Supporting Information Loewen et al. Bioregions are predominantly climatic for fishes of northern lakes

Appendices

Appendix S1 Expanded details for bipartite network modularity analysis and estimation of connectivity by contemporary dispersal routes

Supporting Tables

Table S1.1 List of sampled fishes indicating their occurrence in our database, their native/non-native status in our study region, whether they were considered species at risk, and whether they were used in analysis

Table S1.2 Summary of network modularity results across ten random trails

Table S1.3 Summary statistics of environmental and spatial indicators

Table S1.4 Pearson correlation coefficients between raw environmental and spatial factors

Table S1.5 Detailed results of non-metric partial least-squares structural equation models for primary site modules

Table S1.6 Diagnostics of non-metric partial least-squares structural equation models

Table S1.7 Detailed results of multiple logistic regression analysis of each primary site module on biogeographic constructs (with scores calculated for each module by non-metric partial least-squares structural equation models)

Table S1.8 Detailed results of variation partitioning analysis based on explanatory power of biogeographic constructs (with scores calculated by non-metric partial least-squares structural equation modelling) in multiple logistic regression for each primary site module

Table S1.9 Summary of changes to site modules after removal of non-native species

Supporting Figures

Figure S1.1 Maps (presenting spatial data from Dyke et al. (2003) Deglaciation of North America. Geological Survey of Canada Open File, 1574. Thirty-two digital maps at 1:7,000,000 scale with accompanying digital chronological database and one poster (two sheets) with full map series) showing **(a)** time since glaciation (before present) and **(b)** glacial lake and marine extents with most recent extents on top (chronological) and arrows representing major historical outflows (modified from Kehew et al. (2009) Proglacial megaflooding along the margins of the Laurentide Ice Sheet, *in* Burr, D., Carling, P.A., and Baker, V.R., eds., *Megaflooding on Earth and Mars*. Cambridge University Press, Cambridge, UK, p. 104–127; and Teller (2003) Controls, history, outburst, and impact of large late-Quaternary proglacial lakes in North America. *Developments in Quaternary Science*, 1:45–61)

Figure S1.2 Maps (presenting spatial data derived from ClimateNA v6.40; Wang, Hamann, Spittlehouse, & Carroll (2016) Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE*, 11: e0156720) showing 30-year normals (1961–1990) at approximately 250 m resolution for **(a)** mean warmest month temperature, **(b)** mean coldest month temperature, **(c)** May to September precipitation, and **(d)** precipitation as snow

Figure S1.3 Map (presenting spatial data from Hartmann & Moosdoft (2012) The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, 13:12) showing basic lithological classes used to derive proportion of shield lithology (metamorphic + acid plutonic + basic plutonic + intermediate plutonic + acid volcanic + basic volcanic + intermediate volcanic rocks)

Figure S1.4 Map showing network primary modularity results for individual site nodes

Figure S1.5 Maps showing network sub-modularity results for individual site nodes

Figure S1.6 Maps showing network primary modularity results for site nodes summarized by tertiary drainages to identify bioregions for our main analysis **(a)**, after removal of non-native species from the species pool **(b)**, and after removal of sites with non-native species from the site pool **(c)**

Figure S1.7 Results of NM-PLS-SEM and multiple logistic regression of primary site modules on biogeographic constructs after removal of non-native species from the species pool (see Table S1.1 for list of non-native species). Outer measurement model results **(a)** are presented as weights (darker bars; representing multiple ordinary least-squares regression coefficients) and loadings (lighter bars; representing Pearson correlation coefficients) for standardized indicators. Inner structural model results **(b)** are presented as path coefficients from multiple logistic regression for latent variables. Error bars show the lower 2.5th and upper 97.5th percentiles of bootstrap and profile likelihood confidence intervals for outer and inner model coefficients, respectively. See Figure 5 in main text for results of full analysis

Figure S1.8 Approximately area-proportional Euler plots showing the total, shared, and unique explanatory power of biogeographic constructs (connectivity, climate, and habitat) in logistic regression after removal of non-native species from the species pool (see Table S1.1 for list of non-native species). Fractions are expressed as Tjur's R^2 (coefficients of discrimination) from models of primary site modules (binary responses) predicted by latent variable scores (calculated by NM-PLS-SEM). See Figure 6 in main text for results of full analysis

Figure S1.9 Results of NM-PLS-SEM and multiple logistic regression of primary site modules on biogeographic constructs after removal of sites with non-native species from the site pool (see Table S1.1 for list of non-native species). Outer measurement model results **(a)** are presented as weights (darker bars; representing multiple ordinary least-squares regression coefficients) and loadings (lighter bars; representing Pearson correlation coefficients) for standardized indicators. Inner structural model results **(b)** are presented as path coefficients from multiple logistic regression for latent variables. Error bars show the lower 2.5th and upper 97.5th percentiles of bootstrap and profile likelihood confidence intervals for outer and inner model coefficients, respectively. See Figure 5 in main text for results of full analysis

Figure S1.10 Approximately area-proportional Euler plots showing the total, shared, and unique explanatory power of biogeographic constructs (connectivity, climate, and habitat) in logistic regression after removal of sites with non-native species from the site pool (see Table S1.1 for list of non-native species). Fractions are expressed as Tjur's R^2 (coefficients of discrimination) from models of primary site modules (binary responses) predicted by latent variable scores (calculated by NM-PLS-SEM). See Figure 6 in main text for results of full analysis

Appendix S1 Expanded details for bipartite network modularity analysis and estimation of connectivity by contemporary dispersal routes

Expanded details for bipartite network modularity analysis

Modularity analysis of site-species networks has proven an effective tool for uncovering concise bioregions and their transitional zones across spatial scales (Bloomfield et al., 2018; McGarvey & Veech, 2018). The approach provides unsupervised learning of community structure based solely on network topology, where divisions are made to maximize the number of edges within, rather than between, resultant modules (Newman & Girvan, 2004). To protect against trivial solutions, modularity (Q) is quantified relative to a null expectation where edges are placed at random. Thus, modules are detected independent of site constraints, providing an opportunity to test spatial and environmental factors determining species assemblages. Module detection is complemented by calculation of the participation coefficient, which identifies transitional zones between groups (e.g. Bloomfield et al., 2018). Here, the participation coefficient measures among-module connectivity P_i of node i as:

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{k_{is}}{k_i}\right)^2$$

where N_M is the number of modules in the network, k_{is} is the number of links of i to nodes in its own module s, and k_i is the total degree of node i (Guimerà & Amaral, 2005). While modularity was initially developed for the unipartite case, we used the bipartite formulation of Barber (2007), which modifies that of Newman & Girvan (2004) to allow links only between nodes of opposing types. Barber's definition classifies both sets of nodes simultaneously and produces one-to-one node correspondence (i.e. it will produce the same number of modules for sites and species), which aids interpretation of bioregionalization by associating assemblages directly to their physical locations.

Modularity optimization is computationally challenging (NP-hard; Miyauchi & Sukegawa, 2015), necessitating heuristics to search through solution space of larger networks. We applied the DIRTLPAb+ algorithm of Beckett (2016), which is a label propagation approach to maximizing Barber's bipartite modularity. Briefly, the label propagation algorithm (LPA) for community detection was proposed by Raghavan, Albert, & Kumara (2007), where nodes were initialized with all unique labels, updated iteratively to adopt the most common labels of their neighbours (with ties broken uniformly), and grouped together once no further improvements could be made. The routine was subsequently modified to the bipartite case (LPAb); however, it had a tendency to become trapped in local maxima. To escape such traps, Liu & Murata (2010) incorporated a multistep greedy agglomerative algorithm (LPAb+). Their approach applied asynchronous label propagation (updating node labels of each type in turn) but added a step where resulting groups could be merged if doing so improved the solution, iterating between propagation and aggregation phases. Finally, because label propagation is inherently stochastic, and thus outcomes can vary depending on initialization, Beckett (2016) proposed that LPAb+ should be repeated under different node configurations (i.e. numbers of unique starting labels) and report the greatest ensuing score (DIRTLPAb+).

Expanded details for estimation of connectivity by contemporary dispersal routes

To estimate river network dispersal distances, network origins were obtained by snapping sampling locations to flowlines within each lake polygon. Lakes lacking connection to the broader surface drainage network were visually assessed with recent satellite imagery, and flowlines were manually adjusted where necessary to capture any missing links (i.e. visible drainages or culverts). Regional flowlines were also scanned for discernible errors, such as abrupt breaks along major rivers obscured by bridges or cloud cover. Potential network destination points were obtained by intersecting flowlines with polygons for the secondary drainage features and removing those lacking upstream sampling locations. Finally, we performed origin-destination cost matrix analysis using Network Analyst (an Esri ArcGIS toolbox extension) to calculate shortest paths along the spatial network. Isolated lakes (those lacking connection to the regional flowlines) were assigned distances equal to the longest paths measured among neighbours in their respective tertiary drainages. All waterbody features, flowlines, and drainage boundaries were obtained from the National Hydro Network in Canada (Government of Canada; available at: https://open.canada.ca/data/en/dataset/a4b190fe-e090-4e6d-881e-b87956c07977) and National Hydrologic Dataset in USA (U.S. Geological Survey; available at: https://viewer.nationalmap.gov/basic/#/). All geospatial analyses were estimated from remote sensing data projected to Ontario MNR Lambert (NAD 83 CSRS).

References

Barber, M.J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76, 066102.

Beckett, S.J. (2016). Improved community detection in weighted bipartite networks. *Royal Society Open Science*, 3, 140536.

Bloomfield, N.J., Knerr, N., & Encinas-Viso, F. (2018). A comparison of network and clustering methods to detect biogeographical regions. *Ecography*, 41, 1–10.

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

Liu, X., & Murata, T. (2010). An efficient algorithm for optimizing bipartite modularity in bipartite networks. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 14, 408–415.

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

Miyauchi, A., & Sukegawa, N. (2015). Maximizing Barber's bipartite modularity is also hard. *Optimization Letters*, 9, 897–913.

Newman, M.E.J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69, 026113.

Raghavan, U.N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76, 036106.

