UNIVERSITÉ DU QUÉBEC À RIMOUSKI

L'HISTOIRE DE DEUX NIVEAUX DE BIODIVERSITÉ DEMONTRÉE PAR LE CODE-BARRE D'ADN CHEZ LES CRUSTACÉS DE L'ATLANTIQUE DU NORD

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE EXTENSIONNÉ DE

L'UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PAR

ADRIANA E. RADULOVICI

MARS 2012

UNIVERSITÉ DU QUÉBEC À RIMOUSKI

A TALE OF TWO BIODIVERSITY LEVELS INFERRED FROM DNA BARCODING OF SELECTED NORTH ATLANTIC CRUSTACEANS

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

OF THE DOCTORATE OF BIOLOGY EXTENDED FROM

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BY

ADRIANA E. RADULOVICI

MARCH 2012

Life is like a box of chocolates:

You never know what you're gonna get!

(Forrest Gump)

FOREWORD

This thesis is the product of five years of research under the supervision of France Dufresne (UQAR) and Bernard Sainte-Marie (Maurice-Lamontagne Institute, DFO). My project began in 2006 as part of the Canadian Barcode of Life Network (BOLNET), led by Paul Hebert (University of Guelph, ON), and ended in 2011 as part of the Canadian Healthy Oceans Network (CHONe), led by Paul Snelgrove (Memorial University, NL). Main funding was provided by the Natural Sciences and Engineering Research Council (NSERC) through BOLNET, CHONe (F. Dufresne) and a Discovery Grant awarded to B. Sainte-Marie. Additional funding was provided by Fondation de l'UQAR (Bourse d'excellence en recherche, 2010-2011). Some of my work had the financial support of the Marine Barcode of Life (MarBOL) and the International Barcode of Life (iBOL) projects, Centre de recherche sur les milieux insulaires et maritimes (CERMIM), EnviroNorth, and The Crustacean Society. Additional sources have provided funding for presenting my research at various international conferences: UQAR, Amphipod Newsletter, and The International Biogeography Society.

The four chapters presented here correspond to published articles or manuscripts in preparation. While the work was collaborative between multiple co-authors, the first author was responsible for designing, writing, incorporating comments from co-authors and submitting manuscripts to journals. These chapters are linked by an introduction and general conclusions, while the appendices and references for each chapter are presented at the end of this thesis.

The life of a PhD student is full of "major discoveries" and a lot of frustration. In my case, this challenge would not have been met without the help of a few people deserving credits.

I am grateful to my supervisors for trusting my decisions even when these were not the wisest ones. Merci France for giving me complete freedom on this project and a highly flexible schedule. Merci Bernard for taking this challenge, for waking up early in the morning to write recommendation letters for me and for your unbeatable energy and good mood. I thank the rest of my committee for reading this thesis and giving very helpful feedback: Alison Derry (UQAM), Gary Saunders (UNB) and Pierre Blier (UQAR).

My regular visits to the "Barcoding Mecca" in Guelph helped me gather a very large dataset. I am grateful to the entire staff: technicians who showed me how to use fancy equipment, people helping with various issues (Greg, Nataly, Claudia, Jaclyn, Megan), the most helpful "liaison" person (Rick Turner), postdocs (the best French team: Rodolphe & David), profs (Alex, Sally and Mehrdad) and Paul Hebert who always took the time to listen to my ideas. I am grateful to Dirk Steinke (MarBOL coordinator) for helping with samples, protocols, contacts, analyses and, in general, facilitating barcoding of marine crustaceans. At UQAR/ISMER, I thank the Archambault group and Gesche Winkler for providing valuable samples, Robert Chabot for helping me start with amphipod collection and identification and Richard Cloutier for letting me use his imaging set-up. The staff in the Biology Department (Louise, Danny, Pauline) was very helpful when dealing with bureaucracy.

My close taxonomist collaborators are truly exceptional. Dave Wildish introduced me to the amazing world of "hopping" creatures. Thank you Dave for your enormous patience and for your relaxing British accent! Sara LeCroy always took the time to explain to me taxonomic issues. And Pierre Brunel shared some of his exceptional knowledge on marine amphipods during our meetings at UdeM. Fieldwork along the Eastern Canadian coast would not have been possible without the help of Traian Brad and Fred French. Thank you both for supporting me! Guglielmo Tita (CERMIM) helped with the Magdalen Islands project, grazie mille! In Mexico, Manuel Elías-Gutiérrez and Martha Valdéz (ECOSUR) and their students helped me reach Banco Chinchorro to collect amphipods, gracias!

This past five years included travels around the world to present my work at various conferences and to visit other labs. I thank all the people and organizations who provided bits and pieces of funding to make me a well-traveled grad student. I used desk space in a few labs during my travels: gracias José Guerra García (Sevilla), obrigada Filipe Costa (Braga), thanks Mike Hickerson (NY). Mersi Melania Cristescu (Windsor) for your positive energy, danke Sebastian Klaus (Frankfurt) for always offering your help (Carpathia is waiting!). Muchas gracias Marcos Pérez-Losada (Vairão) for making everything seem so easy ("*theses are never finished, they are just presented*") and helping on the last stretch! I am grateful to everyone who sent specimens (a very long list of nice people!), gave me ideas or feedback for data analysis and... paid for food and drinks or gave me a place to sleep! Also, all the free Wi-Fi areas in airports and cafés, greatly improved my work!

Thank you Sean & Marci for giving me a place to stay in Montreal, for widening my culinary knowledge with "horse-food" and salads and for showing me shoe-shops! Dr. Locke, thanks for all your good advices, from data analysis to proper English, and for all the amphipods you collected for me. Noroc!

In Rimouski: I thank all my friends who helped during my difficult moments: Marie & William, Aude, Aurélia, Irina, Magda, Arnaud, Samantha, Eva, JF & Dominique, the C103 labs (surtout Roland!).... Merci les Cadieux pour votre désir d'étudier le roumain et votre optimisme.

In Romania: my mentors (Vasile Cristea, Ioan Coroiu) with their encyclopedic knowledge taught me from the beginning that science has to be interdisciplinary. My friends back home (and wherever they migrated in the last decade) taught me that distance (spatial and temporal) has nothing to do with friendship.

The last and most importantly, I thank the best mother in the world (mine!), my sister and my big fat Romanian family for always supporting my crazy ideas that make me be away for so long. I hope you are all proud of me as I am proud of you and of being a multicultural hybrid from Transylvania and Banat!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BI	Bayesian inference
BOLD	Barcode of Life Data Systems
Bp	Base pairs
BSC	Biological species concept
COI	Cytochrome *c* oxidase subunit 1
CoML	Census of Marine Life
CBD	Convention of Biological Diversity
DFO	Department of Fisheries and Oceans
DNA	Deoxyribonucleic acid
ESL	Estuary of St. Lawrence
ESU	Evolutionary significant unit
GOM	Gulf of Mexico
GSL	Gulf of St. Lawrence
iBOL	International Barcode of Life
ITS	Internal Transcribed Spacer
K2P	Kimura 2-parameter
LGM	Last Glacial Maximum
MarBOL	Marine Barcode of Life
ML	Maximum-likelihood
MOTU	Molecular operational taxonomic unit

MPA	Marine protected area
Mt	Mitochondrial
MYA	Million years ago
nDNA	Nuclear DNA
NJ	Neighbor-joining
NWA	Northwest Atlantic
PCR	Polymerase chain reaction
PSC	Phylogenetic species concept
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SEM	Scanning electron microscopy
SDM	Spatial distribution modeling
UNEP	United Nations Environment Programme

RÉSUMÉ

La biodiversité est la variété de la vie et elle peut être étudiée à différents niveaux (génétique, espèces, écosystèmes) et à différents échelles (spatiale et temporelle). Les dernières décennies ont montré que la biodiversité marine avait été gravement sous-estimée. Afin d'étudier les caractéristiques de la grande diversité des espèces marines et les processus sous-jacents de l'évolution de ces dernières, il est évident et nécessaire de connaître les espèces. Nous sommes aujourd'hui confrontés aux taux les plus élevés d'extinction depuis la constitution de la société humaine («crise de la biodiversité») et seule une fraction d'espèces a été officiellement décrite (1,9 millions sur 11 millions), en raison, entre autres, d'une pénurie de taxonomistes formés et disponibles pour cet immense travail. Tous ces facteurs ont conduit à la proposition d'outils moléculaires pour permettre et faciliter l'identification des espèces et notamment le barcode moléculaire (le code-barres d'ADN). Il s'agit de séquencer un fragment d'ADN du gène mitochondrial cytochrome *c* oxydase 1 (COI) qui constitue alors un outil rapide, précis et rentable pour identifier les espèces. Ainsi, chaque espèce peut être définie par une étiquette d'identification unique et permanente qui ne sera pas changée par une éventuelle modification taxonomique. Outre l'attribution d'échantillons inconnus à des espèces identifiées a priori, les données fournies par le code-barres d'ADN seront très utiles pour des études phylogéographiques comparatives entre taxons multiples, pour clarifier les relations phylogénétiques à différents niveaux taxonomiques et pour élaborer des patrons évolutifs et de spéciation entre les groupes d'organismes.

Le Chapitre 1 présente une mise en contexte du code-barres d'ADN par une revue des études qui ont été publiées sur le sujet, notamment en ce qui concerne l'identification des espèces marines.

Le Chapitre 2 élabore une bibliothèque pour les crustacés marins de l'estuaire et du golfe du St. Laurent. Toutes les données (taxonomie, informations sur l'échantillonnage, images, séquences d'ADN et chromatogrammes), sont stockées en ligne dans le Barcode of Life Data Systems (BOLD) et sont disponibles pour un usage général. Les spécimens utilisés sont conservés comme 'vouchers' dans des institutions publiques pour des vérifications futures. Les résultats ont montré la présence d'un amphipode invasif dans l'estuaire (mentionné précédemment dans les Grands Lacs et à Montréal, avec des effets sur la faune indigène d'amphipodes), et l'existence d'espèces cryptiques potentielles chez les amphipodes, mysidacés et décapodes.

Le Chapitre 3 est axé sur l'utilisation des séquences COI fournies par le code-barres d'ADN comme un outil complémentaire pour la taxonomie et la phylogénie des amphipodes de la famille Talitridae dans l'Atlantique du Nord. En effet, la distribution et la diversité actuelle des espèces est le résultat de processus d'évolution et

d'interaction avec l'environnement à l'échelle d'une région géographique. Les études phylogénétiques permettent d'appréhender cette problématique en élaborant des scenarios évolutifs des relations entre taxons. Les résultats montrent l'existence d'espèces cryptiques chez trois espèces morphologiques. En outre, les genres anciens ne semblent pas être monophylétiques, suggérant la nécessité d'une révision taxonomique chez cette famille.

Le Chapitre 4 aborde le thème de la diversité génétique qui permet la persistance des populations et des espèces dans le temps en permettant une adaptation continue aux changements environnementaux. À de grandes échelles spatiales, la diversité intraspécifique peut être structurée en généalogies en fonction de la géographie, définissant alors des patrons phylogéographiques, qui peuvent coïncider ou pas avec les divisions biogéographiques. Les séquences COI générées par le code-barres d'ADN ont été utilisées pour déduire des patrons phylogéographiques chez une espèce d'amphipode avec une distribution amphi-Atlantique, *Gammarus oceanicus*. Cette espèce est très abondante et représente une partie importante des communautés intertidales et des réseaux trophiques côtiers. Les résultats ont montré une division profonde au sein de cette espèce avec deux groupes ayant une séparation latitudinale (la région tempérée du Canada Atlantique *versus* la région subarctique du Baie d'Hudson et l'Europe), indiquant la présence des deux espèces cryptiques potentielles.

L'ensemble de ces travaux de recherche a montré que la biodiversité marine, notamment chez les crustacés marins de l'Atlantique du Nord, était sous-estimée. Des espèces cryptiques potentielles ont été trouvées chez huit espèces morphologiques, sachant que seulement les espèces les plus communes ont été échantillonnées pour cette étude. Le taux de diversité augmentera certainement avec l'ajout d'échantillonnes de différents taxons, de divers types d'habitat et de régions marines distinctes.

Mots-clés : biodiversité marine; code-barres d'ADN; identification des espèces; Crustacea; diversité cryptique; Atlantique du Nord

ABSTRACT

Biodiversity is the variety of life and can be studied at different levels (genetic, species, ecosystems) and at different scales (spatial and temporal). The past decades have shown that marine biodiversity has been severely underestimated. To study the characteristics of the great diversity of marine species and the underlying processes of formation and maintenance of marine biodiversity, it is obvious and necessary to know what lives out there. We are now faced with the highest extinction rates since the formation of the human society ("biodiversity crisis") and only a fraction of species was formally described (1.9 million of 11 million), because of a shortage of trained taxonomists available for this immense work, among other things. All these factors have led to the proposal of molecular tools to enable and facilitate the identification of species including DNA barcoding. This method uses a DNA fragment of the mitochondrial gene cytochrome *c* oxidase subunit 1 (COI) as a fast, accurate and cost effective tool to identify species. Thus, each species can be defined by a unique identification tag that will not be changed during taxonomic revisions. In addition to the assignment of unknown specimens to species identified a priori by taxonomists, data generated through barcoding studies will be very useful for comparative phylogeographic studies of multiple taxa, phylogenetic studies at different taxonomic levels and for studies on evolutionary patterns between groups of organisms.

Chapter 1 provides some background on DNA barcoding with a review on studies that were published on the subject, especially those focusing on the identification of marine species.

Chapter 2 develops a reference library for marine crustaceans from the Estuary and the Gulf of St. Lawrence. All data (taxonomy, collection information, images, DNA sequences and chromatograms) are stored online in the Barcode of Life Data Systems (BOLD) and are available for general use. Specimens used for barcoding are kept as “vouchers” in public institutions for future use. The results showed the presence of an invasive amphipod in the estuary (mentioned previously in the Great Lakes and near Montreal, with impact on the native fauna of amphipods), and the existence of potential cryptic species in amphipods, mysids and decapods.

