1　　**Context-dependent interactions and the regulation of species richness in freshwater fish**

2　　MacDougall, A.S. et al.

3

4　　Supplementary Table 1. Species list of (A) Ontario fish species included in this study, based on

5　　species captured by BSM surveys in at least 40 of the 721 lakes, and (B) species grouped into the

6　　category of 'small prey': these species are assumed to be prey for most other fish species, and mostly

7　　inhabit the littoral zone of lakes. Further, their small size means they tend to be under-sampled by the

8　　BSM net-based fish sampling. The remaining species (C) occurred in fewer than 40 lakes, and were

9　　excluded from the analysis due to their infrequency. The 721 lakes occur over an 11-degree latitudinal

10　　gradient (43.06 – 54.52 degrees), range in area from 21-90,484 ha, have maximum depths of 1-214 m,

11　　range in phosphorus levels from 1.7-52.5 units, and have a range of total degree days from 1059-2246.

12　　Average length for each species are from Ontario lakes, and taken from Holm et al (2010)[1]. Thermal

13　　guilds (water temperature preferences) are taken from Hasnain et al. 2013[2].

14

Common Name	Species name	Family	Number of lakes (n = 721)	Average length (cm)	Thermal guild	Functional Group
(A) *Large fishes species*						
Northern Pike	*Esox lucius*	Esocidae	528	61.0	Cool	Predator
Muskellunge	*Esox*	Esocidae	43	96.5	Warm	Predator

	masquinongy					
Walleye	*Sander vitreus*	Percidae	499	42.0	Cool	Predator
Lake Trout	*Salvelinus namaycush*	Salmonidae	298	44.5	Cold	Predator
Burbot	*Lota lota*	Gadidae	250	38.0	Cold/Cool	Predator
Cisco	*Coregonus artedi*	Salmonidae	385	25.0	Cold	Pelagic
Lake Whitefish	*Coregonus clupeaformis*	Salmonidae	374	38.0	Cold	Pelagic
Brook Trout	*Salvelinus fontinalis*	Salmonidae	77	28.0	Cold	Pelagic
Longnose Sucker	*Catostomus catostomus*	Catostomidae	82	33.5	Cold	Littoral
White Sucker	*Catostomus commersonii*	Catostomidae	678	41	Cool	Littoral
Shorthead Redhorse	*Moxostoma macrolepidotum*	Catostomidae	90	41.0	Warm	Littoral
Rock Bass	*Ambloplites rupestris*	Centrarchidae	295	20.0	Cool	Littoral
Pumpkinseed Sunfish	*Lepomis gibbosus*	Centrarchidae	154	18.0	Warm	Littoral
Bluegill	*Lepomis macrochirus*	Centrarchidae	48	19.0	Warm	Littoral
Smallmouth Bass	*Micropterus dolomieu*	Centrarchidae	333	30.0	Warm	Predator
Largemouth	*Micropterus*	Centrarchidae	85	30.0	Warm	Predator

Bass	*salmoides*					
Black Crappie	*Pomoxis nigromaculatus*	Centrarchidae	55	21.5	Cool	Littoral
Brown Bullhead	*Ameiurus nebulosus*	Ictaluridae	133	28.0	Warm	Littoral
Yellow Perch	*Perca flavescens*	Percidae	526	18.0	Cool	Littoral
Sauger	*Sander canadensis*	Percidae	45	33.0	Cool	Predator
(B) *Small prey fishes*						
Mottled Sculpin	*Cottus bairdii*	Cottidae	51	7.5	Cold	Small prey
Slimy Sculpin	*Cottus cognatus*	Cottidae	51	7.5	Cold	Small prey
Spoonhead Sculpin	*Cottus ricei*	Cottidae	22	5.5	Cold	Small prey
Deepwater sculpin	*Myoxocephalus thompsonii*	Cottidae	11	8.0	Cold	Small prey
Lake Chub	*Couesius plumbeus*	Cyprinidae	104	10.0	Cold	Small prey
Common Shiner	*Luxilus cornutus*	Cyprinidae	155	9.0	Cool	Small prey
Northern Pearl Dace	*Margariscus margarita*	Cyprinidae	50	9.0	Cold/cool	Small prey
Golden Shiner	*Notemigonus crysoleucas*	Cyprinidae	8	10.0	Cool	Small prey

Emerald Shiner	*Notropis atherinoides*	Cyprinidae	159	7.5	Cool	Small prey
Blackchin Shiner	*Notropis heterodon*	Cyprinidae	49	6.0	Cool/warm	Small prey
Blacknose Shiner	*Notropis heterolepis*	Cyprinidae	127	6.5	Cool/warm	Small prey
Spottail Shiner	*Notropis hudsonius*	Cyprinidae	344	7.0	Cold/cool	Small prey
Mimic Shiner	*Notropis volucellus*	Cyprinidae	46	6.0	Warm	Small prey
Northern Redbelly Dace	*Phoxinus eos**	Cyprinidae	42	5.5	Cool/warm	Small prey
Bluntnose Minnow	*Pimephales notatus*	Cyprinidae	113	6.5	Warm	Small prey
Fathead Minnow	*Pimephales promelas*	Cyprinidae	42	5.0	Warm	Small prey
Blacknose Dace	*Rhinichthys atratulus*	Cyprinidae	13	8.0	Cool	Small prey
Longnose Dace	*Rhinichthys cataractae*	Cyprinidae	12	7.5	Cool	Small prey
Creek Chub	*Semotilus atromaculatus*	Cyprinidae	39	10.0	Cool	Small prey
Banded Killifish	*Fundulus diaphanus*	Fundulidae	9	7.0	Cool	Small prey
Brook	*Culaea*	Gasterosteidae	40	5.0	Cool	Small prey

Stickleback	*inconstans*					
Ninespine Stickleback	*Pungitius pungitius*	Gasterosteidae	64	6.5	Warm/cold*	Small prey
Iowa Darter	*Etheostoma exile*	Percidae	13	5.0	Cool	Small prey
Johnny Darter	*Etheostoma nigrum*	Percidae	39	5.0	Cool	Small prey
Logperch	*Percina caprodes*	Percidae	180	9.0	Cool/warm	Small prey
Trout-perch	*Percopsis omiscomaycus*	Percopsidae	302	9.0	Cold	Small prey
Central Mudminnow	*Umbra limi*	Umbridae	9	7.5	Cool/warm	Small prey
(C) *Less common species*						
Longnose gar	*Lepisosteus osseus*	Lepisosteidae	10	76.0	Warm	Predator
Alewife	*Alosa pseudoharengus*	Clupeidae	5	15.0	Cold	Pelagic
Goldeye	*Hiodon alosoides*	Hiodontidae	6	28.0	Warm	Pelagic
Mooneye	*Hiodon tergisus*	Hiodontidae	13	28.0	Cool/warm	Pelagic
Rainbow trout	*Oncorhynchus mykiss*	Salmonidae	4	53.0	Cold	Pelagic
Round whitefish	*Prosopium cylindraceum*	Salmonidae	12	25.0	Cold	Pelagic
Bowfin	*Amia calva*	Amiidae	9	54.0	Warm	Littoral

Silver Redhorse	*Moxostoma anisurum*	Catostomidae	21	40.0	Cool	Littoral
Greater Redhorse	*Moxostoma valenciennesi*	Catostomidae	10	41.0	Cool/warm	Littoral
Common Carp	*Cyprinus carpio*	Cyprinidae	14	37.0	Warm	Littoral
Fallfish	*Semotilus corporalis*	Cyprinidae	13	20.0	Cool	Littoral
Yellow Bullhead	*Ameiurus natalis*	Ictaluridae	20	25.5	Warm	Littoral

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29   Supplementary Table 2. Species association values, measured as null-adjusted Pearson correlation

30   coefficients for abundances of 16 species identified as potential 'interaction' species by the SEM

31   analysis (Figs. 1, Suppl. Fig. 3). The association analysis tests the direction and statistical significance

32   of these 'interactions', correcting for bias in zero-inflated data arising because the correlation

33   coefficient is not double-zero symmetrical[3]. Significantly negative (32/120) pairings are grey or black

34   highlighted cells. These results cannot test whether the significant associations derive from species

35   interactions, or if negative associations derive from divergent habitat affinity (e.g., species x is absent

36   when species y is present, because they prefer different lake conditions). However, by organizing

37   these results by temperature affinity (BLUE = cold water species, RED = cool, or GREEN = warm),

38   most negative interactions occur among temperature groups (grey cells) – for example, warm and cold

39   water species tending not to co-occur. Only five negative species pairings were within- group (black

40   cells). There were 37/120 positive pairings and 51/120 non-significant pairings – we assumed these

41   associations to not reflect antagonistic species interactions, and thus could not generate restricted co-

42   occurrence between the species pairs.