Table S1.1 List of sampled fishes indicating their occurrence in our database, their native/non-native status in our study region, whether they were considered species at risk, and whether they were used in analysis

Common Name	Scientific Name	Count	Status	Use	Notes
Alewife	*Alosa pseudoharengus*	45	Non-native	Yes	
American Eel	*Anguilla rostrata*	4		Yes	SAR
Banded Killifish	*Fundulus diaphanus*	568		Yes	
Bigmouth Buffalo	*Ictiobus cyprinellus*	83		Yes	
Black Bullhead	*Ameiurus melas*	622		Yes	
Black Crappie	*Pomoxis nigromaculatus*	1171		Yes	
Blackchin Shiner	*Notropis heterodon*	574		Yes	
Blacknose Shiner	*Notropis heterolepis*	2345		Yes	
Bluegill	*Lepomis macrochirus*	1409		Yes	
Bluntnose Minnow	*Pimephales notatus*	1876		Yes	
Bowfin	*Amia calva*	580		Yes	
Brassy Minnow	*Hybognathus hankinsoni*	84		Yes	
Brook Silverside	*Labidesthes sicculus*	184		Yes	
Brook Stickleback	*Culaea inconstans*	1337		Yes	
Brook Trout	*Salvelinus fontinalis*	1892		Yes	
Brown Bullhead	*Ameiurus nebulosus*	2013		Yes	
Brown Trout	*Salmo trutta*	46	Non-native	Yes	
Burbot	*Lota lota*	1467		Yes	
Central Mudminnow	*Umbra limi*	733		Yes	
Channel Catfish	*Ictalurus punctatus*	146		Yes	
Cisco	*Coregonus*	2408		Yes	Cisco spp. lumped together
Common Carp	*Cyprinus carpio*	490	Non-native	Yes	
Common Shiner	*Luxilus cornutus*	1278		Yes	
Creek Chub	*Semotilus atromaculatus*	1006		Yes	
Emerald Shiner	*Notropis atherinoides*	293		Yes	
Fallfish	*Semotilus corporalis*	97		Yes	
Fathead Minnow	*Pimephales promelas*	1712		Yes	
Finescale Dace	*Chrosomus neogaeus*	893		Yes	
Freshwater Drum	*Aplodinotus grunniens*	57		Yes	
Gizzard Shad	*Dorosoma cepedianum*	19		Yes	
Golden Redhorse	*Moxostoma erythrurum*	36		Yes	
Golden Shiner	*Notemigonus crysoleucas*	1897		Yes	
Grass Pickerel	*Esox americanus*	56		Yes	SAR
Greater Redhorse	*Moxostoma valenciennesi*	35		Yes	
Green Sunfish	*Lepomis cyanellus*	637		Yes	
Iowa Darter	*Etheostoma exile*	2530		Yes	
Johnny Darter	*Etheostoma nigrum*	1951		Yes	
Lake Chub	*Couesius plumbeus*	577		Yes	SAR
Lake Chubsucker	*Erimyzon sucetta*	43		Yes	SAR
Lake Sturgeon	*Acipenser fulvescens*	59		Yes	SAR
Lake Trout	*Salvelinus namaycush*	1807		Yes	
Lake Whitefish	*Coregonus clupeaformis*	2094		Yes	
Largemouth Bass	*Micropterus salmoides*	1869		Yes	
Logperch	*Percina caprodes*	787		Yes	
Longnose Dace	*Rhinichthys cataractae*	321		Yes	
Longnose Gar	*Lepisosteus osseus*	87		Yes	
Longnose Sucker	*Catostomus catostomus*	512		Yes	
Mimic Shiner	*Notropis volucellus*	938		Yes	
Mooneye	*Hiodon tergisus*	23		Yes	SAR
Mottled Sculpin	*Cottus bairdii*	611		Yes	

Common Name	Scientific Name	Count	Status	Use	Notes
Muskellunge	*Esox masquinongy*	259		Yes	
Ninespine Stickleback	*Pungitius pungitius*	454		Yes	
Northern Pearl Dace	*Margariscus nachtriebi*	1042		Yes	Pearl Dace lumped in
Northern Pike	*Esox lucius*	5596		Yes	
Northern Redbelly Dace	*Chrosomus eos*	1788		Yes	
Northern Sunfish	*Lepomis peltastes*	49		Yes	SAR; Longear Sunfish lumped in
Orangespotted Sunfish	*Lepomis humilis*	44	Non-native	Yes	
Pugnose Shiner	*Notropis anogenus*	107		Yes	SAR
Pumpkinseed	*Lepomis gibbosus*	3100		Yes	
Quillback	*Carpiodes cyprinus*	24		Yes	
Rainbow Smelt	*Osmerus mordax*	146	Non-native	Yes	
Rainbow Trout	*Oncorhynchus mykiss*	331	Non-native	Yes	
Redear Sunfish	*Lepomis microlophus*	39	Non-native	Yes	
Rock Bass	*Ambloplites rupestris*	2087		Yes	
Round Goby	*Neogobius melanostomus*	16	Non-native	Yes	
Round Whitefish	*Prosopium cylindraceum*	59		Yes	
Sand Shiner	*Notropis stramineus*	101		Yes	
Sauger	*Sander canadensis*	137		Yes	SAR
Shorthead Redhorse	*Moxostoma macrolepidotum*	359		Yes	
Silver Redhorse	*Moxostoma anisurum*	92		Yes	
Slimy Sculpin	*Cottus cognatus*	245		Yes	
Smallmouth Bass	*Micropterus dolomieu*	2068		Yes	
Splake	*Salvelinus namaycush x Salvelinus fontinalis*	53	Non-native	Yes	
Spoonhead Sculpin	*Cottus ricei*	15		Yes	SAR
Spotfin Shiner	*Cyprinella spiloptera*	80		Yes	
Spottail Shiner	*Notropis hudsonius*	2119		Yes	
Spotted Gar	*Lepisosteus oculatus*	22		Yes	SAR
Starhead Topminnow	*Fundulus dispar*	2		Yes	SAR
Tadpole Madtom	*Noturus gyrinus*	476		Yes	
Tiger Muskellunge	*Esox masquinongy* × *Esox lucius*	11	Non-native	Yes	
Trout-perch	*Percopsis omiscomaycus*	1206		Yes	
Walleye	*Sander vitreus*	2879		Yes	
Warmouth	*Lepomis gulosus*	99		Yes	SAR
Weed Shiner	*Notropis texanus*	15		Yes	SAR
White Bass	*Morone chrysops*	27		Yes	
White Crappie	*Pomoxis annularis*	60		Yes	
White Perch	*Morone americana*	10	Non-native	Yes	
White Sucker	*Catostomus commersonii*	7548		Yes	
Yellow Bullhead	*Ameiurus natalis*	872		Yes	
Yellow Perch	*Perca flavescens*	7125		Yes	
Atlantic Salmon	*Salmo salar*	1		No	Single occurrence
Bigmouth Shiner	*Notropis dorsalis*	18		No	SAR; transient/poorly sampled
Black Buffalo	*Ictiobus niger*	58		No	SAR; transient/poorly sampled
Black Redhorse	*Moxostoma duquesnei*	2		No	SAR; transient/poorly sampled
Blacknose Dace	*Rhinichthys atratulus*	113		No	Transient/poorly sampled
Blackside Darter	*Percina maculata*	19		No	Transient/poorly sampled
Blackstripe Topminnow	*Fundulus notatus*	14		No	SAR; transient/poorly sampled
Bloater	*Coregonus hoyi*	1		No	Single occurrence
Bluntnose Darter	*Etheostoma chlorosomum*	1		No	SAR; single occurrence
Bridle Shiner	*Notropis bifrenatus*	1		No	SAR; single occurrence
Brindled Madtom	*Noturus miurus*	1		No	SAR; single occurrence
Central Stoneroller	*Campostoma anomalum*	2		No	Transient/poorly sampled
Chestnut Lamprey	*Ichthyomyzon castaneus*	6		No	Transient/poorly sampled

Common Name	Scientific Name	Count	Status	Use	Notes
Chinook Salmon	*Oncorhynchus tshawytscha*	4	Non-native	No	Transient/poorly sampled
Coho Salmon	*Oncorhynchus kisutch*	3	Non-native	No	Transient/poorly sampled
Flathead Catfish	*Pylodictis olivaris*	3		No	Transient/poorly sampled
Goldfish	*Carassius auratus*	12	Non-native	No	Transient/poorly sampled
Greenside Darter	*Etheostoma blennioides*	7		No	Transient/poorly sampled
Hornyhead Chub	*Nocomis biguttatus*	48		No	Transient/poorly sampled
Least Darter	*Etheostoma microperca*	152		No	SAR; transient/poorly sampled
Northern Brook Lamprey	*Ichthyomyzon fossor*	1		No	SAR; single occurrence
Northern Hog Sucker	*Hypentelium nigricans*	11		No	Transient/poorly sampled
Pirate Perch	*Aphredoderus sayanus*	1		No	SAR; single occurrence
Plains Topminnow	*Fundulus sciadicus*	1		No	SAR; single occurrence
Pugnose Minnow	*Opsopoeodus emilae*	4		No	SAR; transient/poorly sampled
Rainbow Darter	*Etheostoma caeruleum*	19		No	Transient/poorly sampled
River Carpsucker	*Carpiodes carpio*	2		No	Transient/poorly sampled
River Chub	*Nocomis micropogon*	8		No	Transient/poorly sampled
River Darter	*Percina shumardi*	2		No	SAR; transient/poorly sampled
River Redhorse	*Moxostoma carinatum*	4		No	SAR; transient/poorly sampled
Rosyface Shiner	*Notropis rubellus*	10		No	Transient/poorly sampled
Ruffe	*Gymnocephalus cernua*	3	Non-native	No	Lost after site exclusions
Sea Lamprey	*Petromyzon marinus*	6		No	Transient/poorly sampled
Shortnose Gar	*Lepisosteus platostomus*	6		No	Transient/poorly sampled
Silver Lamprey	*Ichthyomyzon unicuspis*	4		No	SAR; transient/poorly sampled
Smallmouth Buffalo	*Ictiobus bubalus*	5		No	Transient/poorly sampled
Sockeye Salmon	*Oncorhynchus nerka*	1	Non-native	No	Single occurrence
Spotted Sucker	*Minytrema melanops*	5		No	SAR; transient/poorly sampled
Stonecat	*Noturus flavus*	4		No	Transient/poorly sampled
Striped Shiner	*Luxilus chrysocephalus*	1		No	Single occurrence
Threespine Stickleback	*Gasterosteus aculeatus*	27		No	Transient/poorly sampled
Yellow Bass	*Morone mississippiensis*	1		No	SAR; single occurrence

Notes: Counts based on number of occurrences in initial database (prior to site exclusions); use in analysis was based on avoiding species transient in lake communities or otherwise poorly sampled; SAR denotes species at risk, including those threatened, endangered, or of special concern.

