (Edler,
350 Guedes, Zizka, Rosvall, & Antonelli, 2017), Beckett's (2016) label propagation identifies groups
351 of highly associated sites and species simultaneously (without unipartite projection). Thus, our
352 application of site-species network analysis builds on existing knowledge, revealing both lakes
353 with similar fish assemblages and fishes tending to co-occur. By partitioning our network

354   independently of spatial and environmental contexts, we were able to test the relative importance

355   of key formative processes determining different sets of species and generate several hypotheses

356   for future study.

357   As species sorting along geographic gradients involves multiple, often highly correlated factors,

358   we evaluated drivers using a latent variable approach to capture broadly defined climate,

359   connectivity, and habitat filters (Dormann et al., 2013; Loewen et al., 2020). We discovered that

360   climate had the greatest effect in differentiating fish communities, similar to recent reports for

361   subcontinental patterns of terrestrial fauna (Ficetola et al., 2017). Thermal or hydrological

362   regimes may also account for finer-scale differences, such as the sub-modular structure observed

363   within the southern and northern groups (modules 1 and 4; Figure 2) or variation among lakes

364   within tertiary catchments (e.g. Wehrly et al., 2012; Figures S1.4 and S1.5). We found that

365   historical connectivity had a related, albeit lesser, influence on modules 1 and 4, pointing to

366   differences in postglacial colonization opportunities between lakes in more southern and

367   northern areas of our study region (Bailey & Smith, 1981; Legendre & Legendre, 1984; Dias et

368   al., 2014). However, distances to glacial lakes do not necessarily capture dispersal routes and

369   given the degree of overlap between glacial recession and present temperature gradients (Figures

370   S1.1 and S1.2), analysis of shifting recolonization pathways from glacial lakes and their outflows

371   may offer clearer support (e.g. Mandrak & Crossman, 1992). Although habitat factors are often

372   linked to species sorting at smaller scales (Jackson et al., 2001; Lansac-Tôha et al., 2021), we

373   found evidence of broad-scale filtering by lake surface area, Secchi depth, and shield lithology.

374   For instance, the trout group (module 2) showed affinity to smaller lakes with greater Secchi

375   depth, which serves as an integrative proxy for multiple terrestrial factors (e.g. sediment loading)

376   and internal productivity (Tilzer, 1988). Smaller lakes with greater water clarity may thus

377   provide greater open-water foraging success for salmonids (Ortega, Figueiredo, da Graça,

378   Agostinho, & Bini, 2020). In contrast, the northern group (module 4) was linked to larger lakes

379   with greater influence of shield lithology as well as those with reduced water clarity (as

380   encountered in areas overlaying glacial lake beds). Here, species such as walleye may experience

381   optimal foraging in habitats with lower light intensity (Lester, Dextrase, Kushneriuk, Rawson, &

382   Ryan, 2004). Bioregional transition zones also matched southern boundaries of the Laurentian

383   Shield (Figures 2b and S1.3), showing potential influence of geological setting via effects on

384   water chemistry (Conroy & Keller, 1976; Neff & Jackson, 2012).

385 In addition to site typology, we revealed recurrent groups of fishes, implying broadly shared

386 natural histories and niche space. We highlight several sportfishes of interest to fisheries

387 management but provide only a cursory examination of fish module composition. Future

388 research should test for associations among species within and between groups (e.g. Cordero &

389 Jackson, 2019), including their functional trait and phylogenetic similarities. For instance, the

390 predominance of climate filters should sort species into thermal guilds (Magnuson et al., 1979),

391 while connectivity constraints may produce assemblages related by their dispersal ability (i.e.

392 swimming proficiency and tolerance to flow conditions) or common ancestry (Comte & Olden,

393 2018). Habitat filters should further aggregate species with similar life-history strategies,

394 including requirements for forage, reproduction, and overwintering (e.g. Tonn & Magnuson,

395 1982). Explicit consideration of competition and predation among co-occurring species may

396 yield additional mechanistic insights; however, as with habitat filters, we anticipate that species

397 interactions will be most important at smaller scales of inquiry (Jackson et al., 2001).

398 Our findings have applications in conservation biogeography, revealing opportunities for

399 targeted climate adaptation strategies. Setting conservation priorities for regional planning (e.g.

