

Université de Montréal

**LES IMPACTS DE LA DISPERSION HISTORIQUE SUR LA  
VARIABILITÉ GÉNÉTIQUE À DIFFÉRENTES ÉCHELLES  
SPATIALES : CONNAÎTRE L'HISTOIRE POUR MIEUX  
COMPRENDRE LE PRÉSENT.**

par  
Frédéric Cyr

Département de Sciences biologiques  
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Études Supérieures et Postdoctorales  
En vue de l'obtention du grade de Ph.D.  
En Sciences Biologiques

31 août 2012

© Frédéric Cyr

Université de Montréal  
Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée :

**LES IMPACTS DE LA DISPERSION HISTORIQUE SUR LA  
VARIABILITÉ GÉNÉTIQUE À DIFFÉRENTES ÉCHELLES  
SPATIALES : CONNAÎTRE L'HISTOIRE POUR MIEUX  
COMPRENDRE LE PRÉSENT.**

Présentée par :  
Frédéric Cyr

a été évaluée par un jury composé des personnes suivantes :

Bernard Angers, directeur de recherche  
François-Joseph Lapointe, président-rapporteur  
Patrick James, membre du jury  
Dany Garant, examinateur externe  
François-Joseph Lapointe, représentant du doyen de la FES

*L'Histoire, veut la  
boutade, est  
quelque chose qui  
n'est jamais arrivé,  
raconté par  
quelqu'un qui  
n'était pas là.*

*À Louise, Gabrielle et Juliette*

## RÉSUMÉ

La variabilité génétique actuelle est influencée par une combinaison complexe de variables historiques et contemporaines. Dès lors, une interprétation juste de l'impact des processus actuels nécessite une compréhension profonde des processus historiques ayant influencé la variabilité génétique. En se basant sur la prémissse que des populations proches devraient partager une histoire commune récente, nombreuses études, effectuées à petite échelle spatiale, ne prennent pas en considération l'effet potentiel des processus historiques. Cette thèse avait pour but de vérifier la validité de cette prémissse en estimant l'effet de la dispersion historique à grande et à petite échelle spatiale.

Le premier volet de cette thèse avait pour but d'évaluer l'impact de la dispersion historique sur la répartition des organismes à grande échelle spatiale. Pour ce faire, les moules d'eau douce du genre flotteurs (*Pyganodon spp.*) ont servies de modèle biologique. Les moules d'eau douce se dispersent principalement au stade larvaire en tant que parasites des poissons. Une série de modèles nuls ont été développés pour évaluer la co-occurrence entre des parasites et leurs hôtes potenitels. Les associations distinctes du flotteur de Terre-Neuve (*P. fragilis*) avec des espèces de poissons euryhalins permettent d'expliquer sa répartition. Ces associations distinctes ont également pu favoriser la différenciation entre le flotteur de Terre-Neuve et son taxon sœur : le flotteur de l'Est (*P. cataracta*). Cette étude a démontré les effets des associations biologiques historiques sur les répartitions à grande échelle spatiale.

Le second volet de cette thèse avait pour but d'évaluer l'impact de la dispersion historique sur la variabilité génétique, à petite échelle spatiale. Cette fois, différentes populations de crapet de roche (*Ambloplites rupestris*) et de crapet soleil (*Lepomis gibbosus*), dans des drainages adjacents ont servies de modèle biologique. Les différences frappantes observées entre les deux espèces suggèrent des patrons de colonisation opposés. La faible diversité génétique observée en amont des drainages et la forte différenciation observée entre les drainages pour les populations de crapet de roche suggèrent que cette espèce aurait colonisé les drainages à partir d'une source

en aval. Au contraire, la faible différenciation et la forte diversité génétique observées en amont des drainages pour les populations de crapet soleil suggèrent une colonisation depuis l'amont, induisant du même coup un faux signal de flux génique entre les drainages. La présente étude a démontré que la dispersion historique peut entraver la capacité d'estimer la connectivité actuelle, à petite échelle spatiale, invalidant ainsi la prémissse testée dans cette thèse.

Les impacts des processus historiques sur la variabilité génétique ne sont pas faciles à démontrer. Le troisième volet de cette thèse avait pour but de développer une méthode permettant de les détecter. La méthode proposée est très souple et favorise la comparaison entre la variabilité génétique et plusieurs hypothèses de dispersion. La méthode pourrait donc être utilisée pour comparer des hypothèses de dispersion basées sur le paysage historique et sur le paysage actuel et ainsi permettre l'évaluation des impacts historiques et contemporains sur la variabilité génétique. Les performances de la méthode sont présentées pour plusieurs scénarios de simulations, d'une complexité croissante. Malgré un impact de la différentiation globale, du nombre d'individus ou du nombre de loci échantillonné, la méthode apparaît hautement efficace. Afin d'illustrer le potentiel de la méthode, deux jeux de données empiriques très contrastés, publiés précédemment, ont été ré analysés.

Cette thèse a démontré les impacts de la dispersion historique sur la variabilité génétique à différentes échelles spatiales. Les effets historiques potentiels doivent être pris en considération avant d'évaluer les impacts des processus écologiques sur la variabilité génétique. Bref, il faut intégrer l'évolution à l'écologie.

**Mots clés : Patrons de colonisation ; dispersion historique ; dispersion ; connectivité ; associations biologiques ; génétique du paysage ; réseau génétique.**

## ABSTRACT

Current genetic variability depends on a complex combination of historical and contemporary factors. Therefore, an unbiased interpretation of the impact of current processes requires a deep understanding of historical processes that shaped genetic variability. Based on the premise that spatially close populations should share a common recent history, many studies conducted at small spatial scale do not take into account the effect of historical processes. This thesis aims to verify the validity of this assumption by estimating the effect of historical dispersion at large and small spatial scales.

The aim of the first part of this thesis was to evaluate the impact of the historical dispersion on the distribution of organisms, at large spatial scale. The freshwater mussels of the genus Floater (*Pyganodon spp.*) were used as biological model. The dispersal of freshwater mussels occurs when larvae parasitize fish. Also, a series of null models has been developed to evaluate the co-occurrence between parasites and their hosts. The distinct associations of the Newfoundland floater (*P. fragilis*) with euryhaline fish species can explain its distribution. These associations also promoted differentiation with its sister taxon: the Eastern floater (*P. cataracta*). This study demonstrated the effects of historical biological associations on the current distribution of species at a large spatial scale.

The aim of the second part of this thesis was to evaluate the impact of historical dispersal on the genetic variability, at small spatial scale. This time populations of rock bass (*Ambloplites rupestris*) and sunfish (*Lepomis gibbosus*) in adjacent drainages were used as biological model. The striking differences observed between the two species suggest opposing colonization patterns. Rock bass have colonized the drainage from a downstream source, resulting in low genetic diversity upstream and a strong differentiation between drainages. However, the sunfish have colonized the system from upstream, resulting in a high genetic diversity and low differentiation upstream, inducing at the same time a false signal of gene flow between drainages. The present study shows that the historical dispersion may hinder the ability to

estimate the actual connectivity at small spatial scale, thus invalidating the premise tested in this thesis.

The impacts of historical processes on genetic variability are not easy to demonstrate. The aim of the third part of this thesis was to develop a method to facilitate their detection. The proposed method is very flexible and facilitates the comparison between genetic variability and multiple dispersal hypotheses. The method could be used to compare dispersal hypotheses based on the historic or the current landscape and allow the assessment of historical and contemporary processes on genetic variability. The performances of the method are presented in several simulation scenarios, of increasing complexity. Despite an overall impact of differentiation, the number of individuals or the number of loci sampled, the method is highly effective. To illustrate the potential of the method, two contrasted data sets from previously published studies, were re-analyzed.

This thesis demonstrates the impacts of historical dispersal on genetic variability at different spatial scales. Potential historical effects must be taken into account before assessing the impacts of ecological processes on genetic variability. In short, we must bridge the gap between ecology and evolution.

**Keywords:** Patterns of colonization; historical dispersion; dispersal; connectivity; biological associations, landscape genetics, genetic network.

## TABLE DES MATIÈRES

Résumé.....	i
Abstract.....	iii
Table des Matières.....	v
Liste des Tableaux.....	viii
Liste des Figures.....	x
Liste des Encadrés.....	xiv
Les des sigles et des abréviations.....	xiv
Remerciements.....	xvi
Chapitre 1 : Introduction générale.....	1
1.1 La variabilité génétique.....	3
1.2 L'estimation de la variabilité génétique.....	4
1.3 Un déséquilibre entre la perte et le gain de la variabilité génétique.....	5
1.4 L'étude de la variabilité génétique.....	6
La biogéographie.....	6
La phylogéographie.....	7
La génétique du paysage.....	8
1.5 Les processus historiques et les processus contemporains.....	9
1.6 Les processus historiques ou les processus contemporains.....	11
1.7 Les processus historiques contre les processus contemporains.....	12
1.8 Objectifs et structure de la thèse.....	15
Chapitre 2: Allopatric distribution of generalist parasites: interplay between post-glacial dispersal and host species. .....	19
2.1 Résumé/Abstract.....	20
2.2 Introduction.....	22
2.3 Materials and Methods.....	25
Inferring distribution of floater and fish species.....	25
Tests for allopatry among parasite species.....	26
Constrained null model analysis.....	26
Test for host-parasite associations – null model analysis.....	28

2.4 Results.....	29
2.5 Discussion.....	37
2.6 Acknowledgements.....	39
2.7 Author contributions.....	39
Chapitre 3: Historical process lead to false genetic signal of current connectivity among populations. ....	40
3.1 Résumé/Abstract.....	41
3.2 Introduction.....	43
3.3 Materials and Methods.....	45
Sampling.....	45
Genetic analyses.....	49
Genetic diversity distribution.....	49
3.4 Results.....	51
Distribution.....	51
Rock bass.....	51
Sunfish.....	57
Connectivity.....	58
3.5 Discussion.....	59
Impacts of colonization.....	59
The effect of slow changes in genetic diversity.....	61
Concluding remarks.....	62
3.6 Acknowledgements.....	63
3.7 Author contributions.....	63
Chapitre 4: Constrained exploration of population connectivity: a flexible way to assess hypotheses about dispersal processes. ....	64
4.1 Résumé/Abstract.....	65
4.2 Introduction.....	67
4.3 Materials and Methods.....	69
Description of the method.....	69
Simulation study.....	71
Applications.....	73
4.4 Results.....	74

Simulation study.....	74
Application to the creek chub case.....	79
Application to the colonial tunicate <i>Botryllus schlosseri</i> .....	81
4.5 Discussion.....	83
Simulation study.....	83
Illustrations of the proposed procedure with real data.....	85
Concluding remarks.....	87
4.6 Author contributions.....	87
 Chapitre 5 : Conclusion générale.....	88
5.1 L'impact de la dispersion historique à grande échelle spatiale.....	89
5.2 L'impact de la dispersion historique à petite échelle spatiale.....	90
5.3 Les méthodes développées dans cette thèse.....	94
À grande échelle spatiale.....	94
À petite échelle spatiale.....	95
5.5 Perspectives.....	96
 Références.....	99
 Annexes.....	113
A.1 Similarity values between floater and fish species.....	114
A.2 Genetic differentiation among sampling sites.....	118
A.3 The effect of the constraints on the effectiveness of the method.....	119

## LISTE DES TABLEAUX

<b>Table 2.1:</b> Relative C-score index (observed C-score divided by the theoretical maximal value for a given set, i.e., when species are in perfect allopatry) calculated for all species combined (global) and for every species pair, as well as the probabilities associated with their constrained and unconstrained null models. G: Giant floater; E: Eastern floater; N: Newfoundland floater. Moran's I for model residuals for each host species were not significant. ....	30
<b>Table 2.2:</b> Correlations among floater species based on their similarity values with fish species. Correlations based on $S_j$ , $S_p$ and $S_h$ values are shown, respectively, in the first correlation matrix above the diagonal; in the first correlation matrix below the diagonal; and in the second correlation matrix above the diagonal. Significant relationships are shown in bold. ....	31
<b>Table 2.3:</b> Significant associations between fishes and floater mussels according to the three similarity indices used in this study ( $S_j$ ; $S_h$ ; $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge and ?: unknown post-glacial origin. The fish species in bold was previously reported as hosts for floater mussels (Fuller, 1974; Trdan and Hoeh, 1982; Hoggarth, 1992; Watters <i>et al.</i> , 2005) and † indicates euryhaline species. Significant associations after Bonferroni corrections are indicated by the star symbol. ....	34
<b>Table 2.4:</b> Similarity values between floater species and previously reported host species (Fuller, 1974; Hoggarth, 1992; Trdan and Hoeh, 1982; Watters <i>et al.</i> , 2005), according to the three similarity indices used in this study ( $S_j$ ; $S_h$ ; $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge, I: Introduced and ?: unknown post-glacial origin. † indicates euryhaline species. Significant associations after Bonferroni corrections are indicated by the star symbol. ....	35

<b>Table 3.1:</b> Characteristics of the populations analyzed. The following information is provided for each sample site of both species: the sites of capture (letter refers to drainage system); the geographic coordinates and the altitude of the sites; the sample size (n); the total number of alleles (A); the total corrected number of alleles ( $A_c$ ); observed heterozygosity ( $H_O$ ); Nei's gene diversity ( $H_E$ ); $F_{IS}$ values; and Hardy-Weinberg probability ( $H_w$ ). Values in bold type refer to significant probability values before Bonferroni corrections. No probability values were significant after Bonferroni corrections. ....	47
<b>Table 3.2:</b> Hierarchical partitioning of the genetic variation among groups for rock bass and sunfish performed with Arlequin 1.2.3.4.5. Groups were defined according to drainages or groups inferred by STRUCTURE program. d.f., indicates degrees of freedom. Fixation index refers to $F_{CT}$ (variation among group) and $F_{SC}$ (variation among populations within group). Probability values are in parenthesis. ....	53
<b>Table 3.3:</b> Best linear models explaining the variation of Nei's gene diversity ( $H_E$ ) among sampling sites. Backward selection of the variables has been applied on a set of the variables including distance from Ottawa River (down), distance from the most upstream populations (up), altitude (alt), lake or river habitats (lake) and drainages (drainage). Adjusted $R^2$ and are provided. ....	57
<b>Table A1:</b> Similarity values between floater species and all fish species tested in this study, according to the three similarity indices used ( $S_j$ ; $S_h$ ; $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge, I: Introduced and ?: unknown post-glacial origin. Fish species in bold are previously reported hosts for floater mussels (Fuller, 1974; Trdan and Hoeh, 1982; Hoggarth, 1992; Watters <i>et al.</i> , 2005). Significant associations after Bonferroni corrections are indicated by the star symbol. ....	114

## LISTE DES FIGURES

- Figure 1.1 :** Hypothèse proposant qu'à grande échelle spatiale, la biodiversité serait surtout affectée par des processus historiques ; alors qu'à petite échelle spatiale, la biodiversité serait surtout affectée par des processus contemporains. La flèche courbe représente l'impact possible des processus historique à petite échelle spatiale, testé dans le cadre de cette thèse. ..... 15
- Figure 2.1:** Distributions of the Giant, the Eastern and the Newfoundland floater, across Quebec (modified from Clarke, 1981; and Cyr *et al.*, 2007). Dots represent sampled sites (Cyr *et al.*, 2007). A) Distribution of floater species compared to biogeographic zones (I-VIII) based on fish distributions defined by Legendre and Legendre (1984). B) Localization of the studied region in north eastern North Amercia. C) Breaks between genetic lineages of lake trout (*Salvelinus namaycush*) (I) and lake whitefish (*Coregonus clupeaformis*) (II), according to Wilson and Hebert (1998) and Bernatchez and Dodson (1991), respectively. ..... 23
- Figure 3.1:** Species and genetic distribution among sampling sites. The distribution of sunfish and rock bass in R and N drainages inferred from fish captures at each sampling site (A). The distribution of the different genetic groups assessed with STRUCTURE program for rock bass (B) and for sunfish (C) populations. The size of each sampling site refers to genetic diversity ( $H_E$ ). The gray and the black lines refer to the "la rivière Rouge" (R) and "la rivière du Nord" (N), respectively. Both rivers flow into the Ottawa River. ..... 46
- Figure 3.2:** Results of discriminant analyses of principal components (DAPC). Differentiation between drainage systems (the black and the grey lines refer to N and R drainages, respectively) for rock bass (A) and for sunfish (B). The probability to correctly assign an individual to the drainage where it was sampled is 99 % for rock bass and 88% for sunfish. ..... 52
- Figure 3.3:** Number of private alleles in each genetic group. The number (on the Y-axis) and the percentage (numbers above each bar) of private alleles within each

genetic group - identified by the STRUCTURE program - are indicated for rock bass populations (A) and for sunfish populations (B). Each genetic group is identified by the pattern used in Figure 1, B and C. ..... 55

**Figure 3.4:** Relationships between Nei's gene diversity (HE) and geographic variables. Scatter plots showing the correlation between Nei's gene diversity and distance from downstream most point for rock bass (Pearson's  $r = -0.49$ ;  $p = 0.0279$ ) and sunfish (Pearson's  $r = 0.23$   $p = 0.1349$ ) (A). Distance from the upstream most point for rock bass (Pearson's  $r = 0.27$ ;  $p = 0.1519$ ) and sunfish (Pearson's  $r = -0.49$   $p = 0.0075$ ) (B). ..... 56

**Figure 3.5:** Migration among populations and barriers to migration inferred from genetic data. Putative genetic barriers and interconnections are shown for rock bass (A) and for sunfish (B) populations. Black bars represent putative barriers within drainages, as identified with Monmonier's algorithm. Arrows represent putative connections between different arms of the two sampled drainages, for which no genetic boundaries were identified. Finally, recent migration events, as detected with Genclass 2, are shown in panel (C) and (D) for rock bass and sunfish populations, respectively. Intra-drainage migrants are identified with dotted arrows while inter-drainage migrants are identified with solid arrows. The gray and the black lines refer to R and N drainage, respectively. ..... 60

**Figure 4.1:** Schematization of the five steps required to build a minimum spanning network that is constrained spatially and genetically. ..... 69

**Figure 4.2:** The simulations scenarios used in this study. The scenarios are based on a combination of four connectivity patterns (A-D) and three migration rates (0.001, 0.01, and 0.1). Global  $F_{ST}$  for each combination of the four connectivity patterns and three migration rates are provided. ..... 72

**Figure 4.3:** Goodness of fit of the proposed procedure. Goodness of fit is evaluated as the mean proportion ( $\text{mean} \pm \text{SD}$ ) of true and false links correctly identified across samples for each simulated connectivity pattern tested in this study (A-D). For each

connectivity pattern, three migration rates (0.001; 0.01; 0.1) and four sample sizes (i: 10 loci - 20 inds; ii: 10 loci-50 inds; iii: 20 loci-20 inds; iv: 20 loci-50 inds) were tested. .... 76

**Figure 4.4:** Relationships between global differentiation ( $F_{ST}$ ) and the goodness of fit of the method. The combined relationships of all of the connectivity patterns (continuous line) and the relationships for connectivity pattern A (upper line) and D (bottom line) are shown. The performance of the method was evaluated using four distinct sample sizes (A: 10 loci - 20 inds; B: 10 loci-50 inds; C: 20 loci-20 inds; D: 20 loci-50 inds). .... 77

**Figure 4.5:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 10 loci and 20 individuals. .... 78

**Figure 4.6:** Connectivity network among the populations of the creek chub (*Semotilus atromaculatus*) in four drainages (A-D) of the Mastigouche Wildlife Reserve (Quebec, Canada). Solid lines represent connectivity links consistent with the permanent hydrological network. Dotted lines represent connectively links that are not consistent with the permanent hydrological network. Isolation within the connectivity network allowed for the identification of three distinct regions. See the original publication by Boizard *et al.* (2009) for further information about lake identification. .... 80

**Figure 4.7:** (A) Connectivity network inferred among ports (P) and marinas (M) invaded by *B. schlosseri* in Nova Scotia, Canada. The preliminary networks that served to build the final hypothetic network (see method) are shown in (B) and (C). (B) is the connectivity network inferred by the method when using a Delaunay triangulation and (C) is the connectivity network inferred from the boating routes. See the original publication by Lacoursière-Roussel *et al.* (2012) for more information about the identification of marinas and ports. .... 82

**Figure 5.1** : Effets historiques et actuels sur la répartition de la biodiversité. ....98

**Figure A1:** Representation of genetic differentiation among sampling sites. Correspondence analysis was also performed for rock bass (A) and sunfish (B). The first three axes of the analysis respectively explain 23.57%, 19.26%, and 15.79% of the variation among rock bass populations and 17.50 %, 16.38 %, and 10.63 % among sunfish populations. The position of each sampling site on each of the first three axes was coded as a color in the RGB color system (red for the first axis, green for the second axis, and blue for the third axis. ....118

**Figure A2:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 10 loci and 50 individuals. ....119

**Figure A3:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 20 loci and 20 individuals. ....120

**Figure A4:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 20 loci and 50 individuals. ....121

## LISTE DES ENCADRÉS

Encadré I : Les glaciations du Pléistocène.....	10
Encadré II : Le modèle de biogéographie insulaire.....	11

## LISTE DES SIGLES ET DES ABBRÉVIATIONS

A : Nombre total d'allèles / Refuge Atlantique	
AA : Avant aujourd'hui	
Ac : Nombre total d'allèles corrigé pour la taille de l'échantillon	
ADN / DNA : Acide désoxyribonucléique	
AMOVA : Analyse de variance moléculaire	
B : Botryllus	
C : Indice de co-occurrence C-score (Stone and Roberts, 1990)	
ca. : Approximativement	
CA : Analyse des correspondances	
DAPC : Analyse discriminante des composantes principales	
D <sub>CE</sub> : Distance de corde	
d.f. : Degré de liberté	
dNTP : Mélange des quatre désoxyribonucléotides	
E : Flotteur de l'Est	
Éco/Évo : Écologiques et évolutifs	
e.g. : <i>Exempli gratia</i> (par exemple)	
et al. : <i>et alii</i> (et collaborateurs)	
f. : flotteur	
F <sub>CT</sub> : Indice de variation génétique entre les groupes	
F <sub>IS</sub> : Indice de variation génétique des individus à l'intérieur des populations	
FQRNT : Fonds Québécois de Recherche en Nature et Technologie	
F <sub>SC</sub> : Indice de variation génétique entre les populations à l'intérieur des groupes	
F <sub>ST</sub> : Indice de variation génétique inter populations	
G : Flotteur Géant	
gr : Groupe	
H <sub>E</sub> : Hétérozygotie attendue	
H <sub>O</sub> : Hétérozygotie observée	
HW : Hardy Weinberg	
I : Espèce introduite	
i.e. : <i>id est</i> (c'est-à-dire)	
Inc : Incorpore	
inds : Individus	
k : Nombre de groupes	
KAM : Modèle de mutations avec un nombre fini d'allèles	
km : Kilomètres	
km <sup>2</sup> : Kilomètres carrés	
KY : Kilo années	
L : Nombre de loci	

LD : Déséquilibre de liaison gamétique  
m : Mètres / Taux de migration  
M : Refuge Mississippi / Marina  
MCMC : Chaîne de Markov Monte Carlo  
min. : Minutes  
mM : Milimole  
N : Rivière du Nord / Flotteur de Terre-Neuve / Nb d'occurrences  
Nb : Nombre  
Ng : Nano gramme  
nM : Nano mole  
NSERC : Conseil de recherches en sciences naturelles et en génie du Canada  
p : Fréquence de l'allèle p dans le modèle HW / probabilité  
P : Ports / Pyganodon  
PCR : Réaction en chaîne par polymérase  
pmol : Pico mole  
p-value : valeur de probabilité  
r : Coefficient de corrélation de Pearson  
R : Rivière Rouge  
 $R^2$  : Coefficient de determination  
s : Secondes  
SD : Écart-type  
 $S_j$  : Indice de similarité de Jaccard (Jaccard 1912)  
 $S_h$  : Indice de co-occurrence hôte/parasite à l'intérieur de la répartition de l'hôte  
 $S_p$  : Indice de co-occurrence hôte/parasite à l'intérieur de la répartition du parasite  
Spp : Espèces  
SIG : Système d'informations géographiques  
*Taq* : Enzyme polymérase de la bactérie thermophile *Thermus aquaticus*  
USA : États-Unis  
°C : Degré Celsius  
°N : Degré de latitude Nord  
°W : Degré de longitude Ouest  
 $(\delta\mu)^2$  : Distance génétique de Goldstein *et al.*, 1995  
μl : microlitre  
% : Pourcent  
≈ : À peu près égale  
? : Origine inconnue  
\* : Valeur significative de similarité  
† : Espèce euryhaline

## REMERCIEMENTS

Mes premiers remerciements vont à mon directeur de recherche Bernard Angers qui m'a supervisé tout au long de mes études gradués. Bernard, je te remercie d'avoir eu confiance en moi et de m'avoir encadré, sans me limiter. Mes seconds remerciements vont à tous ces gens que j'ai côtoyés au fil des ans, au laboratoire et qui sont aujourd'hui devenus mes amis. Si aujourd'hui j'écris mes remerciements de doctorat, c'est beaucoup grâce à vous ; les nombreuses discussions que nous avons eues m'ont permis de forger mes idées et de les confronter. L'ambiance bon enfant qui a régné au laboratoire toutes ces années a grandement favorisé ma motivation et explique le sentiment d'appartenance très fort que je ressens aujourd'hui envers l'équipe du laboratoire. Je tiens également remercier tous ceux qui m'ont filé un coup de main sur le terrain en laboratoire, ou pendant la rédaction de cette thèse. Votre contribution a grandement enrichi ce travail.

Je tiens à remercier ma famille qui m'a soutenu dans mes choix. J'ai la chance de faire partie d'une famille nombreuse et unie, ça facilite grandement les emprunts d'argent. C'est une blague évidemment, vous êtes une richesse dans ma vie, merci d'être là, simplement. Enfin, au moment où j'ai débuté cette thèse, j'ai rencontré une personne formidable qui m'a accompagné dans cette aventure et qui a eu la décence de me tolérer dans mes moments de distraction. Gabrielle, ta présence m'a permis de garder un certain équilibre entre le travail et la vraie vie, merci. Depuis maintenant 1 an et demi, il y a une petite furie qui fait la pluie et le beau temps dans ma vie. Merci à toi Juliette de m'avoir fait réaliser que le bonheur se trouve dans les yeux d'un enfant.

Bref, merci Bernard, Marie-Claire, Philippe (Girard), Marie-Claude, Rachel (Massicotte), Roland, Daniel, Martin, Claude-Olivier, Émilie (Castonguay), Mathieu, Christelle, Méloé, Simon (Laurin), Simon (Vervaet), Francis, Émilie (Reny-Nolin), Benoit, Antoine, Rachel (Morin), Hugo, Charles, Thomas, Maeva, Philippe (Lamarre), Logan, Léo, Pedro, Anaïs, Gabriel (Lanthier), Guillaume, Jean-Martin, toute l'équipe de Daniel Boisclair, Louise, Grégoire, Jean-François, Viviane, Benoît, Judith, Gabrielle et Juliette.

## **CHAPITRE 1 : INTRODUCTION GÉNÉRALE**

Suite à l'accroissement de la population humaine, la conservation de la biodiversité, littéralement la diversité des organismes vivants, est devenue l'un des enjeux majeurs de notre siècle. L'impact de l'homme étant de plus en plus important, un nombre croissant de chercheurs s'intéressent à quantifier l'effet des perturbations anthropiques sur la biodiversité. En effet, le nombre de recherches touchant de près ou de loin à cette problématique a grimpé en flèche au cours des dernières décennies, donnant naissance à plusieurs disciplines comme par exemple, l'écologie du paysage (voir Wu and Hobbs 2002 pour une revue) ou la génétique du paysage (voir Storfer *et al.* 2006 pour une revue). Sans surprise, les recherches effectuées, à ce jour, ont démontré l'impact des perturbations actuelles sur la biodiversité.

Cependant, le niveau actuel de biodiversité ne dépend pas uniquement de processus contemporains, mais plutôt d'une combinaison complexe de processus historiques et contemporains. Dès lors, pour obtenir un portrait juste et non biaisé des effets des perturbations actuelles sur la biodiversité, il est primordial de bien connaître les processus historiques qui ont modelé la biodiversité. À grande échelle spatiale, la biodiversité s'étant mise en place sur des centaines, voire des milliers d'années, la variabilité observée devrait refléter ces processus historiques. À petite échelle spatiale, toutefois, on peut présumer que les populations partagent une histoire commune récente. En conséquence, il est plus facile de relier la variabilité observée à des processus récents et actuels. Cette prémissse est-elle suffisante pour écarter définitivement les processus historiques comme facteurs explicatifs potentiels? L'emphase mise sur les processus actuels dans les études menées à petite échelle spatiale semble l'indiquer. Cette thèse avait pour but de creuser cette question.

## 1.1 La variabilité génétique

La variabilité génétique dépend à la fois de la diversité génétique et de son organisation spatiale. La variabilité génétique constitue le maillon fondamental de la biodiversité. Le destin des populations, l'apparition, tout comme l'extinction des espèces, dépend de la variabilité génétique (Lacy 1987). La variabilité génétique résulte de l'action de quatre forces évolutives : les mutations ; le flux génique; la sélection et la dérive génétique. Une mutation désigne un changement de la séquence de l'ADN qui résulte en la création d'une nouvelle version d'un gène (i.e. allèle) (Hamilton 2011). Le flux génique peut être défini comme le mouvement des allèles, d'une population vers une autre, soit par le mouvement des individus ou des gamètes (Hamilton 2011). La sélection peut être définie comme la propagation des allèles les mieux adaptés aux dépend des allèles moins bien adaptés. La dérive génétique peut être définie comme l'échantillonnage stochastique des gamètes à chaque génération lors de la reproduction sexuée et résulte en la perte d'allèles au fil du temps (Hamilton 2011). L'action de la dérive génétique est modulée par la taille efficace des populations (i.e. nombre d'individus contribuant au pool génique de la génération suivante) ; plus la taille efficace est petite, plus la dérive génétique agira de manière forte et rapide au sein des populations et inversement plus la taille efficace est grande, plus l'action de la dérive en sera ralentie (Lacy 1987).

Les forces évolutives sont reliées à plusieurs processus écologiques. i) Les caractéristiques environnementales de même que les interactions biotiques constituent des pressions de sélection qui vont influencer la survie et le succès reproducteur des individus. ii) Le flux génique est très souvent fortement corrélé à la capacité de dispersion des espèces (Bohonak 1999). iii) La structure des populations (i.e. ratio mâle femelle ; variance du succès reproducteur ; variance du flux génique entre les populations) couplée aux variations démographiques va influencer la taille efficace et la dérive génétique (Hamilton 2011).

De plus, la trajectoire évolutive des espèces et les processus écologiques sont modulés par des facteurs du paysage historiques et actuels (*e.g.* taille, structure et qualité des habitats, présence de corridors ou de barrières modulant la migration ; Dyer *et al.* 2010) et par des facteurs intrinsèques aux espèces (*e.g.* adaptations au

climat, taux d'accroissement des populations, capacité de dispersion, système de reproduction). Par exemple la répartition actuelle des poissons dans le nord-est de l'Amérique du Nord découle de la topologie historique et actuelle des réseaux hydrographiques, de la capacité de nage des différentes espèces et de leurs préférences environnementales (*e.g.* tolérance au froid) (Mandrak et Crossman 1992 ; Legendre et Legendre 1984 ; voir également le chapitre 2 de cette thèse)

## 1.2 L'estimation de la variabilité génétique

La variabilité génétique d'une espèce dépend de la diversité génétique à l'intérieur et entre les populations. Compte tenu de la taille importante des génomes et de la présence de régions plus ou moins variables à l'intérieur des génomes (Pray 2008), il est difficile d'obtenir une estimation juste de la diversité génétique au sein d'une espèce ou d'une population. Cependant, une estimation relative de la diversité génétique peut facilement être obtenue à partir du nombre d'allèles moyens à plusieurs loci. Un locus pouvant être défini comme une région particulière du génome et les allèles comme les variants présents à cette région.

Cette estimation simple de la diversité génétique est cependant peu précise. Si le nombre d'allèles est important pour déterminer la diversité génétique, la fréquence relative de ces allèles l'est tout autant. Une répartition équitable des allèles à l'intérieur d'une population étant une indice d'une plus grande diversité génétique (*e.g.* plusieurs allèles ayant la même fréquence à l'intérieur d'une population) qu'une répartition inéquitable (*e.g.* un allèle dominant et plusieurs allèles rares). La diversité génétique de Nei (1983) tient à la fois compte du nombre d'allèles et de leur fréquence relative. La diversité génétique de Nei peut donc être considérée comme un estimateur plus précis de la diversité génétique que le nombre d'allèles. On peut facilement faire un parallèle entre ces deux estimateurs et les mesures de richesse et de diversité spécifique couramment utilisées en écologie pour comparer la diversité des communautés.

