UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ELEMENTS OF THE METACOMMUNITY STRUCTURE: COMPARISON ACROSS MULTIPLE METACOMMUNITIES

THESIS
PRESENTED
AS A PARTIAL REQUIREMENT
FOR THE MASTER IN BIOLOGY

BY
RENATO HENRIQUES DA SILVA

MARCH 2011

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉLÉMENTS DE LA STRUCTURE D'UNE METACOMMUNAUTÉ: COMPARAISON À TRAVERS MULTIPLE METACOMMUMAUTÉS

MÉMOIRE
PRESENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN BIOLOGIE

PAR
RENATO HENRIQUES DA SILVA

MARS 2011

ACKNOWLEDGEMENTS

I first want to thank my research advisor, Pedro R. Peres-Neto, Ph. D., professor at Université du Québec à Montréal (UQAM), for the opportunity he provided me, as well as for all his patience, advices and trust in me throughout the masters program. His training and methods were fundamental for my achievements in academia and in life. I would also like to thank Sapna Sharma, Ph. D., and Tariq Gardezi, Ph. D., for providing the data that allowed the feasibility of the project and Mélanie Desrochers for her help with GIS. A thank you to the members of my committee, Beatrix Beisner, Ph. D., and Christian Messier, Ph. D., for their valuable advices and critiques on the present work. I want to specially thank Bailey Jacobson, for all her help, in many ways, in my project, and for her friendship, as well as the other students for all their support: Rich Vogt, Daniel Pires, Caroline Senay, Mehdi Layeghifard, Joachim Prunier, Marie Christine Bellemare and Aline Aguiar. Thank you to all the UQAM employees for their help. I would like to thank my father, Durval Henriques da Silva Filho and the rest of my family for their support, which was crucial to my success. A special thanks goes to my wife Monique Ariane Rezende Henriques da Silva for her patience and love. Finally, I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), which provided financial support for this project.

TABLE OF CONTENTS

CHAPTER I

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The Elements of Metacommunity Structure (EMS) analytical tool is a powerful platform for evaluating patterns of species distribution in geographic or environmental space; however, it is still underused in ecological studies. The objective of this study is to evaluate the structuring mechanisms of boreal lake-fish species distributional patterns at multiple scales by applying the EMS technique on the Ontario Fish Distribution Database (OFDD), a large database that contains presence-absence records of fish species and the geographic position for more than 9000 lakes from Ontario. The environmental information for each lake was assessed in the Lake Inventory Database (LINV) and spatial indices, such as lake connectivity and distance from postglacial refuges, were created from lakes geographic position. Moreover, the phylogenetic relatedness of species as well as their ß-niches were calculated in order to assess the role of species in community assembly and how they affect metacommunity patterns.

In chapter one, the EMS indicated that nestedness and Clementsian gradients are the most common distribution patterns among watersheds and that the main difference between them is species turnover (e.g. change in species composition across space). Most nestedness metacommunities are located in low-energy watersheds, containing larger lakes at higher latitudes whereas Clementsian gradients metacommunities are mostly found in opposite conditions. At the watershed scale, environmental variables explained, in average, 9.1% of the variation in species distribution from both patterns whereas spatial variables accounted for less than 3.5%. At the province scale, the variation in species distribution was best accounted by spatially structured environment (29.26%), followed by pure environmental predictors (10.80%). Statistical tests showed a gradient of low to high species turnover from North to South, influenced mainly by latitude and correlated environmental variables (e.g., temperature).

In chapter two, results indicated that, at the watershed scale, phylogenetic underdispersion is the dominant pattern whereas at the lake scale phylogenetic overdispersion has a stronger signal. Community phylogenetic and niche structure are mainly influenced by lake size, energy-related variables (growing degree-days, temperature, potential evapotranspiration) and latitude. In northern regions, there is higher niche overlap and greater phylogenetic distance between constituents present in the same communities, whereas in southern watersheds, communities are composed of species more closely related but with low niche overlap.

Keywords: EMS, correspondence analysis, Clementsian gradients, species distribution, nestedness, species turnover, phylogenetic structure, niche, environmental gradient

RÉSUMÉ

Les « Éléments de la Structure des Metacommunautés » (EMS) est un outil analytique puissant pour l'évaluation des patrons de distributions d'espèces dans l'espace géographique ou environnementale ; par contre, cette technique est encore sous-utilisée parmi les études écologiques. L'objectif de cette étude est d'évaluer les mécanismes structurants les patrons de distributions d'espèces de poissons de lacs boréaux à des multiples échelles en appliquant la technique EMS sur la *Ontario Fish Distribution Database*, une base de données contenant des informations sur la présence-absence des espèces de poissons de plus de 9000 lacs de l'Ontario ainsi que leurs positions géographiques. Pour chaque lac, l'information sur les variables environnementales on été obtenue grâce au *Lake Inventory Database* (LINV) et des indices spatiaux, comme la connectivité entre les lacs et leur distance aux refuges postglaciaires, ont été calculés à partir d'informations géographiques. Puis, la relation phylogénétique des espèces et leurs niches ß on été estimés pour comprendre le rôle des espèces dans l'assemblage des communautés et formation des metacommunautés.

Dans le premier chapitre, la technique EMS a indiqué que *nestedness* et *Clementsian gradients* sont les patrons de distributions les plus courants parmi les bassins versants. La pluparts des patrons *nestedness* se situent dans des bassins de faible énergie contenant des grands lacs et localisés dans de hautes latitudes tandis que les patrons *Clementsian gradients* sont rencontrés dans des conditions opposés. À l'échelle des bassins, les variables environnementales expliquent en moyenne 9.1% de la variation dans la distribution des espèces pour les deux type de patrons contre moins de 3.5% pour les variables spatiales. À l'échelle provinciale, la variation dans la distribution des espèces est expliquée principalement par les variables environnementales structurées spatialement (29,26%) suivit des variables environnementales indépendantes de l'espace (10.80%). Des tests statistiques suggèrent que le taux de changement dans la composition des communautés, la caractéristique qui mieux distingue les deux patrons, augmente du nord vers le sud, influencé principalement par la latitude et les variables associées (e.g., température).

Dans le second chapitre, les résultats indiquent que, à l'échelle du bassin versant, la sous-dispersion phylogénétique prédomine tandis que la sur-dispersion phylogénétique est plus observée à l'échelle locale. La structure phylogénétique et de niche des communautés sont principalement influencés par la taille des lacs, les variables liées à l'énergie (e.g., température, degré-jour de croissance) et la latitude. Dans les régions du Nord, il y a des taux élevés de chevauchement des niches et de plus grande distance phylogénétique entre les espèces qui cohabitent alors que dans les bassins versants du Sud on rencontre le patron inverse.

Mots clefs : EMS, analyse de correspondance, *Clementsian gradients*, distribution d'espèces, *nestedness*, *species turnover*, structure phylogénétique, niche, gradient environnemental

INTRODUCTION

1.1 - State of knowledge

The processes that select species to assemble into local communities have been a core theme in Ecology as a science. Ecologists well accept that communities are opened entities, in the sense that they are subject to processes of immigration and emigration of organisms (i.e., communities are linked with each other by species dispersal). In this context, the metacommunity theory arose as a prominent framework to explain species distributions based on both local (i.e., abiotic factors and biotic interactions) and regional (i.e., climate and dispersal) factors. One way to study metacommunities is to focus on structural patterns, extracted from a species-by-site matrix. Many species distributional patterns have been described in the ecological literature (e.g., nestedness, checkerboards) and, more importantly, each of them have unique theoretical underpinnings and structuring mechanisms. However, most analytical methods are limited by contrasting only one specific pattern with expectations from a random model. Here, we use a metacommunity framework that can test for six different patterns of distribution simultaneously on a large temperate lake-fish database in order to understand the structuring mechanisms of their distribution. I defined the "metacommunity units" as the tertiary watersheds from Ontario and subjected each of them to the analytical framework originally developed by Leibold & Mikkelson (2002), termed Elements of Metacommunity Structure (EMS), which estimates the pattern that best fit the species distributions in a given metacommunity.

In the first chapter, the metacommunity patterns unveiled by EMS will be analyzed, and the relative importance of spatial and abiotic factors accounting for each pattern will be assessed. In the second chapter, the focus will shift to species properties where phylogenetic and niche relationships will be assessed. Then, this information will be related to the patterns identified by the EMS analysis on each metacommunity in order to determine the possible mechanisms structuring them, such as competitive exclusion *versus* habitat filtering. In both chapters, I also explore patterns across the entire Province to determine how ecological processes such as variation in species composition across space (i.e., species turnover; chapter one), community phylogenetic and niche structure (chapter two) relate to environmental variation. In the following paragraphs, I will introduce the subject of this work in the format of a literature review, outlining the general context in which the questions and issues addressed in this study are discussed. Moreover, the characteristics of the database and the ecological system, the metacommunity paradigm, the EMS technique and the insights acquired from applying phylogenetic approaches to community ecology (and EMS patterns) are presented. The final section is a general conclusion that presents the links between both chapters and the scientific contributions made by this study.

1.2 - Study System

In this study, a local community is defined as all fish species that inhabit a lake (which may potentially interact with each other) and a metacommunity as all lakes (which are potentially linked via fish dispersal through streams) within a watershed. Lake-fish systems have several features that make them a good ecological model to apply the metacommunity paradigm, niche modeling and analysis based on the phylogenetic structure of their communities. Lakes within a watershed can be viewed as "virtual islands" presenting discrete boundaries (Magnuson *et al.*, 1998) and varying in their degree of connectivity (Olden *et al.*, 2001). The connectivity of a lake from the stand-point of a fish is a function of several factors, such as distance from other lakes, presence of streams connections, flow direction of connecting streams, lake elevation and presence of predators in dispersal corridors (Olden *et al.*, 2001). The different degrees of connectivity among lakes and the constraints imposed by the environment within each lake will influence fish dispersal ability and patterns as well as their extinction vulnerability (Jackson *et al.*, 2001). For example, more isolated lakes have reduced colonization rates (Jackson *et al.*, 2001), thus fish populations going through local extinction (i.e., at the lake level) have lower probability to be rescued (Brown & Kodric-Brown, 1977). Other aspects influencing fish biodiversity, among many, are pH (Helmus *et al.*, 2007b), temperature (Shuter & Post, 1990) and lake size, the latter being a surrogate for habitat heterogeneity and highly correlated with fish richness (Eadie *et al.*, 1986).

In order to understand the factors structuring species distribution (Leibold & Mikkelson, 2002; Leibold *et al.*, 2004), the Ontario Fish Distribution Database (OFDD), a database, maintained by the Ontario Ministry of Natural Resources (OMNR) containing the presence-absence of 134 temperate lake-fish species distributed among approximately 9900 lakes was used. Species that are introduced, rare (present in less than 0.5% of the lakes) and hybrids, and lakes without species or geographic coordinates were removed, resulting in 53 extant native species and 8911 lakes divided in 85 tertiary watersheds (considered as metacommunities; see Figure 1.1). The average lake richness was 6.46 ± 3.86 fish species. For more information, see the methodology section of Chapter 1. The mean of each environmental variable for the 85 watersheds are presented in Table 1.1. Abbreviations used in Table 1.1 stands for the following: SA = surface area; P = shoreline perimeter; ISL = island shoreline perimeter; MaxD = max depth; MeanD = mean depth; ELEV = elevation; GDD = growing degree days; SD = secchi depth; TDS = total dissolved solids; MEI = morphoedaphic index; Crown = crown canopy cover; MADT = mean annual daily temperature; MJT = mean July temperature; MAT = mean August temperature; and PET = potential evapotranspiration. ISL refers to the size of the combined shoreline perimeter of all islands present in each lake, hence lakes with no islands will have ISL value of 0. MEI is calculated by dividing the total dissolved solids (TDS) present in a lake by its mean depth (MeanD). Growing degree-days (GDD) were estimated as follow:

$$\sum\left[\left(\left(MeanMonthlyTemperature > 5.5°C\right) - 5.5°C\right) \times 30\right]$$

Where the "30" represent the 30-year recording period of the dataset (Mandrak, 1995). Crown is measured as % of the lake shoreline covered by the canopy of trees (for additional information see Goodchilde & Gale, 1982; Mandrak & Crossman, 1992a).

1.3 - Community ecology and the metacommunity paradigm

In the last century, many approaches were proposed in order to explain the patterns of species distributions and the processes that regulate them. The classic view is that communities assemble according to niche-related processes, such as resource use and competition. This perspective became very popular among ecologists after Hutchinson's (1957) seminal paper unveiling the multidimensional niche (MacArthur & Levins, 1967; Diamond, 1975). Models based on this approach analyze how niche characteristics such as niche breadth (variability in resource use) and marginality (levels of specialization) are affected by the environment and/or biotic interactions (e.g., Mason *et al.*, 2008; Ingram & Shurin, 2009). A few years later, MacArthur & Wilson (1967) proposed the equilibrium theory of island biogeography (IBT), which states that species composition in insular habitats is dictated by differences in the area and isolation of sites, which in turn influence the probability of extinction and colonization of species, respectively (Brown & Kodric-Brown, 1977). Ricklefs (1987) reinforced the importance of this approach, but argued that local diversity was not solely dictated by local environment and competition, as it was also largely dependent on the regional pool of potential colonizers and their evolutionary histories. More recently, Hubbell (2001) developed the Unified Neutral Theory of biodiversity which posits that species differences are not relevant to community assembly and local community composition is dictated mainly by stochastic processes such as dispersal, ecological drift and speciation. Although this model has well fitted some natural systems, space and stochasticity alone cannot explain all the variation in species distribution. Thus recently, the metacommunity theory was developed, including both deterministic (i.e., niche) and spatial processes within the same framework (Leibold *et al.*, 2004).

A metacommunity is a set of local communities that are potentially linked by species dispersal at the regional scale (Leibold *et al.*, 2004), whereas a community is the collection of individuals of all species that potentially interact within a single patch (Holyoak *et al.*, 2005). In this perspective, the spatial distribution of communities and dispersal plays a critical role structuring the species diversity at regional and local scales, which may influence community assembly from what would be expected if only local biotic and abiotic aspects are analyzed (Leibold *et al.*, 2004).

Metacommunities are studied by means of two venues: mechanism (Holyoak *et al.*, 2005; Driscoll & Lindenmayer, 2009) and structure (Leibold & Mikkelson, 2002; Haudsorf & Hennig, 2007). The mechanistic approach seeks to explain species distribution through several spatially mediated models (e.g., species sorting, patch-dynamics, mass-effects and neutral) that have different assumptions about the roles of environment, dispersal rates and stochastic vs. deterministic processes (see Leibold *et al.*, 2004 for a complete review). The structural approach, which is the focus of this work, evaluates species distributions along environmental gradients that result from specific mechanisms and manifest as particular patterns of metacommunity structure (Leibold & Mikkelson, 2002; Haudsorf & Hennig, 2007; Presley *et al.*, 2009). In this approach, metacommunity structure is determined by fitting different non-random patterns to an incidence matrix (i.e., site-by-species matrix). Several non-random patterns have been described (see Leibold & Mikkelson, 2002), with some being quite common in nature, such as nested subsets (Patterson & Atmar, 1986). Studies addressing patterns of species distributions (e.g., nestedness, Clementsian gradients, Gleasonian gradients, evenly-spaced gradients, checkerboards and random) can provide valuable clues about the factors that regulate ecological communities (Presley *et al.*, 2009), because the mechanisms and theory are unique to each pattern (Presley & Willig, 2010).

Recently, several analytical tools have been developed allowing researchers to identify and account for numerous aspects of metacommunity structure (e.g., Hoagland & Collins, 1997; Leibold & Mikkelson, 2002; Haudsorf & Hennig, 2007). In this study, I will use the framework developed by Leibold & Mikkelson (2002), termed Elements of Metacommunity Structure (EMS), which is described below.

1.4 - Elements of Metacommunity Structure (EMS)

The EMS technique focuses on determining which pattern best fits to the species distributions within a metacommunity. Prior to analysis, the incidence matrix (i.e., species by sites matrix) is ordered according to the primary axis extracted via correspondence analysis (Presley *et al.*, 2009), which is a common ordination procedure (Gauch *et al.*, 1977) to detect variation in species distributions that respond to latent processes such as environmental gradients (i.e., variation in unmeasured environmental variables). Correspondence analysis (CA) maximizes the proximity of sites with similar species compositions as well as species with similar distributions (Leibold & Mikkelson, 2002). Therefore, it makes a compromise between minimizing interruptions within species ranges and within community compositions. This reorganization of the data matrix creates a gradient that reflect the integration of multiple factors (abiotic and biotic) that may be important in dictating species distributions (Presley & Willig, 2010).

Table 1.1 Mean lake environmental variables for each watershed. Abbreviations as outlined on page 4. Watersheds are presented in Ontario map of Figure 1.1

Watershed	SA(hec)	P(km)	ISL(km)	MaxD(m)	MeanD(m)	ELEV(m)	GDD	SD(m)	TDS(mg/l)	MEI	CROWN(%)	MADT(°C)	MJT(°C)	MAT(°C)	PET(mm/yr)
1	662.71	26.12	2.58	25.94	9.48	361.89	1501.93	3.70	40.41	7.12	98.49	1.80	19.10	17.47	518.27
2	601.94	21.69	4.59	22.20	8.27	388.70	1493.41	3.70	37.38	8.72	93.57	1.91	18.98	17.15	530.60
3	1645.29	37.71	1.66	13.52	5.00	387.77	1410.69	2.46	43.37	11.56	99.46	0.98	18.38	16.52	509.13
4	766.85	26.07	4.34	16.06	5.22	417.82	1388.00	2.94	38.72	10.82	93.37	1.01	18.14	16.23	508.60
5	554.21	24.79	7.24	26.29	8.49	366.68	1504.57	4.33	36.82	5.96	94.15	2.30	19.23	17.50	538.21
6	406.97	19.41	3.09	21.32	7.65	359.32	1541.11	3.86	33.16	6.98	85.78	2.07	19.35	17.57	538.21
7	471.03	22.01	3.90	23.06	7.64	408.44	1443.86	3.63	29.38	6.07	95.89	1.88	18.46	16.75	525.71
8	1369.65	55.58	19.72	19.89	4.77	411.53	1249.73	2.82	24.65	6.97	96.14	-0.22	16.82	14.93	474.57
9	817.74	38.28	31.90	27.66	9.20	356.26	1502.02	3.72	23.49	3.21	99.70	1.23	18.77	17.24	514.91
10	1322.74	33.05	6.01	15.68	5.52	392.68	1443.05	2.63	35.07	10.03	93.10	1.05	18.55	16.94	510.18
11	2838.94	93.17	22.58	19.53	5.19	316.03	1190.93	2.19	40.50	16.28	96.33	-0.64	17.64	16.08	502.18
12	2057.58	71.58	22.31	18.70	5.74	355.89	1366.44	2.85	27.75	7.21	99.47	0.32	18.20	16.59	513.87
13	2384.35	51.78	11.44	12.13	3.04	330.95	1128.24	2.32	53.29	22.81	99.76	-1.16	17.18	15.45	483.33
14	2445.80	86.97	35.32	19.02	4.45	402.33	1288.09	3.08	31.93	9.73	98.37	0.01	17.64	15.69	502.63
15	1597.09	41.61	7.46	13.75	4.08	320.35	1245.58	3.53	78.68	27.56	99.45	-1.28	16.68	14.72	477.08
16	2133.41	79.29	16.28	11.08	2.64	298.31	1124.00	2.30	58.03	29.89	99.97	-1.48	17.00	15.11	473.35
17	1364.95	36.93	5.21	10.14	2.63	349.69	1214.00	2.42	62.51	32.72	99.57	-0.67	17.43	15.46	495.35
18	528.25	25.80	4.67	25.06	11.65	425.75	1423.82	3.64	24.25	4.81	95.53	1.80	17.39	16.09	514.33
19	967.90	32.17	4.50	13.43	3.07	303.88	1251.38	2.87	88.44	44.93	93.94	-0.54	16.43	14.61	468.45
20	1194.78	41.34	7.70	17.14	4.40	297.63	1251.53	3.02	70.41	22.01	99.94	-1.14	16.56	14.70	463.44
21	493.18	18.49	1.36	16.97	4.53	327.92	1271.34	2.94	97.82	36.79	98.72	0.18	16.15	14.75	481.15
22	150.34	9.45	0.90	15.75	4.47	327.82	1259.68	3.49	86.80	31.03	98.96	0.82	16.19	14.96	479.69
23	671.99	22.28	4.53	6.00	1.59	273.86	1261.14	2.39	55.05	38.52	100.00	-0.91	16.69	14.71	462.23
24	188.19	11.57	0.84	34.50	10.70	340.18	1313.98	4.37	53.63	7.89	97.92	0.94	14.88	14.65	472.09
25	577.49	22.33	4.95	24.35	6.61	364.51	1257.10	3.26	56.99	13.86	95.61	0.17	16.33	14.84	479.22
26	225.56	11.04	0.68	18.46	6.10	348.30	1371.86	3.35	75.96	22.98	83.18	1.23	16.50	15.43	501.00
27	160.67	7.58	0.47	14.92	4.54	308.15	1234.58	4.05	112.80	33.22	62.93	0.31	16.62	14.80	481.18

Table 1.1 (continuation)