400 protected areas or resource extraction) requires defining management units and understanding

401 their formative processes (Brooks et al., 2006; Oikonomou et al., 2014; Montalvo-Mancheno et

402 al., 2020). Applying space-for-time substitution, the central roles of temperature and

403 precipitation provide evidence that fish communities are sensitive to climatic changes (Alofs et

404 al., 2014; Lynch et al., 2016), but future impacts will depend on community type. We also found

405 that connectivity does not appear to have been a major factor preventing lacustrine fishes from

406 matching their temperature preferences within catchments. For instance, the southern group

407 (module 1) was much more associated with warmer temperatures and less precipitation as snow

408 than contemporary dispersal indicators, suggesting potential to expand northward as

409 temperatures rise and connectivity permits. Indeed, there is evidence that certain warm and

410 coolwater sportfishes are already expanding their northern ranges via multiple mechanisms

411 (Alofs et al., 2014), including anthropogenic introductions. Coldwater fishes in the trout group

412 (module 2) were linked to cooler summers and more snow, but habitat size and quality had a

413 greater influence. Thus, this group may be relatively sensitive to hydrological changes and

414 increasing water column turbidity or brownification. In contrast, the Great Lakes group (module

415 3) was clustered in warmer and snowier areas around the Great Lakes in south-central Ontario.

416   These results suggest that Great Lakes effects create conditions for warm and coolwater fishes

417   that may offer partial buffering against climatic changes, though future climates are uncertain

418   and will be regionally variable (Notaro, Bennington, & Vavrus, 2015). Finally, the northern

419   group (module 4) was most sensitive to warming (especially in winter months) and potentially

420   more dispersal limited (based on the greater importance of river network dispersal distance),

421   indicating that these communities may require extra attention for climate change adaptation.

422   However, our estimates of dispersal limitation reflect only the natural connectivity of sites within

423   their respective drainages.

424   Establishment of introduced fishes outside of historical ranges indicates that natural colonization

425   either lags changing conditions or is impeded by other environmental, connectivity, or biotic

426   factors (e.g. Allee effects). Fish movements are constrained by physical barriers (e.g. dams,

427   culverts, and falls) and limited hydrological connections between catchments (especially along

428   continental divides) may induce critical bottlenecks for migration (Jackson et al., 2001; Melles,

429   Chu, Alofs, & Jackson, 2015). Species also have thermal and depth requirements for dispersal,

430   possibly limiting movement by coldwater species as climates warm. While assemblages

431   generally reflect shared colonization histories of dispersal along lake and river networks, effects

432   of increasing fragmentation on future range shifts and persistence may vary (e.g. Herrera-R et al.,

433   2020). Actions such as human-assisted migration or construction of fishways warrant

434   consideration, weighing potential benefits with risks of unintended consequences from

435   introducing harmful species (Olden et al., 2010), including pathogens.

436   Human introductions have already impacted the spatial structure of fish communities globally

437   (Olden et al., 2010; Alofs et al., 2014). For instance, regional translocations contribute to biotic

438   homogenization among lakes with higher angling pressure or nearer urban centres (Olden,

439   Kennard, & Pusey, 2008). Stocking programs and movement of baitfishes may have also

440   contributed to a minor role of connectivity for the trout group (module 2), as introductions

441   obscure how species are naturally filtered from—or serve as additions to—the regional pool

442   (Alofs et al., 2014; Bernardo-Madrid et al., 2019; Cazelles et al., 2019). Although we were

443   unable to account for regional translocations, the small effect of excluding records for non-native

444   species (from outside the study region) on our understanding of community differentiation

445   highlights the robustness of our results to potential sampling errors. While biases associated with

446   site selection (e.g. targeting lakes with fishes of greater management interest) or imperfect
447   detection can confound species distribution patterns, errors were mitigated by our large sample
448   size and the relative insensitivity of our analytical approach to missing data. However, removing
449   a species record does not remove its community impact, and our sensitivity analysis showed a
450   major role of invaded communities in the differentiation of the southern and Great Lakes groups
451   (modules 1 and 3). Whether driven by species interactions, loss of coverage, or some other
452   aspect of invaded sites, results excluding lakes with non-native species both revealed the imprint
453   of introductions on fish communities in the region and upheld our interpretations of key
454   formative processes and climatic sensitivities. Despite non-native species typically establishing
455   by anthropogenic means (such as shipping, recreational stocking, or escape from
456   aquaculture/aquaria), they contribute to contemporary bioregions and are often management
457   priorities as fisheries or biotic stressors on native communities (e.g. Loewen et al., 2020).

458   Finally, as we focus on natural drivers of community differentiation, important questions remain
459   about how anthropogenic factors overlay observed patterns. For instance, regional land-use can
460   cause changes in downstream water quality and fish community structure (e.g. Mantyka-Pringle
461   et al., 2014). In our study region, most urban and agricultural developments are concentrated at
462   southern latitudes where fishes generally contend with heightened angling pressure (Lester et al.,
463   2003). Many thousands of lakes in central Ontario have also experienced anthropogenic
464   acidification, leaving lasting impacts on aquatic food webs (Keller et al., 2019) and contributing
465   to a suite of waterscape stressors affecting freshwater ecosystems at multiple scales (Heino et al.,
466   2021). Further integration of these and other human activities will be needed to reconcile natural
467   patterns of freshwater biogeography with ongoing global changes and spread of invasive species.