Table S1.2 Summary of network modularity results across ten random trails

Analysis	Number of computed modules	Lowest modularity (Q)	Greatest Modularity (Q)
Categorical	3–4 median = 4	0.3318334	0.3324972* (0.2800484)
Module 1	4–9 median = 6	0.1753108	0.1813550* (0.1269607)
Module 2	5–10 median = 7	0.290066	0.3025909* (0.2750079)
Module 3	6–7 median = 6	0.2325408	0.2392342* (0.2327642)
Module 4	4–8 median = 6	0.1778775	0.1883207* (0.1714832)

Notes: Number of computed modules from repeated analysis of unweighted network graphs using Beckett's DIRTLPAb+ algorithm to optimize Barber's bipartite formulation of Newman & Girvan's modularity index (with default settings for random initializations); * denotes $P<0.01$ from one-sided randomization tests based on 100 null networks obtained using the curveball algorithm (mean modularity of simulations in brackets).

Table S1.3 Summary statistics of environmental and spatial indicators

Variable	Units	Minimum	Median	Mean	Maximum	SD
Connectivity						
Regional habitat density	ha (inland habitat) / ha (watershed)	0.002	0.091	0.104	0.339	0.061
River network dispersal distance	m (shortest path)	0	325,953	376,413	1,244,527	285,933
Glacial lake/marine distance	m (Euclidean distance)	0	10,230	19,562	214,550	24,652
Time since glaciation	millennia	8	11	11	17	1.5
Climate						
Mean warmest month temperature (mean 1957–2017)	Celsius (Kelvin)	14.1 (287.27)	18.3 (291.48)	18.4 (291.54)	23.0 (296.15)	1.5 (1.5)
Mean coldest month temperature (mean 1957–2017)	Celsius (Kelvin)	-25.2 (247.92)	-15.5 (257.65)	-15.4 (257.74)	-5.0 (268.13)	3.6 (3.6)
May to September precipitation (mean 1957–2017)	mm	307	420	420	536	26
Precipitation as snow (mean 1957–2017)	mm	67	251	235	406	71
Habitat						
Surface area	ha	0.2	76.5	373.74	140,943.3	2081.8
Maximum depth	m	0.3	13.5	16.8	213.5	13.9
Secchi depth	m	0.15	3.20	3.53	22.00	1.85
Proportion shield lithology	proportion	0.00	1.00	0.78	1.00	0.40

Notes: SD denotes standard deviation and temperature statistics are presented in both Celsius and Kelvin (in brackets), the latter of which was used in analysis to avoid negative values.

Table S1.4 Pearson correlation coefficients between raw environmental and spatial factors

	Regional habitat density	River network dispersal distance	Glacial lake/marine distance	Time since glaciation	Mean warmest month temperature	Mean coldest month temperature	May to September precipitation	Precipitation as snow	Surface area	Maximum depth	Secchi depth	Proportion shield lithology
Regional habitat density	1.00	0.19	-0.13	0.02	0.05	-0.27	0.16	-0.16	0.08	0.22	0.11	0.27
River network dispersal distance	0.19	1.00	0.02	-0.30	-0.17	-0.54	0.10	-0.18	0.06	0.00	-0.02	0.15
Glacial lake/marine distance	-0.13	0.02	1.00	0.20	0.15	0.23	0.24	0.04	-0.04	-0.03	0.04	0.00
Time since glaciation	0.02	-0.30	0.20	1.00	0.81	0.66	0.33	-0.52	-0.06	-0.03	0.00	-0.26
Mean warmest month temperature	0.05	-0.17	0.15	0.81	1.00	0.61	0.26	-0.62	-0.03	-0.06	-0.08	-0.30
Mean coldest month temperature	-0.27	-0.54	0.23	0.66	0.61	1.00	0.02	0.02	-0.12	-0.08	0.07	-0.35
May to September precipitation	0.16	0.10	0.24	0.33	0.26	0.02	1.00	-0.16	-0.02	0.01	-0.06	0.06
Precipitation as snow	-0.16	-0.18	0.04	-0.52	-0.62	0.02	-0.16	1.00	-0.07	0.01	0.19	0.24
Surface area	0.08	0.06	-0.04	-0.06	-0.03	-0.12	-0.02	-0.07	1.00	0.19	-0.04	0.02
Maximum depth	0.22	0.00	-0.03	-0.03	-0.06	-0.08	0.01	0.01	0.19	1.00	0.48	0.08
Secchi depth	0.11	-0.02	0.04	0.00	-0.08	0.07	-0.06	0.19	-0.04	0.48	1.00	0.03
Proportion shield lithology	0.27	0.15	0.00	-0.26	-0.30	-0.35	0.06	0.24	0.02	0.08	0.03	1.00

Table S1.5 Detailed results of non-metric partial least-squares structural equation models for primary site modules

Variable	Stat.	Boot mean	Std. error	2.5th perc.	97.5th perc.
Categorical					
Indicators (weights)					
Regional habitat density	-0.308	-0.308	0.012	-0.330	-0.285
River network dispersal distance	-0.116	-0.116	0.016	-0.148	-0.085
Glacial lake/marine distance	0.093	0.093	0.010	0.073	0.114
Time since glaciation	0.879	0.879	0.008	0.862	0.895
Mean warmest month temperature	-0.007	-0.007	0.026	-0.059	0.043
Mean coldest month temperature	0.705	0.704	0.018	0.670	0.737
May to September precipitation	0.135	0.135	0.011	0.113	0.157
Precipitation as snow	-0.649	-0.648	0.032	-0.713	-0.584
Surface area	0.035	0.035	0.046	-0.056	0.121
Maximum depth	0.054	0.056	0.040	-0.027	0.134
Secchi depth	0.146	0.143	0.044	0.054	0.228
Proportion shield lithology	0.962	0.960	0.009	0.942	0.975
Indicators (loadings)					
Regional habitat density	-0.340	-0.339	0.014	-0.366	-0.312
River network dispersal distance	-0.409	-0.409	0.017	-0.443	-0.378
Glacial lake/marine distance	0.293	0.292	0.012	0.267	0.316
Time since glaciation	0.934	0.934	0.004	0.926	0.942
Mean warmest month temperature	0.892	0.892	0.005	0.882	0.900
Mean coldest month temperature	0.725	0.725	0.020	0.685	0.763
May to September precipitation	0.300	0.299	0.018	0.264	0.334
Precipitation as snow	-0.702	-0.701	0.021	-0.743	-0.658
Surface area	0.105	0.106	0.044	0.020	0.195
Maximum depth	0.258	0.258	0.027	0.207	0.312
Secchi depth	0.259	0.256	0.036	0.188	0.327
Proportion shield lithology	0.982	0.980	0.006	0.968	0.990
Latent variables (path coefficients)					
Connectivity	-0.183	-0.183	0.021	-0.222	-0.142
Climate	-0.628	-0.627	0.021	-0.669	-0.590
Habitat	0.019	0.019	0.007	0.005	0.033
Module 1					
Indicators (weights)					
Regional habitat density	-0.389	-0.388	0.015	-0.417	-0.358
River network dispersal distance	0.037	0.037	0.017	0.002	0.068
Glacial lake/marine distance	0.028	0.028	0.011	0.007	0.049
Time since glaciation	0.923	0.923	0.007	0.908	0.937
Mean warmest month temperature	-0.273	-0.272	0.020	-0.312	-0.234

Variable	Stat.	Boot mean	Std. error	2.5th perc.	97.5th perc.
Mean coldest month temperature	0.550	0.549	0.016	0.520	0.580
May to September precipitation	0.214	0.214	0.010	0.195	0.236
Precipitation as snow	-1.005	-1.005	0.013	-1.029	-0.980
Surface area	-0.294	-0.295	0.029	-0.355	-0.240
Maximum depth	0.013	0.015	0.038	-0.056	0.092
Secchi depth	0.390	0.388	0.037	0.317	0.461
Proportion shield lithology	0.850	0.849	0.016	0.818	0.878
Indicators (loadings)					
Regional habitat density	-0.393	-0.392	0.016	-0.422	-0.360
River network dispersal distance	-0.285	-0.285	0.020	-0.323	-0.249
Glacial lake/marine distance	0.250	0.249	0.013	0.224	0.273
Time since glaciation	0.922	0.922	0.006	0.910	0.933
Mean warmest month temperature	0.790	0.789	0.007	0.776	0.802
Mean coldest month temperature	0.421	0.421	0.011	0.400	0.442
May to September precipitation	0.396	0.396	0.015	0.368	0.424
Precipitation as snow	-0.895	-0.894	0.004	-0.902	-0.886
Surface area	-0.251	-0.252	0.024	-0.298	-0.206
Maximum depth	0.240	0.240	0.025	0.191	0.286
Secchi depth	0.478	0.476	0.026	0.428	0.529
Proportion shield lithology	0.867	0.866	0.015	0.838	0.893
Latent variables (path coefficients)					
Connectivity	0.082	0.083	0.013	0.057	0.108
Climate	0.697	0.696	0.012	0.672	0.721
Habitat	-0.046	-0.046	0.008	-0.061	-0.031
Module 2					
Indicators (weights)					
Regional habitat density	-0.014	-0.009	0.042	-0.087	0.081
River network dispersal distance	-0.342	-0.298	0.170	-0.413	0.376
Glacial lake/marine distance	0.919	0.818	0.413	-0.878	0.951
Time since glaciation	-0.682	-0.600	0.325	-0.736	0.730
Mean warmest month temperature	-0.619	-0.619	0.051	-0.721	-0.518
Mean coldest month temperature	0.606	0.606	0.033	0.548	0.672
May to September precipitation	0.329	0.329	0.020	0.289	0.366
Precipitation as snow	0.580	0.580	0.038	0.501	0.653
Surface area	0.950	0.949	0.012	0.925	0.973
Maximum depth	-0.233	-0.233	0.027	-0.286	-0.179
Secchi depth	-0.298	-0.298	0.025	-0.347	-0.253
Proportion shield lithology	-0.087	-0.087	0.020	-0.125	-0.049
Indicators (loadings)					
Regional habitat density	-0.120	-0.102	0.075	-0.188	0.171