La répartition spatiale de la diversité génétique est une composante importante de la variabilité génétique. Les indices de fixation développés par Wright (1969) permettent d'estimer dans quel compartiment la variabilité génétique se

retrouve : à l'intérieur des populations ; entre les populations ; ou entre les groupes de populations (*e.g.* différentes régions ou sous-espèces ; Excoffier *et al.* 1992). À titre d'exemple, l'indice  $F_{ST}$  peut être défini comme la corrélation des allèles entre les individus d'une même population par rapport à l'ensemble des populations. En pratique il se calcule à partir du ratio entre la variabilité allélique inter populations sur la variabilité allélique totale (Weir & Cockerham 1984). Le  $F_{ST}$  est borné entre 0 et 1 et augmente avec la différenciation génétique. La variabilité génétique augmente avec la différenciation et ce même si la diversité génétique demeure stable.

### **1.3 Un déséquilibre entre la perte et le gain de la variabilité génétique**

La variabilité génétique résulte de l'action combinée des forces évolutives et peut être vue comme un équilibre entre la perte et le gain de diversité génétique dans le temps et l'espace. Cependant, la perte de la variabilité génétique est généralement beaucoup plus rapide que le gain. Ainsi, les événements provoquant la perte de variabilité génétique laisseront une empreinte durable, car ils ne seront contrebalancés que très lentement par le gain (Amos and Harwood 1998). La variabilité génétique actuelle reflète donc principalement les événements actuels ou historiques qui ont mené à une perte de variabilité génétique.

À titre d'exemple, une population ayant subi une importante baisse d'effectif va voir sa diversité génétique chutée, simplement par la perte d'individus. En contrepartie, si la population retrouve sa taille originale, l'augmentation de diversité génétique résultant des mutations se fera dans un laps de temps beaucoup plus long. Aussi, les chutes d'effectifs historiques seront décelables très longtemps, alors que les expansions démographiques (récentes) seront beaucoup plus difficiles à détecter (Kimmel *et al.* 1998).

Le niveau de différenciation entre des populations est généralement le résultat d'un équilibre entre la migration et la dérive génétique. La migration peut homogénéiser le pool génique de différentes populations très rapidement et très efficacement. En effet, il suffit d'un migrant par génération pour que des populations maintiennent la même composition allélique (Allendorf 1993). En revanche, la différenciation des populations par dérive génétique est potentiellement très lente.

Lorsque la migration est faible et les populations sont grandes, l'atteinte d'un équilibre entre la migration et la dérive génétique peut s'échelonner sur des milliers de générations (Whitlock 1992). Aussi, la migration entre deux populations est beaucoup plus facile à détecter que l'isolement.

#### **1.4 L'étude de la variabilité génétique**

Historiquement, l'étude de la variabilité génétique s'est d'abord effectuée sur la base de la répartition des espèces (Coxe and Moore 2000). Les comparaisons entre des espèces proches taxonomiquement permettant d'inférer les processus évolutifs ayant conduit à la spéciation. Par exemple, l'étude de la répartition d'espèces apparentées a permis de révéler l'effet de la dérive des continents sur l'évolution des espèces (Cracraft 1973). Le développement des techniques moléculaires a permis de porter les inférences à un grain taxonomique plus fin, soit au niveau des populations et des individus. Enfin, le développement des analyses écologiques spatiales a permis de relier la variabilité génétique à la composition et à la configuration du paysage. Parallèlement à ces avancées techniques, les disciplines de la biogéographie, la phylogéographie et la génétique du paysage se sont développées. Ces trois disciplines se distinguent par les échelles taxonomiques, spatiales et temporelles considérées.

#### *La biogéographie*

La biogéographie est la science qui tente d'expliquer la répartition des espèces dans le temps et l'espace. Les études biogéographiques sont très souvent conduites à grande échelle spatiale dans une perspective historique. L'approche préconisée est souvent multidisciplinaire et fait entre autre appel à l'écologie, à l'évolution, à la géologie et à la climatologie (Cox et Moore 2000). La biogéographie actuelle doit beaucoup aux travaux de Sir Alfred Russel Wallace qui en 1872 a divisé le monde en 6 grandes régions sur la base de la répartition des espèces animales. Les travaux de Wallace ont permis de souligner les effets des caractéristiques environnementales (climat ; végétation) et des barrières à la dispersion (chaîne de montagnes ; océans) sur la répartition des espèces animales.

La biogéographie a fourni les premiers modèles permettant d'expliquer la divergence génétique, dans un cadre géographique. En accord avec les résultats obtenus par Wallace, le modèle de spéciation allopatrique met l'emphase sur la présence de barrières physiques à la dispersion pour expliquer l'apparition de nouvelles espèces (Mayr 1963). Ce modèle reconnaît deux processus pouvant mener à la spéciation : soit la dispersion ou la vicariance (Avise 2000). La spéciation par dispersion suppose que certains individus, d'une première espèce, se seraient dispersés au-delà d'une barrière géographique pré-existante, pour donner naissance à une seconde espèce. La spéciation vicariante suppose que l'apparition d'une barrière dans le paysage aurait scindé la répartition d'une espèce (jusqu'alors continue) en deux ; donnant ainsi naissance à deux espèces distinctes. D'autres modèles de spéciation ont été bien sûr proposés. La spéciation péripatique et la spéciation parapatique sont dérivées du modèle allopatrique, mais suppose la présence de barrières géographiques plus ou moins étanches et des niveaux différents de flux génique et de sélection. Enfin, le modèle de spéciation sympatrique (Maynard Smith 1966) propose que la spéciation puisse arriver en l'absence de barrière géographique, alors que des adaptations divergentes sont progressivement renforcées par la sélection naturelle.

### *La phylogéographie*

Avec le développement des marqueurs moléculaires est née la phylogéographie. La phylogéographie est une discipline très jeune (le terme est apparu pour la première fois dans la littérature en 1987 (Avise *et al.* 1987)). Issue de la biogéographie, la phylogéographie est basée sur la répartition des populations et des individus, plutôt que sur la répartition des espèces. Malgré un grain taxonomique plus fin, les études phylogéographiques, à l'instar des études biogéographiques, sont la plupart du temps conduites à grande échelle spatiale, dans une perspective historique. Le terme phylogéographie se décompose en deux parties soit le préfixe « phylo » issu du grec signifiant tribu ou race et le terme « géographie ». Cette décomposition illustre bien les deux objectifs principaux de cette discipline soit, dans un premier temps, l'identification des lignées évolutives distinctes, au sein des espèces, à l'aide de méthodes phylogénétiques et dans un deuxième temps, l'étude de la répartition géographique de ces lignées évolutives. Outre l'identification de futures

espèces potentielles, l'identification de lignées évolutives distinctes peut servir de base à l'élaboration de projets écologiques ou de conservation d'une plus grande acuité taxonomique (*e.g.* Moritz 2002).

La phylogéographie s'est développée au rythme des techniques moléculaires ; la puissance des analyses phylogéographiques augmentant avec l'amélioration de ces techniques. Historiquement, les études phylogéographiques utilisaient surtout des marqueurs mitochondriaux. De nos jours, nombreux marqueurs moléculaires sont disponibles. Les différents marqueurs ont évidemment différentes propriétés (héritabilité ; taux de mutations ; polymorphisme ; neutralité ; dominance ; représentativité du génome ; voir Vignal *et al.* 2002 pour une revue), le choix d'un type de marqueur devra donc se faire en fonction des besoins de l'étude.

### *La génétique du paysage*

La génétique du paysage est encore plus jeune que la phylogéographie ; le terme est apparu pour la première fois dans la littérature en 2003 (Manel *et al.* 2003). La génétique du paysage tire partie du développement des analyses écologiques spatiales pour tester formellement l'effet de la composition et de la configuration du paysage sur la variabilité génétique (Storfer *et al.* 2007). De nombreuses analyses écologiques spatiales ont pu servir dans l'interprétation de données génétiques, comme par exemple le test de Mantel (1967 ; voire *g.* Boizard *et al.* 2009 pour un exemple d'application), l'analyse canonique des correspondances (*e.g.* Angers *et al.* 1999), les corréogrammes, basés sur l'indice de Moran ou le « r » de Pearson (*e.g.* Arnaud 2003) ainsi que plusieurs analyses de modélisation (*e.g.* Arens *et al.* 2007).

Contrairement aux études biogéographiques et phylogéographiques, les études de génétique du paysage sont réalisées sur de petites échelles spatiales, dans une perspective récente. Une autre distinction souvent évoquée entre ces trois disciplines est le traitement des variables explicatives. En effet, dans les études biogéographiques et phylogéographiques, à l'opposé de ce qui est observé dans les études de génétique du paysage, certaines variables sont souvent proposées comme explication ad hoc et ne sont pas testées formellement. Toutefois, les analyses

phylogéographiques et biogéographiques incorporent de plus en plus des analyses spatiales complexes, rendant cette distinction caduque.

### **1.5 Les processus historiques et les processus contemporains**

À ce jour, les études biogéographiques et phylogéographiques nous ont permis de constater l'importance des changements climatiques historiques et de la capacité de dispersion des espèces sur la répartition de la biodiversité. Les glaciations du Pléistocène (Encadré I) constituent sans nul doute un événement historique majeur expliquant la répartition actuelle de la biodiversité. En accord avec le modèle de spéciation allopatrique (Mayr 1963), les populations ayant occupé des refuges glaciaires différents ont divergé les unes des autres ; donnant ainsi naissance à des lignées évolutives différentes caractérisées par des trajectoires évolutives différentes. La répartition actuelle des populations et des espèces a grandement été influencée par la localisation des refuges glaciaires (Crossman et McAllister 1986) ; les organismes se dispersant depuis différents refuges glaciaires n'ayant pas pu bénéficier des mêmes opportunités de colonisation. En accord avec les prédictions du modèle de biogéographie insulaire de MacArthur et Wilson (1967 ; Encadré II), des pertes de diversité spécifique (Legendre et Legendre 1984) et génétique (Bernatchez et Wilson 1998) sont observées à mesure qu'on s'éloigne des refuges glaciaires. Aussi, les communautés, occupant aujourd'hui un territoire anciennement prisonnier des glaciers, ont une faible diversité spécifique et génétique et elles ont pour la plupart expérimenté de graves effondrements démographiques entraînant d'importants effets fondateurs au sein des populations Bernatchez et Wilson 1998).

Les études en génétique du paysage ont permis de révéler les impacts du paysage actuel sur les processus évolutifs. Jusqu'à présent, les études en génétique du paysage se sont surtout attardées à caractériser l'effet des éléments du paysage sur le flux génique. Les principaux patrons spatiaux observés sont l'isolement par la distance (*e.g.* Broquet *et al.* 2006) ; la présence de barrières partielles ou complètes au flux génique (*e.g.* Leclerc *et al.* 2008) ; et l'influence de la qualité de l'habitat sur la dispersion (*e.g.* Vos *et al.* 2001). Cependant, certaines études ont pu révéler l'effet du paysage sur d'autres forces évolutives. À titre d'exemple, mentionnons les

travaux de Hitching et Beebee (1997) qui ont observé une diversité génétique plus faible chez des populations de grenouilles communes, vivant en milieu urbain, en comparaison aux populations vivant en milieu rural. Une telle observation suggère que les populations urbaines ont de plus petites tailles efficaces et sont davantage isolées. Mentionnons également les travaux de Reush *et al.* (2001) qui ont observé des populations d'épinoches à trois épines divergeant selon le type d'habitat plutôt que la proximité géographique, suggérant la présence d'adaptations locales et de la sélection d'habitats.

### Encadré I : Les glaciations du Pléistocène

L'époque du Pléistocène a débuté il y a 1,8 MA, par un refroidissement général du climat ayant provoqué d'intenses bouleversements dans la répartition des espèces (Pielou 1991). D'immenses glaciers ont pris d'assaut les continents emprisonnant sous les glaces d'immenses territoires, provoquant tour à tour l'écrasement isostatique de la croûte terrestre et la baisse du niveau des océans. L'avènement de ces immenses glaciers a provoqué la destruction de nombreux écosystèmes ; les espèces, extirpées de ces territoires, ont dû trouver refuge en marge des glaciers.

Le Pléistocène est entrecoupé par de brèves périodes interglaciaires d'environ 50 000 ans caractérisées par un réchauffement du climat et par la fonte des glaces. La fonte de gigantesques masses de glace a provoqué l'avènement d'immenses lacs, la brusque remontée du niveau des océans et conséquemment d'importantes incursions marines sur les continents, encore écrasés par le poids des glaciers (Pielou 1991).

Les glaciations du Pléistocène ont donc constituées une période d'intenses bouleversements dans le paysage par la compression des habitats, des biomes, et par des changements climatiques et hydrologiques dramatiques. Les impacts géologiques des premières périodes glaciaires et interglaciaires ont été effacés par l'avancée des glaciers lors de la dernière période glaciaire (dites Wisconsinne), de 80 000 à 18 000 AA (Mandrack et Crossman 1992). Nous vivons actuellement la dernière période interglaciaire (l'Holocène), alors que la fonte des glaces, en Amérique du Nord, s'est échelonnée de 18 000 à 7000 ans AA (Dyke et Prest 1987).

## Encadré II : Le modèle de biogéographie insulaire

Le modèle de biogéographie insulaire de MacArthur et Wilson (1967) constitue un des premiers modèles prédictifs en biogéographie, ce modèle tente de prédire la richesse en espèces d'une île en fonction de sa taille et de la distance qui la sépare du continent. Bien que conçu pour des îles, ce modèle peut être appliqué à n'importe quel habitat fragmenté. Le modèle en île sous-tend que les espèces sont écologiquement équivalentes (i.e. neutralité) et que la dispersion explique la répartition de la biodiversité. À défaut d'être exhaustif le modèle en île peut servir de modèle nul afin de savoir si oui ou non d'autres processus sont en cause.

## 1.6 Les processus historiques ou les processus contemporains

Les études phylogéographiques ont démontré l'incidence des processus historiques sur la variation génétique spatiale à grande échelle spatiale, alors que les études en génétique du paysage ont démontré l'influence du paysage actuel, à petite échelle spatiale.

La différence de cadre temporel entre les études phylogéographiques et les études de génétique du paysage s'explique partiellement par la résilience des caractéristiques du paysage à grande échelle spatiale. Les éléments du paysage ayant eu, historiquement, un impact sur la répartition de la biodiversité à grande échelle spatiale sont soit encore présents (*e.g.* chaîne de montagnes ; océans), soit suffisamment importants pour avoir laissé des traces géologiques tangibles (*e.g.* glaciers ; lacs glaciaires ; mers intérieures). Il devient, dès lors, aisément de corrélérer ces éléments historiques du paysage à la répartition actuelle de la biodiversité. En contrepartie, à petite échelle spatiale, les caractéristiques historiques du paysage n'ont laissé que peu de traces perceptibles. Or, il est plus facile d'estimer les effets visibles du paysage actuel que les effets cryptiques du paysage historique. Cependant, il est probable que certains effets historiques viennent brouiller l'interprétation qu'on puisse faire des données génétiques en termes de signaux contemporains.

Bénéficiant de données sur le paysage historique, certaines études ont néanmoins cherché à discriminer les effets des processus historiques et contemporains sur la variabilité génétique, à petite échelle spatiale (*e.g.* Vandergast *et al.* 2007). Ainsi, certaines études ont démontré un effet marqué des processus historiques (Fraser and Bernatchez 2005), un effet marqué des processus actuels (Zellmer and Knowles 2009), ou les deux (Dyer *et al.* 2010). Par exemple, la détection de deux groupes génétiques d'omble de fontaine (*Salvelinus fontinalis*) dans le lac Mistassini suggère la présence de deux groupes fondateurs. Cette hypothèse est consistante au changement de l'exutoire du lac, suite au soulèvement isostatique dans la région (Fraser and Bernatchez 2005). Spear and Storfer (2008) ont trouvé que la variabilité génétique entre des populations de grenouilles à queue (*Ascaphus truei*) était davantage représentative du paysage historique ( $\approx 1970$ ) que du paysage actuel. Au contraire, Zellmer and Knowles (2009) ont démontré que la divergence entre des populations de grenouille des bois (*Rana sylvatica*) est davantage corrélée au paysage actuel qu'au paysage historique ( $\approx 1800$ ). Enfin, la différentiation génétique des populations d'*Orthonyx temmii*, une espèce d'oiseaux de la forêt pluviale, est corrélée à la fois au paysage actuel et au paysage historique ( $\approx 1800$ ) (Pavlacky *et al.* 2010).

### **1.7 Les processus historiques contre les processus contemporains.**

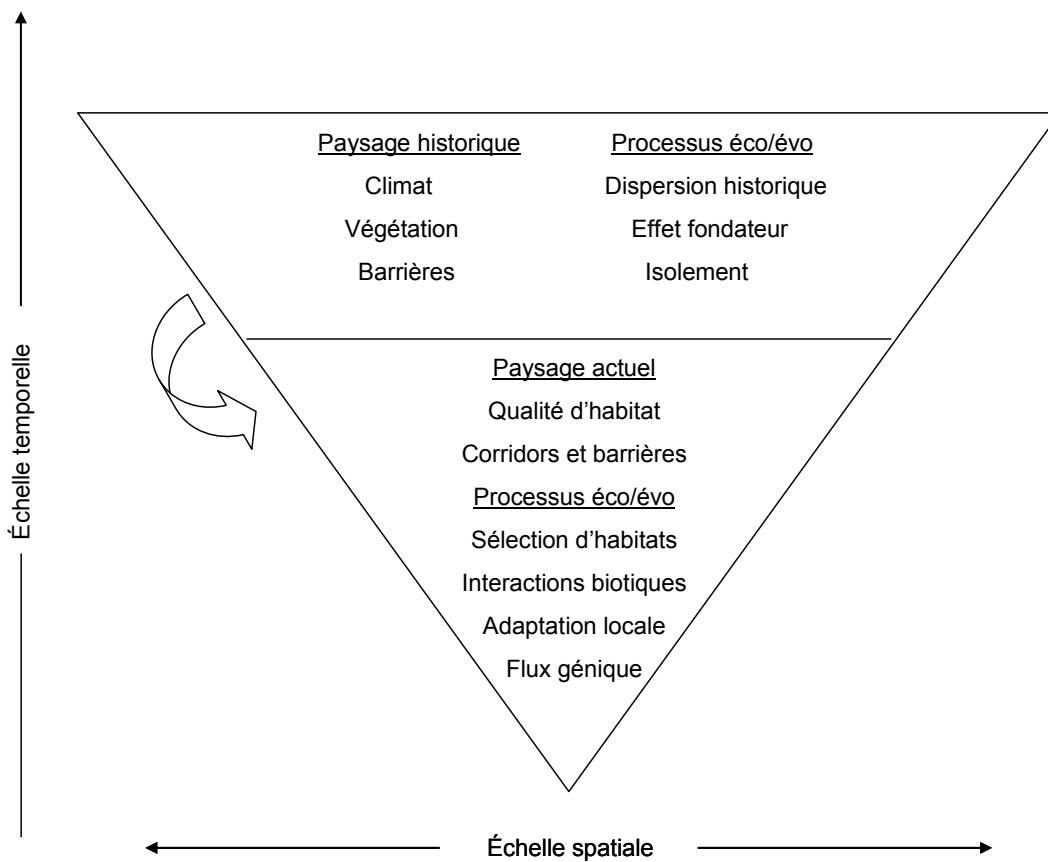
La variabilité génétique reflète des processus historiques et des processus contemporains. L'adoption d'un cadre temporel trop restreint apparaît donc comme un biais potentiel important dans l'interprétation des données génétiques (Allendorf 2010). Whitlock and McCaulay (1999) suggèrent même que l'estimation des processus contemporains à partir de la variabilité génétique n'est pas valable en raison des effets historiques potentiels. La génétique du paysage a connu un essor fulgurant, au cours de la dernière décennie (Sork and Waits 2010, Storfer *et al.* 2010). Cependant, de nombreuses études de génétique du paysage ne prennent pas en compte les effets potentiels des processus historiques, alors que l'importance et l'étendue des effets historiques demeurent à quantifier. Les processus historiques peuvent potentiellement induire : i) des inconsistances entre la variabilité génétique et le paysage actuel, provoquant une baisse du niveau de corrélation ; ii) de fausses relations de causalité entre la variabilité génétique et le paysage actuel. En effet, si

des processus écologiques actuels sont corrélés à des processus historiques, il devient pertinent de contrôler pour les effets historiques avant d'évaluer la portion de la variabilité génétique expliquée par les processus actuels (Dyer *et al.* 2010).

En revanche, plusieurs facteurs pourraient potentiellement atténuer l'ampleur des effets historiques : i) la stabilité du paysage ; ii) la nature des perturbations ; iii) l'espèce visée par l'étude ; iv) les marqueurs génétiques ; v) les méthodes statistiques et vi) l'échelle spatiale.

- i) La variabilité génétique a davantage de chance d'être corrélé au paysage actuel si ce dernier est stable, depuis longtemps (Anderson *et al.* 2010). La stabilité du paysage ne diminue pas les effets historiques sur la variabilité génétique, mais elle engendre la stabilité des processus écologiques et évolutifs. Dès lors, l'estimation des processus écologiques et évolutifs à partir de la variabilité génétique est représentative des processus courants.
- ii) La variabilité génétique actuelle reflète davantage les évènements qui ont conduit à la perte de variabilité génétique (Amos and Harwood 1998). On peut donc supposer que certaines perturbations humaines, ayant par exemple conduit à une réduction de la taille des populations naturelles, seront détectables rapidement au niveau de la variabilité génétique.
- iii) L'ampleur des effets historiques dépend de l'espèce étudiée. Le décalage entre une perturbation et son effet sur la variabilité génétique se calcule en nombre de générations. Ainsi, plus une espèce à un temps de génération court, plus sa variabilité génétique est susceptible de refléter des processus récents (Anderson *et al.* 2010).

- iv) L'ampleur des effets historiques dépend du type de marqueur utilisé. En effet, des marqueurs caractérisés par un taux de mutations et un polymorphisme élevés, comme par exemple les microsatellites, vont refléter plus rapidement les processus évolutifs et seront donc davantage influencés par les processus récents que des marqueurs moins variables (Estoup and Angers 1998 ; Holderegger and Wagner 2008).
- v) Les méthodes statistiques utilisées seront influencées à un degré variable par les processus historiques. Certaines méthodes intègrent les processus évolutifs sur nombreuses générations et les résultats obtenus à partir de ces méthodes sont susceptibles d'être biaisés par les processus historiques (*e.g.* Hudson *et al.* 1987 ; Rousset 1997). En revanche, certaines méthodes ont une fenêtre d'inférence beaucoup plus étroite et sont donc moins sensibles aux effets historiques (*e.g.* Nei and Tajima 1981 ; Paetkau *et al.* 2004). Cependant, l'efficacité des analyses génétiques étant généralement influencée par la variabilité génétique (Waples 1998), les effets historiques vont néanmoins influencer indirectement les performances de ces méthodes.
- vi) Il est probable que l'importance relative des processus historiques par rapport aux processus contemporains dépende de l'échelle spatiale considérée. Les différences génétiques que ce soit entre deux individus ; deux populations ; ou deux espèces intègrent les changements survenus depuis le plus récent ancêtre commun. Puisque des populations proches sont davantage susceptibles d'avoir une histoire commune récente que des populations éloignées, de façon générale les différences génétiques entre des populations proches devraient refléter des processus récents, alors que les différences génétiques entre des populations éloignées devraient refléter des processus historiques. Cette hypothèse, illustrée à la figure 1.1, permet d'expliquer partiellement pourquoi, dans les études à grande échelle spatiale, l'emphase est mise sur les processus historiques, alors que dans les études à petite échelle spatiale, l'emphase est plutôt mise sur les processus contemporains.



**Figure 1.1 :** Schéma illustrant l'hypothèse qu'à grande échelle spatiale, la biodiversité serait surtout affectée par des processus historiques ; alors qu'à petite échelle spatiale, la biodiversité serait surtout affectée par des processus contemporains. La flèche courbe représente l'impact possible des processus historique à petite échelle spatiale, testé dans le cadre de cette thèse.

## 1.8 Objectifs et structure de la thèse

L'objectif principal de cette thèse était d'évaluer l'impact de la dispersion historique sur la répartition de la biodiversité, à grande et à petite échelle spatiale. L'hypothèse conceptuelle, présentée à la figure 1.1, a été utilisée comme hypothèse de travail. Ainsi, à grande échelle spatiale, un impact important de la dispersion historique était escompté, alors qu'à petite échelle spatiale un impact plus marginal de la dispersion historique était attendu. Plusieurs questions connexes ont également été soulevées : i) Quels éléments du paysage, quels processus évolutifs ou écologiques ont influencés la dispersion historique ? ii) Quelle est l'influence de la dispersion historique sur les processus écologiques et évolutifs contemporains ? iii) Quels sont les interactions entre les processus locaux et les processus à grande échelle spatiale ?

Dans le premier volet de cette thèse, l'interaction entre la dispersion historique et les interactions biotiques chez un parasite généraliste a été testée. Les moules d'eau douce du genre Flotteur (*Pyganodon spp*) ont été utilisées comme modèles biologiques. Les moules d'eau douce se dispersent principalement au stade larvaire, comme parasite des poissons. Les espèces du genre flotteur peuvent parasiter un grand nombre d'espèces de poissons ; à ce jour, plus d'une trentaine d'espèces hôtes ont été répertoriés (Fuller 1974). L'objectif de cette étude était d'évaluer si les associations hôtes parasites permettent d'expliquer la répartition des espèces de Flotteur retrouvées sur le territoire québécois.

Une approche de modèles nuls a été développé afin de déterminer : i) si des espèces de flotteurs, présentant des répartitions disjointes, sont associées à des espèces différentes de poissons ; ii) et si ces associations permettent d'expliquer la répartition des flotteurs. Les questions soulevées dans ce premier volet rejoignent les questions générales de la thèse. Quel est l'impact de la dispersion historique sur la répartition de la biodiversité à large et à petite échelle spatiale ? Quel est l'effet de la dispersion historique sur les interactions biotiques ? Et réciproquement, quel est l'impact des interactions biotiques sur la dispersion historique ?

À prime abord, le choix d'un parasite généraliste comme organisme modèle peut sembler contre-intuitif puisque la détection des associations hôtes-parasites se complexifie avec l'augmentation du nombre d'hôtes potentiels. En revanche, la capacité d'infecter plusieurs hôtes rend d'autant plus intéressant les différences de répartition observées entre les différences espèces parasites. D'une part, ces différences peuvent révéler des préférences distinctes chez les espèces parasites quant au choix des hôtes ou même une évolution dans la capacité d'infecter tel ou tel hôte. Un changement dans le choix des hôtes est d'ailleurs rapporté comme un mécanisme évolutif majeur expliquant la diversification des taxons parasites (e.g. Funk *et al.*, 1995; Zuber and Widmer, 2000; Thorogood *et al.*, 2008). D'autre part, ces différences si elles ne sont pas couplées à l'infection d'hôtes distincts peut s'avérer révélatrice de patrons de dispersion distincts au sein des populations d'hôtes. En effet, la présence d'une faune parasite distincte chez des populations d'hôtes peut être considérée comme une indication d'un isolement à tout le moins partiel entre ces populations. En ce sens, plus la capacité d'infection d'un parasite est grande plus la

différence de répartition entre des espèces parasites peut être révélateur de divergences profondes quant à la dispersion au sein de la communauté de poissons.

L'objectif du second volet de cette thèse était de déterminer les impacts de la dispersion historique sur la répartition de la variation génétique, à petite échelle spatiale. Pour ce faire, la variation génétique spatiale à l'intérieur de deux réseaux de drainages adjacents a été déterminée pour deux espèces morphologiquement semblables, soit le crapet soleil (*Lepomis gibbosus*) et le crapet de roche (*Ambloplites rupestris*). Ces deux espèces ont des niches écologiques et des capacités natatoires semblables (Brinsmead and Fox 2002) ; une répartition semblable de la variation génétique chez ces deux espèces était donc attendue, sous réserve d'une histoire de colonisation commune. Ainsi, les différences observées entre les deux espèces ont pu être attribuées à une histoire distincte de colonisation. Il n'est pas aisés de distinguer les effets de la dispersion historique de ceux de la dispersion contemporaine. À cette fin, plusieurs indices génétiques ont été utilisés et plusieurs prédictions ont été formulées quant aux attendus de ces indices sous différents scénarios de dispersion. Ce second volet a permis de tester à l'aide de données empiriques la prédiction centrale du modèle conceptuel proposé (Figure 1.1) À savoir : à petite échelle spatiale, est-ce que les effets contemporains sont plus importants que les effets historiques pour expliquer la variation génétique spatiale ? Ainsi, les résultats obtenus pourraient permettre d'évaluer la crédibilité du modèle proposé, bien qu'une seule étude soit probablement insuffisante pour le valider ou l'invalider complètement.

Pour ce second volet de thèse, deux espèces de centrarchidés ont été sélectionnées comme organismes modèles. Deux arguments expliquent ce choix. Premièrement, tel que rapporté dans une étude récente ces deux espèces sont très fréquentes et très abondantes dans la région échantillonnée (Senay 2009). Deuxièmement – et c'est là notre argument principal - ces deux espèces présentent des niches écologiques très similaires, de sorte, que les différences observées entre les deux espèces pourront plus facilement être associées à des différences de dispersion historique entre les deux espèces. En effet, ces deux espèces ont un habitat, une alimentation et un comportement, lors de la reproduction, similaires (Bernatchez et Giroux, 2000 ; Brinsmead and Fox 2002). De plus, ces deux espèces

ne présentent pas de différences significatives quant à leur capacité de nage (Senay 2009). Enfin, dans la région échantillonnée, ces deux espèces présentaient des préférences environnementales tout à fait similaires (Senay 2009).

L'approche proposée dans le second volet de cette thèse, bien que rigoureuse, est très intuitive. Afin de combler cette lacune, l'objectif du troisième volet de cette thèse était de développer une approche formelle permettant l'identification des routes de dispersion entre des populations. Le résultat obtenu est un réseau spatial minimisant la distance génétique. Ce résultat, très visuel, facilite l'identification des éléments du paysage affectant la dispersion. De plus, la souplesse de la méthode permet la comparaison entre plusieurs hypothèses de dispersion. La méthode pourrait donc être utilisée pour comparer des hypothèses de dispersion basées sur le paysage historique et sur le paysage actuel. Ainsi, la méthode pourrait permettre, dans une large mesure, d'évaluer les impacts historiques et contemporains sur la variabilité génétique des espèces. L'algorithme a été testé sous plusieurs scénarios simulés de dispersion ; ces scénarios représentant un gradient de complexité. Afin d'illustrer le large potentiel de la méthode, les données de deux études très contrastantes, ont également été ré-analysées.

**CHAPITRE 2 : ARTICLE 1****ALLOPATRIC DISTRIBUTION OF GENERALIST PARASITES:  
INTERPLAY BETWEEN POST-GLACIAL DISPERSAL AND  
HOST SPECIES**

Frédéric Cyr, Pedro Peres-Neto, Bernard Angers

Publié dans la revue Canadian Journal of Fisheries and Aquatic Sciences (2012) **69**:  
1491-1498.

## 2.1 Résumé

Les moules d'eau douce se dispersent principalement au stade larvaire en tant que parasites des poissons. Dans le nord-est de l'Amérique du Nord, les moules du genre flotteurs (*Pyganodon spp.*) présentent des répartitions allopatriques dans un patron est-ouest qui correspond peu au patron biogéographique observé pour les communautés de poissons dans la région. Cette étude vise à déterminer si des associations avec des espèces distinctes de poissons peuvent expliquer les répartitions des espèces de flotteurs. Nous avons donc développé une série de modèles nuls conçus pour élucider la dispersion des espèces dont la répartition est liée à celles de d'autres espèces. Le patron de répartition des espèces de flotteurs a été testé en contrôlant pour les associations hôte-parasite, alors que ces associations ont été déterminées à partir de modèles nuls basés sur les patrons de co-occurrence. La répartition disjointe du Flotteur Géant (*P. grandis*) et du Flotteur de l'Est (*P. cataracta*) ne peut être expliquée par des associations distinctes et résulte probablement d'associations avec des populations distinctes plutôt que des espèces distinctes. En contrepartie, les associations distinctes du Flotteur de Terre-Neuve (*P. fragilis*) avec des espèces de poissons euryhalins permettent d'expliquer sa répartition. Ces associations distinctes ont également pu favoriser la différenciation entre le Flotteur de Terre-Neuve et son taxon sœur : le Flotteur de l'Est. Cette étude a démontré les effets des associations biologiques positives sur les patrons de répartitions et le grand potentiel des modèles nuls pour les détecter.