Watershed	SA(hec)	P(km)	ISL(km)	MaxD(m)	MeanD(m)	ELEV(m)	GDD	SD(m)	TDS(mg/l)	MEI	CROWN(%)	MADT(°C)	MJT(°C)	MAT(°C)	PET(mm/yr)
28	481.52	15.78	4.43	15.63	5.22	445.91	1419.79	2.77	44.71	16.27	81.45	1.35	16.88	15.54	502.69
29	188.28	9.15	0.70	15.44	4.63	373.09	1248.06	3.53	74.07	23.61	98.77	0.80	15.81	14.61	476.58
30	280.08	7.47	0.54	15.33	6.01	456.13	1480.05	3.32	45.55	12.37	84.85	1.82	17.16	15.86	506.12
31	570.08	26.29	3.33	18.11	5.20	374.08	1316.63	3.10	45.88	13.97	96.04	1.00	16.16	14.87	483.51
32	260.48	14.25	2.07	22.69	6.65	374.60	1316.14	3.97	49.91	13.05	66.88	1.40	15.87	15.11	474.01
33	111.21	8.46	0.94	22.35	6.80	382.81	1378.13	4.89	39.03	8.55	80.83	2.42	16.73	15.99	509.37
34	81.39	6.58	0.51	8.86	2.41	303.49	1353.33	2.56	63.54	44.48	92.84	1.18	17.39	15.70	494.41
35	218.25	11.68	1.94	12.43	3.78	375.48	1371.99	3.53	64.73	27.78	97.45	1.73	17.34	15.79	500.62
36	223.09	12.32	1.76	17.20	4.97	417.67	1362.78	4.12	69.42	19.78	80.87	1.69	16.69	15.50	499.58
37	205.35	13.3	2.6	17.3	6.2	384.9	1545.0	4.4	28.7	6.4	88.4	3.3	18.2	16.6	526.39
38	273.33	17.23	2.64	20.80	6.30	445.40	1448.08	4.22	33.61	7.41	98.49	2.75	17.47	16.08	517.84
39	47.44	3.71	0.07	13.14	3.63	287.83	1303.10	2.95	79.67	34.58	91.28	0.49	16.76	15.12	485.47
40	200.75	8.00	0.46	16.35	4.41	270.23	1273.27	2.61	82.39	23.85	95.76	0.02	16.46	14.87	482.28
41	1586.30	10.05	5.43	13.09	4.55	297.17	1356.06	3.30	47.29	14.31	97.48	0.74	16.95	15.25	487.08
42	118.13	4.08	0.41	15.54	4.29	296.31	1372.22	3.70	77.75	26.50	80.10	1.09	17.34	15.68	495.25
43	1054.02	20.92	1.32	5.52	1.78	276.72	1271.80	1.52	36.32	34.40	100.00	-0.50	15.82	14.40	471.73
44	68.06	3.64	0.26	11.40	3.93	321.18	1429.67	3.98	68.43	24.85	88.75	1.34	17.56	15.87	506.17
45	94.78	7.48	0.75	20.61	6.39	386.08	1506.80	5.06	58.90	12.34	90.47	2.60	18.44	16.52	524.96
46	140.08	10.01	1.33	14.52	4.37	358.16	1440.08	3.68	56.71	23.82	95.36	1.78	17.90	16.14	511.87
47	149.32	11.97	0.82	18.81	6.71	312.27	1576.62	4.90	37.36	8.09	81.78	3.70	18.79	17.14	528.13
48	142.41	9.40	1.11	15.59	4.96	349.75	1494.38	4.30	44.52	12.95	89.71	2.26	18.25	16.54	524.46
49	736.93	17.38	4.60	12.69	4.25	288.76	1253.97	3.18	108.03	29.55	56.69	0.29	16.52	14.87	474.65
50	256.53	10.24	0.20	8.66	2.86	245.06	1269.03	2.07	70.09	37.08	67.66	0.17	16.68	14.96	473.18
51	34.22	2.74	0.04	13.44	3.93	179.26	1273.62	3.44	111.65	33.81	40.88	0.14	16.96	15.17	473.61
52	178.62	7.15	0.60	9.25	3.06	239.70	1297.43	2.76	99.43	39.05	98.11	0.27	16.81	15.26	475.94
53	83.32	3.73	0.10	10.54	3.40	226.86	1309.89	2.78	98.06	35.35	77.62	0.39	16.94	15.36	479.31
54	21.91	2.37	0.20	11.07	3.54	224.62	1293.52	3.13	107.19	47.69	85.71	0.10	16.80	15.25	485.33
55	89.46	4.31	0.06	9.70	3.77	258.48	1328.86	2.49	84.66	32.16	48.62	0.77	16.95	15.34	481.79
56	132.43	6.56	0.77	10.88	3.99	308.23	1484.56	2.83	51.09	20.79	59.30	1.52	17.58	15.95	510.75

Table 1.1 (continuation)

Watershed	SA(hec)	P(km)	ISL(km)	MaxD(m)	MeanD(m)	ELEV(m)	GDD	SD(m)	TDS(mg/l)	MEI	CROWN(%)	MADT(°C)	MJT(°C)	MAT(°C)	PET(mm/yr)
57	251.40	9.14	0.73	19.83	6.19	324.61	1593.29	4.21	45.02	12.51	87.92	3.45	18.44	17.00	532.03
58	312.32	14.49	2.74	21.53	6.57	308.37	1599.33	4.83	38.47	8.64	90.88	3.13	18.54	16.93	527.42
59	181.24	9.59	0.67	24.06	8.55	312.43	1636.28	5.32	39.86	8.15	87.37	4.31	18.60	17.27	530.19
60	121.24	6.41	0.23	22.80	8.02	387.75	1512.92	5.59	30.08	7.13	89.66	3.56	18.12	16.65	527.48
61	138.13	7.10	0.65	17.74	6.70	325.94	1514.57	4.07	29.43	7.40	65.77	3.65	17.44	16.60	524.00
62	151.31	11.41	1.30	16.27	5.12	277.09	1708.07	3.00	28.96	8.60	60.80	4.42	18.92	17.53	537.63
63	33.66	3.31	0.07	10.53	3.60	237.23	1745.64	2.71	28.33	11.43	82.08	3.85	18.70	17.33	542.03
64	71.15	5.91	0.56	16.66	5.21	400.23	1454.87	4.24	25.53	7.82	81.41	3.01	16.83	16.07	512.72
65	99.64	5.61	0.44	15.72	5.18	397.95	1704.14	3.89	56.07	16.36	74.05	4.07	18.38	16.95	542.73
66	231.33	11.41	1.32	19.32	6.31	348.93	1678.99	3.99	29.07	7.30	96.07	3.74	18.22	16.81	542.73
67	132.18	5.08	0.23	14.07	4.71	242.48	1828.00	3.66	65.44	20.34	50.79	4.40	18.99	17.59	546.35
68	181.78	8.60	1.54	16.32	5.75	346.58	1689.04	3.86	24.24	6.81	52.42	4.65	18.74	17.47	542.73
69	150.39	7.33	0.61	22.26	7.82	365.57	1714.45	4.99	42.41	9.00	26.93	4.47	18.55	17.26	543.45
70	141.18	8.94	1.11	14.20	5.02	297.54	1723.79	3.57	25.37	7.58	53.67	4.67	18.85	17.56	542.01
71	112.26	7.12	0.96	14.11	4.75	325.26	1805.65	3.89	93.80	31.79	38.73	5.07	18.99	17.78	545.13
72	297.82	12.14	2.11	15.03	4.76	306.25	1819.75	3.47	64.22	30.46	65.58	5.36	19.24	18.13	540.63
73	145.56	8.0	1.1	12.7	4.9	244.0	1886.2	2.7	104.6	35.5	46.9	5.4	19.4	18.3	548.83
74	132.72	6.53	0.80	11.18	4.06	277.44	1808.81	2.99	61.74	42.00	52.71	5.38	19.31	18.15	550.88
75	86.72	4.99	0.23	14.52	5.24	289.92	1867.06	4.17	114.40	43.17	73.35	4.77	19.40	18.01	543.54
76	136.78	7.37	0.87	13.11	4.45	247.58	1930.55	3.71	102.89	64.25	36.35	5.11	19.60	18.25	543.76
77	336.11	17.43	1.10	16.73	5.69	148.10	1999.46	3.50	133.04	115.41	2.86	5.94	20.32	19.00	557.92
78	223.22	14.19	2.96	19.79	7.28	130.11	1999.69	3.57	114.81	38.45	21.82	6.07	20.29	19.03	563.02
79	120.38	9.0	0.5	17.6	5.7	174.0	1990.4	3.6	108.1	45.6	33.7	5.8	19.8	18.6	550.89
80	62.60	2.56	0.37	4.76	1.95	328.00	2188.58	1.69	227.42	185.75	5.69	6.92	20.21	19.16	571.91
81	13.39	1.74	0.05	5.35	1.91	347.35	2119.91	2.35	237.26	223.76	0.00	6.61	19.93	18.93	565.68
82	9.29	1.47	0.04	8.87	2.68	256.17	2207.61	1.72	242.43	186.03	8.26	7.09	20.54	19.50	580.69
83	78.71	4.35	0.36	3.38	1.15	207.06	1782.63	2.05	195.54	331.69	11.84	5.87	19.05	18.12	546.06
84	11.05	1.70	0.06	5.50	2.20	340.03	1889.49	2.64	253.97	195.57	5.31	5.95	19.04	18.01	556.50
85	46.66	3.64	0.30	4.77	1.92	353.22	1891.57	2.35	233.74	202.67	15.22	5.88	18.96	18.10	551.28

EMS is based on three fundamental elements (see Figure 1.2) of the binary incidence matrix, after reordered through correspondence analysis: coherence, species turnover and boundary clumping (Leibold & Mikkelson, 2002). Coherence is calculated by counting the number of embedded absences (i.e., absences between presences) within species ranges or community compositions. Turnover is evaluated by counting the number of times two sites exchange two species. Finally, boundary clumping is the assessed by the Morisita index (Morisita, 1971), representing the degree of coincident range or community boundaries in the matrix (Leibold & Mikkelson, 2002). The significance of each element is assessed by a null model analysis, which is described in the methodology section of Chapter one. Using the interaction between these three basic elements of the incidence matrix, six different patterns can be distinguished: checkerboards, Clementsian gradients, Gleasonian gradients, evenly-spaced gradients, nested subsets and random distributions (See Figure 1.2).

Checkerboard patterning follows from Diamond (1975) fifth assembly rule which states that "some pairs of species never co-occur, either by themselves or in larger combinations, mainly due to competition" (Diamond, 1975). In this case the incidence matrix has significantly negative coherence (i.e., more embedded absences than expected by chance), which means that the metacommunity is composed by pairs of mutually exclusive species that occurs independently of one another (Presley *et al.*, 2009).

Clementsian and Gleasonian gradients come from an historical debate in community ecology that mainly focused on vegetation communities (Hoagland & Collins, 1997). One side argued that biotic communities are a discrete group of species that show similar responses to environmental factors (Clements, 1916) and

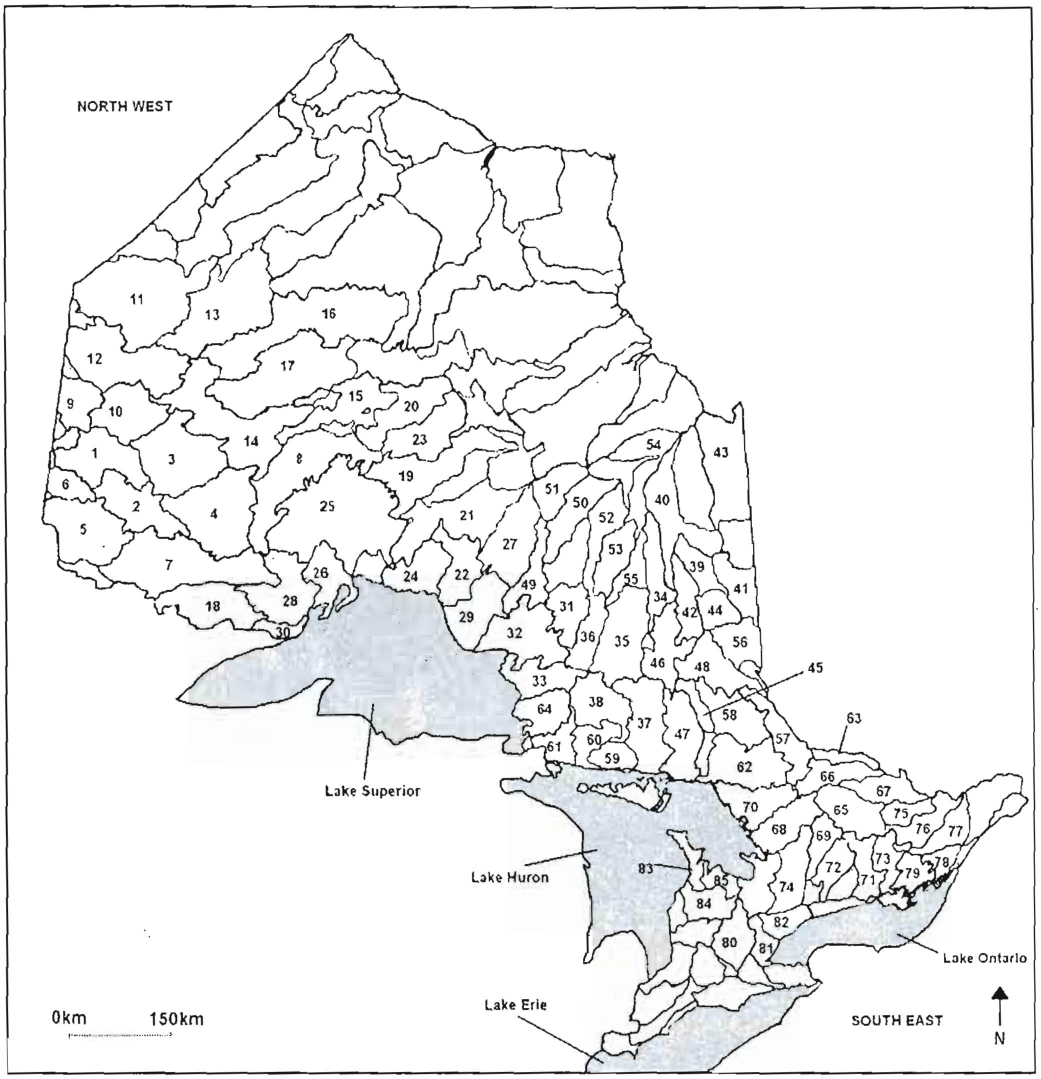

Figure 1.1 Map of Ontario divided by tertiary watersheds (Cox, 1978), with corresponding codes from Table 1.1. Watersheds without codes have no data available. Map adapted from Ministry of Environment (2004).

replace each other across space (Hoagland & Collins, 1997), whereas the other suggested that species have somewhat individual responses to abiotic factors and communities form a continuum of gradually changing compositions along the environmental gradient (Gleason, 1926). These ideas were recently extended to deal with animal communities as well (Leibold & Mikkelson, 2002, Heino, 2005; Presley *et al.*, 2009). Evenly-spaced gradients occur in metacommunities where species are competing along an environmental gradient and species distribution is dictated by trade-offs in their ability to explore alternative resources (Tilman, 1982; Leibold & Mikkelson, 2002). All three patterns appear in coherent metacommunities that exhibit positive turnover. The difference is in boundary clumping, where boundaries can be either clumped in Clementsian gradients, randomly distributed in the Gleasonian gradients or hyperdispersed in evenly-space gradients (Leibold & Mikkelson, 2002).

Nested subsets arise in sets of sites where poor-species biota are predictable subsets of the species composition from richer biota, i.e., common species occur at most sites and rare species occur only in the most diverse communities (Patterson & Atmar, 1986). Biotic nestedness has been found to be structuring the distribution of a large number of taxa (Wright *et al.*, 1998) and appears to be a common pattern in fragmented landscapes such as islands, isolated mountain tops and fragmented forest patches (Cook & Quinn, 1995; Patterson & Atmar, 1986; Honnay *et al.*, 1999; Férnandez-Juridic, 2002). Many studies have found that the main ecological processes driving nested patterns are selective colonization and selective extinction (e.g., Cook & Quinn, 1995; McDonald & Brown, 1992) which are related to differences in patches isolation and area, respectively. However, other causes for nested distributions have been also proposed, such as passive sampling, nested habitats, selective environmental tolerances and environmental harshness (see Ulrich *et al.*, 2009 for a complete review). In this case, the metacommunity is coherent and exhibits low turnover rates among communities due to some degree of overlap between their species composition. Finally, a metacommunity can exhibit non-significant coherence, which suggests that species are not responding to the same environmental gradient and are classified as random (Presley *et al.*, 2009).

Although these patterns are well studied in the ecological literature (e.g., Hoagland & Collins, 1997; Leibold & Mikkelson, 2002; Haudsorf & Hennig, 2007; Presley *et al.*, 2009; Ulrich *et al.*, 2009), there has been no attempt to compare different patterns across several metacommunities within the same system (but see Presley & Willig, 2010). Part of the problem is that large data sets which encompass several metacommunity systems over an entire biogeographic region are rare. This issue will be overcome by using the OFDD, because of the large number of lakes (n=8911) and the possibility to divide the set among somewhat discrete metacommunity units (e.g., watersheds), allowing comparisons among EMS patterns. Because each pattern can be considered a different "metacommunity trait", that are patterned by different structuring mechanisms, comparisons among patterns can increase our knowledge about how biological communities are structured over space, history and environmental gradients (Leibold & Mikkelson, 2002).

1.5 - The phylogenetic structure of ecological communities

Another important issue to consider in metacommunity studies is the evolutionary history of species (Loeuille & Leibold, 2008). In the last decade, the phylogenetic aspects of community assembly have gained increasing attention from ecologists (Webb *et al.*, 2002; Cavender-Bares *et al.*, 2009; Peres-Neto, 2004; Kraft *et al.*, 2007). Because species that diverged recently (i.e., species close within a phylogenetic tree) tend to be ecologically similar (Cavender-Bares *et al.*, 2004; Peres-Neto, 2006), there may be a link between the phylogenetic relatedness of taxa and the factors that determine their distributions (Leibold *et al.*, 2010). Assuming that species niches are somewhat conserved through time (i.e., closely related species diverge less through time than of what would be expected in an unconstrained evolutionary

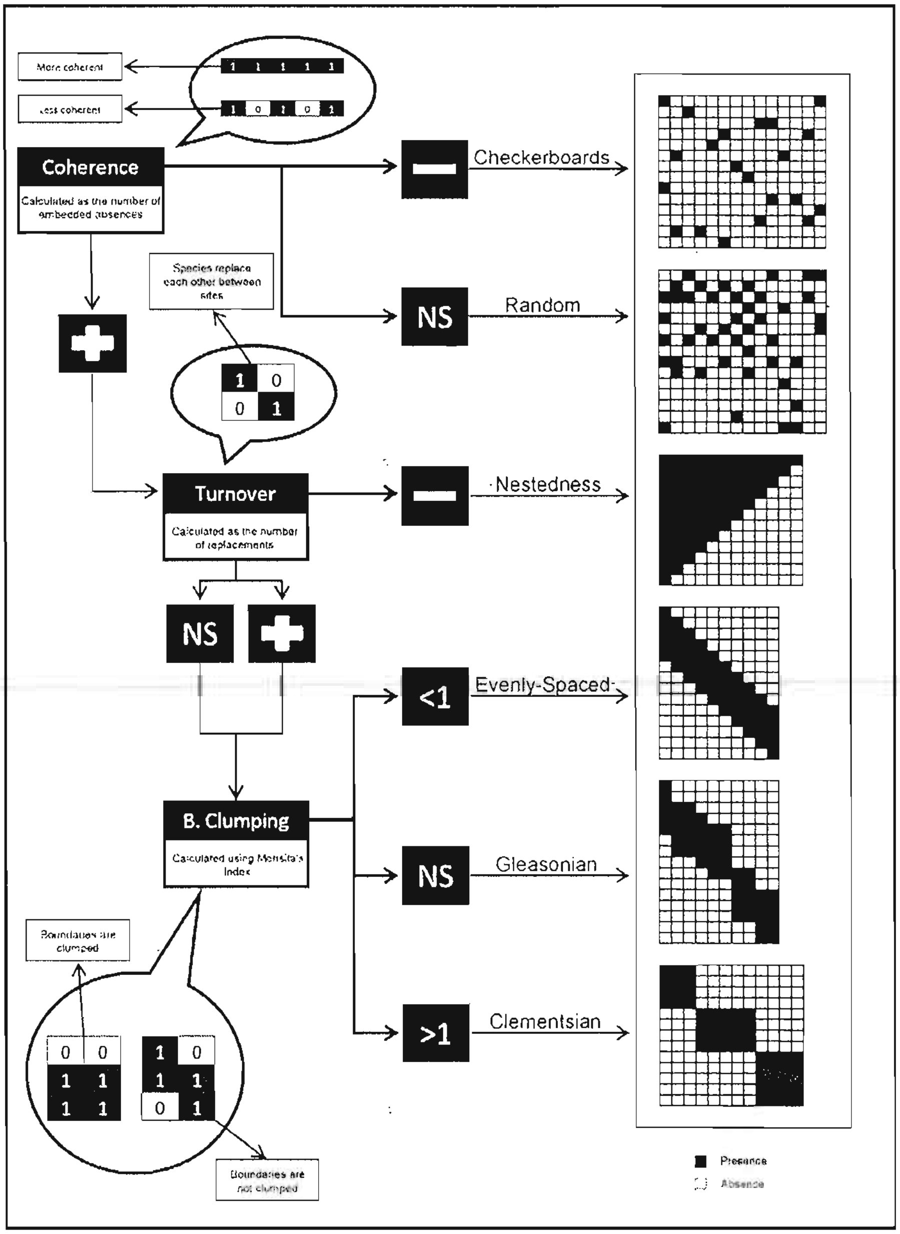

Figure 1.2 General framework for the Elements of Metacommunity Structure (coherence, turnover and boundary clumping). Columns represent sites and rows represent species. NS = non significant. See Table 1.2 for EMS results of the hypothetical matrices.

Table 1.2 Results of EMS analyses on hypothetical matrices from Figure 1.2. Analyses were performed using the first ordination axis extracted via reciprocal averaging and based on community perspective. Abs = number of embedded absences; Re = number of replacements; Mo = Morisita's index; μ = mean value of each element for the random distribution; σ = standard deviation of each element for the random distribution; p = significance probability; SV = standardized value. Significant results ($p \leq 0.05$) are in bold. *Coherence standardized values were multiplied by -1 ** Mo is statistically tested with a two-tailed test, thus when $p \geq 0.95$, the result is significant and indicate an evenly-spaced gradient metacommunity pattern.

	Coherence				
Pattern	Abs	μ	σ	p	SV*
Checkerboards	73	17.007	11.1711	**<0.001**	-5.01230855
Random	30	43.137	7.5232	0.0808	1.74619843
Nestedness	0	26.171	4.2718	**<0.001**	6.12645723
Evenly-Spaced	0	36.257	5.4515	**<0.001**	6.65083005
Gleasonian	0	38.482	5.841	**<0.001**	6.58825544
Clementsian	0	42.724	6.0589	**<0.001**	7.05144498

	Turnover				
Pattern	Re	μ	σ	p	SV
Checkerboards	.	.	.	.	.
Random	.	.	.	.	.
Nestedness	0	200.657	66.3235	**<0.001**	-3.02542839
Evenly-Spaced	575	520.38	93.9057	0.5608	0.58164733
Gleasonian	649	563.356	91.6	0.3498	0.93497817
Clementsian	880	614.996	86.6173	**0.0022**	3.05948119

	Boundary Clumping				
Pattern	Mo	μ	σ	p	SV
Checkerboards	.	.	.	.	.
Random	.	.	.	.	.
Nestedness	.	.	.	.	.
Evenly-Spaced	0.4706	0.1179	0.2882	0.975**	1.22380291
Gleasonian	1.2426	0.2183	0.4489	0.185	2.28179996
Clementsian	2.2917	0.2334	0.5533	**0.005**	3.72004338

process; Wiens & Graham, 2005), phylogeny can assist in disentangling two main opposing processes of community assembly: habitat filtering, where communities are composed of species that share similar environmental tolerances (Webb, 2000), and competitive-exclusion, where species that are ecologically similar cannot co-exist in the same local communities due to high overlap in resource use (MacArthur & Levins, 1967; Mason *et al.*, 2008). If the first process is dominating, co-occurring species are expected to exhibit a pattern of phylogenetic clustering (Cavender-Bares *et al.*, 2006); however, if the second process is the most important, communities will be composed of distant-related species (i.e., phylogenetic overdispersion) due to competitive exclusion of species that use similar resources. If niche is not conserved through time due to, for instance, convergent evolution due to natural selection, phylogenetic results have less power to detect non-random community assembly processes (Losos, 2008).