Variable	Stat.	Boot mean	Std. error	2.5th perc.	97.5th perc.
River network dispersal distance	-0.105	-0.087	0.070	-0.182	0.148
Glacial lake/marine distance	0.758	0.677	0.338	-0.711	0.805
Time since glaciation	-0.390	-0.342	0.192	-0.448	0.436
Mean warmest month temperature	-0.552	-0.552	0.018	-0.585	-0.517
Mean coldest month temperature	0.210	0.209	0.021	0.166	0.250
May to September precipitation	0.041	0.041	0.023	-0.006	0.086
Precipitation as snow	0.892	0.892	0.008	0.875	0.907
Surface area	0.873	0.873	0.010	0.853	0.892
Maximum depth	-0.097	-0.096	0.022	-0.142	-0.052
Secchi depth	-0.473	-0.473	0.019	-0.508	-0.436
Proportion shield lithology	-0.082	-0.082	0.022	-0.124	-0.039
Latent variables (path coefficients)					
Connectivity	0.065	0.058	0.030	-0.058	0.083
Climate	0.221	0.222	0.010	0.203	0.241
Habitat	-0.371	-0.371	0.009	-0.388	-0.355
Module 3					
Indicators (weights)					
Regional habitat density	0.172	0.169	0.047	0.101	0.242
River network dispersal distance	-0.634	-0.618	0.139	-0.685	-0.572
Glacial lake/marine distance	0.009	0.011	0.031	-0.051	0.072
Time since glaciation	0.617	0.603	0.128	0.551	0.670
Mean warmest month temperature	0.995	0.994	0.032	0.928	1.059
Mean coldest month temperature	0.224	0.224	0.025	0.171	0.272
May to September precipitation	-0.351	-0.351	0.016	-0.381	-0.321
Precipitation as snow	0.881	0.880	0.028	0.826	0.936
Surface area	0.284	0.280	0.097	0.117	0.440
Maximum depth	0.620	0.601	0.146	0.407	0.774
Secchi depth	-0.989	-0.954	0.225	-1.101	-0.821
Proportion shield lithology	0.398	0.385	0.108	0.223	0.544
Indicators (loadings)					
Regional habitat density	0.067	0.066	0.037	-0.007	0.137
River network dispersal distance	-0.782	-0.762	0.169	-0.819	-0.736
Glacial lake/marine distance	0.097	0.097	0.036	0.023	0.161
Time since glaciation	0.798	0.780	0.164	0.751	0.835
Mean warmest month temperature	0.456	0.457	0.012	0.432	0.480
Mean coldest month temperature	0.798	0.798	0.009	0.779	0.815
May to September precipitation	-0.311	-0.312	0.016	-0.342	-0.282
Precipitation as snow	0.293	0.293	0.012	0.269	0.317
Surface area	0.548	0.536	0.122	0.412	0.669
Maximum depth	0.165	0.165	0.073	0.022	0.307

Variable	Stat.	Boot mean	Std. error	2.5th perc.	97.5th perc.
Secchi depth	-0.590	-0.567	0.153	-0.698	-0.436
Proportion shield lithology	0.399	0.387	0.105	0.227	0.547
Latent variables (path coefficients)					
Connectivity	0.059	0.058	0.016	0.042	0.077
Climate	0.538	0.537	0.011	0.516	0.557
Habitat	0.052	0.050	0.017	0.020	0.076
Module 4					
Indicators (weights)					
Regional habitat density	-0.221	-0.221	0.014	-0.246	-0.194
River network dispersal distance	-0.351	-0.351	0.013	-0.377	-0.326
Glacial lake/marine distance	0.370	0.370	0.014	0.342	0.398
Time since glaciation	0.658	0.657	0.013	0.632	0.683
Mean warmest month temperature	0.084	0.083	0.028	0.023	0.135
Mean coldest month temperature	0.941	0.941	0.017	0.910	0.978
May to September precipitation	0.129	0.129	0.012	0.106	0.151
Precipitation as snow	0.109	0.109	0.021	0.069	0.148
Surface area	0.827	0.827	0.014	0.800	0.855
Maximum depth	-0.083	-0.082	0.028	-0.138	-0.030
Secchi depth	-0.298	-0.300	0.024	-0.347	-0.253
Proportion shield lithology	0.458	0.456	0.018	0.418	0.489
Indicators (loadings)					
Regional habitat density	-0.306	-0.305	0.016	-0.334	-0.273
River network dispersal distance	-0.554	-0.554	0.012	-0.576	-0.530
Glacial lake/marine distance	0.506	0.506	0.016	0.475	0.536
Time since glaciation	0.838	0.837	0.009	0.820	0.854
Mean warmest month temperature	0.621	0.621	0.010	0.600	0.640
Mean coldest month temperature	0.990	0.990	0.002	0.986	0.993
May to September precipitation	0.137	0.136	0.014	0.109	0.163
Precipitation as snow	-0.012	-0.012	0.014	-0.038	0.016
Surface area	0.835	0.834	0.011	0.813	0.855
Maximum depth	0.071	0.071	0.022	0.028	0.113
Secchi depth	-0.334	-0.334	0.019	-0.370	-0.297
Proportion shield lithology	0.472	0.470	0.017	0.436	0.501
Latent variables (path coefficients)					
Connectivity	-0.192	-0.193	0.010	-0.211	-0.174
Climate	-0.405	-0.404	0.011	-0.426	-0.382
Habitat	0.232	0.233	0.009	0.217	0.251

Notes: Site modules were treated as either unordered multinomial (multilevel categorical) or binary dependent variables (modules 1–4); indicator weights and latent variable path coefficients are multiple ordinary least-squares regression coefficients; indicator loadings are Pearson correlation coefficients; boot means are from bootstrap validation with 1,000 resamples (without replacement); Std. error is the standard error of bootstrap samples; and the lower 2.5th and upper 97.5th percentiles are bounds of bootstrap confidence intervals.

Table S1.6 Diagnostics of non-metric partial least-squares structural equation models

Variable	Categorical	Mod 1	Mod 2	Mod 3	Mod 4
Goodness-of-fit	0.477	0.451	0.259	0.282	0.390
R^2	0.642	0.610	0.278	0.305	0.483
Connectivity Block Communality	0.310	0.287	0.188	0.315	0.339
Climate Block Communality	0.476	0.440	0.286	0.257	0.346
Habitat Block Communality	0.277	0.275	0.251	0.209	0.259

Table S1.7 Detailed results of multiple logistic regression analysis of each primary site module on biogeographic constructs (with scores calculated for each module by non-metric partial least-squares structural equation models)

Model	**Coefficient**	**Stat**	**Std. error**	***z* score**	**2.5^{th} perc.**	**97.5^{th} perc.**
Module 1						
ABC	Intercept	-3.830	0.084	-45.517	-3.999	-3.669
	Connectivity	0.517	0.069	7.508	0.382	0.652
	Climate	2.445	0.081	30.105	2.289	2.608
	Habitat	0.171	0.059	2.879	0.056	0.288
	AUC	0.965				
AB	Intercept	-3.814	0.084	-45.394	-3.983	-3.653
	Connectivity	0.449	0.064	7.001	0.323	0.575
	Climate	2.408	0.080	30.071	2.255	2.569
	AUC	0.965				
AC	Intercept	-3.082	0.058	-52.975	-3.198	-2.969
	Connectivity	1.918	0.053	36.278	1.816	2.024
	Habitat	-0.321	0.038	-8.345	-0.396	-0.245
	AUC	0.951				
BC	Intercept	-3.789	0.083	-45.466	-3.956	-3.629
	Climate	2.782	0.072	38.391	2.643	2.927
	Habitat	0.032	0.055	0.594	-0.074	0.140
	AUC	0.961				
A	Intercept	-3.126	0.059	-53.170	-3.243	-3.012
	Connectivity	2.098	0.051	41.370	2.000	2.199
	AUC	0.945				
B	Intercept	-3.788	0.083	-45.430	-3.955	-3.628
	Climate	2.766	0.067	41.010	2.637	2.902
	AUC	0.961				
C	Intercept	-2.274	0.038	-59.690	-2.350	-2.201
	Habitat	-0.939	0.029	-31.950	-0.996	-0.881
	AUC	0.756				
Module 2						
ABC	Intercept	-1.377	0.033	-41.573	-1.443	-1.313
	Connectivity	0.173	0.033	5.227	0.108	0.237
	Climate	0.827	0.039	21.112	0.750	0.904
	Habitat	-1.217	0.035	-34.587	-1.287	-1.149
	AUC	0.846				
AB	Intercept	-1.095	0.027	-40.679	-1.148	-1.043
	Connectivity	0.085	0.029	2.915	0.028	0.142
	Climate	1.125	0.035	32.150	1.057	1.194
	AUC	0.758				
AC	Intercept	-1.244	0.030	-41.500	-1.303	-1.185

Model	Coefficient	Stat	Std. error	*z* score	2.5th perc.	97.5th perc.
	Connectivity	0.530	0.028	18.660	0.475	0.586
	Habitat	-1.382	0.034	-40.450	-1.450	-1.316
	AUC	0.823				
BC	Intercept	-1.371	0.033	-41.670	-1.436	-1.307
	Climate	0.921	0.035	26.440	0.853	0.990
	Habitat	-1.206	0.035	-34.430	-1.275	-1.138
	AUC	0.845				
A	Intercept	-0.908	0.023	-39.330	-0.953	-0.863
	Connectivity	0.559	0.024	22.900	0.511	0.607
	AUC	0.644				
B	Intercept	-1.092	0.027	-40.720	-1.145	-1.040
	Climate	1.168	0.032	36.680	1.106	1.231
	AUC	0.757				
C	Intercept	-1.172	0.028	-41.340	-1.228	-1.117
	Habitat	-1.406	0.034	-41.760	-1.472	-1.340
	AUC	0.804				
Module 3						
ABC	Intercept	-2.304	0.045	-50.897	-2.394	-2.216
	Connectivity	0.055	0.044	1.245	-0.032	0.142
	Climate	1.769	0.049	36.431	1.675	1.866
	Habitat	0.044	0.034	1.312	-0.022	0.110
	AUC	0.886				
AB	Intercept	-2.310	0.045	-51.143	-2.400	-2.223
	Connectivity	0.055	0.044	1.231	-0.033	0.141
	Climate	1.761	0.048	36.471	1.668	1.858
	AUC	0.885				
AC	Intercept	-1.671	0.030	-56.270	-1.729	-1.613
	Connectivity	0.728	0.028	25.940	0.673	0.783
	Habitat	-0.275	0.028	-9.780	-0.330	-0.220
	AUC	0.748				
BC	Intercept	-2.309	0.045	-51.099	-2.399	-2.222
	Climate	1.802	0.041	43.595	1.722	1.884
	Habitat	0.044	0.034	1.299	-0.022	0.110
	AUC	0.886				
A	Intercept	-1.645	0.029	-56.580	-1.703	-1.589
	Connectivity	0.753	0.028	26.830	0.698	0.808
	AUC	0.748				
B	Intercept	-2.316	0.045	-51.370	-2.406	-2.229
	Climate	1.794	0.041	43.850	1.715	1.875
	AUC	0.885				
C	Intercept	-1.518	0.027	-57.290	-1.570	-1.466
	Habitat	-0.331	0.027	-12.460	-0.383	-0.279