468   **REFERENCES**

469   Abell, R.A., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., … Petry, P.
470   (2008). Freshwater ecoregions of the world: a new map of biogeographic units for freshwater
471   biodiversity conservation. BioScience, 58, 403–414.

472   Alofs, K.M., Jackson, D.A., & Lester, N.P. (2014). Ontario freshwater fishes demonstrate
473   differing range-boundary shifts in a warming climate. Diversity and Distributions, 20, 123–136.

474   Bailey, R.M., & Smith, G.R. (1981). Origin and geography of the fish fauna of the Laurentian

475   Great Lakes Basin. Canadian Journal of Fisheries and Aquatic Sciences, 38, 1539–1561.

476   Barbosa, A.M., Real, R., Muñoz, A.-R., & Brown, J.A. (2015). New measures for assessing

477   model equilibrium and prediction mismatch in species distribution models. Diversity and

478   Distributions, 19, 1333–1338.

479   Beckett, S.J. (2016). Improved community detection in weighted bipartite networks. Royal

480   Society Open Science, 3, 140536.

481   Bernardo-Madrid, R., Calatayud, J., González-Suárez, M., Rosvall, M., Lucas, P.-M., Rueda, M.,

482   … Revlla, E. (2019). Human activity is altering the world's zoogeographical regions. Ecology

483   Letters, 22, 1297–1305.

484   Bloomfield, N.J., Knerr, N., & Encinas-Viso, F. (2018). A comparison of network and clustering

485   methods to detect biogeographical regions. Ecography, 41, 1–10.

486   Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., Gerlach, J., Hoffmann, M., Lamoreux,

487   J.F., … Rodrigues, A.S.L. (2006). Global biodiversity conservation priorities. Science, 313, 58–

488   61.

489   Carstensen, D.W., & Olesen, J.M. (2009). Wallacea and its nectarivorous birds: nestedness and

490   modules. Journal of Biogeography, 36, 1540–1550.

491   Cazelles, K., Bartley, T., Guzzo, M.M., Brice, M.-H., MacDougall, A.S., Bennett, J.R., …

492   McCann, K.S. (2019). Homogenization of freshwater lakes: recent compositional shifts in fish

493   communities are explained by gamefish movement and not climate change. Global Change

494   Biology, 25, 4222–4233.

495   Comte, L., & Olden, J.D. (2018). Evidence for dispersal syndromes in freshwater fishes.

496   Proceedings of the Royal Society B, 285, 20172214.

497   Conroy, N., & Keller, W. (1976). Geological factors affecting biological activity in Precambrian

498   shield lakes. The Canadian Mineralogist, 14, 62–72.

499   Cordero, R.D., & Jackson, D.A. (2019). Species-pair associations, null models, and test of

500   mechanisms structuring ecological communities. Ecosphere, 10, e02797.

501 D'Arcy, P., & Carignan, R. (1997). Influence of catchment topography on water chemistry in

502 southeastern Québec Shield lakes. Canadian Journal of Fisheries and Aquatic Sciences, 54,

503 2215–2227.

504 Dias, M.S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J.-F., … Tedesco, P.A.

505 (2014). Global imprint of historical connectivity on freshwater fish biodiversity. Ecology Letters,

506 17, 1130–1140.

507 Dodge, D.P., Goodchild, G.A., Tilt, J.C., Waldriff, D.G., & MacRitchie, I. (1987). Manual of

508 instructions: Aquatic habitat inventory surveys. Ontario Ministry of Natural Resources, Toronto,

509 Canada.

510 Dormann, C.F., Gruber, B., & Fründ, J. (2008). Introducing the bipartite package: analysing

511 ecological networks. R News, 8, 8–11.

512 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G, … Lautenbach, S. (2013).

513 Collinearity: a review of methods to deal with it and a simulation study evaluating their

514 performance. Ecography, 36, 27–46.

515 Drake, M.T., & Pereira, D.L. (2002). Development of a fish-based index of biotic integrity for

516 small inland lakes in Central Minnesota. North American Journal of Fisheries Management, 22,

517 1105–1123.

518 Drake, M.T., & Valley, R.D. (2005). Validation and application of a fish-based index of biotic

519 integrity for small central Minnesota lakes. North American Journal of Fisheries Management,

520 25, 1095–1111.

521 Dyke, A.S. (2004). An outline of North American deglaciation with emphasis on central and

522 northern Canada. Geological Survey of Canada, Ottawa, Canada.

523 Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2017). Infomap bioregions:

524 interactive mapping of biogeographical regions from species distributions. Systematic Biology,

525 66, 197–204.

526 Ficetola, G.F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical

527 boundaries. Nature Ecology and Evolution, 1, 0089.