Both processes (e.g., habitat filtering and limiting similarity) can influence patterns and dynamics observed at the metacommunity level (Pillar & Duarte, 2010). For example, Clementsian gradient is a pattern that occurs in metacommunities where clusters of communities are formed by a discrete group of species that can either show a similar response to environmental gradients (Presley *et al.*, 2009) or be a result of "clumped competitive exclusion" (Gilpin & Diamond, 1982). Evidence for these mechanisms can be obtained by a community phylogenetic analysis: if co-occurring species are closely related and present phylogenetic niche conservatism than it might suggest that environmental filtering is selecting species with similar tolerances to the environment (Cavender-Bares *et al.*, 2009) and structuring the Clementsian gradient pattern at the metacommunity level. However, if co-occurring species are more distantly related in the phylogeny, than two processes can be suggested: 1) competitive interactions are precluding long-term co-existence between species that present phylogenetic niche conservatism and thus similar ecological characters (Cavender-Bares *et al.*, 2004); 2) Phylogenetic distant related species present similar ecological characteristics due to convergent evolution resulting from environmental filtering (Losos, 2008). Thus, after finding the pattern that best fit the species distributions within a metacommunity, we can use the phylogenetic approach to seek for possible structure mechanisms that dictate these patterns.

CHAPTER I

A COMMUNITY OF METACOMMUNITIES: EXPLORING PATTERNS IN SPECIES DISTRIBUTIONS ACROSS LARGE GEOGRAPHICAL AREAS

Une communauté de metacommunautés: Exploration des patrons de distribution d'espèces à travers de larges aires géographiques

2.1 - Introduction

2.1.1 - Metacommunities: Structural *versus* Mechanistic approach

Unraveling and disentangling multiple mechanisms influencing the composition and variation of ecological communities across space is a central problem and intellectual challenge in biology (Ricklefs, 1987; Gaston, 2000; Holyoak *et al.*, 2005). Research based on spatial patterns of biodiversity seeks to identify the set of abiotic and biotic process (and how they interact) that define how subsets of species (i.e., local communities) are filtered down from those found in the larger regional species pool of potential colonizers (Gotelli and Graves, 1996; Jackson *et al.*, 2001; see also Figure 3.2 in chapter two).

There is a growing consensus among ecologists that both large-scale processes and local factors need to be considered in order to understand spatial patterns of biodiversity (Ricklefs, 2004; Holyoak *et al.*, 2005). In this context, metacommunity is the fastest advancing framework in spatial ecology because it accounts for both local (e.g., environment) and regional (e.g., dispersal) processes (Leibold *et al.*, 2004). In this framework, a metacommunity is defined as a set of local communities that are potentially linked, but not necessarily, by dispersal of individuals of species across local communities (Leibold *et al.*, 2004; Holyoak *et al.*, 2005).

Metacommunities have been mainly studied by two approaches: the mechanistic approach (Holyoak *et al.*, 2005; Muneepeeraku *et al.*, 2008; Driscoll & Lindenmayer, 2009), in which the focus is on determining the predominance (or their relative importance, e.g., Cottenie 2005) of distinct model processes (e.g., species sorting, patch dynamics, mass-effects and neutral) regarding different assumptions underlying metacommunity dynamics. Some of these assumptions involve species trade-offs, dispersal rate differences across species and communities, the presence of environmental gradients and stochastic *versus* deterministic processes (Leibold *et al.*, 2004); and 2) the structural approach (Leibold & Mikkelson, 2002; Heino, 2005;

Hausdorf & Henning, 2007; Presley *et al.*, 2009; Presley & Willig, 2010), which focuses on understanding non-random patterns in the structure of the distributions of species across communities represented by species incidence or abundance matrices. Uncovering large-scale distributional patterns is an essential source of inferences about the causes driving variation in species composition across communities (Gotelli and Graves 1996; Peres-Neto, 2004; Werner *et al.*, 2007). Identifying the forces structuring these patterns can provide clues about the underlying processes driving species co-occurrence and site occupancy mechanisms. In this work, I will focus on the latter approach given that under only certain conditions we can estimate the likelihood of the different models under the mechanistic approach (Legendre *et al.*, 2008; but see Cottenie, 2005).

The analysis of patterns in incidence matrices in order to determine the degree of negative and positive associations across species has a long history of applications in ecology (Diamond, 1975; Connor & Simberloff, 1979; see Gotelli & Graves, 1996 for a review). Although this species-guided approach has provided important insights into ecological processes structuring ecological communities, it has been somewhat limiting because it does not take into account the organization of sites that may also influence species co-occurrences patterns. For instance, species along environmental gradients may be organized into blocks of species overlapping in their distribution. Species within blocks may appear positively associated due to common habitat affinities, whereas species across blocks may appear negatively associated due to differences in habitat across environmental gradients. Indeed, more recently, ecologists have been exploring a dual species and community (site) perspective (Leibold & Mikkelson, 2002). Perhaps the most well known of these incidence patterns is the nested species subsets in which communities of successively fewer species contains subsets of those species found on the next richer community (Patterson & Atmar, 1986).

2.1.2 - Patterns of distribution and the Elements of Metacommunity Structure

Leibold & Mikkelson (2002) developed a framework termed Elements of Metacommunity Structure (EMS) that analyzes several possible patterns that take into account this dual arrangement of communities and species. In their EMS framework, the interaction between three different elements (i.e., coherence, turnover and boundary clumping) generates six different possible patterns of species distribution across communities, which have been referred as to nestedness, Clementsian gradients, Gleasonian gradients, evenly-spaced gradients, checkerboards and random (Figure 1.2). In this approach, each pattern assumes that species distributions are a result of particular species responses to abiotic and biotic factors along a major distributional gradient of species across sites (Leibold & Mikkelson, 2002; although multiple interacting gradients could provide greater refinement, little analytical progress has been made in this direction; but see Presley *et al.*, 2009). The interactions of each metacommunity element described above (i.e., coherence, turnover and boundary clumping) make predictions about the six above mentioned metacommunity patterns. When species do not respond to the same environmental gradient (i.e., different habitat affinities), the metacommunity (i.e., distribution of species across communities embedded in local sites) will present a random structure (Presley & Willig, 2010). If metacommunities are composed of pairs of mutually exclusive species that occur independently of other pairs along the gradient, they are classified as checkerboards (Diamond, 1975). Nestedness occurs on metacommunities with low turnover rates, where the composition of poor-species sites represents proper subsets of progressively richer sites (Ulrich *et al.*, 2009; see above). When turnover rates (i.e., changes of species compositions across communities) are higher than expected, metacommunities can be classified as Clementsian, Gleasonian or evenly-spaced gradients. The first one indicates that biotic communities are a discrete group of species that shows similar responses to the gradient and replace each other on space across the metacommunity (Clements, 1916). Gleasonian gradients represent communities composed of species that show idiosyncratic responses to the gradient, yielding a metacommunity with a form of a continuum of gradually changing composition (Gleason, 1926). Finally, metacommunities defined as evenly-spaced gradients are composed by species supposedly competing along a gradient and their distribution will be dictated by trade-offs in their ability to explore alternative resources (Tilman, 1982; Leibold & Mikkelson, 2002).

Although a large number of studies have assessed some of these EMS patterns separately (e.g., nestedness, Cook & Quinn, 1995; Wright *et al.*, 1998; Fernández-Juricic, 2002; Leprieur *et al.*, 2009; checkerboards, Diamond, 1975; Connor & Simberloff, 1979; Gilpin & Diamond, 1982), to date only a few studies have applied this approach to test which pattern best fits to a given metacommunity data (studies were reviewed by Presley *et al.*, 2009), or compared patterns within metacommunities across different systems (e.g., Leibold & Mikkelson, 2002).

2.1.3 - Looking further into EMS framework

The EMS approach is extremely promising because it allows characterizing metacommunity patterns across different taxa, metacommunities and ecosystems, providing an exceptional venue to search for general rules in determining the structure of community assemblages across space. For instance, two bird and two plant metacommunities, each taxa combination (i.e., one bird and one plant) being in different climatic zones (tropical *versus* temperate) may show different EMS patterns across taxa (i.e., bird *versus* plant) but similar within regions (e.g., nested plant and bird in tropical region and Gleasonian gradients plant and bird in the temperate region). In this case, the conclusion would be that the climatic zone is driving the pattern. Although promising, this comparative approach either across taxa or region has yet to be explored. To my knowledge, only one study has looked at how these elements of metacommunities compare across different regions for the same taxa (i.e., bat metacommunity structure on Caribbean islands; Presley & Willing, 2010). However, their study was also limited by the fact that they only have three regions and therefore little inference can be made about how differences between regions could have explained the observed patterns.

Given that each EMS pattern can be considered as a different "metacommunity trait" with unique underlying structuring mechanisms and theory (Leibold & Mikkelson, 2002; Hoagland & Collins, 1997), exploring and comparing such patterns across large geographical regions has the potential to enhance our understanding of how biological communities respond to environmental (Presley *et al.*, 2009) and biogeographical variation. Moreover, insights on key ecological patterns such as β-diversity (Leprieur *et al.*, 2009) might be acquired throughout such comparisons (e.g., nestedness *versus* turnover; Hausdorf & Hennig, 2007).

2.1.4 - Lake-fish systems as metacommunities

Lakes within a watershed can be considered as "virtual islands" (Magnuson *et al.*, 1998) varying in size, environmental features (Eadie *et al.*, 1986) and degree of isolation (Olden *et al.*, 2001), which may impose different environmental and spatial constraints which in turn will influence fish dispersal and probability of establishing viable populations, as well as their extinction vulnerability (Magnuson *et al.*, 1998; Olden *et al.*, 2001). Indeed, some studies have found that local environment was the most important predictor of lake-fish species distribution (e.g., Magnuson *et al.*, 1998) whereas others have found that spatial (i.e., regional) factors were the most prevalent (e.g., Beisner *et al.*, 2006). This dichotomy shows that different lake-fish metacommunities can be structured by different factors, but little is known whether there are general assembly patterns emerging from these processes. The search for general rules that may dictate patterns that best reflect the distribution of species within a metacommunity should increase our understanding about the underlying mechanism structuring metacommunities (Leibold & Mikkelson, 2002; Heino, 2005). To date no study has investigated whether and how local and regional features generate consistent patterns across different metacommunities.

2.1.5 - Chapter objectives

I used a unique data set containing environmental and presence-absence data on fish distribution on about 9000 boreal lakes from Ontario, Canada, across 85 metacommunities (watersheds). The approach used here was the following: 1) classify each fish metacommunity (i.e., watershed) according to EMS patterns; 2) determine the relative influence of spatial and environmental factors within and across metacommunity EMS patterns.

2.2 - Methodology

2.2.1 - Ontario Fish Distribution Database (OFDD)

I used a lake-fish database, the Ontario Fish Distribution Database (OFDD), maintained by the Ontario Ministry of Natural Resources, which contains presence-absence records of 134 fish species (including 7 hybrids) and geographic positions for approximately 9900 boreal lakes (inland lakes only) from Ontario. Records span from 1900 to 1992, however most lakes were sampled between 1968 and 1985. The OFDD is known to have sampling biases, where sport fishes are overrepresented and small-bodied species, such as cyprinids, are underrepresented (Minns, 1986). The full history of the dataset can be found in Mandrak and Crossman (1992a) and the sampling methods in Goodchilde and Gale (1982). Despite the potential sample biases and the fact that collection spanned over a long period, this dataset has been providing important insights in many different types of ecological research (Mandrak, 1995; Gonzalez & Gardezi, 2008; Sharma *et al.*, 2009). Finally, given that I am interested in broad regional-scale patterns, sampling biases should be diluted across regions.

2.2.2 - Lake Inventory Database (LINV)

Information about the local environment in each lake was assessed using the Lake Inventory Database (LINV), a dataset that includes the following environmental variables for each lake in the OFDD: surface area (SA), shoreline perimeter (P), island shoreline perimeter (ISL), mean depth (MeanD), maximum depth (MaxD), secchi depth (SD), growing degree days (GDD), elevation (Elev), total dissolved solids (TDS), morpho-edaphic index (MEI), mean annual daily temperature (MADT), canopy cover (Crown), mean July temperatures (MJT) and mean August temperatures (MAT). Missing values were replaced by the mean value of that variable within the watershed; 0.2% of the lakes on average per variable were replaced. Note that the environmental information of lakes treated in this way became uninformative, specially compared to the total number of lakes used in the analysis (n ≈ 9000). I have also considered values of potential evapotranspiration (PET) which serves as a proxy of thermal energy that is in turn correlated with lake productivity (Gonzalez & Gardezi, 2008). This variable was missing for a large number of lakes, but has low variability within watersheds (see Gonzalez & Gardezi 2008 for details on how this measure was estimated), and therefore was only used in analyses among watersheds In this case, I used the mean PET from all lakes that were available for any given watershed. Finally, lakes with missing geographic coordinates or without species were removed from the analyses.

2.2.3 - Species used in the analyses

Species that were present in less than on an arbitrary value of 0.5 % of all lakes in the data set were removed. Rare and endemic species are somewhat uninformative due to its idiosyncratic nature, but they can affect EMS analysis in ways that will not be discussed here (but see Presley & Willig, 2010). Introduced species were also removed because they do not follow any historical contingency experienced by the native species. In total, 53 extant native species across all lakes were used in analyses (Table 3.1; Chapter 2).

2.2.4 - Watersheds as metacommunities

There are three primary watersheds in Ontario, 28 secondary watersheds nested within the primary, and the 144 tertiary watersheds that are nested within the secondary watersheds (Cox, 1978). As in other studies based on this data set (e.g., Chu *et al.*, 2005; Gardezi & Gonzalez, 2008), I have considered tertiary watersheds as the scale unit of the analysis (i.e., they represent group of fish communities). Tertiary watersheds are nested within secondary watersheds thus sharing hierarchical topological rules, though unlike secondary watersheds, the delineation of tertiary ones presents some arbitrary level regarding size convenience for management purposes (called ecoregions; see Minns, 1989). As a consequence, some lakes across two spatially close tertiary watersheds could potentially share a greater fish dispersal history than within their own designated watersheds; though lakes within tertiary watersheds should still share in average a greater dispersal history than compared to lakes across watersheds. Finally, the results showed very strong EMS patterns within tertiary watersheds (see "Results" section), indicating that this scale does represent an important ecological unit for their fish assemblages. I defined each metacommunity unit as being composed by the fish communities within lakes of any given tertiary watershed (Figure 1.1). Finally, I have excluded all tertiary watersheds with less than 20 lakes as the EMS analysis would have low statistical power based on these small incidence matrices (Leibold & Mikkelson, 2002), resulting in a database of 8911 boreal lakes distributed across 85 tertiary watersheds. Hereafter, metacommunity and watershed will be used interchangeably.

2.2.5 - Statistical Analyses

2.2.5.1 - Ordination

Community composition and species distributions across communities are probably regulated by multiple environmental characteristics (e.g., area, temperature, habitat heterogeneity) and biotic interactions. Multivariate ordination techniques can be used to order species along mathematical gradients (i.e., gradient constructs) that integrate common patterns of variation in community composition. Given that common or opposite patterns of variation is the product of the integration of multiple environmental factors and biotic interactions, they can provide clues about the factors that structured these gradients in the first place (Gauch *et al.*, 1977). In this study, each metacommunity was analyzed via a separate correspondence analysis (also known as reciprocal averaging). Correspondence analysis creates orthogonal axes (gradients) in which species and sites are ordinated (Presley & Willig, 2010). This method re-orders rows and columns using repeated averaging of species and sites scores, maximizing their correspondence. Thus, it maximizes the positioning of sites along axes based on the degree in which their communities share species compositions and the positioning of species sharing similar distributional ranges (i.e., across multiple sites) (Leibold & Mikkelson, 2002). The final solution is a compromise between minimizing interruptions within ranges and minimizing interruptions within communities (Leibold & Mikkelson. 2002). The eigenvalue of any correspondence analysis axis represents the correlation between species and sites scores (Gauch *et al.*, 1977). As in previous studies (e.g., Presley & Willig, 2010), sites and species within incidence matrices were ranked according to their position along the primary ordination axis, which maximizes this correlation (Gauch *et al.*, 1977). Other axes of ordination were not used as they did not explain much of the variation in species distribution.

2.2.5.2 - Null model

In order to test the significance of EMS patterns, I applied a null model that permutes species across lakes (sites) but that kept the total number of species in lakes as fixed (i.e., equal to the observed values). Thus, the chosen null model included some site property such as species richness, which, in lakes, is highly correlated with surface area, (e.g., Eadie *et al.*, 1986; Barbour & Brown, 1974; Gardezi & Gonzalez, 2008). Note that there are several ways in which to permute incidence matrices and the procedure used has correct Type I error rates and appropriate levels of power in detecting distributional patterns (Gotelli & Graves, 1996). Conversely, other types of algorithms have performances that do not provide a good compromise between Type I and II errors. For instance, Leibold and Mikkelson (2002) used an equiprobable-equiprobable (i.e., no constraints in occurrences throughout the incidence matrix) and a fixed-fixed (i.e., row marginal totals and columns marginal totals fixed to the number observed in the empirical matrix) null models. I chose not to use these approaches due to an elevated chance of type I errors in the former and increased type II errors in the latter (Gotelli & Graves, 1996). This occurs because the liberal null model assumes no structure in its randomization (i.e., increased Type I errors) whereas the conservative assumes too much structure (i.e., increased Type II errors) (Gotelli & Graves, 1996).

2.2.5.3 - EMS analyses

The elements of metacommunity structure are evaluated in an hierarchical way (see Figure 1.2). The first element of the metacommunity structure, coherence, was evaluated as the number of embedded absences (Abs) in species range and community composition of each watershed. I generated 1000 random matrices (for each watershed) based on the above described null model which were also ordinated using the primary axis extracted via correspondence analysis (i.e., each random matrix was analyzed by a separate correspondence analysis). The probability (*p*) was assessed using a one-tailed test with an alpha of 0.05: (number of $\mathrm{Abs}_{\mathrm{OBS}}$ equal or larger than $\mathrm{Abs}_{\mathrm{RND}}$ +1) / (number of randomizations +1) or (number of $\mathrm{Abs}_{\mathrm{OBS}}$ equal or smaller than $\mathrm{Abs}_{\mathrm{RND}}$ +1) / (number of randomizations +1), where 1 represents the observed value and is also included as a possible outcome of the randomization process. If coherence was significantly lower than in the null distribution (i.e., build from the calculated values of the 1000 random matrices), it would indicate that the metacommunity is composed by pairs of mutually exclusive species that were independently occurring of other pairs (Figure 1.2), such as the checkerboard pattern (Diamond, 1975; Leibold & Mikkelson, 2002). Non-significant coherence would suggest that the metacommunity is composed of species that are not responding to the same environmental gradient, creating a random structure (Figure 1.2). Finally, a significantly positive coherence suggests that species are distributed according to the same environmental gradient (Leibold & Mikkelson, 2002).

The two other elements, turnover and boundary clumping, can be evaluated using two perspectives: species range or community (site) composition (Leibold & Mikkelson, 2002). In this study, metacommunities were evaluated only by the community perspective (i.e., composition) because I wanted to assess the potential roles of lake characteristics in structuring metacommunity structure. Moreover, I was interested in understanding the role of environmental and spatial processes accounting for some of their ecological features, such as richness, composition and turnover. Despite some of the reservations regarding this perspective (Presley *et al.*, 2009), I argue that the common environmental gradients to which species are responding have to be represented by the communities in which they inhabit and not by the species that are present or absent. Turnover is evaluated counting the number of times a pair of sites "exchanged" (i.e., replacements) between two species (Figure 1.2). The empirical number of replacements (Re) was compared to the number of replacements calculated from the null distribution. The probability (*p*) was assessed using a one-tailed test with an alpha of 0.05: (number of Re_{OBS} equal or larger than Re_{RND} +1) / (number of randomizations +1) or (number of Re_{OBS} equal or smaller than Re_{RND} +1) / (number of randomizations +1), where 1 represents the observed value and is also included as a possible outcome of the randomization process. If metacommunities showed significantly low turnover, it was a sign of nested distribution (Figure 1.2); and conversely, if it exhibited moderate turnover (i.e., not significant) or high turnover, I evaluated boundary clumping to distinguish it among the remaining patterns (Figure 1.2).

The degree of community boundary clumping for each watershed (i.e., the third and last EMS) was assessed with Morisita's index (Morisita, 1971) which has a null expectation of 1. If the value obtained was not significantly different from 1, it indicated that community boundaries are randomly distributed, suggesting Gleasonian gradients ($0.05 < p < 0.95$). However, if the index value was significantly greater (clumped boundaries) or smaller (over-dispersed boundaries) than 1, it suggested that the metacommunity was distributed according to a Clementsian gradient ($p < 0.05$) or over evenly-spaced gradient ($p > 0.95$), respectively (Leibold & Mikkelson, 2002).

In order to compare metacommunities according to EMS patterns, I standardized (i.e., mean= 0 and variance = 1) each of the EMS within watershed units using the mean and standard deviation of their respective null distribution. Because in coherent metacommunities the mean number of embedded absences in the null distribution is higher than the observed value, when standardized, coherent metacommunities presented negative values. To facilitate interpretation, coherence standardized values were multiplied by -1 so that more coherent metacommunities presented positive values.

2.2.6 - Biotic and abiotic lake indices

2.2.6.1 - Community similarity

In order to access the similarity between pairs of lakes within any given watershed, I computed the Jaccard Index as follows:

$$J = \frac{S_{ij}}{S_i + S_j + S_{ij}}$$

where S_i is the number of species only present at site *i*, S_j is the number of species only present at site *j* and S_{ij} is the number of species present at both sites. To obtain an overall measure of similarity among lakes within each watershed, I averaged the values from all pairs of lakes within each watershed.

2.2.6.2 - Watershed connectivity

I have modified Hanski's (1994) connectivity measure to assess the levels of spatial connectivity across lakes within watersheds as follows:

$$AC_{ik} = \frac{\sum_{\substack{i=1 \\ i \neq j}}^{n} p_{jk} \exp(-d_{ij})}{n}$$

where AC measures the average geographic distance (based on latitude/longitude) across lake i for the k^{th} species across all other n-1 lakes within any given watershed. p indicates the presence (1) and absence (0) of the k^{th} species in the j^{th} lake. Note that in cases where species i was found only in one lake, I assigned for that species the maximum distance between any two sampled lakes within the watersheds as its connectivity value (i.e., maximum isolation or smallest connectivity). A non-linear response was used to reduce the contribution of lakes far away from the focal lake as it should be increasingly hard for lakes beyond a certain distance to contribute to the connectivity of any given lake (Kadoya, 2009). For each lake, overall connectivity was calculated as the average connectivity values for all species for any given lake. The overall connectivity of any given watershed was the average connectivity across all lakes within any given watershed.

2.2.6.3 - Postglacial dispersal

The province of Ontario underwent a recent process of glaciation (8000-10000 years ago; Mandrak & Crossman, 1992b). Ontario lakes and rivers were formed from the meltwaters of the receding Wisconsin glacial sheet and actual patterns of fish species distributions were contingent upon the process of re-colonization from fishes that dwelled in refugees south of the ice sheet (Mandrak & Crossman, 1992b). I expected that the distance between postglacial routes and lakes may have influenced the likelihood of their colonization by fish species (Olden *et al.*, 2001), affecting the species found in any particular watershed and ultimately influencing metacommunity structure. Mandrak & Crossman (1992b) suggested five postglacial dispersal routes, used by fish to re-colonize Ontario (Figure 2.1 in the result section). In order to assess the importance of distance from these refugees in structuring metacommunity patterns, I calculated the following two indices: MeanDP as the mean Euclidian distance between any given watershed and all five postglacial routes, and MinDP as the minimal Euclidian distance between any given watershed and its closest postglacial route.