Model	Coefficient	Stat	Std. error	*z* score	2.5th perc.	97.5th perc.
	AUC	0.596				
Module 4						
ABC	Intercept	-0.922	0.034	-27.490	-0.988	-0.857
	Connectivity	-0.808	0.058	-13.920	-0.924	-0.696
	Climate	-1.450	0.062	-23.270	-1.573	-1.329
	Habitat	1.031	0.037	28.010	0.959	1.104
	AUC	0.923				
AB	Intercept	-0.895	0.032	-27.934	-0.958	-0.833
	Connectivity	-0.449	0.053	-8.426	-0.555	-0.346
	Climate	-1.838	0.060	-30.487	-1.957	-1.720
	AUC	0.895				
AC	Intercept	-0.917	0.032	-28.680	-0.980	-0.855
	Connectivity	-1.984	0.043	-45.670	-2.071	-1.900
	Habitat	1.191	0.036	33.450	1.122	1.262
	AUC	0.909				
BC	Intercept	-0.910	0.033	-27.400	-0.975	-0.845
	Climate	-2.132	0.045	-47.280	-2.222	-2.045
	Habitat	0.905	0.035	26.130	0.838	0.974
	AUC	0.919				
A	Intercept	-0.826	0.029	-28.620	-0.883	-0.770
	Connectivity	-1.984	0.040	-49.550	-2.063	-1.906
	AUC	0.867				
B	Intercept	-0.900	0.032	-28.090	-0.963	-0.838
	Climate	-2.228	0.043	-52.180	-2.312	-2.145
	AUC	0.894				
C	Intercept	-0.588	0.024	-24.670	-0.635	-0.542
	Habitat	1.206	0.030	40.800	1.148	1.264
	AUC	0.778				

Notes: AUC denotes area under the curve and the lower 2.5th and upper 97.5th percentiles are bounds of profile likelihood confidence intervals.

Table S1.8 Detailed results of variation partitioning analysis based on explanatory power of biogeographic constructs (with scores calculated by non-metric partial least-squares structural equation modelling) in multiple logistic regression for each primary site module

Latent construct	Fraction	Tjur R^2
Module 1		
Total Connectivity	--	0.4895
Total Climate	--	0.7282
Total Habitat	--	0.1431
Pure Connectivity	a	0.0052
Pure Climate	b	0.2335
Pure Habitat	c	0.0019
Pure Connectivity | Climate overlap	d	0.3518
Pure Climate | Habitat overlap	e	0.0087
Pure Connectivity | Habitat overlap	f	-0.0017
Connectivity | Climate | Habitat overlap	g	0.1342
Unexplained	h	0.2663
Module 2		
Total Connectivity	--	0.0547
Total Climate	--	0.1760
Total Habitat	--	0.2440
Pure Connectivity	a	0.0034
Pure Climate	b	0.0424
Pure Habitat	c	0.1431
Pure Connectivity | Climate overlap	d	0.0306
Pure Climate | Habitat overlap	e	0.0803
Pure Connectivity | Habitat overlap	f	-0.0021
Connectivity | Climate | Habitat overlap	g	0.0227
Unexplained	h	0.6796
Module 3		
Total Connectivity	--	0.0745
Total Climate	--	0.3401
Total Habitat	--	0.0151
Pure Connectivity	a	0.0002
Pure Climate	b	0.2571
Pure Habitat	c	0.0008
Pure Connectivity | Climate overlap	d	0.0687
Pure Climate | Habitat overlap	e	0.0087
Pure Connectivity | Habitat overlap	f	<0.0001
Connectivity | Climate | Habitat overlap	g	0.0057
Unexplained	h	0.6589
Module 4		
Total Connectivity	--	0.3927
Total Climate	--	0.4776
Total Habitat	--	0.2206
Pure Connectivity	a	0.0180

Latent construct	**Fraction**	**Tjur R^2**
Pure Climate	b	0.0503
Pure Habitat	c	0.0733
Pure Connectivity | Climate overlap	d	0.2690
Pure Climate | Habitat overlap	e	0.0416
Pure Connectivity | Habitat overlap	f	-0.0110
Connectivity | Climate | Habitat overlap	g	0.1167
Unexplained	h	0.4421

Table S1.9 Summary of changes to site modules after removal of non-native species

Changes	**Main analysis** ***n* = 10,016**	**Counts of changes after removal of non-native species** ***n* = 9,993**	**Counts of changes after removal of sites with non-native species** ***n* = 9,072**
Module 1 total	1,258	1,297	2262
Module 1 to 2		1	0
Module 1 to 3		16	0
Module 1 to 4		4	0
Module 2 total	3,010	3,271	2938
Module 2 to 1		38	19
Module 2 to 3		20	0
Module 2 to 4		57	0
Module 3 total	1,851	1,773	2
Module 3 to 1		10	1386
Module 3 to 2		134	170
Module 3 to 4		30	97
Module 4 total	3,897	3,652	3870
Module 4 to 1		12	73
Module 4 to 2		261	20
Module 4 to 3		63	1

Notes: See Table S1.1 for list of non-native species; 23 sites were lost after removal of non-native species from the species pool (module 2 = 20 and module 3 = 3); 944 sites were lost after removal of sites with non-native species from the site pool (module 1 = 474, module 2 = 243, module 3 = 197, and module 4 = 30); and modularity (Q) values of sensitivity analyses after removal of non-native species from the species pool and removal of sites with non-native species from the site pool were 0.3284089 and 0.3258256, respectively.

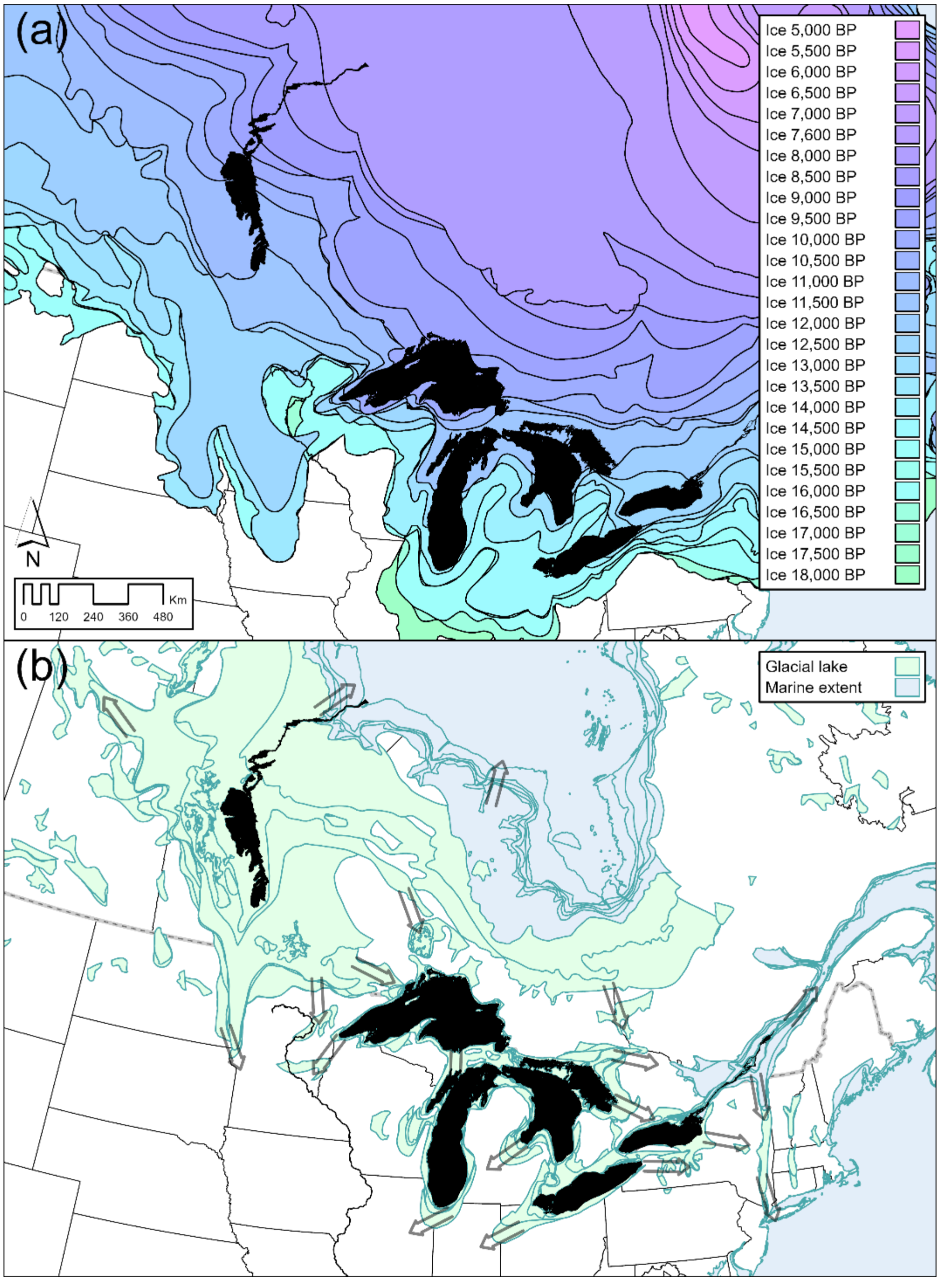

Figure S1.1 Maps (presenting spatial data from Dyke et al. (2003) Deglaciation of North America. Geological Survey of Canada Open File, 1574. Thirty-two digital maps at 1:7,000,000 scale with accompanying digital chronological database and one poster (two sheets) with full map series) showing **(a)** time since glaciation (before present) and **(b)** glacial lake and marine extents with most recent extents on top (chronological) and arrows representing major historical outflows (modified from Kehew et al. (2009) Proglacial megaflooding along the margins of the Laurentide Ice Sheet, *in* Burr, D., Carling, P.A., and Baker, V.R., eds., *Megaflooding on Earth and Mars*. Cambridge University Press, Cambridge, UK, p. 104–127; and Teller (2003) Controls, history, outburst, and impact of large late-Quaternary proglacial lakes in North America. *Developments in Quaternary Science*, 1:45–61)