## Abstract

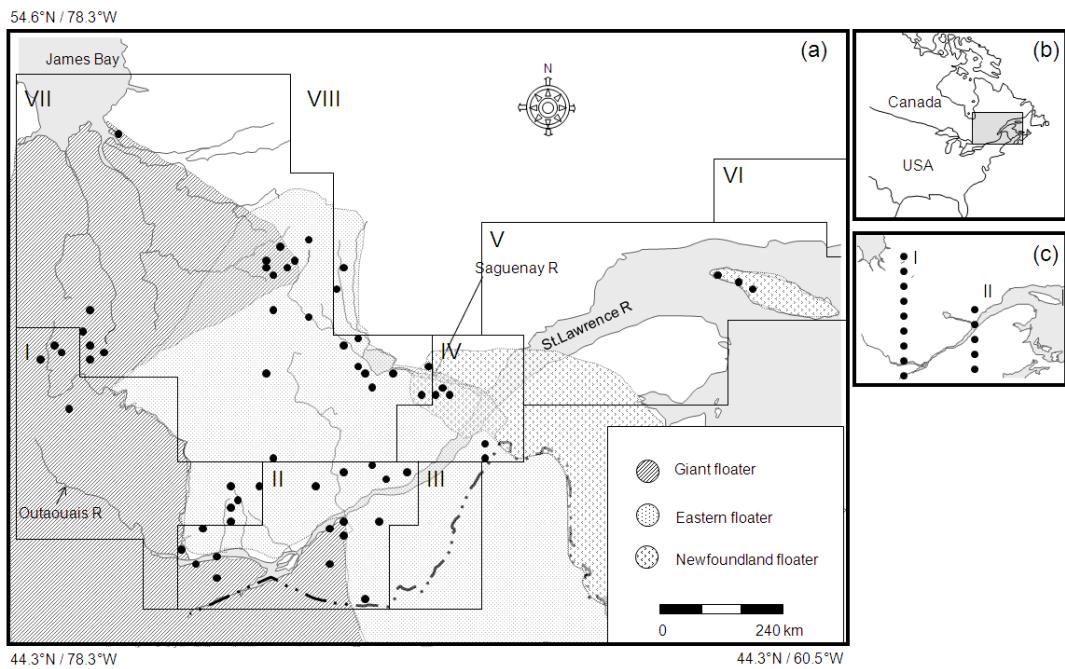
Dispersal of freshwater mussels occurs when the larvae parasitize fishes. In northeastern North-America, distributions of floater mussels (*Pyganodon spp.*) are allopatric and arranged in an east-west pattern that corresponds poorly with the biogeographic patterns of the regional fish fauna. This study aims at determining whether associations with distinct fish species can explain the distribution of floaters. Here, we devised a framework based on a series of novel distributional null models to elucidate dispersal of species that are dependent on host species. The distributional patterns of floaters were tested by controlling for associations with fishes, while host-parasite associations were assessed with null models based on co-occurrence patterns. The disjoint distribution of the Giant (*P. grandis*) and the Eastern floaters (*P. cataracta*) could not be explained by their association with distinct host fish and is likely the result of associations with distinct host populations rather than distinct species. In contrast, associations of the Newfoundland floater (*P. fragilis*) with euryhaline fish species could explain its distribution. Such associations may have also promoted differentiation with its sister taxa, the Eastern floater. This study demonstrated the effects of positive biological associations on distributional patterns and the utility of a null model framework to uncover them.

## 2.2 Introduction

Dispersal is an important factor determining the distribution of species (e.g. Belyea and Lancaster, 1999). However, a large number of organisms rely on other species for their dispersal (Vaughn and Taylor, 2000 and references therein). The distribution of those hitchhiker species is linked to the distribution of their hosts, while host specialization or host shift appears as a major evolutionary mechanism, promoting allopatry, isolation and genetic divergence among hitchhiker populations (e.g. Funk et al., 1995; Zuber and Widmer, 2000; Thorogood et al., 2008).

Larvae of freshwater mussels (*Unionidae*) must parasitize a fish host to complete their life cycle. Mussel dispersal occurs mainly during the parasitic stage, whereas adult mussels are benthic and display limited dispersal capability. As a result, the distribution of mussels is expected to be strongly linked to the distribution of their fish hosts (Watters, 1994; Haag and Warren, 1998; Vaughn and Taylor, 2000). Unionid mussels range from generalists that use a taxonomically diverse range of fishes to strict specialists that use only one or a few closely related fish species (e.g. Fuller, 1974).

Floater species (*Pyanodon spp.*) are generalists, lacking any mechanism to attract or to recognize a specific host and are known to infect more than 30 different species of fish distributed across 11 different families (Fuller, 1974; Hoggarth, 1992; Watters et al., 2005). The number of host species is potentially greater because we still lack information about all potential hosts across the entire range of the parasite species. The distribution of floater species poorly fits the main biogeographic pattern observed for fish communities in north eastern North America, which is characterized by a loss of species, eastward and northward, while the coastal zones are also characterized by the presence of euryhaline fish species (Legendre and Legendre 1984). In contrast, floaters display wide but nearly allopatric distributions arranged in an east-west pattern (Fig. 2.1; see also Clarke, 1981; Cyr et al., 2007). Indeed, breaks in distribution among floater species are inconsistent with the biogeographic zones described by Legendre and Legendre (1984).



**Figure 2.1:** Distributions of the Giant, the Eastern and the Newfoundland floater, across Quebec (modified from Clarke, 1981; and Cyr et al., 2007). Dots represent sampled sites (Cyr et al., 2007). A) Distribution of floater species compared to biogeographic zones (I-VIII) based on fish distributions defined by Legendre and Legendre (1984). B) Localization of the studied region in north eastern North America. C) Breaks between genetic lineages of lake trout (*Salvelinus namaycush*) (I) and lake whitefish (*Coregonus clupeaformis*) (II), according to Wilson and Hebert (1998) and Bernatchez and Dodson (1991), respectively.

Allopatry among floater species could be a consequence of distinct patterns of historical dispersal during and after glaciation. The Pleistocene period was marked by several glaciation episodes (Pielou, 1991) during which northern species found refuge in distinct regions at the southern margin of the glacier, promoting allopatry and divergence among species and also among populations (e.g., Bernatchez and Wilson, 1998). The current distribution of floater species suggests that they colonized north eastern North America from different refuges (Kat, 1985; 1986; Cyr et al., 2007). Thus, different Pleistocene distributions and opportunities for recolonization may have induced the present day allopatry. However, whether or not allopatry among floaters was driven by the use of different host species remains unexplored. In this study, we set out to test whether allopatry among floater species

can be explained by associations with distinct host species. On the one hand, associations with distinct hosts may suggest that either host shift or host specialization promoted isolation and differentiation among floater species. If so, the use of distinct hosts may have shaped the current distribution of mussels, leading to allopatry. Conversely, historical allopatry may have pushed mussel species to use different hosts. On the other hand, if mussel species use the same host species despite their current allopatry, their distributions are likely to be mainly (or solely) due to isolation during the Pleistocene period.

To test the hypothesis of distinct associations with hosts, we developed a null model framework in which allopatry among mussel species was assessed by contrasting the results of constrained and unconstrained permutations tests. The constrained tests controlled alternately for fish (host) species composition and spatial autocorrelation among sampling sites. If allopatry among floater species could be explained by association with distinct fish species, we may expect that allopatry - otherwise significant - should become non significant under the constrained null model (i.e., controlling for fish species composition). Potential hosts specific to each floater species were identified using a null model based on co-occurrence patterns between mussel and fish species. If allopatry among floater species is a function of their associations with distinct fish species, floater species should be associated with distinct sets of species. Accordingly, given their current distributions (Fig. 2.1), we expected that the Giant (*Pyganodon grandis*) and the Eastern floaters (*Pyganodon cataracta*) should be associated with stenohaline freshwater fish species that recolonized the territory from Mississippi and Atlantic refuges, respectively, while the Newfoundland floater (*Pyganodon fragilis*) should be associated with euryhaline fish species also coming from the Atlantic refuge.

## 2.3 Materials and Methods

### *Inferring distribution of floater and fish species*

A total of 471 individual floater mussels were hand-collected across 66 sites (lakes and rivers) across southern Quebec (Canada; Fig. 2.1); species identities were confirmed by genetic markers (Cyr et al., 2007). Sampling sites were selected to cover a large territory, and whenever possible, sites from distinct drainages were selected. Details about the sites and identification method can be found in Cyr et al. (2007). This survey confirmed the distribution of the Giant floater, the Eastern floater and the Newfoundland floater across the region, described by Clarke (1981) based on shell morphology. This survey also detected the presence of cryptic hybrid floater lineages. However, in the present study, those cryptic lineages were not considered since their presence on the territory is marginal and their evolutionarily relationships with other floater species remain unclear (Cyr et al., 2007, but see Stanton, 2009)

Fish species composition at sampling sites was determined from distribution maps found in Scott and Crossman (1973) and Bernatchez and Giroux (2000). Although distribution maps will not necessarily confirm a species' presence at a particular site, the use of such maps is still appropriate because the fish species, as a potential host, could have been or could be present in nearby sites. Moreover, because our framework takes into account host distributions (constrained null model), by using distribution maps, in which the entire host distribution is considered, the analyses of positive and negative patterns of mussel co-occurrence should be in fact more conservative (i.e., more difficult to reject in contrast to random mussel associations) whenever their host distribution overlaps.

*Tests for allopatry among parasite species*

Allopatry among parasite species was estimated using the C-score index (Stone and Roberts, 1990):

$$(1) \quad C_{ij} = [(S_i - Q_{ij})(S_j - Q_{ij})] / (S_i S_j)$$

where  $S_i$  and  $S_j$  are the number of sites where the  $i^{\text{th}}$  and  $j^{\text{th}}$  species are present respectively, and  $Q_{ij}$  is the number of sites where both species occur. Note that the denominator ( $S_i S_j$ ) assures that the index varies between 0 (when species are in perfect sympatry) and 1 (perfect allopatry). The C-score statistic calculates the number of checkerboard units (i.e., the number of sites for which species  $i$  is present and species  $j$  is absent and vice-versa) such that large C-scores indicate an increasing degree of mutual exclusivity between species. A global score was obtained by averaging the C-scores for all species-pair combinations.

To test the significance of the C-score statistic, the incidence (presence/absence) of parasite species across sites was permuted. Species were randomly allocated across sites 9999 times, and for every permuted set the C-score was calculated and compared to the observed value. The number of sites occupied by a particular parasite species was kept constant across permutation sets (i.e., column sums), but we allowed the number of parasite species to vary across sites (i.e., row sums). These tests by permutations, as well as all permutation tests performed in this study were performed using the R software (R development core team 2009). Significant statistic could be interpreted as the evidence that some non random process must have influenced the dispersal of the parasite species. Otherwise, a non significant pattern could be interpreted as the evidence that similar distributional pattern could have emerged from random dispersal of the parasite species.

*Constrained null model analysis*

To determine whether allopatry observed among parasite species could be explained by associations with distinct host species, the C-score values were tested using a constrained null model analysis (Peres-Neto et al., 2001) based on their hosts' distributions. In this null model, the probability that a parasite species is present or absent from a particular site was not equiprobable and was determined by

the distribution of host species. If allopatry among parasite species could be explained by associations with distinct host species, the C-score values obtained from this null model should be comparable with C-score values that we observed among floater species. First, we applied a correspondence analysis (CA) on the host distributional matrix (66 sites by 100 species). Next, the distribution of each parasite species was regressed against the axes from the correspondence analyses using logistic regression and a stepwise selection procedure. Because our main goal was to control for host distribution and not interpret how mussels are related to particular species, we used the correspondence axes to assure orthogonality and reduce issues related to collinearity and model selection (see Peres-Neto and Legendre 2010 for a discussion). Finally, the predictive values from the logistic regression model were then used as probabilities in the constrained permutation test (Peres-Neto et al., 2001). Unlike the first null model where species were randomly permuted across sites, the constrained version used the model-predicted site probabilities of occurrence of parasite species as a function of their hosts distributions as the likelihood of their occupancy during the generation of permuted sets. If allopatry of a particular parasite species is explained by fish species, C-score values should become non-significant under the constrained permutations procedure in contrast to the unconstrained permutations described above.

To assess whether spatial autocorrelation after accounting for host distribution (logistic regression above) could affect patterns of negative co-occurrence of parasite distribution, we calculated Moran's I spatial autocorrelation coefficient and associated probabilities using permutation tests (as in Griffith and Peres-Neto, 2006) for each of the retained (i.e., after selection) logistic regression models. Spatial autocorrelation in null model analyses of species co-occurrence are known to inflate the type I error rate of such tests (i.e., the number of false rejections; Hardy, 2008) and needs to be taken into account while assessing test significance. Because Moran's I was not significant for any of the model residuals (see result section), we did not correct for spatial autocorrelation. If autocorrelation has been significant, we could have either considered spatial predictors into the model (see Peres-Neto and Legendre (2010) for a discussion) or controlled for the spatial structure using spatially-constrained permutations as in the next set of null models below.

*Test for host-parasite associations – null model analysis*

Associations with host species were assessed by null model analyses based on a suite of similarity indices measuring different aspects of these associations (see below). These null models identify significant positive co-occurrence patterns between host and parasite species and allow identification of putative distinct host species for each parasite lineage.

To compare parasite species on the basis of their associations with host species, three statistics were used to measure co-occurrence:

$$(2) \quad S_j = a / (a + h + p)$$

The statistic  $S_j$  is the Jaccard similarity coefficient (Jaccard, 1912) between a host and a parasite species and is measured as the number of sites where both species are present ( $a$ ) divided by the number of sites where at least one species is present;  $h$  represents the number of sites where only the host species is present, whereas  $p$  is the number of sites where only the parasite species is present. In addition to the Jaccard similarity coefficient, the following two indices were used:

$$(3) \quad S_h = a / (a + h)$$

$$(4) \quad S_p = a / (a + p)$$

$S_h$  measures the co-occurrence of the host and parasite species within the host distribution, ignoring the sites where only the parasite species is present ( $h$ ), whereas  $S_p$  measures the co-occurrence of the host and parasite species within the parasite distribution, hence not taking into account the sites where only the host species is present ( $p$ ). Combined, the three statistics can help to distinguish among different scenarios of parasitism. A parasite species can be associated with distinct species in distinct regions and thus have a wider distribution than its hosts, thus producing large  $S_h$  values. Conversely a parasite species could be restricted in some portions of the

range of its hosts, thus producing high  $S_p$  values. Finally, if a parasite species has a distribution similar to its hosts, a high similarity value should be expected for all three statistics. Moreover, the number of significant associations detected could be used to assess the infectious capacity of a parasite species. A generalist parasite species is expected to be associated with multiple hosts; whereas a specialist parasite species should be associated with a low number of species.

The significance of each of the above similarity indices was assessed by null model analyses in which parasite species were randomly assigned to the sites across 9999 permutations. The number of sites occupied by a particular parasite species was kept constant (i.e., column sums), but we allowed the number of parasite species to vary across sites (i.e., row sums). For each permuted data set, the similarity values were recalculated and compared to the observed values. Probability values were adjusted using Bonferroni corrections for multiple testing. To control for spatial autocorrelation while assessing null model significance, we used a permutation procedure in which the probability of permutation across sites was based on their geographic distance (see Lapointe and Garland 2001). If similarity patterns between host and parasites had been solely due to spatial distribution, then the test would not be significant.

## 2.4 Results

Allopatry among floater species is evident (Fig. 2.1) and indeed the C-score values were significant for every species pair and, consequently, for the global test based on the average C-scores for all species-pairs (Table 2.1). Because Newfoundland and the Giant floaters do not share any site in common, they presented the highest C-score value.

**Table 2.1:** Relative C-score index (observed C-score divided by the theoretical maximal value for a given set, i.e., when species are in perfect allopatry) calculated for all species combined (global) and for every species pair, as well as the probabilities associated with their constrained and unconstrained null models. G: Giant floater; E: Eastern floater; N: Newfoundland floater. Moran's I for model residuals for each host species were not significant.

	Global	G-E	G-N	E-N
Relative C-score	0.78	0.68	1.00	0.74
Unconstrained null model	p<0.001	p<0.001	p=0.009	p=0.023
Constrained null model	p<0.001	p<0.001	p=0.444	p=0.100
	G	E	N	
Moran's I for model residuals -	-	-0.012	0.031	-0.023
Moran's significance		p=0.789	p=0.765	p=0.823

Under the constrained model, the C-score value between the Giant and the Eastern floaters was significant (Table 2.1), indicating that no specific association was responsible for the allopatry between these species. On the other hand, the C-score values between the Newfoundland and the Eastern floaters and between the Newfoundland and the Giant floaters were not significant under the constrained null model (Table 2.1). There was no significant spatial autocorrelation in model residuals for any of the logistic regression produced for the constrained null model, indicating that test results were not the product of increased false rejections due to autocorrelation in parasite distributions (Table 2.1).

The Giant and the Eastern floaters were often associated with the same fish species, despite their allopatric distribution. Fish species presenting a high  $S_j$  or  $S_p$  value with the Giant floater often presented a high similarity value with the Eastern floater or vice-versa (Table A1). When compared on the basis of  $S_j$  or  $S_p$  values, the Giant and the Eastern floater were highly correlated (Pearson's r of 0.51 and 0.85, respectively; Table 2.2). The opposite trend was observed on the basis of the  $S_h$  index, for which the Giant and the Eastern floater did not have high  $S_h$  values with the same host species. Based on this index, the Giant and the Eastern floater were inversely correlated (Pearson's r of -0.73).

**Table 2.2:** Correlations among floater species based on their similarity values with fish species. Correlations based on  $S_j$ ,  $S_p$  and  $S_h$  values are shown, respectively, in the first correlation matrix above the diagonal; in the first correlation matrix below the diagonal; and in the second correlation matrix above the diagonal. Significant relationships are shown in bold.

	Giant f.	Eastern f.	Newfoundland f.
Giant f.	1	<b>0.51</b>	<b>-0.35</b>
Eastern f.	<b>0.85</b>	1	0.06
Newfoundland f.	<b>0.32</b>	<b>0.59</b>	1
	Giant f.	Eastern f.	Newfoundland f.
Giant f.	1	<b>-0.73</b>	<b>-0.57</b>
Eastern f.		1	-0.06
Newfoundland f.			1

The Newfoundland floater appeared associated with distinct fish species. Fish species presenting high  $S_j$  or  $S_h$  values with the Newfoundland floater did not present high similarity values with either the Giant or the Eastern Floater (Table A1). Based on these two indexes, the Newfoundland floater appeared inversely correlated with the Giant Floater and uncorrelated with the Eastern floater (Table 2.2). The Newfoundland floater also seemed to diverge from the other two floater species using the  $S_p$  index as a basis, though to a lesser degree. Based on this index, the Newfoundland floater was correlated with both the Giant and the Eastern floaters. However, correlations were weaker than the correlation observed between the Giant and the Eastern floater.

Differences between the Giant and the Eastern floater emerge when only fish species having distributions significantly similar to floater distributions are compared. All the associations detected as significant on the basis of the three similarity indexes ( $S_j$ ,  $S_p$  and  $S_h$ ) were also significant on the basis of the constrained spatial permutation test, thus indicating that these tests were not affected by spatial autocorrelation in host and parasite distributions. Permutation tests revealed that the Giant, the Eastern and the Newfoundland floaters are associated with 5, 1 and 7 fish species, respectively, and do not have any potential host in common (Table 2.3). The Newfoundland floater was associated with Atlantic euryhaline fish species, whereas the two other floater species displayed more continental distributions and appeared associated with strictly freshwater fish species. The Giant floater was mostly associated with fish species with a presumed Mississippi glacial refuge, while the only fish species significantly associated with the Eastern floater (Round Whitefish) is presumed to originate from both Mississippi and Atlantic refuges (Mandrak and Crossman, 1992). Even though these particular fish species were significantly associated with a single floater species, they presented distributions that partially overlapped the distributions of at least two floater species. The only exception is the goldeye (*Hiodon alosoides*), which has a distribution completely nested within the distribution of the Giant floater.

All fish species that were significantly associated with one or another floater species presented a high and significant similarity value for the three indexes used in this study (Table 2.3). Interestingly, fish species associated with either the Giant or the Eastern floater did not present maximum  $S_p$  values as 17 fish species out of the 100 tested presented higher  $S_p$  value with both the Eastern and the Giant floater (Table A1). Fish species significantly associated with the Giant floater presented  $S_p$  values ranging between 0.39 and 0.87, whereas the round whitefish, the only fish associated with the Eastern floater, presented an  $S_p$  value of 0.74. However, seventeen species had both an  $S_p$  value with the Giant floater higher than 0.87 and an  $S_p$  value with the Eastern floater higher than 0.74. Each one of those fish species was characterized by a high occurrence frequency and were found in at least 52 sites (which represents 78% of the sites). Most of those fish species are also characterized by a presumed glacial origin from both the Mississippi and the Atlantic refuges (Mandrak and Crossman, 1992)

Finally, one of the fish species significantly associated with the Giant floater was also a confirmed host for that species (Fuller, 1974). Among previously reported hosts of floater species present in the region (Table 2.4), two presented a distribution completely nested within the distribution of the Giant floater ( $S_h = 1$ ), 7 presented distributions that overlapped nearly completely with the distributions of both giant and eastern floater ( $S_p$  value  $> 0.95$ ) and 18 presented distributions that partially overlapped the distributions of those two mussel species.

**Table 2.3:** Significant associations between fishes and floater mussels according to the three similarity indices used in this study ( $S_j$ ;  $S_h$ ;  $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge and ?: unknown post-glacial origin. The fish species in bold was previously reported as hosts for floater mussels (Fuller, 1974; Trdan and Hoeh, 1982; Hoggarth, 1992; Watters et al., 2005) and † indicates euryhaline species. Significant associations after Bonferroni corrections are indicated by the star symbol.

Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	$S_j$	$S_h$	$S_p$	$S_j$	$S_h$	$S_p$	$S_j$	$S_h$	$S_p$
Lake Sturgeon	36	?	Acipenseridae	0.51*	0.56*	0.87*	0.30	0.47	0.45	0.00	0.00	0.00
Mooneye	20	M	Hiodontidae	0.48*	0.70*	0.61*	0.12	0.30	0.16	0.00	0.00	0.00
Goldeye	9	M	Hiodontidae	0.39*	1.00*	0.39*	0.00	0.00	0.00	0.00	0.00	0.00
<b>Longnose dace</b>	<b>35</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.49*</b>	<b>0.54*</b>	<b>0.83*</b>	<b>0.26</b>	<b>0.43</b>	<b>0.39</b>	<b>0.02</b>	<b>0.03</b>	<b>0.11</b>
Sauger	20	M	Percidae	0.59*	0.80*	0.70*	0.07	0.20	0.11	0.00	0.00	0.00
Round whitefish	35	AM	Salmonidae	0.18	0.26	0.39	0.62*	0.80*	0.74*	0.05	0.06	0.22
Atlantic sturgeon †	10	M	Acipenseridae	0.00	0.00	0.00	0.09	0.40	0.11	0.73*	0.80*	0.89*
Alewife †	12	A	clupeidae	0.06	0.17	0.09	0.09	0.33	0.11	0.62*	0.67*	0.89*
American Eel †	26	A	Anguillidae	0.14	0.23	0.26	0.23	0.46	0.32	0.35*	0.35*	1.00*
American Shad †	12	A	clupeidae	0.06	0.17	0.09	0.09	0.33	0.11	0.62*	0.67*	0.89*
Tomcod †	16	?	percopsidae	0.00	0.00	0.00	0.23	0.63	0.26	0.47*	0.50*	0.89*
Atlantic salmon †	21	A	salmonidae	0.02	0.05	0.04	0.31	0.67	0.37	0.43*	0.43*	1.00*
Fourspine stickleback †	4	?	Gasterosteidae	0.00	0.00	0.00	0.02	0.25	0.03	0.44*	1.00*	0.44*

**Table 2.4:** Similarity values between floater species and previously reported host species (Fuller, 1974; Hoggarth, 1992; Trdan and Hoeh, 1982; Watters et al., 2005), according to the three similarity indices used in this study ( $S_j$ ;  $S_h$ ;  $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge, I: Introduced and ?: unknown post-glacial origin. † indicates euryhaline species. Significant associations after Bonferroni corrections are indicated by the star symbol.

Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	$S_j$	$S_h$	$S_p$	$S_j$	$S_h$	$S_p$	$S_j$	$S_h$	$S_p$
Blacknose dace	63	AM	Cyprinidae	0.37	0.37	1.00	0.60	0.60	1.00	0.09	0.10	0.67
Longnose sucker	63	AM	Catostomidae	0.37	0.37	1.00	0.60	0.60	1.00	0.09	0.10	0.67
White Sucker	62	AM	Catostomidae	0.37	0.37	1.00	0.61	0.61	1.00	0.08	0.08	0.56
Brook Stickleback	61	M	Gasterosteidae	0.38	0.38	1.00	0.57	0.59	0.95	0.09	0.10	0.67
Yellow Perch	61	AM	Percidae	0.38	0.38	1.00	0.57	0.59	0.95	0.09	0.10	0.67
Pearl Dace	62	AM	Cyprinidae	0.35	0.35	0.96	0.61	0.61	1.00	0.09	0.10	0.67
Creek chub	60	AM	Cyprinidae	0.36	0.37	0.96	0.58	0.60	0.95	0.10	0.10	0.67
Longnose dace	35	AM	Cyprinidae	0.49*	0.54*	0.83*	0.26	0.43	0.39	0.02	0.03	0.11
Blacknose Shiner	31	M	Cyprinidae	0.38	0.48	0.65	0.21	0.39	0.32	0.03	0.03	0.11
Iowa darter	28	M	Percidae	0.34	0.46	0.57	0.22	0.43	0.32	0.00	0.00	0.00
Golden Shiner	36	AM	Cyprinidae	0.20	0.28	0.43	0.37	0.56	0.53	0.10	0.11	0.44
Common Shiner	32	AM	Cyprinidae	0.22	0.31	0.43	0.37	0.59	0.50	0.00	0.00	0.00
Rock Bass	28	M	Centrarchidae	0.24	0.36	0.43	0.29	0.54	0.39	0.00	0.00	0.00

Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Largemouth Bass	27	M	Centrarchidae	0.25	0.37	0.43	0.27	0.52	0.37	0.00	0.00	0.00
Johny darter	23	AM	Percidae	0.28	0.43	0.43	0.20	0.43	0.26	0.00	0.00	0.00
Freshwater Drum	14	M	Scianidae	0.28	0.57	0.35	0.13	0.43	0.16	0.00	0.00	0.00
Pumpkinseed	24	AM	Centrarchidae	0.18	0.29	0.30	0.29	0.58	0.37	0.00	0.00	0.00
Threespine stickleback †	41	A	Gasterosteidae	0.10	0.15	0.26	0.52	0.66	0.71	0.22	0.22	1.00
Banded Killifish †	26	AM	Cyprinodontidae	0.14	0.23	0.26	0.28	0.54	0.37	0.13	0.15	0.44
Bluntnose Minnow	15	AM	Cyprinidae	0.19	0.40	0.26	0.15	0.47	0.18	0.00	0.00	0.00
Carp	12	I	Cyprinidae	0.21	0.50	0.26	0.11	0.42	0.13	0.00	0.00	0.00
Longnose gar	8	AM	Lepisosteidae	0.19	0.63	0.22	0.07	0.38	0.08	0.00	0.00	0.00
Blackchin Shiner	5	M	Cyprinidae	0.17	0.80	0.17	0.02	0.20	0.03	0.00	0.00	0.00
Bluegill	5	M	Centrarchidae	0.12	0.60	0.13	0.05	0.40	0.05	0.00	0.00	0.00
Gizzard Shad	4	M	Clupeidae	0.08	0.50	0.09	0.05	0.50	0.05	0.00	0.00	0.00
White bass	4	M	Percichthyidae	0.08	0.50	0.09	0.05	0.50	0.05	0.00	0.00	0.00
Brook Silverside	2	M	Atherinidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Longear sunfish	2	M	Centrarchidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Yellow Bullhead	2	?	Ictaluridae	0.04	0.50	0.04	0.03	0.50	0.03	0.00	0.00	0.00

## 2.5 Discussion

The framework proposed here allows testing whether allopatry among floater species can be explained by their associations with distinct fish set of species. Based on their current distributions, it appears that the Giant floater invaded the Quebec region, after the last glaciation from the Mississippi refuge, while the Eastern floater likely came from Atlantic slopes (Kat, 1985; 1986; Cyr et al., 2007). As expected, both floater species were associated with distinct fish species. However, those associations appeared insufficient to explain the distribution pattern of the two floater species as: i) their allopatry remained significant even when permutations were controlled for fish species composition at each site; ii) despite few specific associations, the Giant and the Eastern floater often co-occurred with the same fish species; and iii) several fish species, including many previously reported hosts of floater species, presented distributions that overlapped nearly completely with the distribution of both the Giant and the Eastern floater.

One important point to consider is that the power of the permutation procedures to detect significant associations can be reduced when the frequency of occurrence of the species being tested is too high. A host species present at most sites will probably have a high  $S_p$  value with parasite species; however, the  $S_p$  value would also be high for every permuted data set because occurrence of the species is held constant. As a consequence, although such high similarity values would not be significant, they should not necessarily be interpreted as an absence of association between a parasite and a host species. Such a high and non-significant similarity value should rather be interpreted as a limitation of the test procedure. Therefore, fish species with high, but non-significant  $S_p$  values should not be necessarily discarded as potential hosts of floater species. Consequently, the Giant and the Eastern floaters could have colonized the territory using distinct populations of the same host species.

As expected, our analyses revealed that the Newfoundland floater was associated with euryhaline fish species. Associations of the Newfoundland floater with euryhaline fish species may have shaped its distribution as: i) the allopatry of the Newfoundland floater has become non-significant under the constrained null model in which the Newfoundland floater was restricted to the sites where euryhaline

fish species were present; ii) the Newfoundland floater presented divergent  $S_j$  values with fish species when compared to other floater species; iii) fish species with distributions that completely overlapped the distributions of both the Giant and the Eastern floater, only partially overlapped the distribution of the Newfoundland floater.

The Newfoundland floater is weakly genetically divergent from the Eastern floater (Clarke and Rick, 1963; Kat, 1983; Cyr et al., 2007) and both lineages likely invaded the region after the last glaciation from the Atlantic slope (Kat and Davis, 1984; Cyr et al., 2007). Associations between the Newfoundland floater and the euryhaline community of hosts indicate a shift in selective regime for that species, as different host species should have distinct immune responses. Associations between the Newfoundland floater and euryhaline fish species could be a consequence of the spatial heterogeneity of host communities in the Atlantic refuge. Previous studies have inferred that both euryhaline and stenohaline fish species came from the Atlantic refuge (e.g. Mandrak and Crossman, 1992). Because the habitats of euryhaline and stenohaline species partially overlap, the switch of Newfoundland floater toward euryhaline fish appears possible. The spatial heterogeneity of host communities may have been maintained during and after post-glacial dispersal. Such an association with a distinct fish community may have promoted allopatric isolation and prevented hybridization between the Newfoundland and the Eastern floaters.

Interestingly, some consistencies were observed between the distribution of floater species and the distribution of genetic lineages observed in some fish species. However, the comparisons that could be made are limited as the distribution of genetic lineages has been studied only for a few fish species (mostly salmonid species); moreover, sampling effort and the position of sampling sites vary greatly across these studies. Longitudinal breaks coinciding with those observed between the Giant and the Eastern floaters and between the Newfoundland and the Eastern floaters have been documented for Lake Trout (*Salvelinus namaycush*; Wilson and Hebert, 1998) and for Lake Withefish (*Coregonus clupeaformis*; Bernatchez and Dodson, 1991), respectively. Moreover, an important contact zone between the Giant and the Eastern floaters is found in northern sites which have been previously documented as the region where Brook Trout populations (*Salvelinus fontinalis*)

from the Atlantic and the Mississippi glacial refuges, co-occur (Fraser and Bernatchez, 2005). This contact zone is consistent with the presence of interconnected river networks in the region (Legendre and Legendre, 1984).

The consistency between longitudinal breaks observed between parasite species and genetic breaks previously reported in salmonid fish species may suggest that those breaks could be widespread within the regional fish fauna. Indeed the presence of distinct parasites fauna could be taken as an indication of isolation among fish populations. Moreover, although it is possible that current associations do not reflect historical ones, it is likely that fish populations harbouring distinct parasites species have been isolated since the last glacial period. Obviously, more phylogeographic studies are needed to definitively conclude on that topic,

The quantitative framework used in this study allows an assessment of the impact of positive biological association on distributional patterns. This framework can be used to assess dispersal of other hitchhiker species as well as other species that could be limited in its dispersion by positive biological associations. Such assessments could be in the context of post-glacial dispersal or in more recent contexts, such as the invasion of pest species with several possible hosts in the context of global warming.

## **2.6 Acknowledgements**

We are grateful to Andre Martel from the Canadian Museum of Nature (CMN) for his support and Philippe Girard and Bailey Jacobson for their valuable editorial advice. FC was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and “Fonds Quebecois de Recherche en Nature et Technologie” (FQRNT, Quebec). This work was supported by a “Fonds Quebecois de Recherche en Nature et Technologie” (FQRNT, Quebec) new researchers grant to BA.