Chapter 3 focuses on the use of COI sequences provided through DNA barcoding as a complementary tool for taxonomy and phylogeny of the amphipod family Talitridae in the North Atlantic. The current distribution and diversity of species is the result of evolutionary processes and interaction with the environment across a geographic region. Phylogenetic studies can investigate this issue by developing evolutionary scenarios on the relationships between taxa. The results show the existence of cryptic species in three morphological species. In addition, older genera do not

cryptic species in three morphological species. In addition, older genera do not appear to be monophyletic, suggesting the need for taxonomic revisions in this family.

Chapter 4 addresses the issue of genetic diversity which enables the persistence of populations and species over time, allowing continuous adaptation to environmental changes. At large spatial scales, diversity within species can be structured in genealogies according to geography, thus defining phylogeographic patterns, which may coincide or not with biogeographic divisions. COI sequences generated by DNA barcoding were used to infer phylogeographic patterns in an amphipod species with amphi-Atlantic distribution, *Gammarus oceanicus*. This species is very abundant and an important part of the intertidal communities and coastal food webs. The results showed a deep division within this species with two divergent groups corresponding to a latitudinal segregation (temperate region of Atlantic Canada versus the subarctic Hudson Bay and Europe), indicating the presence of two potential cryptic species.

This research showed that marine biodiversity, as seen in marine crustaceans from North Atlantic, was underestimated. Potential cryptic species were found in eight morphological species, knowing that only the most common species were sampled for this study. The level of diversity will certainly increase with the addition of different taxa, different types of habitat and distinct marine regions.

Keywords: marine biodiversity; DNA barcoding; species identification; Crustacea; cryptic species; North Atlantic

GENERAL INTRODUCTION

"In all cultures, taxonomic classification means survival. The beginning of wisdom, as the Chinese say, is calling things by their right names"

E.O. Wilson, *The Diversity of Life*

What's in the "biodiversity" name?

"Biodiversity" is a heavily used term in science and very popular with the general public (>48 million results on Google, March 2012). It is a shorthand form of "biological diversity" and it was defined as "the variability among living organisms from all sources including, *inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are a part; this includes diversity within species and of ecosystems" (Convention on Biological Diversity, CBD, 1992) or, in simple words, "the variety of life". Conventionally, three levels of biodiversity are recognized (genetic, species, ecosystems) but only one is usually investigated, namely the species level. Reasons for this trend probably include "ease of reach" of species diversity (e.g., observations in nature or experiments, relatively cheap to conduct) and the "ease of understanding" its more intuitive numbers (of species, of individuals etc.). Generally, geographic areas with many species are considered more interesting for conservation than species-poor areas. The species level is, however, more than an easy-to-grasp category due to its practical value: it is a check-list of extant species, a baseline against which to compare future changes towards biodiversity gain or, more likely, biodiversity loss.

All biodiversity levels are interconnected and impacts on any level will trigger responses from the other biodiversity components. For example, genetic variation,

considered to be related to population size (Frankham, 1996; but see Bazin, Glémin and Galtier, 2006), can act as a buffer against environmental changes (natural and anthropogenic), allowing the persistence of populations and species in time. Theory predicts correlations between genetic and species diversity that are either positive (due to environmental heterogeneity and/or time since disturbance) or negative (species richness associated with reduced niche breadth per species, allowing fewer genotypes to coexist) (Lankau, 2011). Ecosystem functioning (e.g., pelagic ecosystem processes) is related to biodiversity in genes, species and functional groups (e.g., richness of producers and consumers) (Duffy and Stachowicz, 2006). Experiments have shown that intraspecific genetic diversity of foundation species (i.e., dominant primary producers) may influence the community structure (i.e., species richness and abundance at higher trophic levels), ecosystem processes and resistance to disturbance (Hughes and Stachowicz, 2004; Reusch *et al.*, 2005; Crutsinger *et al.*, 2006), although the spatial scale has to be considered as well (Crutsinger, Cadotte and Sanders, 2009).

Biodiversity and ecosystem functioning are directly connected to human well-being through ecosystem services, thus the need to protect biodiversity for the existence of the human society (Figure 1). Humans (*Homo sapiens*) should not be considered an external factor but an intrinsic part of biodiversity as we are one species among the ~8.7 million estimated to exist (Mora *et al.*, 2011). Human activities have large impacts on all levels of global diversity but they are also variable across cultures (although differences between cultures might decrease due to globalization). In this context, cultural diversity can be considered as an important factor in biodiversity *sensu stricto*, and even as another level of biodiversity rather than a research subject for a separate field (anthropology). As an index for this diversity, ~7,000 languages are spoken worldwide (Davis, 2010), mostly by small groups of indigenous people with livelihoods directly depending on natural resources, thus involved in shaping local biodiversity and continuously evolving with their environment.

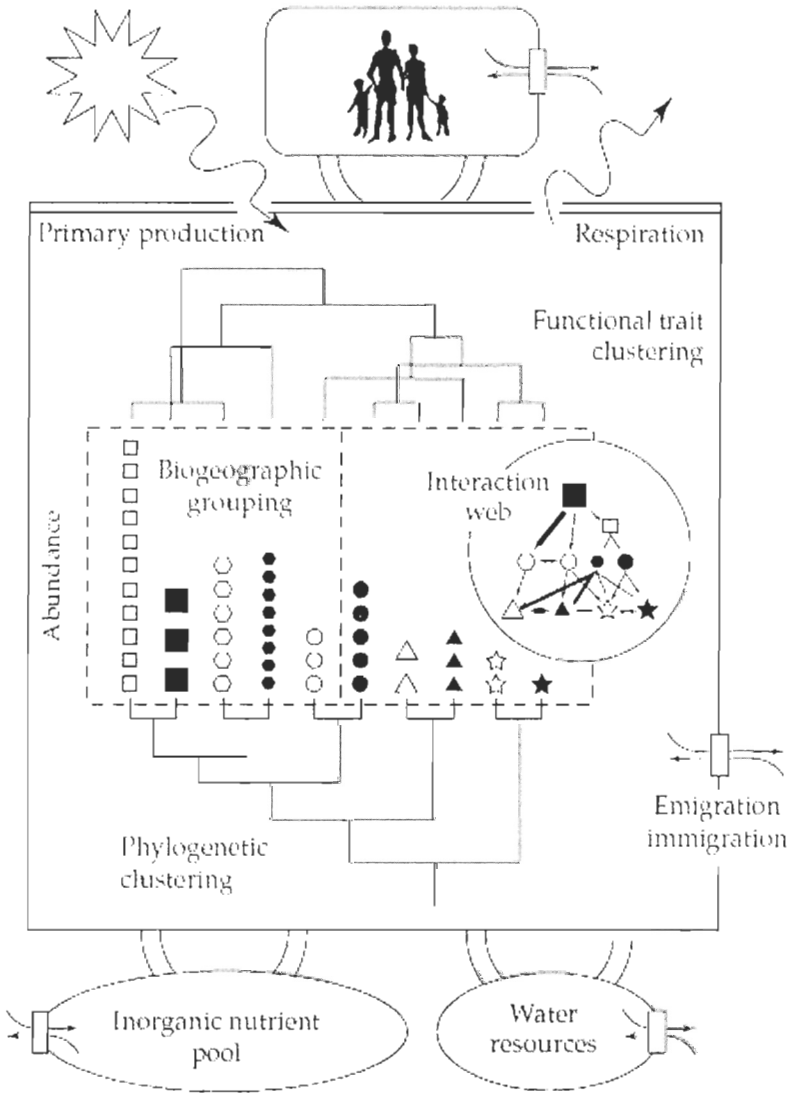


Figure 1 Relationship between biodiversity, ecosystem functioning and human well-being. Species are represented in the center by black and white objects with various shapes and sizes. (Source: Naeem *et al.*, 2009)

Marine biodiversity: a few characteristics

Marine biodiversity has long been underestimated due to the general belief that oceans are homogeneous with limited habitat diversity (compared to land), hence limited species diversity and infrequent speciation events. About 250,000 marine eukaryote species have been described (First Census of Marine Life, CoML, 2010). The estimated numbers range, however, from 500,000 (Gray, 1997) to over 10 million (Grassle and Maciolek, 1992) with recent estimates reaching 2.2 million species, which means that ~90% of marine species are still to be discovered (Mora *et al.*, 2011). At higher taxonomic levels, marine diversity is much higher than the terrestrial counterpart (35 marine phyla versus 11 terrestrial phyla) due to the fact that life appeared in the sea, and hence has had a longer time for evolutionary diversification (Gray, 1997). The differences in species numbers between land and sea are believed to be quite recent (~110 million years ago, MYA), coinciding with an increase in productivity on land, and explained by: i) higher primary productivity on land, on average (although marine kelp forests have higher productivity per surface unit); ii) narrower specialization of terrestrial species; iii) more effective barriers to dispersal on land; iv) greater 3D complexity and niche availability on land; and v) greater viability of low-density populations (and consequently rare species) (Vermeij and Grosberg, 2010).

Biodiversity has a heterogeneous distribution on the planet, with some areas being more diverse than others (Gaston, 2000). Some marine groups (e.g., bivalves) show a latitudinal diversity gradient with tropics as centers of origin and diversification and poles as species-poor areas (Gaston, 2000 and references therein; Valentine and Jablonski, 2010). The Arctic regions are less diverse compared to the Antarctic regions due to historical differences such as age and glacial history (Gray, 1997). Other patterns include an increase of species richness from shallow-waters to the deep-sea in soft sediments, higher diversity in the benthic

compared to the pelagic realm, a diversity peak for coastal species in the western Pacific and for oceanic groups at mid-latitudes (Gray, 1997; Tittensor *et al.*, 2010).

Both species and genes (within species) have a heterogeneous distribution in nature, thus the importance of geography for biodiversity. Species-rich and endemic species-rich areas are considered hotspots of biodiversity, in need of conservation. Likewise, genetically diverse or evolutionary distinct populations (evolutionary significant units, ESU) are hotspots of intraspecific diversity and should be considered as such in conservation plans (Crandall *et al.*, 2000; Rauch and Bar-Yam, 2004).

Current status: threats and mitigation measures

Many species are currently going extinct (or are predicted to do so) leading scientists to declare a state of emergency, or "biodiversity crisis". The world is dynamic with species being formed and lost through natural processes. Large extinction events have occurred throughout Earth's history (five mass extinctions between 440 MYA – 65 MYA; Futuyma, 1998). The genus *Homo* has also caused marked changes in ecosystems and species extinctions since the formation of primitive human communities. There have been a number of recent calls for defining a "sixth mass extinction" in the current era, in light of extinction rates between 100-1,000 times higher than pre-human values and estimated future rates 10 times the current rates (Figure 2) (Pimm *et al.*, 1995; Pimm and Raven, 2000; Millennium Ecosystem Assessment, MA, 2005). In the sea, humans have so far directly caused the global extinction of more than 20 species including mammals, seabirds, fishes, invertebrates and algae, along with many more local or regional extinctions (Sala and Knowlton, 2006 and references therein). As species do not live in isolation but in interactions, the extinction of one species triggers effects at other biodiversity levels. However, the functional role might be more important than the number of species going extinct *per se* (O'Connor and Crowe, 2005).

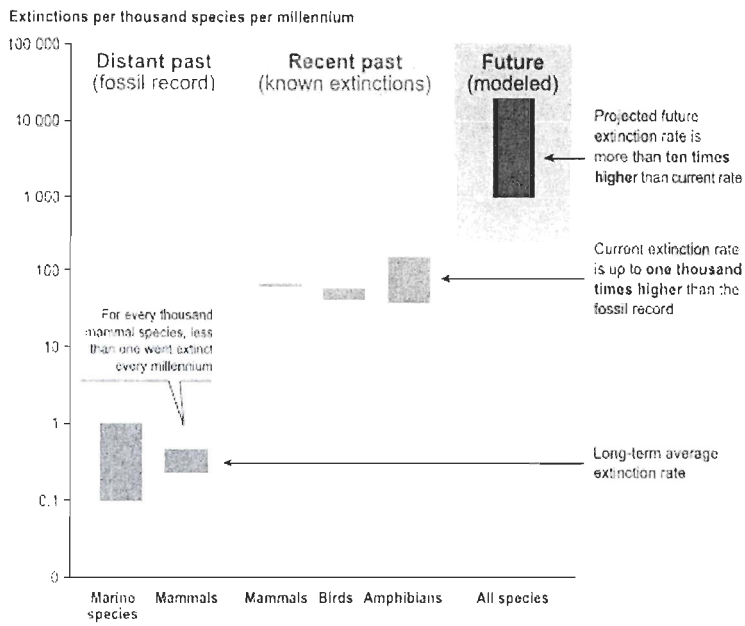


Figure 2 Species extinction rates from fossil data, real data and predictions for the future. (Source: MA, 2005)

The greatest threats for marine biodiversity include overharvesting, pollution (partly due to agricultural run-off), habitat destruction, climate change (increasing surface sea temperatures, acidification due to an increase in CO_2) and invasive species (UNEP report on marine biodiversity, 2010). Coastal systems are more susceptible to be affected due to a growing human population concentrating on coastlines (Gray, 1997). Indeed, it has been estimated that no pristine marine area is left and that 41% of oceans are heavily impacted by humans (Figure 3; Halpern *et al.*, 2008). This view stands in opposition to the view of oceans as open systems, less susceptible to be seriously affected by human activities (at least pollution) compared to land (Gray, 1997; Boero, 2009). Marine fisheries are predicted to collapse by the mid-21st century (Worm *et al.*, 2006), while local collapses of small fish species can have ecosystem-wide impacts by reducing food supply for larger fish, seabirds, and marine mammals (Pinsky *et al.*, 2011).