528 Finigan, P.A., Mandrak, N.E., & Tufts, B.L. (2018). Large-scale changes in the littoral fish
529 communities of lakes in southeastern Ontario, Canada. Canadian Journal of Zoology, 96, 753–
530 759.

531 Guimerà, R, & Amaral, L.A.N. (2005). Functional cartography of complex metabolic networks.
532 Nature, 433, 895–900.

533 Harmann, J., & Moosdoft, N. (2012). The new global lithological map database GLiM: a
534 representation of rock properties at the Earth surface. Geochemistry, Geophysics, Geosystems,
535 13, Q12004.

536 Heino, J., Alahuhta, J., Bini, L.M., Cai, Y., Heiskanen, A.-S., Hellsten, S., … Angeler, D.G.
537 (2021). Lakes in the era of global change: moving beyond single-lake thinking in maintaining
538 biodiversity and ecosystem services. Biological Reviews, 96, 89–106.

539 Herrera-R, G.A., Oberdorff, T., Anderson, E.P., Brosse, S., Carvajal-Vallejos, F.M., … Tedesco,
540 P.A. (2020). The combined effects of climate change and river fragmentation on the distribution
541 of Andean Amazon fishes. Global Change Biology, 26, 5509–5523.

542 Hitt, N.P., & Angermeier, P.L. (2008). Evidence for fish dispersal from spatial analysis of stream
543 network topology. Journal of the North American Benthological Society, 27, 304–320.

544 Jackson, D.A., & Harvey, H.H. (1997). Qualitative and quantitative sampling of lake fish
545 communities. Canadian Journal of Fisheries and Aquatic Sciences, 54, 2807–2813.

546 Jackson, D.A., Peres-Neto, P.R., & Olden, J.D. (2001). What controls who is where in freshwater
547 fish communities–the roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries
548 and Aquatic Sciences, 58, 157–170.

549 Keller, W.B., Heneberry, J., & Edwards, B.A. (2019). Recovery of acidified Sudbury, Ontario,
550 Canada, lakes: a multi-decade synthesis and update. Environmental Reviews, 27, 1–16.

551 Lansac-Tôha, F.M., Bini, L.M., Heino, J., Meira, B.R., Segovia, B.T., Pavanelli, C.S., … Velho,
552 L.F.M. (2021). Scale-dependent patterns of metacommunity structuring in aquatic organisms
553 across floodplain systems. Journal of Biogeography, 48, 872–885.

554 Larsson, J. (2020). eulerr: area-proportional Euler and Venn diagrams with ellipses. R package
555 version 6.1.0.

556 Legendre, P., & Legendre, V. (1984). Postglacial dispersal of freshwater fishes in the Québec
557 Peninsula. Canadian Journal of Fisheries and Aquatic Sciences, 41, 1781–1802.

558 Lester, N.P., Marshall, T.R., Armstrong, K., Dunlop, W.I., & Ritchie, B. (2003). A broad-scale
559 approach to management of Ontario's recreational fisheries. North American Journal of
560 Fisheries Management, 23, 1312–1328.

561 Lester, N.P., Dextrase, A.J., Kushneriuk, R.S., Rawson, M.R., & Ryan, P.A. (2004). Light and
562 temperature: key factors affecting walleye abundance and production. Transactions of the
563 American Fisheries Society, 133, 588–605.

564 Leroy, B., Dias, M.S., Giraud, E., Hugueny, B., Jézéquel, C., Leprieur, F., Oberdorff, T., &
565 Tedesco, P.A. (2019). Global biogeographical regions of freshwater fish species. Journal of
566 Biogeography, 46, 2407–2419.

567 Loewen, C.J.G., Strecker, A.L., Gilbert, B., & Jackson, D.A. (2020). Climate warming
568 moderates the impacts of introduced sportfish on multiple dimensions of prey biodiversity.
569 Global Change Biology, 26, 4937–4951.

570 Lynch, A.J., Myers, B.J.E., Chu, C., Eby, L.A., Falke, J.A., Kovach, R.P., … Whitney, J.E.
571 (2016). Climate change effects on North American inland fish populations and assemblages.
572 Fisheries, 41, 346–361.

573 MacArthur, R.H., & Wilson, E.O. (1963). An equilibrium theory of insular zoogeography.
574 Evolution, 17, 373–387.

575 Magnuson, J.J, Crowder, L.B., & Medvick, P.A. (1979). Temperature as an ecological resource.
576 American Zoology, 19, 331–343.

577 Mandrak, N.E. (1995). Biogeographic patterns of fish species richness in Ontario lakes in
578 relation to historical and environmental factors. Canadian Journal of Fisheries and Aquatic
579 Sciences, 52, 1462–1474.