2.2.6.4 - Environmental gradient length

Regarding environmental characteristics of each watershed, I calculated an index that measures the variation in environmental conditions across lakes within a watershed (hereafter referred to as EnvDis). First, all environmental variables were standardized (mean=0 and variance=1) across all the lakes from the database (i.e., all lakes across all watersheds and not only within watersheds). Then, a Euclidean distance matrix based on these variables but contrasting lakes within watersheds was calculated. Finally, distance values within watersheds were averaged across lakes in order to produce a global measure of environmental variability within watersheds and can be seen as the "length" (or extension) of their environmental gradients.

2.2.6.5 - Abiotic integration

Additionally, I computed a measure that I refer to as abiotic integration (AbINT), which assesses the degree of correlation among environmental variables within watersheds. For instance, a watershed can have a small or large gradient (measured by EnvDis), but variables across lakes in that watershed may have small or large correlations across them. Abiotic integration was computed based on a principal component analysis of the correlation matrix of environmental variables within a particular watershed. AbINT was then based on the variance across eigenvalues (this is akin to morphological integration, see Peres-Neto & Magnan, 2004). High variance levels indicate that the environmental variables were more correlated within a watershed when compared to low levels. In order to remove any bias across watersheds having different number of lakes, I used a null distribution based on 1000 permutations where values within environmental variables were permuted across lakes. Then, I subtracted the mean of the null distribution from the observed value to obtain corrected values of AbINT. The rationale for using this integration metric is based on the fact that species use environmental cues to perform important activities such as dispersal and reproduction, which ultimately influence their distribution and the metacommunity structure, at broader scales. It follows that if environmental variables are highly correlated, they may aid fish to make more coherent decisions across different lakes or detect important changes that will have important fitness consequences as correlated environmental characteristics may provide a stronger signal (cue) to initiate certain activities and/or make life-history and dispersal decisions.

2.2.7 - Environmental *versus* spatial variation

In order to determine how much of the variation in lake-fish distribution was explained by environmental *versus* spatial factors (e.g., missing environmental variables or dispersal, see Peres-Neto & Legendre 2010, and Jacobson & Peres-Neto 2010 for a recent discussion on the potential factors driving spatial patterns of species distribution) within each watershed, I applied a variation partitioning scheme in order to estimate unique and combined contributions of environmental and spatial predictors (Borcard *et al.*, 1992). Environmental variables from all lakes within watersheds were used as environmental predictors and I used the MEM (Moran's Eigenvector Maps) method to describe spatial variation (Peres-Neto & Legendre, 2010). Species data (the response variable) were Hellinger transformed (Legendre & Gallagher, 2001), as it has been shown to provide unbiased estimates for variation partitioning based on redundancy analysis (Peres-Neto *et al.*, 2006). Peres-Neto et al. (2006) has also shown that variation partitioning is affected by the number of spatial and environmental predictors as well as by sample size, so results were based on adjusted fractions of variation, which is analogous to the adjusted R^2 in multiple regression models. Statistical significance was based on 1000 permutations based on an alpha of 0.05 to assess significance. When a fraction was not found significant, I assigned a zero to its value instead of the observed value, which can be negative, especially in the case of adjusted values. Finally, I averaged the explained variation from spatial and environmental predictors across watersheds per pattern uncovered by EMS analysis.

In order to assess any relationship between EMS patterns and the environment, I first calculated the mean value of each environmental variable within watersheds. All environmental variables were log-transformed, except elevation, mean depth, max depth and secchi depth, which appeared, by visual inspection, to have linear correlations with the standardized EMS values. Next, a redundancy analysis (RDA) was performed and Pearson correlations were calculated between the first two canonical axis and the environmental variables and indices. To test if the overall environment was different among EMS distributional patterns, I performed a discriminant function (Legendre & Legendre, 1998).

EMS patterns and their associated null models were performed using a Matlab code developed by Presley et al. (2009). This code is available at http://www.tarleton.edu/~higgins/EMS.htm. The Jaccard index was calculated using a Matlab script developed by Strauss (2008). The connectivity index, EnvDIS and variation partitioning were performed with functions written in Matlab 7, Release 14 (The Mathworks Inc).

2.3 - Results

2.3.1 - Environmental and spatial drivers of metacommunity patterns

Among the 85 watersheds analyzed with the EMS methodology, four patterns were uncovered: 42 watersheds were consistent with nested distributions, 35 with Clementsian gradients, 5 with Gleasonian gradients and 3 random (Figure 2.1). Given that only a few watersheds were found to be random or representing Gleasonian gradients, I restrained all the analyses to the two most common patterns: nestedness and Clementsian gradients. RDA based on the three elements (i.e., coherence, turnover and boundary clumping) uncovered the two predominant metacommunity patterns (Figure 2.2), Clementsian Gradients and nestedness. RDA-1 and RDA-2 accounted for 55.94 % and 35.64% of the total variation, respectively. The first canonical axis (RDA-1) was positively correlated with turnover ($r = 0.52$) and negatively correlated with coherence ($r = -0.93$) and boundary clumping ($r = -0.72$) (Figure 2.3). The second canonical axis (RDA-2) was positively correlated with turnover ($r = 0.82$) and clumping ($r = 0.62$) and was uncorrelated ($r = -0.002$) with coherence (Figure 2.3). In general, watersheds that were more consistent with nestedness had a trend to have moderate to high negative scores for RDA-1 and negative scores for RDA-2, which accounted for high values of coherence, low turnover and moderate clumping (Figure 2.2 and Figure 2.3) whereas Clementsian watersheds had moderate positive scores of RDA-1 and high positives scores for RDA-2, which accounted for moderate values of coherence and high values of turnover and clumping (Figure 2.2 and Figure 2.3). The discriminant analysis indicated that the overall environment was significantly different ($F (14.62) = 2.6023$, $p < 0.0051$) between nestedness and Clementsian watersheds.

Nestedness was predominant in higher latitudes (Figure 2.1), but appeared in considerable numbers in the southern region of Ontario as well. Abiotic integration was higher in nested watersheds, suggesting that environmental variables were more correlated between lakes within those watersheds (Figure 2.4). As expected, lakes from nested watersheds were also highly positively correlated with their similarity in species composition (i.e., Jaccard index; Figure 2.4) given the lower levels of species turnover across communities within watersheds. Average connectivity (AC) was positively correlated with turnover (Figure 2.4) indicating that nested watersheds, in general, have less connected lakes than Clementsian watersheds. Nested watersheds were also composed of lakes with greater surface area, shoreline perimeter and island perimeter (Figure 2.3). Most nested watersheds showed low values of total dissolved solids (TDS), morpho-edaphic index (MEI) (which are lake productivity proxies) and EnvDIS (indicating a tendency for a smaller variation across the environmental gradient of these watersheds), which were all negatively correlated with PC-1 (Figure 2.3). Finally, these watersheds have higher values of MeanDP, meaning that in average they are located farther from postglacial routes (Figure 2.4).

Clementsian gradients appeared in watersheds mostly located at lower latitudes and near the Great Lakes (Figure 2.1). They presented lower values for AbINT (abiotic integration), suggesting that environmental variables were less correlated (Figure 2.4). As expected, turnover was highly negatively correlated with similarity in species composition (Jaccard Index; Figure 2.4). Moreover, Clementsian metacommunities were in general positively related with MEI, TDS and EnvDIS (Figure 2.3 and Figure 2.4). Furthermore, this pattern appeared, on average, in watersheds closer to postglacial routes (i.e., lower MeanDP; Figure 2.1 and Figure 2.4).

The variation partitioning did not reveal any apparent differences between nested and Clementsian watersheds. In both cases, environmental predictors explained, on average, 9.1% of the variation in species composition across lakes within watersheds (Table 2.1). Space explained 3.5% and 2.9% of the variation in species composition of nested and Clementsian watersheds, respectively (Table 2.1).

To test if nested patterns are being driven by lake connectivity (e.g., more connected lakes are richer than less connected lakes) I calculated the correlation between lake mean connectivity and their richness (log transformed) for all watersheds (including Clementsian ones). For nested watersheds, 11 revealed positive significant correlations between these two variables ($r = 0.12$ to 0.49 / $p < 0.05$) and 3

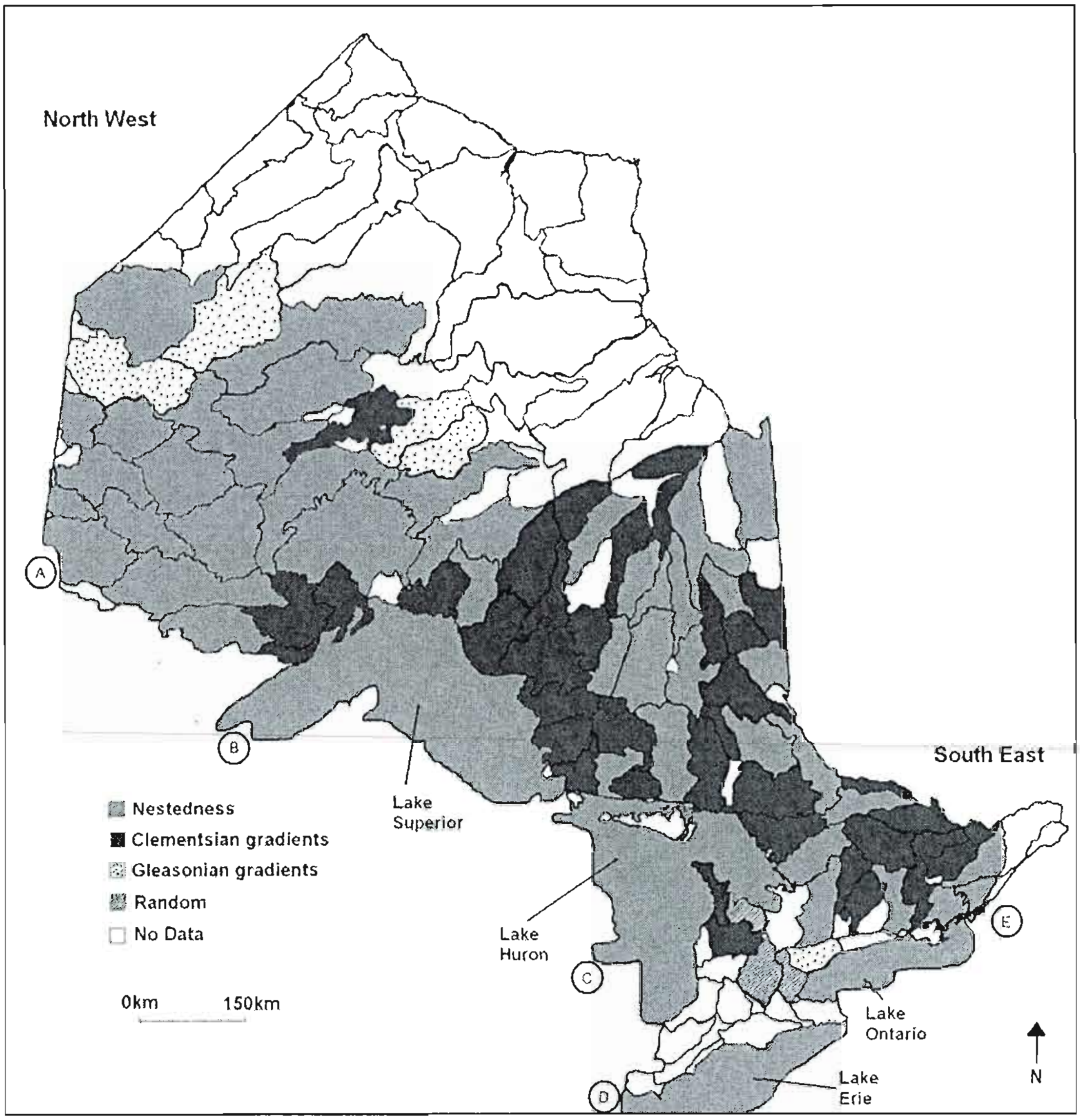

Figure 2.1 EMS results on Ontario tertiary watersheds. Map modified from Ministry of Environment (2004). The letters refers to postglacial dispersal routes. A = Glacial Lake Agassiz; B = Brule-Portage Outlet; C = Grand Valley Outlet; D= Fort Wayne; E = Champlain Outlet (Mandrak & Crossman, 1992b).

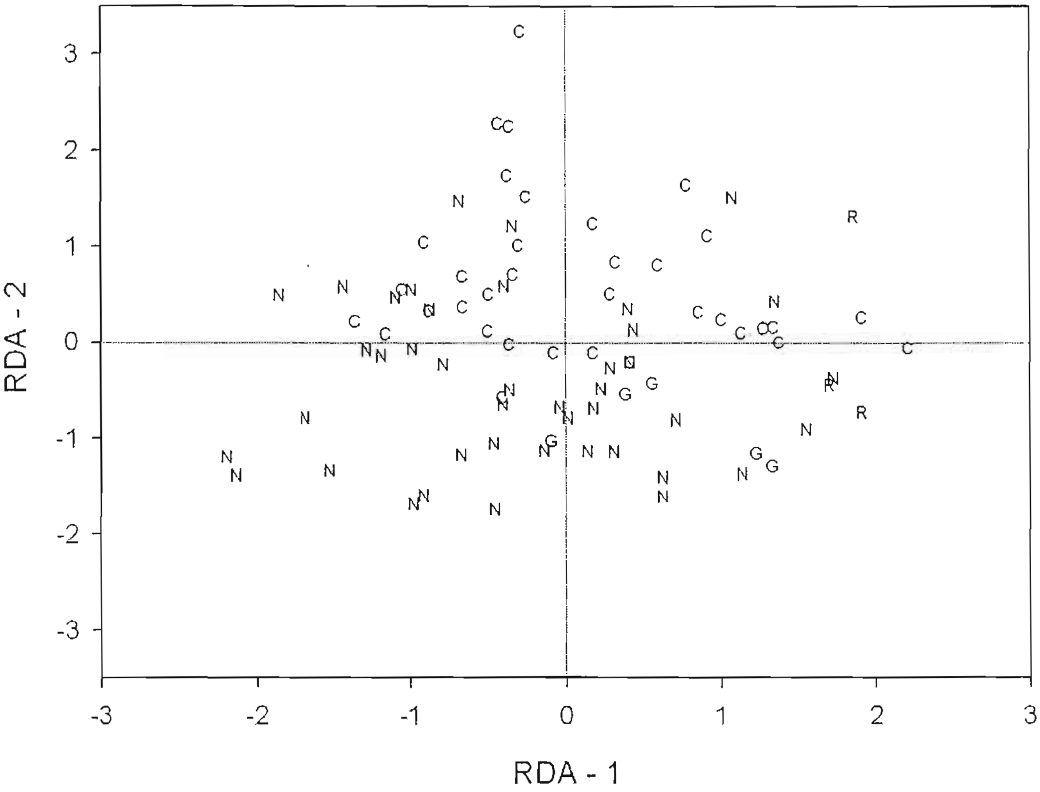

Figure 2.2 Redundancy analysis on Elements of metacommunity structure (coherence, turnover and boundary clumping) standardized values. Score coordinates of each metacommunity according to axes. N = nestedness, C = Clementsian gradients, G = Gleasonian gradients and R = random.

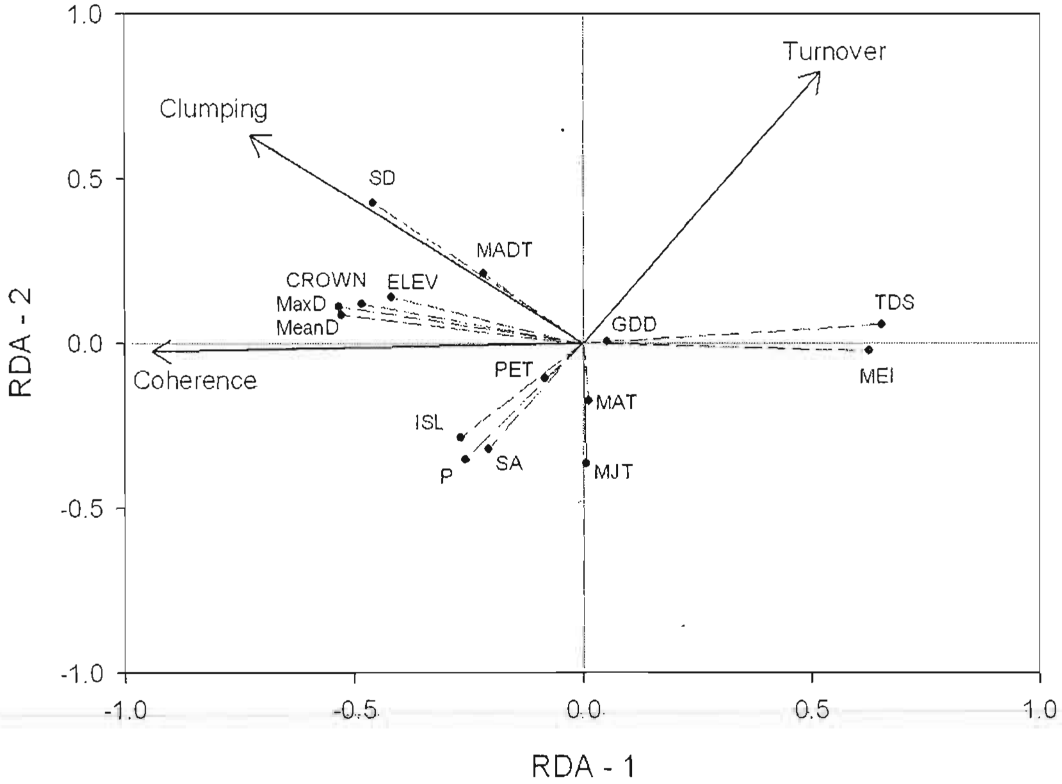

Figure 2.3 Redundancy analyses from EMS and correlation with environmental variables. Solid lines with arrows represent the EMS. Dashed lines with small circles represent the environmental variables. SA = surface area, P = shoreline perimeter, ISL = island perimeter, SD = secchi depth, MaxD = maximum depth, MeanD = mean depth, CROWN = crown canopy cover, ELEV = elevation, GDD = growing degree days, TDS = total dissolved solids, MEI = morpho-edaphic index, MADT = mean annual daily temperature, MJT = mean July temperature, MAT = mean August temperature and PET = potential evapotranspiration.

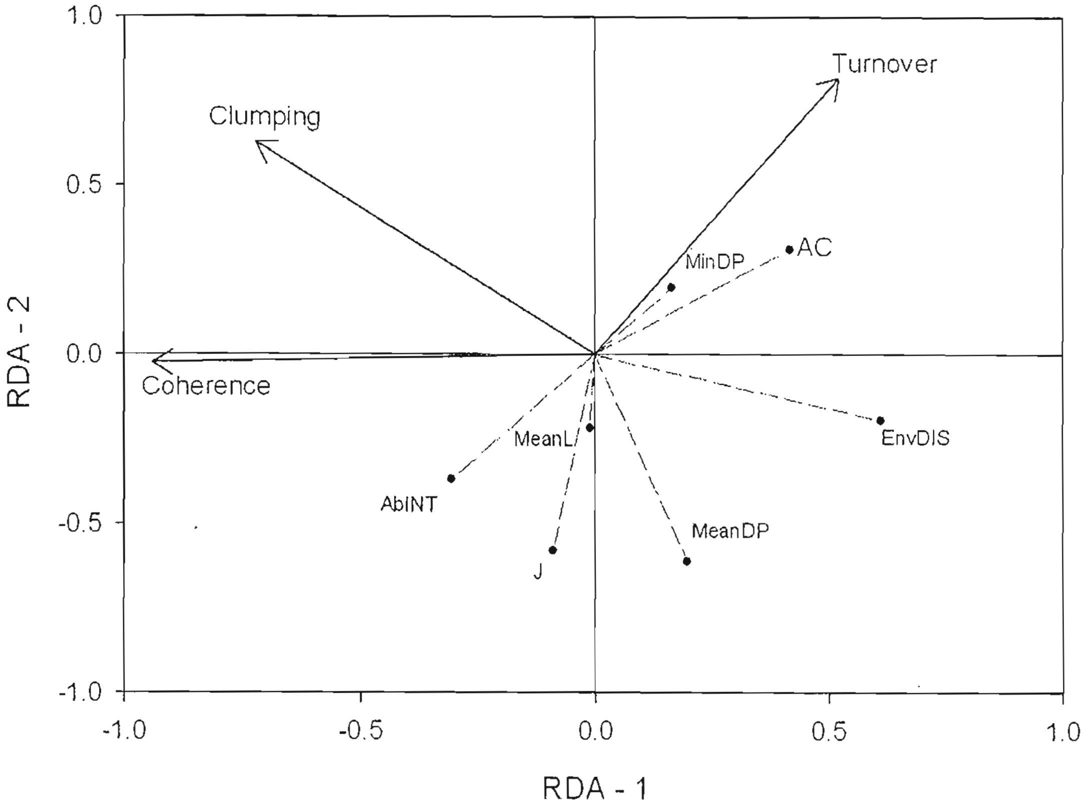

Figure 2.4 Redundancy analyses from EMS and correlation with indices. Solid lines with arrows represent the EMS. Dashed lines with small circles represent the indices. MeanL = mean latitude, AC = average connectivity, AbINT = abiotic integration, EnvDis = environmental distance, J = Jaccard index, MinDP = minimal distance from postglacial route, MeanDP = mean distance from postglacial routes.

showed negative significant correlations ($r = -0.27$ to -0.35 / $p < 0.05$). For Clementsian watersheds, only one had positive significant correlation ($r = 0.12$ / $p < 0.05$) and 4 negative significant correlations ($r = -0.16$ to -0.46/ $p < 0.05$). These results indicate that in some nested watersheds, lake connectivity present a positive effect on fish species richness (and in some affect it negatively).

2.3.2 - Species turnover at the Provincial scale

As pre-defined by the EMS analysis, turnover is the element that mainly distinguishes nestedness and Clementsian gradients and among the environmental factors considered in the analysis, lake surface area and correlated variables (e.g., shoreline perimeter, island perimeter) differed significantly (Figure 2.2 and Figure 2.3) between these two patterns; (on average, nested watersheds presented larger lakes than Clementsian ones) but climatic-related factors (e.g., temperature, PET and GDD) presented little variation across lakes within the same watershed. Thus, to further investigate the causes of species turnover, I scaled up some analyses to the entire province (i.e., across watersheds) and correlated the environmental variables at this scale with Jaccard index (i.e., inverse of species turnover) (results in Table 2.2).

I also performed a variation partitioning using mean environmental variables of each watershed as environmental predictors; mean lake geographic coordinates as spatial predictors and watershed composition as the response variable (results in Table 2.2).

The variation partitioning performed at the Provincial scale (i.e., across watersheds) indicated that the variation in species distributions was better explained by spatially autocorrelated environmental variables (e.g., $[E+S] = 29.26\%$) followed by the purely environmental ($[E] = 10.8\%$) and pure spatial components ($[S] = 3.5\%$; see Table 2.2). Energy-related variables such as potential evapotranspiration (PET),

Table 2.1 Results for variation partitioning Values presented are the average values of explained variation for all watersheds within each pattern. [E] = the fraction of variation explained solely by the environment; [S] = the unique fraction of variation explained by space; [E+S] = the common fraction of the variation shared by space and environment; [R] = residual variation.