## **2.7 Author contribution**

Original idea: FC and BA; Permutation tests: FC; Autocorrelation analyses: PPN; Redaction: FC, PPN and BA

**CHAPITRE 3 : ARTICLE 2**

**HISTORICAL PROCESS LEAD TO FALSE GENETIC SIGNAL  
OF CURRENT CONNECTIVITY AMONG POPULATIONS**

Frédéric Cyr, Bernard Angers

Publié dans la revue Genetica (2012) **139**: 1417-1428

### 3.1 Résumé

La détermination des effets des processus historiques sur la variation génétique spatiale est d'une importance majeure en matière de génétique du paysage, en particulier dans les systèmes récents où le signal d'isolement est souvent à peine perceptible. Le but de cette étude était d'évaluer comment les différences dans les patrons de colonisation spatiale pourraient influencer la variation génétique en utilisant deux espèces de Centrarchidae, le crapet soleil (*Lepomis gibbosus*) et le crapet de roche (*Ambloplites rupestris*), de deux drainages adjacents comme modèle biologique. La différence frappante entre les variations génétiques spatiales des deux espèces suggère un patron de colonisation complètement opposé. Le crapet de roche aurait colonisé les drainages en provenance d'une source en aval, entraînant une perte de diversité génétique en amont et une différenciation forte entre les drainages. En revanche, la réduction de la diversité génétique et l'augmentation de la différenciation vers l'aval tel qu'observé entre les populations de crapets soleils suggèrent une colonisation de l'amont vers l'aval. Ce patron de colonisation résulte en une différenciation faible entre la plupart des sites en amont des deux drainages, conduisant à un faux signal de flux génétique entre les drainages. La présente étude démontre empiriquement que les patrons de colonisation peuvent interferer avec notre capacité à estimer la connectivité actuelle.

## Abstract

Identification of the effects of historical processes on spatial genetic variation is of major importance in landscape genetics, especially in recent systems where the signal of recent isolation is difficult to perceive. The goal of this study was to assess how differences in colonization patterns could influence spatial genetic variation using two centrarchidae species, the pumpkinseed sunfish (*Lepomis gibbosus*) and the rock bass (*Ambloplites rupestris*), from two adjacent drainage systems. The striking difference between the spatial genetic variations of the two species suggests completely opposite patterns of colonization. Rock bass colonized the drainage system from a downstream source, which resulted in a loss of diversity in upstream populations and a strong differentiation between drainage systems. In contrast, the reduction of genetic diversity and increase of differentiation toward downstream populations that were observed among sunfish populations suggest colonization from upstream to downstream. The colonization pattern observed in sunfish, which result in low differentiation between upstream most sites of the two drainages, leads to a false genetic signal of current inter-drainage gene flow. The present study demonstrates through empirical evidence that colonization patterns may impede the capacity to estimate current connectivity.

### 3.2 Introduction

The expanding field of landscape genetics focuses on the interrelated effects of landscape elements and evolutionary processes on spatial genetic variation (Angers *et al.* 1999; Manel *et al.* 2003). Recent developments in this field have been motivated by the challenge of maintaining the evolutionary potential of species within an increasingly human-impacted world (Sork and Waits 2010). Previous studies have provided numerous examples of how landscape elements shape genetic patterns through neutral and non-neutral processes (Sork and Waits 2010; Storfer *et al.* 2010).

However, attempts to characterize the impact of current landscape elements can be complicated by the fact that evolutionary processes can affect genetic diversity at different rates (*e.g.* Lacy 1987; Whitlock and McCauley 1999). While demographic events that lead to a loss of genetic diversity rapidly affect populations, events that result in increased genetic diversity could extend over thousands of generations. Similarly, the effect of gene flow in homogenizing populations will occur rapidly, but genetic differentiation resulting from mutation and genetic drift could be extremely slow. Consequently, the influence of the landscape on genetic data in recent systems (*e.g.* recently deglaciated areas) is often biased toward rapid changes such as founder events (or bottleneck) or reduction of differentiation by gene flow.

Historical processes are of major importance as potential explanatory factors of current spatial genetic variation (*e.g.* Bernatchez and Wilson 1998; Taberlet *et al.* 1998; Angers *et al.* 1999; Gagnon and Angers 2005; Boizard *et al.* 2008). Phylogeographic studies have revealed how historical factors have played a major role in partitioning genetic diversity across large spatial scales. As an example, the joint influence of refuge origins and post-glacial colonization routes substantially affected the spatial genetic variation in formerly glaciated areas (*e.g.* Bernatchez and Wilson 1998; Taberlet *et al.* 1998; Turgeon and Bernatchez 2001; Fraser and Bernatchez 2005; Girard and Angers 2006). Moreover, populations located in recently glaciated areas show a reduction of genetic diversity with increases in latitude as a result of past founder events that occurred along colonization routes

(Bernatchez and Wilson 1998). Stepping-stone colonization, which results in numerous founder events, has also had an important influence on the spatial genetic variation at a regional scale (e.g. Fraser and Bernatchez 2005; Gagnon and Angers 2005; Boizard *et al.* 2008).

Fish species are often targeted in landscape genetics due to the expected correlation between differentiation and the hydrological network (e.g. Angers *et al.* 1999; Castric *et al.* 2001). However, a lack of congruence between the pattern of genetic variation and the current drainage system is often reported and explained by the changes in the hydrologic network (Angers and Bernatchez 1998; Gagnon and Angers 2005; Boizard *et al.* 2008; Fraser and Bernatchez 2005). In some instances, the current hydrologic network may incorrectly represent colonization routes taken by the various species; populations may then present either a strong and unexpected genetic similarity between drainages or strong differentiation within drainage (e.g. Gagnon and Angers 2005). Such results raise the question of the influence of colonization patterns on our capacity to estimate current processes and particularly current connectivity among populations.

This study is designed to address how differences in colonization patterns should influence spatial genetic variation. More precisely, the genetic variation of populations from two adjacent river networks was determined for two morphologically similar species of the family centrarchidae: the pumpkinseed sunfish (*Lepomis gibbosus*) and the rock bass (*Ambloplites rupestris*). While both species have comparable ecological niches and swimming capacity (Scott and Crossman 1974; French III 1998; Brinsmead and Fox 2002), pumpkinseed sunfish exhibit important morphological diversity related to trophic polymorphism (Robinson and Wilson 1996). To our knowledge, such polymorphism has never been reported in rock bass.

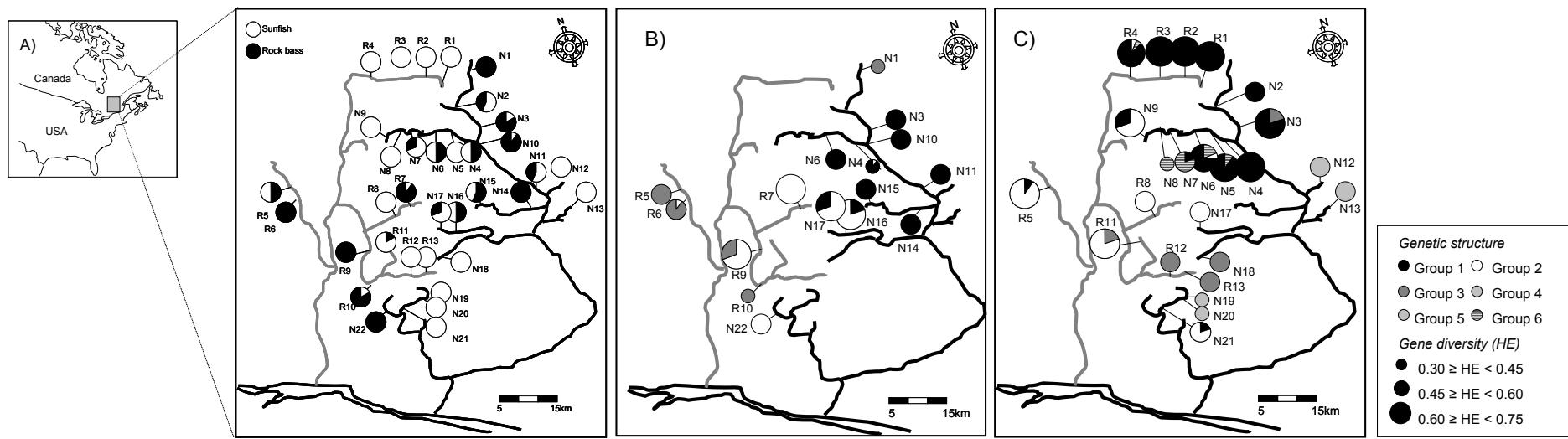
The colonization pattern of each species was inferred from the spatial distribution of genetic diversity. For a given species, a loss of genetic diversity and an increase of genetic differentiation between drainages are expected along the colonization route as the result of recurrent founding events (e.g. Ramachandran *et al.* 2005). We may expect that populations from different drainages with strong

genetic similarity exchange migrants through permanent or temporary connections between drainages (*e.g.* headwater lake with two outlets, intermittent connections due to seasonal freshets, or changing water levels). Alternatively, recently isolated populations could have retained genetic similarity due to close ancestry and slow changes in genetic diversity. In this respect, comparison among several species having similar dispersal capacities could be particularly informative; differences in genetic patterns among species could be interpreted as differences in colonization patterns or population sizes rather than differences in current dispersal among species.

### **3.3 Materials and methods**

#### *Sampling*

The study site is located on the southern margin of the Canadian Shield in the drainage system of the St. Lawrence River in Quebec, Canada. Forty sites were sampled during the summer 2008 for the pumpkinseed sunfish (*Lepomis gibbosus*) and the rock bass (*Ambloplites rupestris*). These sites belong to the drainages of the Rouge River (R) and Du Nord River (N) and both drainages flow into the Ottawa River (Fig. 3.1). The sampled region covers a territory located from 45.77°N to 46.19°N and from -74.12°W to -74.77°W and covers a total of approximately 2350 km<sup>2</sup>. Between six and twenty-one individuals were sampled for each population, for a total of 623 individuals (Table 3.1). Sites with less than 5 individuals of a given species were not analyzed.



**Figure 3.1:** Species and genetic distribution among sampling sites. The distribution of sunfish and rock bass in R and N drainages inferred from fish captures at each sampling site (A). The distribution of the different genetic groups assessed with STRUCTURE program for rock bass (B) and for sunfish (C) populations. The size of each sampling site refers to genetic diversity ( $H_E$ ). The gray and the black lines refer to the "la rivière Rouge" (R) and "la rivière du Nord" (N), respectively. Both rivers flow into the Ottawa River.

**Table 3.1:** Characteristics of the populations analyzed. The following information is provided for each sample site of both species: the sites of capture (letter refers to drainage system); the geographic coordinates and the altitude of the sites; the sample size (n); the total number of alleles (A); the total corrected number of alleles (Ac); observed heterozygosity ( $H_O$ ); Nei's gene diversity ( $H_E$ );  $F_{IS}$  values; and Hardy-Weinberg probability (HW). Values in bold type refer to significant probability values before Bonferroni corrections. No probability values were significant after Bonferroni corrections.

Sites	Latitude	Longitude	altitude (m)	n	A	AC	$H_O$	$H_E$	$F_{IS}$	HW
Rock bass										
N1	46.18	-74.28	460	17	15	14.3	0.37	0.40	0.08	0.076
N3	46.12	-74.28	380	17	20	17.4	0.50	0.50	0.02	0.533
N4	46.10	-74.36	380	16	22	16.7	0.42	0.44	0.06	0.376
N6	46.09	-74.39	400	18	23	18.5	0.51	0.51	0.01	0.145
N10	46.08	-74.28	380	17	20	16.8	0.46	0.50	0.07	0.260
N11	45.98	-74.12	220	15	21	17.7	0.49	0.47	-0.04	0.618
N15	46.00	-74.39	380	16	18	15.4	0.51	0.47	-0.07	0.874
N14	45.96	-74.17	320	16	17	15.8	0.57	0.53	-0.09	0.811
N16	45.93	-74.40	360	6	21	21.0	0.67	0.65	-0.03	0.775
N17	45.93	-74.40	380	9	23	21.3	0.61	0.67	0.09	0.178
N22	45.77	-74.51	320	16	20	18.2	0.55	0.56	0.02	0.212
R5	46.00	-74.77	220	16	26	21.5	0.60	0.60	0.00	0.421
R6	45.92	-74.57	200	15	36	25.9	0.67	0.68	0.01	0.630
R7	45.97	-74.52	240	16	30	23.5	0.70	0.67	-0.06	0.647
R9	45.91	-74.77	200	16	28	22.3	0.58	0.57	-0.02	0.404
R10	45.84	-74.47	280	18	14	12.9	0.31	0.32	0.01	0.407

Sites	Latitude	Longitude	altitude (m)	n	A	AC	H <sub>O</sub>	H <sub>E</sub>	F <sub>IS</sub>	HW
<i>Sunfish</i>										
N2	46.13	-74.30	380	16	25	24.1	0.49	0.63	0.24	<b>0.004</b>
N3	46.12	-74.28	380	7	25	28.0	0.60	0.66	0.10	0.373
N4	46.10	-74.36	380	10	24	27.6	0.72	0.70	-0.03	0.987
N5	46.09	-74.33	380	16	31	31.1	0.71	0.70	-0.01	0.470
N6	46.10	-74.41	400	16	30	28.2	0.63	0.64	0.02	0.607
N7	46.09	-74.43	400	16	20	19.7	0.41	0.47	0.14	0.509
N8	46.07	-74.44	420	16	14	14.8	0.28	0.32	0.14	0.613
N9	46.09	-74.46	400	16	31	28.5	0.72	0.70	-0.02	0.828
N12	45.97	-74.07	320	19	20	20.4	0.46	0.49	0.07	0.205
N13	45.95	-74.09	300	21	20	21.0	0.50	0.57	0.13	0.476
N17	45.93	-74.40	360	16	20	24.4	0.54	0.52	-0.03	0.738
N18	45.87	-74.34	400	16	20	19.8	0.59	0.54	-0.11	0.322
N19	45.83	-74.34	360	16	14	13.3	0.31	0.38	0.19	0.059
N20	45.83	-74.35	360	15	12	12.7	0.36	0.38	0.07	0.897
N21	45.82	-74.34	320	16	18	18.9	0.46	0.56	0.19	0.241
R1	46.17	-74.37	400	16	30	31.6	0.73	0.74	0.01	0.624
R2	46.19	-74.38	380	16	31	31.1	0.68	0.72	0.06	0.152
R3	46.19	-74.43	380	16	28	29.6	0.65	0.70	0.08	0.645
R4	46.19	-74.45	320	16	30	29.9	0.63	0.71	0.11	<b>0.026</b>
R5	46.00	-74.77	220	19	27	27.8	0.63	0.70	0.10	0.074
R8	45.96	-74.42	420	16	15	19.8	0.46	0.53	0.15	0.391
R11	45.92	-74.47	300	16	24	17.3	0.60	0.65	0.08	0.159
R12	45.85	-74.41	400	16	19	19.3	0.50	0.53	0.06	0.179
R13	45.86	-74.38	380	16	18	18.1	0.52	0.50	-0.04	0.200

### *Genetic analyses*

A piece of the caudal fin was sampled and preserved in 95% ethanol for DNA extraction by proteinase K digestion followed by phenol-chloroform purification and ethanol precipitation (Sambrook *et al.* 1989). Nuclear DNA variation was characterized with microsatellite loci selected on the basis of their high variability from previous studies. Microsatellite loci used for pumpkinseed sunfish were Lepo10; Lepo14 (Schable *et al.* 2002); Lma116 (Neff *et al.* 1999); Lma29; RB7 and RB20 (McElroy *et al.* 2003). Microsatellite loci used for rock bass were AinA218; AinA117; AinC212; AinA2; AinA108 (Schwartz and May 2004); and RB20 (McElroy *et al.* 2003).

Polymerase chain reactions were conducted from a template of 12.5 µl including 1.25 µl of 10x buffer (Feldan Bio inc.,); 2.5 nM of each dNTP; 10 pmol of each primer; 1.5 mM of MgCl<sub>2</sub>; 0.2 Unit of *Taq* polymerase; and 100 ng of DNA. PCR reactions were as follow: an initial denaturation of 30 s at 92°C, followed by 45 cycles of 3 steps (15 s at 92°C; 15 s at 52°C, and 15 s at 68°C), and a final extension of 2 min. at 68°C. Amplification products were electrophoresed on a denaturing urea polyacrylamide gel and visualized using silver nitrate staining (Bassam *et al.* 1991).

### *Genetic diversity distribution*

Genetic diversity was assessed with the total number of alleles (A) and the total number of alleles corrected for the sampling size (Ac) using FSTAT (Goudet 2001). Observed heterozygosity ( $H_O$ ) and Nei's gene diversity ( $H_E$ ) of each population (Nei 1987) were estimated using Genetix version 4.05.2 (Belkhir *et al.* 1996–2004).  $F_{IS}$  values and pairwise  $F_{ST}$  were estimated using GenePop version 4.0 (Raymond and Rousset 1995). Departure from Hardy-Weinberg (HW) expectations and linkage disequilibrium (LD) among loci were tested using GenePop version 4.0 (Raymond and Rousset 1995). The presence of null alleles was tested by comparing  $F_{IS}$  values with  $H_E$  values for each locus and each sampling site with the software Fdist2 (Beaumont and Nichols 1996), as suggested by Girard (2011).

Population structure was inferred using the STRUCTURE 2.3.1. (Pritchard *et al.* 2000) using no admixture model with correlated allelic frequencies. Twenty five runs for k-group (k

= 1 to 15 groups) were performed with a burn in length and a Markov chain Monte Carlo (MCMC) of 100 000 iterations each. The number of groups (k) was determined with the statistic  $\Delta K$  related to the second order rate of change of the log probability of the data (Evanno *et al.* 2005; Earl and vonHoldt 2011). Population structure was also investigated with a correspondence analysis (CA) (Benzecri 1973) performed from average allelic frequencies of each sampling site with the package adegenet (Jombart 2008) in R software (R development core team 2009). Comparison of the genetic variation partitioning between the different populations structures was performed with the software Arlequin version 3.5.1.2 (Schneider *et al.* 2000). In a first partitioning, sampling sites were grouped according to their drainage system. In a second partitioning, sampling sites were sorted into STRUCTURE groups according to the majority rule (i.e. each site was classified on the basis of affiliation of individuals in each group). Finally, in order to assess whether individuals could be correctly assigned to the drainage where they had been sampled, discriminant analysis of principal components was performed (DAPC) (Jombart *et al.* 2010) with the package adegenet (Jombart 2008) in R software (R development core team 2009).

Under a scenario of stepping stone colonization process, gene diversity should be higher near source population and should decrease with distance from that source (Ramachandran *et al.* 2005). In order to assess the likelihood of such a scenario for the studied populations, Pearson's correlations between Nei's gene diversity ( $H_E$ ) and a set of variables describing the landscape were performed using the package vegan (Oksanen *et al.* 2010) in R software (R development core team 2009). The variables considered were the distance from Ottawa River (to assess upstream dispersal), distance from the most upstream populations (to assess downstream dispersal), altitude, the type of habitats (lake or river) and the drainage. Multiple regressions between gene diversity and the same set of variables were also performed using a backward selection of the variables. Backward selection of the variables was intended to keep in the model the variables that contribute the most to explaining the response variable.

Both migration and barriers to migration among populations were indirectly inferred from the genetic data. Connectivity among populations was assessed visually from the distribution of individuals into groups inferred from the STRUCTURE partition. Recent migrants were detected with the assignment algorithm of Paetkau *et al.* (2004) with 1000 simulated individuals and a threshold of 0.01, as implemented in the software GeneClass 2.0 (Piry *et al*

2004). Discontinuities in the pairwise  $F_{ST}$  matrix were identified as proposed by Manni *et al.* (2004) using the Monmonier's algorithm (1973) on a Gabriel's graph (Gabriel and Sokal 1969), as implemented in the package adegenet (Jombart 2008) in R software (R development core team 2009).

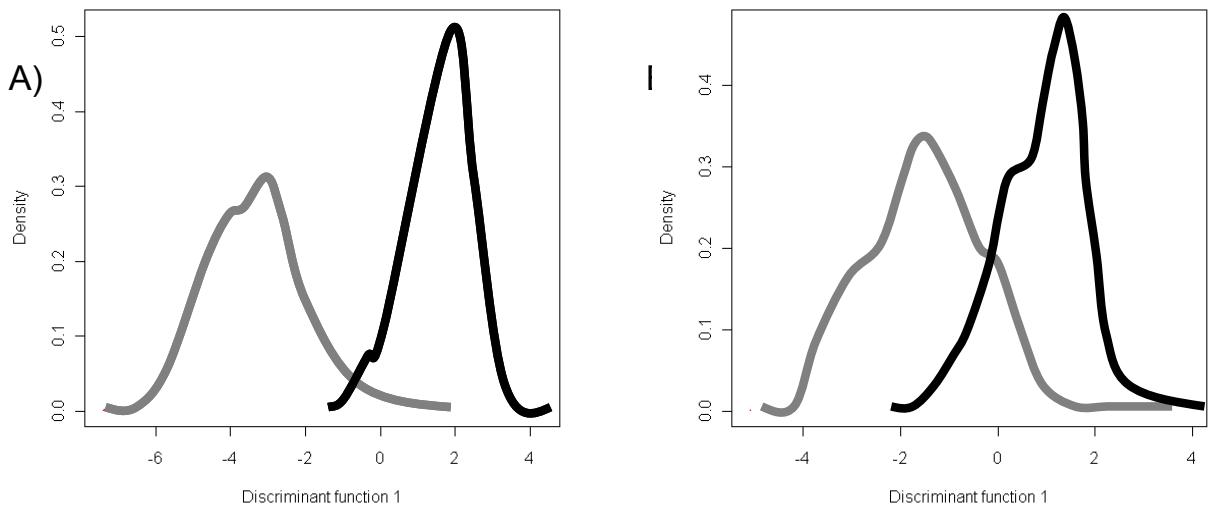
### 3.4 Results

#### *Distribution*

Of the 35 sites sampled, both species were detected in 14 sites, 16 sites contained only sunfish, and 5 sites contained only rock bass (Fig. 3.1A). Sunfish appeared more evenly distributed throughout the sampled area, whereas rock bass was not caught in some river sections (notably in the northern section of the R drainage).

#### *Rock bass*

The total number of alleles per sample site for the rock bass species varied from 14 to 36 and the Nei's gene diversity ranged from 0.32 to 0.68 (Table 3.1). There was neither significant deviation from HW expectations within sampling sites nor LD among loci. Global  $F_{ST}$  index for rock bass populations was 0.1762, which is highly significant ( $p<0.001$ ). The strong global  $F_{ST}$  value is comparable with what have been observed for other fish species in the region (Angers and Bernatchez 1998; Gagnon and Angers 2005) and suggests historical isolation among populations. DAPC analysis indicated that the drainage systems were well differentiated and allowed for differentiation between the two sampled drainages with a global probability of correct assignment of 99% (Fig. 3.2A). Grouping populations according to drainages explains 9.91% of the variation observed in rock bass populations (Table 3.2).



**Figure 3.2:** Results of discriminant analyses of principal components (DAPC). Differentiation between drainage systems (the black and the grey lines refer to N and R drainages, respectively) for rock bass (A) and for sunfish (B). The probability to correctly assign an individual to the drainage where it was sampled is 99 % for rock bass and 88% for sunfish.

**Table 3.2:** Hierarchical partitioning of the genetic variation among groups for rock bass and sunfish performed with Arlequin 1.2.3.4.5. Groups were defined according to drainages or groups inferred by STRUCTURE program. d.f., indicates degrees of freedom. Fixation index refers to  $F_{CT}$  (variation among group) and  $F_{SC}$  (variation among populations within group). Probability values are in parenthesis.

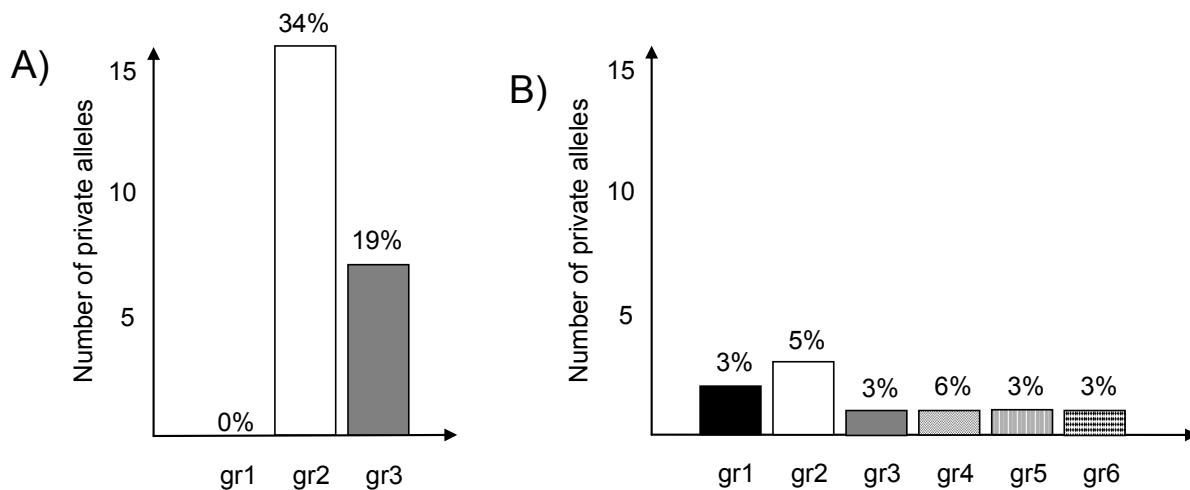
Species	Source of variation	d.f.	Sum of squares	Variance component	% of variation	Fixation index
	Among drainages	1	51.290	0.19466	9.91 (0.002)	0.09909
	Among populations within drainages	14	122.077	0.23725	12.08 (< 0.001)	0.13406
	Within populations	472	723.375	1.53257	78.01 (< 0.001)	
<hr/>						
Rock bass						
	Among groups <sup>1</sup>	2	85.219	0.23011	11.67 (< 0.001)	0.11673
	Among populations within groups	13	89.702	0.17742	9.00 (< 0.001)	0.10190
	Within populations	472	730.354	1.56371	79.33 (< 0.001)	

Species	Source of variation	d.f.	Sum of squares	Variance component	% of variation	Fixation index
	Among drainages	1	28.119	0.02253	0.98 (0.269)	0.00976*
	Among populations within drainages	22	426.473	0.56228	24.36 (< 0.001)	0.24600
	Within populations	732	1261.549	1.72343	74.66 (< 0.001)	
Sunfish						
	Among groups <sup>2</sup>	5	299.064	0.43303	17.95 (< 0.001)	0.17947
	Among populations within groups	18	159.200	0.23103	9.57 (< 0.001)	0.11669
	Within populations	732	1270.616	1.74883	72.48 (< 0.001)	

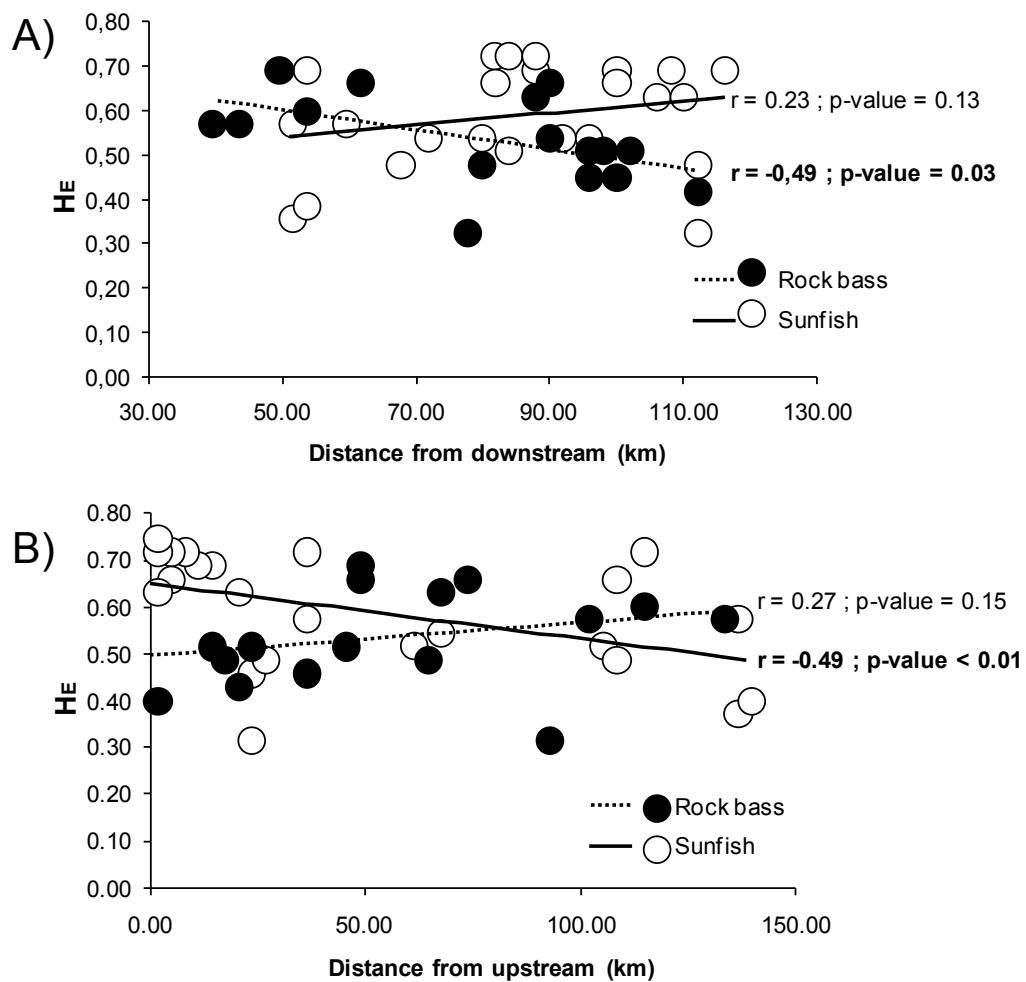
<sup>1</sup> Genetic groups for rock bass: R5, R6, R10, N1/ R7, R9, N16, N17, N22/ N3, N4, N6, N10, N11, N14.

<sup>2</sup> Genetic groups for sunfish: N19, N20/ R12, R13, N18/ R5, R8, R11, N9, N17/ R1, R2, R3, R4, N2, N3, N4, N5, N6 / N7, N8 / N12, N13.

The best model obtained with the STRUCTURE program retained three genetic groups (Fig. 3.1B). Grouping populations according to their affiliation to the different genetic groups explains 11.67 % of the total variation (Table 3.2). Population structure was confirmed by correspondence analysis obtained from allelic frequencies of each sampling site (see appendix: Fig. A1). The two groups located downstream were characterized by high genetic diversity and a large proportion of private alleles (i.e. alleles found exclusively in one genetic group; Group 2: 34%; Group 3: 19%) (Fig. 3.3A). In contrast, the upstream-most group displayed a low genetic diversity and no private alleles. In other words, all of the alleles detected in the upstream populations were also found in downstream sites. As a result of this net loss of alleles toward upstream populations, genetic diversity was significantly inversely correlated with the distance from the entrance point of each river (Table 3.3; Fig. 3.4A).



**Figure 3.3:** Number of private alleles in each genetic group. The number (on the Y-axis) and the percentage (numbers above each bar) of private alleles within each genetic group - identified by the STRUCTURE program - are indicated for rock bass populations (A) and for sunfish populations (B). Each genetic group is identified by the pattern used in Figure 3.1, B and C.



**Figure 3.4:** Relationships between Nei's gene diversity ( $H_E$ ) and geographic variables. Scatter plots showing the correlation between Nei's gene diversity and distance from downstream most point for rock bass (Pearson's  $r = -0.49$ ;  $p = 0.0279$ ) and sunfish (Pearson's  $r = 0.23$   $p = 0.1349$ ) (A). Distance from the upstream most point for rock bass (Pearson's  $r = 0.27$ ;  $p = 0.1519$ ) and sunfish (Pearson's  $r = -0.49$   $p = 0.0075$ ) (B).

Interestingly, two groups (Fig. 3.1B) overlapped between both river systems. This is suggesting that the drainages were not completely isolated from each other and some gene flow may have occurred between them, via an interconnection or via the most downstream sections of the drainages.

**Table 3.3:** Best linear models explaining the variation of Nei's gene diversity ( $H_E$ ) among sampling sites. Backward selection of the variables has been applied on a set of the variables including distance from Ottawa River (down), distance from the most upstream populations (up), altitude (alt), lake or river habitats (lake) and drainages (drainage).

	Equation	Adjusted R <sup>2</sup>	p-value
Rock bass	$H_E = 0.787 - 0.003 \text{ down} - 0.076 \text{ lake}$	0.276	0.048
Sunfish	$H_E = 0.972 - 0.001 \text{ alt} - 0.002 \text{ up} + 0.089 \text{ drainage}$	0.453	0.002

---

All variables are significant ( $p<0.05$ ) except for the variable lake

### *Sunfish*

The total number of alleles per sunfish populations ranged from 12 to 31, and the gene diversity of the sunfish populations ranged from 0.28 to 0.73. No deviation from HW expectations was detected except in the case of two populations (N2 and R4), which showed a significant heterozygosity deficiency ( $P<0.05$ ) (Table 3.1). However, this deviation was no longer significant when Bonferroni correction is applied (Hill 1966). Global  $F_{ST}$  index for sunfish populations is 0.2482 and is highly significant ( $p<0.001$ ). The two drainages were only weakly differentiated (Fig. 3.1C). Differentiation accounts for 0.98 % of the total variation and is not significantly different from zero (Table 3.2). Moreover, DAPC analysis allowed for only partial differentiation between the two sampled drainage systems and the global probability of correct assignment was 88%, far less than the global probability obtained for rock bass populations (Fig. 3.2B).