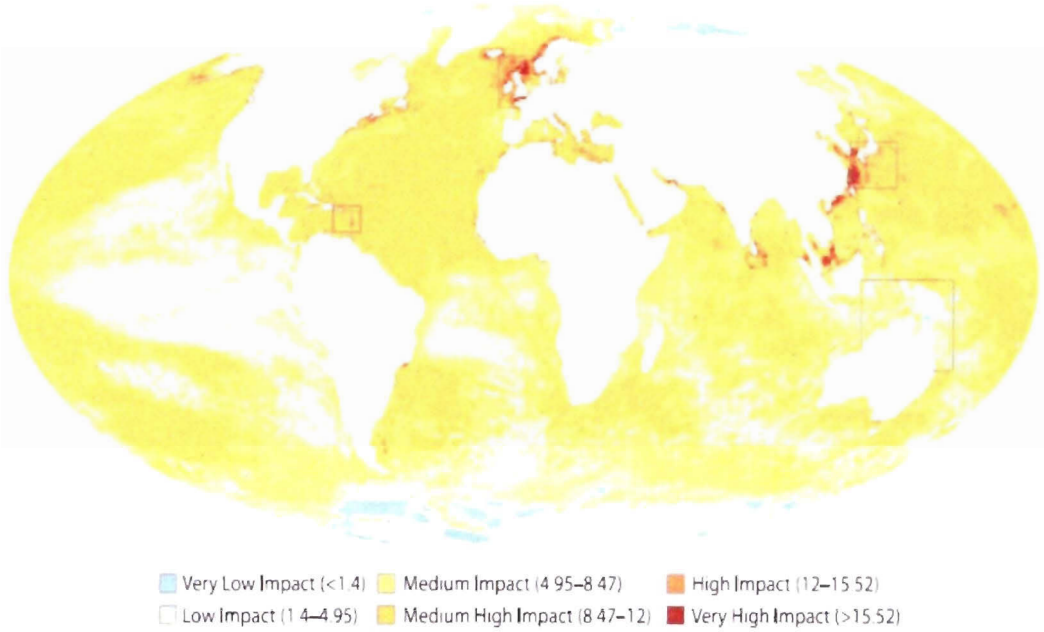


Figure 3 Anthropogenic impact on the world's oceans (investigation performed in 20 ecosystem types). Colors correspond to impact intensity, see above. (Source: Halpern *et al.*, 2008)

The present-day extinction rates for species (but also for plant varieties and for domestic animal breeds) are worrisome. One species, our own, has had an enormous influence, directly and indirectly, on the rest of the biodiversity. Equally impressive are changes in our own cultural diversity. About 50% of existing languages are predicted to disappear within 1-2 human generations (Davis, 2010). Since languages can be considered as markers of distinct cultures, this implies that we stand to lose "half of humanity's social, cultural and intellectual legacy" (Davis, 2010). Notably, the globalized, industrialized culture, which is in great ascension, poses the greatest risk to biodiversity in the conventional sense.

In order to mitigate global biodiversity loss there is a need for sound conservation measures which usually consist of creating protected areas based on

species level attributes (e.g., richness, endemism). However, genetic data should be included as well when designing marine protected areas (MPA) due to its capability to infer population connectivity in protected species (Palumbi, 2003) and to identify populations with different evolutionary histories in need of protection (Crandall *et al.*, 2000).

The economics of biodiversity

Humans are an intrinsic part of global biodiversity and our very existence depends heavily upon biodiversity preservation. We are living in a dynamic environment and we are witnessing a shift in our perception on biodiversity and its importance. Since the formal recognition of the term, biodiversity was considered important and worthy of conservation measures due to its role in supplying food, raw materials, biotechnological resources, ecosystem health and many other services (Table 1), although the overall value was difficult to grasp. In this context, a new approach focused on applying economic concepts to biodiversity valuation has been proposed and a synthesis on the global economic benefits of biodiversity and the costs of biodiversity loss has been published (TEEB, 2010). The overall goal of this emerging direction is to provide a link between science, policy making and business, thus a new vision for managing natural resources.

Ecosystem "goods and services" (provisioning, regulating, cultural and supporting services; MA, 2005) have been valued at US$ 16-54 trillion per year (average of US$ 33 trillion/year) for the entire biosphere while the global gross domestic product was ~US$ 18 trillion per year (Costanza *et al.*, 1997). The marine environment contributes ~63% of the estimated value with most services coming from coastal systems (US$ 10.6 trillion/year). For instance, half a billion people depend on coral reefs for their livelihoods and the monetary value of reefs was

estimated at US$ 6,000/ha/year (Constanza *et al.*, 1997) or US$ 360 million/year for Hawaiian reefs alone (TEEB 2010).

Table 1 Ten economic, ecological, moral, and legal reasons why society needs to protect and manage biodiversity (Modified from Costello, 1998)

	Economic
1	It is essential for the assimilation and recycling of wastes derived from human activity.
2	It is the source of food for humans and domestic animals.
3	It provides valuable recreational resources.
4	It contains biotechnological resources of increasing commercial importance.
5	It produces nonliving resources of commercial importance.
	Ecological
6	It supports economic resources through the food web and interaction between species.
7	It maintains local-to-global ecosystem health through its interaction with the physical and chemical environment (e.g., atmospheric carbon dioxide, oxygenation) and can buffer the world against climate change.
	Moral and Ethical
8	It is generally accepted that other life forms have a right to exist, and that humans have a responsibility of stewardship to protect our natural inheritance for future generations. Indeed, a review of history suggests that we can have little idea of what uses and values future generations may discover in biodiversity.
9	The production of unnecessary waste, and thus pollution, can be considered immoral.
	Legal
10	The Convention on Biological Diversity and other laws now place a legal obligation on most countries and their citizens to protect and sustainably use biodiversity. This is essential because *some people will either not have the ability or willingness to understand the importance of biodiversity, or their short-term selfishness and greed will result in their activities reducing biodiversity.*

TEEB might be seen as giving a price to something priceless. However, it might be an effective direction for sustainable development because the human mind can deal better with numbers (e.g., US$ 19,580/ha/year for swamp/floodplains) than with ecological/ethical reasons when protecting nature. As global biodiversity and services it provides are dynamic, there is a need to forecast modifications associated with climate change and globalization, to update CBD to the current pace of global change and to find viable solutions at local, regional and global levels (Bayon and Jenkins, 2010; Mooney, 2010; TEEB 2010). However, a prerequisite for all conservation plans is to know the extent of biodiversity, how it was developed and what processes maintain it.

The origin of marine biodiversity – how do species arise?

Those people who believe that life is dynamic and continuously evolving have wondered about the underlying mechanisms of diversification (an intrinsic part of biodiversity). Marine organisms are not uniformly distributed but they are rather grouped into local populations connected by dispersal. The ability to maintain population connectivity in the sea will affect the genetic structure, which ranges from lack of structure, indicating panmixia, to various degrees of differentiation, which will eventually culminate in the formation of new species (Figure 4) (Hedgecock, 1986; Palumbi, 1994; Bohonak, 1999). Dispersal capability in the sea is mainly determined by biological factors such as the developmental mode of organisms and by environmental factors such as the oceanographic features. Pelagic species (the less diverse component of marine biodiversity; Gray, 1997) are usually highly dispersive through ocean currents, and are therefore believed to be panmictic. Benthic species (accounting for 98% of marine species; Brunel, 2005) usually have an adult benthic phase and a larval pelagic phase. Larvae are released into the water column and can disperse over large spatial scales via oceanographic currents depending on the amount of time spent in the plankton, their behavior, the spawning season and the

rate and direction of the currents (Scheltema, 1986; Hohenlohe, 2004). Other benthic species are direct developers with eggs hatching into juveniles or non-dispersive larval forms (often associated with maternal care), thus lacking a pelagic larval phase and apparently being highly restricted in their dispersal. Alternative modes of dispersal for benthic taxa include adult active dispersal (by swimming or crawling) and passive dispersal through rafting on floating objects or transport by human vectors (e.g., shipping) (Scheltema, 1986; Thiel and Gutow, 2005).

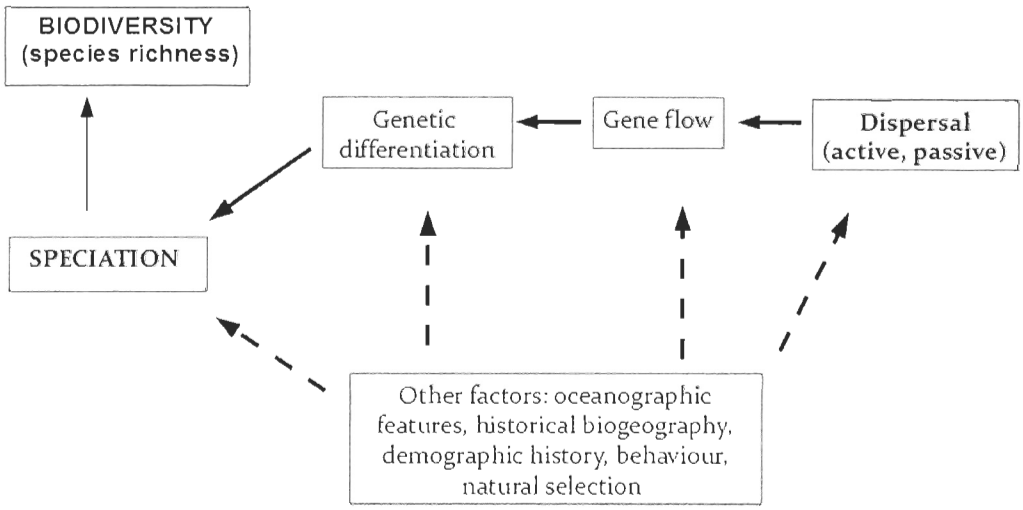


Figure 4 Factors affecting the genetic differentiation, and thus speciation and biodiversity, in the sea. (Partially compiled from Palumbi, 1994 and Grosberg and Cunningham, 2000)

Measuring dispersal in the marine environment is a difficult task but is crucial in determining the size of spatial neighborhoods to be considered in management plans (Palumbi, 2004). Historical patterns of dispersal can be indirectly inferred from fossil data (when such data exist), while the present-day dispersal can be directly measured by tagging organisms or indirectly inferred from genetic data. Tracking individuals with various electronic devices is used mainly for marine vertebrates (mammals, turtles, seabirds, fishes) (Block *et al.*, 2011), less so for invertebrates

(Freire and Gonzalez-Gurriaran, 1998; Gilly *et al.*, 2006) and only for the adult or late-juvenile phases. Therefore, genetic studies are widely employed to investigate the influence of marine dispersal on gene flow and genetic structure with the prediction that direct developers (or species with abbreviated larval development) will show stronger genetic structure (potentially leading to isolation by distance and even allopatric fragmentation) compared to species with dispersive larval phases. Indeed, support for these theoretical expectations has been found in studies of bryozoans (Watts and Thorpe, 2006), gastropods (Kyle and Boulding, 2000; Collin, 2001; Johnson and Black, 2006) and crustaceans (Teske *et al.*, 2007). However, many other genetic studies found various patterns not concordant with the developmental mode – gene flow hypothesis (Costa *et al.*, 2004; Richards *et al.*, 2007; Weetman *et al.*, 2007; Luttikhuizen *et al.*, 2008). Based on genetic evidence, dispersal (i.e., successful movement to a new location) cannot be equalled with gene flow (i.e., successful reproduction of migrants in the new location), although direct developers are obviously less connected at the geographic and genetic levels (Hedgecock, 1986; Scheltema, 1986; Bohonak, 1999). Genetic differentiation and marine speciation are also influenced by environmental factors (e.g., oceanographic features, climatic oscillations, plate tectonics, topography) as well as demographic history, or behavioral, ecological and genetic factors (Figure 4) (review in Palumbi, 1994 and Grosberg and Cunningham, 2000).

The evolution of genetically divergent populations into closely related species (i.e., reproductively isolated units) is based on the appearance of pre/post-zygotic reproductive barriers (e.g., oceanographic features, environmental tolerance, habitat specialization, mate preference and recognition, spawning synchrony, fertilization, offspring viability) (Palumbi, 1994), even if some external barriers are temporary (Hohenlohe, 2004). Depending on the spatial scale involved in the formation of reproductive barriers, speciation can be allopatric, parapatric, peripatric and sympatric (Figure 5) and while the allopatric mechanism seems more likely to occur (reproductive isolation is "helped" by geographic separation), other mechanisms are

also occurring in the sea and might be even more common than previously believed (Malay and Paulay, 2010; Miglietta, Faucci and Santini, 2011 and references therein).