43

Species	LNSk	Whtfh	BTrt	Burb	Cisco	YPer	Pike	Wall	WSk	Red	SMB	Pmp	LMB	RBss	BBul
Lake Trout	0.07	0.11	-0.08	0.08	0.02	0.00	-0.20	-0.27	0.18	-0.09	0.08	-0.07	-0.08	0.12	-0.05
LNSucker		0.28	0.22	-0.02	0.00	-0.05	-0.09	-0.08	0.13	-0.03	-0.08	-0.03	-0.02	-0.03	-0.04
Whitefish			-0.01	0.22	0.10	0.03	0.08	0.11	-0.02	-0.04	-0.05	-0.11	-0.10	-0.03	-0.09
Brook Trout				-0.09	-0.06	-0.11	-0.17	-0.17	0.39	-0.06	-0.12	-0.05	-0.05	-0.05	-0.04
Burbot					0.15	0.02	0.00	0.16	-0.04	0.01	0.01	-0.07	-0.02	-0.01	-0.05
Cisco						0.16	0.30	0.25	0.05	0.02	0.04	-0.03	0.01	-0.02	0.06
Yellow Perch							0.20	-0.01	-0.01	-0.04	0.10	0.24	0.20	0.03	0.18
Pike								0.42	-0.12	0.07	-0.12	0.01	0.09	-0.12	0.09
Walleye									-0.15	0.16	-0.14	-0.11	-0.08	-0.16	-0.06
White Sucker										-0.07	-0.17	-0.05	-0.08	-0.08	-0.05
Redhorse											-0.03	-0.04	-0.04	-0.04	-0.04
SM Bass												0.31	0.18	0.30	0.12
Pumpkinseed													0.27	0.08	0.27
LM Bass														0.06	0.26
Rock Bass															0.03
Brown Bullhead															

45   Supplementary Table 3. Boosted regression tree (BRT) model results, for species with one or more

46   significantly negative species associations corrected for null expectation (Supplementary Table 2). The

47   results from the association analysis cannot distinguish between species interactions versus habitat

48   affinity – the BRT analysis tests the relative influence (%) of 22 possible drivers of variation affecting

49   the presence of each species, including species composition (the presence of each of the other species)

50   and abiotic factors including degree days, pH, phosphorus, minimum depth, maximum depth, lake

51   area, Secchi depth, and small prey species richness. Correlation refers to how much a predictor model

52   estimates the variation within the raw data (%). The seven grey-shaded cells are the only species

53   pairings where the explanatory species (top row) was one of the ten strongest factors <u>negatively</u>

54   <u>associated</u> (see Suppl. Fig. 5) with the presence of the response species (vertical column on the left

55   side of the table). The species are organized by temperature affinity: blue = cold, green = cool, red =

56   warm.

57

Focal species	lake whitefish	cisco	lake trout	burbot	brook trout	walleye	pike	small mouth bass	yellow perch	rock bass	white sucker	brown bullhead	pumpkin seed	large mouth bass	Degree days	pH	Total P	Depth min (m)	Depth Max (m)	Area (ha)	secchi depth	small prey richness	No. of trees included in final model	Correlation Between observed and predicted response variables
lake whitefish		20	0.02	15	0	5	0.3	0.1	2	0.02	1	0	0.2	1	22	1	4	1	2	24	1	1	750	76
cisco	21		0	0.10	1	0.10	8	0.03	3	0.05	0.01	0.3	0.10	0	7	3	3	7	5	28	4	10	1000	77
lake trout	0	0		0.01	1	5	1	0	0.01	0	1	0	0.03	0	3	2	9	46	24	2	4	0.3	1050	85
burbot	20	0	0		0.1	0.01	1	0.3	0.3	0.1	0.1	0	0.02	0.1	4	3	1	7	34	23	3	2	900	77
brook trout	0	0.05	6	1		11	11	5	56	0.1	0	0	0	0	3	1	0.2	3	1	0.5	1	1	350	81
pike	1	4	3	0.2	11	24		0.10	22	0.5	0.4	1	0.10	0.2	12	3	2	4	2	6	3	1	1400	82
walleye	3	0	5	0	0.01		36	4	0.07	0.5	1	0	0.10	0.03	12	1	4	2	2	10	1	2	1000	79
small mouth bass	0.02	0.2	0.3	2	3	5	0.2		1	8	0.00	2	10	0.02	39	1	1	4	3	11	4	10	1050	80
yellow perch	2	2	0.2	1	44	0.2	14	3		1	0.2	0.01	0.06	0	10	2	4	4	3	6	2	3	1600	86
rock bass	0	4	0.3	0.2	0.3	2	1	10	3		0	0.05	14	4	39	3	2	3	3	6	3	4	1450	79
white sucker	7	0	5	0	0.2	4	0	0	0.4	0		0	3	0.3	31	2	1	4	3	37	6	1	300	49
brown bullhead	0.1	0.02	0	0.05	0.04	0	2	1	0.01	0.1	0		17	8	60	1	3	1	0.4	1	5	0.3	700	79
Pumpkin seed	4	0	0	0	0.1	0.1	0	2	0.02	4	0	11		11	61	1	1	2	1	1	2	1	650	81
large mouth bass	0	0	0	2	0	0	0	0	0	2	3	6	7		73	2	2	2	2	3	0	0	550	83

58

59   Supplementary Table 4. Environmental range comparisons of the negatively associated species pairs

60   identified in Supplementary Table 2. Species AB refers to lakes where both species co-occur. Ten

61   pairs had <u>range overlap</u> in environmental factors yet rarely co-occur. Nine pairs had <u>range</u>

62   <u>compression</u> in lakes of co-occurrence, where the range of environmental conditions in lakes where a

63   species occurs in the absence of the other is wider than when the two species co-occur. BRT refers to

64   pairs where the presence of one species significantly and directly predicts the absence of another (see

65   Suppl. Table 3, Suppl. Fig. 5). Cold water species, cool water species, warm water species - most

66   rarely co-occurring pairs or those with compressed co-occurrence involve species of different

67   temperature groupings.

species pair	% lakes co-occur	RANGE OVERLAP* Species A vs Species B (ANOSIM)	RANGE COMPRESSION† Species A vs AB (PERMDISP)	RANGE COMPRESSION Species B vs AB (PERMDISP)	BRT The presence of A predicts B and vice versa	CO-OCCURRENCE RANKING
Brooktrout/Pike	0.50%	0.138 (0.02%)	ns	ns	Yes	*species rarely co-*
Brooktrout/Walleye	0.30%	0.150 (0.02%)	ns	ns	Yes	*occurring (Fig.*
Brooktrout/SMBass	0.60%	0.193 (0.02%)	ns	ns	Yes	*2A)*
SMBass/LongnoseSucker	3.50%	0.203 (0.02%)	ns	ns		
Brooktrout/Burbot	1.40%	0.293 (0.02%)	ns	ns		
Brooktrout/Brownbullhead	1.08%	0.361 (0.02%)	ns	ns		
Brooktrout/LakeTrout	4.80%	0.381 (0.02%)	ns	ns	Yes	
LMBass/Whitefish	3.10%	0.448 (0.02%)	ns	ns		
Brooktrout/LMBass	0.15%	0.565 (0.02%)	NA^	NA^		
LakeTrout/LMBass	4.50%	0.641 (0.02%)	ns	ns		
LakeTrout/Pike	27.90%	0.318 (0.02%)	ns	compressed (p<0.001)	Yes	*one or both*
Pike/Rockbass	36.40%	0.383 (0.02%)	ns	compressed p=0.049)		*species occur in a*
Pike/SMBass	38.20%	0.389 (0.02%)	ns	compressed (p=0.008)		*compressed*
LakeTrout/Pumpkinseed	12.20%	0.475 (0.02%)	ns	compressed (p<0.001)		*range of*
LakeTrout/Walleye	23.70%	0.287 (0.02%)	compressed (p=0.01)	compressed (p<0.001)	Yes	*environments*

	Overlap	R value				
Rockbass/Whitefish	23.30%	0.305 (0.02%)	compressed (p<0.001)	compressed (p=0.006)		*when co-*
Burbot/Pumpkinseed	9.70%	0.348 (0.02%)	compressed (p=0.01)	compressed (p<0.001)		*occurring,*
Whitefish/Pumpkinseed	10.50%	0.363 (0.02%)	compressed (p=0.021)	compressed (p<0.001)		*compared to*
LakeTrout/ShrtnoseRedhorse	3.70%	0.428 (0.02%)	compressed (p=0.003)	compressed (p=0.017)		*when found alone*
						(Fig. 2B)
Pike/Whitesucker	71.50%	0.223 (0.10%)	ns	ns		*no compression*
Pumpkinseed/Walleye	7.90%	0.276 (0.02%)	ns	ns		*of environmental*
Rockbass/Walleye	36.50%	0.326 (0.02%)	ns	ns		*lake conditions*
LMBass/Walleye	12.20%	0.376 (0.02%)	ns	ns		*when together*
SMBass/Walleye	39.60%	0.407 (0.02%)	ns	ns		*compared to*
SMBass/Whitesucker	45.10%	0.448 (0.02%)	ns	ns		*when alone (Fig.*
Whitesucker/Rockbass	41.14%	0.428 (0.02%)	ns	ns		*2C)*
Brownbullhead/Whitefish	6.30%	0.425 (0.02%)	ns	ns		
LMBass/Whitesucker	12.90%	0.448 (0.02%)	ns	ns		
LongnoseSucker/Pike	7.60%	0.030 (3.2%)	expanded (p<0.001)	ns		*species is found*
Brooktrout/YellowPerch	2.20%	0.104 (0.3%)	expanded (p=0.04)	ns	Yes	*in a wider range*
LongnoseSucker/Walleye	6.90%	0.123 (0.5%)	expanded (p=0.017)	ns		*of environmental*
WhiteSucker/Walleye	69.20%	0.311 (0.02%)	expanded (p=0.036)	ns		*conditions when*
						the other is
						present (Fig. 2D)

68   * Overlap refers to the degree of overlap in the multivariate environmental conditions of lakes where Species A and

69   Species B occur without each other - R values closer to 0 indicate greater similarity in multivariate environmental

70   conditions. Numbers in brackets are the percentage of 5000 randomizations for which an R value as high or higher

71   occurred (i.e., a R value equal or greater to the calculated value less than 5% of the time is considered significant).