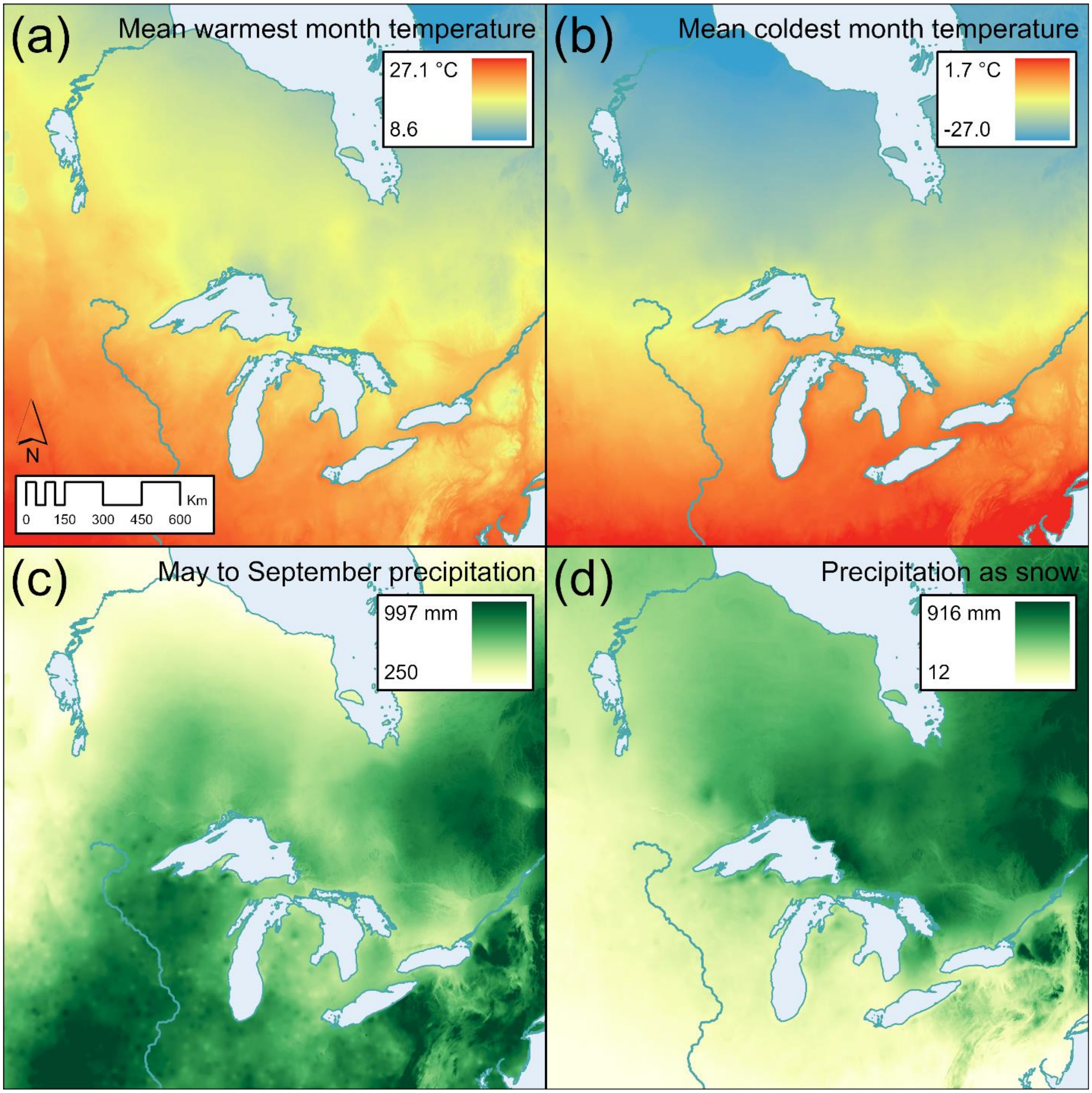

Figure S1.2 Maps (presenting spatial data derived from ClimateNA v6.40; Wang, Hamann, Spittlehouse, & Carroll (2016) Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE*, 11: e0156720) showing 30-year normals (1961–1990) at approximately 250 m resolution for **(a)** mean warmest month temperature, **(b)** mean coldest month temperature, **(c)** May to September precipitation, and **(d)** precipitation as snow

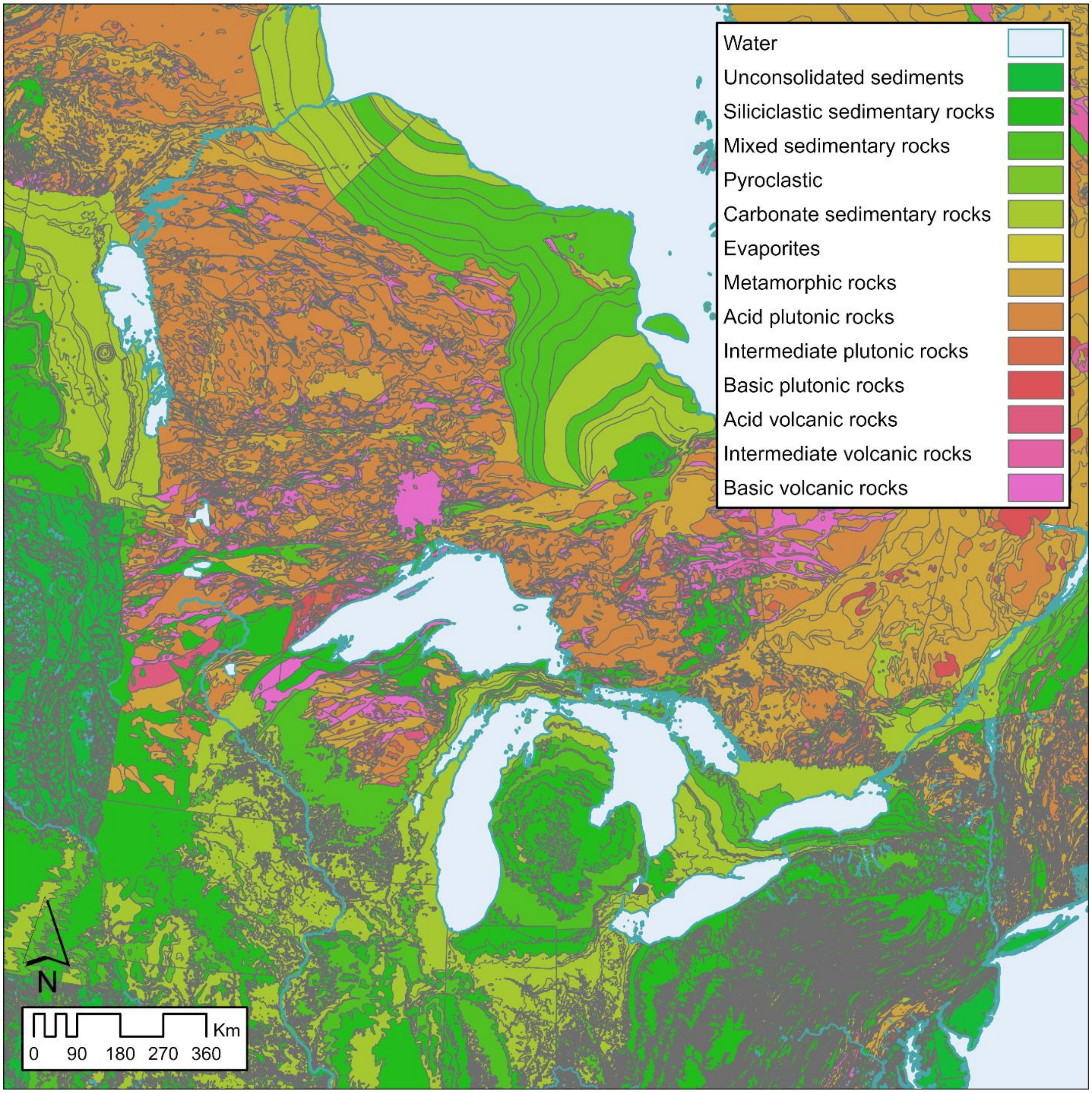

Figure S1.3 Map (presenting spatial data from Hartmann & Moosdoft (2012) The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, 13:12) showing basic lithological classes used to derive proportion of shield (metamorphic + acid plutonic + basic plutonic + intermediate plutonic + acid volcanic + basic volcanic + intermediate volcanic rocks) versus sedimentary rock (carbonate sedimentary + mixed sedimentary + siliciclastic sedimentary)

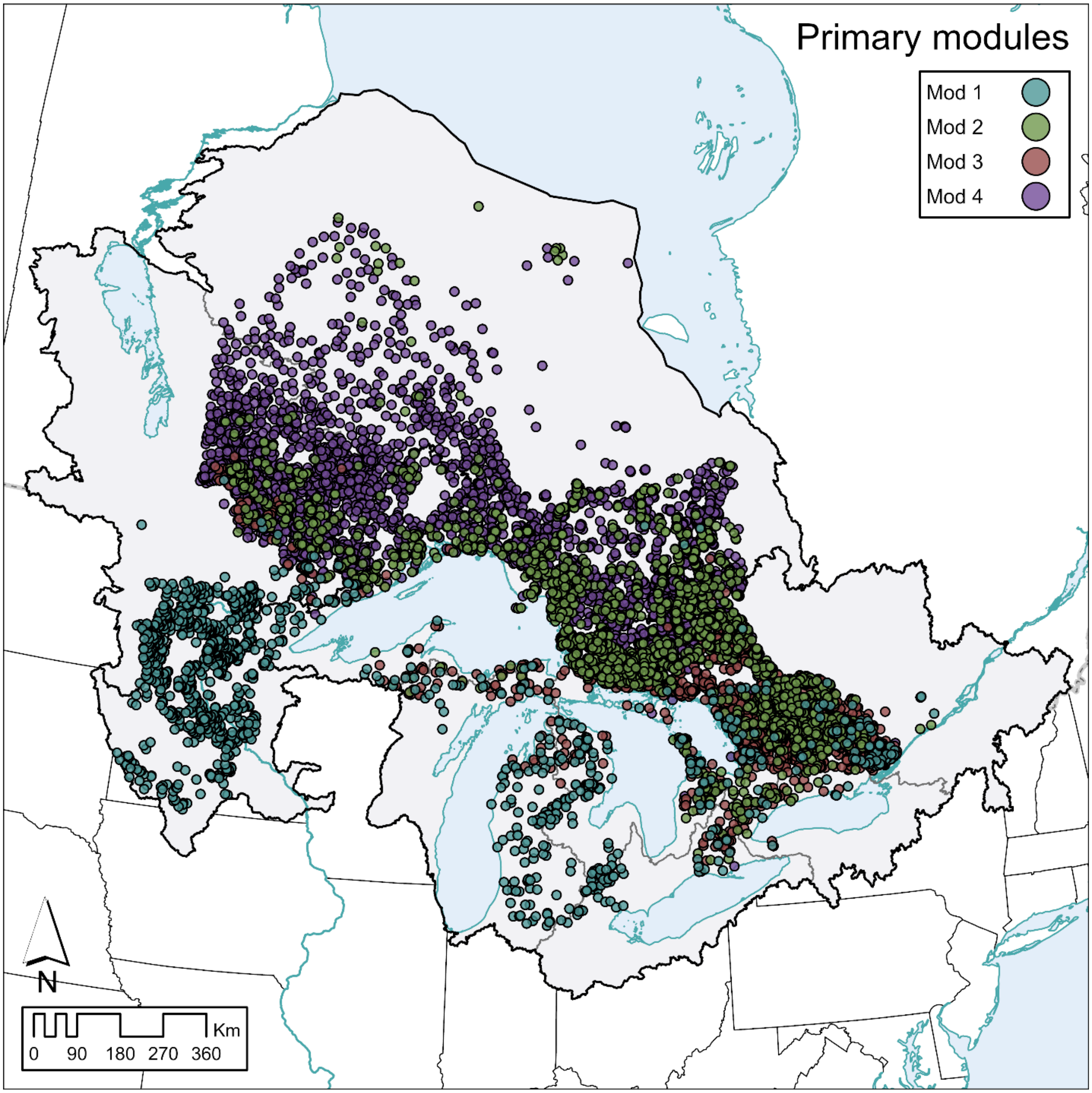

Figure S1.4 Map showing network primary modularity results for individual site nodes