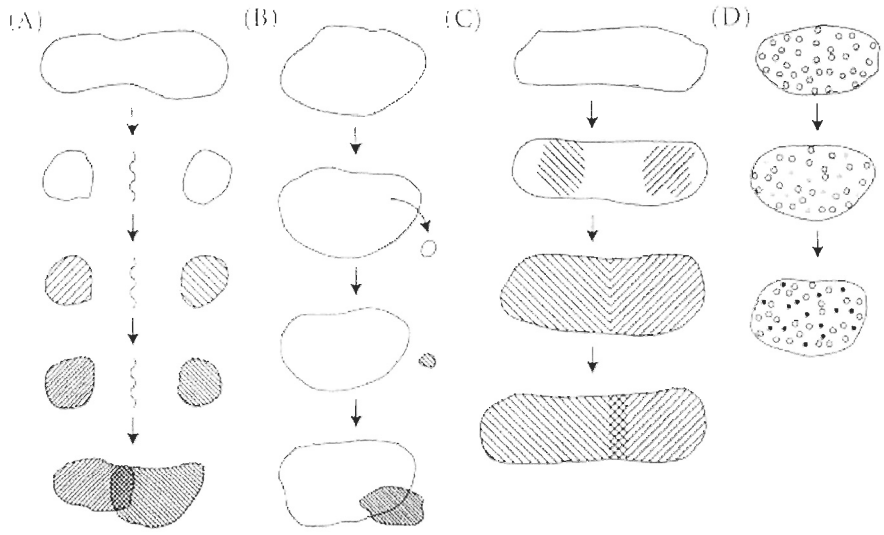


Figure 5 Speciation models depend on the spatial scale involved. A: allopatric model when reproductive isolation occurs between populations physically isolated; B: peripatric model (founder effect) with a small population being physically separated and evolving towards reproductive isolation; C: parapatric model with reproductive barriers occurring between contiguous populations due to low dispersal; D: sympatric model with reproductive barriers developing within the same geographic area. (Source: Futuyma, 1998)

The application of molecular techniques to the study of marine biodiversity and speciation has challenged the once widely-held view of oceans as homogenous environments with few barriers to dispersal, and of marine species as truly panmictic with large population sizes, high fecundity and high dispersal capability. In the light of genetic evidence, speciation appears to be very common in the sea (review in Palumbi, 1994 and Miglietta, Faucci and Santini, 2011). Dispersal in some groups is more limited than theoretical predictions with adult movements of only a few km to

up to 10-100 km (e.g., adult demersal fishes and invertebrates) and larval dispersal of only 10-100 km in invertebrates and of only up to 50-200 km in fishes (review in Palumbi, 2004). Cryptic speciation, due to ecological divergence of species without morphological differentiation, is also common in the sea (Knowlton, 1993). The end product of this process, namely cryptic species, cannot be identified based on morphological characters but just by using molecular methods. Many taxa previously considered cosmopolitan are actually complexes of cryptic species with geographical separation. Other cryptic complexes followed a sympatric model of speciation with reproductive barriers resulting from differences in habitat choice or resource use (Miglietta, Faucci and Santini, 2011 and references therein). Cryptic species are a hidden aspect of marine biodiversity and seem to occur across all marine groups, therefore the extent of marine biodiversity (i.e., species richness) might eventually rival that of the terrestrial realm after more detailed investigation. The identification of cryptic species can be highly controversial (see the following sections) but genetic data can reveal at least the existence of intraspecific genetic groups that are very divergent from one another and, usually, separated geographically (i.e., phylogeographic groups) (Avise *et al.*, 1987) according to marine biogeographic divisions (Dawson, 2001). Whether or not a cryptic species is formally recognized, the occurrence of such intraspecific divergent groups indicates a separate evolutionary history (i.e., ESU), and hence of importance for biodiversity and for management strategies.

Towards a global inventory

The need to have a global inventory of extant species is not provoked only by our curiosity and ambition to organize nature in a professional way (e.g., a stamp collection). Such a checklist will act as a baseline for assessing future biodiversity changes with implications on ecosystem services and, consequently, on human well-being. Moreover, it will help us understand the ecological and evolutionary processes

which are generating and maintaining biodiversity. Surprisingly for the general public, the number of world extant species (or even described species) is unknown. Not surprisingly for scientists, the difficulty in adding up numbers comes from the weakness of extrapolation methods to estimate richness, the scarce sampling of the Earth, the multitude of synonyms (2 million names for 1.6 million described species; Stork, 1997) and taxonomic splitting (i.e., division of one species into two or more). It has been suggested that the number of species on Earth lies anywhere between 3 and 100 million species (Wilson, 2003 and references therein), but most likely around 11 million species inhabit the planet (Chapman, 2009). The latest estimate is a bit lower, ~8.7 million species, but still indicates a large amount of species awaiting discovery (Mora *et al.*, 2011). In addition, we lack sound information on most species that do have names (e.g., distribution ranges, threat of extinction; Stork, 1997).

Traditionally, species have been classified, named and described according to their morphological characteristics within the field of taxonomy (i.e., alpha-taxonomy). This procedure follows a strict protocol according to the International Codes of Nomenclature by which species have unique binomial scientific names (genus and species) and are linked to type specimens (from type localities) preserved in museum collections. Establishing this Linnaean taxonomic system is a very laborious task, which involves the analysis of, ideally, hundreds or thousands of specimens per species in order to assess the extent of intraspecific morphological variation. Consequently, only a fraction of presumed species richness has been described in 250 years. About 6,000 taxonomists are believed to practise worldwide (Wilson, 2003) and their number is rapidly decreasing due to shortage in funding and to the lack of interest in pursuing a "dead" specialization on the job market. The resulting "taxonomic impediment" and the current progress in classifying life (~1,600 species described every year; Bouchet, 2006) predict a timeframe of >1,000 years for an inventory of marine biodiversity alone. Considering also the rates of biodiversity loss, it is evident that many species will go extinct before we even know they existed (Mora *et al.*, 2011).

With the lack of trained personnel and the inherent difficulties to identify many invertebrate groups (especially the various life history phases of species with complex life histories and groups with highly plastic morphology such as corals), it is no wonder that marine faunal inventories usually fail to identify one third of specimens to the species level (Schander and Willassen, 2005). In addition, cryptic species will add to the species level of biodiversity (once they are validated) but also to the difficulty in compiling such lists as they are almost impossible to detect by morphological characters. Therefore, molecular methods have been proposed for species identification (DNA barcoding) as well as for a new taxonomic system (DNA-taxonomy; Tautz *et al.*, 2003).

A new tool: DNA barcoding

The term "DNA barcoding" was coined by analogy with the Universal Product Codes, in which every product has a unique barcode, and it was proposed as a fast, reliable and cost-effective identification tool that uses DNA sequences unique to each species (Hebert *et al.*, 2003). In most animals, this approach uses a fragment of the mitochondrial (mt) gene cytochrome *c* oxidase subunit 1 (COI) to assign unidentified specimens to known species (previously identified by experts and stored in a reference DNA library). The choice of mtDNA over nuclear DNA is based on a few characteristics: i) large copy numbers in each cell, therefore easier to amplify from small amounts of tissue or when DNA is degraded; ii) maternal inheritance, therefore no recombination (but see Galtier *et al.*, 2009); iii) higher evolutionary rate; and iv) lack of introns (Hebert *et al.*, 2003). In most animals, the circular mt genome includes 24 genes for mtDNA translation (2 ribosomal RNAs: 12S, 16S; 22 transfer RNAs) and 13 protein-coding genes for the electron transport chain (Figure 6). These 37 genes interact with ~1,500 genes encoded by nuclear DNA (nDNA) (Gershoni, Templeton and Mishmar, 2009).

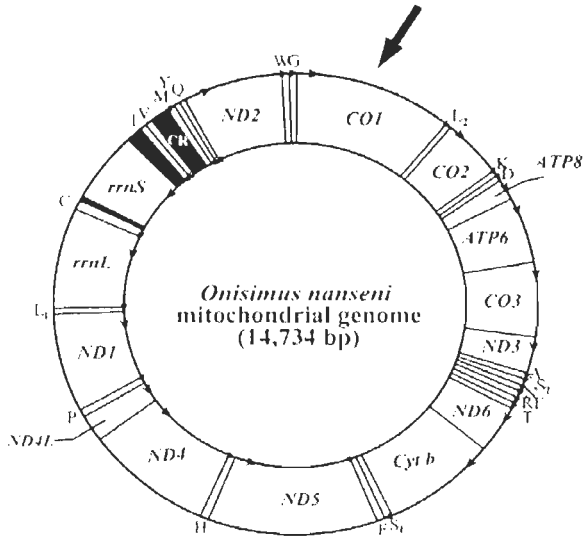


Figure 6 Mitochondrial genome of the Arctic amphipod *Onisimus nanseni.* The position of COI is indicated by an arrow. (Modified from Ki *et al.*, 2010)

Protein-coding genes have the advantage of lacking insertions and deletions and COI was selected due to its slow mutation rate (relative to other mt genes), thus a higher probability of being amplified in a wide range of species with standard protocols, while previous research found this gene to distinguish between closely related species and to identify intraspecific phylogeographic groups (Hebert, Ratnasingham and deWaard, 2003; Hebert *et al.*, 2003). Mitochondria are the "powerhouse" of cells generating energy through the electron transport chain which consists of multiple protein complexes situated in the inner mitochondrial membrane. The protein coded by COI has a functional role as part of the Complex IV of the respiratory chain (Figure 7). The mitochondrial respiratory chain is more efficient in producing energy than the nuclear-controlled glycolysis for instance (30 versus 2 ATP molecules per molecule of glucose oxidized), but it also generates toxic products (reactive oxygen species, ROS), which can have a negative effect on DNA, protein and lipids (Ballard and Whitlock, 2004).

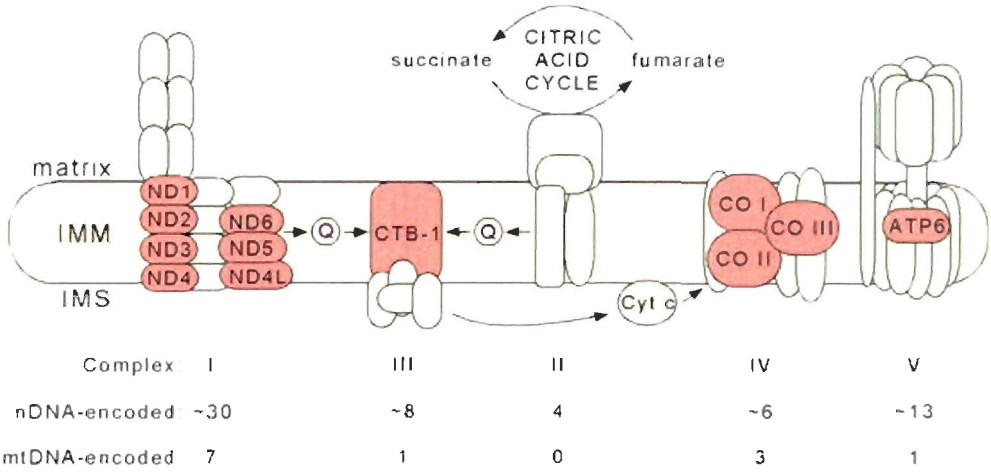


Figure 7 Mitochondrial respiratory chain in the nematode *Caenorhabditis elegans* with five complexes of proteins encoded by mtDNA (red) and nDNA (green). IMM: inner mitochondrial membrane; IMS: intermembrane space; Q: ubiquinone; Cyt c: cytochrome *c*. (Note: in *C. elegans* there are only 12 protein-coding mt genes). (Source: Lemire, 2005)

This molecular identification method has stirred an unprecedented debate since its inception, with opponents constructing a long list of shortcomings for COI, the use of only one diagnostic character, taxonomic inflation by over-splitting traditional species, potential loss of interest for morphological taxonomy, and alleged anti-intellectualism or competition for funding with other biology fields (Ebach and de Carvalho, 2010; Will and Rubinoff, 2004; Will, Mishler and Wheeler, 2005; Rubinoff, 2006; Rubinoff, Cameron and Will, 2006). However, almost one decade of research has shown that DNA barcoding did not cause the extinction of classical taxonomy and that many advantages can arise from its use. The capacity to identify anything that contains DNA has multiple practical applications: food traceability (Marko *et al.*, 2004; Wong and Hanner, 2008; Barbuto *et al.*, 2010); detection of pests, disease vectors, parasites (Locke *et al.*, 2010), endangered species traded illegally, and invasive species (Radulovici, Sainte-Marie and Dufresne, 2009; Saunders, 2009); diet analysis (gut content or feces) (Deagle *et al.*, 2010; Stech *et al.*, 2011; Zeale *et al.*, 2011). DNA extraction protocols are evolving towards non-invasiveness by swabbing bird eggs (Schmaltz *et al.*, 2006), using cetacean blows (Frère *et al.*,

2010), amplifying DNA leaked into the water (Ficetola *et al.*, 2008) or ethanol (mescal "worms"; Shokralla, Singer and Hajibabaei, 2010). New protocols also allow for the recovery of small specimens after DNA isolation in order to preserve the vouchers (Porco *et al.*, 2010).

DNA barcoding is more than just another method of molecular identification in that, as its name implies, it involves standardization. In practice, in any given taxonomic group, there are always markers that are as good as or even better for resolving species than the COI barcode. However, the issue is not which marker is best for each particular group. By sequencing optimal markers for each group, there will be a vast, diverse, but non-comparable array of genetic data. The issue is whether the COI barcode performs sufficiently well across the broadest possible range of taxa. In addition, barcoding fosters links to various non-genetic data such as collection information, specimen images, accessions for vouchers stored in public institutions. All data are uploaded on-line (Barcode of Life Data Systems, BOLD; Ratnasingham and Hebert, 2007) and publicly available following project publication. The importance of DNA barcoding for marine biodiversity will be discussed at length in Chapter I.

Goals of this thesis

The general goal of my PhD thesis was to use molecular methods (specifically DNA barcodes represented by COI sequences) as a means to assess biodiversity in the marine environment. As it is impossible to investigate the entire extent of marine biodiversity at the global scale, a case-study was chosen: shallow-water crustaceans from the northwest Atlantic (NWA). Two biodiversity levels were tackled: genes and species.

As DNA barcoding is an emerging tool, the first step in any study of this kind consists of building a reference library of DNA sequences. A reliable database has to be built by performing COI sequencing on specimens previously identified by a taxonomist. Therefore, a pre-requisite for genetic investigations in this study was the technical step of building a database for crustaceans from NWA.

Species level

An intrinsic part of DNA barcoding is species-hypothesis testing (i.e., does any given morphologically defined species consist of one or multiple barcode clusters?) and for this purpose I used various taxa with different potential for dispersal (hence different potential for genetic divergence and speciation; Figure 4). Results of DNA barcoding usually include detection of cryptic species, which will translate into higher species richness once validated by taxonomists.