72

73   †Compression refers to a significant change in the multivariate spread or range of environmental conditions, with

74   'compressed' meaning that the range is NARROWER when the species is in the presence of the other species compared to

75   when alone; 'expanded' means the range is WIDER for lakes with both species present compared to when the species is

76   alone

77

78   ^ The PERMDISP test is not applicable (NA) in this case, because only 1 lake contained both species.
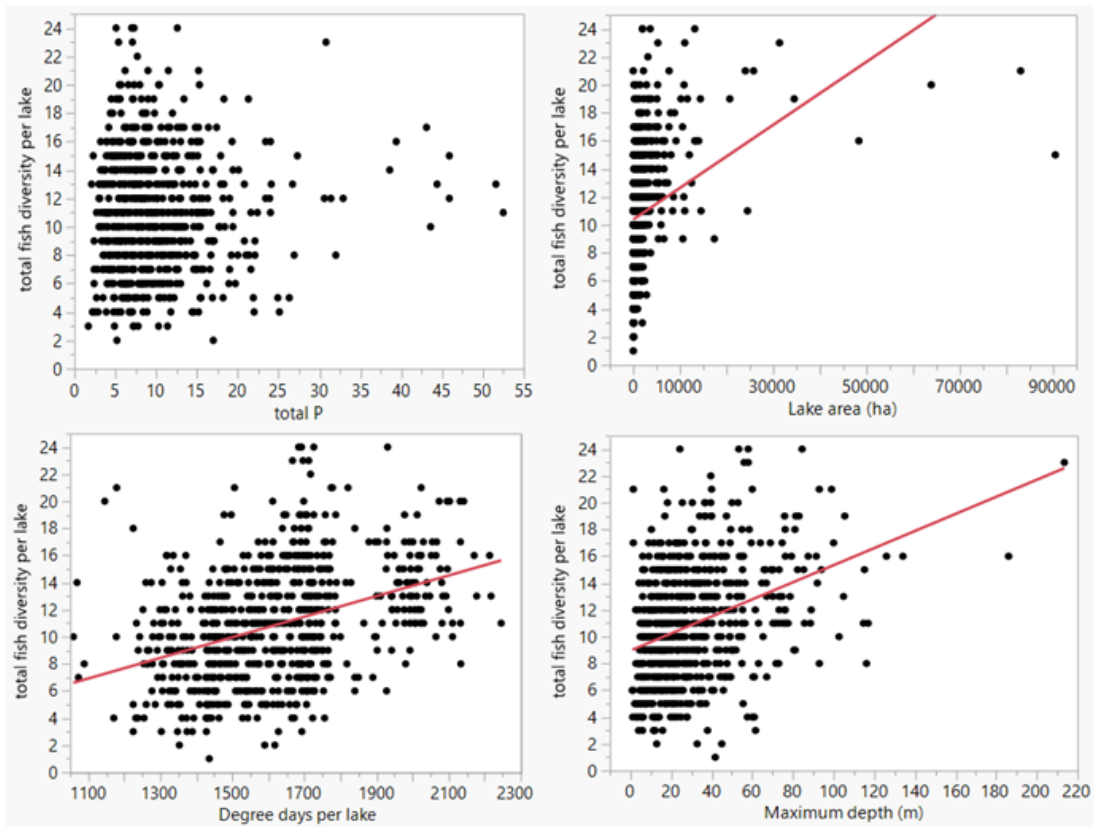
79

80

81

82   Supplementary Figure 1: Bivariate relationships (with linear fits where significant) between total fish

83   diversity and major explanatory drivers of fish occurrences in lakes. These graphs reveal the potential

84   influence of each explanatory factor on fish.

85

86

87

88

89

90   Supplementary Figure 2: Bivariate relationships between abiotic factors in 721 lakes. All relationships

91   were significant with linear fits shown, except latitude despite the concentration of human populations

92   in the southern regions of Ontario. Again, these graphs reveal how any given combination of bivariate

93   factors can significantly interact in lakes, suggesting that integrative analyses testing all

94   simultaneously should better capture how they affect each other, and by extension, the fish

95   communities found within.

96
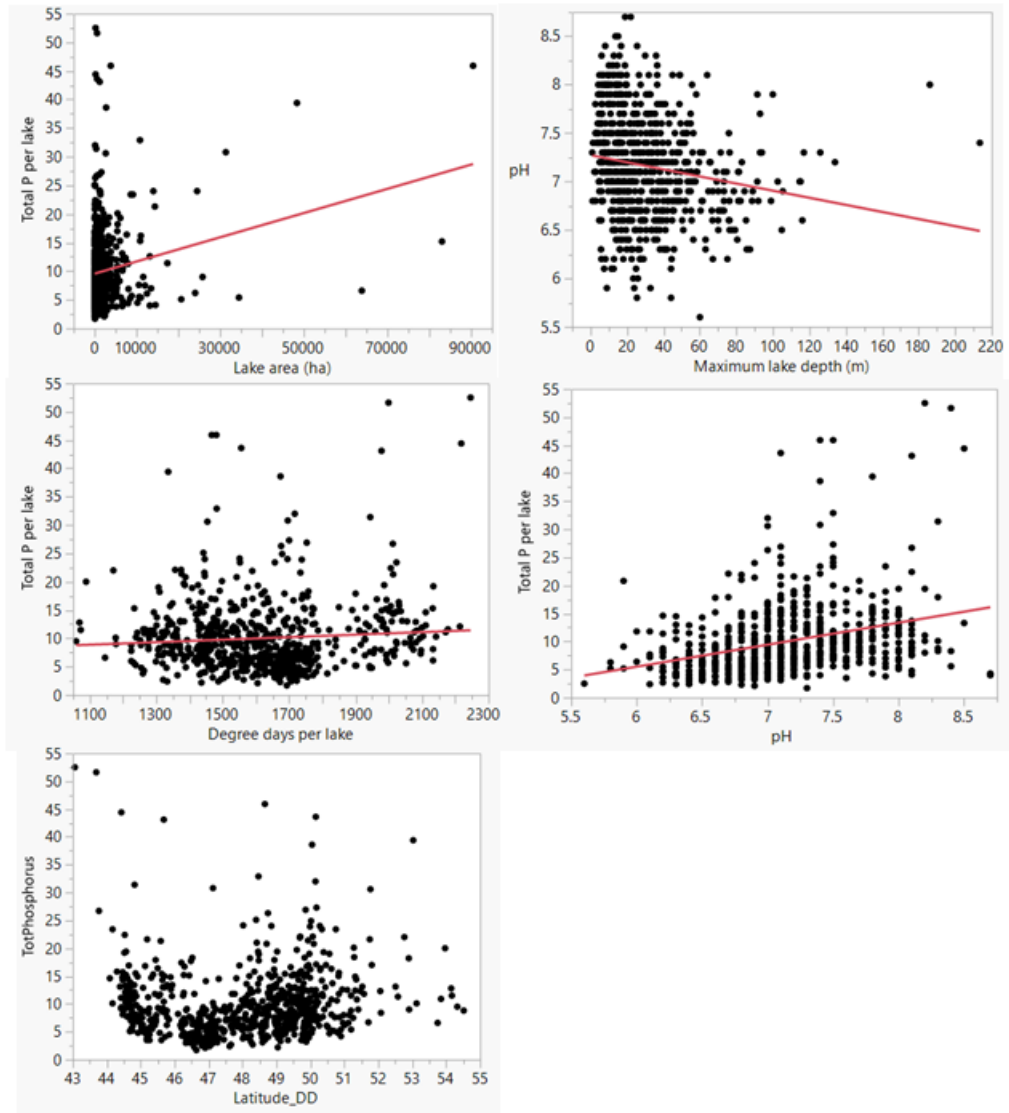
97

98

99   **Supplementary Note 1 – Structural Equation Modeling**

100   Our model construction for the SEM analysis began with the entire presence/absence dataset, for

101   22 fish species in 721 lakes with ten associated environmental measures relating to climate (degree

102   days), lake morphometry (area [ha], maximum depth [m], mean depth [m]), and lake water quality

103   (total P concentration, ph, conductivity, dissolved solids, Secchi depth). To minimize strong co-

104   linearity in our model and maximize parsimony, we screened the correlation matrix among the

105   environmental variables (e.g., see some correlations displayed in Suppl. Fig. 2). This allowed us to

106   reduce the number of environmental variables in the SEM from 10 to 5, with these 5 factors

107   corresponding to our three hypothesized abiotic-based models of influence: climate (degree days), lake

108   morphometry (lake area, maximum lake depth), and water quality (P, pH). All SEM analyses were

109   conducted with R 3.1.2 (R Core Team 2016), using the lavaan package version 0.5-20[4].