580 Mandrak, N.E., & Crossman, E.J. (1992). Postglacial dispersal of freshwater fishes into Ontario.
581 Canadian Journal of Zoology, 70, 2247–2259.

582 Mantyka-Pringle, C.S., Martin, T.G., Moffatt, D.B., Linke, S., & Rhodes, J.R. (2014).
583 Understanding and predicting the combined effects of climate change and land-use change on
584 freshwater macroinvertebrates and fish. Journal of Applied Ecology, 51, 572–581.

585 McGarvey, D.J., & Veech, J.A. (2018). Modular structure in fish co-occurrence networks: a
586 comparison across spatial scales and grouping methodologies. PLoS One, 13, e0208720.

587 Melles, S.J., Chu, C., Alofs, K.M., & Jackson, D.A. (2015). Potential spread of Great Lakes
588 fishes given climate change and proposed dams: an approach using circuit theory to evaluate
589 invasion risk. Landscape Ecology, 30, 919–935.

590 Mitchell, T.D., & Jones, P.D. (2005). An improved method of constructing a database of
591 monthly climate observations and associated high-resolution grids. International Journal of
592 Climatology, 25, 693–712.

593 Montalvo-Mancheno, C.S., Ondei, S., Brook, B.W., & Buettel, J.C. (2020). Bioregionalization
594 approaches for conservation: methods, biases, and their implications for Australian biodiversity.
595 Biodiversity and Conservation, 29, 1–17.

596 Neff, M.R., & Jackson, D.A. (2012). Geology as a structuring mechanism of stream fish
597 communities. Transactions of the American Fisheries Society, 141, 962–974.

598 Notaro, M., Bennington, V., & Vavrus, S. (2015). Dynamically downscaled projections of lake-
599 effect snow in the Great Lakes Basin. Journal of Climate, 28, 1661–1684.

600 Oikonomou, A., Leprieur, F., & Leonardos, I.D. (2014). Biogeography of freshwater fishes of
601 the Balkan Peninsula. Hydrobiologia, 738, 205–220.

602 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H.
603 (2020). vegan: community ecology package. R package version 2.5-7.

604 Olden, J.D., Kennard, M.J., & Pusey, B.J. (2008). Species invasions and the changing
605 biogeography of Australian freshwater fishes. Global Ecology and Biogeography, 17, 25–37.

606 Olden, J.D., Kennard, M.J., Leprieur, F., Tedesco, P.A., Winemiller, K.O., & García-Berthou, E.
607 (2010). Conservation biogeography of freshwater fishes: recent progress and future challenges.
608 Diversity and Distributions, 16, 496–513.

609 Ortega, J.C.G., Figueiredo, B.R.S., da Graça, W.J., Agostinho, A.A., & Bini, L.M. (2020).
610 Negative effect of turbidity on prey capture for both visual and non-visual aquatic predators.
611 Journal of Animal Ecology, 89, 2427–2439.

612 Petrarca, F., Russolillo, G., & Trinchera, L. (2017). Integrating non-metric data in partial least
613 squares path models: methods and application. In H. Latan & R. Noonan (Eds.), Partial Least
614 Squares Path Modeling (pp. 259–279). Cham, Switzerland: Springer International Publishing.

615 Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., … Cooke, S.J..
616 (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity.
617 Biological Reviews, 94, 849–873.

618 Rosvall, M., & Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal
619 community structure. Proceedings of the National Academy of Sciences of the United States of
620 America, 105, 1118–1123.

621 Sanchez, G. (2013). PLS Path Modeling with R. Berkeley, CA: Trowchez Editions.

622 Sanchez, G., Trinchera, L., & Russolillo, G. (2017). plspm: tools for partial least squares path
623 modeling (PLS-PM). R package version 0.4.9.

624 Sandstrom, S., Rawson, M., & Lester, N. P. (2013). *Manual of instructions for broad-scale fish*
625 community monitoring using North American (NA1) and Ontario small mesh (ON2) gillnets.
626 Peterborough, ON: Ontario Ministry of Natural Resources.

627 Ministry of Natural Resources.Shuter, B.J., & Post, J.R. (1990). Climate, population viability,
628 and the zoogeography of temperate fishes. Transactions of the American Fisheries Society, 119,
629 314–336.

630 Smith, C.L., & Powell, C.R. (1971). The summer fish communities of Brier Creek, Marshall
631 County, Oklahoma. No. 2458. New York, NY: American Museum of Natural History.

632 Strona, G., Nappo, D., Boccacci, F., Fattorini, S., & San-Miguel-Ayanz, J. (2014). A fast and

633 unbiased procedure to randomize ecological binary matrices with fixed row and column totals.

634 Nature Communications, 5, 4114.

635 Thébault, E. (2013). Identifying compartments in presence-absence matrices and bipartite

636 networks: insights into modularity measures. Journal of Biogeography, 40, 759–768.