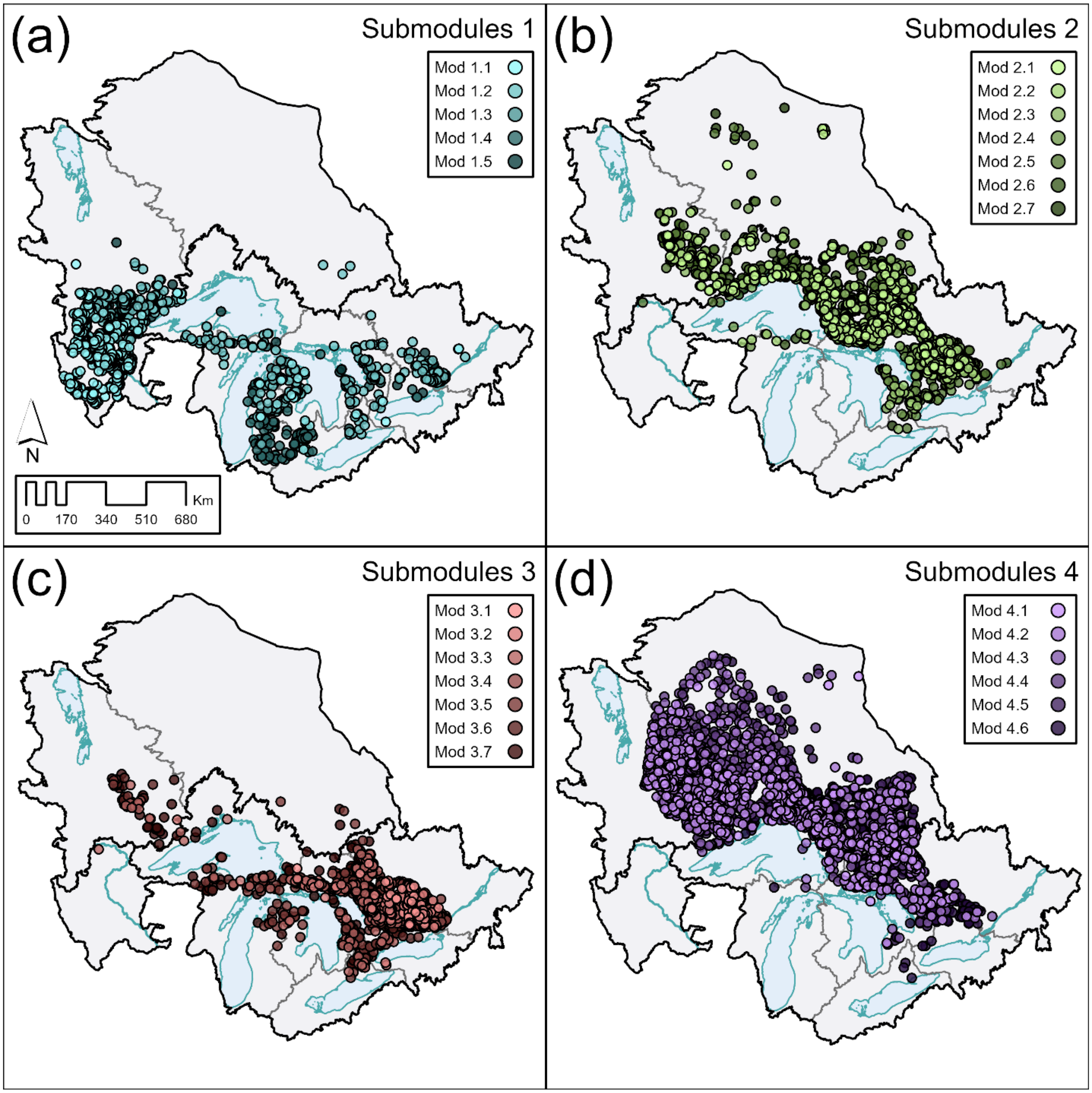

Figure S1.5 Maps showing network sub-modularity results for individual site nodes

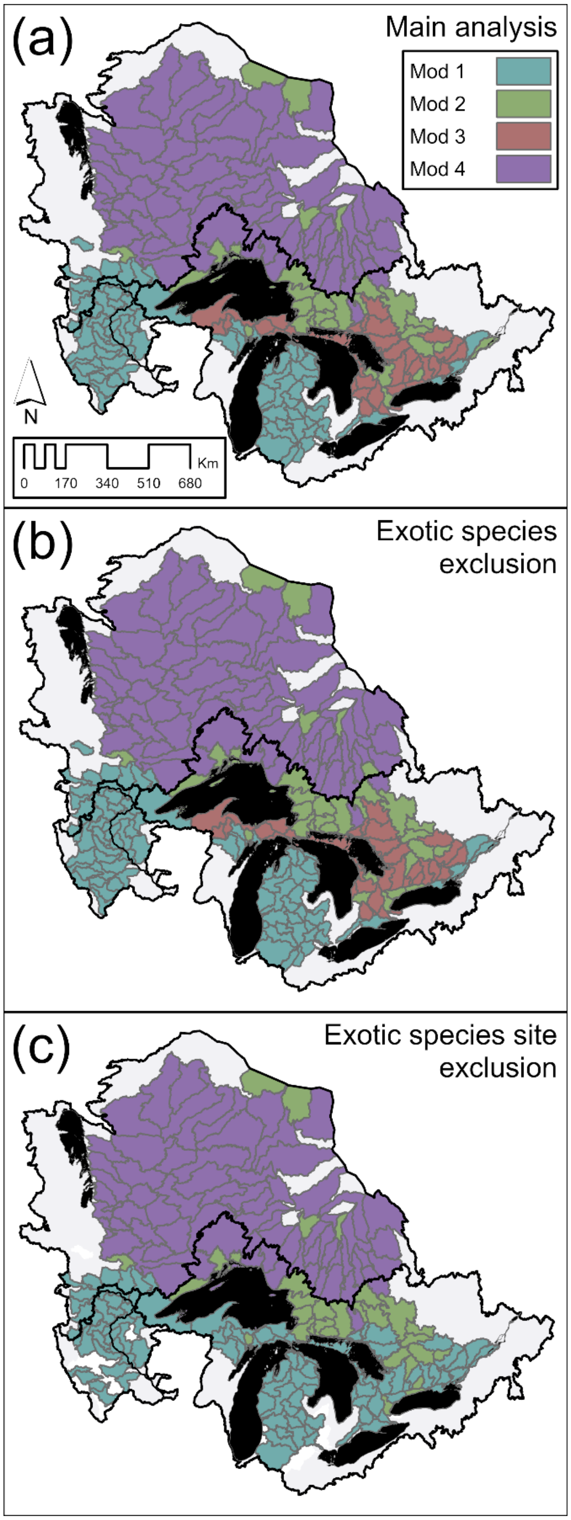

Figure S1.6 Maps showing network primary modularity results for site nodes summarized by tertiary drainages to identify bioregions for our main analysis **(a)**, after removal of non-native species from the species pool **(b)**, and after removal of sites with non-native species from the site pool **(c)**