110   For the 'biotic' components of the SEM, we broadly classified the 22 large-sized species

111   (average length > 10 cm) into three functional groups based on habitat affinity and foraging strategy as

112   described in Holm et al.[1] – major predators, littoral species, and pelagic species (Suppl. Table 1). We

113   also grouped the 28 small-sized species (mean length < 10 cm) into a single functional group "small

114   prey species", as described above in the Methods. We tested for influences on total numbers of species

115   per functional group ('functional diversity') and also shifts in species composition (see Methods). For

116   the latter (composition shifts), we identified which functional groups in the SEM were associated with

117   significant species composition changes among lakes using PCA analyses on Hellinger transformed

118   fish data (see Suppl. Fig. 3 and Methods for details). That is to say, which species were significantly

119   associated with the presence of other species. Finally, to satisfy the SEM assumption of multi-

120   normality we log-transformed variables with skewed distributions.

121   Our SEM analysis was based on the recognized combined importance of abiotic and biotic

122   factors in shaping fish diversity in freshwater lakes[5]. Starting from our initial model that incorporated

123   all levels of interaction among degree days, measures of lake morphometry and water quality, and

124   numbers of species per major functional group, we used modification indices[4] to guide decisions about

125   adding missing paths to the model. We used the "modindices" function from the lavaan package[4],

126   which provides a list of all missing linkages in the model and the expected effect of additions on

127   model fit (Maximum likelihood estimator [ML_{EST}]). We used the modification indices incrementally,

128   adding one path at a time, until no modification indices were higher than 2. After this process, we

129   scanned path regressions and incrementally pruned the less significant ones until there was no more

130   gain in model fit (decreasing ML_{EST}[6]). Our final model converged after 171 iterations (n = 648,

131   ML_{EST} = 4.91, Degree of freedom = 13, *P*=0,977).

132   Our SEM analysis focused on the hypothesis that an integrated model capturing abiotic and

133   biotic interactions within and among functional groups would best capture the regulation of species

134   richness[5]. Yet, there is also evidence in the fish literature of more singular models shaping the

135   regulation of richness, specifically (I) top-down regulation by major predators, (II) bottom-up models

136   where richness in lower levels of the food chain influences diversity levels above, and (III) abiotic-

137   based models where climate and resources primarily affect species richness at all trophic levels. To

138   further disentangle the relative contribution of abiotic and biotic associations, we decomposed our

139   'Integrative' model into its different components each representing a contrasting hypothesis on the

140   regulation of fish diversity (i.e., no biotic associations between functional groups, top-down

141   associations only, bottom-up associations only).

142   To make our models comparable we used the same variable structure for all candidate models

143   but altered the direction of effects among functional groups. The models of 'Top-down' effects of

144   consumers on prey (Suppl. Fig. 4b) and the 'Bottom-up' effects of prey on consumers (Suppl Fig. 4c)

145   were started from an initial model including all possible top-down or bottom-up associations among

146   fish functional groups. The abiotic factors were then added using modification indices and the less

147   significant regression paths were pruned (using a similar approach as described above). The model

148   without biotic associations (thereafter 'Abiotic only' model) included no effects (arrows) between

149   functional groups so that only direct effects of environmental factors were possible (Suppl Fig. 4a).

150   We compared the four models using Akaike Information Criterion (see Suppl. Table 5 and Suppl. Fig.

151   4). By examining our main 'Integrative' model against its reduced components, we were interested in

152   contrasting the magnitude of each factor in the SEM (effect size) versus its contribution to the overall

153   fit of the model (e.g., some factors might have large effect size, but not contribute much to our overall

154   understanding of the system).

155   Results from our model comparison indicate that the model including only abiotic factors

156   performed the weakest, suggesting that despite being important drivers in terms of effect size, abiotic

157   factors are insufficient to explain the system (i.e., lowest model fit, Suppl. Table 5). That is to say,

158   ignoring biotic associations among fish functional groups led to a significant decline in analytical

159   power (Suppl. Table 5). Alternatively, top-down factors had smaller effect sizes compared to some

160   abiotic factors yet strongly contributed to model fit. This contribution centered largely on two top-

161   down associations: from 'predator PCA1' to 'small prey richness' and from 'predator richness' to

162   'pelagic richness'. Removing either of those associations from the 'Top-Down' model lead to a jump

163   in the Maximum Likelihood estimator of 40 and 30 points respectively, sending the model at the

164   bottom of the ranking in the AICc comparisons. Overall, the 'Integrative' model balanced the factors

165   with the greatest effect sizes while maximizing model fit, including the described top down

166 associations, the strong effects of degree days and lake morphometry, and the bottom-up effect of

167 littoral richness on predator richness (Fig. 1).

168   The R code for the SEM analysis is provided in full, in Supplementary Note 4.

169

170 Supplementary Table 5. Comparison of our best model with more simple alternative hypotheses.

Model	K	AICc	Δ AICc	Difference support
Top-down	48	40119.91	0.00	
Integrative	50	40120.85	0.94	None
Bottom-up	52	40130.71	10.80	Strong
Abiotic	57	40135.64	15.73	Strong

171 Critical values for model difference support are: **none** ΔAICc < 2; **substantial** 4 < ΔAICc < 7; and

172 **strong**: ΔAICc > 10[7].

173
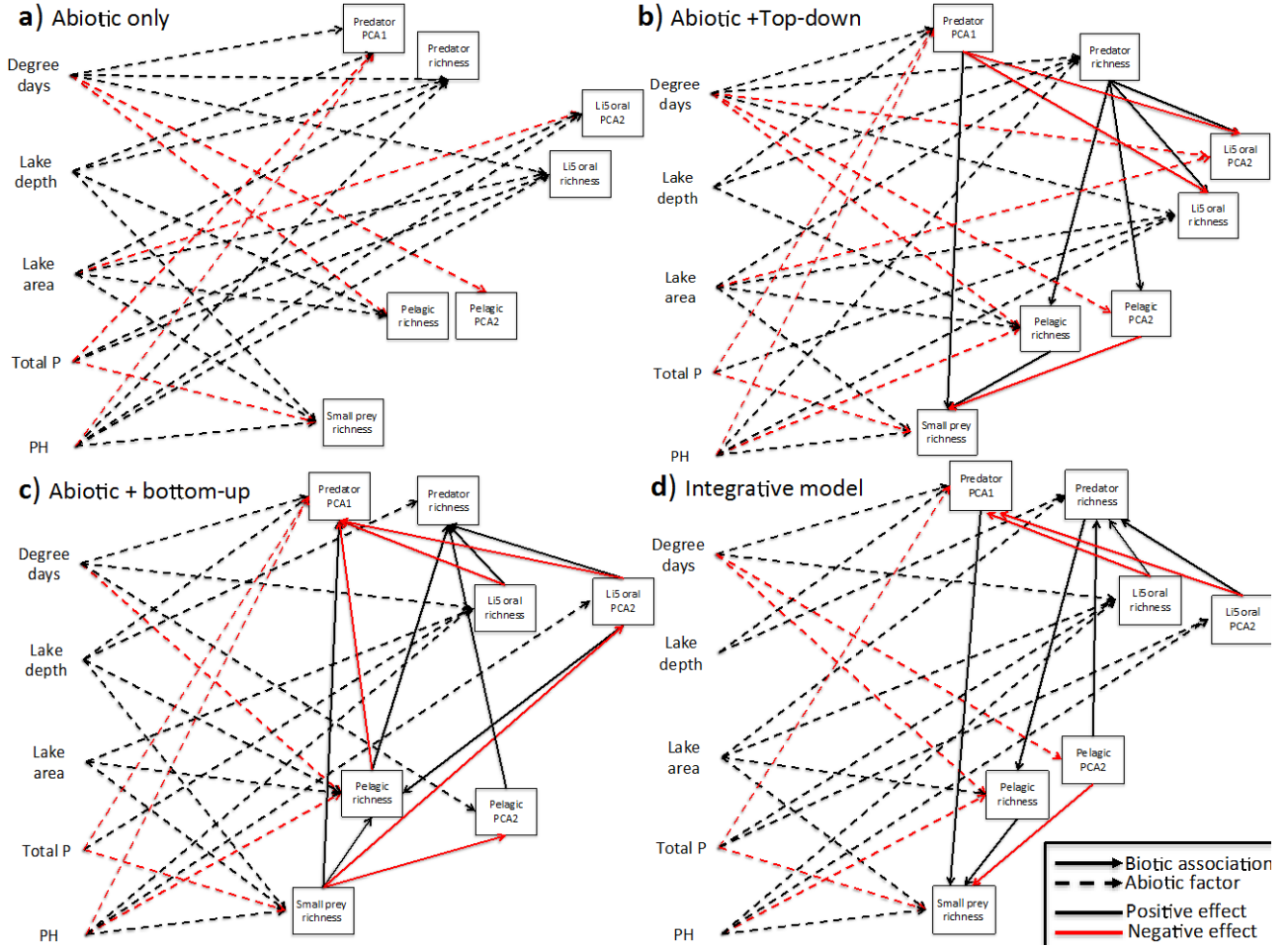
174

175

176

177

178

179

180

181

182

183

184   Supplementary Figure 3. PCA analysis of Hellinger transformed species associated with significant

185   shifts in species composition in the SEM model (Fig. 1), divided into three functional groups based on

186   broad lake habitat associations. The mechanisms underlying the association shifts cannot be

187   determined by the SEM, and may possibly reflect species interactions, habitat affinity not relating to

188   species interactions, or some combination of the two. All subsequent analyses tested the likelihood of

189   interactions versus habitat in explaining these association shifts, targeting these 16 species with the