637 Tilzer, M.M. (1988). Secchi disk – chlorophyll relationships in a lake with highly variable

638 phytoplankton biomass. Hydrobiologia, 162, 163–171.

639 Tjur, T. (2009). Coefficients of determination in logistic regression models—a new proposal: the

640 coefficient of discrimination. The American Statistician, 63, 366–372.

641 Tonn, W.M., & Magnuson, J.J. (1982). Patterns in the species composition and richness of fish

642 assemblages in northern Wisconsin lakes. Ecology, 63, 1149–1166.

643 Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially

644 customizable climate data for historical and future periods for North America. PLoS One, 11,

645 e0156720.

646 Wehrly, K.E., Breck, J.E., Wang, L. & Szabo-Kraft, L. (2012). A landscape-based classification

647 of fish assemblages in sampled and unsampled lakes. Transactions of the American Fisheries

648 Society, 141, 414–425.

649 Wehrly, K.E., Carter, G.S., & Breck, J.E. (2021). Standardized sampling methods for the inland

650 lakes status and trends program. Fisheries Special Report. Lansing, MI: Michigan Department

651 of Natural Resources.

652 **DATA AVAILABILITY STATEMENT**

653 Processed data (including site and species modules and derived spatial and environmental

654 indicator variables) and code supporting our results are archived in a dedicated GitHub

655 repository (https://github.com/loewenecology/Fish-modularity-and-nmplssem). Raw fish data

656 (with species at risk removed for their protection) are available from the Dryad Digital

657 Repository (https://doi.org/10.5061/dryad.44j0zpcfc). As contractual requirements prevent us

658 from releasing detailed sampling records for Ontario directly, these data are provided with

659 approximate location coordinates. The full set of fish data (including species at risk) are readily

660   available for non-commercial purposes from the Ministry of Northern Development, Mines,

661   Natural Resources, and Forestry in Ontario (https://geohub.lio.gov.on.ca/datasets/aquatic-

662   resource-area-polygon-segment-) and Departments of Natural Resources in Michigan

663   (https://www.michigan.gov/dnr/0,4570,7-350-79137_79765_84005_84013---,00.html) and

664   Minnesota (https://www.dnr.state.mn.us/waters/surfacewater_section/lake_ibi/index.html)

665   through their standard data-sharing agreements.

666   **FIGURE LEGENDS**

667   **Figure 1.** Maps presenting **(a)** lake sampling locations across Ontario, Minnesota, and

668   Michigan; **(b)** portions of the major Southwestern Hudson Bay (green), Mississippi River

669   (yellow), and St. Lawrence River (orange) drainage areas and their secondary drainages within

670   our study region; **(c)** tertiary drainage areas (sub-sub-drainage or United States Geological

671   Survey hydrologic unit code 8) in the Lake Huron watershed showing regional habitat density;

672   and **(d)** a single tertiary drainage area showing river network dispersal distance between a lake

673   sampling location (green point) and its nearest secondary drainage feature (Lake Huron; red

674   point).

675   **Figure 2.** Maps presenting network modularity results for site nodes summarized by tertiary

676   drainages to identify bioregions. Bioregions were assigned to the dominant primary module **(a)**

677   and submodule **(c-f)** across sites in each drainage. Participation coefficients measure diversity of

678   connections between primary modules and were assigned to the mean across sites, with greater

679   values indicating bioregional transition zones **(b)**.

680   **Figure 3.** Hierarchical edge bundling plot showing network modularity results for fish nodes to

681   identify recurrent assemblages. Connections (edges) denote fish co-occurrences, with darker

682   lines indicating more observations. Around the outside of the plot, node sizes indicate relative

683   frequencies of fish occurrence, node colours indicate fish submodule assignment, and label

684   colours indicate fish primary module assignment. Colours correspond to bioregions in Figure 2.

685   **Figure 4.** Non-metric partial least-squares structural equation (NM-PLS-SEM) meta-model.

686   Major biogeographic constructs (connectivity, climate, and habitat) were formed as linear

687   combinations of observed spatial and environmental factors to explain variation in primary site

688   module assignment. Biogeographic constructs and site modules were treated as composite and

689   single-indicator latent variables (circles), respectively, for the inner structural model. Spatial and

690   environmental factors (and nominal indicators of fish modules) were treated as formative