The best model obtained with STRUCTURE retained six genetic groups (Fig. 3.1C); 17% of the variation observed in sunfish populations is explained when populations were partitioned according to their affiliation to the different genetic groups (Table 3.2). Population structure was supported by correspondence analysis obtained from allelic frequencies of each sampling site (see appendices Fig. A1B). Three groups were shared between the two drainages (Fig. 3.1C), while three groups were exclusively found in the N drainage (except for one individual of the group 6 found in the sampling site R4). In contrast to rock bass, the highest level of genetic diversity is found in the most upstream sampling sites (Fig. 3.1C). Genetic groups of sunfish located downstream were characterized by lower genetic diversity and lower number of private alleles. However, private alleles were more evenly distributed among genetic groups in sunfish populations than in rock bass populations. This loss of diversity toward downstream sampling sites is consistent with the observation that genetic diversity was inversely correlated with distance from the most upstream sampling site ( $p < 0.007$ ) but not to the distance from Ottawa River ( $p > 0.1347$ ) (Fig. 3.4B; Table 3.3).

### *Connectivity*

Monmonier's algorithm (1973) allowed for the detection of discontinuities within drainages for both species (Figs. 3.5A and 3.5B). One barrier was detected among rock bass populations, which isolated the most upstream population (Fig. 3.5A, N1) in the N drainage. Two barriers were detected among sunfish populations, which isolated upstream populations (Fig. 3.5B, N9, N19, and N20) in the N drainage. These populations are also characterized by distinct genetic composition when compared to neighbouring populations (Figs. 3.1B and 3.1C).

The following evidence suggests that gene flow occurred between the two drainages: 1) Putative interconnections were detected, yet Monmonier's algorithm failed to detect genetic boundaries between some drainage sections that were not otherwise connected by the permanent river network (Figs. 3.5A and 3.5B). Interestingly, a putative interconnection was shared by the two species (arrows c1 and c4 for rock bass and sunfish, respectively). 2) The assignment method detected 15 and 8 recent migrants between rock bass and sunfish populations, respectively

(Figs. 3.5C and 3.5D), among which 3 migrants of each species appear to have crossed from one drainage to another. Thus, nearly half of the migration events detected between sunfish populations occurred between drainages. 3) Finally, the distribution of populations from a given genetic group into different drainages also suggests historical and/or current interconnections between R and N drainages (Figs. 3.1B and 3.1C). Genetic groups encompassing the two sampled drainages could be compared on the basis of pairwise  $F_{ST}$  values observed among populations from different drainages. An increase of inter-drainage differentiation toward downstream region is observed in sunfish populations, while inter-drainage  $F_{ST}$  ranges between [0.002 - 0.08] in the group 1; [0.11 - 0.30] in the group 2 and [0.19 - 0.20] in the group 3.

### **3.5 Discussion**

#### *Impacts of colonization*

Differences observed between the geographic distribution of rock bass and sunfish species suggest that they did not have the same opportunities to colonize the different sites (Fig. 3.1A). Moreover, several fish species present a distribution in the region similar to the one of rock bass or sunfish (Chamberland 2011), supporting the idea of two distinct colonization patterns.

More importantly, the spatial organization of genetic diversity is clearly different between species and suggests opposite patterns of colonization (Figs. 3.1B and 3.1C). The pattern of genetic diversity indicates that rock bass colonized the study area from downstream causing a loss of diversity in upstream populations, while independent founder events in both drainages increased inter-drainage differentiation. Sunfish populations revealed an opposite trend; upstream populations were less differentiated and were characterized by a higher genetic diversity and more private alleles. Such a pattern is consistent with a downstream colonization by sunfish. Downstream colonization is not unlikely as the former glacial lake Ojibway Barlow, which was an important colonization route into the territory, was located north of the study area, (Legendre and Legendre 1984; Mandrack and Crossman 1992).



**Figure 3.5:** Migration among populations and barriers to migration inferred from genetic data. Putative genetic barriers and interconnections are shown for rock bass (A) and for sunfish (B) populations. Black bars represent putative barriers within drainages, as identified with Monmonier's algorithm. Arrows represent putative connections between different arms of the two sampled drainages, for which no genetic boundaries were identified. Finally, recent migration events, as detected with Genclass 2, are shown in panel (C) and (D) for rock bass and sunfish populations, respectively. Intra-drainage migrants are identified with dotted arrows while inter-drainage migrants are identified with solid arrows. The gray and the black lines refer to R and N drainage, respectively.

The specific pattern of genetic differentiation between drainages may be explained by differences in dispersal capabilities between species. Thus, rock bass could have been unable to migrate from one drainage to another, upstream of the drainages, because of a lower dispersal capability. Alternatively, environmental characteristics may have prevented rock bass to successfully colonize some sections of the R drainage and limited inter-drainage migration. However, these hypotheses require important inter-drainage migration among sunfish populations which is not supported by the examination of the current hydrological networks. Indeed, no current interconnection has been observed in the most upstream part of the drainages. More importantly, these hypotheses could not explain the opposite distribution of genetic diversity observed between species (Fig. 3.4).

#### *The effect of slow changes in genetic diversity on gene flow estimates*

Observed spatial genetic variation is often strongly influenced by historical processes and biased toward past demographic events. For instance, in the present study, as observed in many other studies performed in recently glaciated areas (e.g. Bernatchez and Wilson 1998; Boizard *et al.* 2008), genetic diversity is related to the distance along colonization routes and does not reflect the current effective size of the populations. In addition, such a very recent system (ca. 10 KY) may not be at drift-migration equilibrium and genetic differentiation estimates may not represent current gene flow between drainages (Whitlock and McCauley 1999), but historical processes. Thus, the net increase of inter-drainage genetic similarity toward upstream regions observed among sunfish populations could be a consequence of colonization from upstream regions; historical gene flow and incomplete lineage sorting rather than the consequence of strong current gene flow between drainages. As a result, the presence of populations with similar genetic composition in distinct drainages (Figs. 3.1B and 3.1C), suggests gene flow between drainages, leads to potentially false genetic signal of current connectivity.

Moreover, it is difficult to estimate current gene flow based solely on assignment methods. Even if the detection of recent migrants is reliable, the identification of source populations remains problematic, particularly when some populations have not been sampled or when differentiation among populations is weak and multiple sources are likely (Castric and Bernatchez 2004). Such difficulty is demonstrated in this study by the identification of a recent migrant in the population R5, which came from the population N18 (Fig. 3.5D). Such a remote source population is at best surprising and at worst unlikely.

The use of a limited number of loci - as is often the case in studies on non-model organisms - may also introduce an important bias as  $F_{ST}$  estimates varied considerably among loci (Nei et al. 1977). This bias could be particularly misleading when  $F_{ST}$  is small among populations (Waples 1998). In this study, the use of a limited number of loci may have prevented to find sharper inter-drainage differentiation within each genetic group. The problem is that, regardless of the number of loci used, it is difficult to conclude between a lack of resolution of sampling and an absence of differentiation. It is therefore even more contentious to conclude a high level of current gene flow based solely on an absence of genetic differentiation.

### *Concluding remarks*

In regions formerly covered by glaciers, current spatial genetic variation could be the result of different opportunities of dispersal from distinct refuges (*e.g.* Girard and Angers 2006), secondary contacts between distinct glacial races (*e.g.* Turgeon and Bernatchez 2001), past founder events along colonization route (*e.g.* Boizard *et al.* 2008), former landscape configuration (*e.g.* Gagnon and Angers 2005), and, as evidenced in the present study, the patterns of colonization along a hydrographic network. Thus, it is important to differentiate historical from current processes in order to disentangle potential effects of each process on current spatial genetic variation. The present study demonstrates through empirical evidence that such historical factors may blur our capacity to estimate current processes among populations, as illustrated by the estimates of connectivity generated by this study within and between the two sampled drainages.

Genetic differentiation among populations could be interpreted as a good proxy of gene flow among populations (Rousset 1997; Templeton 1999; but see Whitlock and McCauley 1999). However, caution should be applied to such interpretations, as the potential effects of historical factors must also be acknowledged. There has been an exponential growth in the field of landscape genetics in the last two decades (Sork and Waits 2010, Storfer *et al.* 2010). An important contribution of this discipline has been to assess the connectivity among populations according to the partitioning of genetic diversity. Future developments in landscape genetics should include formal tests that (1) use objective criteria to detect historical effects on current distribution of genetic diversity and (2) disentangle the effect of historical and current processes on populations. For instance, when possible, it is wise to test the relationship between current landscape and genetic variability by controlling for historical landscape configuration. Historical, geological, geomorphological and paleontological data being more accessible, this type of analysis should be more feasible and promoted.

### **3.6 Acknowledgements**

We are grateful to Guillaume Bourque for his early comments on this work. We are also grateful to Dr. Daniel Boisclair's team for their help in the field. FC was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was supported by a "Fonds Québécois de Recherche en Nature et Technologie" (FQRNT, Quebec) grant to BA.

### **3.7 Author contribution**

Original idea: FC and BA; Analyses: FC; Redaction: FC and BA

**CHAPITRE 4 : ARTICLE 3**

**CONSTRAINED EXPLORATION OF POPULATION CONNECTIVITY: A**

**FLEXIBLE WAY TO ASSESS HYPOTHESES ABOUT DISPERSAL**

**PROCESSES**

**Frédéric Cyr, Bernard Angers, Anaïs Lacoursière-Roussel, Philippe Girard**

Soumis à la revue Genetics

#### 4.1 Résumé

Évaluer les voies de dispersion reliant des populations est essentielle pour comprendre leur évolution et à en assurer la gestion. Dans ce travail, nous proposons une méthode souple pour inférer les voies de dispersion entre des sites en considérant les processus externes qui influencent la dispersion des individus dans un paysage donné. Notre algorithme i) permet l'identification des éléments du paysage ou des facteurs extérieurs influençant la dispersion et ii) contribue à l'identification des effets historiques ou des relations inattendues entre la dispersion et certains éléments du paysage. Nous présentons les résultats de notre méthode dans plusieurs scénarios simulés d'une complexité croissante. Malgré les effets de la différenciation globale et de la taille de l'échantillon, la méthode semble très efficace pour tous les scénarios de simulations testés. Pour illustrer son potentiel dans les systèmes naturels, l'approche a été appliquée sur deux jeux de données empiriques. Le premier jeu de données représente la variabilité génétique au sein de populations de poissons dont la dispersion est limitée au sein d'un réseau hydrologique. Le second jeu de données représente la variabilité génétique au sein de populations d'une espèce envahissante avec une dynamique de dispersion complexe et rapide et qui s'étend sur une côte littorale.

## Abstract

Assessing the dispersal pathways linking populations is essential to understanding population evolution and to ensuring population management. In this work, we propose a flexible framework to infer dispersal pathways among sites by considering the external processes affecting dispersal of individuals in any given landscape. Our algorithm i) enables the identification of landscape elements or external factors impacting dispersal and ii) contributes to the identification of historical effects or unexpected relationships to missed, or overlooked, landscape elements. We present the performance of our method in several simulated scenarios of increasing complexity. Despite impacts of global differentiation and sample size, the method appears quite effective for all of the simulations scenarios tested. To illustrate its potential in natural systems, the approach was applied on two empirical data sets: i) a fish species whose dispersal is restricted within a hydrological network and ii) an invasive species with complex and rapid dispersal dynamics that span across a littoral coast.

## 4.2 Introduction

Dispersal among populations plays a fundamental role in the ecology and evolution of natural systems (Garant *et al.* 2007; Gibbs, Saastamoinen *et al.* 2010). Such processes may bring local and regional benefits to a system by increasing average fitness across a heterogeneous landscape (Olivieri *et al.* 1995), reducing kin competition (Hamilton and May 1977) and inbreeding depression (Greenwood *et al.* 1978) in the source populations, allowing persistence of small populations, or spreading adaptive traits (Manel *et al.* 2003). However, dispersal may also have substantial disadvantages. Dispersers may suffer fitness reduction either because of higher mortality risk during transit (Hanski *et al.* 2000) or reduced survival and reproduction success after reaching the host population (Donovan *et al.* 1995). Dispersal may also destabilize the dynamics of the entire system if it occurs toward lower quality sink populations (Garant *et al.* 2007) and may favour system extinction through disease spread (Alexander and Antonovics 1988) and the introduction of invasive species (Wilcove *et al.* 1998).

As such, assessing the dispersal pathways (i.e. connectivity) linking natural populations is essential to understanding their evolution and managing their fate. In recent years, several methods have been proposed to estimate connectivity from genetic data. These methods have focused primarily on the development of genetic metrics (Slatkin 1993; Rousset 1997; Piry *et al.* 2004; Dyer *et al.* 2010), distance metrics that adequately represent landscape resistance (Adriaensen *et al.* 2003; Epps *et al.* 2007; McRae and Beier 2007), or statistical methods that link genetic data to landscape data (Angers *et al.* 1999; Murphy *et al.* 2008; Cushman and Landguth 2010). These approaches are highly successful in assessing dispersal corridors (LaRue and Nielsen 2008) and invasion routes by exotic species (Estoup *et al.* 2004) as well as estimating the permeability of fragmented landscapes (Coulon *et al.* 2006).

One difficulty of this kind of approaches based on correlation between landscape and genetic differentiation is that the factors that could impact dispersal are numerous and multiple dispersal hypotheses should be considered. Landscape features and external vectors facilitating dispersal can be dynamic in time and/or species-specific. If ecological or historical data are lacking, central features affecting

dispersal could be missed or overlooked (Cyr and Angers 2012). To circumvent such limitations, methods have been developed to infer genetic discontinuity directly from matrices of pairwise genetic differentiation (Manni *et al.* 2004; Paquette and Lapointe 2009). The advantage of these methods lies in their independence from the landscape. However, these methods may be less powerful precisely because they leave out relevant information about the landscape. Moreover, they might be perceived as incomplete by some researchers as they do not identify true dispersal routes among populations, which is valuable information for management decisions.

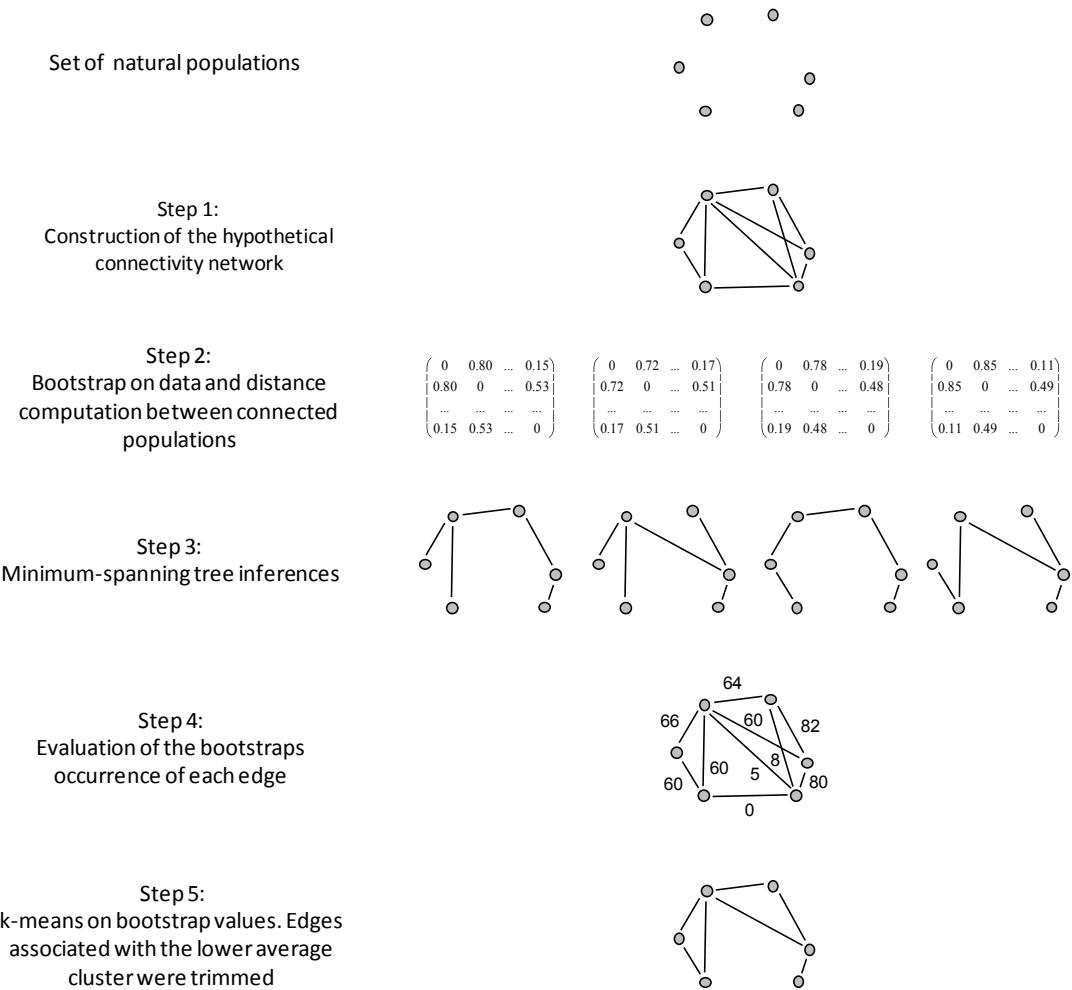
Another limitation of the approaches proposed in the literature come from the use of global correlation tests. Global correlation tests are poorly informative: one dispersal hypothesis may receive a stronger support; it does not necessarily imply that the other hypotheses are all wrong; or that the best hypothesis is truly the best in the whole sampled region. Of course, the use of partial correlation may allow controlling for possible correlation among rival hypotheses. However, there is a need for an index which provides site by site evaluation of connectivity.

In this study, we propose a flexible new framework to infer dispersal pathways among sites given any hypothetical dispersal process. Our algorithm has been developed to facilitate the identification of i) the dispersal routes according to genetic differentiation across sample sites and ii) the external processes facilitating or restricting individual movements in any given landscape. Our algorithm is efficient whether or not prior information about landscape is available. Moreover, the combination of several dispersal hypotheses in a composite hypothesis could lead to an efficient site by site comparison among hypotheses. In the following section, we provide a detailed description of the assumptions behind our approach. We then evaluate its potential with simulations using several connectivity structures, migration rates, sample sizes, and numbers of loci. Finally, in order to illustrate the application of the method, we reanalyze two microsatellite datasets: i) creek chubs (*Semotilus atromaculatus*) in a complex hydrological network of Québec (Canada) published by Boizard *et al.* (2009) and ii) the invasive tunicate *Botryllus schlosseri* sampled in ports and marinas of Nova Scotia (Canada) by Lacoursière-Roussel *et al.* (2012).

### 4.3 Materials and Methods

#### *Description of the method*

The method consists of determining the routes that minimize the genetic distance among a set of sampling locations over a network constrained according to a given dispersal hypothesis. The method is divided into five steps (Fig. 4.1):



**Figure 4.1:** Schematization of the five steps required to build a minimum spanning network that is constrained spatially and genetically.

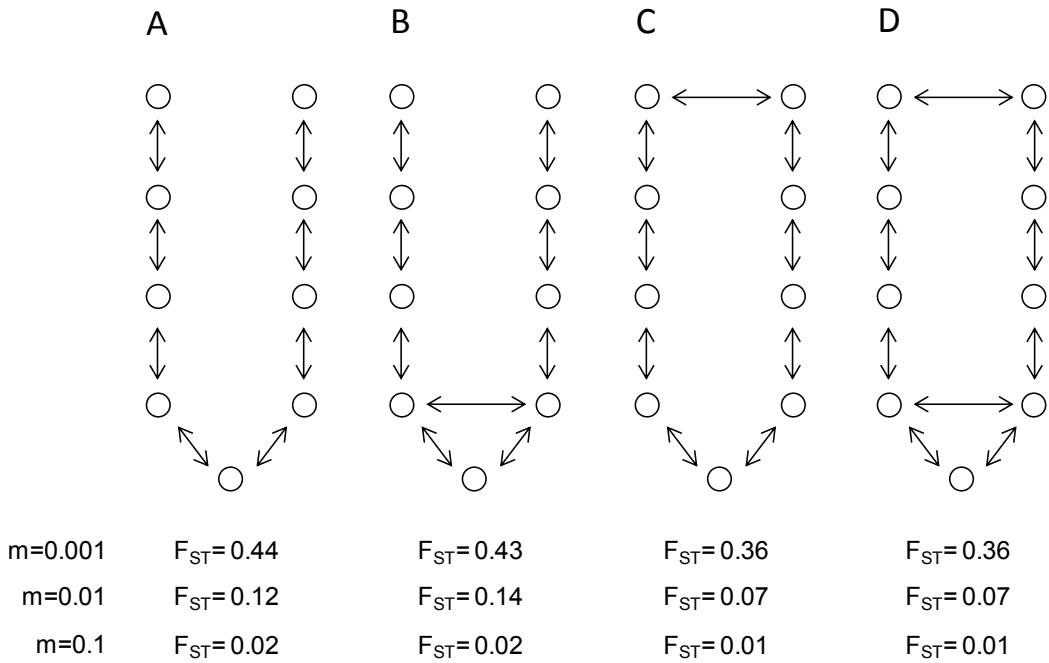
1. The first step consists of building a connectivity network to link all of the sampled sites together (Fig. 4.1, Step 1). The construction of this network can be adjusted to any dispersal hypothesis. If no prior information is available, a first exploration run consists of using a standard geographic triangulation (*e.g.* Delaunay triangulation), which only takes into account the spatial coordinates of the sites and does not consider external facilitating dispersal mechanisms (*e.g.* human translocation, hydrological network) or landscape barriers (*e.g.* mountain chains, landmass for fish). At this step, one may also choose to test all possible links among sites. As a result, this step has tremendous flexibility; the only limitation is that all sites being considered must be connected to each other by at least by one path.
2. The second step consists of calculating pairwise genetic distances among the sites that are connected by the hypothetical network built in Step 1 (Fig. 4.1, Step 2). This step is repeated multiple times through bootstraps over loci. Depending on the temporal scale one wishes to underline, different genetic distances can be used (Takezaki and Nei 1996). The choice of a specific distance will influence the connectivity structure inferred by the method. In this study, the chord distance (Cavalli-Sforza and Edwards 1967) was used. This distance, which is based on differences in allelic frequencies, is related to mid- to short-term temporal scales (Takezaki and Nei 1996; Angers and Bernatchez 1998).
3. The third step consists of calculating a minimum spanning tree (Prim 1957), which represents the path of minimal genetic length and connects all of the points together for each of the distance matrices produced in Step 2. Since we do not know the exact number of links that connect populations, the use of a less stringent graph algorithm, at this step, may represent a source of alpha error by adding false links among populations.
4. The fourth step involves calculating the occurrence of every link of the hypothetical network built in Step 1 in the set of trees produced in Step 3.

5. The fifth step consists of determining which links of the hypothetical network should be considered as dispersal routes. This identification can be performed visually based on bootstrap values, but the user may choose, as an option, to use an objective criterion instead. We suggest using a k-means algorithm (MacQueen 1967) to cluster bootstrap values in  $k=2$  groups. All of the edges associated with the cluster that are characterized by the higher bootstrap average are considered to be connectivity routes. This latter step is optional and should be used cautiously. Considering that the k-means algorithm will always cluster the values into two groups, it is wise to compare the automatic clustering discrimination with a visual analysis of the frequency distribution of the bootstrap values to ensure that the distribution is truly bimodal. If it is not, the connectivity network is unresolved and the k-means results should be ignored.

An R script performing this algorithm is available for free download at [http://www.bio.umontreal.ca/personnel/ANGERS\\_Bernard/index.html](http://www.bio.umontreal.ca/personnel/ANGERS_Bernard/index.html).

#### *Simulation study*

The reliability of the method was assessed at various rates and patterns of dispersal. Twelve scenarios, based on a combination of four connectivity patterns and three migration rates (0.001, 0.01, and 0.1), were simulated. The simulated patterns represented increasing levels of complexity (Fig. 4.2 A-D), in which one, two, or three possible paths linked nine populations of 500 individuals together. Genetic datasets were generated using the software Nemo v2.1.4 (Guillaume and Rougemont 2006) and consisted of 20 independent loci, each of which had 20 possible allelic states. Loci evolved under a KAM model (K-allele model) and mutation rate was fixed to  $10^{-4}$ /loci/generation.



**Figure 4.2:** The simulations scenarios used in this study. The scenarios are based on a combination of four connectivity patterns (A-D) and three migration rates (0.001, 0.01, and 0.1/individual/population/generation). Global  $F_{ST}$  for each combination of the four connectivity patterns and three migration rates are provided.

For each scenario, the migration rate was fixed in time and across all of the connections. Bi-directional and symmetric migration was set on all of the edges. At the beginning of each simulation, only the root population was occupied and evolved as a closed system for 1000 generations; in doing so, the population moved toward drift-mutation equilibrium. After that period, colonization of other populations followed the edges and migration rate implemented in the connectivity structure. The simulation stopped 5000 generations after the completion of the colonization process. The migration was kept active until the end of the simulation. Random subsamples of  $N$  individuals per population (20 or 50) and of  $L$  loci (10 or 20) were performed for each simulation scenario. Fifty replicates of each combination ( $N, L$ ) were performed for all of the scenarios.

To illustrate the usefulness of a constraint network, our method was applied on each subsample whereby starting networks were built either with a Delaunay triangulation (spatial constraint) or using all possible links (no constraint). The performance of the method was evaluated for each simulated dataset and for both of the hypothetical networks by its capacity to correctly recover the connectivity pattern under which the data were generated.

### *Applications*

- 1) Boizard *et al.* (2009) assessed the role of interconnections and barriers on the genetic structure of the creek chub (*Semotilus atromaculatus*) in a complex hydrological network located on the southern margin of the Canadian Shield (46°40'N, 73°25'W). This hydrographic system has a typical radial organization, but the main drainages (Fig. 4.7: A, B, C, and D) are only connected 50 km downstream in the St. Lawrence River (Fig. 4.7). Interestingly, the spatial proximity of their headwaters lead the authors to hypothesize that seasonal interconnection may have emerged among them due to seasonal freshets or changing water levels. The genetic datasets of Boizard *et al.* (2009) consist of 32 lakes from which 9 to 20 individuals were sampled, for a total of 538 individuals, each of which was characterized by eight nuclear microsatellites loci. The global differentiation in this system was estimated as  $F_{ST} = 0.53$ . Our method was applied on this dataset using a Delaunay network, which uses only the spatial coordinates of the sampling sites, but does not consider the inherent connectivity that is caused by hydrological configuration.
  
- 2) Lacoursière-Roussel *et al.* (2012) sought to clarify the ways in which commercial shipping and recreational boating contributed to the marine invasion of the colonial tunicate *Botryllus schlosseri* in Nova Scotia, Canada (43-46°N and 60-66°W). This invasive filter feeder has a cosmopolitan distributional range (Van Name 1945) and has recently become the dominant fouling invasive species in Nova Scotia (Sephton, *et al.* 2011). These characteristics, combined with a low self-dispersal capacity (Svane and Young 1989; Ayre *et al.* 1997), have made this species an important component of the study of anthropogenic invasive dispersal mechanisms. Lacoursière-Roussel *et al.* collected a total of 874 individuals from 5 ports and 21 marinas and performed

analysis using 10 nuclear microsatellite markers. The global differentiation of this system was evaluated as  $F_{ST}=0.11$ . Our method was used to describe the connectivity structure among these sample sites. This was a challenge since the connectivity among these sites most likely results from a complex interaction between space, natural dispersal, and anthropogenic transportation. As such, the method was implemented with the following two hypothetic networks, which were used as starting points: i) a strict spatially constrained network built with a Delaunay triangulation and ii) a network built according to boating routes (see Lacoursière-Roussel *et al.* for more details). Each network was then processed separately by the method. To obtain a more precise portrait of the connectivity structure and the dispersal mechanisms behind the invasion of this tunicate, the dispersal routes resulting from each hypothetic network were then overlapped onto a third network, which was also processed by the method.

#### 4.4 Results

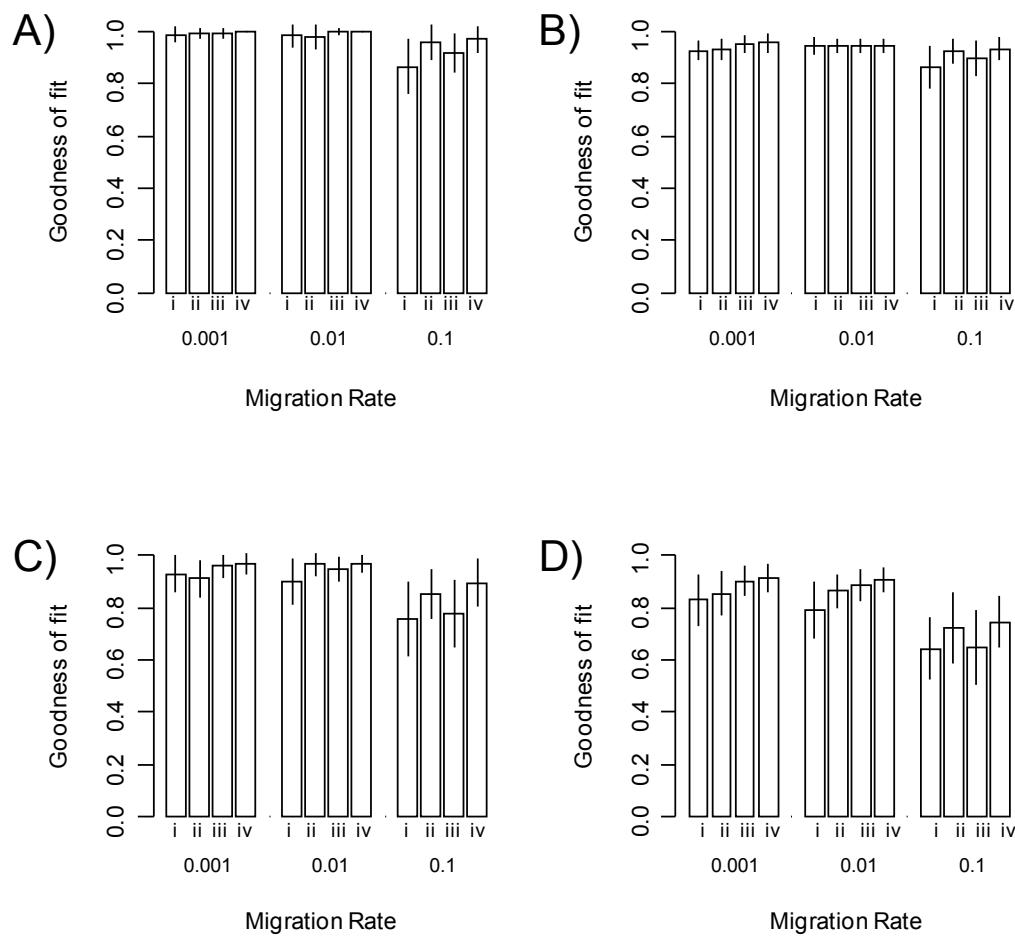
##### *Simulation study*

Overall, the proportion of the links correctly recovered by the method across the different simulation scenarios ranged from 0.64 to 1.0, with a very satisfactory average of 0.90 (Fig. 4.3). As expected from genetically-based methods, the results highlighted the influence of sample size, migration rates, and connectivity patterns on the reliability of the method. Increasing sample size increases the effectiveness of the method, regardless of the connectivity pattern tested. This effect became especially noticeable in cases of complex connectivity structures or high migration rates. Interestingly, our simulations suggest that the effectiveness of the method is improved more significantly by increasing the number of individuals than by increasing the number of loci (Fig. 4.3). For a given connectivity pattern, the method was more reliable when migration rates were lower. When the migration rate was 0.1, the performance of the method decreased noticeably, with an average of 0.83 links correctly recovered. Finally, accuracy of the method decreased as the complexity of the connectivity pattern increased. The performance of the method was higher for connectivity patterns characterized by a single path linking populations (on average the proportion of links correctly recovered for the

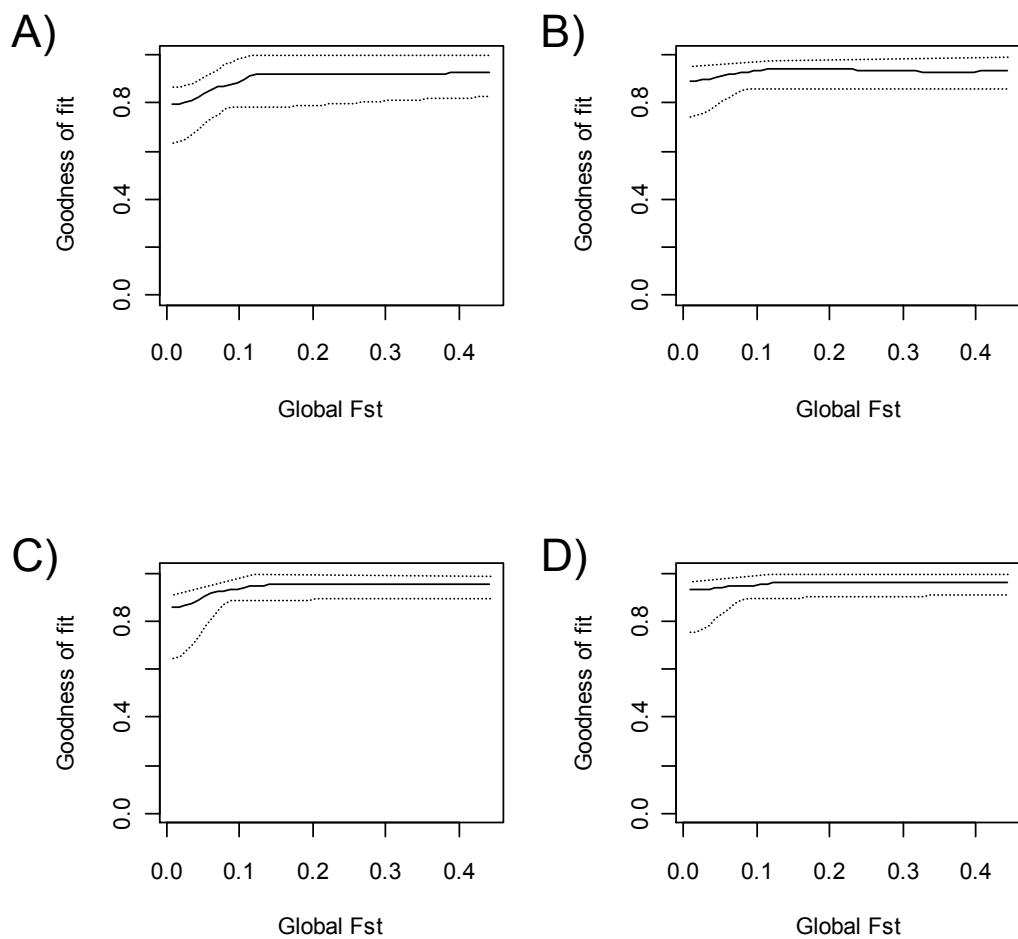
connectivity pattern A was 0.97; Fig. 4.3A). The accuracy of the approach was approximately equivalent for the two connectivity patterns characterized by two possible paths, regardless of the position of the supplementary path (with averages of 0.93 and 0.90; Figs. 4.3B and 4.3C). The lowest reliability was obtained with the connectivity pattern that had three possible paths (average of 0.81; Fig. 4.3D).