190   greatest vector-length separation out of the initial 22 species used in the study.
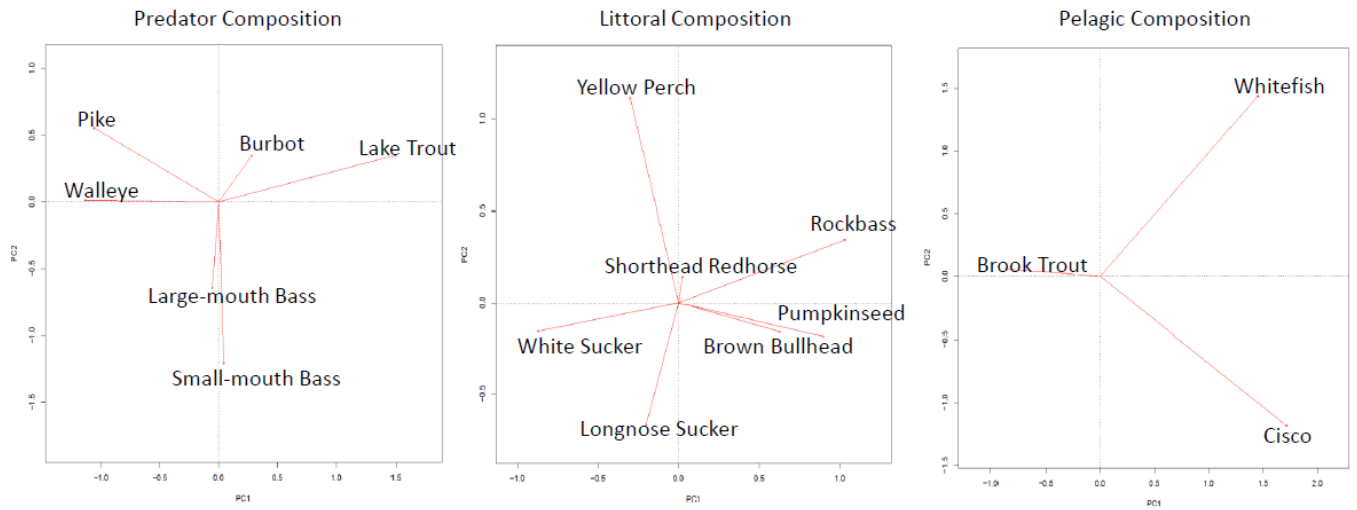
191

192

193

194

195

196

197

198

199

200

Supplementary Figure 4. Visual representation of all the alternative hypotheses considered in this study to explain fish functional group richness and compositional shifts. Each model was built iteratively with modifications indices[4] starting from an initial model which structure was constrained by specific rules as a function of the hypothesis tested (see Supplementary Notes 1). The a) Abiotic only model represents the effects of abiotic factors of fish functional group richness and composition in the absence of any biotic associations (n = 648, ML_{EST} = 3.03, Degree of freedom = 6, P=0.805), b) Top-down model represents the effects of abiotic factors interacting with top-down associations on richness and composition (n = 648, ML_{EST} = 8.66, Degree of freedom = 15, P=0.895), c) Bottom-up model represents the effects of abiotic factors interacting with bottom-up associations on richness and composition (n = 648, ML_{EST} = 10.04, Degree of freedom = 11, P=0.527), and d) Integrative model represents the combined effects of abiotic factors, bottom-up and top-down associations on richness and composition (n = 648, ML_{EST} = 4.91, Degree of freedom = 13, P=0,977).

a) Abiotic only

Degree days · Lake depth · Lake area · Total P · PH

Predator PCA1 · Predator richness · Li5 oral PCA2 · Li5 oral richness · Pelagic richness · Pelagic PCA2 · Small prey richness

b) Abiotic +Top-down

Degree days · Lake depth · Lake area · Total P · PH

Predator PCA1 · Predator richness · Li5 oral PCA2 · Li5 oral richness · Pelagic richness · Pelagic PCA2 · Small prey richness

c) Abiotic + bottom-up

Degree days · Lake depth · Lake area · Total P · PH

Predator PCA1 · Predator richness · Li5 oral richness · Li5 oral PCA2 · Pelagic richness · Pelagic PCA2 · Small prey richness

d) Integrative model

Degree days · Lake depth · Lake area · Total P · PH

Predator PCA1 · Predator richness · Li5 oral richness · Li5 oral PCA2 · Pelagic PCA2 · Pelagic richness · Small prey richness

Biotic association
Abiotic factor
Positive effect
Negative effect

213

214

215

216

217

218

219

220

221

222

223 **Supplementary Note 2 - Individual fish species boosted regression tree (BRT) models**

224   We developed BRT models[8] to assess the relative importance of 22 biotic and abiotic predictor

225 variables for explaining the variation in fish species presence in lakes for 14 of the fish species, i.e.

226 brown bullhead, pumpkin seed, walleye, lake whitefish, pike, cisco, yellow perch, rock bass, white

227 sucker, large-mouth bass, lake trout, small-mouth bass, burbot, and brook trout. In the model

228 developed for each species, we included presence data for the other 13 species to look for possible

229 species relationships, as well as a biotic predictor of the species richness of small prey and seven

230 abiotic characteristics of the lakes, i.e. degree days, pH, phosphorous levels, minimum depth of water,

231 maximum depth of water, area of the lake, and Secchi depth.

232   All models were fitted in R statistical computing program V. 3.2.1 using the gbm package

233 version 2.1.1[9]. BRT models are simple classification or rule-based models that partition observations

234 into groups based on similar values in the response variables using binary splits that are based on

235 predictor variables[10]. The boosting algorthim then iteratively develops a final model using a forward

236 stage-wise approach that progressively adds trees to the model. BRT models are advantageous for

237 analyzing our data because they can be used to assess different types of predictor variables in the same

238 model, which fits our questions well[11]. We used a tree complexity rate of two (which allows for

239 possible two way interactions) and a learning rate of 0.05 to allow for a suitable predictive

240 performance[10]. Because response variables were based on presence and absence data from surveys

241 conducted within each of the lakes we used a Bernoulli distribution. We show the relative influence of

242 each of the 22 possible predictors of variation in the presence of the 14 fish species in Suppl. Table 3.

243 We also show the number of trees included in the final model and estimates of the correlation between

244 the observed and predicted response variables. Variation in the presence of the different fish species

245 predicted by each BRT model is shown in Suppl. Figure 5, where values are predicted for each

246   variable by keeping all other variables at mean values.

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268 Supplementary Figure 5: The BRT tests if the occurrence of a species was associated with significant

269 negative changes in the occurrence of the target species, versus all abiotic factors used in our analyses.

270 Only two target species – brook trout (A) and lake trout (B) showed negative relationships versus the

271 occurrence of other species, indicated by the red boxes. These data are the same as shown

272 in Suppl. Fig. 3. Negative relationships are indicated by declining slopes, with the x-axis ranging from

273 0 (absent) the 1 (present). For the remaining 14 species tested with the BRT (Suppl. Fig. 3), all major

274 drivers of occurrence were abiotic.
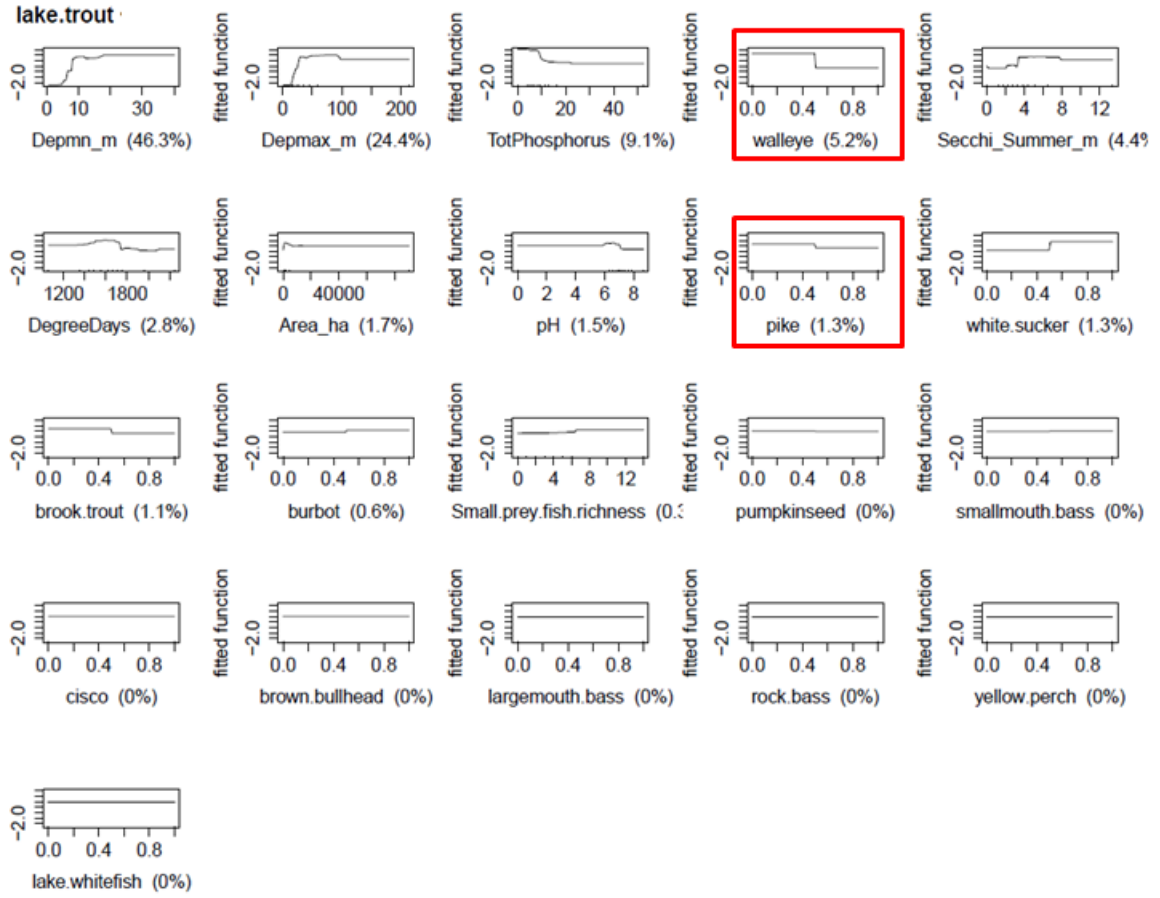
275