Besides species richness, another important aspect in biodiversity is the phylogenetic diversity, involved in calculating the taxonomic distinctness index (Warwick and Clarke, 1995). Underestimating this type of genetic variation will affect diversity indices and, consequently, biodiversity assessments. Phylogenetic analyses were conducted within one crustacean family, the semi-terrestrial Talitridae, in order to investigate the monophyly of genera (i.e., all congeneric species are descending from one common ancestor). Non-monophyly, implying different evolutionary histories, will lead to taxonomic splitting into multiple genera which will translate into higher diversity above the species level (higher taxonomic distinctness), once validated by taxonomists.

Specific questions at the species level:

i) How common are cryptic species among NWA crustaceans?

ii) How many cryptic species exist within one crustacean family, the Talitridae?

iii) What are the phylogenetic relationships within Talitridae? Are talitrid genera monophyletic or not?

Genetic level

Focusing on individual species, DNA sequences can be used to infer phylogeographic patterns at large spatial scales and/or genetic structure at smaller spatial scales. Strong population differentiation will have reverberations at the superior (species) level on an evolutionary time scale (Figure 4). Phylogeographic patterns were investigated in one littoral amphipod species, *Gammarus oceanicus*, with amphi-Atlantic distribution (Steele and Steele, 1972), and most likely affected by the glacial history of the North Atlantic. Besides genetic differentiation, the goal was to explain the present-day distribution pattern (survival on both coasts or on only one with subsequent colonization of the other coast).

Study area: North Atlantic

The North Atlantic originated in the Jurassic period during the break-up of Pangaea and it was influenced by climatic oscillations with rapid cooling in the late Eocene (from subtropical to temperate and cold). These changes lead to biological diversification in the marine environment in relation to emerging environmental conditions (Golikov and Tzvetkova, 1972). During the Pliocene, the North Atlantic was invaded by Pacific taxa via the Arctic due to the opening of the Bering Strait (Vermeij, 1991). More recently the North Atlantic communities were influenced by the Pleistocene glaciations, during their glacial and interglacial phases. At the last glacial maximum (LGM), North America and Europe were covered by massive ice sheets

(Figure 8) while the sea level decreased to -130 m (Mix, Bard and Schneider, 2001) uncovering the continental shelves and forcing organisms to migrate south or survive in glacial refugia.

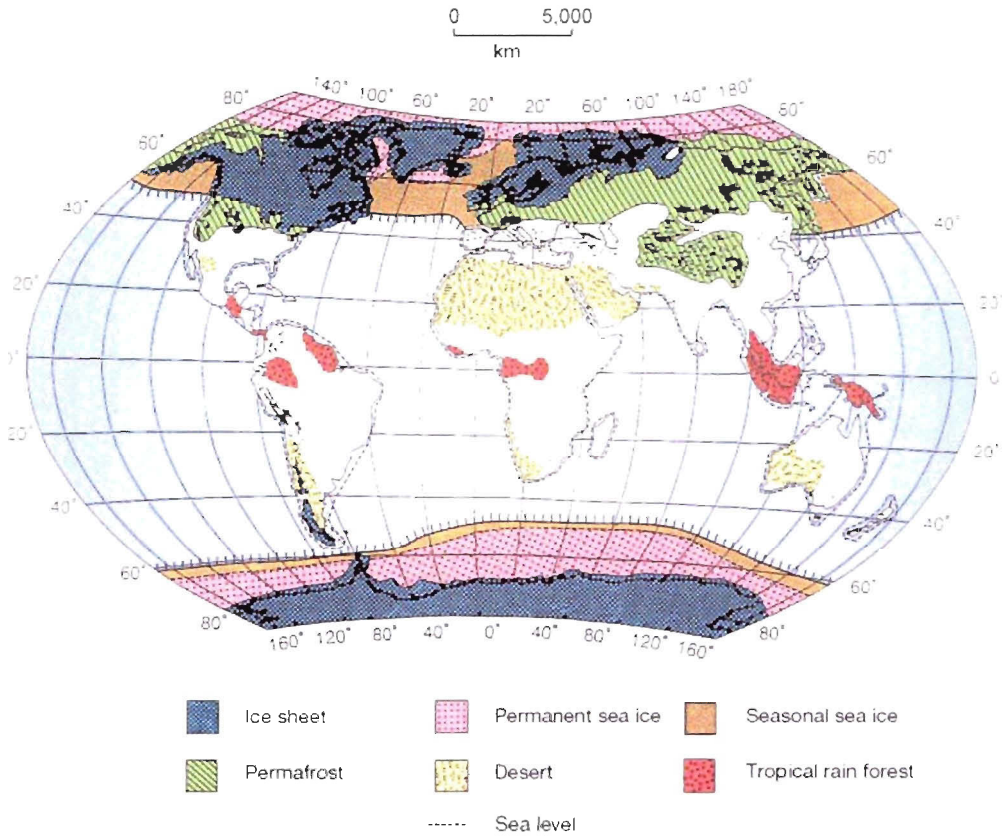


Figure 8 The extent of ice and sea level retreat during LGM. (Source: Hewitt, 2000)

The present-day North Atlantic communities are the result of the above-mentioned historical events. Moreover, the ocean circulation (Figure 9) is one of the main factors influencing genetic differentiation, and therefore biodiversity (Figure 4).

Figure 9 Present-day circulation in the North Atlantic. Currents: red – warm, blue – cold. GIN – Greenland/Iceland/Norway. Black rectangle: Atlantic Canada. (Source: www.planetastronomy.com)

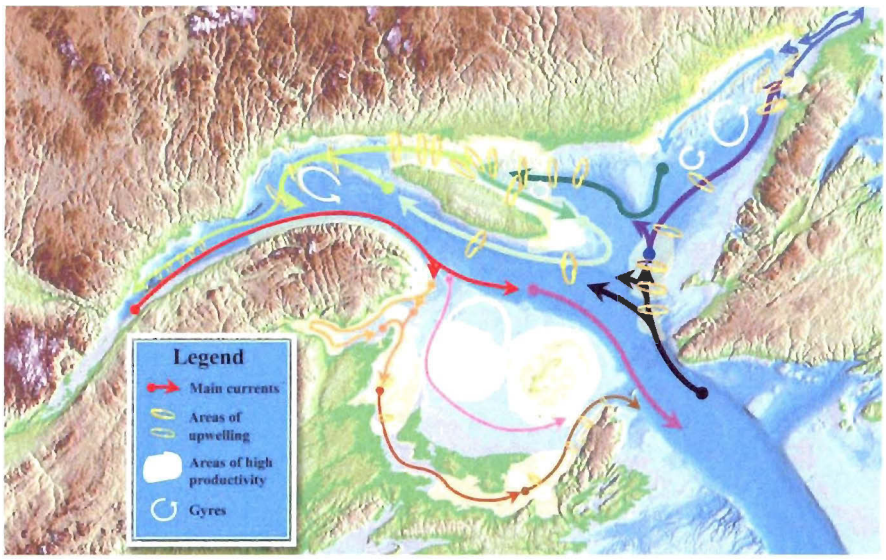


Figure 10 Circulation patterns within the Estuary and the Gulf of St. Lawrence. (Source: DFO)

In Atlantic Canada, special focus was oriented towards the Estuary (ESL) and the Gulf of St. Lawrence (GSL), a region with complex physiographic, oceanographic and bathymetric characteristics, which has been divided into 20 biogeographical zones (Figure 10) (Brunel, Bossé and Lamarche, 1998).

Study group: Malacostraca, Crustacea

Crustacea is a subphylum currently composed of six classes, 42 orders, 849 families and ~52,000 described species but estimated to be much more diverse (Martin and Davis, 2001 and references therein). Living in marine, freshwater and terrestrial systems, crustaceans are an ancient group, dating back to the Cambrian, and from a morphological and ecological point of view, it is the most diverse metazoan group (Martin and Davis, 2001). Recent phylogenies based on multiple genetic markers (62 single-copy nuclear protein-coding genes) have shown the non-monophyletic character of crustaceans, placed together with terrestrial insects (Hexapoda) in a Pancrustacea phylum (Regier *et al.*, 2010).

Marine crustaceans exhibit a wide variety of body shapes, sizes and life styles (from free-living to tube-dwelling, sessile, commensal or parasites on invertebrates or vertebrates), and biological and ecological characteristics. Crustaceans occupy diverse habitats in both the pelagic and benthic realm, at all latitudes and depths. As a result, they play an important role in marine ecosystems, often being a key part of food webs (e.g., copepods in the Northern Oceans, krill in the Southern Oceans) or being harvested as a food source on large spatial scales. For this study, crustaceans were chosen as a target group for the following reasons: i) taxonomic difficulty, often requiring the help of highly-trained personnel for identification; ii) unsettled systematic; and iii) importance (ecological and economic). The use of DNA barcoding for crustacean identification has multiple practical applications: identification of eggs and larvae (consequent use in stock assessment

of harvested species), invasive species, parasites, cryptic species or fraudulent substitutions in processed seafood.

This study focused on two superorders of malacostracan crustaceans from shallow-water North Atlantic: Peracarida and Eucarida (Figure 11). Peracarids are generally short-lived organisms with low fecundity and they are characterized by direct development (eggs hatch directly into juveniles within the maternal brooding pouch with no larval phases), and hence limited capabilities for large-scale dispersal (with implications at the genetic level, see Figure 4). Among peracarids, amphipods are a species-rich order that is also an important component of the marine food webs. Other peracarids targeted in this thesis include isopods and mysids. Eucarids are generally long-lived organisms with high-fecundity and larval development, with larvae usually spending various amounts of time in the plankton, hence their potential for large-scale dispersal. Among eucarids, decapods are the most important group. They include species with economic importance (e.g., lobsters, shrimps, and crabs) which bring high revenues to Atlantic Canada. Decapods are also ecologically important as top predators in marine benthic ecosystem. Genetic studies for North Atlantic crustaceans have shown various degrees of population connectivity and genetic structure (Sévigny, Savard and Parsons, 2000; Martinez *et al.*, 2006; Puebla *et al.*, 2008) culminating with cryptic speciation (Kelly, MacIsaac and Heath, 2006), which might be a frequent phenomenon in crustaceans (Knowlton, 1993, 2000).

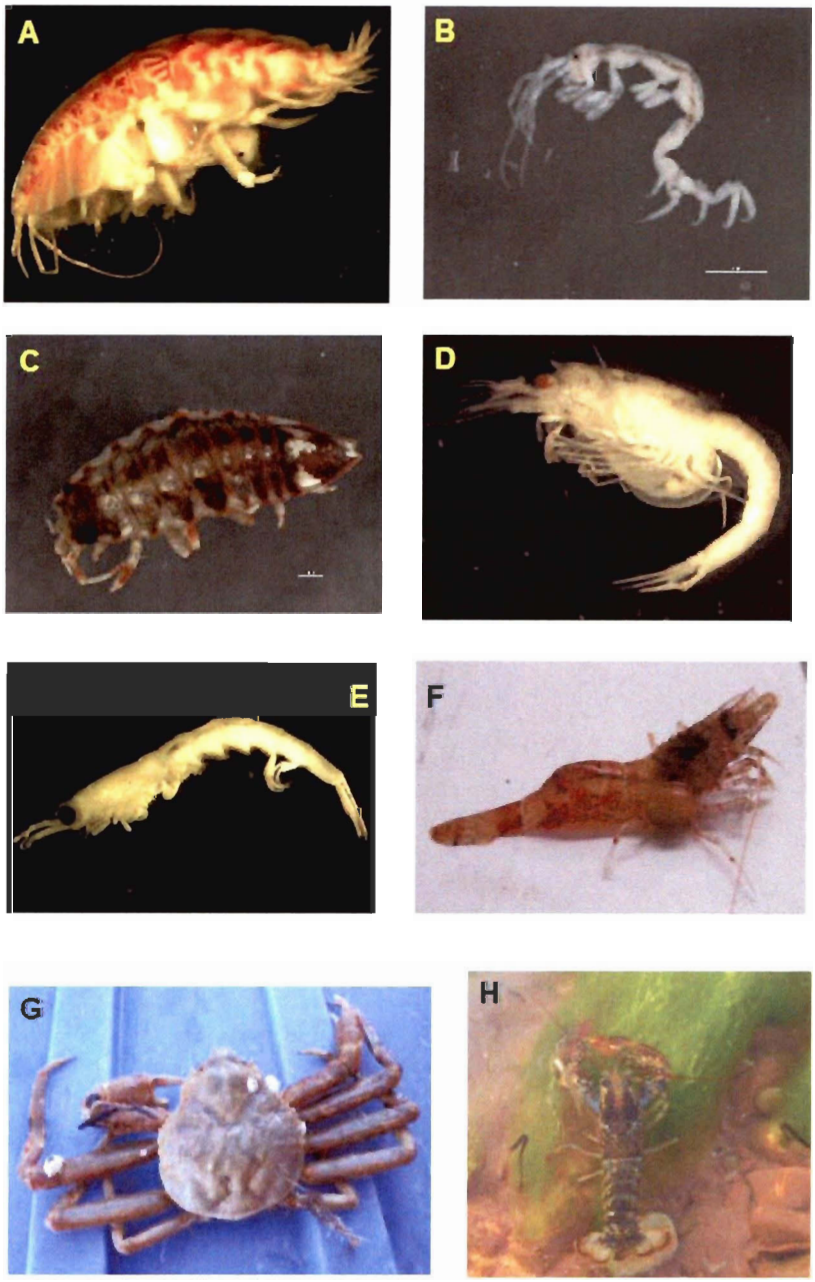


Figure 11 Representatives of the main crustacean groups targeted in this study. A-D: peracarids (A, B: amphipods, C: isopod, D: mysid), E-H: eucarids (E: euphausiid, F-H: decapods).