	Fractions			
Pattern	[E]	[S]	[E+ S]	[R]
Clementsian	9,1%	2,9%	5,7%	82,0%
Nestedness	9,1%	3,5%	3,9%	83,5%
Gleasonian	1,8%	NS	0,6%	97,6%
Random	NS	NS	NS	NS

Table 2.2 Results for “selective colonization” and “species turnover at the provincial scale” sections. (A) Pearson correlations. PET = potential evapotranspiration; MADT = mean annual daily temperature; GDD = growing degree days; SA = surface area; MeanDP = mean distance from postglacial routes. All correlations in the table are significant ($p < 0.05$); r = correlation coefficient. (B) Results for variation partitioning across watersheds. All fractions are significant ($p < 0.05$). [E] = the unique fraction of variation explained by the environment; [S] = the unique fraction of variation explained by space; [E+S] = the common fraction of the variation shared by space and environment; and [R] = residual variation.

	(A)		(B)	
	Jaccard Index	Mean Latitude	Variation Partitioning	
	r	r	Fractions	% Explained Variation
PET	-0.38	-0.78	[E]	10.80%
MADT	-0.51	-0.76	[E+S]	29.26%
GDD	-0.44	-0.88	[S]	3.50%
SA	0.71		[R]	57.12%
MeDP	0.55			

mean annual daily temperature (MADT) and growing degree days (GDD) were, as expected, strongly negatively correlated with latitude (r = -0.78, -0,76 and -0.88 respectively; Table 2.2). Moreover, these variables were also strongly negatively related to the Jaccard similarity index, indicating that watersheds with more energy had higher levels of species turnover (Table 2.2). Surface area (SA) and mean distance from postglacial routes (MeanDP) were both positively related with the Jaccard index, hence watersheds that contained larger lakes and were more distant from postglacial routes had less species turnover across their lakes (Table 2.2).

2.4 – Discussion

2.4.1 - Nestedness *versus* Clementsian gradient

The results indicate that the studied metacommunities are mainly organized according to two main patterns of species distributions, namely nestedness and Clementsian gradient. The spatial distribution of these patterns was also quite clear, as Clementsian watersheds were mostly encountered in the south-eastern part of Ontario and near to the Great Lakes (Figure 2.1) and nested watersheds were the dominant pattern in north-western Ontario, but they still occurred in considerable numbers in the south-eastern region.

Lakes within watersheds at higher latitudes undergo more severe winters, with ice-cover lasting longer periods of time, with greater periods of oxygen depletion and shorter growing seasons relative to lakes occurring at lower latitudes (Magnuson *et al.*, 1998; Fang & Stefan, 2000). This process can account for the majority of watersheds classified as nested in north-west Ontario, because lakes (i.e., communities) with harsher environmental conditions preclude species with young-of-the-year that cannot attain a minimum size to survive winter starvation in the first year (Shuter *et al.*, 1980). Communities under unfavourable environmental conditions are more deterministic and, coupled with a hierarchy in breadth of species-specific tolerances, could result in a strongly nested system (Fernández-Juricic, 2002; Smith

& Brown, 2002; Chase, 2007), where the species that tolerate the broadest range of conditions would be more widespread whereas the least tolerant species would present a restricted distribution. This possibility is consistent with the "selective environmental tolerances" hypothesis (Smith & Brown, 2002). Sites (lakes) differ in their suitability due to differences in resources availability, stability, habitat diversity and environmental conditions, which affect the probability of species extinction and/or ability to establish (Kodric-Brown & Brown, 1993; Smith & Brown, 2002). In the present case, differences in suitability presented by lakes might also be a reflection of surface area, where larger lakes are more stable (Shurin *et al.*, 2010), might present more winter refuges for species to survive through this period (Magnuson *et al.*, 1998) and have higher habitat diversity (Eadie *et al.*, 1986). If there is a hierarchical relationship between lakes suitability and species capacities to survive and reproduce, nestedness is likely to occur (Azeria & Kolasa, 2008). Moreover, regions at higher latitudes have lower primary productivity, where less energy is available to be partitioned by multiple species (Mandrak, 1995; Gardezi & Gonzalez, 2008; Table 2.2). This expectation was supported, because potential evapotranspiration (PET), growing degrees days (GDD) and mean annual daily temperature (MADT), which are energy-related variables, were negatively correlated with latitude. These variables were also negatively correlated with the Jaccard similarity index (Table 2.2), suggesting that communities located in unfavourable environments are more similar, given that there is less variation in the environmental gradient (Mandrak, 1995) and thus less opportunities for specialist species to live in it (Chase, 2007; Gardezi & Gonzalez, 2008).

Nested watersheds observed higher values of abiotic integration (Figure 2.4), indicating that the environmental variables are more correlated, which might increase the likelihood of co-occurrence patterns. Another possible cause is the low degree of connectivity among lakes in nested metacommunities, which constrain poor dispersers to colonize only a few lakes (Cook & Quinn, 1995). If there is a hierarchy in the colonization ability of the species, in which more isolated lakes are colonized only by the best dispersers, and their composition represents proper subsets of subsequently less isolated lakes that are colonized by good and poor dispersers it could lead to a nested system (Cook & Quinn, 1995). This hypothesis was supported by the correlation between lake mean connectivity and lakes richness. In general, lakes that are more connected can be accessed by a greater number of species than more isolated ones (i.e., species are dispersal limited; Olden *et al.*, 2001). This could explain why nested watersheds are found at lower latitudes (Figure 2.1), and indeed, 6 out of the 11 nested watersheds with significant correlations were located in latitudes below 48°.

Some nested watersheds presented high values of boundary clumping (Figure 2.1; Figure 2.2). This pattern occurs when species disappear as a group (instead of one by one in standard nestedness pattern; Ulrich *et al.*, 2009) from richer lakes to poorer ones (see Presley *et al.*, 2010 for more details and schemes), suggesting that entire groups of species that share environmental preferences disappear when these conditions are not met. Another piece of evidence for this process is that nested watersheds have greater values of abiotic integration. For example, if in a group of species, each uses a different environmental cue to fulfill its biological needs (e.g., reproduction), they would tend to have greater levels of co-occurrence across lakes within the watershed where these environmental variables are more correlated (i.e., higher abiotic integration) than in watersheds where these variables vary more independently (i.e., lower abiotic integration). They may also disappear all together when these environmental conditions are not met, creating a clumped nested pattern (Presley *et al.*, 2010).

Clementsian watersheds were found generally in the south-east region of Ontario, near the Great Lakes (Figure 2.1). These watersheds are located at lower latitudes, observing lower winter severity (i.e., less oxygen depletion through winter under ice-cover, larger growing season), with higher productivity and being closer to post-glacial dispersal routes. As expected, Clementsian gradient watersheds presented low values of across-lakes species composition similarities (Figure 2.4). Their metacommunities were highly positively correlated with average connectivity (Figure 2.4), thus suggesting species were not dispersal-limited. Moreover, the correlation between lake connectivity and species richness (see section "selective colonization) was not significant for almost all Clementsian watersheds. If species are not dispersal-limited but lakes differ in their compositions and richness, than local environment should be acting as a filter and species are distributed according to their niches which may suggest species sorting dynamics (*sensu* Leibold *et al.*, 2004). Another possible explanation for Clementsian gradients is that pairs of competing species (i.e., "forbidden combinations"; Diamond, 1975) are not occurring independently of each other, forming "clusters of forbidden combinations" (Gilpin & Diamond, 1982). However, in order to detect the competition signal, further research is needed using a phylogenetic or trait-based approach to determine if species that inhabit those lakes are more or less similar than expected by chance in terms of their ecological niches and associated traits (Mason *et al.*, 2008).

Despite the different mechanisms invoked in the structure of the metacommunity pattern it is probable that both environmental and spatial components are important and complementary, explaining the results in variation partitioning. It unveiled variation in species composition explained by both environmental (i.e., extinction) and spatial factors (i.e., colonization) (Table 2.1); the explained proportion was larger for the former. This concurs with the results of Magnuson et al. (1998) who found that for fishes in boreal lakes, environmental factors are more important than spatial factors in explaining lake-fish composition, because extinctions are likely to occur at a faster rate than colonization events in boreal lake-fish assemblages (Magnuson *et al.*, 1998). Although the explained variance seems low (around 9% and 3.5% for environmental and spatial predictors, respectively), I argue that these result are not negligible in absolute numbers, considering the size of the dataset (n = 8911 lakes). Moreover, important environmental variables that are known to influence fish species composition, such as pH (Magnuson *et al.*, 1998; Helmus *et al.*, 2007b), were not available. Space, likewise, had low explanatory power, but it is important to point out that the connectivity metric used in this study and the spatial variables used in variation partitioning ignored key spatial predictors such as waterway connections (Olden *et al.*, 2001) and geographic barriers (Leprieur *et al.*, 2009). Waterway connections were not measured due to issues of feasibility for the amount of lakes (n= 8911) in the dataset. Moreover, many waterways are perennial, freezing in winter season. Thus, these important spatial structures should increase the explained variation of regional factors, probably enhancing the difference between spatially explained variation between nestedness and Clementsian gradient watersheds.

Finally, watersheds classified as random were highly correlated to the EnvDis index (Figure 2.4), indicating that they contain a greater environmental variation. Due to the ordination of matrices performed according to the first axis of the analysis of correspondence and because in general this axis does not represent all the existent environmental variation (Presley *et al.*, 2009), it is likely that these larger environmental gradients might be not entirely represented by this ordination method as species may be responding to other gradients, resulting in non-coherent matrices (Leibold & Mikkelson, 2002) and thus presenting random distributions (Presley & Willig, 2010).

2.4.2 - Large-scale patterns

Productivity related variables (e.g., temperature, PET and GDD) showed little correlation with the canonical axes (RDAs) that distinguished between nested and Clementsian patterns (Figure 2.3). Given that there is not enough variation in climatic variables within watersheds and because nested watersheds were distributed throughout the Province (Figure 2.1), it possibly obscured some environmental differences that might have structured these two metacommunity types. Thus, increasing the scale (see Results "Large-Scale patterns" section) shed some light on the mechanisms that produced species turnover, which is the main difference between Clementsian and nestedness, as outlined in EMS methodology. All productivity related variables were negatively correlated with latitude and were negatively correlated with similarity in species composition (Table 2.2). This indicates that increasing temperatures, growing season duration and consequently the energy available on the system increases turnover rates between communities. It is possible that a more productive environment increases the numbers of resources that can be exploited by species (Chase & Leibold, 2002). Moreover, southern Ontario has a relatively more "benign" environment (Mandrak, 1995), allowing species to specialize in the exploitation of different resources and increasing the importance of stochasticity in community assembly, which may lead to a higher species turnover among communities (Chase, 2007). In northern Ontario, only species adapted to the harsher environmental conditions can colonize lakes (Shuter *et al.*, 1980), increasing the importance of deterministic processes in community assembly (Chase, 2007), which explains the higher degree of similarity across lake communities in this region. Indeed, the variation partitioning using all watersheds (Table 2.2) pointed to this trend. The largest fraction explaining species distribution (almost 30%) across watersheds was [E + S], which is the variation in the environment that is spatially structured. However, Peres-Neto and Legendre (2010) stated that, in variation partitioning analysis, the fraction of common explained variation between space and environment can be the result of either measured spatialized environmental variables or unmeasured spatialized environmental variables. The findings of this study suggest the former, where productivity-related variables (e.g. GDD, temperature, PET) are spatially autocorrelated, increasing within a north-south gradient and affecting community assembly at a local scale and species turnover at a regional scale.

CHAPTER II

COMMUNITY PHYLOGENETIC STRUCTURE AND SPECIES NICHE: IMPLICATIONS FOR METACOMMUNITY STRUCTURE

La structure phylogénétique des communautés et la niche des espèces: Implications dans la structure des metacommunautés

3.1 - Introduction

3.1.1 - Why use patterns of species distributions?

Determining the processes that regulate species assembly within communities has been one of the main goals of community ecology (Chesson, 2000; Leibold *et al.*, 2004; Ricklefs, 2004). The general consensus is that large-scale processes (e.g., speciation, dispersal, glaciations events) dictate the number and identity of species in the regional pool whereas local processes (e.g., local abiotic and biotic factors) determine which species from the regional pool will be assembled to compose local communities (Ricklefs, 1987; Ricklefs, 2004).

One way in which ecologists have been tackling community assembly is by describing patterns of species distributions (e.g., Clementsian gradients, checkerboards, nestedness; Diamond, 1975; Hoagland & Collins, 1997; Wright *et al.*, 1998; Leibold & Mikkelson, 2002). These patterns are revealed at the metacommunity level, i.e., set of local communities potentially linked by dispersal of species dwelling locally (Leibold *et al.*, 2004), and are characterized by a site-by-species matrix of presence-absences or abundances (Leibold & Mikkelson, 2002; Presley *et al.*, 2009). It follows that species distributed in a non-random manner, should provide valuable clues about the processes that regulate community assembly (Leprieur *et al.*, 2009; Presley & Willig, 2010). For example, nested subsets (i.e., in a set of sites, composition of species-poor sites represent subsets of more specious communities; Patterson & Atmar, 1986) is characterized by low species turnover between sites (Leibold & Mikkelson, 2002) which is a product of many possible factors such as differential dispersal capabilities, differential niche breadths, selective extinctions and others (see Ulrich *et al.*, 2009 for a review). Thus, when a pattern is unveiled and given some knowledge of the studied system, the range of plausible hypotheses to test about the processes that regulate species distribution is drastically reduced (Leibold & Mikkelson, 2002; Presley *et al.*, 2009).

3.1.2 - Influence of niche on distributional patterns

Factors that influence the distribution of a species (e.g., resource affinities, species interactions; climatic tolerance) are integrated into the concept of niche (Dolédec *et al.*, 2000; Wiens & Graham, 2005; Soberón, 2007, Costa & Schlupp, 2010). Following Hutchinson (1957), the niche of a species can be represented by a n-dimensional hypervolume in which populations of a given species can persist. Abiotic factors and biotic interactions with other species define the niche hypervolume (Dolédec *et al.*, 2000). The niche can be divided into two components: the α-niche comprises the resource use of species (Soberón, 2007) whereas the β-niche consists of the range of environmental conditions that species tolerate and their dispersal abilities (Costa & Schlupp, 2010). The information about α-niche is used to disentangle two opposing assembly rules: competitive exclusion (MacArthur & Levins, 1967; Diamond, 1975) and habitat filtering (Mason *et al.*, 2007). In the former, due to competition for resources, species that co-exist in local communities present low niche overlap (i.e., species that had high niche overlap with the species that are present in the community where excluded through competition in the past). According to the second process, local environmental conditions are more important than biotic interactions, and species that share the same physiological tolerances would be more likely to co-exist, even if they present high overlap in their resource use (i.e., high niche overlap; Mason *et al.*, 2008). In contrast, the β-niche provides information to estimate species geographic distributions at broader scales (Soberón, 2007; Costa & Schlupp, 2010).

The α-niche can be estimated using species trophic position (Ingram & Shurin, 2009) and/or by a set of functional traits or patterns of phylogenetic relationships that are potentially related to their niche axes (Webb, 2000; Cavender-Bares *et al.*, 2004; Mason *et al.*, 2008; Ingram & Shurin, 2009). The β-niche is often estimated by a set of broad-scale environmental variables that characterize all the environments where the species are present, which in turn serve as an indirect surrogate for unmeasured physiological traits related to their environmental tolerances (Dolédec *et al.*, 2000; Costa & Schlupp, 2010). Moreover, species phylogenetic relatedness can be used as a proxy to infer about β-niche similarities among species as well (Wiens & Graham, 2005).

3.1.3 - Phylogenetic relatedness and community assembly

Given that species niches (α and β) are the result of processes occurring during their evolutionary history (Ackerly *et al.*, 2006), their phylogenetic relationship can serve as a proxy (but see Losos, 2008) to infer niche similarities among them (Webb, 2000; Cavender-Bares *et al.*, 2004; Helmus *et al.*, 2007a). Because species that diverged recently tend to be more ecologically similar (i.e., niche conservatism; Wiens & Graham, 2005), there may be a link between the phylogenetic relatedness of a taxa and the factors that determine the distribution of the species within that taxa. In this context, if niches are conserved (i.e., they exhibit strong phylogenetic signal), community phylogenetics have two basic predictions: (1) if closely related species are co-occurring (i.e., phylogenetic underdispersion), it suggests that environmental filtering is allowing only the species that have similar environmental tolerances to inhabit a particular community (Cavender-Bares *et al.*, 2004); (2) if species assembled in a community are more phylogenetic distantly related (i.e., phylogenetic overdispersion), it suggests that competitive interactions are precluding the species that are ecologically similar to co-occur (Cavender-Bares *et al.*, 2006). However, if niche components that are important for community assembly are convergent (i.e., low phylogenetic signal), tests for phylogenetic patterns (i.e., underdispersion or overdispersion) have a less straightforward interpretation and caution is needed when inferring processes that causes these patterns (Webb *et al.*, 2002; Losos, 2008).

3.1.4 - Integrating phylogeny and niche into metacommunity patterns

Although many theoretical and empirical advances have been made in these three venues (e.g., niche modelling, phylogenetic approaches and metacommunity paradigm; Webb, 2000; Webb *et al.*, 2002; Leibold & Mikkelson, 2002; Wiens & Graham, 2005; Cavender-Bares *et al.*, 2006; Helmus *et al.*, 2007a; Ingram & Shurin, 2009; Costa & Schlupp, 2010; Presley & Willig, 2010), there has been few attempts to combine phylogenetic and niche properties into metacommunity analyses. In this study, I propose to integrate these three approaches in order to promote an integrated view of the processes that regulate community assembly (i.e., local scale processes) and large-scale distributional patterns. As the niche of species is a product of their evolutionary history (Ackerly *et al.*, 2006), and the distribution of species are dependent on their niches (Soberón, 2007), there may be a link between the phylogenetic relationship of species and their patterns of distribution that arise from non-random associations of species across communities. In chapter one, I have described and assessed the environmental correlates of the metacommunity elements; in the present chapter, I will link these elemental patterns to phylogeny as a surrogate of niche relationships as well as other aspects of niche such as marginality and breadth.

3.1.5- Studied system

In order to combine these elements (i.e., niche, phylogeny and metacommunity), I used a dataset of lake-fish distribution from Ontario (details are given in the introduction of the thesis and in the methodology section of chapter 1). The present Ontario fish-fauna is a result of recent processes of colonization (and re-colonization) at the end of the Pleistocene glacial period where lakes and rivers were formed from the meltwaters of the retreating glacial sheet (Mandrak & Crossman, 1992). Moreover, Ontario, more precisely the southern region, has the greatest fish species richness of Canada, which is a result of postglacial dispersal, human activities and climate (Chu *et al.*, 2003). Finally, there is a strong gradient of environmental variation, with more extreme environmental conditions (i.e., harsher winter) in the north (Mandrak, 1995) which can act as a strong environmental filter for many species that do not have the physiological adaptations to survive in these conditions (Shuter *et al.*, 1980) and a more favourable climatic condition in the southern part of the province (Mandrak, 1995). These three conditions and the size of the dataset (8911 lakes) make it an ideal system to disentangle the processes of limiting similarity and environmental filtering that affect species association, which will influence community structure at local scale and ultimately yield to the distributional patterns at the metacommunity scale.

3.1.6- Chapter objectives

Here I assess phylogenetic relatedness (which was used as a proxy for the α-niche), niche indices (marginality and breadth) regarding the β-niche and the influence of environmental gradients on the patterns of species distributions, at three hierarchical spatial scales: lake, watershed and the whole province. The specific goals are as follows: 1) Determine how patterns of phylogenetic structure change across scales (e.g., local *versus* regional); 2) Asses the relationship between phylogenetic relatedness, niche community structure and environmental gradients across watersheds; and 3) Test for differences among metacommunity patterns regarding these factors.

3.2 - Methodology

The data set used here is the same as in chapter one. I used all 8911 lakes divided into 85 tertiary watersheds (Cox, 1978) and applied the metacommunity patterns observed at each watershed as detected by the EMS framework (methodology section, chapter 1). In the next sections, I describe the framework used to measure niche properties, phylogenetic relatedness and environmental gradients. A summarized description of these indices is presented in Table 3.2.

3.2.1 - Niche indices

In this study, properties related to the β-niche were estimated using the abiotic factors (see Table 1.1 for a list of environmental variables) from all lakes where each species were present. I used two niche aspects, namely niche breadth (range of environments used by the species) and niche marginality (level in which an environment used by a species differs from what is available in the entire landscape). Because I used the environment of lakes to measure species preferences, species would tend to have more similar niches due to their common co-occurrence across many lakes within the same watershed. In order to reduce this potential bias, I calculated niche breadth and marginality for each watershed separately; however, estimates were based on all lakes across the entire data set except the ones for the specific watershed at hands. First, all environmental variables were standardized (mean = 0 and variance = 1) across the reduced data set (i.e., without the lakes of the target watershed) and then, a pairwise Euclidian distance matrix across lakes was calculated. For each species, niche breadth was calculated as the average Euclidian distance across all their occupied lakes. Niche breadths across species were averaged within watersheds; a watershed that presents a large mean niche breadth suggests that is composed, in average, by more generalist species whereas if it exhibits a small value, it indicates that is composed by more specialist species.

The second niche measure, marginality, was based on the average Euclidian distance across all occupied lakes for a given species (except the ones for the target watershed) and the median of all environmental variables across lakes (i.e., occupied or non-occupied). I then calculated the Euclidian distance among species marginalities and averaged them across species within each watershed. Large mean values indicate that the watershed is composed by species with a relatively small niche overlap among them, whereas small mean values indicate that the watershed is composed by species with relatively large niche overlap. Note that because all environmental variables were used in the calculation of Euclidian distances, breadth and marginality represent multivariate measures of niche.

3.2.2 - Characterizing environmental gradients

Akin to a species niche breadth and marginality, I calculated environmental breadth (variance) and environmental marginality (uniqueness) for each watershed. These measures aim at assessing whether gradient variability may be driving differences in community assembly patterns across watersheds. The two indices were calculated based on the standardized (mean=0 and variance =1) environmental matrix across the entire dataset and also the standardized environmental matrix within each watershed. The former are hereafter referred as to environmental breadth (across watersheds) and environmental marginality (across watersheds) and the latter are hereafter referred as to environmental breadth (within watersheds) and environmental marginality (within watersheds). For each lake, environmental breadth was calculated as the mean Euclidean distance across lakes within watersheds. Environmental marginality was based on the average Euclidian distance between all lakes within a watershed and the average of all environmental variables across all lakes across all watersheds. Because all variables were standardized prior to analysis, the mean of each environmental variable is zero and hence the marginality of each lake is simply the square root of the sum-of-squared values across variables. I then averaged the marginality values for all the lakes within any given watershed. Standard deviation of marginalities across lakes within a watershed was used as a measure of environmental overlap across lakes.