276 (A)

277
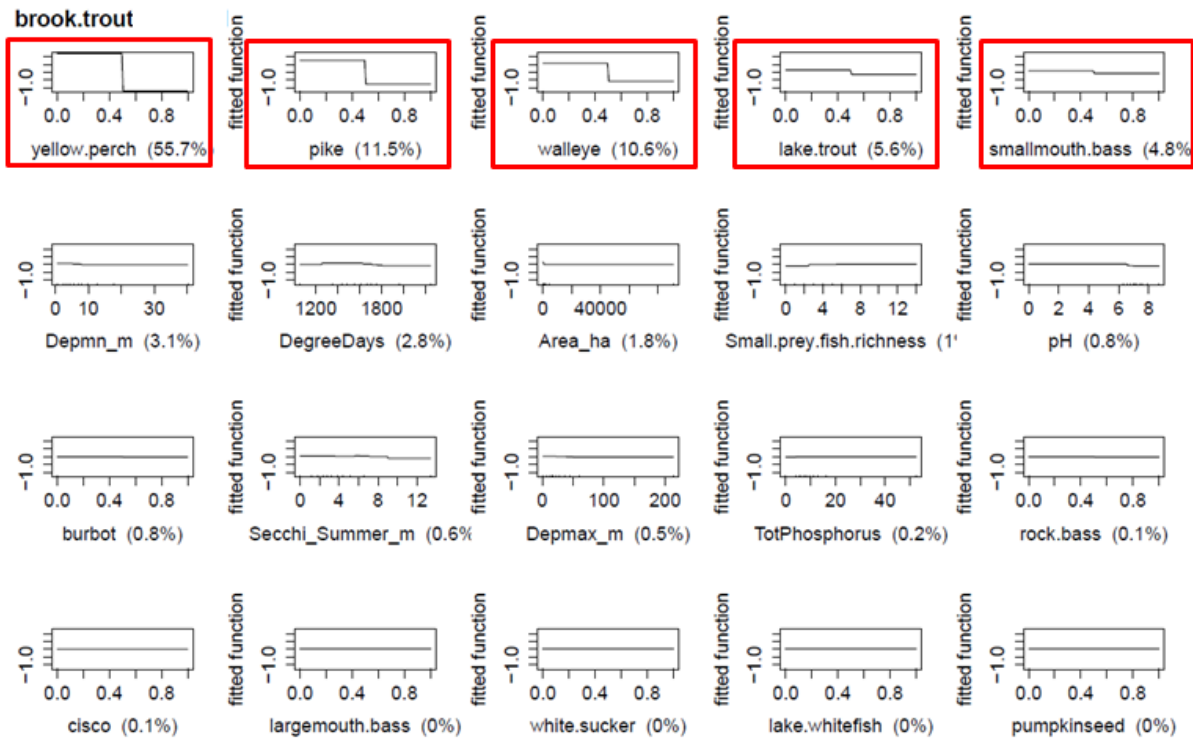
278

279

280 (B)

281

282

283

284

285

286

287

288

289

290

291

292 **Supplementary Note 3 - Multivariate analyses of environmental conditions for species pair**

293   **combinations**

294   For all species pairs, we calculated the Euclidian distance between three lake combinations

295 (Species A only, Species B only, both) based on seven of the measured abiotic/biogeographic factors

296 that were minimally inter-correlated. These factors are lake area, maximum depth of the lake, summer

297 Secchi depth, conductivity, pH, total phosphorus, and degree days. We eliminated lakes for which one

298 or more of these factors were not measured, leaving 649 of 721 lakes. We log transformed and

299 normalized all variables before the analysis.

300   To determine the overall degree of overlap in environmental space between lakes with species

301 A only versus lakes with species B only, we used an ANOSIM analysis[12]. This analysis tests the null

302 hypothesis that there are no differences in terms of multivariate environmental conditions, and is

303 equivalent to a non-parametric MANOVA, with permutation to test for significance[12]. The test

304 computes the R statistic, which ranges from 0 (complete overlap in environmental space) to 1

305 (complete separation in environmental space). As recommended by Clarke et al. (2014)[12], we

306 considered the pairwise test between lakes with species A only versus lakes with species B only after

307 confirming that the global ANOSIM test (considering all lake types: species A only, species B only,

308 both A and B, and neither) was significant. This was true in all cases. In all cases we used 5,000

309 random permutations to test for significant differences.

310   If a difference in multivariate environmental conditions was found via the ANOSIM analysis

311 above, it could be due to a shift in location (multivariate average of environmental conditions) and/or a

312 difference in dispersion (multivariate variation or range of environmental conditions). To test for the

313 latter possibility, we determined for each species pair whether lakes with both species co-occurring

314 had a reduced range of environmental conditions compared to lakes with species A only or species B

315   only. This could happen, for example, if a subordinate competitor or prey species can only co-occur

316   with a dominant competitor or predator in deep and large lakes, with those lake conditions allowing

317   the subordinate more possibilities to persist. Relatedly, we would predict that the subordinate could

318   occur in smaller and shallower lakes, if the dominant species is absent from those lakes. To do this, we

319   used a test of difference in multivariate dispersions (PERMDISP[13]). This is a multivariate extension of

320   Levene's test, measuring the average distance from each lake to the centroid of the group of lakes in

321   the same category, and comparing this average distance between different categories[14]. If coexistence

322   of species A and B is limited to a smaller range of environmental conditions than lakes inhabited by

323   species A without species B, then we would expect to find a significant decline in multivariate

324   dispersion of lakes with both species compared to lakes with species A only. This test also uses

325   permutation to determine significance of the test. We again set the number of permutations to 5,000.

326   We note that for species pairs where the number of lakes containing both species was very small, our

327   power to detect differences in multivariate dispersion was likely limited. For both the ANOSIM and

328   PERMDISP analyses we used the program PRIMER[14]. These tests are also available in the 'vegan'

329   package in R, but R does not allow pairwise tests in the ANOSIM function.

330   To visualize the similarity among lakes in terms of their environmental conditions, we used

331   PCA to ordinate all lakes in environmental space. We used the 'vegan' package in R to create

332   ordination plots, with vectors to visualize the correlation of each environmental factor with the first

333   two PCA axes.