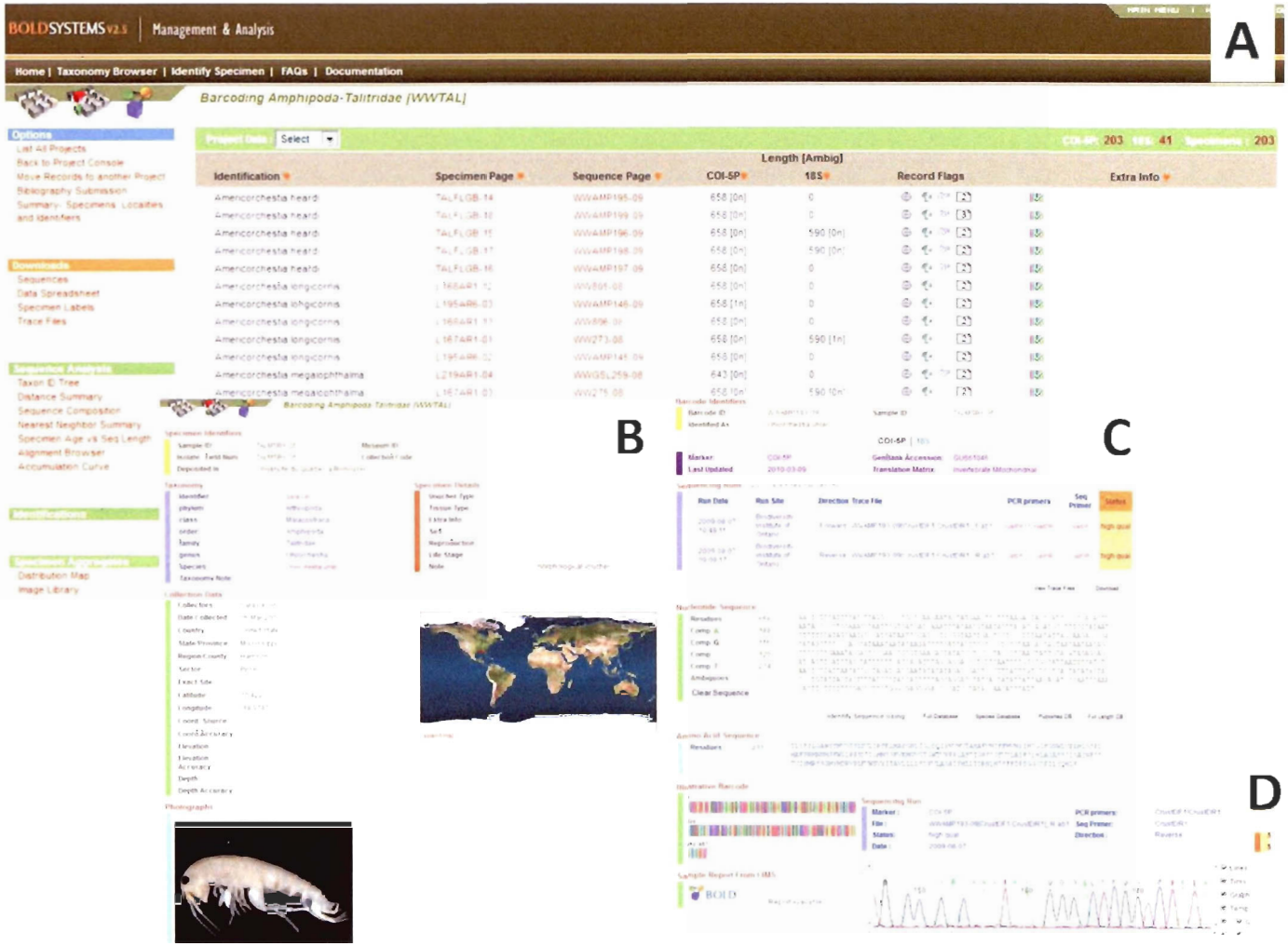
Methodology

DNA barcoding is a unique and rapidly expanding method for the molecular identification of organisms. The workflow includes a few mandatory steps required for "true" DNA barcoding studies as opposed to other methods of molecular identification: vouchers stored as reference in public institutions, taxonomy, images and collection details uploaded on BOLD and publicly available. This study followed the barcoding workflow as closely as possible. Crustaceans were collected at low-tide in multiple habitat types (rocky shores, mudflats, sandy beaches, salt marshes, seagrass beds) or during research surveys of Fisheries and Oceans Canada (DFO) and stored in a DNA-friendly manner (e.g., fixed and stored in 95% ethanol). Metadata included collection details (date, GPS coordinates, locality name for low-tide sampling, and depth for DFO missions) and taxonomic data. Specimens were photographed, identified by qualified personnel (at least one specimen per species) and stored as vouchers for future reference. All data were uploaded to BOLD and they are publicly available (published projects, see Chapter II) or will become so after publication (Chapter III and IV) (Figure 12).

The laboratory operations were carried out at the Canadian Center for DNA Barcoding (University of Guelph, Canada). Specific protocols are explained in greater detail in Chapter II. COI sequences and trace files were uploaded to BOLD and in some cases data were analyzed directly in BOLD (Chapter II) by calculating genetic distances and building neighbor-joining trees. Genetic distances are usually calculated by incorporating the Kimura-two-parameter (K2P; Kimura 1980), which takes into account multiple substitutions per site and different rates for transitions (A-G, C-T) versus transversions (A/G-C/T), but invariable substitution rate between sites and equal frequency for the four nucleotides. Although there are multiple models of molecular evolution, K2P has been proposed for DNA barcoding studies involving the COI gene as the best metric for low distances (Hebert *et al.*, 2003). By using these standard methods, large-scale comparisons across taxa will be easily

conducted in future studies. Neighbor-joining trees based on K2P distances are usually built in order to rapidly visualize large datasets of DNA barcodes and the assortment of individuals into clusters (Hebert *et al.*, 2003). Genetic distances are used for species delimitation based on a threshold of 3% (seen to deliver 98% success for species delimitation in Lepidoptera; Hebert *et al.*, 2003) or 10× the mean intraspecific value (Hebert *et al.*, 2004). However, cut-off approaches have to be carefully considered due to variable mutation rates across taxa (Galtier *et al.*, 2009) or incomplete taxonomic sampling (Meyer and Paulay, 2005). Another method for species delimitation in crustaceans takes into account the number of substitutions per site (0.16 substitutions per site; patristic distances) but requires an *a priori* phylogeny (Lefébure *et al.*, 2006), therefore it was not used in this thesis. Partial datasets of COI sequences generated for barcoding purposes were used in phylogenetic and phylogeographic analyses and methodological details are given in Chapters III and IV.

Figure 12 (page 29) Example of data and metadata related to a DNA barcoding project in BOLD. All data can be downloaded. A: Project page (code: WWTAL) with a list of specimens included in the project, links to specimen and sequence details, and to various analyses that can be conducted directly in BOLD. B: Specimen page with metadata for a specimen of *Uhlorchestia uhleri* (voucher details, taxonomy, image, and collection details with GPS coordinates and site map). C: Sequence page for the same *U. uhleri* specimen with details about the sequencing step (primers used, DNA sequence and amino acid translation). D: Chromatogram (forward reaction) for the same *U. uhleri* specimen.


BOLDSYSTEMS v2.5 | Management & Analysis
Home | Taxonomy Browser | Identify Specimen | FAQs | Documentation
A
Barcoding Amphipoda-Talitridae [WWTAL]
Options
Downloads
Sequence Analysis
Taxon ID Tree
Distance Summary
Sequence Composition
Nearest Neighbor Summary
Specimen Age vs Seq Length
Alignment Browser
Accumulation Curve
Distribution Map
Image Library
Project Data | Select
Identification
Specimen Page
Sequence Page
COI-5P
18S
Length [Ambig]
Record Flags
Extra Info
Americorchestia heardi
Americorchestia longicornis
Americorchestia megalophthalma
B
C
D

Thesis outline

This thesis includes an introduction to the basic principles explored, a review of past and current literature (Chapter I), three research chapters (Chapter II-IV) and general conclusions.

Chapter I gives a partial introduction to molecular methods for marine biodiversity assessments. The focus of this chapter is on the species level of biodiversity and it reviews multiple studies involving DNA barcoding of various marine groups, from seaweeds and diatoms, to invertebrates and ending with mammals. This chapter provides also a philosophical view on the importance of species and future directions for collaborative work between taxonomists and barcoders.

Chapter II begins the research part of this thesis at a medium spatial scale, namely GSL, and its malacostracan fauna (amphipods, isopods, mysids, decapods, and euphausiids). The most common species and those with economic importance (shrimps, crabs, and lobster) were included. While it can be considered a technical chapter in which sequences were generated for a regional database, it includes an intrinsic goal of testing species boundaries (a universal theme in barcoding studies). Therefore, routine barcoding studies reveal cases of cryptic species (species-splitting) or taxonomic synonymy (species-lumping).

Chapter III increases the spatial scale southward by including the east coast of Canada and the US, the Gulf of Mexico (GOM), and eastward by including a few localities from Europe. The focus of this chapter is still on the species level but with investigations at higher taxonomic levels (within and between genera). The investigation of species boundaries (DNA barcoding) and phylogenetic relationships

(maximum-likelihood, Bayesian inference) were conducted within one family, Talitridae, a unique amphipod family with semi-terrestrial distribution.

The last chapter extends the spatial scale northward by including amphi-Atlantic localities (GSL, open Atlantic coast in Canada, Iceland, Norway, Poland) as well as the Arctic (Hudson Bay, Hudson Strait). This chapter specifically targets the genetic (intraspecific) level by conducting large phylogeographic analyses in one amphipod species, *Gammarus oceanicus*, a very common and abundant species in intertidal and subtidal communities.

The general conclusions review the main findings of my thesis on barcoding marine crustaceans from North Atlantic. This final component includes a broad discussion on possible limitations of the present study and future directions in understanding and protecting marine biodiversity.

CHAPTER I

DNA BARCODES FOR MARINE BIODIVERSITY: MOVING FAST FORWARD?

Diversity 2010, 2, 450-472

Adriana E. RADULOVICI[1], Philippe ARCHAMBAULT[2], France DUFRESNE[1]

[1] Département de Biologie, Université du Québec à Rimouski, Rimouski (QC) G5L 3A1, Canada
[2] Institut des sciences de la mer de Rimouski (ISMER), Université du Québec à Rimouski, Rimouski (QC) G5L 3A1, Canada

Author contribution
Concept: AER
Manuscript writing: AER, PA
Comments & funding: FD

1.1 Résumé

Le terme «Biodiversité» désigne la diversité du vivant. Elle peut être étudiée à différents niveaux (génétique, espèces, écosystèmes) et à différentes échelles (spatiale et temporelle). Les dernières décennies ont montré qu'à tous les niveaux, la biodiversité marine a été gravement sous-estimée. Afin d'étudier plusieurs modèles représentatifs de cette biodiversité marine et les mécanismes de spéciation, il est nécessaire d'identifier les espèces présentes dans l'écosystème marin. Un nouvel outil d'identification des espèces, le code-barres d'ADN, peut attribuer sans ambiguïté des échantillons inconnus à des espèces connues, révélant aussi le potentiel cryptique de certaines espèces ou la présence de populations génétiquement éloignées. Ce chapitre passe en revue le rôle du code-barres d'ADN dans l'étude de la biodiversité marine au niveau des espèces.

Mots-clés: biodiversité; marine; code-barres d'ADN; identification des espèces

1.2 Abstract

"Biodiversity" means the variety of life and it can be studied at different levels (genetic, species, ecosystem) and scales (spatial and temporal). Recent decades showed that marine biodiversity has been severely underestimated at all levels. In order to investigate diversity patterns and their underlying processes, there is a need to know what species live in the marine environment. An emerging tool for species identification, DNA barcoding, can reliably assign unknown specimens to known species, also flagging potential cryptic species and genetically distant populations. This paper will review the role of DNA barcoding for the study of marine biodiversity at the species level.

Keywords: biodiversity; marine; DNA barcoding; species identification

1.3 Introduction

"Biodiversity" is a broad and abstract concept, widely used by the scientific world but with reverberations at the economic, political and social levels. With more than 17,000,000 hits on the Google search engine (February 2010), the concept of biodiversity is becoming a commonplace name, even more so in 2010 – The International Year of Biodiversity as proposed by the United Nations. But what does "biodiversity" mean? Shorthand form of "biological diversity", it literally means the "variety of life" (Gk. "bios", Lat. "diversitas"). It was officially mentioned for the first time at the National Forum on Biodiversity held in 1986 at Washington D.C. (Wilson, 1988) and it became a funded research field in 1992 through the Convention on Biological Diversity (http://www.cbd.int). With three main levels accepted and usually investigated (genes, species, ecosystems), biodiversity must be conserved in order for our society to prosper, even more so that a "biodiversity crisis" (highest human-induced extinction rates ever) was shown to occur (Pimm *et al.*, 1995). However, a required step prior to protection is biodiversity assessment, usually conducted at the species level of biodiversity. Therefore, species identification has a paramount importance.

How many species are there and how do we recognize them? No precise species number can be provided but it is believed to approximate 1.9 million described species out of 11 million estimated (Chapman, 2009). Traditionally, morphology was a key factor in describing and naming species within the field of taxonomy. This long-standing approach, starting with Aristotle and becoming organized due to Linnaeus, can be very tedious and a matter of subjectivity since it is up to the taxonomist to choose those morphological characters believed to delineate species (whatever "species" meant according to different views; Coyne and Orr, 2004). As a result, it took 250 years for traditional taxonomy to provide descriptions for less than a quarter of the world species using as tools a variety of morphological keys, sometimes "written by those who don't need them for those who can't use

them" (Packer *et al.*, 2009a). After centuries of acquiring knowledge, taxonomy started to lose popularity to other fields resulting in a worldwide shortage of trained personnel. Paradoxically enough, every biological study requires some taxonomic knowledge.