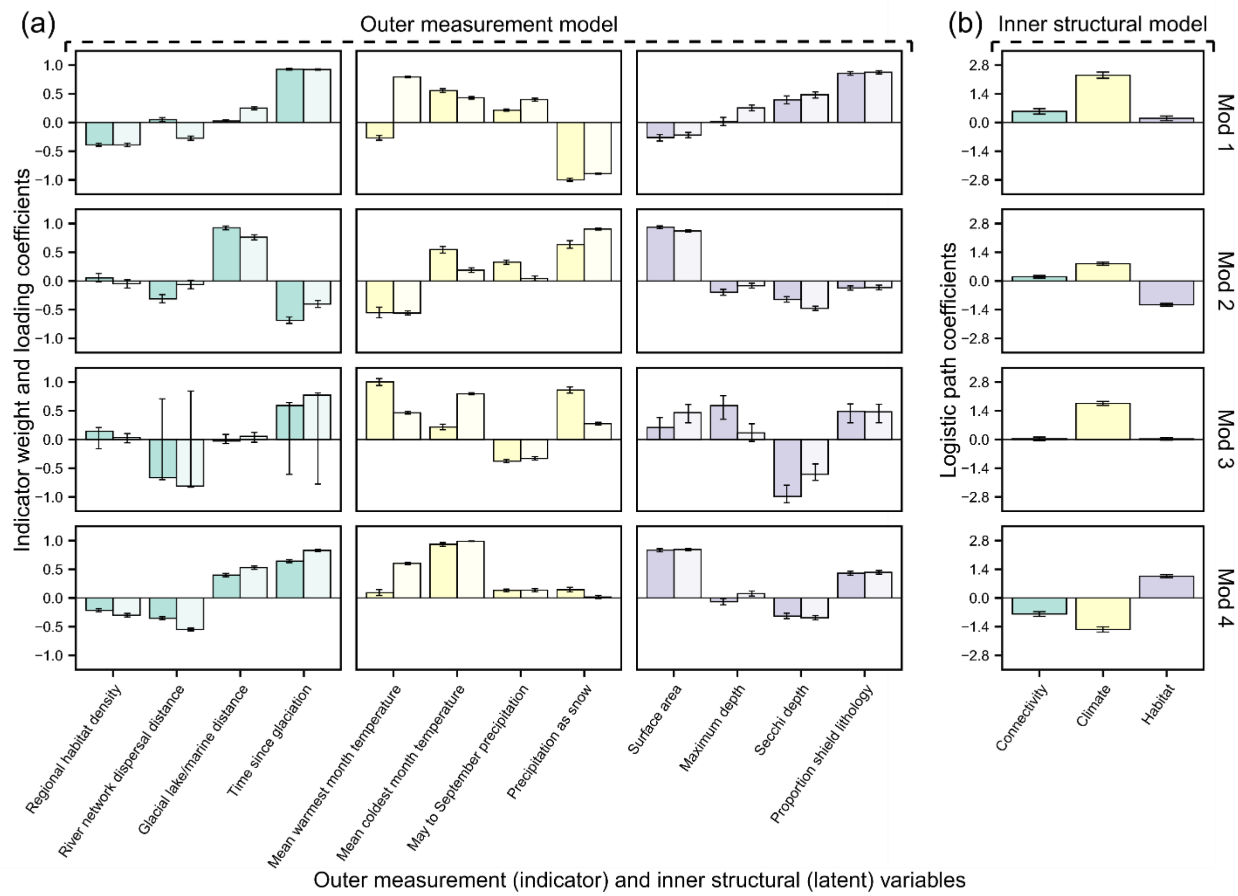

Figure S1.7 Results of NM-PLS-SEM and multiple logistic regression of primary site modules on biogeographic constructs after removal of non-native species from the species pool (see Table S1.1 for list of non-native species). Outer measurement model results **(a)** are presented as weights (darker bars; representing multiple ordinary least-squares regression coefficients) and loadings (lighter bars; representing Pearson correlation coefficients) for standardized indicators. Inner structural model results **(b)** are presented as path coefficients from multiple logistic regression for latent variables. Error bars show the lower 2.5th and upper 97.5th percentiles of bootstrap and profile likelihood confidence intervals for outer and inner model coefficients, respectively. See Figure 5 in main text for results of full analysis

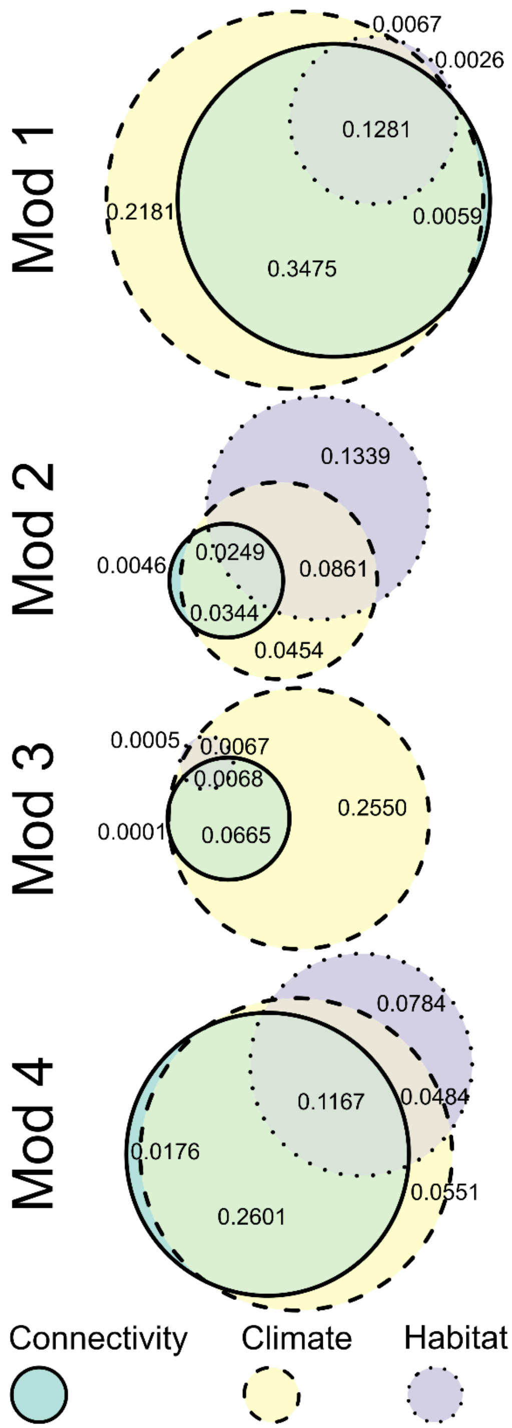

Figure S1.8 Approximately area-proportional Euler plots showing the total, shared, and unique explanatory power of biogeographic constructs (connectivity, climate, and habitat) in logistic regression after removal of non-native species from the species pool (see Table S1.1 for list of non-native species). Fractions are expressed as Tjur's R^2 (coefficients of discrimination) from models of primary site modules (binary responses) predicted by latent variable scores (calculated by NM-PLS-SEM). See Figure 6 in main text for results of full analysis

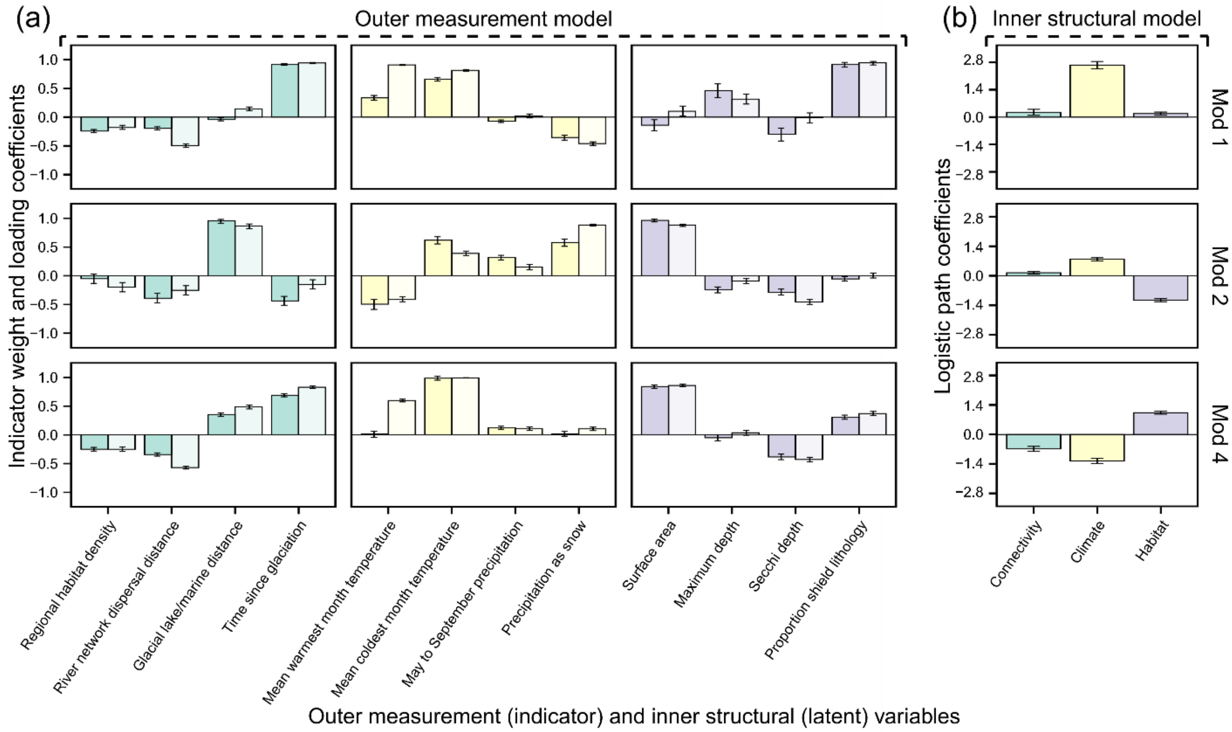

Figure S1.9 Results of NM-PLS-SEM and multiple logistic regression of primary site modules on biogeographic constructs after removal of sites with non-native species from the site pool (see Table S1.1 for list of non-native species). Outer measurement model results **(a)** are presented as weights (darker bars; representing multiple ordinary least-squares regression coefficients) and loadings (lighter bars; representing Pearson correlation coefficients) for standardized indicators. Inner structural model results **(b)** are presented as path coefficients from multiple logistic regression for latent variables. Error bars show the lower 2.5th and upper 97.5th percentiles of bootstrap and profile likelihood confidence intervals for outer and inner model coefficients, respectively. See Figure 5 in main text for results of full analysis

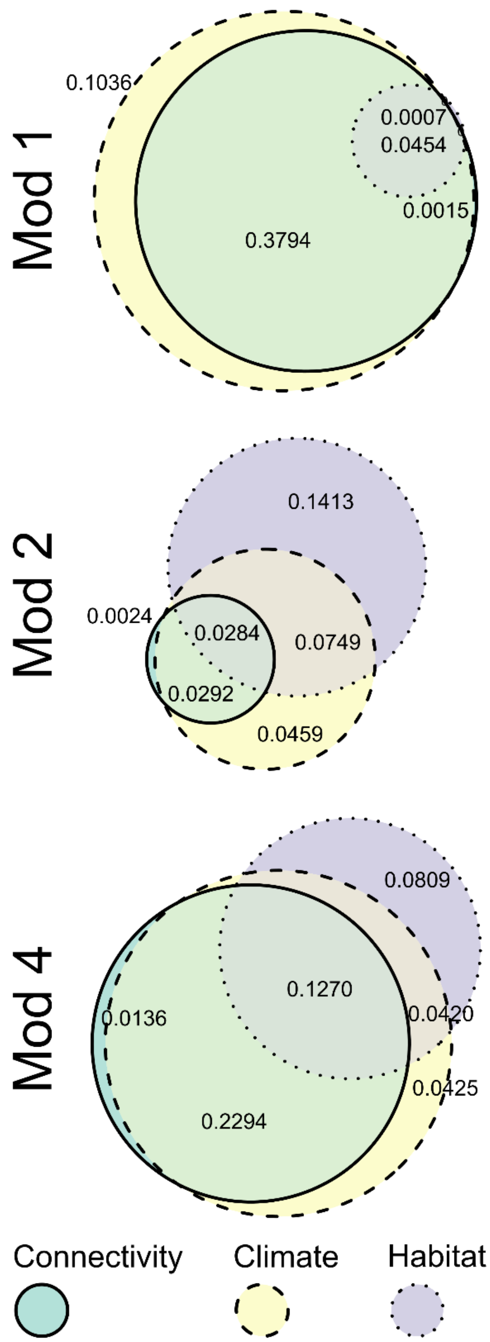

Figure S1.10 Approximately area-proportional Euler plots showing the total, shared, and unique explanatory power of biogeographic constructs (connectivity, climate, and habitat) in logistic regression after removal of sites with non-native species from the site pool (see Table S1.1 for list of non-native species). Fractions are expressed as Tjur's R^2 (coefficients of discrimination) from models of primary site modules (binary responses) predicted by latent variable scores (calculated by NM-PLS-SEM). See Figure 6 in main text for results of full analysis