334

335   **Supplementary Note 4 – R code for SEM analysis**

336

```
337   #......................................#
338   #..........BSM project (SEM)..............#
339   #......................................#
```

```
340    #.........................................................................................................................
341    #
342    #... Collaborators: Andrew MacDougall et al#
343    #
344    #... Author of the script: Eric Harvey (http://ericharvey.weebly.com/)
345    #
346    #... Creation date: January 05 2016
347    #
348    #... Latest edits: November 07 2016
349    #
350    #.........................................................................................................................
351    #
352    #... Clear variables
353    rm(list=ls())
354    #... Directories paths
355    datapath = "~/Documents/Research/1.Projects/3.BSM_project/3.Analysis/"
356    #...Load needed packages
357    library(vegan)
358    library(ade4)
359    library(lavaan)
360    library(corrplot)
361    #... Load the data
362    setwd(datapath)
363    presence.dat = read.delim("BSM_presence.txt")
364    presence.dat = na.omit(presence.dat)
365    #... Identify species of interest for each functional group
366    predator.of.interest =
367    c("walleye","lake.trout","pike","smallmouth.bass","largemouth.bass","burbot")
368    bentho.pelagic.species = c("cisco","brook.trout","round.whitefish","lake.whitefish")
369    bentho.litoral.species =
370    c("shorthead.redhorse","silver.redhorse","goldeye","longnose.sucker","white.sucker","br
371    own.bullhead","channel.catfish","yellow.bullhead","black.crappie","bluegill","pumpkinsee
372    d","rock.bass","yellow.perch","mooneye","fallfish")
373    small.prey.species = c("blackchin.shiner","blacknose.shiner","bluntnose.minnow",
374    "brook.stickleback","central.mudminnow","common.shiner",
375    "creek.chub","deepwater.sculpin","eastern.blacknose.dace",
376    "emerald.shiner","fathead.minnow","finescale.dace",
377    "golden.shiner","Iowa.darter","johnny.darter","lake.chub",
378    "logperch","longnose.dace","mimic.shiner","Mottled.Sculpin","ninespine.stickleback",
379    "northern.redbelly.dace","Pearl.dace","sculpin.spp","Slimy.Sculpin",
380    "spoonhead.sculpin","spottail.shiner","trout.perch")
381    #... Generate presence/absence matrix with species of interest
382    pred.comp.mat = presence.dat[,predator.of.interest]
383    pelagic.comp.mat = presence.dat[,bentho.pelagic.species]
384    litoral.comp.mat = presence.dat[,bentho.litoral.species]
385    small.prey.comp.mat = presence.dat[,small.prey.species]
386    #############################################################################
387    ###################################
```

```
388   #...Extract variables for SEM
389   #.....Ordinations analyses for compositional shifts
390   #.......Change in species composition with PCA
391   pred.PCA = rda(decostand(pred.comp.mat,"hell"))
392   pelagic.PCA = rda(decostand(pelagic.comp.mat,"hell"))
393   litoral.PCA = rda(decostand(litoral.comp.mat,"hell"))
394   #..........PCA results
395   pred.PCA
396   cumsum(pred.PCA$CA$eig/sum(pred.PCA$CA$eig)*100)
397   pelagic.PCA
398   cumsum(pelagic.PCA$CA$eig/sum(pelagic.PCA$CA$eig)*100)
399   litoral.PCA
400   cumsum(litoral.PCA$CA$eig/sum(litoral.PCA$CA$eig)*100)
401   #..........PCA diagnosis plots
402   source ('http://www.davidzeleny.net/anadat-r/doku.php/en:numecolr:cleanplot.pca?
403   do=export_code&codeblock=0')
404   cleanplot.pca(pred.PCA)
405   cleanplot.pca(pelagic.PCA)
406   cleanplot.pca(litoral.PCA)
407   dev.off()
408   #define "evplot" function first:
409   source ("http://www.davidzeleny.net/anadat-r/doku.php/en:numecolr:evplot?
410   do=export_code&codeblock=0")
411   # select the data frame with eigenvalues of particular axes:
412   ev.pred <- pred.PCA$CA$eig
413   ev.pelagic <- pelagic.PCA$CA$eig
414   ev.litoral <- litoral.PCA$CA$eig
415   # calculate axis-importance and draw the barplots:
416   evplot (ev.pred)
417   evplot (ev.pelagic)
418   evplot (ev.litoral)
419   dev.off()
420   #.....Create variables for SEM
421   #.......Axes representing changes in species composition
422   pred.PCA1 = scores(pred.PCA)$sites[,1]
423   pred.PCA2 = scores(pred.PCA)$sites[,2]
424   pelagic.PCA1 = scores(pelagic.PCA)$sites[,1]
425   pelagic.PCA2 = scores(pelagic.PCA)$sites[,2]
426   litoral.PCA1 = scores(litoral.PCA,choices=c(1:4))$sites[,1]
427   litoral.PCA2 = scores(litoral.PCA,choices=c(1:4))$sites[,2]
428   #.......Species richness per functional group
429   pred.richness = specnumber(pred.comp.mat)
430   small.prey.richness = specnumber(small.prey.comp.mat)
431   pelagic.richness = specnumber(pelagic.comp.mat)
432   litoral.richness = specnumber(litoral.comp.mat)
433   #.......Environmental variables
434   #..........Geo
435   Latitude = presence.dat$Latitude_DD
```

```
436    Longitude = presence.dat$Longitude_DD
437    DD = presence.dat$DegreeDays
438    #..........Morphology
439    Area = presence.dat$Area_ha
440    Perimeter = presence.dat$Perimeter_km
441    Depth = presence.dat$Depmn_m
442    Depth.max = presence.dat$Depmax_m
443    #..........Productivity
444    Secchi = presence.dat$Secchi_Summer_m
445    Conductivity = presence.dat$Conductivity..uS.cm.s.
446    TDSolids = presence.dat$TDSolids_mgL
447    TotPhosphorus = presence.dat$TotPhosphorus
448    PH = presence.dat$pH
449    ###########################################################################
450    ###################################
451    #...Correlations among variables
452    #Among biotic factors
453    biotic.cor =
454    cor(cbind(litoral.PCA1,litoral.PCA2,pelagic.PCA1,pelagic.PCA2,pred.PCA1,pred.PCA2,
455    pred.richness,litoral.richness,pelagic.richness,small.prey.richness))
456    corrplot(biotic.cor,method="number",type="full")
457    #Among environmental variables
458    env.cor =
459    cor(na.omit(cbind(Latitude,Longitude,DD,Area,Perimeter,Depth,Secchi,Conductivity,TD
460    Solids,TotPhosphorus,PH)))
461    corrplot(env.cor,method="number",type="full")
462    ###########################################################################
463    #####################################
464    #...Data frame for SEM ananalysis
465    #........variables are log transformed if necessary
466    #........variables need to be all at the same variance scale - to achieve this some are
467    divided/multiplied by a factor of 10 or 100
468    data.mod = data.frame(DD/100,Latitude,
469    Area/1000,Perimeter/100,Depth,
470    Secchi,TotPhosphorus,Conductivity/10,PH,TDSolids/10,
471    as.numeric(pred.richness),as.numeric(small.prey.richness),pelagic.richness,litoral.richne
472    ss,
473    pred.PCA1*100,pred.PCA2*100,pelagic.PCA1*100,pelagic.PCA2*100,litoral.PCA1*100,
474    litoral.PCA2*100)
475    colnames(data.mod) = c("DD","Latitude",
476    "Area","Perimeter","Depth",
477    "Secchi","TotPhosphorus","Conductivity","PH","TDSolids",
478    "pred.richness","small.prey.richness","pelagic.richness","litoral.richness",
479    "pred.PCA1","pred.PCA2","pelagic.PCA1","pelagic.PCA2","litoral.PCA1","litoral.PCA2")
480    #........Correlation matrix of the data
481    round(cov2cor(cov(data.mod)),3)
482    ###########################################################################
483    ##################################
```

```
484   #...Structural equation models
485   #Each model is built by starting from a simplified model from which paths are
486   sequentially added based on the modification indices until no modification indices are
487   >2.
488   #The initial simplified model for each alternative model is built as follow:
489   #Model 1 - Abiotic factors only
490   #Model 2 - Top-down regulation: Initial model was only top-down interactions and then
491   abiotic factors are added sequentially
492   #Model 3 - Bottom-up regulation: Initial model was only bottom-up interactions and then
493   abiotic factors are added sequentially
494   #Model 4 - Integrative model (main model from the paper): Abiotic + Top-down +
495   Bottom-up (based on our intial meta-model - see Supp.Mat)
496   #If the modification indice proposed to add a path that is forbidden according to the
497   model (e.g., between trophic groups in the abiotic model), the proposition is ignored and
498   the next best proposition is implemented.
499   #If the modification indice value of adding a path between an abiotic factor and a
500   functional group is the same as the value of adding a path between said trophic group
501   and another trophic group, then it was interpreted as a signal that the effect of said
502   abiotic factor might be mediated by a trophic interaction
503   library(semTools)
504   #.......Model 1 --> Model with only abiotic factors
505   BSM.mod1 <- "
506   # Regression
507   small.prey.richness ~ DD + PH + Area + TotPhosphorus + Depth
508   pelagic.richness ~ DD + PH + Depth + Area + TotPhosphorus
509   pelagic.PCA2 ~ DD + TotPhosphorus
510   litoral.richness ~ DD + Area + TotPhosphorus + PH + Depth
511   litoral.PCA2 ~ DD + Area + PH + TotPhosphorus
512   pred.richness ~ PH + Depth + Area + DD
513   pred.PCA1 ~ DD + PH + TotPhosphorus + Depth
514   "
515   BSM.mod1.fit = sem(BSM.mod1,data=data.mod) #run the SEM
516   BSM.mod1.fit #visualize fitting results
517   summary(BSM.mod1.fit,stand=T,rsq=T) #visualize path coefficients
518   modindices(BSM.mod1.fit) #modification indices
519   # To inspect model estimates and parameters
520   varTable(BSM.mod1.fit) #Variance table
521   inspect(BSM.mod1.fit,"cov.ov") #co-variance matrix
522   inspect(BSM.mod1.fit,"est") #regression estimates
523   #
524   #.......Model 2 Top-down regulation --> paths only from higher to lower trophic groups +
525   abiotic factors
526   #Notice that Depth - a factor that was very important in Model 1 is now mainly important
527   for top-predators - it suggests that the effect of depth on other groups is mediated by
528   top-predators
529   BSM.mod2 <- "
530   # Regression
531   small.prey.richness ~ pred.PCA1 + pred.richness + pelagic.richness + pelagic.PCA2 +
```

```
532    litoral.richness + PH + Area + TotPhosphorus
533    pelagic.richness ~ pred.richness + DD + Depth + PH + Area + TotPhosphorus +
534    pred.PCA1
535    pelagic.PCA2 ~ pred.richness + litoral.richness + DD
536    litoral.richness ~ pred.richness + pred.PCA1 + DD + TotPhosphorus + Area + PH
537    litoral.PCA2 ~ pred.richness + pred.PCA1 + DD + Area + pelagic.PCA2
538    pred.richness ~ PH + Depth + Area + DD
539    pred.PCA1 ~ DD + PH + TotPhosphorus + Depth
540    # Residual correlations
541    pred.richness ~~ pred.PCA1
542    pelagic.richness ~~ pelagic.PCA2
543    litoral.richness ~~ litoral.PCA2
544    "
545    BSM.mod2.fit = sem(BSM.mod2,data=data.mod)
546    BSM.mod2.fit
547    summary(BSM.mod2.fit,stand=T,rsq=T)
548    modindices(BSM.mod2.fit)
549    # To inspect model estimates and parameters
550    varTable(BSM.mod2.fit) #Variance table
551    inspect(BSM.mod2.fit,"cov.ov") #co-variance matrix
552    inspect(BSM.mod2.fit,"est") #regression estimates
553    #
554    #.......Model 3 Bottom-up regulation --> Ppaths only from lower to higher trohpic groups+
555    abiotic factors
556    BSM.mod3 <- "
557    # Regression
558    small.prey.richness ~ PH + Depth + Area + TotPhosphorus
559    pelagic.richness ~ small.prey.richness + DD + Depth + PH + Area + litoral.PCA2
560    pelagic.PCA2 ~ small.prey.richness + litoral.richness + Area + DD + TotPhosphorus
561    litoral.richness ~ small.prey.richness + DD + TotPhosphorus + Area + PH
562    litoral.PCA2 ~ small.prey.richness + DD + PH
563    pred.richness ~ small.prey.richness + pelagic.richness + pelagic.PCA2 + litoral.richness
564    + litoral.PCA2 + Depth + DD + Area + TotPhosphorus
565    pred.PCA1 ~ small.prey.richness + pelagic.richness + pelagic.PCA2 + litoral.richness +
566    litoral.PCA2 + Depth + DD + TotPhosphorus + PH
567    # Residual correlations
568    pred.richness ~~ pred.PCA1
569    pelagic.richness ~~ pelagic.PCA2
570    litoral.richness ~~ litoral.PCA2
571    pelagic.richness ~~ litoral.richness
572    "
573    BSM.mod3.fit = sem(BSM.mod3,data=data.mod)
574    BSM.mod3.fit
575    summary(BSM.mod3.fit,stand=T,rsq=T)
576    modindices(BSM.mod3.fit)
577    # To inspect model estimates and parameters
578    varTable(BSM.mod3.fit) #Variance table
579    inspect(BSM.mod3.fit,"cov.ov") #co-variance matrix
```

```
580    inspect(BSM.mod3.fit,"est") #regression estimates
581    #
582    #.......Model 4 Integrative model (main model from our meta-model) --> Abiotic factors +
583    top-down + bottom-up
584    BSM.mod4 <- "
585    # Regression
586    small.prey.richness ~ PH + Area + TotPhosphorus + pred.PCA1 + pelagic.richness +
587    pelagic.PCA2
588    pelagic.richness ~ DD + PH + Depth + Area + TotPhosphorus + pred.richness
589    pelagic.PCA2 ~ DD + TotPhosphorus + pred.richness
590    litoral.richness ~ DD + Area + TotPhosphorus + PH + pred.richness
591    litoral.PCA2 ~ DD + Area + PH + TotPhosphorus
592    pred.richness ~ PH + Depth + litoral.richness + litoral.PCA2 + pelagic.richness +
593    pelagic.PCA2 + TotPhosphorus + small.prey.richness + Area
594    pred.PCA1 ~ DD + PH + TotPhosphorus + Depth + litoral.richness + litoral.PCA2
595    # Residual correlations
596    litoral.richness ~~ litoral.PCA2
597    pelagic.richness ~~ pelagic.PCA2
598    pred.richness ~~ pred.PCA1
599    litoral.richness ~~ small.prey.richness
600    "
601    BSM.mod4.fit = sem(BSM.mod4,data=data.mod)
602    BSM.mod4.fit
603    summary(BSM.mod4.fit,stand=T,rsq=T)
604    modindices(BSM.mod4.fit) #modification indices
605    resid(BSM.mod4.fit,type="raw")
606    # To inspect model estimates and parameters
607    varTable(BSM.mod4.fit) #Variance table
608    inspect(BSM.mod4.fit,"cov.ov") #co-variance matrix
609    inspect(BSM.mod4.fit,"est") #regression estimates
610    #
611    ########################
612    #Model goodness of fit and comparaison
613    #######
614    #AIC comparaisons
615    library(AICcmodavg)
616    aictab(list(BSM.mod1.fit,BSM.mod2.fit,BSM.mod3.fit,BSM.mod4.fit),c("Abiotic","TD","BU
617    ","Integrative"))
618    # Model selection based on AICc:
619    #
620    # K AICc Delta_AICc AICcWt Cum.Wt LL
621    # TD 48 40119.91 0.00 0.61 0.61 -20008.03
622    # Integrative 50 40120.85 0.94 0.38 1.00 -20006.15
623    # BU 52 40130.71 10.80 0.00 1.00 -20008.72
624    # Abiotic 57 40135.64 15.73 0.00 1.00 -20005.22
625    #After removing 'Predator PCA1 --> 'Small prey richness' from the 'Top-down' model
626    # Model selection based on AICc:
627    #
```