3.2.3 – Phylogenetic tree

In order to assess the phylogenetic relationships between the 53 fish species in this study, a phylogenetic tree was created following Hubert et al. (2008). The genomic information for each species was obtained at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). Among other information, this database contains species names, voucher data, collection record, barcode sequence, PCR primers and trace files (Hubert *et al.*, 2008). The GenBank accession number for each species is shown in Table 3.1. For each species, the molecular tag utilized was a 652-bp segment of 5' region from a mitochondrial cytochrome c oxidase I (COI) gene, which was coded in format FASTA (Hubert *et al.*, 2008). Mitochondrial DNA (mtDNA) was used due to its high mutation rate and small effective population size, which make it a reliable genome about evolutionary processes and patterns (Brown *et al.*, 1979).

Phylogenetic analysis was performed using Mega4 v4.0 (Takamura *et al.*, 2007), where the FASTA format sequence codes were imputed. In order to calculate the sequence divergence between species, the Kimura 2-parameter (K2P) model (Kimura, 1980) was applied and the graphic representation of the evolutionary tree (Figure 3.1) was created using the mid-point rooted Neighbour-joining tree technique (Saitou & Nei, 1987) on K2P distances. The evolutionary tree was converted into a phylogenetic covariance matrix based on nodal covariance between species (Webb, 2000; Peres-Neto, 2006). Nodal distance methods are subject to some limitations, such as dependence on clade richness, i.e., two species drawn randomly from a species-rich clade will appear more related than two species drawn from a species-poor clade. However, this problem is more likely to bias statistical tests between pairs of species, and having less effect when communities are increasingly richer (Webb, 2000). Moreover, the clades here have, in general, similar number of species (Figure 3.1), thus I expect that a high number of nodes shared by two species will well reflect species that have a more recent common ancestor and thus are more ecologically similar (Webb, 2000; Helmus *et al.*, 2007b).

Finally, although the phylogenetic covariance matrix unrealistically assumes that distances between species are a function of a single trait under "Brownian motion" evolution (Felsenstein, 1985), it gives a way to translate evolutionary history of species into a measure of phylogenetic relatedness (Peres-Neto, 2006; Helmus *et al.*, 2007a), which in turn can be related to assembly processes at the community level

(Cavender-Bares *et al.*, 2004). The phylogenetic covariance matrix was used to calculate the indices described in the next section.

3.2.4 - Phylogenetic indices

For lakes within each watershed, two indices were calculated: mean nearest taxon distance (MNTD) and mean pairwise distance (MPD). The first one measures the average relatedness between each species and its closest relative whereas the other assesses the average relatedness between all co-occurring species in a community (Webb, 2000). See below:

$$MNTD = \frac{\sum DN_{ij}}{n}$$

where DN_{ij} is the phylogenetic distance between species i and its nearest relative j and n is the number of species.

$$MPD = \frac{\sum D_{ij}}{n}$$

where DN_{ij} is the phylogenetic distance between species i and species j.

Thus, if we assume that phylogenetic distance is a proxy of ecological similarity (Webb, 2000; Peres-Neto, 2004), MNTD is a proxy of how ecologically similar two co-occurring species are and MPD is a proxy of how ecologically similar is an entire community. In this analysis, lakes containing only one species were removed.

3.2.5 - Null models for phylogenetic data

3.2.5.1 – Regional null model

In order to test whether species that colonized a particular watershed presented or not a phylogenetic pattern (i.e., overdispersed, random or underdispersed) I performed (Figure 3.2) the following null model (hereafter referred to as regional null model). For each watershed, a number of species equal to the observed were randomly chosen from all species present in the entire dataset (i.e., regional pool; n = 53) and both indices (i.e., MNTD and MPD) were calculated. The regional null model allows us to look for global trends of species colonization capacities and whether this property is related to phylogenetic structure. This null model was performed based on 999 random species samples (from the entire regional pool) and the values of the indices from each permutation were compared to the observed values. Probability values were calculated based on the number of random values equal or smaller than the observed value divided by 1000 (i.e., 999+1, which includes the observed value as a possible outcome of the randomization test) or on the number of random values equal or larger than the observed value. The test was one-tailed, with an alpha-level of 0.05, where $p < 0.05$ was used as indicative of phylogenetic underdispersion whereas $p > 0.95$ was indicative of phylogenetic overdispersion. If the p-value fell between 0.05 and 0.95, species present at any given watershed were considered to be phylogenetically non-patterned. This null model was used to assess whether species sharing similar environmental tolerances or dispersal abilities (assuming that these ecological properties are conserved through evolutionary history; Wiens & Graham, 2005) are colonizing the same watersheds or not.

3.2.5.2 –Local null model

The second null model (see Figure 3.2) represents a filter at the local community scale (i.e., within lakes). Within each watershed, 999 random species samples were drawn for each lake, with the observed number of species in each lake held constant, but the identity of species drawn randomly. Note, however, that here, instead, the species pool was based only on the species present in that particular watershed. Thus, the assumption here is that species in a watershed are capable of colonizing every lake and what determine their successful establishment depends on

Table 3.1 Species name, prevalence (number of lakes present) and GenBank accession number.

Species	Prevalence	GenBank n°	Species	Prevalence	GenBank n°
Acipenser fulvescens	45	EU524392.1	*Semotilus atromaculatus*	901	EU525136.1
Salvelinus fontinalis	1788	EU522409.1	*Semotilus corporalis*	96	EU525145.1
Salvelinus namaycush	1753	EU522422.1	*Margariscus margarita*	1009	EU524128.1
Coregonus clupeaformis	2018	EU523957.1	*Ameiururs nebulosus*	1154	EU523909.1
Coregonus artedi	2083	EU523939.1	*Noturus gyrinus*	50	EU525043.1
Prosopium cylindraceum	53	EU524288.1	*Anguilla rostrata*	40	EU524440.1
Esox lucius	4284	EU524578.1	*Fundulus diaphanus*	110	EU524058.1
Umbra limi	163	EU522446.1	*Lota lota*	1378	EU524753.1
Catostomus catostomus	482	EU524462.1	*Culaea inconstans*	1117	EU524532.1
Catostomus commersoni	6430	EU524478.1	*Pungitius pungitius*	451	EU524319.1
Moxostoma anisurum	53	EU524846.1	*Percopsis omiscomaycus*	481	EU524261.1
Moxostoma macrolepidotum	272	EU524889.1	*Ambloplites rupestris*	1337	EU524407.1
Phoxinus eos	1659	EU525058.1	*Lepomis gibbosus*	1930	EU524714.1
Phoxinus neogaeus	816	EU525064.1	*Lepomis macrochirus*	225	EU524732.1
Couesius plumbeus	564	EU524523.1	*Lepomis megalotis*	68	EU524124.1
Hypbognathus hankinsoni	71	EU524081.1	*Micropterus dolomieu*	1647	EU524810.1
Notemigonus crysoleucas	1180	EU524930.1	*Micropterus salmoides*	719	EU524132.1
Notropis atherinoides	253	EU524950.1	*Pomoxis nigromaculatus*	101	EU524285.1
Luxilus cornutus	1029	EU524768.1	*Perca flavescens*	5717	EU524240.1
Notropis heterodon	166	EU524981.1	*Sander Canadensis*	132	EU524373.1
Notropis heterolepis	1914	EU524999.1	*Sander vitreus*	2590	EU524374.1
Notropis hudsonius	1720	EU525003.1	*Etheostoma exile*	1783	EU524024.1
Notropis volucellus	737	EU524183.1	*Etheostoma nigrum*	1253	EU524045.1
Pimephales notatus	1079	EU525076.1	*Percina caprodes*	595	EU524246.1
Pimephales promelas	1362	EU525085.1	*Cottus bairdi*	498	EU522459.1
Rhinichthys atratulus	99	EU524322.1	*Cottus cognatus*	235	EU524511.1
Rhinichthys cataractae	302	EU524323.1			

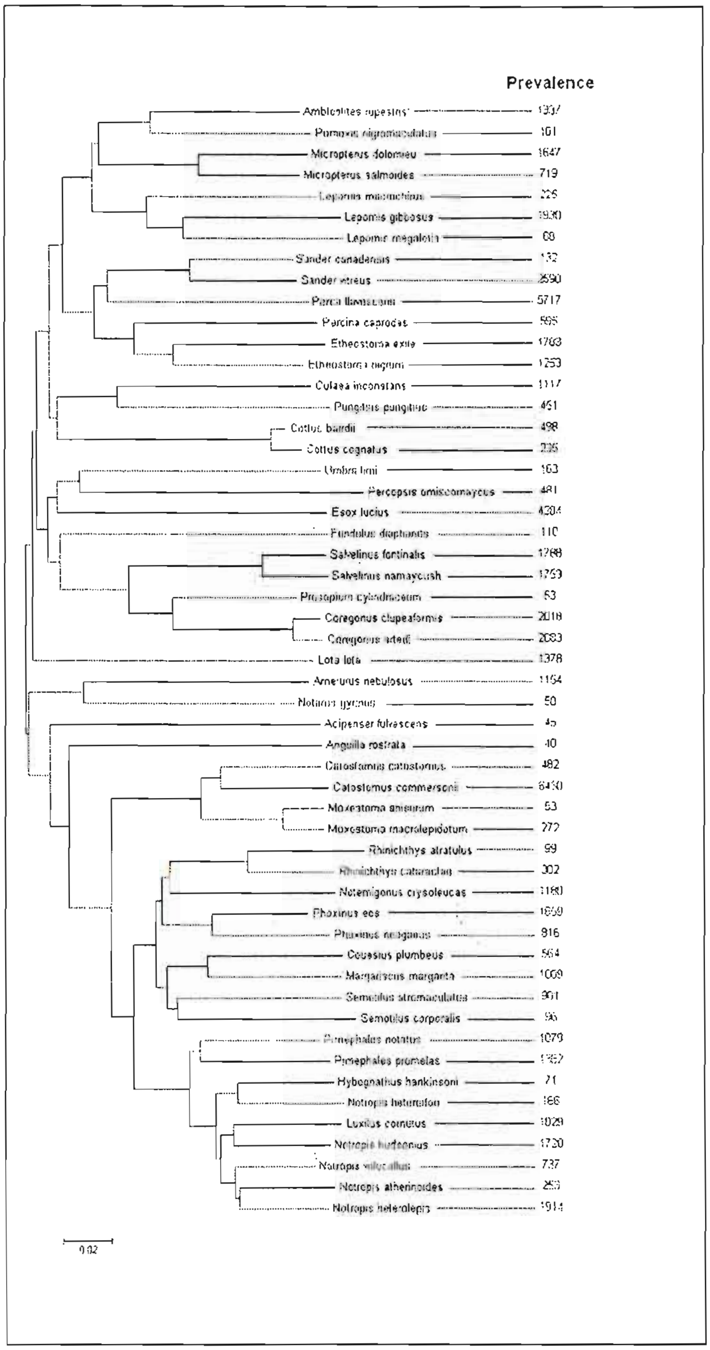

Figure 3.1 Phylogenetic tree of the 53 extant species created using the mid-point rooted Neighbour-joining tree technique (Saitou & Nei, 1987) applied on K2P distances (Kimura, 1980). Prevalence = number of lakes where the species is present.

their capabilities to cope with local abiotic and biotic factors. Moreover, in permuted sets, I constrained lake species richness to be equal to the observed given that fish species richness is highly correlated with lake size (e.g., Barbour & Brown, 1974; Eadie *et al.*, 1986). Indices (i.e., MNTD and MPD) were calculated for each random drawn. As earlier, the probability of rejection was estimated as the proportion of the 999 null communities with values more extreme than the observed value. Significance was assessed based on a one-tailed test: if $p < 0.05$ it indicated that co-occurring species were phylogenetically underdispersed, if $p > 0.95$ it indicated that co-occurring species were phylogenetically overdispersed and if *p*-value fell between 0.05 and 0.95, it suggested that there was no evident phylogenetically-related process structuring patterns of species co-occurrence.

MPD and MNTD values were standardized ($MNTD_{STAND}$ and MPD_{STAND} in the equation bellow) to allow for a better contrast across lakes having different number of species as follows (Gotelli & Mcabe, 2002):

$$MNTD_{STAND} = \frac{MNTD_{OBS} - mean(MNTD_{RND})}{sd(MNTD_{RND})}$$

$$MPD_{STAND} = \frac{MPD_{OBS} - mean(MPD_{RND})}{sd(MPD_{RND})}$$

where $MNTD_{OBS}$ and MPD_{OBS} represent the observed value of MNTD and MPD, respectively, $MNTD_{RND}$ and MPD_{RND} represent the random value estimated by the null model for each index and *sd* stands for standard deviation. The indices of each lake were standardized using the null distribution of the local null model. For the metacommunity scale (i.e., for each watershed), I averaged the values of standardized indices from all lakes present at each watershed. Negative values of standardized indices indicate an overall tendency to phylogenetic attraction (i.e., underdispersion), whereas positive values suggest trends for phylogenetic repulsion (i.e., overdispersion).

3.2.6 - Statistical Analyses

3.2.6.1 - Provincial scale patterns – Relation between environment and indices

In order to assess the regional patterns in terms of how watershed environment structures community assembly, I used watersheds (metacommunities) as observations (i.e., values for all statistics were averaged or generated at the watershed level). First, environmental variables (see chapter one) were log-transformed when necessary to assure assumptions of normality (Legendre & Legendre, 1998) and averages across lakes were used. Then, I reduced the number of environmental predictors by conducting a principal components analysis (Legendre & Legendre, 1998) on the 15 environmental variables. I used the two first principal components (PC) which accounted for most of the environmental variability among watersheds. Finally, I performed multiple regression models using the scores of the two first principal components (i.e., PC-1 and PC-2) as predictors and each index as response variable (i.e., niche marginality, niche breadth, environmental marginality (within and across watersheds), environmental breadth (within and across watersheds), MNTD and MPD).

3.2.6.2 - Provincial scale patterns – Variation in phylogenetic structure, community niche structure and environment across watersheds

In order to explore the phylogenetic basis of community structure I calculated Pearson correlations between phylogenetic and niche indices. Moreover, to understand how properties of the environmental gradients within watersheds, such as their uniqueness (i.e., marginal *versus* common environments) and their range of environmental variation, may affect community niche and phylogenetic structure, Pearson correlations between environmental indices (i.e., marginality and breadth)

Table 3.2 A short description for the indices used throughout this study.

Index	Description
Environmental marginality (within watersheds)	This index measures the difference between a lake environment and the mean environment from the landscape. The environmental matrix is standardized based on lakes within the target watershed.
Environmental marginality (across watersheds)	This index measures the difference between a lake environment and the mean environment from the landscape. The environmental matrix is standardized based on lakes from the entire Province.
Environmental breadth (within watersheds)	This index measures the variance (i.e., size) of the environmental gradient of any given watershed. Environmental matrix is standardized based on lakes within the target watershed.
Environmental breadth (across watersheds)	This index measures the variance (i.e., size) of the environmental gradient of any given watershed. Environmental matrix is standardized based on lakes from the entire Province.
Niche marginality	This index measures the difference between a species mean environment and the mean environment from the entire Province. Estimates were based on all lakes except the ones within the target watershed.
Niche breadth	This index measures the range of environments used by the species. Estimates were based on all lakes except the ones within the target watershed.
MPD	This index measures the average relatedness between all co-occurring species within a community (i.e., lake). The standard value is calculated using the null distribution of the local null model.
MNTD	This index measures the average relatedness between each species and its closest relative within a community (i.e., lake). The standard value is calculated using the null distribution of the local null model.

and community structure indices (i.e., phylogeny and niche) were performed. Correlations were calculated across watersheds and the results for this section are summarized in Table 3.6.

3.2.6.3 - Watershed scale patterns – Differences in species and environmental properties between metacommunity patterns

Differences regarding all indices at the watershed level (average values of lakes present in each watershed) were analyzed using one-way ANOVAs. In order to remove the effects of geographic positioning on these analyses, I also performed an ANCOVA on the same indices using latitude as a covariate due to the presence of nestedness watersheds both in south-eastern and north-western Ontario (see results chapter one).

3.3 – Results

3.3.1 - Regional and local null models

The results of both null models (local and regional) are summarized in Table 3.3. The regional null model suggests that several metacommunities are phylogenetic underdispersed regarding both indices (MNTD n=18; MPD n=10), but most watersheds were not significantly phylogenetically structured. Conversely, the local null model, which represented a small scale filter at the lake level indicates that several metacommunities display phylogenetic overdispersion (MNTD n=25; MPD n=21) in addition to phylogenetic underdispersion (MNTD n=9; MPD n=15), though several were found as phylogenetically unstructured (MNTD n=51; MPD n=49). The principal components analysis indicates the presence of two major environmental gradients (Table 3.4). PC-1 and PC-2 accounted for 44.36 % and 26.72% of the total variation, respectively. PC-1 represents a gradient composed mainly by surface area and energy related variables: watersheds containing larger lakes and colder

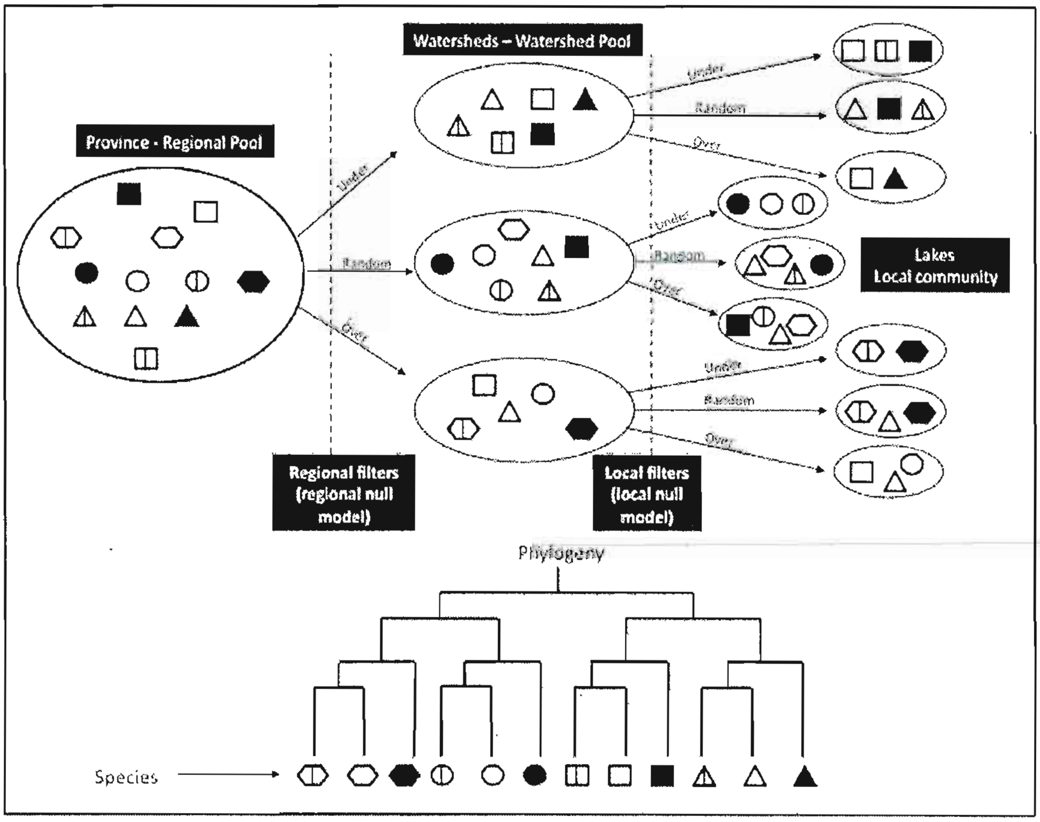

Figure 3.2 Conceptual framework for the regional and local null models, which represent the filters that species need to surpass in order to assemble in local communities. The regional filter represents broad-scale factors (e.g., climate variables and/or dispersal limitation such as geographic barrier). The local filter represents local dispersal limitation (e.g., community isolations), local environment (e.g., pH) and/or biotic interactions (e.g., competition). Species are represented by symbols. "Under" stands for underdispersion (i.e., phylogenetic clustering), "Over" stands for phylogenetic overdispersion (i.e., phylogenetic evenness) and "Random" stands for random phylogenetic structure.

temperatures (i.e., lower energy) load positively whereas watersheds containing smaller lakes and higher temperatures load negatively (Table 3.4). PC-2 represents a gradient for water clarity (high positive loadings from SD), lake morphometry variables (i.e., depth measures) and energy related variables (Table 3.4). Watersheds containing deep lakes, clear water and high energy load positively and watersheds containing shallow lakes, turbid water and low energy load negatively with PC-2. Both PCs presented also moderate positive loadings for elevation.

3.3.2 - Environmental gradients, phylogenetic structure and community niche structure at the Provincial scale

Environmental marginality (across watersheds) was negatively associated with PC-1 ($\beta = -0.29$, $p = 0.005$), indicating greater environmental overlaps in low-energy watersheds with larger lakes and lower environmental overlaps (i.e., more heterogeneous environments) in high-energy watersheds containing smaller lakes. Environmental marginality (across watershed) was positively associated with PC-2, emphasizing the positive relation between environmental heterogeneity (i.e., low overlap) and energy-related variables. However, neither PCs were significant predictors of environmental marginality (calculated within watersheds) index (PC-1, $p = 0.60$; PC-2, $p = 0.20$). These results suggest greater environmental homogeneity within than across watersheds. Environmental marginality is more evident across watersheds and significantly related with southern high-energy watersheds, which probably have more marginal environments (i.e., warmer) compared to the "average environment" of the entire province. Environmental breadth (within watersheds) was only positively related with PC-2 ($\beta = 0.36$, $p = 0.0005$) indicating a positive relationship between the size (i.e., variance) of the environmental gradient and energy-related variables. Neither PCs were significantly related to environmental breadth (across watersheds) (Table 3.5). As for the phylogenetic structure, both PC-1 and PC-2 exhibited an opposite relationship with MPD, with PC-1 being positively associated ($\beta = 0.48$, $p < 0.0001$) and PC-2 negatively associated ($\beta = -0.39$, $p <$

0.0001) (Table 3.5). In the case of PC-1, the result can be interpreted as communities containing, in average, species more distantly related (i.e., higher MPD) within larger lakes in low-energy watersheds to communities containing on average more closely-related species (i.e., lower MPD) within smaller lakes in high-energy watersheds. PC-2 again emphasized the negative relationship between MPD and energy-related variables (i.e., the greater the energy available in the environment the more the community is phylogenetically clustered). Both PCs were negatively related to MNTD (PC-1, $\beta = -0.34$, $p < 0.0001$; PC-2, $\beta = -0.21$, $p = 0.039$), indicating that in larger lakes the average phylogenetic distance between species and their closest species is smaller than in progressively smaller lakes, but the result is less clear when analyzing the relationship of this index (MNTD) and energy (e.g., growing degree-days, mean annual daily temperature and others), because MNTD is negatively associated with both PCs and they load in opposite ways for energy-related variables. Niche breadth showed a negative relationship with both PC-1 ($\beta = -0.33$, $p = 0.00016$) and PC-2 ($\beta = -0.53$, $p < 0.0001$), indicating that species having smaller niche breadths inhabit larger lakes (Table 3.5); again, for energy, the pattern was not clear because niche breadth was also negatively related to both PCs, which load in opposite ways for energy-related variables. Finally, niche marginality was negatively related with PC-1 ($\beta = -0.59$, $p < 0.0001$) and positively related with PC-2 ($\beta = 0.62$, $p < 0.0001$) (Table 3.5), suggesting that species with higher overlap (i.e., less distant) in their environmental preferences (β-Niche) tend to co-exist in low-energy watersheds within larger lakes.