At the turn of the last century, the original blend of "biodiversity crisis" and "taxonomic impediment" brought a stringent flavour to biodiversity studies. Although a solution is not envisaged yet, new approaches based on molecular markers might be of great help in advancing our knowledge of biodiversity. As opposed to morphological identifications and their "mediocrity" in some cases (Packer *et al.*, 2009a), molecular methods are better tools for the identification of early life stages or partial specimens. One method in particular, DNA barcoding, was the incentive for a large debate on the current and future status of taxonomy. Here, we review the role of DNA barcoding for marine biodiversity studies at the species level. For this goal, we searched the Web of Science by using "DNA barcod*" and "marine" as keywords and we retained only those papers that specifically dealt with species diversity and reference libraries of DNA barcodes. We provide an update regarding the progress in barcoding various marine groups and some future directions, as well as a plea for collaboration between barcoders and classical taxonomists.

1.4 Marine biodiversity

By numbers, biodiversity in the sea seems to be quite low, varying between 167,817 valid species (or 318,004 taxa, species to phyla) according to the World Register of Marine Species (WoRMS; http://www.marinespecies.org) (February 2010), and 229,602 marine species described (Bouchet, 2006) (Table 1.1), but estimated to exceed 10 million (Grassle and Maciolek, 1992).

Table 1.1 Global numbers of marine species per taxon according to Bouchet (2006) and WoRMS. Only taxa present on both lists were included

Marine group	Bouchet (2006)	WoRMS (February 2010) Valid species
Bacteria	4,800	625
Fungi	500	1,061
Rhodophyta	6,200	6,302
Acanthocephala	600	410
Annelida	12,148[a]	12,631
Arthropoda	47,217[b]	44,591
Brachiopoda	550	386
Bryozoa	5,700[c]	1,525
Chaetognatha	121	208
Cnidaria	9,795	11,071
Ctenophora	166	170
Cycliophora	1	2
Echinodermata	7,000	5,764
Echiura	170	203
Entoprocta	165–170	161
Gastrotricha	390–400	524
Gnathostomulida	97	97
Hemichordata	106	106
Mesozoa	106[d]	115
Mollusca	52,525	23,689
Nematoda	12,000	5,889
Nemertea	1,180–1,230	1,371
Phoronida	10	11
Platyhelminthes	15,000	3,348
Porifera	5,500	8,174
Rotifera	50	185
Sipuncula	144	158
Tardigrada	212	170
Chordata	21,517[e]	21,944
Total	**203,887**	**150,891**

[a] includes Pogonophora (separate taxon in Bouchet, 2006)
[b] as two taxa, Crustacea and Chelicerata
[c] as Ectoprocta
[d] as two taxa, Rhombozoa and Orthonectida
[e] includes Urochordata, Cephalochordata, Pisces and Mammalia (no reptiles)

The belief that oceans are a homogeneous environment in which speciation is not a common process resulted in only a fraction of the scientific attention being oriented towards marine compared to terrestrial biodiversity (Figure 1.1).

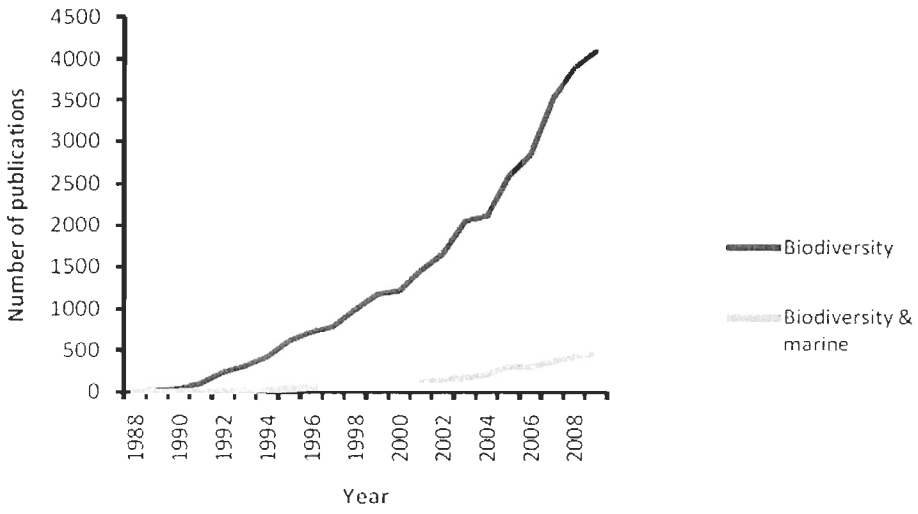


Figure 1.1 The amount of articles focusing on marine biodiversity since 1988 ("biodiversity" and "marine" used as keywords in Web of Science).

However, oceans cover more than 70% of our planet and it was a matter of improving technologies until new explorations of new habitats, especially deep-sea, allowed the discovery of new species (Vrijenhoek, 2009), while cryptic species (morphologically similar but genetically distinct) were shown to be a common presence in marine systems (Knowlton, 1993). Consequently, a more careful look at the world oceans might show, even by numbers, that biodiversity in the sea is as great as on land. On the other hand, an opposite situation occurs at higher taxonomic levels. Of the 35 animal phyla that have been described so far, all but one has living representatives in the oceans, while 14 phyla are marine endemics (Briggs, 1994; Gray, 1997). Within marine ecosystems, most diversity is benthic, consisting of invertebrates residing in (infauna) and on (epifauna) sediments. Brunel

(2005) mentioned that benthic animals, seaweeds and protists account for 98% of marine species diversity and the remaining 2% is pelagic. Other patterns of marine biodiversity include an increase in species diversity from Arctic to tropics and from coastal waters to deep-sea (Gray, 1997).

The importance of marine biodiversity can be translated at the economic or ecological level: source of food, biotechnological and non-living resources, as well as indicator of environmental health and ecosystem functioning (food webs). Major threats to marine biodiversity include overharvesting, habitat degradation, pollution, global warming, biological invasions and other anthropogenic stressors, most of them impacting coastal areas rather than the open ocean (Gray, 1997). For instance, overfishing is predicted to cause a collapse of all fished taxa within the next 50 years (Worm *et al.*, 2006), while marine invaders have already increased their ranges and are present in at least 84% of marine ecoregions worldwide (Molnar *et al.*, 2008). Given these major concerns, it becomes more important than ever to know how many species are present in an ecosystem in order to understand and conserve species diversity.

There are significant disparities across marine taxa in terms of knowledge and status of taxonomic inventory. Larger organisms (e.g., fishes, mammals) are represented by fewer taxa in the world oceans and are usually well-studied groups. However, surprising findings can sometimes emerge, challenging our views on current knowledge. For instance, the number of marine mammals from Canadian waters currently reaches 52 species (Archambault *et al.*, 2010) compared to 10 species listed in 1995 (Mosquin, Whiting and McAllister, 1995). Considering how comparatively well known marine mammals are relative to most marine invertebrates, the inferred gaps in knowledge are particularly disconcerting when attempting to estimate the biodiversity of smaller organisms in poorly-sampled taxonomic groups, such as benthic and pelagic invertebrates, phytoplankton, and microbes. For marine invertebrates, the extent of taxonomic knowledge, including

the number of species described every year, depends on the size of the taxonomic community studying various groups (Figure 1.2) (Bouchet, 2006). For instance, molluscs and crustaceans are the largest groups but probably due to large communities of malacologists and carcinologists, while polychaetes, believed to be one of the most abundant and species-rich macrobenthic taxa (Grassle and Maciolek, 1992), are in great need of taxonomic work. With so many difficulties for biodiversity assessment, there is no wonder that marine faunal inventories usually fail to identify one third of specimens to the species level when using morphological methods (Schander and Willassen, 2005).

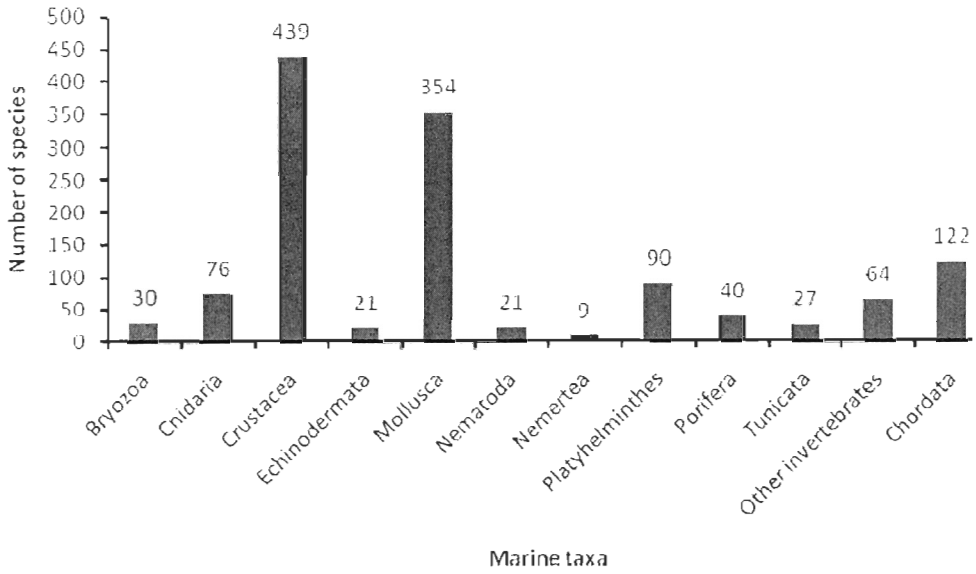


Figure 1.2 Average number of marine animal species per taxon described every year. (Modified from Bouchet, 2006)

1.5 Molecular methods for species diversity

Given that morphological diagnosis poses a problem for the identification of all life stages (e.g., eggs, larvae), for sexually dimorphic species or those with large phenotypic plasticity and considering that cryptic species are widely distributed in marine systems (Knowlton, 1993), it is no surprise that scientists took the opportunity provided by the development of molecular methods to clarify many ambiguities in traditional taxonomy. Allozymes, alternative forms of enzymes coded by alleles at the same locus, were the first molecular markers widely used in population genetics to document patterns of genetic diversity in populations and also served as a useful tool in early molecular systematic studies (Avise, 1975). For instance, Sévigny *et al.* (1989) used the information provided by glucose phosphate isomerase to distinguish between closely related species of the planktonic copepod *Pseudocalanus*. Although electrophoretic patterns were not useful for species discrimination due to shared alleles, genetic analyses (heterozygosity, allele frequency, private alleles) showed that organisms previously grouped into species based on subtle morphological differences were also genetically isolated. Better resolution was found for larval identification of three oyster species (Hu, Lutz and Vrijenhoek, 1992). However, protein-based approaches soon lost popularity in systematic studies due to several drawbacks such as the need to work with tissues that were either fresh or frozen and available in relatively large quantity (i.e., very small eggs or larvae could not be analyzed). Furthermore, as this technique only detects nonsynonymous substitutions, the revealed polymorphism was often low. Consequently, the advent of polymerase chain reaction (PCR) allowing the amplification of various genes from small amounts of tissue, either fresh or preserved in ethanol, led to a boost in molecular-based identification of organisms. Various methods have been developed, including DNA hybridization, species-specific PCR, random amplified polymorphic DNA, restriction fragment length polymorphism, single strand conformational polymorphic DNA and sequencing of PCR products, with their advantages and disadvantages (see Table 1 in Wong and Hanner, 2008). Of all these, sequencing

methods, providing access to the most accurate genetic information (i.e., the string of nucleotides), were soon to become the method of choice for species identification.

One of the early sequencing-based studies in marine species looked at a mitochondrial gene, cytochrome *b* oxidase (cyt-b), and found that four species of tuna could be distinguished based on these sequences (Bartlett and Davidson, 1991), while Medeiros-Bergen *et al.* (1995) successfully identified three holothurian species with other mitochondrial sequences (16S). Bucklin *et al.* (1999) sequenced yet another mitochondrial gene, cytochrome *c* oxidase subunit I (COI), in eight species from three genera of planktonic copepods and found the method to reliably discriminate even among sibling species. The authors acknowledged the need for a "rapid, simple, inexpensive and reliable" molecular protocol for marine species identification.

1.6 DNA barcoding for species identification and discovery

1.6.1 The concept: advantages and limitations

A ground-breaking approach to species identification was brought by Hebert *et al.* (2003) who proposed the use of a small fragment from the mitochondrial genome for species identification across phyla from the entire animal kingdom and coined the term "DNA barcoding" for this approach. Reasons for choosing mitochondrial (mtDNA) over nuclear DNA include uniparental inheritance (in a majority of animal phyla, but see Breton *et al.*, 2007), high evolutionary rate, lack of introns, large copy numbers in every cell, and limited recombination (but see Galtier *et al.*, 2009). The proposal of COI as the target gene for DNA barcoding was not an arbitrary choice since decades of research showed a useful phylogenetic signal for both above- and below-species level and that "universal" primers could recover the 5'end of COI in most animal phyla. According to the barcoding approach, species

could be identified based on a "barcoding gap" between intra- and interspecific genetic distances by using a threshold value of ≥3% (Hebert *et al.*, 2003) or a 10-fold value of mean intraspecific distance (Hebert *et al.*, 2004) for species delimitation.