691   indicators (rectangles) of their respective latent variables.
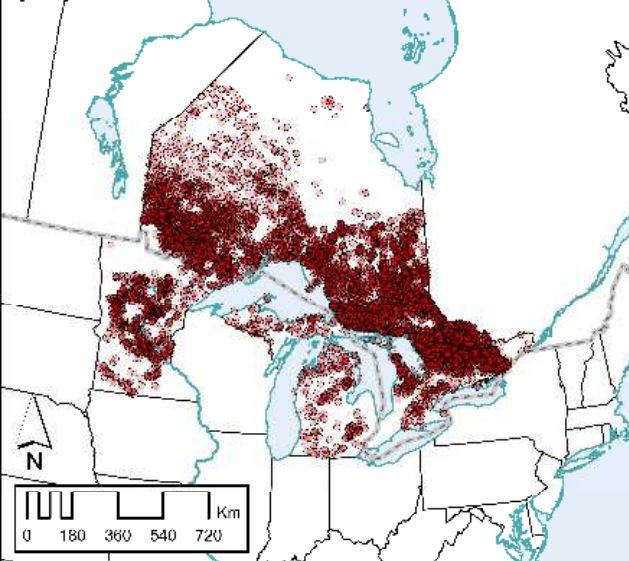
692   **Figure 5.** Results of NM-PLS-SEM and multiple logistic regression of primary site modules on

693   biogeographic constructs. Outer measurement model results **(a)** are presented as weights (darker

694   bars; representing multiple ordinary least-squares regression coefficients) and loadings (lighter

695   bars; representing Pearson correlation coefficients) for standardized indicators. Inner structural

696   model results **(b)** are presented as path coefficients from multiple logistic regression for latent

697   variables. Error bars show the lower 2.5^{th} and upper 97.5^{th} percentiles of bootstrap and profile

698   likelihood confidence intervals for outer and inner model coefficients, respectively. Detailed

699   results are available in Supplementary Tables S1.5-S1.8. Colours and model structure correspond

700   to the meta-model presented in Figure 4.

701   **Figure 6.** Approximately area-proportional Euler plots showing the total, shared, and unique

702   explanatory power of biogeographic constructs (connectivity, climate, and habitat) in logistic

703   regression. Fractions are expressed as Tjur's R^2 (coefficients of discrimination) from models of

704   primary site modules (binary responses) predicted by latent variable scores (calculated by NM-