Results for correlations between environmental indices *versus* phylogenetic and niche indices are summarized in Table 3.6 as well as correlations between phylogenetic *versus* niche indices. Among all environmental indices, only environmental marginality (across watersheds) was significantly correlated with niche marginality ($\beta = 0.37$, $p = 0.001$), suggesting that watersheds presenting more marginal (i.e., extreme) environments, in contrast with the "average environment" of the Province, are inhabited by species that also present marginal environmental preferences (i.e., marginal β-niches). Niche breath was only significantly correlated with environmental breadth (within watersheds; $\beta = -0.24$, $p = 0.022$). MPD was only significantly correlated to environmental marginality (across watersheds) ($\beta = -0.33$, $p = 0.002$), indicating that watersheds presenting more marginal environments were composed of communities structured by phylogenetic underdispersion. MNTD was associated with environmental marginality (across watersheds) ($\beta = -0.26$, $p = 0.013$) and with environmental breadth (across watersheds) ($\beta = -0.44$, $p < 0.0001$; Table 3.6). The first association corroborates previous results, suggesting a decrease in community phylogenetic distance (i.e., phylogenetic underdispersion) as environments within watersheds become more marginal (in relation to the landscape average environment). The second correlation suggests a decrease in community phylogenetic distance (i.e., phylogenetic underdispersion) as the "size" (i.e., variance) of the environmental gradient increase.

The contrast between phylogenetic structure and species niche (i.e., marginality and breadth) indicates a positive correlation between MPD and niche breadth ($\beta = 0.23$, $p = 0.027$) and a negative one with niche marginality ($\beta = -0.62$, $p < 0.001$; Table 3.6). Thus, watersheds composed of communities with species presenting greater differentiation in their niches have a lower average community phylogenetic distance. This suggests that communities composed by different groups of closely related-species occupy different portions of the niche-space available. Moreover, watersheds composed of communities with species presenting broader niches have a larger average community phylogenetic distance. This trend was also supported by the positive correlation between niche breadth and MNTD ($\beta = 0.61$, $p < 0.0001$; Table 3.6), suggesting that, in average, communities that have a tendency for phylogenetic overdispersion (i.e., species are more distantly related), are composed by species that have broader niches (i.e., generalist). This indicates that generalist species (i.e., greater niche breadth) are evenly-distributed across the many

Table 3.3 Results for the regional and local null models: Number of watersheds per type of phylogenetic structure. "Under" stands for phylogenetic underdispersion, "Over" stands for phylogenetic overdispersion. N.S. = non significant (*p*-values between 0.05 and 0.95), which is interpreted as a random phylogenetic structure.

Null Model (Phylogenetic Index)	Number of watersheds		
	Under ($p \leq 0.05$)	N.S.	Over ($p \geq 0.95$)
Regional null model (MNTD)	18	67	0
Regional null model (MPD)	10	75	0
Local null model (MNTD)	9	51	25
Local null model (MPD)	15	49	21

Table 3.4 Principal components analysis on environmental variables. Variables loadings represent the correlation coefficient between each variable and the principal component (i.e., PC-1 and PC-2). See introduction (page 4) for significance of abbreviations for environmental variables.

	PC-1	PC-2
EigenValue	6.65	4
Variance Explained (%)	44.36	26.72

Variables	Variables Loading	
MaxD	0.542737	0.703877
MeanD	0.393744	0.809226
ELEV	0.484218	0.331714
SD	0.187726	0.702508
PET	-0.709209	0.620995
SA	0.790989	-0.023441
P	0.790791	0.06324
ISL	0.722936	0.165447
GDD	-0.824655	0.488282
TDS	-0.690928	-0.510892
MEI	-0.714173	-0.609143
CROWN	0.831263	0.032296
MADT	-0.599261	0.587034
MJT	-0.652116	0.535517
MAT	-0.718375	0.577706

Table 3.5 Multiple regression models using indices as response variables and both PCs (e.g., PC-1 and PC-2) as predictors. β represent the regression coefficients. The α-level used was 0.05; *p*-values in bold are significant.

Indices			PC-1		PC-2	
	F (2.82)	R^2 adjusted	β	*p*	β	*p*
Environmental marginality (within watersheds)	0.96	-0.0008	-0.05	0.6	-0.14	0.2
Environmental marginality (across watersheds)	6.86	0.12	-0.29	**0.005**	0.23	**0.02**
Environmental breadth (within watersheds)	6.53	0.11	0.05	0.57	0.36	**0.0005**
Enviromnental breadth (across watersheds)	2.21	0.02	0.16	0.12	0.15	0.16
MPD	25.95	0.37	0.48	**<0.0001**	-0.39	**<0.0001**
MNTD	7.82	0.13	-0.34	**0.001**	-0.21	**0.039**
Niche marginality	121.93	0.74	-0.59	**<0.0001**	0.62	**<0.0001**
Niche breadth	27.67	0.38	-0.33	**0.00016**	-0.53	**<0.0001**

Table 3.6 Pearson correlations between indices calculated at the watershed level. The top of the table refers to the relationship between watershed environmental properties and species properties (both niche and phylogenetic structure). The bottom of the table refers to the niche-phylogenetic structure relationship. All indices were calculated for each community (i.e., lake) and the average value of all lakes was taken as a measure for any given watershed. Numbers represent coefficient of correlations between indices and significant values ($p \leq 0.05$) are in bold.

	Niche indices		**Phylogenetic indices**	
Environmental indices	Niche marginality	Niche breadth	MPD	MNTD
Environmental marginality (within watersheds)	-0.05	0.08	-0.05	-0.08
Environmental marginality (across watersheds)	**0.37**	-0.09	**-0.34**	**-0.27**
Environmental breadth (within watersheds)	0.21	**-0.25**	-0.12	0.01
Environmental breadth (across watersheds)	0.01	-0.2	-0.03	**-0.44**

	Phylogenetic Indices	
Niche indices	MPD	MNTD
Niche marginality	**-0.62**	-0.1
Niche breadth	**0.24**	**0.62**

clades of the phylogenetic tree (Figure 3.1). Indeed, prevalence (number of lakes in which a species is present; Figure 3.1 and Table 3.1) supports this result because a great number of sister species are frequently composed by a common (present in many lakes) and a rare species (present in few lakes). Thus, when a lake is composed by many generalist species, there is a greater chance that these species are distantly related, which will increase the overall community phylogenetic distance.

3.3.3 - Differences between nestedness and Clementsian gradients regarding indices

ANOVAs revealed little differences between nestedness and Clementsian gradients. Environmental marginality and breadth (both, across and within watersheds) were not different between the two patterns (Figure 3.3), indicating that the amount of environmental variation and environmental overlap within a watershed is not related to the pattern uncovered by the EMS analysis. Relative to the phylogenetic indices, nested watersheds presented significant higher values of MPD (F=4.5, p=0.036; Figure 3.4), suggesting that in general, lakes in nested watersheds present greater phylogenetic distances across co-occurring species. MNTD was also significant (F=11.991, p<0.001) but the result was the reverse, in which Clementsian watersheds exhibited greater values (Figure 3.4). Thus, in Clementsian watersheds lakes the average phylogenetic distance between the species and their closest relative is greater than within lakes of nested watersheds. In relation to the niche indices, the only significant difference between Clementsian and nestedness watersheds was regarding the niche breadth (F=5.622, p=0.02), where species of Clementsian watersheds present an average niche breadth greater than species inhabiting nested watersheds (Figure 3.4). Niche marginality was not different between the two patterns (F=0.55, p=0.44; Figure 3.4). The results from the ANCOVA, using mean latitude as covariate, maintained almost all findings but for MPD. After accounting for spatial location, MPD turned out not to be significant (F=0.6191, p=0.43), indicating that variation for this measure could be explained by latitude and not necessarily by the structuring metacommunity process (i.e., nestedness *versus* Clementsian).

3.4 - Discussion

3.4.1 - Regional and local filters

The main results of this study show a sharp contrast in community phylogenetic structure across scales. At the regional scale, watersheds were colonized by species that are more closely related than expected by chance (i.e., phylogenetic clustering) suggesting environmental filtering (Cavender-Bares *et al.*, 2004). Note that at this scale, no watershed presented a pattern of phylogenetic overdispersion. This result suggests that species from some clades were either incapable to tolerate the environmental conditions of these watersheds (Mandrak, 1995; Gardezi & Gonzalez, 2008) or did not have enough capacity (or time) to disperse into them (Ingram & Shurin, 2009). Moreover, this may be the case, as 78% of the watersheds that were phylogenetically clustered for either MPD or MNTD were localized in latitudes above 47° whereas only 22% watersheds presented this pattern at latitudes lower than 47°. High-latitude watersheds present harsher winters, with shorter growing seasons and lower temperatures (Mandrak, 1995), which can prevent species without the necessary physiological adaptations to survive in these conditions (Shuter *et al.*, 1980). Another possible explanation for this pattern of underdispersion is the increased importance of environmental filtering at larger scales because greater environmental variation is encompassed at greater spatial scales (i.e., there is more environmental variation across a watershed than within a lake), providing opportunities for species to sort themselves across these broad environmental gradients (Cavender-Bares *et al.*, 2006). Finally, phylogenetic clustering at large spatial scales can arise from biogeographic processes of local radiation and limited dispersal (Pennington *et al.*, 2006), such as in Hawaii islands (e.g., *Drosophila*; Carson & Kaneshiro, 1976). However, this hypothesis is unlikely the case in Ontario

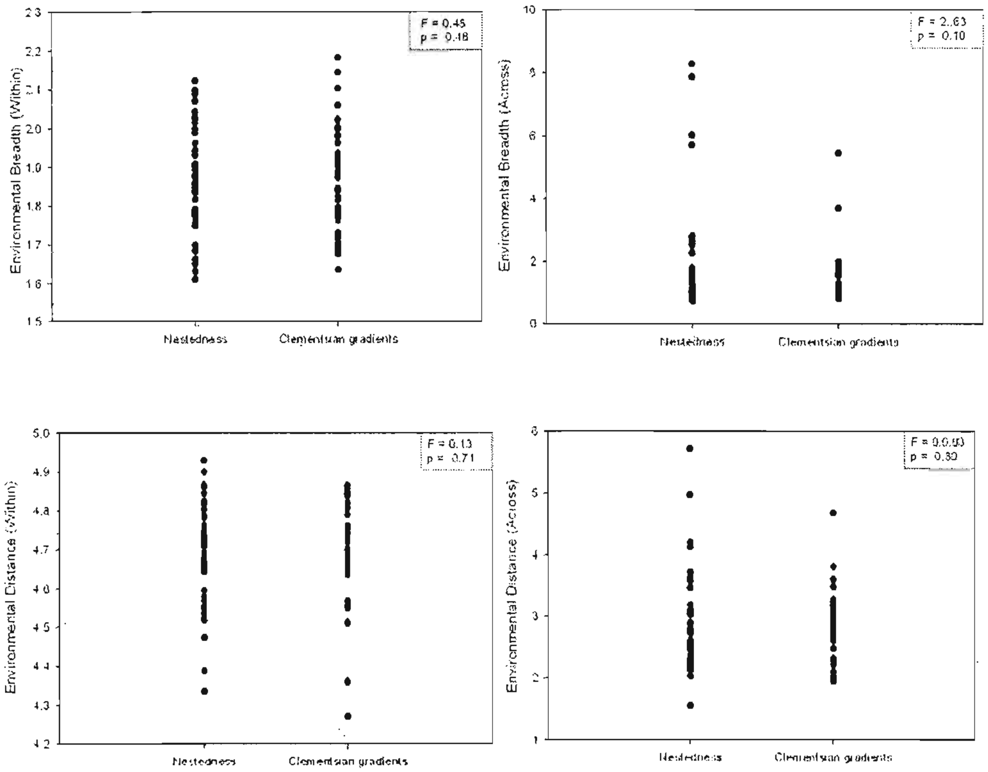

Figure 3.3 Results from ANOVAs on environmental gradients indices between the two main metacommunity patterns (nestedness and Clementsian gradient). The α-level used was 0.05.

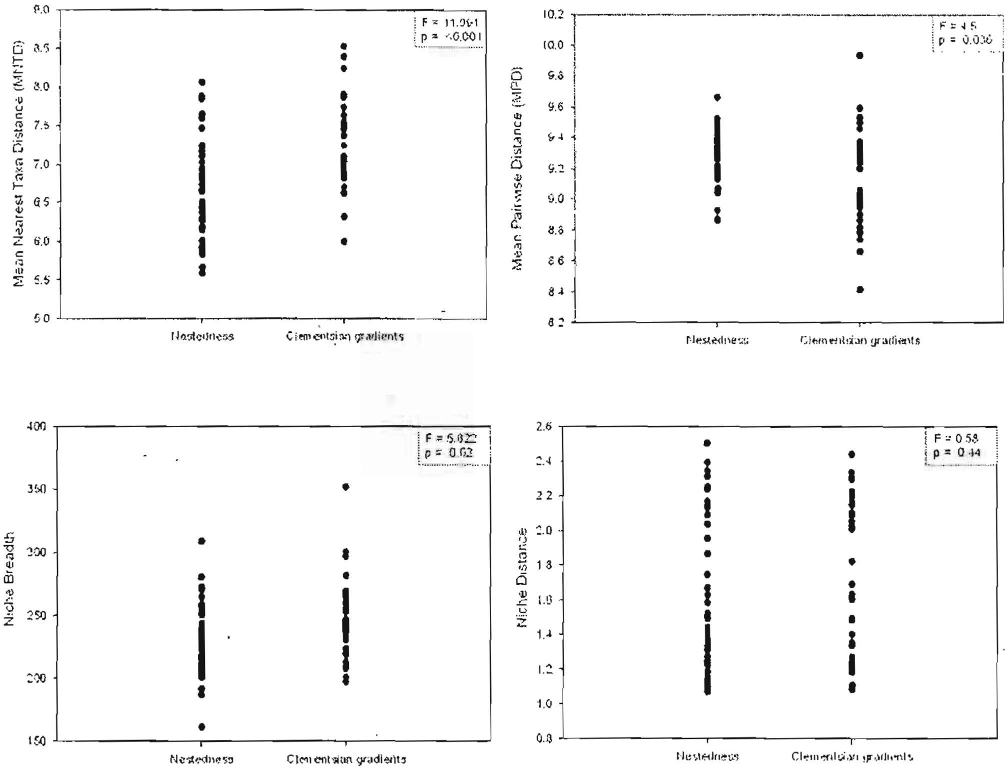

Figure 3.4 Results for ANOVAs on watersheds phylogenetic and niche indices between the two main metacommunity patterns (nestedness and Clementsian gradient). The α-level used was 0.05.

since it was only recently colonized by fish species, just after the end of Pleistocene glaciations (10000 years ago) and watersheds at higher latitudes were the last to be colonized (Mandrak & Crossman, 1992b). Therefore, speciation events are unlikely in such short period of time and indeed there is no endemic species in the region.

At the local scale, phylogenetic patterns were quite different, in which phylogenetic overdispersion was quite predominant (Table 3.3). Indeed, based on MNTD, 10 of the 25 watersheds that presented phylogenetic overdispersion locally were found as phylogenetically clustered regionally and the same pattern was encountered in 4 of the 21 watersheds for MPD. Moreover, with the local filter, 90% of the watersheds presenting overdispersion where above the latitude 47° whereas only 10% where below for MPD, and 64% were located above and 36% below for MNTD. The differences in terms of phylogenetic patterns at the local and regional scales indicate different filtering processes at each scale (Cavender-Bares *et al.*, 2006). Here I propose the following explanation for these contrasting community assembly patterns: Winter harshness is greater in northern watersheds (Mandrak, 1995) and filter out species that do not have the necessary physiological adaptations to survive and reproduce under these conditions (Shuter *et al.*, 1980). At reduced spatial scales (i.e., lake scale) local competition becomes more important, precluding species with high resource-use overlap to inhabit the same lakes (Helmus *et al.*, 2007b). Because latitude is negatively correlated with energy (e.g., PET: $r = -0.77$, $p < 0.0001$) and energy is correlated with ecosystem primary productivity (Gardezi & Gonzalez, 2008), a proxy for resource availability, I expect that resource availability also decreases at high-latitude watersheds. This reduction in resource-availability may increase local competition, generating a pattern of phylogenetic evenness (overdispersion) at the lake scale, even though the watershed species pool is phylogenetically clustered (i.e., at the regional scale). Tendency for phylogenetic overdispersion in higher latitudes was also found by Shurin & Igram (2009) while studying phylogenetic structure of northeast Pacific rockfish assemblages from the genus *Sebastes*.

3.4.2 - Differences between metacommunity distributional patterns

Indices related to environmental gradients showed no differences between Clementsian gradients and nestedness watersheds suggesting that their properties, such as marginality or breadth, are not differently related to the structuring process of each pattern. When standardizing the environment across watersheds, the environmental breadth presented slightly more variation in nestedness watersheds, mostly due to extreme values (Figure 3.3). This is probably due to the fact that nestedness watersheds are encountered throughout the province whereas Clementsian watersheds are found mostly in the south eastern region (Figure 2.1).

There was a trend that distinguished the average phylogenetic structure of communities between the watersheds belonging to each pattern: Clementsian watersheds revealed a higher average MNTD and a lower average MPD than nested watersheds (Figure 3.4). The Clementsian gradient is characterized by several groups of species in which each shows an idiosyncratic response to a portion of the environmental gradient that assemble in different communities (Hoagland & Collins, 1997). If species maintain their environmental preferences through evolutionary times (i.e., niche conservatism; Wiens & Graham, 2005), species closely related might have a more similar β-niche than species that are more distantly related (Cavender-Bares *et al.*, 2006). Because watersheds represent a larger scale which encompasses a considerable amount of environmental variation and due to the greater environmental variability within southern (Mandrak, 1995) watersheds (where the Clementsian gradients pattern appeared the most), species might be filtered according to their optimal environment across lakes. This process result in groups of closely related species that tend to co-occur more than expected by chance across the landscape (i.e., lower MPD) and result in a Clementsian gradient pattern at the metacommunity level

(Leibold & Mikkelson, 2002). Taking into account that the studied fish families are composed by both rare and common species (see prevalence; Figure 3.1 and Table 3.1), I suggest two hypothesis for the fact that nested watersheds are in average composed by more distantly related species: 1) A large number of these watersheds are located in the northern region and hence distant from postglacial refugees (Mandrak & Crossman, 1992b); thus they were probably colonized by the best dispersers (which in general are common species) from different clades (Figure 3.1); 2) Northern watersheds have lower energy and cannot sustain rare species which are generally specialists due to lack of resource availability. However, these watersheds can be inhabited by common species that are generalist and can maintain viable populations in low-energy environments (Gardezi & Gonzalez, 2008). Because these species are in average distantly related (i.e., from different clades; Figure 3.1), these watersheds show in average greater community phylogenetic distance (i.e., higher MPD).

Finally, the two metacommunity patterns did not present any real differences in the average environmental niche (marginality or breadth) of their species. Niche marginality did not differ significantly between patterns and niche breadth was only significantly higher in Clementsian watersheds due to an outlier (Figure 3.4). I suggest the following: 1) These niche measures take into account only the environment in which species inhabit; 2) They are calculated based on the average of all species niches that inhabit the metacommunity; and 3) Most common species (e.g., *Perca flavescens*, *Catostomus commersoni*) are distributed across most watersheds, regardless of metacommunity pattern. These three properties may mask any possible difference between nested and Clementsian watersheds in relation to species β-niches.

3.4.3 - Linking environmental gradients to communities phylogenetic structure and species niches

The analyses of the two main environmental gradients showed a trend for higher environmental heterogeneity (and extreme environments) in high-energy watersheds containing smaller lakes, which are more concentrated in south-eastern Ontario (see results chapter 1). Studies elsewhere support this correlation between energy and environmental heterogeneity (Bonn *et al.*, 2004 and references therein). Moreover, Mandrak (1995) studying Ontario lakes showed that south-eastern Ontario lakes have a greater local variability in precipitation and seasonally variable temperature, suggesting greater environmental variability in contrast to north-western lakes, which exhibit a more stable climate.

The phylogenetic structure of communities varied along environmental gradients. MPD was greater in high-energy watersheds with large lakes and, thereby, positively correlated with PC-1 and negatively correlated with PC-2 (Table 3.5); MNTD was negatively related to both axes (Table 3.5). Such correlations with PC-1 gradient (to which I will restrain my interpretations) suggests that low-energy watersheds, which contain more homogenous (i.e., less resources or habitat types; Guégan *et al.*, 1998) and harsher environments (i.e., cold temperatures; Mandrak, 1995), might not be colonized by clades which house species intolerant to these conditions (e.g., hot water species) or specialist species that require specific resources potentially lacking within such low-energy environments. Lakes within these watersheds, however, are larger and perhaps they possess greater abundances of the few resources they provide allowing pairs of closest related species across different clades to co-exist and accounting for the higher MPD but lower MNTD found within these watersheds.

Conversely, the high-energy watersheds found within southern Ontario encompass lakes with more heterogeneous and yet benign environments (e.g., higher temperatures, longer growing season; Mandrak, 1995). These conditions allow the survivorship of a highly diverse fish-fauna (Chu *et al.*, 2003), which incorporates species from all the clades housed within the regional pool including rare specialist species (Gardezi & Gonzalez, 2008). Indeed, southern Ontario supports cool water species and encompasses the lower distributional limit of cold water species and the upper distributional limit of warm water species (Chu *et al.*, 2008). Despite such inclusive conditions, lakes within these watersheds are smaller and thereby each potentially houses only a small portion of the environmental gradients. Moreover, these smaller lakes also provide to constituent species less abundant resources. Such characteristics then account for communities within these watersheds having on average, a lower MPD (i.e., lower average phylogenetic diversity) and a higher MNTD (i.e., greater effect of competition between closest relative species due to low abundance of available resources). For example, species might assemble across lakes according to the specific environmental tolerances (i.e., habitat filtering) which characterize their clades (e.g., species from cyprinadae family used on this study are mostly cool or hot water species) but closest relative species which overlap greatly in their resource use (α-niche) do not co-occur in the same lakes (e.g., *Rhinichthys atratulus* and *Rhinichthys cataractae* are both invertivores; Coker *et al.*, 2001). Niche marginality corroborates these results for MPD, showing greater values in low-energy environments containing lakes with larger surface area. Assuming that β-niche is conserved (Wiens & Graham, 2005), a larger phylogenetic diversity in low-energy watersheds should also lead to greater difference between species niche within lakes.