628 # K AICc Delta_AICc AICcWt Cum.Wt LL
629 # Integrative 50 40120.85 0.00 0.99 0.99 -20006.15
630 # BU 52 40130.71 9.86 0.01 1.00 -20008.72
631 # Abiotic 57 40135.64 14.79 0.00 1.00 -20005.22
632 # TD 47 40157.95 37.10 0.00 1.00 -20028.22
633 #After removing 'Predator richness' --> 'Pelagic richness' from the 'Top-down model'
634 # K AICc Delta_AICc AICcWt Cum.Wt LL
635 # Integrative 50 40120.85 0.00 0.99 0.99 -20006.15
636 # BU 52 40130.71 9.86 0.01 1.00 -20008.72
637 # Abiotic 57 40135.64 14.79 0.00 1.00 -20005.22
638 # TD 47 40150.85 30.00 0.00 1.00 -20024.66
639 #Observations on AIC comparisons
640 # 1. Abiotic only is last
641 # 2. The difference between TD and ALL is not substantial
642 # 3. The difference between TD/ALL and BU/Abiotic is substantial
643 # 4. It is important to note that no interactions in the TD model does not indicate
644 of actual top-down interactions (none of them are negative!) except for one strong topdown
645 interaction between top-predators and small prey richness
646 # 5. In the context of observation #4 - It seems that the only reason why BU does not
647 perform well is because it is lacking the very important top-down effect of top-predators
648 on small prey richness
649 # 6. Observation #5 can be verified easily by removing the top-down path from toppredators
650 to small prey richness in the top-down model ('Predator PCA1' --> 'Small
651 prey richnes':ML increases by 40 points and the P-value = 0.0000, 'Predator richness' --
652 > 'Pelagic richness': ML increase 33 points) - therefore the good fit of the dop-down
653 model is driven by this keystone interaction in the model
654 #Conclusion:
655 # 1. A mixture of all variables is essential to understand the system. Both bottom-up
656 and top-down effects are very important, and in many case they seem to mediate the
657 effect of abiotic factors (e.g., Lake depth).
658 # 2. Thus, Model 4 is confidently the best model to understand the system because it
659 integrates all information and the mediated effects in the most parsimonious way
660 # 3. Model 4 could be analyzed further to gain better insights about the system.
661 #END
662 #####

663

664 **Supplementary References**

665

666   1. Holm, E., Mandrak, N. E., & Burridge, M. E. (2010). Freshwater fishes of Ontario. *Royal*

667   *Ontario Museum, Toronto*.

668   2.   Hasnain, S. S., Shuter, B. J., & Minns, C. K. (2013). Phylogeny influences the relationships

669   linking key ecological thermal metrics for North American freshwater fish species. *Canadian*

670   *Journal of Fisheries and Aquatic Sciences*, *70*, 964-972.

671   3.   Legendre, P., and de Cácares, M. 2013. Beta diversity as the variance of community data:

672   dissimilarity coefficients and partitioning. Ecology Letters, 16: 951-963.

673   4.   Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. Journal of statistical

674   software 48:1–36

675   5.   Jackson, D. A., Peres-Neto, P. R., & Olden, J. D. (2001). What controls who is where in

676   freshwater fish communities the roles of biotic, abiotic, and spatial factors. *Canadian journal*

677   *of fisheries and aquatic sciences*, *58*(1), 157-170.

678   6.   Grace, J. B., D. R. Schoolmaster, G. R. Guntenspergen, A. M. Little, B. R. Mitchell, K. M.

679   Miller, and E. W. Schweiger. 2012. Guidelines for a graph-theoretic implementation of

680   structural equation modeling. Ecosphere 3:art73.

681   7.   Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A

682   Practical Information-Theoretic Approach. Springer Science & Business Media.

683   8.   Friedman, J.H. 2001. Greedy function approximation: the gradient boosting machine. Annal

684   Stat. 29:1189-1232.

685   9.   R Development Core Team (2013). R: language and environment for statistical computing. R.

686   F. f. S. Computing. Vienna, Austrria, http://www.R-project.org/.

687   10. Ridgeway 2015 Package GBM. CRAN http://code.google.com/p/gradientboostedmodels/

688   11. Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species.

689   *Methods in ecology and evolution*, *1*(4), 330-342.

690   12. Clarke, K.R., Gorley, R.N., Somerfield, P.J., Warwich, R.M. (2014) Change in marine

691   communities: an approach to statistical analysis and interpretation, 3rd edition. PRIMER-E:

692   Plymouth, UK.

693   13. Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure

694   of beta diversity. Ecology Letters, 9, 683–693.

695   14. Anderson, M.J., Gorley, R.N. & Clarke, K.R. (2008) PERMANOVA+ for PRIMER: Guide to

696   Software and Statistical Methods. PRIMER-E, Plymouth, UK.

697