705   PLS-SEM). Colours correspond to the meta-model presented in Figure 4.

706   **SUPPORTING INFORMATION**

707   Additional supporting information may be found online in the Supporting Information section at
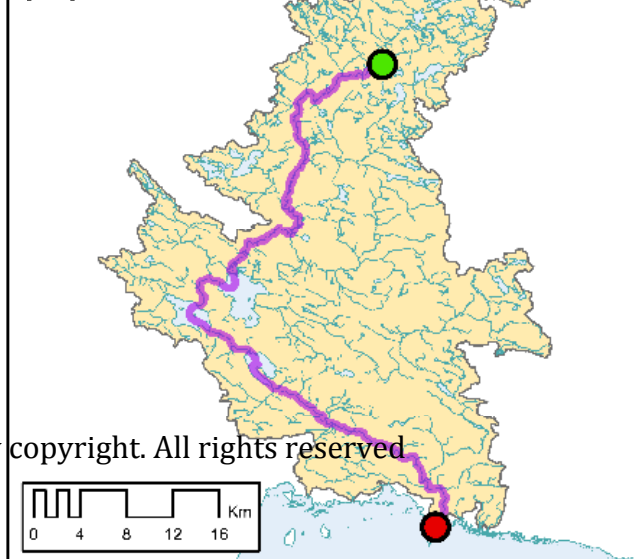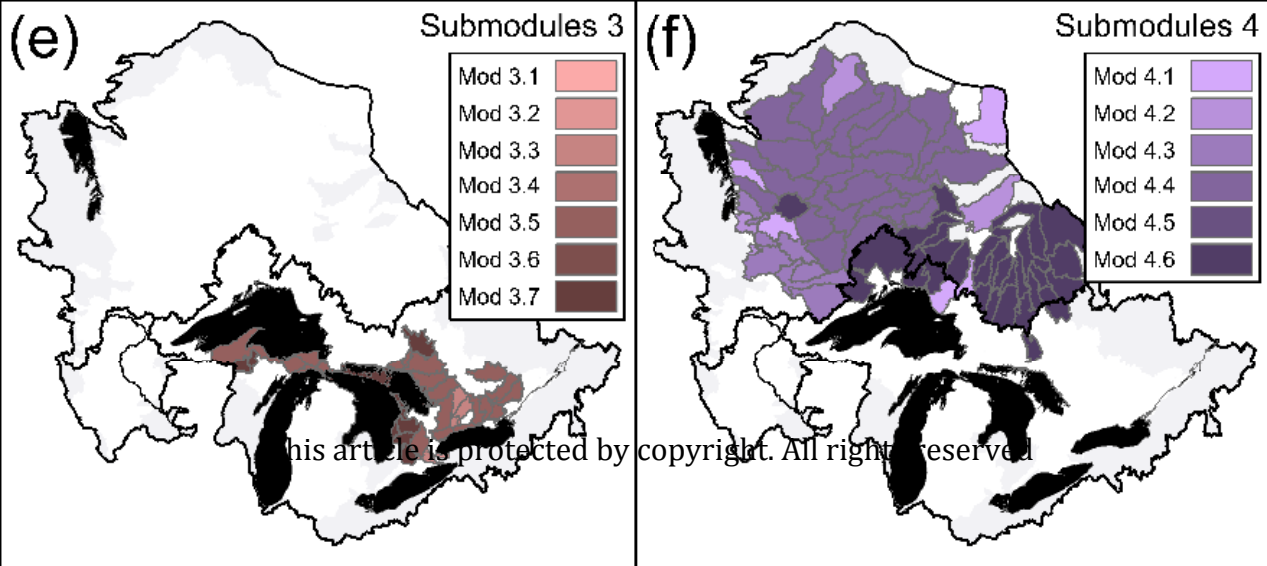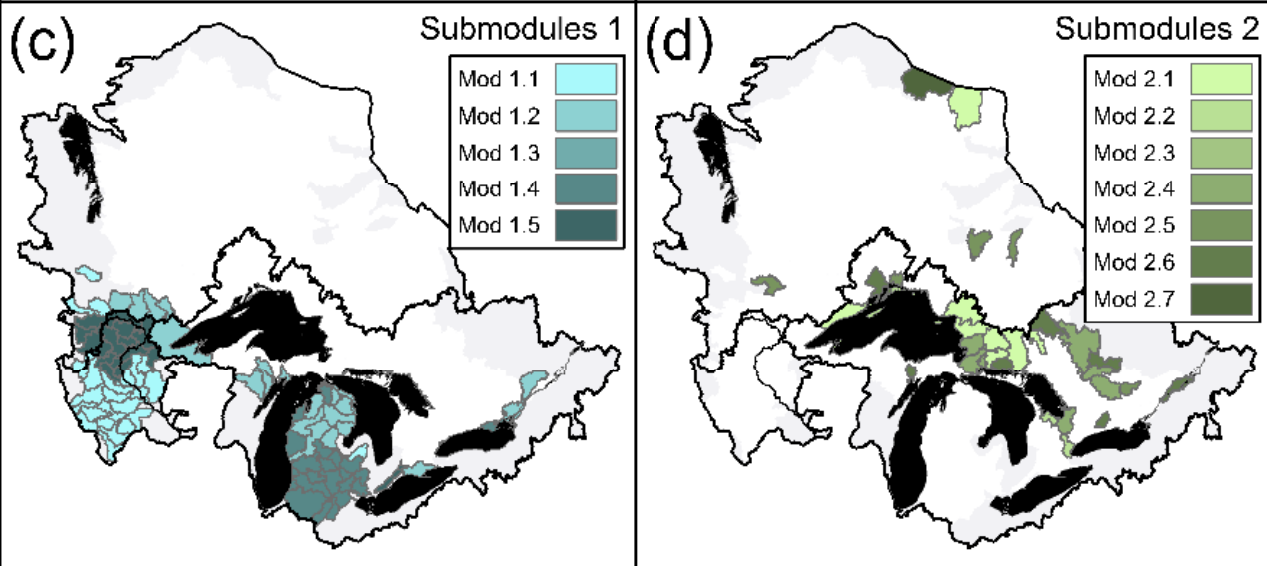
708   the end of the article.

(a) Primary modules

Mod 1
Mod 2
Mod 3
Mod 4

N

0   170   340   510   680   Km

(b) Participation coefficient

≤0.2197
≤0.3757
≤0.5100
≤0.6001
≤0.7111

(c) Submodules 1

Mod 1.1
Mod 1.2
Mod 1.3
Mod 1.4
Mod 1.5

(d) Submodules 2

Mod 2.1
Mod 2.2
Mod 2.3
Mod 2.4
Mod 2.5
Mod 2.6
Mod 2.7

(e) Submodules 3

Mod 3.1
Mod 3.2
Mod 3.3
Mod 3.4
Mod 3.5
Mod 3.6
Mod 3.7

(f) Submodules 4

Mod 4.1
Mod 4.2
Mod 4.3
Mod 4.4
Mod 4.5
Mod 4.6

geb_13424_f3.pdf

geb_13424_f4.pdf

geb_13424_f6.pdf

Mod 1

0.0087
0.0019
0.1342
0.0052
0.2335
0.3518

Mod 2

0.1431
0.0034
0.0227
0.0803
0.0306
0.0424

Mod 3

0.0008
0.0087
0.0057
0.0002
0.2571
0.0687

Mod 4

0.0733
0.1167
0.0416
0.0180
0.0503
0.2690

Connectivity   Climate   Habitat