Environmental marginality (within watersheds) did not present any significant correlation with the species indices (Table 3.6) whereas environmental marginality (across watersheds) was correlated significantly with niche marginality, MPD and MNTD (Table 3.6). Because lakes within any given watershed are more likely to be similar in environment than lakes across watersheds, perhaps there is not enough variation in environmental conditions across lakes within the same watershed to pattern phylogenetic community structure. This result reinforces the idea that processes of habitat filtering are more likely to occur at larger spatial scales (Cavender-Bares *et al.*, 2006). Both MPD and MNTD were negatively related with environmental marginality (across watersheds), supporting the idea that increased environmental heterogeneity in a given watershed is likely to sort species according to their environmental preferences, clustering the overall community phylogenetic structure (Cavender-Bares *et al.*, 2006). Moreover, niche marginality was negatively correlated with MPD, indicating that communities composed by species that have more marginal niches tend to have a clustered phylogenetic structure. Thus, species from the same clades that have marginal niches in relation to the mean environment of the landscape should be more affected by habitat filtering. For example, species from the Centrarchidae family present in this study, which are mostly hot water species (Coker *et al.*, 2001), tend to co-occur in watersheds with marginal environmental conditions (in relation to the average cold temperatures from the entire Province), such as the southern watersheds, decreasing their overall average phylogenetic distance (MPD).

CONCLUSION

Multiple factors dictate the success (or failure) of species from the regional pool to assemble into local communities (e.g., Jackson *et al.*, 2001), hence, it is expected that different communities across space will be affected by different processes. This issue has led ecologists to perceive community ecology as a plethora of unique systems, where each is contingent on specific processes (Lawton, 1999). Thus, advances in theory and analytical tools should be aimed at the development of general frameworks in which ecologists cab base their studies (e.g., Chesson, 2000; Leibold & Mikkelson, 2002; Holyoak *et al.*, 2005; Vellend, 2010). In this context, EMS analysis is an effective tool to distinguish among six metacommunity patterns (Leibold & Mikkelson, 2002; Presley *et al.*, 2009) and represents substantial progress in contrast to analyses that are restricted to one distributional patterns (e.g., traditional analyses of nestedness; Patternon & Atmar, 1986; checkerboards; Diamond, 1975; Gilpin & Diamond, 1982). Despite the uniqueness of ecological systems (Lawton, 1999), the species within should ultimately be distributed according to one of the metacommunity patterns (or some variant of them) (Presley *et al.*, 2010). Even when particular species distributions do not coincide exactly with a particular pattern, there is always one that best fits the data, facilitating the exploration of mechanisms underlying species distributions. This study was the first to compare multiple metacommunities across a large region and among the 85 watersheds analyzed which contained more than 8900 lakes, the fish species were distributed according to two main patterns: nestedness (Ulrich *et al.*, 2009) and Clementsian gradients (Hoagland & Collins, 1997). Thus, contingency should only be a main concern at fine-scale studies.

Moreover, a large number of ecological studies have been addressing patterns of diversity and distribution correlating them with the variation in the physical environment (e.g., Presley *et al.*, 2009) and, while these approaches have provided a good deal of insights regarding the factors driving patterns of species co-existence within metacommunities, we still lack knowledge regarding abiotic and biotic that may determine differences across metacommunities. Such analysis was performed in chapter one; however, other perspectives have to be integrated to complete the ecological puzzle. Each species present somewhat unique characteristics; the presence (or absence) of a particular species in a community may affect other species in several ways (e.g., competition; Diamond, 1975) and may even change the environmental conditions of a site, hence influencing the outcome of community dynamics. In chapter two, the species perspective was incorporated, in which ecological properties of species such as their niches and phylogenetic relatedness were utilized to infer processes underlying community assembly (Cavender-Bares *et al.*, 2004; Ingram & Shurin, 2009). The consideration of species phylogeny within ecological studies is becoming increasingly relevant, especially as molecular techniques to estimate phylogenetic relationships among species increase in availability. This study was one of the first to integrate all of these views in order to generate a more complete picture of the processes affecting species co-existence at large scales. Future studies within this venue should also work with species functional traits (McGill *et al.*, 2006) such as those relevant for foraging, physiological tolerance and dispersal capacity (e.g., Webb, 1984; Cavender-Bares *et al.*, 2006; Mason *et al.*, 2008; Ingram & Shurin, 2009). This information should then be related back to species phylogenetic relatedness in order to better distinguish processes of evolution such as niche conservatism *versus* convergence (Wiens & Graham, 2005; Losos, 2008). This would allow further advancement of our knowledge about the factors structuring mechanisms of communities and ultimately, metacommunities.

Finding the pattern of best fit within distributional data can be of great utility when one has to make decisions regarding conversation planning; for instance, target site selection should depend on the pattern found. As nestedness is characterized by low species turnover and hierarchical richness, one should focus conservation efforts at richer sites, which should protect most of the regional biodiversity. However, if a Clementsian gradient is the pattern of best fit, focusing efforts on the richest site would not be sufficient for regional biodiversity conservation because of high species turnover rates across sites; instead, one should focus on the environmental heterogeneity across communities.

I have shown that EMS analysis can be a head start to test multiple hypotheses. As presence-absence datasets are becoming increasingly available, ecologists should explore this analytical tool more often, as it is simple to use and provides great insights. An interesting future step for the evaluation of species distribution along environmental gradients is the improvement of this analytical tool in order to incorporate abundance data. This would allow one, for instance, to weight species influences on structural patterns according to their abundances and would certainly increase the explanatory power of EMS technique in relation to the structuring mechanisms of the different patterns. Nevertheless, I suggest that EMS analysis should be applied to a wide variety of taxa in order to find general associations between idealized patterns and species types.

REFERENCES

Azeria, E. T. and Kolasa, J., 2008. Nestedness, niche metrics and temporal dynamics of a metacommunity in a dynamic natural model system. Oikos, **117**: 1006-1019.

Barbour, C. D. and Brown, J. H., 1974. Fish species diversity in lakes. The American Naturalist, **108**: 473-489.

Bonn, A., Storch, D. And Gaston, K. J., 2004. Structure of the species-energy relationship. Proc. R. Soc. Lond. B, **271**: 1685-1691.

Borcard, D., Legendre, P. and Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology, **73**: 1045-1055.

Brown, J. H. and Kodric-Brown, A., 1977. Turnover rates in insular biogeography: effect of immigration on extinction. Ecology, **58**: 445-449.

Brown, W. M., George Jr., M. and Wilson, A. C., 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of Natural Academy of Sciences, **76**: 1967-1971.

Carson, H. L., and K. Y. Kaneshiro. 1976. Drosophila of Hawaii: systematics and ecological genetics. Annual Review of Ecology and Systematics 7: 311-346.

Cavender-Bares, J., Ackerly, D. D., Baum, D. A. and Bazzaz F. A., 2004. Phylogenetic overdispersion in Floridian oak communities. The American Naturalist, **163**(6): 823-843.

Cavender-Bares, J., Keen, A. and Miles, B., 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. Ecology, **87**: S109-S122.

Cavender-Bares, J., Kozak, K. H., Fin, P. V. A. and Kembel, S. W., 2009. The merging of community ecology and phylogenetic biology. Ecology Letters, **12**: 693-715.

Chase, J. M., 2007. Drought mediates the importance of stochastic community assembly. Proceedings of the National Academy of Sciences, **104**(44): 17430-17434.

Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annual Review of Ecology, Evolution and Systematics, **31**: 343–366.

Chu, C., Minns, C. K. and Mandrak, N. E., 2003. Comparative regional assessment of factors impacting freshwater fish biodiversity in Canada. Canadian Journal of Fisheries and Aquatic sciences, **60**: 624-634.

Chu, C., Mandrak, N. E. and Minns, C. K., 2005. Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Canada. Diversity and Distributions, **11**: 299-310.

Chu, C., Jones, N. E., Mandrak, N. E., Piggott, A. R. and Minns, C. K., 2008. The influence of air temperature, groundwater discharge, and climate change on the thermal diversity of stream fishes in southern Ontario watersheds. Canadian Journal of Fisheries and Aquatic Sciences, **65**: 297-308.

Clements, F. E., 1916. Plant succession, an analysis of the development of vegetation – Carnegie Institution.

Cooker, G. A., Portt, C. B., and Minns, C. K., 2001. Morphological and ecological characteristics of Canadian freshwater fishes. Fisheries and Oceans Canada, Burlington, Ontario.

Connor, E. F. and Simberloff, D. S. 1979. The assembly of species communities: chance or competition? Ecology, **60**: 1132–1140.

Cook, R. R. and Quinn, J. F., 1995. The influence of colonization in nested subsets. Oecologia, **102**(4): 413-424.

Costa, G. C. and Schupp, I., 2010. Biogeography of the Amazon molly: ecological niche and range limits of an asexual hybrid species. Global Ecology and Biogeography, **19**: 442-451.

Cottenie, K., 2005. Integrating environmental and spatial processes in ecological community dynamics. Ecology letters, **8**: 1175-1182.

Cox, E. T., 1978. Counts and measures of Ontario lakes: watershed unit summaries based on maps of various scales by watershed unit. Ontario ministry of Natural Resources, Fisheries Branch, Toronto.

Diamond, J. M., 1975. Assembly of species communities. Pages 342-444. *In* Ecology and evolution of communities. *Edited by* M. L. Cody and J. M. Diamond. Harvard University Press, Cambridge, Massachusetts, USA.

Dolédec, S., Chessel, D. and Gimaret-Carpentier, C., 2000. Niche separation in community analysis: A new method. Ecology, **81**(10): 2914-2927.

Driscoll, D. A. and Lindenmayer, D. B., 2009. Empirical tests of metacommunity theory using an isolation gradient. Ecological Monographs, **79**(3): 485-501.

Eadie, J . M., Hurly, T. A., Montgomerie, R. D. and Teather, K. L., 1986. Lakes and rivers as islands: species-area relationships in the fish faunas of Ontario. Env. Bio. of Fishes, **15**(2): 81-8.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist, **125**: 1-15.

Fernández-Juricic, E., 2002. Can human disturbance promote nestedness? A case study with breeding birds in urban habitat fragments. Oecologia, **131**: 269-278.

Fang, X. and Stefan, H. G., 2000. Projected climate change effects on winterkill in shallow lakes in the northern United States. Environmental Management, **25**(3): 291-304.

Gardezi, T. and Gonzalez, A., 2008. Scale dependence of species-energy relationships: Evidence from fishes in thousands of lakes. The American Naturalist, **171**(6): 800-815.

Gaston, K.J. (2000). Global patterns in biodiversity. Nature, **405**: 220-227.

Gauch, H. G., Whittaker, R. H. and Wentworth, T. R., 1977. A comparative study of reciprocal averaging and other ordinations techniques. Journal of Ecology, **65**: 157-174.

Gilpin. M. E. and Diamond, J .M., 1982. Factors contributing to non-randomness in species co-occurrences on islands. Oecologia, **52**(1): 75-84.

Gleason, H. A., 1926. The individualistic concept of the plant association. – Bull. Torrey Bot. Club, **53**: 7-26.

Goodchilde, G. A., and Gale, G. E., 1982. Aquatic habitat inventory: the Ontario approach to lake surveys. Pages 304-311 *in* N. B. Armantrout, ed. Acquisition and utilization of aquatic habitat inventory information. American Fisheries Society, Wester Division, Bethesda, MD.

Gotelli, N. J. and Graves, G. R., 1996. Null Models in Ecology. Smithsonian Institution Press, Washington, DC.

Gotelli, N.J. and McCabe, D. J., 2002. Species co-occurrence: a meta-analysis for J. M. Diamond's assembly rules model. Ecology, **83**: 2091-2096.

Guégan, J., Lek, S. and Oberdorff, T., 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. Nature, **391**: 382-384.

Hanski, I., 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology, **63**: 151-162.

Haudsorf, B. and Hennig, C., 2007. Null model test of clustering of species, negative co-occurrence patterns and nestedness in meta-communities. Oikos, **116**: 818-828.

Heino, J., 2005. Metacommunity patterns of highly diverse stream midges: gradients, chequerboards, and nestedness, or is there only randomness. Ecological Entomology, **30**: 590-599.

Helmus, M. R., Bland, T. J., Williams, C. K., and Ives, A. R., 2007a. Phylogenetic measures of biodiversity. The American Naturalist, **69**: E68-E83.

Helmus, M. R., Savage, K., Diebel, M. W., Maxted, J. T. and Ives, A. R., 2007b. Separating determinants of community structure. Ecology Letters, **10**: 917-925.

Hoagland, B. W. and Collins, S. L., 1997. Gradient Models, Gradient Analysis, and Hierarchical Structure in Plant Communities. Oikos, **78**(1): 23-30.

Holyoak, M., Leibold, M. A. and Holt, R. D., editors (2005). Metacommunities: Spatial Dynamics and Ecological Communities. University of Chicago Press, Chicago.

Honnay, O., Hermy, M. and Coppin, P., 1999. Nested plant communities in deciduous forest fragments: species relaxation or nested habitats? Oikos, **84**(1): 119-129.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian Freshwater Fishes through DNA Barcodes. PLoS ONE, **3**(6): e2490.

Hutchinson, G. E., 1957. Concluding remarks - Cold Spring Harbor Symposia on Quantitative Biology., **22**: 415-427. Reprinted in 1991: Classics in Theoretical Biology. Bull. of Math. Biol., **53**: 193-213.

Kadoya, T., 2009. Assessing functional connectivity using empirical data. Population Ecology, **51**: 5-15.

Ingram, T. and Shurin, J. B., 2009. Trait-based assembly and phylogenetic structure in northeast Pacific rockfish assemblages. Ecology, **90**(9): 2444-2453.

Jackson, D. A., P. R. Peres-Neto, and Olden, J. D., 2001. What controls who is where in freshwater fish communities - the role of biotic, abiotic and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences, **58**: 157-170.

Jacobson, B. and Peres-Neto, P. R., 2010. Quantifying and disentangling dispersal in metacommunities: How close have we come? How far is there to go? Landscape Ecology, **25**: 494-507.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, **15**: 111-120.

Kodric-Brown, A. and Brown, J. H., 1977. Turnover rates in insular biogeography: effect of immigration on extinction. Ecology, **58**(2): 445-449.

Kraft, N. J. B., Cornwell, W. K., Webb, C. O. and Ackerly, D. D., 2007. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. The American Naturalist, **170**(2): 271-283.

Layeghifard, M., Peres-Neto, P. R. and Makarenkon, V., 2010. HTG detection method in historical biogeography. Global Ecology and Biogeography (submitted).

Lawton, J. H., 1999. Are there general laws in ecology? Oikos, **84**(2): 177-192.

Leibold, M. A. and Mikkelson, G. M., 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. Oikos, **97**: 237-250.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M. and Gonzalez, A., 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters, **7**: 601-613.

Leibold, M.A., Economo, E.P. and Peres-Neto, P.R., 2010. Metacommunity phylogenetics: Separating the roles of environmental filters and historical biogeography. Ecology Letters (in press).

Legendre, P and Legendre, L., 1998. Numerical ecology. 2nd English edition. Elsevier Science BV, Amsterdam.

Legendre, P. and Gallagher, E. D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, **192**: 271-280.

Legendre, P, Borcard, Peres-Neto, P. R., 2008. Analyzing or explaining beta-diversity. Ecology, **89**: 3238-3244.

Leprieur, F., Olden, J. D., Lek, S. and Brosse, S., 2009. Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. Journal of Biogeography, **36**: 1899-1912.

Loeuille, N. and Leibold, M. A., 2008. Evolution in metacommunities: On the relative importance of species sorting and monopolization in structuring communities. The American Naturalist, **171**(6): 788-798.

Losos, J. B., 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecology Letters, **11**: 995-1007.

MacArthur, R. and Levins, R., 1967. The limiting similarity, convergence and divergence of coexisting species. The American Naturalist, **101**: 377-385.

MacArthur, R. H. and Wilson, E. O., 1967. The Theory of Island Biogeography. Princeton: Princeton University Press.

Magnuson, J. J., Tonn, W. M., Banerjee, A., Toivonen, J., Sanchez, O. and Rask, M., 1998. Isolation vs. extinction in the assembly of fishes in small northern lakes. Ecology, 79(8): 2941-2956.

Mandrak, N. E. and Crossman, E. J., 1992*a*. A checklist of Ontario freshwater fishes: annotated with distribution maps. Royal Ontario Museum, Toronto.

Mandrak, N. E and Crossman, E. J., 1992*b*. Postglacial dispersal of freshwater fishes into Ontario. Canadian Journal of Zoology, **70**: 2247-2259.

Mandrak, N. E., 1995. Biogeographic patterns of fish species richness in Ontario lakes in relation to historical and environmental factors. .Canadian Journal of Fisheries and Aquatic Sciences, **52**: 1462-1474.

Mason, W. H. N., Lanoiselée, C., Mouillot, D., Irz, P. and Argillier, C., 2007. Functional characters combined with null models reveal inconsistency in mechanisms of species turnover in lacustrine fish communities. Oecologia, **153**: 441-452.

Mason, W. H. N., Lanoiselée, C., Mouillot, D., Wilson, J. B. and Argillier, C., 2008. Does niche overlap control relative abundance in French lacustrine fish communities? A New method incorporating functional traits. Journal of Animal Ecology, **77**: 661-669.

McDonald, K. A. and Brown, J. H., 1992. Using montane mammals to model extinctions due to global change. Conservation Biology, **6**(3): 409-415.

McGill, B.J., Enquist, B.J., Weiher, E. and Westoby, M., 2006. Rebuilding community ecology from functional traits. Trends in Ecology and Evolution, **21**: 178–185.

Ministry of Environment, Integrated Environmental Planning Division and Strategic Policy Branch, 2004. White paper on watershed-based source protection planning. Ontario: Queen's printer for Ontario, pp 34.

Minns, C. K., 1986. A model of bias in lake selection for survey. Can. Tech. Rep. Fish. Aquat. Sci., **1496**: 21p.

Minns, C. K., 1989. Factors affecting fish species richness in Ontario lakes. Transactions of the American Fisheries Society, **119**: 282-291.

Mims M. C., Olden J. D., Shattuck Z. R., Poff N.L., 2010. Life history trait diversity of native freshwater fishes in North America. Ecology of Freshwater Fish, **19**: 390-400.

Morisita, M., 1971. Composition of the I-index. Researches on Population Ecology, **13**: 1-27.

Muneepeeraku, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A. and Rodriguez-Iturbe, I., 2008. Neutral metacommunity model predict fish diversity patterns in Mississipi-Miosuri basin. Nature, **453**: 220-222.

Olden, J. D., Jackson, D. A. and Peres-Neto, P. R., 2001. Spatial isolation and fish communities in drainage lakes. Oecologia, **127**: 572-585.

Patterson, B. D. and Atmar, W., 1986. Nested Subset and the structure of insular mammalian faunas and archipelagos. Biological Journal of Linnean Society, **28**: 65-82.

Pennington, R. T., J. A. Richardson, and M. Lavin. 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, phylogenetic community structure and neutral ecological theory. New Phytologist, **172**: 605–616.

Peres-Neto, P. R., 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. Oecologia, **140**: 352-360.

Peres-Neto, P. R., and P. Magnan, 2004. The influence of swimming demand on phenotypic plasticity and morphological integration: a comparison of two polymorphic charr species. Oecologia, **140**: 36-45.

Peres-Neto, P.R. 2006. A unified strategy for estimating and controlling spatial, temporal and phylogenetic autocorrelation in ecological models. Oecologia Brasiliensis, **10**: 105-119.

Peres-Neto, P. R., Legendre, P., Dray, S. and Borcard, D., 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology, **87**: 2614-2625.

Peres-Neto, P. R and Legendre, P., 2010. Estimating and controlling for spatial structure in ecological communities, Global Ecology and Biogeography, **19**: 174-184.

Pillar, V. D. and Duarte, L. D. S., 2010. A framework for metacommunity analysis of phylogenetic structure. Ecology Letters, **13**: 587-596.

Presley, S. J., Higgins, C. L., López-González, C. and Stevens, R. D., 2009. Elements of metacommunity structure of Paraguayan bats: multiple gradients require analysis of multiple ordination axes. Oecologia, **160**: 781-793.

Presley, S. J. and Willig, M. R., 2010. Bat metacommunities structure on Caribbean islands and the role of endemics. Global Ecology and Biogeography, **19**: 185-199.

Presley, S. J., C. L. Higgins, and M. R. Willig, 2010. A comprehensive framework for the evaluation of metacommunity structure. Oikos, **119**: 908-917.

Ricklefs, R. E., 1987. Community Diversity: Relative Roles of Local and Regional Processes. Science, **235**(4785): 167-171.

Ricklefs, R. E., 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters, **7**: 1-15.

Saitou, N. And Nei, M., 1987. The neighbour-joining method: a new method for reconstructing evolutionary trees. Molecular Biology and Evolution, **4**: 406-425.

Sharma, S., Jackson, D. A. and Minns, C. K., 2009. Quantifying the potential effects of climate change and the invasion of smallmouth bass on native lake trout populations across Canadian lakes. Ecography, **32**: 517-525.

Shurin, J. B., Winder, M., Adrian, R., Keller, W., Matthews, B., Paterson, A. M, Pinel-Alloul, B., Rusak, J. A. and Yan, N. D., 2010. Environmental stability and lake zooplankton diversity - contrasting effects of chemical and thermal variability. Ecology Letters, **13**: 453-463.

Shuter, B. J., MacLean, J. A. Fry, F. E. J. and Regier H. A., 1980. Stochastic simulation of temperature effect on first-year survival of smallmouth bass. Transactions of the American Fisheries Society, **109**: 1-34.

Shuter, B. J. and Post, J. R., 1990. Climate, population viability, and the zoogeography of temperate fishes. Transactions of the American Fisheries Society, **199**: 314-336.

Smith, K. F. and Brown, J. H. 2002. Patterns of diversity, depth range and body size among pelagic fishes along a gradient of depth. Global Ecology and Biogeography, **11**: 313-322.

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters, **10**: 1115–1123.

Takamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** Molecular Biology and Evolution, **24**: 1596-1599.

Tilman, D., 1982. Resource competition and community structure – Princeton University Press.

Tonn, W. M., Magnuson, J. J., Rask, M. and Toivonen J. 1990. Intercontinental Comparison of Small-Lake Fish Assemblages: The Balance between Local and Regional Processes. The American Naturalist, **136**(3): 345-375.

Ulrich, W., Almeida-Neto, M. And Gotelli, N. J., 2009. A consumer's guide to nestedness analysis. Oikos, **118**(1): 3-17.

Vellend, M., 2010. Conceptual synthesis in community ecology. The quarterly review of biology, **85**(2): 183-206.

Webb, P. W., 1984. Body form, locomotion and foraging in aquatic vertebrates. American Zoologist, **24**(1): 107-120.

Webb, C. O., 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. The American Naturalist, **156**(2): 145-155.

Webb, C. O., Ackerly, D. D., McPeek, M. A. and Donoghue, M. J., 2002. Phylogenies and community ecology. Annual Review of Ecology, Evolution and Systematics, **33**: 475-505.

Werner, E. E., Yurewicz, K. L., Shelly, D. K. and Relyea, R. A., 2007. Turnover in an amphibian metacommunity: the role of local and regional factors. Oikos, **116**: 1713-1725.

Wiens, J. J. and Graham, C. H., 2005. Niche conservatism: Integrating evolution, ecology and conservation biology. Annual Review of Ecology, Evolution and Systematics, **36**: 519-539.

Wright, D. H., Patterson, B. D., Mikkelson, G. M., Cutler, A. and Atmar, W., 1998. A comparative analysis of nested subset patterns of species composition. Oecologia, **113**: 1-20.