Given that migration rate and connectivity complexity can be related to differentiation among populations, the accuracy of the method can be correlated to global  $F_{ST}$  (Fig. 4.4). A slight increase in differentiation leads to a steep increase in the performance of the method, which reaches its optimal reliability when  $F_{ST}$  is above 0.1. However, the performance of the method remains influenced by the complexity of the connectivity structures regardless of differentiation. Indeed, the optimal reliability was different depending on the connectivity pattern (Fig. 4.4). Also, when controlling for  $F_{ST}$ , there was a significant partial correlation between goodness of fit and the number of paths.

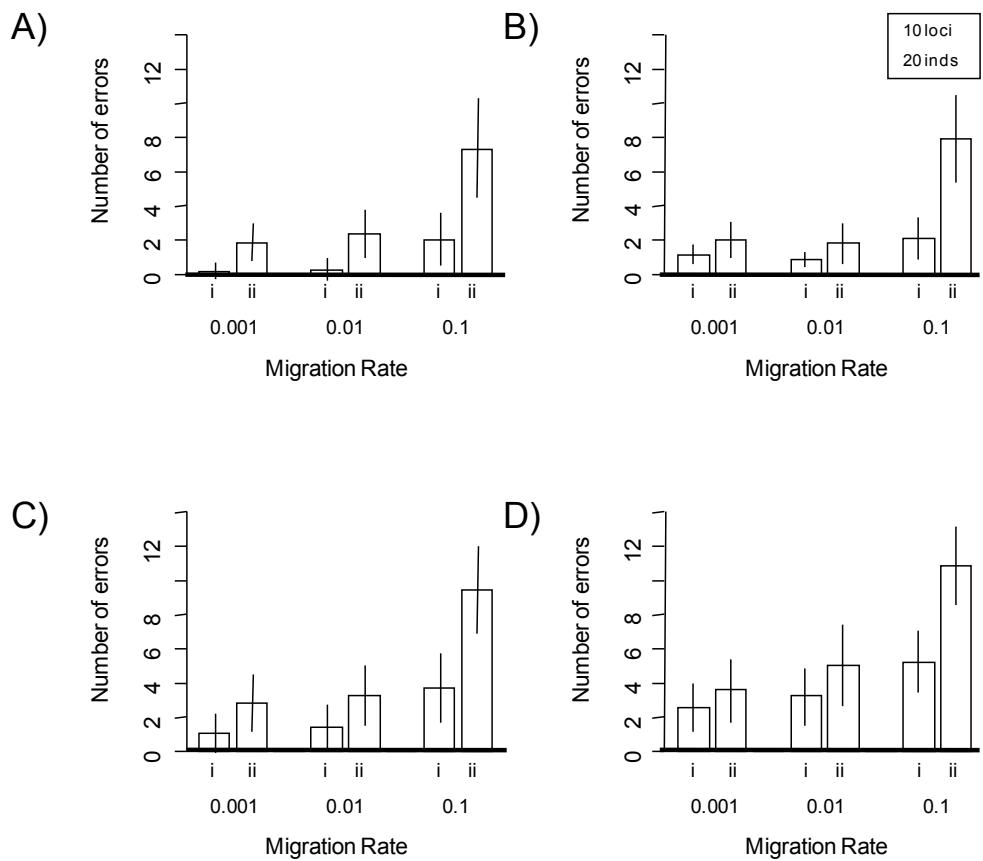
When compared to analysis without constraint, a Delaunay constraint increases the capacity to recover the connectivity pattern linking populations together and diminishes the impact of connectivity pattern and the migration rate implemented in the simulations (Fig. 4.5). Nevertheless, the impact of the constraint was less important when the migration rate was weak and there was a large sample size. However, the Delaunay constraint was clearly helpful for the simulation that involved lower genetic differentiation, which were caused either by strong gene flow (0.1; Fig. 4.5) or complex connectivity structures (Fig. 4.5, pattern D).



**Figure 4.3:** Goodness of fit of the proposed procedure. Goodness of fit is evaluated as the mean proportion ( $\text{mean} \pm \text{SD}$ ) of true and false links correctly identified across samples for each simulated connectivity pattern tested in this study (A-D). For each connectivity pattern, three migration rates (0.001; 0.01; 0.1) and four sample sizes (i: 10 loci - 20 inds; ii: 10 loci-50 inds; iii: 20 loci-20 inds; iv: 20 loci-50 inds) were tested.



**Figure 4.4:** Relationships between global differentiation (FST) and the goodness of fit of the method. The combined relationships of all of the connectivity patterns (continuous line) and the relationships for connectivity pattern A (upper line) and D (bottom line) are shown. The performance of the method was evaluated using four distinct sample sizes (A: 10 loci - 20 inds; B: 10 loci-50 inds; C: 20 loci-20 inds; D: 20 loci-50 inds).

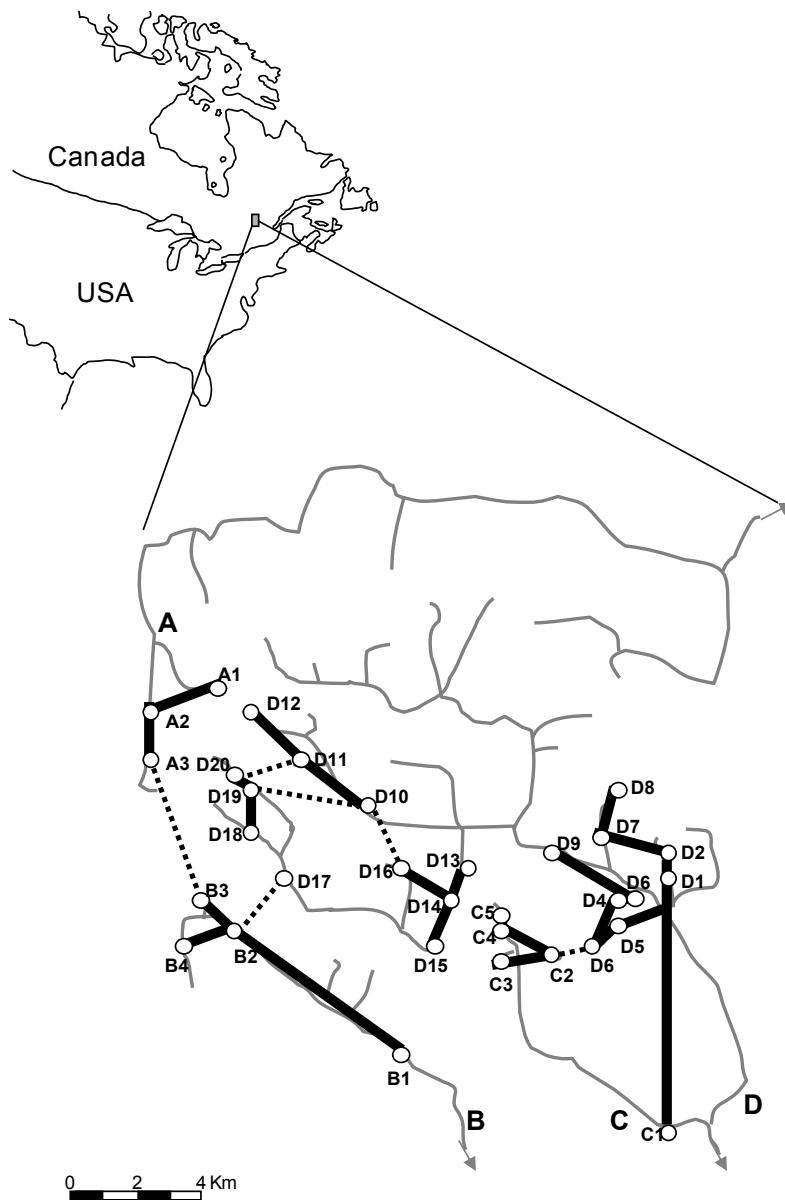


**Figure 4.5:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 10 loci and 20 individuals.

*Application to the creek chub case*

The dispersal network inferred by the method in the creek chub case was discontinuous and revealed three isolated zones (Fig. 4.6). Interestingly, these zones were not entirely consistent with the drainages. The first zone was composed of all of the sites from the drainage C along with those from the eastern part of the drainage D (D1-D9). A second group was comprised of the sites located in the western edge of the drainage D (D10-D20, except D17). Finally, the third group was composed of the sites from the drainages A and B and the site D17.

The inconsistencies between the distinct zones and the current drainages were the result of four links. The method identified three interconnections between sites that were spatially close but belonged to different drainages (C2-D6, B1-D17, and A3-B3). In contrast, the method did not retain the link between sites D9 and D10, even though they were linked together in the hypothetical Delaunay network (not shown). Finally, some inconsistencies were also detected within the western edge of drainage D; the permanent hydrological network did not support links D10-D16, D10-D19, and D11-D20.

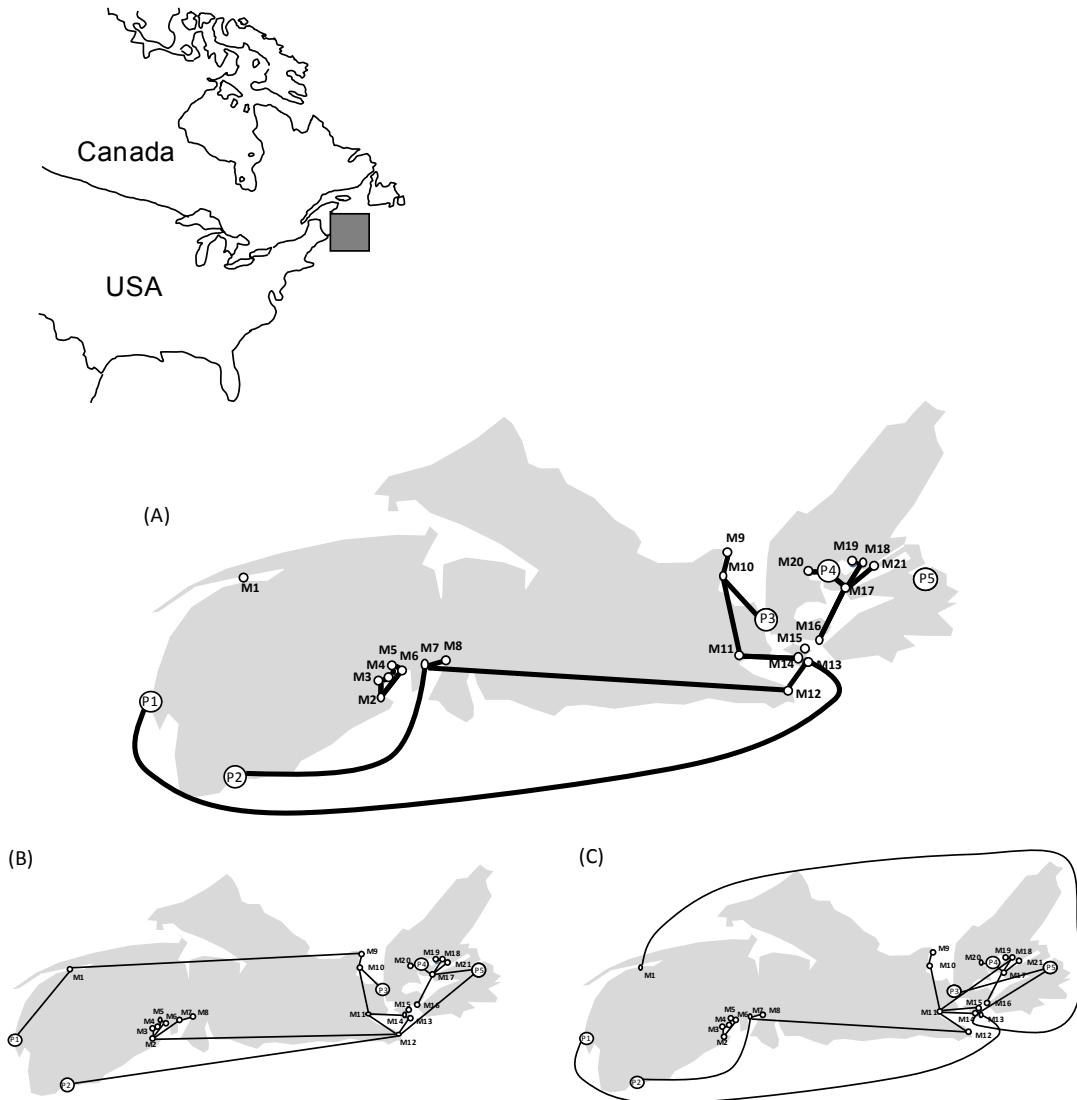


**Figure 4.6:** Connectivity network among the populations of the creek chub (*Semotilus atromaculatus*) in four drainages (A-D) of the Mastigouche Wildlife Reserve (Quebec, Canada). Solid lines represent connectivity links consistent with the permanent hydrological network. Dotted lines represent connectively links that are not consistent with the permanent hydrological network. Isolation within the connectivity network allowed for the identification of three distinct regions. See the original publication by Boizard *et al.* (2009) for further information about lake identification.

*Application to the colonial tunicate Botryllus schlosseri*

The case of the invasive tunicate *B. schlosseri*, (Lacoursière-Roussel *et al.*, 2012) was investigated using Delaunay triangulation and a network built according to boating routes. Globally, most links kept by the method in the Figure 4.7A were shared by both the spatial and the boating networks (Figs. 4.7B and 4.7C). However, noticeable differences were observed between the two outputs, especially in regard to the edges that connected geographically distant sites. While some of these links were simply removed in the final solution (see M1 and P5), all of those retained were associated with the boating connectivity pattern (P1/M13, P2/M7, and M7/M12).

The tunicate dispersal network suggested by our method was characterized by three distinct connectivity networks (Fig. 4.7A) and three isolated locations (M1, M15, and P5). Interestingly, three ports (P1, P2, and P3) were found at the tips of the larger connectivity network (Fig. 4.7A). In contrast, the two other networks appeared highly localized and consisted mainly of marinas. The most connected site (four links) was a small marina (M17) located in the center of the Bras d'Or Lake region, a semi-closed saltwater environment heavily crossed by recreational boaters (Lacoursière-Roussel *et al.*, 2012).



**Figure 4.7:** (A) Connectivity network inferred among ports (P) and marinas (M) invaded by *B. schlosseri* in Nova Scotia, Canada. The preliminary networks that served to build the final hypothetic network (see method) are shown in (B) and (C). (B) is the connectivity network inferred by the method when using a Delaunay triangulation and (C) is the connectivity network inferred from the boating routes. See the original publication by Lacoursière-Roussel *et al.* (2012) for more information about the identification of marinas and ports.

## 4.5 Discussion

Defining the connectivity structure that links populations or sample sites together is a widespread objective in many scientific disciplines, from landscape genetics to evolutionary ecology (Storfer, *et al.* 2010). Therefore, we developed a method that is applicable to a wide variety of systems, is easy to implement, and rapidly provides readily interpretable outputs. By combining existing and widely known analytical tools from graph theory and population genetics with bootstrap and clustering analyses, this approach is elegant, intuitive, and also extremely malleable. Simulation analyses have demonstrated the robustness of our approach in various population differentiation and sampling conditions. Similarly, the empirical applications have clearly demonstrated its great potential for use in various landscapes and ecological systems.

### *Simulation study*

While, overall, the method is highly effective in determining the connectivity network among a set of populations, its effectiveness remains influenced by population differentiation. Modifying the migration rate or the connectivity pattern has a noticeably effect on the differentiation among populations and affects the capacity of our method to correctly retrieve the connectivity network used to produce the simulations. However, such an effect is unsurprising since, like other population genetic methods, the effectiveness of the method is derived at least in part from genetic distinctiveness among the populations (Waples 1998; Manel *et al.* 2005; Waples and Gaggiotti 2006). If populations share highly similar allelic composition, the minimum spanning trees should vary greatly across bootstraps, leading the method to produce a network where populations are linked to all of their neighbours, irrespective of the true migration channels. Such unresolved connectivity patterns (which are star-like or island-like) are likely to be common if strong migration occurs among sites or if populations share a recent common ancestry.

Accuracy of the method is also lower when the connectivity pattern among populations is more complex, irrespective of the level of population differentiation. These observations clearly suggest that the performance of the method is reduced when the connectivity pattern to recover is a network (i.e. sites connected through multiple paths). Indeed, if more than one alternative path is equally likely, only one path will be included at the time in the minimum-spanning tree inferred from each iteration. Therefore, the bootstrap values associated with each path will be reduced. A greater complexity of the connectivity structure results in a greater number of alternative paths available and a lower bootstrap value associated with each path; this in turn reduces the effectiveness of the method.

The simulation experiments also suggest that an unresolved network could occur with insufficient sampling either in terms of individuals per sites or number of loci. In these conditions, the genetic resolution of the data could affect the performance of the method. Therefore, users must be cautious when facing unresolved networks and make sure that they result from a true biological signal and not from weak genetic data. Unfortunately, the only means to discriminate between these two hypotheses is perhaps to increase the sampling effort either in adding some loci or individuals. Overall, the simulations suggest that the potential of the method is close to perfect with simple connectivity networks; in these cases, highly resolved networks are produced even with sampling conditions as low as 20 individuals per sites and 10 loci. However, if the connectivity structure is more complex, it may be preferable to increase the number of individuals per site, the number of loci, or both (Waples 1998).

One of the novelties of our method is the use of dispersal constraints. The addition of an appropriate constraint greatly improves the performance of the method. Indeed, it eliminates improbable links, thus increasing the effectiveness of the method. The advantage of such constraints is especially important if differentiation among populations is low, if sample size per site is limited, or if the true connectivity structure is complex. However, one must be cautious and keep in mind that the results of the method will be biased toward the constraint. If the imposed constraint is inappropriate for a given system, the resulting connectivity network will be obviously erroneous. To avoid such a problem, we strongly suggest

using either a dispersal hypothesis that is based on prior knowledge of the biological system; a spatial constraint (such as a Delaunay triangulation) that does not require particular knowledge of the biological system but instead takes into account the distance between the sample sites (recognizing that distance impacts dispersal, regardless of the species); or a combination of both.

#### *Illustrations of the proposed procedure with real data*

The analysis of the creek chub dataset underlined some inconsistencies in the permanent drainages. Interestingly, Boizard *et al.* (2009) reported the presence of intermittent streams connecting drainages; those intermittent streams corresponded with the three interconnections detected by the method. Boizard *et al.* also reported the presence of an impassable waterfall, which corresponded with the break detected by the method between the sites D9 and D10. The strong consistency between the overall portrait given by the method and the conclusions provided by Boizard *et al.* strongly suggest that the method is reliable despite the relatively small sample size (8 loci and 9-20 individuals per site). While the conclusions provided by Boizard *et al.* are based on a careful study of the landscape, the analysis conducted here has been performed without geographic *a priori* except spatial proximity. Therefore, the method appears to be a tool of choice if the landscape elements that impacted dispersal are unknown, or to get an independent confirmation of a hypothesis based on landscape.

On the other hand, some inconsistencies with the permanent river networks detected by the method have not been reported by Boizard *et al.* (2009). Potential explanations for those inconsistencies are numerous; they could correspond to floodplain areas, a historical configuration of the hydrological network, or a recent common history and incomplete lineage sorting or could simply be due to the low number of loci analysed. In these cases, further investigation of this part of the network would be necessary to determine the nature of these inconsistencies.

Lacoursière-Roussel *et al.* (2012) have concluded that ports are the principal location for the introduction of the invasive tunicate *B. schlosseri*, which arrive on commercial ships and most likely spread along the Nova Scotia coast as a result of recreational boating activities. The connectivity network presented here suggests similar conclusions but with a better resolution. Dispersal originates mainly from ports and subsequently follows boating routes, but then most marinas are grouped in highly localized connectivity networks. This latter result may complement two hypotheses that are not mutually exclusive: (i) natural dispersal may occur in nearby marinas in close or semi-close environments (bays, lakes) and (ii) invasion is increased by shorter recreational boating trips. Moreover, the fact that two distinct networks consisting mainly of marinas were defined is consistent with the hypothesis put forward by Lacoursière-Roussel *et al.* that occasional primary introductions can occur by recreational boats. This is especially relevant since these networks included the marinas that experience higher amounts of nautical activity, including international recreational boating (M6, M16, and M18; Lacoursière-Roussel *et al.*). Identification of the source location(s), pathway(s), and vector(s) is a prerequisite for the management of invasive species (Grosholz 2002). The results we obtained with the *B. schlosseri* example demonstrate that our method can be used to identify a complex and dynamic connectivity structure, even with little prior knowledge. The use of two distinct hypothetical networks, both based on the characteristics of this biological system, has helped eliminate many useless links. This increases the effectiveness of the method despite the low number of markers and the low global differentiation. As a result, we were able to confirm most of the conclusions by Lacoursière-Roussel *et al.*, but with a finer resolution and a straightforward visual output that facilitates the interpretation of the data in regards to both the landscape particularities and the boating dynamics of the region.

The conclusions provided by Lacoursière-Roussel *et al.* (2012) were mainly based on the observations that typically marina populations show high genetic similarity with the nearest port and lower genetic diversity than port populations. In addition to supporting the results obtained by Lacoursière-Roussel *et al.* the analysis performed here has the great advantage of highlighting source populations and identifying the dispersal links between ports and marinas.

### *Concluding remarks*

Our algorithm was developed to illustrate the connectivity structure linking populations and to facilitate its interpretation in terms of dispersal factors. By doing so, this algorithm may help to identify i) elements that facilitate or restrict dispersal and/or ii) hidden patterns in the genetic data that are inconsistent with prior hypotheses and, as a result, that reveal historical effects or unexpected relationships with missed or overlooked landscape elements. As such, this method can either contribute to refuting prior hypotheses or to formulating new ones. The reanalysis of empirical datasets done in this study demonstrates that our approach is advantageous in comparison to other currently available methods. In addition, our approach functions without requiring an excess of prior knowledge or the need for model parameterization, a prerequisite of many time-consuming and complex population genetics analyses associated with Bayesian inferences.

### **4.6 Author contributions**

Original idea: FC and BA; Conception of the algorithm: FC and PG; Computer programming: PG; Simulations programming: FC Redaction: FC, ALR, BA and PG.

## **CHAPITRE 5 : CONCLUSION GÉNÉRALE**

Les études présentées dans le cadre de cette thèse ont permis de mettre en évidence certains des effets de la dispersion historique sur la répartition actuelle de la biodiversité, au niveau inter et intra spécifique ; à grande et à petite échelle spatiale. De plus, une méthode a été proposée pour caractériser la connectivité d'un ensemble de populations, facilitant la comparaison entre la variabilité génétique et le paysage historique ou actuel. Il apparaît clair que les études cherchant à caractériser l'influence du paysage et des processus écologiques actuels doivent tenir compte des effets historiques potentiels et du cadre temporel réel affectant la biodiversité en général et la variabilité génétique en particulier.

### **5.1 L'impact de la dispersion historique à grande échelle spatiale**

En accord avec les précédentes études biogéographique effectuées dans le nord-est de l'Amérique du Nord, le premier volet de cette thèse a permis de mettre en évidence, les impacts des refuges glaciaires et de la dispersion post-glaciaire sur la répartition des espèces (*e.g.* Legendre and Legendre 1984 ; Black *et al.* 1986 ; Mandrak and Crossman 1992 ; Curry 2007). Par ailleurs, l'étude de la répartition de parasites généralistes des poissons a permis d'inférer un patron phylogéographique potentiellement très répandu chez ces derniers. En effet, des populations de poissons associées à des parasites différents sont vraisemblablement isolées les unes des autres. Cet isolement pourrait dater du Pléistocène, bien que cette dernière hypothèse ne puisse pas être validée sur la seule base des répartitions d'espèces parasites. L'existence d'un patron phylogéographique commun à plusieurs espèces a été répertoriée en Europe où les barrières des Alpes et des Pyrénées ont modulé la dispersion historique des espèces (Taberlet *et al.* 1998). L'existence d'un tel patron commun, dans l'est de l'Amérique du Nord, en l'absence de barrières aussi importantes serait donc inattendue et pourrait déboucher sur l'élaboration d'une stratégie de conservation commune à la faune ichtyologique. Des études phylogéographiques plus poussées sont évidemment requises pour confirmer l'existence d'un tel patron.

L'étude effectuée dans ce premier volet de thèse a également permis de mettre en lumière l'impact des changements d'hôtes sur l'émergence de la différentiation entre les taxons parasites (Funk et al. 1995). La répartition disjointe du Flotteur de Terre-Neuve est concomitante à l'utilisation d'hôtes distincts, ce qui a vraisemblablement contribué à promouvoir l'isolement et la divergence génétique entre le Flotteur de Terre-Neuve et son taxon sœur le Flotteur de l'Est. Le changement d'hôtes chez le Flotteur de Terre-Neuve a pu se produire avant, après ou de façon simultanée à l'adoption d'une répartition disjointe. C'est le paradoxe de l'œuf et de la poule, rejoué façon hôte-parasite.

## **5.2 L'impact de la dispersion historique à petite échelle spatiale**

Bien que mené à grande échelle spatiale, le premier volet de cette thèse permet néanmoins, d'apprécier l'impact de la dispersion historique à petite échelle spatiale. La présence de plusieurs vagues de colonisation sur le territoire et surtout la présence de zones de contact entre ces vagues permet de souligner une limite importante de l'hypothèse de travail adoptée dans cette thèse. S'il est vraisemblable que des populations proches géographiquement aient une histoire commune récente, cette hypothèse ne tient pas la route dans les zones de contact entre plusieurs vagues de colonisation où des populations très divergentes, ne partageant pas du tout la même histoire phylogéographique, se rencontrent. Pour être valable, l'interprétation de la variabilité génétique en terme actuel doit donc se faire entre des populations issues d'une même vague de colonisation. Heureusement, la présence de plusieurs vagues de colonisation dans un territoire donné peut facilement être détectée au préalable, à partir du niveau de divergence génétique entre les populations (Dyer *et al.* 2010).

Le second volet de cette thèse avait pour objectif de caractériser les effets de la dispersion historique sur la variabilité génétique de deux espèces de crapets, à petite échelle spatiale. Les différences observées entre les deux espèces au niveau de la variabilité génétique suggèrent un patron de colonisation diamétralement opposé. Le crapet de roche aurait colonisé les réseaux de drainages depuis l'aval occasionnant une perte d'allèle et une augmentation de la différentiation génétique à mesure qu'on progresse dans les réseaux de drainages. Au contraire, le crapet soleil

aurait colonisé les réseaux de drainages de l'amont vers l'aval, entraînant une forte similarité génétique et une forte diversité à la tête des réseaux. Cette forte similarité génétique est à l'origine d'un faux signal de connectivité actuelle à la tête des réseaux, démontrant du même coup l'importance des effets historiques comme source d'erreurs potentielles dans les études de génétique du paysage.

Les deux espèces échantillonnées ont comme point commun une variabilité génétique qui reflète davantage des processus historiques. La stabilité temporelle du paysage (i) ; l'espèce étudiée (iii) ; le type de marqueurs utilisé (iv) ; les méthodes utilisées (v) de même que l'échelle spatiale considérée (vi) ne semblent pas avoir eu un impact significatif sur l'importance relative des processus historiques par rapport aux processus actuels. En revanche, le type de perturbations (ii) semble avoir eu un impact important.

- i.) Dans cette étude, la stabilité des processus évolutifs ne semble pas reliée à la stabilité temporelle du paysage. En effet, même si le crapet de roche et le crapet soleil ont évolué dans le même paysage, la stabilité temporelle des processus évolutifs, pour ces deux espèces, ne semble pas concorder. La variabilité génétique observée chez le crapet de roche semble corrélée au paysage actuel, alors que la connectivité historique et la connectivité actuelle chez cette espèce semblent correspondre. À l'inverse, la connectivité historique et la connectivité actuelle du crapet soleil semblent complètement différentes. Aussi, la variabilité génétique de cette espèce est peu corrélée au paysage actuel.
  
- ii.) Des perturbations de différentes natures vont influencer plus ou moins rapidement la variabilité génétique. Les résultats obtenus dans ce deuxième volet permettent d'illustrer la difficulté de détecter l'isolement entre les populations sur la base de la différentiation génétique. La dérive génétique est un processus lent, particulièrement lorsque la taille efficace des populations est grande (Whitlock and McCauley 1999) ; la différentiation génétique entre des populations isolées peut alors s'échelonner sur des milliers de générations (Whitlock 1992). De plus, lorsque peu de loci sont utilisés, une importante incertitude reliée à l'estimation de la différentiation génétique peut empêcher

de détecter la différentiation récente entre des populations (Waples 1998). Ainsi, l'utilisation d'un faible nombre de loci a pu nous empêcher de détecter la différentiation génétique entre les populations de crapets soleil, en amont des réseaux de drainages.

- iii.) Le choix des espèces étudiées n'a pas eu d'impact sur l'importance relative des processus historiques. En génétique des populations le temps se calcule en nombre de générations, les deux espèces échantillonnées ayant un temps de génération identique ; cette variable n'a donc pas eu d'impact sur les différences observées entre les deux espèces. De plus, puisque ces deux espèces ont une capacité de nage similaire, il semble que les différences au niveau génétique soient attribuables à une histoire de colonisation différente.
- iv.) Malgré l'utilisation de marqueurs microsatellites, la variabilité génétique semble davantage refléter des processus historiques que des processus contemporains. Le taux de mutations et le polymorphisme élevé des microsatellites en font des marqueurs de choix pour détecter les fluctuations démographiques récentes (Beaumont 1999) ou la migration récente (Paetkau *et al.* 2004). Cependant, ce polymorphisme élevé associé à une taille efficace élevée fait également en sorte que l'isolement entre deux populations est difficilement décelable. Quoiqu'il en soit, peu importe le marqueur utilisé, il y aura un décalage entre la perturbation et son effet sur la variabilité génétique.
- v.) L'utilisation de différentes méthodes (directe ou indirecte) pour estimer la connectivité n'a pas permis de bien cerner la connectivité actuelle entre les populations. Les méthodes indirectes se basent sur la différentiation génétique et permettent d'estimer la migration sur plusieurs générations (*e.g.* Rousset 1997). Les méthodes directes sont basées sur le génotype multilocus des individus et sur des tests d'assignments, ces méthodes permettent la détection de la migration récente (*e.g.* Paetkau *et al.* 2004). Dans notre étude, ces deux types de méthodes ont été influencés par les processus historiques. D'un côté les méthodes indirectes supposent que le niveau de différentiation génétique entre les populations reflète l'équilibre entre l'homogénéisation des

populations par migration et leur différentiation par dérive génétique (Lowe et Allendorf 2010). Or, comme la dérive génétique est un processus lent, le niveau de différentiation génétique observée entre les populations de crapets reflète la généalogie des populations plutôt que la migration actuelle. D'un autre côté, l'efficacité de la méthode directe dépend de la différentiation génétique entre les populations (Waples 1998) ; alors que cette dernière est influencée par les processus historiques. Si les populations sont peu différencierées l'identification des migrants et des populations sources sera beaucoup plus difficile. De plus, la présence de populations non échantillonnées peut fortement biaiser les résultats obtenus (Manel *et al.* 2005). Ainsi, dans ce deuxième volet, l'identification des populations sources s'est avérée particulièrement problématique.

- vi.) Enfin, la petite échelle spatiale considérée ne constitue pas une garantie contre les processus historiques. Conjointement à d'autres études, cette étude permet de mettre en évidence l'impact des processus historiques, à petite échelle spatiale. À ce jour, plusieurs effets historiques ont été rapportés : le contact secondaire entre des vagues de colonisation (*e.g.* Fraser et Bernatchez 2005) ; l'influence de la configuration historique du paysage (*e.g.* Gagnon et Angers 2006) ; les effets fondateurs successifs sur les populations (*e.g.* Boizard *et al.* 2009) et l'impact de la généalogie des populations et du patron de colonisation (cette étude). Ces études n'invalident pas la prédiction que des populations proches devraient partager une histoire commune récente, en comparaison de populations éloignées. Cependant, elles remettent en question la possibilité d'interpréter la variabilité génétique sans tenir compte des effets historiques potentiels et ce peu importe l'échelle spatiale considérée.

### 5.3 Les méthodes développées dans cette thèse

#### *À grande échelle spatiale*

Il est plus facile de détecter les effets historiques à grande échelle spatiale. D'une part, les patrons observés ont souvent pris des milliers d'années pour se mettre en place, alors que les processus récents ont surtout un impact local. D'autre part, à cette échelle, les éléments du paysage significatifs ont laissé des traces tangibles. Néanmoins, puisque de multiples facteurs peuvent potentiellement expliquer la variabilité génétique, il peut-être difficile de séparer le grain de l'ivraie. Les modèles nuls constituent un outil statistique dédié à cette tâche.

Les modèles nuls sont basés sur les permutations aléatoires des données. Ils permettent notamment de générer les patrons attendus sous l'effet du hasard, en absence de mécanisme écologique (Gotelli and Graves 1996), alors que l'ajout d'une contrainte spatiale permet de modéliser et de tester l'effet d'un mécanisme écologique particulier (Peres-Neto *et al.* 2001). Historiquement, nombreux modèles nuls permettant de déterminer l'impact de la compétition et de la prédation sur les répartitions des espèces ont été proposés (*e.g.* Stone and Roberts 1990), alors qu'il existe un biais historique de la recherche envers les associations biologiques négatives (*e.g.*, Diamond, 1975). Dans le premier volet de cette thèse, l'utilisation de modèles nuls, basés sur la co-occurrence des espèces s'est avérée très efficace pour tester les associations biologiques positives et leur impact sur la répartition des espèces. La méthodologie proposée est fort prometteuse et pourrait être utilisée pour n'importe quelle espèce dont la dispersion dépend de d'autres espèces ; comme par exemple nombreuses espèces d'invertébrés ou de végétaux.

### *À petite échelle spatiale*

La méthode développée dans le troisième volet de cette thèse permet l'identification des connexions génétiques entre des populations, à partir d'un réseau hypothétique de dispersion (spatial ou autre). L'utilisation d'un tel réseau, comme point de départ, permet à la méthode d'être extrêmement flexible en fonction de l'espèce et des connaissances disponibles sur le système biologique étudié. De plus, il est possible de combiner différents réseaux de départ, permettant l'évaluation conjointe de plusieurs hypothèses de dispersion. Dans l'éventualité où des données historiques seraient disponibles, la méthode proposée pourrait permettre de comparer des hypothèses de dispersion basées sur le paysage historique et sur le paysage actuel. Dans l'éventualité où aucune donnée historique ne serait disponible, la méthode pourrait tout de même faciliter l'identification d'effets historiques potentiels en révélant les inconsistances entre le paysage actuel et la variabilité génétique.

L'évaluation des performances de la méthode sous des scénarios simulés nous permet de tirer plusieurs conclusions. La méthode s'est avérée très efficace pour retrouver le patron de connexions qui a servi à générer les données simulées, et ce pour l'ensemble des conditions testées. L'utilisation d'une contrainte de dispersion adéquate augmente systématiquement l'efficacité de la méthode en éliminant d'emblée plusieurs connexions improbables. Néanmoins, les performances de la méthode sont affectées par la différentiation génétique entre les populations ; la complexité du patron de connectivité sous-jacent, le nombre de loci et le nombre d'individus échantillonnés. La méthode a également été appliquée sur deux jeux de données empiriques très contrastés. En premier lieu, la connectivité a été estimée entre des populations de mulets à cornes (*Semotilus atromaculatus*), dans les réseaux de drainages de la Mastigouche. Pour ce premier jeu de données, la connectivité a été estimée à partir d'un réseau spatial, sans tenir compte au préalable du réseau hydrographique. Ceci a permis de révéler les correspondances et les inconsistances entre la variabilité génétique et le réseau hydrographique permanent. En second lieu, la connectivité a été estimée entre des populations d'Ascidies coloniales (*Botryllus schlosseri*) qui ont récemment envahi les zones littorales de la Nouvelle-Écosse. Pour ce second jeu de données, la connectivité a été estimée à partir d'un réseau spatial ; à partir d'un réseau représentant le trafic maritime et un réseau recoupant ces deux

hypothèses de dispersion. Ceci a permis de constater que, globalement, la variabilité génétique de cette espèce correspondait davantage au trafic maritime, qu'à la proximité spatiale entre les sites. L'analyse a également permis d'évaluer l'importance relative du trafic maritime commercial par rapport au trafic maritime de villégiature.

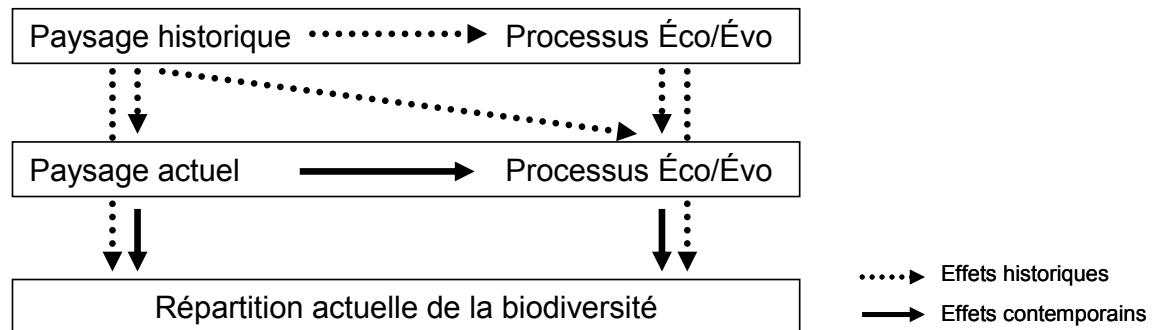
#### **5.4 Perspectives**

Les études de génétique du paysage se doivent d'adopter un cadre temporel qui englobe l'ensemble des processus (historiques et actuels) qui ont influencé la variabilité génétique (Figure 5.1). Plusieurs recommandations peuvent être formulées en ce sens :

- i) L'adoption, dans un premier temps, d'une large perspective spatiale devrait permettre l'identification des groupes de populations ne partageant pas une histoire commune récente et de certains effets historiques régionaux, comme par exemple des effets fondateurs. Les données génétiques recueillies à petite échelle spatiale pourraient, par la suite, être analysées en tenant compte des effets historiques soit en contrôlant pour la divergence historique (*e.g.* Dyer et al. 2010) ou soit en analysant séparément les groupes de populations présentant une histoire phylogéographique divergente.
- ii) Lorsque possible, il est pertinent de corrélérer la variabilité génétique au paysage actuel en contrôlant pour le paysage historique. Les données historiques, géologiques, géomorphologiques et paléontologiques sont de plus en plus accessibles, rendant du même coup ce type d'analyse plus fréquemment envisageable (*e.g.* Pavlacky *et al.* 2010).

- iii) Avant de conclure à un effet récent, il est important d'exclure les effets historiques potentiels. Ainsi, les processus historiques potentiels devraient servir à la construction d'hypothèses nulles, alors que la formulation de prédictions associées à ces hypothèses nulles devraient permettre de les tester (*e.g.* voir le deuxième volet de cette thèse). L'utilisation de logiciels de simulations peut aider à formuler des prédictions quant à l'impact de la colonisation sur la variabilité génétique (Le Corre 1997 ; Bohonak 2001). Ces prédictions doivent demeurer simples. En effet, en raison de la complexité inhérente des systèmes naturels, des prédictions très précises peuvent être associées à une très grande marge d'erreur.
- iv) L'utilisation de différentes méthodes d'analyses caractérisées par des cadres temporels différents demeure justifiée. Par exemple, les méthodes d'assignement permettent d'identifier les migrants de première génération (Piry *et al.* 2004), alors que les méthodes basées sur les fréquences alléliques permettent d'estimer le flux de gène sur plusieurs générations. Ainsi, une différence entre ces deux estimations pourrait indiquer une différence dans le flux génique historique et contemporain (*e.g.* Castric and Bernatchez 2004). Cependant, il faut faire attention de ne pas confondre une différence réelle entre le flux génique récent et historique et une erreur associée à l'une ou l'autre de ces méthodes (*e.g.* voir le deuxième volet de cette thèse). La congruence ou plutôt les inconsistances entre le flux génique récent et le paysage actuel devrait permettre d'identifier les erreurs d'assignement (*e.g.* voir le deuxième volet de cette thèse). De plus, l'utilisation d'une contrainte spatiale, tel que suggéré dans le troisième volet de cette thèse, devrait diminuer le nombre d'erreurs associées à l'estimation de la connectivité historique ou actuelle. Ainsi, la méthode présentée dans le troisième volet de cette thèse pourrait favoriser ce type de comparaison entre différents indices de distance génétique.

Il y a présentement en écologie une dichotomie importante entre les études à large et à petite échelle spatiale (Wiens and Donoghue 2004). Pourtant, les patrons observés à grande échelle spatiale sont l'amalgame des patrons locaux, alors que ces derniers sont à la fois influencés par des processus à différentes échelles spatiales et temporelles. Cette thèse aura contribué à faire le pont entre les études à large et à petite échelle spatiale, à caractère historique ou actuel. En démontrant l'impact de la dispersion historique à large et à petite échelle spatiale, cette thèse aura permis de révéler un biais potentiel important dans l'interprétation de la variabilité génétique, à petite échelle spatiale. Vraisemblablement, la variabilité génétique à petite échelle spatiale reflète davantage des processus récents que la variabilité génétique à grande échelle spatiale. Néanmoins, les processus historiques doivent être pris en compte, peu importe l'échelle spatiale considérée. Cette thèse avait pour objet l'étude de la variabilité génétique. Cependant, l'influence des processus historiques ne se cantonne pas à la variabilité génétique et d'autres descripteurs de la biodiversité (*e.g.* occurrence des espèces) sont susceptibles d'être affectés. Bref, en douze mots comme en mille : il faut tenir compte de l'évolution lorsqu'on étudie l'écologie.



**Figure 5.1 :** Effets historiques et actuels sur la répartition de la biodiversité.

## RÉFÉRENCES

- Aday, D. D., J. J. Parkos et D. H. Wahl (2009). Population and Community Ecology of Centrarchidae. *Dans : Centrarchid Fishes: Diversity, Biology and Conservation.* Édité par : Cooke S. J., Philipp D. P., Wiley-Blackwell, Oxford, Ch. 6.
- Adriaensen, F., J. P. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulinck et E. Matthysen (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and urban planning* **64**: 233-247.
- Alexander, H. M. et J. Antonovics (1988). Disease spread and population dynamics of anther-smut infection of *Silene alba* caused by the fungus *Ustilago violacea*. *Journal of Ecology* **76**: 91-104.
- Allendorf, F. W. (1983). Isolation, Gene Flow, and Genetic Differentiation Among Populations. *Genetics and conservation: a reference for managing wild animal and plant populations* **1**: 51-65.
- Amos, W. et J. Harwood (1998). Factors affecting levels of genetic diversity in natural populations. Philosophical Transactions of the Royal Society of London. Series B: *Biological Sciences* **353**: 177-186.
- Anderson, C. D., B. K. Epperson, M. J. Fortin, R. Holderegger, P. James, M.S. Rosenberg, K. T. Scribner et S. Spear (2010). Considering spatial and temporal scale in landscape genetic studies of gene flow. *Molecular Ecology* **19**(17): 3565-3575.
- Angers, B. et L. Bernatchez (1998). Combined use of the SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution* **15**:143–159.
- Angers, B., P. Magnan, M. Plante, et L. Bernatchez (1999). Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology* **8**(6): 1043-1053.
- Arens, P., T. van der Sluis, W. P. C. Van't westende, B. Vosman, C. C. Vos et M. J. M. Smulders (2007). Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. *Landscape Ecology* **22**(10): 1489-1500.
- Arnaud, J. F. (2003). Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology* **18**(3): 333-346.
- Avise, J. C. (2000). Phylogeography: the history and formation of species, Harvard Univ Pr.
- Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb et N. C. Saunders (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics* **18**: 489-522.

- Ayre, D., A. Davis, M. Billigham, T. Lorens et C. Styan (1997). Genetic evidence for contrasting patterns of dispersal in solitary and colonial ascidians. *Marine Biology* **130**: 51-61.
- Bassam, B. J., G. Caetano-Anolles et P. M. Gresshoff (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochemistry* **196**: 80–83.
- Beaumont, M. A. (1999). Detecting population expansion and decline using microsatellites. *Genetics* **153**(4): 2013-2029.
- Beaumont, M. A. et R. A. Nichols (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceeding of Royal Society of London Series B Biology* **263**: 1619-1626.
- Belkhir, K., P. Borsa et L. Chikki (1996–2004). Genetix Version 4.05.2, Logiciel sous Windows pour la Génétique des Populations. <http://kimura.univ-montp2.fr/genetix/>. Accédé le 9 novembre 2011.
- Belyea, L.R. et J. Lancaster (1999). Assembly rules within a contingent ecology. *Oikos* **86**(3): 402-416.
- Benzécri, J.P. (1973). Analyse des données. T. 2 : l'analyse des correspondances, Dunod, Paris.
- Bernatchez, L. et J. J. Dodson (1991). Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution* **45**(4): 1016-1035.
- Bernatchez, L. et M. Giroux (2000). Les poissons d'eau douce du Québec et leur répartition dans l'est du Canada, Broquet, Québec (Québec).
- Bernatchez, L. and C. C. Wilson (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* **7**(4): 431-452.
- Black, G. A., J. B. Dempson, et W. Bruce (1986). Distribution and postglacial dispersal of freshwater fishes of Labrador. *Canadian Journal of Zoology* **64**(1): 21-31.
- Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quarterly review of biology*: **21**-45.
- Bohonak, A. J., N. Davies, f. X. Villablanca et G. K. Roderick (2001). Invasion genetics of New World medflies: testing alternative colonization scenarios. *Biological Invasions* **3**(2): 103-111.
- Boizard, J., P. Magnan, et B. Angers (2009). Effects of dynamic landscape elements on fish dispersal: the example of creek chub (*Semotilus atromaculatus*). *Molecular Ecology* **18**(3): 430-441.

- Brinsmead, J. et M. Fox (2002). Morphological variation between lake and stream dwelling rock bass and pumpkinseed populations. *Journal of Fish Biology* **61**(6): 1619-1638.
- Broquet, T., N. Ray, E. Petit, J. M. Fryxell et F. Burel (2006). Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology* **21**(6): 877-889.
- Castric, V. and L. Bernatchez (2004). Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology* **13**(5): 1299-1312.
- Castric, V., F. Bonney et L. Bernatchez (2001). Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* **55**: 1016–1028.
- Cavalli-Sforza, L. L. et A. W. F. Edwards (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics* **19**: 233-257.
- Chamberland, J. M. (2011). Échantillonnage et modélisation de l'habitat des communautés de poissons de rivière des basses Laurentides. Mémoire de maîtrise (M.Sc), Université de Montréal, Montréal (Québec).
- Chardon, J. P., F. Adriaensen, et E. Matthysen (2003). Incorporating landscape elements into a connectivity measure: a case study for the Speckled wood butterfly (*Pararge aegeria* L.). *Landscape Ecology* **18**(6): 561-573.
- Clarke, A. H. 1981. The freshwater molluscs of Canada. National Museums of Canada, Ottawa (Ontario).
- Clarke, A.H. et A. M. Rick (1963). Supplementary records of Unionacea from Nova Scotia with a discussion of the diversity of *Anodonta fragilis* Lamarck. *Natural Museum Canada Bulletin* **199**: 15-27.
- Coulon, A., G. Guillot, J. F. Cosson, J. M. A. Angibault, S. Aulagnier, B. Cargnelutti, M. Galan et A. J. M. Hewison (2006). Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology* **15**: 1669-1679.
- Cox, C. B. et P. D. Moore (2010). Biogeography: an ecological and evolutionary approach, Wiley.
- Cracraft, J. (1973), Continental drift, paleoclimatology, and the evolution and biogeography of birds. *Journal of Zoology* **169**: 455–543.
- Crossman, E. J. et D. E. McAllister (1986). Zoogeography of freshwater fishes of the Laurentian Great Lakes, the St. Lawrence lowlands, Newfoundland and Labrador. *Dans : The zoogeography of North American freshwater fishes. Édité par : C. H. Hocutt, E.O., John Wiley & Sons, New York : 105-135.*

- Curry, A. R. (2007). Late glacial impacts on dispersal and colonization of Atlantic Canada and Maine by freshwater fishes. *Quaternary Research* **67**(2): 225-233.
- Cushman, S. A. et E. L. Landguth (2010). Spurious correlations and inference in landscape genetics. *Molecular Ecology* **19**: 3592-3602.
- Cyr, F., A. Paquet, A. L. Martel et B. Angers (2007). Cryptic lineages and hybridization in freshwater mussels of the genus *Pyganodon* (*Unionidae*) in northeastern North America. *Canadian Journal of Zoology* **85**: 1216-1227.
- Diamond, J. M. (1975). The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biological conservation* **7**(2): 129-146.
- Donovan, T. M., F. R. Thompson, J. Faaborg et J.R. Probst (1995). Reproductive success of migratory birds in habitat sources and sinks. *Conservation Biology* **9**: 1380-1395.
- Dyer, R. J., J. D. Nason, et R. C. Garrick (2010). Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**(17): 3746-3759.
- Dyke, A. S. et V. K. Prest (1987). Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie physique et Quaternaire* **41**(2): 237-263.
- Earl, D. A. et B. M. vonHoldt (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* DOI 10.1007/s12686-011-9548-7.
- Epps, C. W., J. D. Wehausen, V. C. Bleich, S. G. Torres et J. S. Brashares (2007). Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*. **44**: 714-724.
- Estoup, A. et B. Angers (1998). Microsatellites and Minisatellites for Molecular Ecology: Theoretical and empirical considerations. *Dans : Advances in Molecular Ecology. Édité par : G. R. Carvalho* : 55-79.
- Estoup, A., M. Beaumont, F. Sennedot, C. Moritz et J. M. Cornuet (2004). Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* **58**: 2021-2036.
- Evanno, G., S. Regnaut et J. Goudet (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611-2620.
- Excoffier, L., G. Laval, et S. Schneider (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online* **1**: 47-50.

- Excoffier, L., P. E. Smouse, et J. M. Quattro. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**(2): 479-491.
- Fraser, D. et L. Bernatchez (2005). Allopatric origins of sympatric brook charr populations: colonization history and admixture. *Molecular Ecology* **14**(5): 1497-1509.
- French III, J. R. P. (1988). Effect of submersed aquatic macrophytes on resource partitioning in yearling rock bass (*Ambloplites rupestris*) and Pumpkinseeds (*Lepomis gibbosus*) in lake St. Clair. *Journal of Great Lakes research* **14**: 291-300.
- Fuller, S. L. H. (1974). Clams and mussels (Mollusca: Bivalvia). *Dans : Pollution ecology of freshwater invertebrates. Édité par : C. W. Hart and S. L. H. Fuller*, Academic Press, New York. 389: 215-273.
- Funk, D. J., D. J. Futuyma, G. Ortí et A. Meyer (1995). A history of host associations and evolutionary diversification for Ophraella (Coleoptera: Chrysomelidae): new evidence from mitochondrial DNA. *Evolution* **49**(5): 1008-1017.
- Gabriel, K. R. et R. R. Sokal (1969). A new statistical approach to geographic variation analysis. *Systematic Zoology* **18**: 259–270.
- Gagnon, M. C. et B. Angers (2006). The determinant role of temporary proglacial drainages on the genetic structure of fishes. *Molecular Ecology* **15**(4): 1051-1065.
- Garant, D., S. E. Ford et A. P. Hendry (2007), The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functionnal Ecologoy* **21**: 434-443.
- Gibbs, M., M. Saastamoinen, A. Coulon et V. M. Stevens (2010). Organisms on the move: ecology and *Evolution* of dispersal. *Biology letters* **6**: 146-148.
- Girard, P. (2011). A robust statistical method to detect null alleles in Microsatellite and SNP Datasets in Both Panmictic and Inbred Populations. *Statistical Applications in Genetics and Molecular Biology* **10**(1): Article 9.
- Girard, P et B. Angers (2006). The impact of postglacial marine invasions on the genetic diversity of an obligate freshwater fish, the longnose dace (*Rhinichthys cataractae*), on the Québec peninsula. *Canadian Journal of Fisheries and Aquatic Sciences* **63**: 1429–1438.
- Gotelli, N. J. et G. R. Graves (1996). Null models in ecology, Smithsonian Institution Press.
- Goudet, J. (2001). Fstat, a program to estimate and test gene diversities and fixation indices (version 2.9.3). <http://www2.unil.ch/popgen/softwares/fstat.htm>. Accédé le 9 novembre 2011.
- Greenwood, P. J., P. H. Harvey et C. M. Perrins (1978). Inbreeding and dispersal in the great tit. *Nature* **271**: 52-54.

- Griffith, D.A. et P. R. Peres-Neto (2006). Spatial modelling in ecology: the flexibility of eigenfunction spatial analyses in exploiting relative location information. *Ecology* **87**: 2603-2613.
- Grosholz, E. (2002). Ecological and *Evolutionary* consequences of coastal invasions. *Trends in Ecology and Evolution* **17**: 22-27.
- Guillaume, F. et J. Rougemont (2006). Nemo: an *Evolutionary* and population genetics programming framework. *Bioinformatics* **22**: 2556-2557.
- Haag, W.R. et M. L. Jr. Warren (1998). Role of ecological factors and reproductive strategies in structuring freshwater mussel communities. *Canadian Journal of Fisheries and Aquatic Sciences* **55**: 297-306.
- Hamilton, M. (2011). Population genetics. John Wiley & Sons Editors.
- Hamilton, W. et R. M. May (1977). Dispersal in stable habitats. *Nature* **269**: 578-581.
- Hanski, I., J. Alho, et A. Moilanen (2000). Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology* **81**: 239-251.
- Hardy, O. J. (2008) Testing the spatial phylogenetic structure of local communities : statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology* **96**: 914–926.
- Hill, A. B. (1966). Principles of medical statistics, Oxford University Press, London.
- Hitchings, S. et T. Beebee (1998). Loss of genetic diversity and fitness in common toad (*Bufo bufo*) populations isolated by inimical habitat. *Journal of Evolutionary Biology* **11**(3): 269-283.
- Hoggarth, M. A. (1992). An examination of the glochidia-host relationships reported in the literature for North American species of Unionacea (Mollusca-Bivalvia). *Malacology Data Net* **3**: 1-30.
- Holderegger, R. et H. H. Wagner (2008). Landscape genetics. *BioScience* **58**(3): 199-207.
- Hutchison, D. W. et A. R. Templeton (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**: 1898-1914.
- Hudson, R. R., M. Kreitman et M. Aguadé (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153-159.
- Jaccard, P. (1912). The distribution of flora in the alpine zone. *New Phytologist* **11**: 37-50.

- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403-1405
- Jombart, T., S. Devillard et F. Balloux (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94
- Kat, P.W. (1983). Genetic and morphological divergence among nominal species of North American Anodonta. *Malacologia* **23**: 361-374.
- Kat, P.W. (1985). Historical evidence for fluctuation in levels of hybridization. *Evolution* **39**: 1164-1169.
- Kat, P. W. (1986). Hybridization in a Unionid faunal suture zone. *Malacologia* **27**: 107-125.
- Kat, P.W. et G. M. Davis (1984). Molecular genetics of peripheral populations of Nova Scotian Unionidae (Mollusca: Bivalvia). *Biological Journal of Linnean Society* **22**: 157-185.
- Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins et L. B. Jorde (1998). Signatures of population expansion in microsatellite repeat data. *Genetics* **148**(4): 1921-1930.
- Lacoursière-Roussel, A., D. G. Bock, M. E. Cristescu, F. Guichard, P. Girard, P. Legendre et C. W. McKindsey (2012). Disentangling invasion processes in a dynamic shipping-boating network. *Molecular Ecology* **21**: 4227-4241.
- Lacy, R. C. (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation biology* **1**: 143-158.
- Lapointe, F.-J. et T. Jr. Garland (2001). A generalized permutation model for the analysis of cross-species data. *Journal of Classification* **18**: 109-127.
- LaRue, M. A. et C. K. Nielsen (2008). Modelling potential dispersal corridors for cougars in midwestern North America using least-cost path methods. *Ecological modelling* **212**: 372-381.
- Leclerc, E. M., Y. Mailhot, M. Mingelbier et L. Bernatchez (2008). The landscape genetics of yellow perch (*Perca flavescens*) in a large fluvial ecosystem. *Molecular Ecology* **17**(7): 1702-1717.
- Le Corre, V., N. Machon, R. J. Petit et A. Kremer (1997). Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetical Research* **69**: 117-125.
- Legendre, P. et V. Legendre (1984). Postglacial dispersal of freshwater fishes in the Québec peninsula. *Canadian Journal of Fisheries and Aquatic Sciences* **41**(12): 1781-1802.

- Lowe, W. H. et F. W. Allendorf (2010). What can genetics tell us about population connectivity? *Molecular Ecology* **19**(15): 3038-3051.
- MacArthur, R. H. et E. O. Wilson (1967). The theory of island biogeography, Princeton Univ Pr.
- MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical, University of California, California.
- Mandrak, N. E. et E. Crossman (1992). Postglacial dispersal of freshwater fishes into Ontario. *Canadian Journal of Zoology* **70**(11): 2247-2259.
- Manel, S., O. E. Gaggiotti, et S. Waples (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution* **20**(3): 136-142.
- Manel, S., M. K. Schwartz, G. Luikart et P. Taberlet (2003). Landscape genetics: combining *Landscape Ecology* and population genetics. *Trends in Ecology & Evolution* **18**(4): 189-197.
- Manni, F., E. Guerard et E. Heyer (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* **76**: 173–190
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.
- Maynard-Smith, J. (1966). Sympatric speciation. *American Naturalist* **100**: 637-650.
- Mayr, E. (1963). Animal species and evolution, Harvard University Press.
- McElroy, T. C., K. L. Kandl, J. Garcia et J. L. Trexler (2003). Extinction colonization dynamics structure genetic variation of spotted sunfish (*Lepomis punctatus*) in the Florida Everglades. *Molecular Ecology* **12**: 355–368.
- McRae, B. H. et P. Beier (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceeding of Natural Academy of Sciences-Biology* **104**: 19885-19890.
- Monmonier, M. S. (1973). Maximum-difference barriers: an alternative numerical regionalization method. *Geography Anal* **5**: 245–261.
- Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* **51**(2): 238-254.
- Murphy, M. A., J. S. Evans, S. A. Cushman et A. Storfer (2008). Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* **31**: 685-697.

- Neff, B. D., P. Fu et M. R. Gross (1999). Microsatellite evolution in sunfish (Centrarchidae). *Canadian Journal of Fisheries and Aquatic Sciences* **56**: 1198-1205.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**(12): 3321-3323.
- Nei, M. (1987) Molecular Evolutionary Genetics, Columbia University Press, New York.
- Nei, M. et F. Tajima (1981). Genetic drift and estimation of effective population size. *Genetics* **98**: 625-640.
- Oksanen, J. F., G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens et H. Wagner (2010). vegan: Community Ecology Package. R package version 1.17-5. <http://CRAN.R-project.org/package=vegan>. Accédé le 11 novembre 2011.
- Olivieri, I., Y. Michalakis et P. H. Gouyon (1995). Metapopulation genetics and the Evolution of dispersal. *American Naturalist* **146**: 202-228.
- Paetkau, D., R. Slade, M. Burden et A. Estoup (2004). Genetic assignment methods for the direct, real time estimation of migration rate: a simulation based exploration of accuracy and power. *Molecular Ecology* **13**(1): 55-65.
- Pavlacky JR, D. C., A. W. Goldizen, P. J. Prentis, J. A. Nicholls et A. J. Lowe (2009). A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology* **18**(14): 2945-2960.
- Peres-Neto, P.R., et P. Legendre. (2010). Estimating and controlling for spatial autocorrelation in the study of ecological communities. *Global Ecology and Biogeography* **19**: 174-184.
- Peres-Neto, P. R., J. D. Olden, et D. A. Jackson (2001). Environmentally constrained null models: site suitability as occupancy criterion. *Oikos* **93**(1): 110-120.
- Pielou, E. C. (1991). After the ice age: the return of life to glaciated North America, University of Chicago Press.
- Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin et A. Estoup (2004). GeneClass2: a software for genetic assignment and first generation migrants detection. *Journal of Heredity* **95**: 536-539.
- Pray, L. (2008). DNA replication and causes of mutation. *Nature Education* **1**(1).
- Prim, R. C., 1957 Shortest connection networks and some generalizations. *Bell System Technical Journal* **36**: 1389-1401.

- Pritchard, J. K., M. Stephens et P. Donnelly (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg et M. W. Feldman (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceeding of Natural Academy of Sciences* **102**: 15942–15947.
- Raymond, M. et F. Rousset (1995). GenePop version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org/>. Accédé le 11 novembre 2011.
- Reusch, T. B. H., K. M. Wegner, et M. Kalbe (2001). Rapid genetic divergence in postglacial populations of threespine stickleback (*Gasterosteus aculeatus*): the role of habitat type, drainage and geographical proximity. *Molecular Ecology* **10**(10): 2435–2445.
- Rioux Paquette, S. et F. J. Lapointe (2009). A statistical procedure to assess the significance level of barriers to gene flow. *Journal of Genetics and Genomics* **36**: 685–693.
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**(4): 1219–1228.
- Sambrook, J., E. F. Fritsch et T. Maniatis (1989). Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, New York
- Schable, N. A., Fischer R. U. et T. C. Glenn (2002). Tetranucleotide microsatellite DNA loci from the dollar sunfish (*Lepomis marginatus*). *Molecular Ecology Notes* **2**: 509–511.
- Schneider, S., D. Roessli et L. Excoffier (2000). Arlequin: a software for population genetics data analysis. User Manual Version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.
- Schwartz, R. S. et B. May (2004). Characterization of microsatellite loci in Sacramento perch (*Archoplites interruptus*). *Molecular Ecology Notes* **4**: 694–697.
- Scott, W. B. et E. J. Crossman (1973). Freshwater fishes of Canada, *Bulletin of Fisheries Research Board Canada* No. **184**.
- Senay, C. (2009) Quelle contraintes structurent les communautés de poissons ? Influence de l'environnement, de la structure spatiale multiéchelles et des traits fonctionnels sur l'assemblage des communautés de poissons en rivière. Mémoire de maîtrise (M.Sc), Université du Québec à Montréal, Montréal (Québec).

- Sephton, D., B. Vercaemer, J. M. Nicolas et J. Keays (2011). Monitoring for invasive species tunicates in Nova Scotia, Canada (2006–2011). *Aquatic Invasions* **6**: 391-403.
- Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264-279.
- Sork, V. L. et L. Waits (2010). Contributions of landscape genetics—approaches, insights, and future potential. *Molecular Ecology* **19**(17): 3489-3495.
- Stanton, L.M. (2009). Population genetic structure and taxonomic evaluation of two closely related freshwater mussel species, the Eastern floater, *Pyganodon cataracta* and the Newfoundland floater, *P. fragilis* in Atlantic Canada. M.Sc thesis, Acadia University, Wolfville (Nouvelle-Écosse).
- Stone, L. et A. Roberts (1990). The checkerboard score and species distributions. *Oecologia* **85**(1): 74-79.
- Storfer, A., M. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani, E. Delmelle, L. Vierling et L. P. Waits (2006). Putting the ‘landscape’ in landscape genetics. *Heredity* **98**(3): 128-142.
- Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger et L. P. Waits (2010). Landscape genetics: where are we now? *Molecular Ecology* **19**(17): 3496-3514.
- Svane, I. et C. Young (1989). The ecology and behaviour of ascidian larvae. *Oceanography and Marine Biology* **27**: 45-90.
- Taberlet, P., L. Fumagalli, A. G. Wust-Saucy et J. F. Cosson (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**(4): 453-464.
- Takezaki, N. et M. Nei (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**: 389-399.
- Thorogood, C.J., F. J. Rumsey, S. A. Harris et S. J. Hiscock (2008). Host-driven divergence in the parasitic plant *Orobanche minor* Sm. (Orobanchaceae). *Molecular Ecology* **17**: 4289-4303.
- Trdan, R.J. et W. R. Hoeh (1982). Eurytopic host use by two congeneric species of freshwater mussel (*Pelecypoda*: *Unionidae*: *Anodonta*). *American Midland Naturalist* **108**: 381-388.
- Turgeon, J. et L. Bernatchez (2001). Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial races. *Molecular Ecology* **10**: 987-1001.
- Vandergast, A. G., A. J. Bohonak, D. B. Weissman et R. N. Fisher (2007). Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern

California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus). *Molecular Ecology* **16**(5): 977-992.

Van Name, W. (1945). The north and south American ascidians. *Bulletin of the American Museum of Natural History* **84**: 1-463.

Vaughn, C.C. et C. M. Taylor (2000). Macroecology of host-parasite relationship. *Ecography* **23**: 11-20.

Vignal, A., D. Milan, M. SanCristobal et A. Eggen (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**(3): 275-306.

Waples, R. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**(5): 438-450.

Waples, R. S. et O. Gaggiotti (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**: 1419-1439.

Watters, G. T. (1994). An annotated bibliography of the reproduction and propagation of the Unionoidea (primarily of North America). *Ohio Biological Survey Miscellaneous Contributions No. 1*.

Watters, G. T., T. Menker, S. Thomas et K. Kuehn (2005). Host identifications or confirmations. *Ellipsaria* **7**: 11-12.

Weir, B. S. et C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Whitlock, M. C. (1992). Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**: 608-615.

Whitlock, M. C. et D. E. McCauley (1999). Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm + 1)$ . *Heredity* **82**(2): 117-125.

Wiens, J. J. et M. J. Donoghue (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* **19**(12): 639-644.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips et E. Losos (1998). Quantifying threats to imperiled species in the United States. *BioScience* **48**: 607-615.

Wilson, C.C. et P. D. N. Hebert (1998). Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences* **55**: 1010-1024.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**: 395-420.

Zellmer, A. J. et L. L. Knowles (2009). Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology* **18**(17): 3593-3602.

Zuber, D. et A. Widmer (2000). Genetic evidence for host specificity in the hemi-parasitic *Viscum album* L. (Viscaceae). *Molecular Ecology* **9**: 1069-1073.

## **ANNEXES**

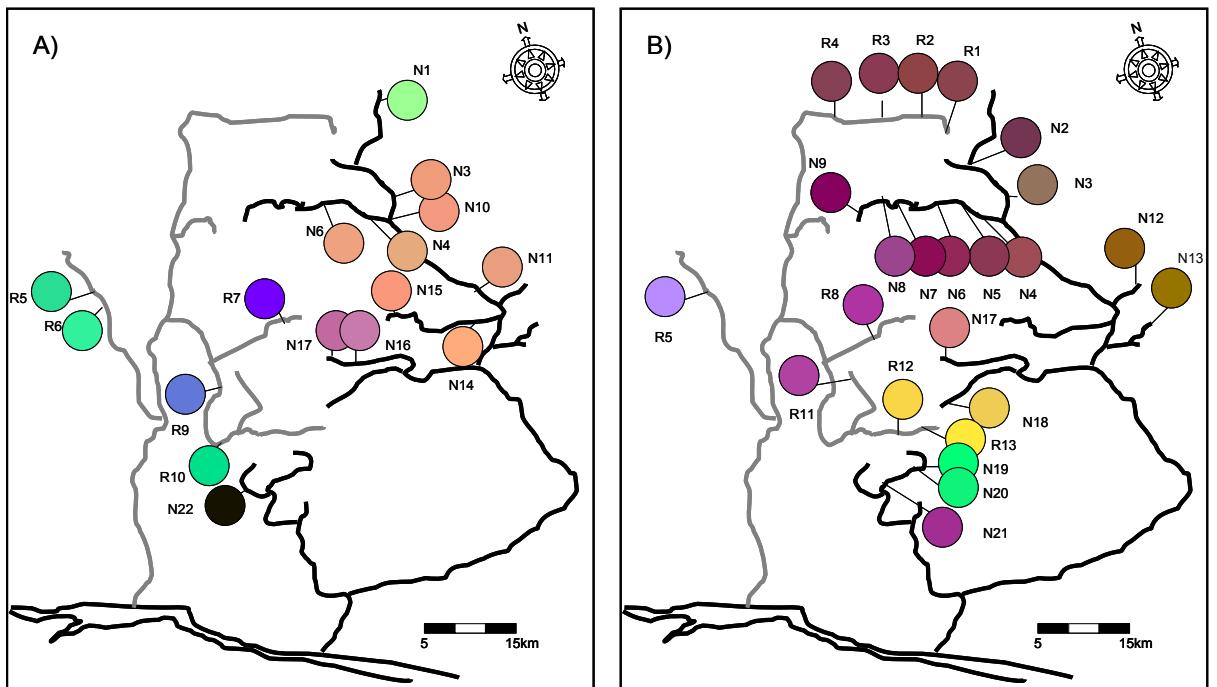
**Table A1:** Similarity values between floater species and all fish species tested in this study, according to the three similarity indices used ( $S_j$ ;  $S_h$ ;  $S_p$ ). N = number of sites within the range of each species out of the 66 sites examined; Refuge indicates origin of each species according to Mandrak and Crossman (1992); A: Atlantic refuge, M: Mississippi refuge, I: Introduced and ?: unknown post-glacial origin. Fish species in bold are previously reported hosts for floater mussels (Fuller, 1974; Trdan and Hoeh, 1982; Hoggarth, 1992; Watters et al., 2005). Significant associations after Bonferroni corrections are indicated by the star symbol.

Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	Sj	Sh	Sp	Sj	Sh	Sp	Sj	Sh	Sp
<b>Blacknose dace</b>	<b>63</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.37</b>	<b>0.37</b>	<b>1.00</b>	<b>0.60</b>	<b>0.60</b>	<b>1.00</b>	<b>0.09</b>	<b>0.10</b>	<b>0.67</b>
<b>Brook Stickleback</b>	<b>61</b>	<b>M</b>	<b>Gasterosteidae</b>	<b>0.38</b>	<b>0.38</b>	<b>1.00</b>	<b>0.57</b>	<b>0.59</b>	<b>0.95</b>	<b>0.09</b>	<b>0.10</b>	<b>0.67</b>
Brook Trout	66	AM	Salmonidae	0.35	0.35	1.00	0.58	0.58	1.00	0.14	0.14	1.00
Burbot	63	M	Gadidae	0.37	0.37	1.00	0.60	0.60	1.00	0.09	0.10	0.67
Fallfish	62	A	Cyprinidae	0.37	0.37	1.00	0.61	0.61	1.00	0.08	0.08	0.56
Lake chub	63	AM	Cyprinidae	0.37	0.37	1.00	0.60	0.60	1.00	0.09	0.10	0.67
Lake Whitefish	63	AM	Salmonidae	0.37	0.37	1.00	0.60	0.60	1.00	0.09	0.10	0.67
Logperch	60	M	Percidae	0.38	0.38	1.00	0.58	0.60	0.95	0.08	0.08	0.56
<b>Longnose sucker</b>	<b>63</b>	<b>AM</b>	<b>Catostomidae</b>	<b>0.37</b>	<b>0.37</b>	<b>1.00</b>	<b>0.60</b>	<b>0.60</b>	<b>1.00</b>	<b>0.09</b>	<b>0.10</b>	<b>0.67</b>
Northern Pike	60	M	Esocidae	0.38	0.38	1.00	0.58	0.60	0.95	0.08	0.08	0.56
Trout-Perch	52	M	Percopsidae	0.44	0.44	1.00	0.53	0.60	0.82	0.00	0.00	0.00
Walleye	54	AM	Percidae	0.43	0.43	1.00	0.56	0.61	0.87	0.00	0.00	0.00
<b>White Sucker</b>	<b>62</b>	<b>AM</b>	<b>Catostomidae</b>	<b>0.37</b>	<b>0.37</b>	<b>1.00</b>	<b>0.61</b>	<b>0.61</b>	<b>1.00</b>	<b>0.08</b>	<b>0.08</b>	<b>0.56</b>
<b>Yellow Perch</b>	<b>61</b>	<b>AM</b>	<b>Percidae</b>	<b>0.38</b>	<b>0.38</b>	<b>1.00</b>	<b>0.57</b>	<b>0.59</b>	<b>0.95</b>	<b>0.09</b>	<b>0.10</b>	<b>0.67</b>
<b>Creek chub</b>	<b>60</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.36</b>	<b>0.37</b>	<b>0.96</b>	<b>0.58</b>	<b>0.60</b>	<b>0.95</b>	<b>0.10</b>	<b>0.10</b>	<b>0.67</b>
Ninespine Stickleback	62	M	Gasterosteidae	0.35	0.35	0.96	0.56	0.58	0.95	0.15	0.15	1.00
<b>Pearl Dace</b>	<b>62</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.35</b>	<b>0.35</b>	<b>0.96</b>	<b>0.61</b>	<b>0.61</b>	<b>1.00</b>	<b>0.09</b>	<b>0.10</b>	<b>0.67</b>
Lake Sturgeon	36	?	Acipenseridae	0.51*	0.56*	0.87*	0.30	0.47	0.45	0.00	0.00	0.00

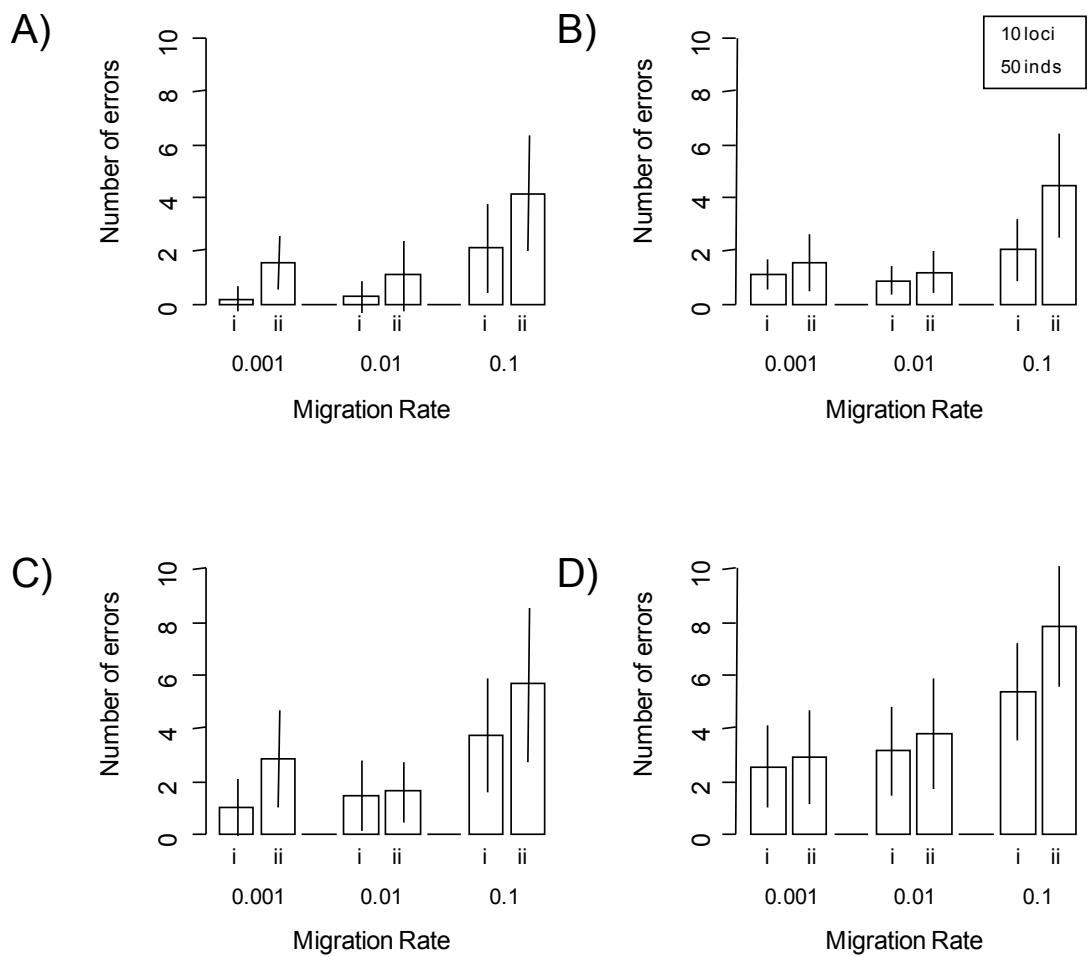
Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	Sj	Sh	Sp	Sj	Sh	Sp	Sj	Sh	Sp
Mottled Sculpin	46	AM	Cottidae	0.41	0.43	0.87	0.53	0.63	0.76	0.00	0.00	0.00
Slimy Sculpin	58	AM	Cottidae	0.33	0.34	0.87	0.63	0.64	0.97	0.10	0.10	0.67
<b>Longnose dace</b>	<b>35</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.49*</b>	<b>0.54*</b>	<b>0.83*</b>	<b>0.26</b>	<b>0.43</b>	<b>0.39</b>	<b>0.02</b>	<b>0.03</b>	<b>0.11</b>
Cisco	39	AM	Salmonidae	0.38	0.44	0.74	0.45	0.62	0.63	0.12	0.13	0.56
Sauger	20	M	Percidae	0.59*	0.80*	0.70*	0.07	0.20	0.11	0.00	0.00	0.00
<b>Blacknose Shiner</b>	<b>31</b>	<b>M</b>	<b>Cyprinidae</b>	<b>0.38</b>	<b>0.48</b>	<b>0.65</b>	<b>0.21</b>	<b>0.39</b>	<b>0.32</b>	<b>0.03</b>	<b>0.03</b>	<b>0.11</b>
Brown Bullhead	30	AM	Ictaluridae	0.39	0.50	0.65	0.21	0.40	0.32	0.00	0.00	0.00
Emerald Shiner	30	M	Cyprinidae	0.39	0.50	0.65	0.24	0.43	0.34	0.00	0.00	0.00
Fathead Minnow	26	M	Cyprinidae	0.44	0.58	0.65	0.14	0.31	0.21	0.00	0.00	0.00
Mimic Shiner	26	M	Cyprinidae	0.44	0.58	0.65	0.14	0.31	0.21	0.00	0.00	0.00
Spottail Shiner	30	AM	Cyprinidae	0.39	0.50	0.65	0.24	0.43	0.34	0.00	0.00	0.00
Mooneye	20	M	Hiodontidae	0.48*	0.70*	0.61*	0.12	0.30	0.16	0.00	0.00	0.00
<b>Iowa darter</b>	<b>28</b>	<b>M</b>	<b>Percidae</b>	<b>0.34</b>	<b>0.46</b>	<b>0.57</b>	<b>0.22</b>	<b>0.43</b>	<b>0.32</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Lake trout	44	AM	Salmonidae	0.24	0.30	0.57	0.58	0.68	0.79	0.08	0.09	0.44
<b>Common Shiner</b>	<b>32</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.22</b>	<b>0.31</b>	<b>0.43</b>	<b>0.37</b>	<b>0.59</b>	<b>0.50</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Deepwater Sculpin	17	AM	Cottidae	0.33	0.59	0.43	0.17	0.47	0.21	0.00	0.00	0.00
<b>Golden Shiner</b>	<b>36</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.20</b>	<b>0.28</b>	<b>0.43</b>	<b>0.37</b>	<b>0.56</b>	<b>0.53</b>	<b>0.10</b>	<b>0.11</b>	<b>0.44</b>
<b>Johny darter</b>	<b>23</b>	<b>AM</b>	<b>Percidae</b>	<b>0.28</b>	<b>0.43</b>	<b>0.43</b>	<b>0.20</b>	<b>0.43</b>	<b>0.26</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<b>Largemouth Bass</b>	<b>27</b>	<b>M</b>	<b>Centrarchidae</b>	<b>0.25</b>	<b>0.37</b>	<b>0.43</b>	<b>0.27</b>	<b>0.52</b>	<b>0.37</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<b>Rock Bass</b>	<b>28</b>	<b>M</b>	<b>Centrarchidae</b>	<b>0.24</b>	<b>0.36</b>	<b>0.43</b>	<b>0.29</b>	<b>0.54</b>	<b>0.39</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Shorthead redhorse	18	AM	Catostomidae	0.32	0.56	0.43	0.17	0.44	0.21	0.00	0.00	0.00
Goldeye	9	M	Hiodontidae	0.39*	1.00*	0.39*	0.00	0.00	0.00	0.00	0.00	0.00
Round whitefish	35	AM	Salmonidae	0.18	0.26	0.39	0.62*	0.80*	0.74*	0.05	0.06	0.22
<b>Freshwater Drum</b>	<b>14</b>	<b>M</b>	<b>Scianidae</b>	<b>0.28</b>	<b>0.57</b>	<b>0.35</b>	<b>0.13</b>	<b>0.43</b>	<b>0.16</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Finescale Dace	33	M	Cyprinidae	0.14	0.21	0.30	0.45	0.67	0.58	0.02	0.03	0.11
Northern RedbellyDace	33	M	Cyprinidae	0.14	0.21	0.30	0.45	0.67	0.58	0.02	0.03	0.11
<b>Pumpkinseed</b>	<b>24</b>	<b>AM</b>	<b>Centrarchidae</b>	<b>0.18</b>	<b>0.29</b>	<b>0.30</b>	<b>0.29</b>	<b>0.58</b>	<b>0.37</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Rainbow Smelt	45	A	Osmeridae	0.11	0.16	0.30	0.54	0.64	0.76	0.20	0.20	1.00

Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	Sj	Sh	Sp	Sj	Sh	Sp	Sj	Sh	Sp
American Eel	26	A	Anguillidae	0.14	0.23	0.26	0.23	0.46	0.32	0.35*	0.35*	1.00*
<b>Banded Killifish</b>	<b>26</b>	<b>AM</b>	<b>Cyprinodontidae</b>	<b>0.14</b>	<b>0.23</b>	<b>0.26</b>	<b>0.28</b>	<b>0.54</b>	<b>0.37</b>	<b>0.13</b>	<b>0.15</b>	<b>0.44</b>
<b>Bluntnose Minnow</b>	<b>15</b>	<b>AM</b>	<b>Cyprinidae</b>	<b>0.19</b>	<b>0.40</b>	<b>0.26</b>	<b>0.15</b>	<b>0.47</b>	<b>0.18</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<b>Carp</b>	<b>12</b>	<b>I</b>	<b>Cyprinidae</b>	<b>0.21</b>	<b>0.50</b>	<b>0.26</b>	<b>0.11</b>	<b>0.42</b>	<b>0.13</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Central Mudminnow	22	M	Umbridae	0.15	0.27	0.26	0.28	0.59	0.34	0.00	0.00	0.00
Fantail Darter	19	M	Percidae	0.17	0.32	0.26	0.24	0.58	0.29	0.00	0.00	0.00
Rosyface Shiner	24	M	Cyprinidae	0.15	0.25	0.26	0.32	0.63	0.39	0.00	0.00	0.00
Sand Shiner	13	M	Cyprinidae	0.20	0.46	0.26	0.13	0.46	0.16	0.00	0.00	0.00
Silver Redhorse	10	M	Catostomidae	0.22	0.60	0.26	0.09	0.40	0.11	0.00	0.00	0.00
Smallmouth Bass	18	AM	Centrarchidae	0.17	0.33	0.26	0.22	0.56	0.26	0.00	0.00	0.00
Spotfin Shiner	15	?	Cyprinidae	0.19	0.40	0.26	0.18	0.53	0.21	0.00	0.00	0.00
Stonecat	9	M	Ictaluridae	0.23	0.67	0.26	0.07	0.33	0.08	0.00	0.00	0.00
<b>Threespine sticklebac</b>	<b>41</b>	<b>A</b>	<b>Gasterosteidae</b>	<b>0.10</b>	<b>0.15</b>	<b>0.26</b>	<b>0.52</b>	<b>0.66</b>	<b>0.71</b>	<b>0.22</b>	<b>0.22</b>	<b>1.00</b>
Black Crappie	10	M	Centrarchidae	0.18	0.50	0.22	0.12	0.50	0.13	0.00	0.00	0.00
Brassy Minnow	18	M	Cyprinidae	0.14	0.28	0.22	0.24	0.61	0.29	0.00	0.00	0.00
Brown trout	15	I	Salmonidae	0.15	0.33	0.22	0.18	0.53	0.21	0.00	0.00	0.00
Channel Catfish	18	M	Ictaluridae	0.14	0.28	0.22	0.24	0.61	0.29	0.00	0.00	0.00
Cutlips Minnow	16	A	Cyprinidae	0.15	0.31	0.22	0.23	0.63	0.26	0.00	0.00	0.00
<b>Longnose gar</b>	<b>8</b>	<b>AM</b>	<b>Lepisosteidae</b>	<b>0.19</b>	<b>0.63</b>	<b>0.22</b>	<b>0.07</b>	<b>0.38</b>	<b>0.08</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Muskellunge	15	M	Esocidae	0.15	0.33	0.22	0.18	0.53	0.21	0.00	0.00	0.00
Silvery Lamprey	7	M	Petromyzontidae	0.20	0.71	0.22	0.05	0.29	0.05	0.00	0.00	0.00
Silvery Minnow	20	?	Cyprinidae	0.13	0.25	0.22	0.29	0.65	0.34	0.00	0.00	0.00
Tadpole Madtom	11	AM	Ictaluridae	0.17	0.45	0.22	0.09	0.36	0.11	0.00	0.00	0.00
American Lamprey	6	AM	Petromyzontidae	0.16	0.67	0.17	0.05	0.33	0.05	0.00	0.00	0.00
<b>Blackchin Shiner</b>	<b>5</b>	<b>M</b>	<b>Cyprinidae</b>	<b>0.17</b>	<b>0.80</b>	<b>0.17</b>	<b>0.02</b>	<b>0.20</b>	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Bowfin	6	M	Amiidae	0.16	0.67	0.17	0.05	0.33	0.05	0.00	0.00	0.00
Quillback	6	AM	Catostomidae	0.16	0.67	0.17	0.05	0.33	0.05	0.00	0.00	0.00
Rainbow Trout	15	I	Salmonidae	0.12	0.27	0.17	0.20	0.60	0.24	0.00	0.00	0.00

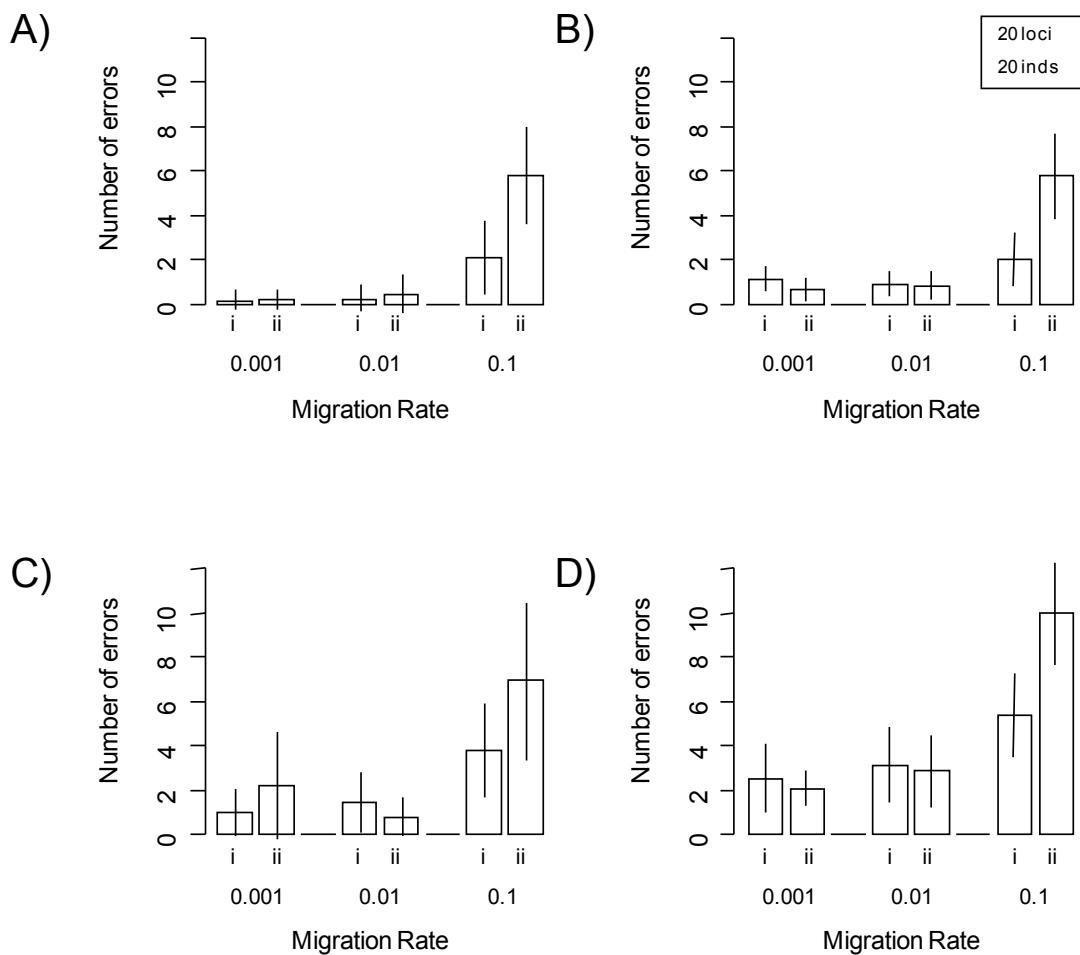
Fishes				Giant floater			Eeastern floater			Newfoundland floater		
Species name	N	Refuge	Family	Sj	Sh	Sp	Sj	Sh	Sp	Sj	Sh	Sp
<b>Bluegill</b>	<b>5</b>	<b>M</b>	<b>Centrarchidae</b>	<b>0.12</b>	<b>0.60</b>	<b>0.13</b>	<b>0.05</b>	<b>0.40</b>	<b>0.05</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Channel Darter	8	M	Percidae	0.11	0.38	0.13	0.10	0.50	0.11	0.00	0.00	0.00
Greater redhorse	3	M	Catostomidae	0.13	1.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
Tessellated darter	4	A	Percidae	0.13	0.75	0.13	0.02	0.25	0.03	0.00	0.00	0.00
Alewife	12	A	Clupeidae	0.06	0.17	0.09	0.09	0.33	0.11	0.62*	0.67*	0.89*
American Shad	12	A	Clupeidae	0.06	0.17	0.09	0.09	0.33	0.11	0.62*	0.67*	0.89*
Bridled Shiner	3	A	Cyprinidae	0.08	0.67	0.09	0.03	0.33	0.03	0.00	0.00	0.00
<b>Brook Silverside</b>	<b>2</b>	<b>M</b>	<b>Atherinidae</b>	<b>0.09</b>	<b>1.00</b>	<b>0.09</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Chain Pickerel	4	A	Esocidae	0.08	0.50	0.09	0.05	0.50	0.05	0.00	0.00	0.00
Copper Redhorse	2	?	Catostomidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Eastern Sand Darter	2	M	Percidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
<b>Gizzard Shad</b>	<b>4</b>	<b>M</b>	<b>Clupeidae</b>	<b>0.08</b>	<b>0.50</b>	<b>0.09</b>	<b>0.05</b>	<b>0.50</b>	<b>0.05</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<b>Longear sunfish</b>	<b>2</b>	<b>M</b>	<b>Centrarchidae</b>	<b>0.09</b>	<b>1.00</b>	<b>0.09</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Northern lamprey	2	M	Petromyzontidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Redfin Pickerel	4	AM	Esocidae	0.08	0.50	0.09	0.05	0.50	0.05	0.00	0.00	0.00
River redhorse	2	M	Catostomidae	0.09	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Sea Lamprey	9	A	Petromyzontidae	0.07	0.22	0.09	0.09	0.44	0.11	0.38	0.56	0.56
<b>White bass</b>	<b>4</b>	<b>M</b>	<b>Percichthyidae</b>	<b>0.08</b>	<b>0.50</b>	<b>0.09</b>	<b>0.05</b>	<b>0.50</b>	<b>0.05</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
White perch	3	A	Percichthyidae	0.08	0.67	0.09	0.03	0.33	0.03	0.00	0.00	0.00
Atlantic salmon	21	A	salmonidae	0.02	0.05	0.04	0.31	0.67	0.37	0.43*	0.43*	1.00*
Brook trout	16	?	salmonidae	0.03	0.06	0.04	0.26	0.69	0.29	0.32	0.38	0.67
<b>Yellow Bullhead</b>	<b>2</b>	<b>?</b>	<b>Ictaluridae</b>	<b>0.04</b>	<b>0.50</b>	<b>0.04</b>	<b>0.03</b>	<b>0.50</b>	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Atlantic sturgeon	10	A	Acipenseridae	0.00	0.00	0.00	0.09	0.40	0.11	0.73*	0.80*	0.89*
Blackspotted stickleback	3	?	Gasterosteidae	0.00	0.00	0.00	0.00	0.00	0.00	0.33	1.00	0.33
Fourspine stickleback	4	?	Gasterosteidae	0.00	0.00	0.00	0.02	0.25	0.03	0.44*	1.00*	0.44*
Tomcod	16	?	percopsidae	0.00	0.00	0.00	0.23	0.63	0.26	0.47*	0.50*	0.89*



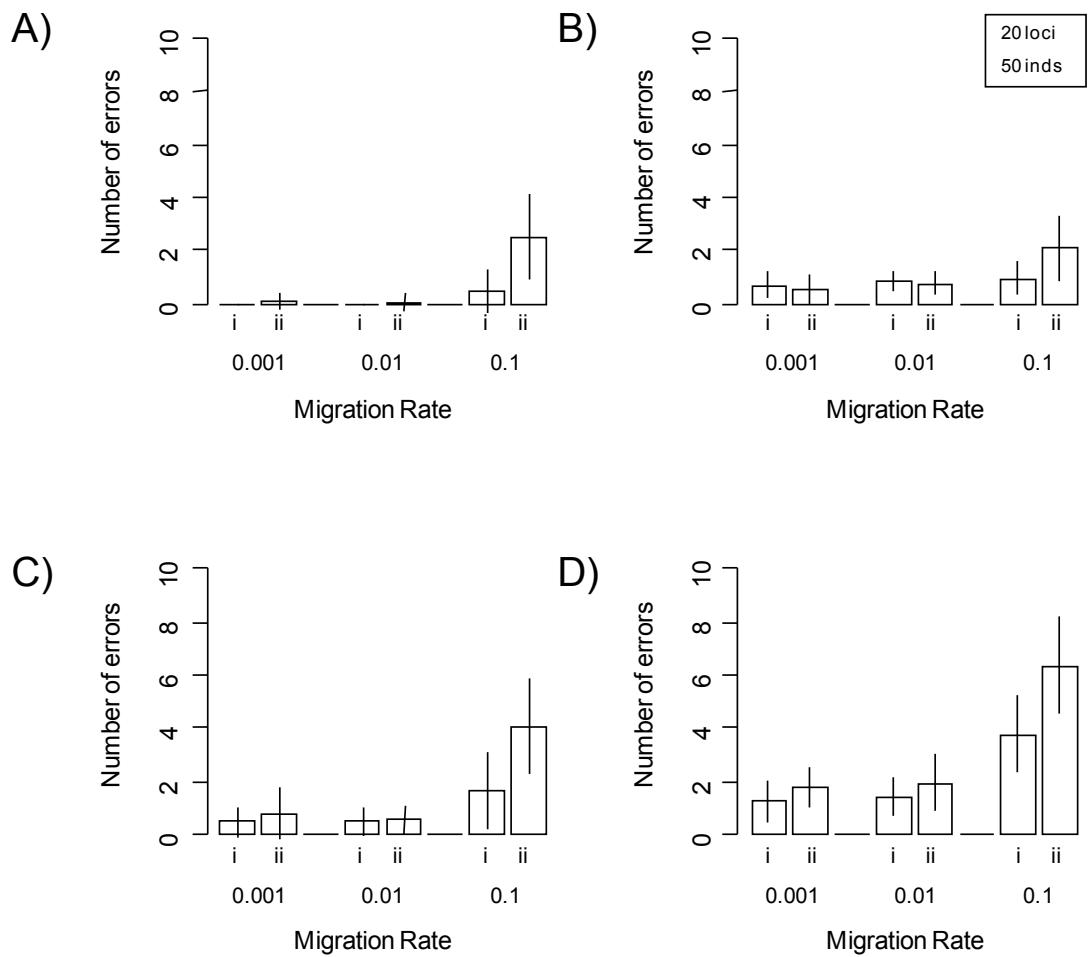
**Figure A1:** Representation of genetic differentiation among sampling sites. Correspondence analysis was also performed for rock bass (A) and sunfish (B). The first three axes of the analysis respectively explain 23.57%, 19.26%, and 15.79% of the variation among rock bass populations and 17.50 %, 16.38 %, and 10.63 % among sunfish populations. The position of each sampling site on each of the first three axes was coded as a color in the RGB color system (red for the first axis, green for the second axis, and blue for the third axis).



**Figure A2:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 10 loci and 50 individuals.



**Figure A3:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 20 loci and 20 individuals.



**Figure A4:** The effect of the spatial constraints on the effectiveness of the method. The absolute number of errors (either a link falsely identified or a link missed by the method) inferred from a genetic network with (i) and without (ii) a spatial constraint is compared. Results were gathered from three distinct migration rates (0.001; 0.01; 0.1), four connectivity patterns (A-D), and a sample size of 20 loci and 50 individuals.