Although numerous studies used molecular methods for species identification prior to the DNA barcoding era, it is still a unique concept with manifold attributes. Initially proposed only for animal taxa, a DNA-based identification system was soon found to be successful in land plants (Hollingsworth *et al.*, 2009), algae (Saunders, 2005), fungi (Seifert *et al.*, 2007), whether using only COI and/or other DNA regions (mitochondrial, plastid, nuclear) for better resolution. Besides the global scale involved, DNA barcoding brings a few major assets. It implies standardization (i.e., the same DNA fragment(s) used within a taxon), which allows comparisons between datasets of various researchers, revealing cases of synonymy, potential cryptic species or genetically distinct populations. Vouchers are permanently stored, ideally in a DNA-friendly manner, in museum collections, publicly accessible for future reference. This step is in contrast to most molecular studies conducted so far, which lack the possibility of specimen retrieval for sequences deposited in public databases (GenBank), therefore resulting in impossible taxonomic verifications and growing concerns about the documentation of scientific data (Pleijel *et al.*, 2008 and references therein). Vouchers can be stored under different forms (specimens, tissue, detailed photographs or stained slides for microscopy) and preservation methods (frozen, ethanol-preserved or dried specimens). DNA extracted from these vouchers is permanently stored in DNA banks available for future usage (e.g., inferring evolutionary patterns in different genes or proteins among taxa or habitats). The DNA Barcode of Life Data Systems (BOLD; http://www.boldsystems.org Ratnasingham and Hebert, 2007) provides a unifying protocol for data acquisition, storage and analysis. Data stored in BOLD include sampling details with GPS coordinates, images, taxonomic information, DNA barcodes, primer sequences, electropherogram "trace" files, and even detailed laboratory operations (with

protocols for each step and gel images) for specimens processed at the Biodiversity Institute of Ontario (BIO, http://www.biodiversity.uoguelph.ca). Above all, this database if freely accessible and all data can be downloaded after publication or analyzed directly in BOLD with distance-based methods. Future taxonomic updates are possible. These attributes make BOLD a more advantageous tool to use when dealing with DNA barcodes than GenBank (notorious for hosting erroneous data; Harris, 2003), proved by an eight-fold greater amount of barcodes produced at BIO and directly stored in BOLD (>650,000 barcodes) compared to GenBank (>90,000 barcodes) (February 2010).

Data scrutiny is vital since errors can occur at every step of the DNA barcoding protocol, from sampling in the field to COI amplification, leading to surprising results such as amphipods identified as decapods according to DNA barcodes (A. Radulovici, unpublished). Any evidence of misidentification, mislabelling, cross-contamination between samples due to leaked DNA in ethanol jars with mixed samples (Shokralla, Singer and Hajibabaei, 2010) or during COI amplification, other contaminations (e.g., human, mouse, bacteria) or pseudogenes (nuclear copies of COI), is routinely investigated in barcoding studies. Once through the cleansing step, DNA barcodes can be used in various analyses.

DNA barcoding was initially faced with great criticism (Will and Rubinoff, 2004; Will, Mishler and Wheeler, 2005; Rubinoff, 2006; Rubinoff, Cameron and Will; 2006) by people who feared that a universal DNA-based approach for species identification would gain exclusivity over traditional methods and taxonomists would go extinct while funding would be vacuumed by high-throughput facilities in order to provide "barcode-species" (i.e., species seen as strings of nucleotides). As with any other method, DNA barcoding has limitations, acknowledged by barcoders: low resolution in some cases (hybrids, recently diverged species, species complexes or slow evolving groups); the presence of pseudogenes (Song *et al.*, 2008); contaminants amplified with "universal" primers (Siddall *et al.*, 2009); or cases of

mitochondrial introgression (Kemppainen *et al.*, 2009) (see barcoding reviews, Frézal and Leblois, 2008 and Mitchell, 2008). Also, the functional group of many organisms is impossible to identify with DNA-barcodes. Thresholds have to be carefully considered due to variable mutation rate across taxa (Galtier *et al.*, 2009) or incomplete sampling of taxa (Meyer and Paulay, 2005; Ekrem, Willassen and Stur, 2007). Distance-based methods have been criticized and they are sometimes used in combination with character-based ones, but analytical tools are constantly being developed to incorporate the large body of information produced by DNA barcoding (Nielsen and Matz, 2006). Moreover, critics have been oriented towards a new "barcode-species" concept which will lead to an extreme amount of divergent clusters being arbitrarily raised to the species level (taxon over-splitting). On the other hand, reproductive isolation, the requirement for the popular biological species concept, is a very difficult investigation in marine systems. However, Gómez *et al.* (2007) tested this case in a cosmopolitan marine bryozoan and showed that divergent barcode clusters might indeed correspond to reproductively isolated groups, providing a link between DNA barcoding and the biological species concept.

Despite its limitations, DNA barcoding has become an appealing tool for biodiversity investigations, by identifying specimens during all life stages, from fresh or preserved material, and cases of sexual dimorphism or potential cryptic species. Non-specialists are able to have a fast (express-barcoding in less than two hours; Ivanova, Borisenko and Hebert, 2009), cheap and reliable identification tool with many practical and fundamental applications. Moreover, there is an international Consortium for the Barcode of Life (CBOL; http://www.barcoding.si.edu) dedicated to establish DNA barcoding as a standard tool for species identification. The largest project currently envisaged is the International Barcode of Life Project (iBOL, http://www.ibol.org), launched in October 2010, with the goal of acquiring DNA barcodes for 500,000 species by 2015.

1.6.2 Practical applications for the marine environment

In recent years, DNA barcodes have proved to be a valuable asset in identifying marine organisms, especially in the obvious cases where morphological identification is not possible, namely processed seafood. The famous example of fish sold as "red-snapper" in the US and actually consisting of other species in 77% of cases (cyt-b sequences; Marko *et al.*, 2004) was soon followed by other studies, which proved that seafood substitutions are common. The extent of this phenomenon on the global market of fresh, smoked or dried fish products varies across continents (Smith, McVeagh and Steinke, 2008; Wong and Hanner, 2008; Holmes, Steinke and Ward, 2009; Barbuto *et al.*, 2010) and the possible explanations include genuine mislabelling due to morphological similarities between closely related species or fraudulent substitution of expensive species with cheaper variants. An extreme case of fish substitution had drastic consequences for public health, leading to food poisoning due to puffer fish toxin and the consequent recall of products (Cohen *et al.*, 2009). With its power to reveal mislabelled products, DNA barcoding will have multiple implications from food safety and public health, to fisheries management (depletion of fish stocks) and conservation (protected species caught illegally).

Most marine organisms have larval stages difficult to identify based on morphological characters and DNA barcoding could have a great impact in this field, provided that a complete reference library for adults is developed (Barber and Boyce, 2006; Pegg *et al.*, 2006; Webb *et al.*, 2006,). Reliable identification of adults could have economic implications, for instance in aquarium fish trade regulations since many species originate in coral reefs (Steinke, Zemlak and Hebert, 2009), a highly threatened ecosystem. Moreover, routine DNA barcoding of marine organisms could identify invasive species (Saunders, 2009), with special importance in cases of partial specimens which lost their key diagnostic characters (Radulovici, Sainte-Marie and Dufresne, 2009).

1.6.3 Progress in DNA-based inventories of marine groups

Many marine taxa represent an ideal target for DNA barcoding due to a lack of reliable morphological characters for easy diagnosis. Marine algae represent such a group due to simple morphology, phenotypic plasticity and alternative heteromorphic generations, among other factors (Saunders, 2005). The same standard marker as for animals (COI) proved to work well in red algae and revealed the presence of an invasive species in Canadian waters (Saunders, 2009) as well as a large proportion of cryptic species (Saunders, 2008). Other invasive red algae with a negative impact on coral reefs were identified in Hawaii based on a multi-gene approach including COI (Conklin, Kurihara and Sherwood, 2009). Successful results with COI were shown in brown algae (McDevit and Saunders, 2009) but less so in green algae where other markers are being tested (G. Saunders, pers. comm.).

Diatoms represent a large component of the marine microbiota and another group where COI was not successful on large scale. A recent study including 114 diatom species found the internal transcribed spacer (ITS) to have 99.5% identification success (Moniz and Kaczmarska, 2010), a result that will surely lead to an increase in DNA-based inventories for this important marine group.

Due to low substitution rate in mtDNA, plant barcoding had a lower success rate compared to barcoding the animal kingdom. Alternative regions have been proposed and a final recommendation for a two-locus approach (plastid coding genes: matK and rcbL) has recently been made (Hollingsworth *et al.*, 2009). Consequently, seagrass species (e.g., *Zostera* spp., *Posidonia* spp.) with no reference in BOLD yet (February 2010), will soon be targeted by barcoding studies.

Sponges are an ancestral metazoan group with simple morphology but complex and important roles in marine ecosystems and pharmaceutical industry

(Worheide and Erpenbeck, 2007). Currently, this is the only invertebrate phylum to be barcoded through a global campaign (Sponge Barcoding Project, http://www.spongebarcoding.org), although a COI fragment downstream of the "Folmer" region was found to be more variable, and hence more appropriate for species identification in sponges (Erpenbeck, Hooper and Worheide, 2006).

Cnidarians (e.g., corals, sea anemones) and sponges constitute the most important components of coral reefs. COI seems to evolve too slowly in both groups, therefore lacking the power to reliably identify species. And while in sponges another COI fragment than the standard 5'end might be useful, cnidarian barcoding might need another gene (<2% interspecific divergences in scleractinian corals (Shearer and Coffroth, 2008) (Table 1.2). Moura *et al.* (2008) assessed the efficacy of 16S and showed that this gene could be a useful marker at the species and even population, genus and family levels in hydrozoans. Combining their own sequences with public ones from GenBank, the authors flagged problematic issues for hydroid systematics: potential cryptic species, conspecificity (low divergence between species) or cosmopolitan species consisting of species complexes. However, recent advances involving planktonic hydrozoans (Bucklin *et al.*, 2010) indicate that this group might actually be successfully COI barcoded.

Table 1.2 Levels of genetic divergence in marine taxa. Only studies using the 5' end of COI and giving average K2P genetic divergences were included. NoS: number of species barcoded; Intra: mean genetic distances within species; Inter: mean genetic distances between species

Marine group	NoS	Intra (%)	Inter (%)	Reference
Crustaceans				
Malacostracans	80	0.91[a]	13.6	Radulovici *et al.*, 2009
Decapods	54	0.46	17.16	Costa *et al.*, 2007
Copepods	24	0.75[b]	27.05	Bucklin *et al.*, 2010
Molluscs				
Heteropods	9	3.28	21.7	Jennings *et al.*, 2010
Pteropods	31	3.02	17.6	Jennings *et al.*, 2010
Corals	30	0.05	1.90	Shearer and Coffroth 2008
Chaetognaths	14	1.45	34.5	Jennings, Bucklin and Pierrot-Bults, 2010
Echinoderms	191	0.62	15.33	Ward, Holmes and O'Hara, 2008
Fishes	207	0.39	9.93	Ward *et al.*, 2005

[a]if deeply divergent clusters are removed, the mean value becomes 0.51%.
[b]mean intraspecific for the entire dataset (crustaceans, cnidarians, chaetognaths, one nemertean).

Molluscs represent the largest marine group with more than 50,000 described species (Table 1.1). One of the early studies to draw attention on the risks of using thresholds and incomplete sampling in barcoding approaches was tested on cowries, a very diverse and well-studied group of marine gastropods (Meyer and Paulay, 2005). Results showed that overlap between intra- and interspecific divergences might lead to large errors in species identification when a taxon is undersampled. Two species of intertidal gastropods were found to share haplotypes in NE Atlantic, potentially due to introgression or incomplete lineage sorting (Kemppainen *et al.*, 2009), while gastropod eggs from Philippines could not be identified to the species level due to a lack of comprehensive barcode databases (Puillandre *et al.*, 2009). Local-scale barcoding of species from four genera of Norwegian bivalves was a

successful case, although larger datasets are needed to prove the applicability of barcodes in identifying bivalves (Mikkelsen, Schander and Willassen, 2007). A barcoding study of planktonic gastropods (pteropods and heteropods) from six oceans revealed the highest average values (> 3%) for genetic distances between individuals of the same species reported to date (Table 1.2) (Jennings *et al.*, 2010). This is a strong indication that divisions below the species level (e.g., subspecies) might represent valid species and a taxonomic revision should be conducted.

Crustaceans are one of the largest (Table 1.1) and most diverse, morphologically and ecologically, marine groups. Playing important roles in marine food webs, crustaceans have representatives in all marine habitats. Costa *et al.* (2007) used their own sequence data and public data from GenBank to perform a large-scale analysis in crustaceans (150 species from 23 orders). Besides successful species identification (Table 1.2), their study revealed cases of potentially overlooked species and the need for taxonomic revisions (e.g., valid species that should be lumped). Taxon-specific barcoding studies were conducted on euphausiids (Bucklin *et al.*, 2007) and stomatopod larvae (Barber and Boyce, 2006). While the former could identify all specimens to the species level, the latter showed that a large part of stomatopod species from Indo-Pacific coral reefs is unknown as adults. Reef-associated crustaceans, mainly decapods, stomatopods and peracarids, from French Polynesia have been recently barcoded, revealing a large proportion of singletons (i.e., species represented by one specimen) living in *Pocillopora* dead heads (Plaisance *et al.*, 2009). While undersampling is usually the cause for a bias towards singletons, this study used a semi-quantitative sampling design to show that associated fauna in coral reefs is largely composed of low-abundance species. In addition, no species barcoded in this study had a match in GenBank, highlighting once more the need for comprehensive reference libraries. Radulovici, Sainte-Marie and Dufresne (2009) used a regional approach in barcoding malacostracan crustaceans from the Gulf of St. Lawrence and revealed the existence of an invasive amphipod species, *Echinogammarus ischnus*, which

expanded its distribution since previous studies. Cryptic speciation was not found to be common (5% of cases) but it might be a result of incomplete taxon sampling (80 species representing only 20% of the regional malacostracan fauna) or geographical scale.

A large barcoding study was conducted on echinoderms (191 species from five classes) by including also public data from GenBank (70% of the final dataset) (Ward, Holmes and O'Hara, 2008). Based on shallow intraspecific versus deep congeneric divergences (Table 1.2), a large amount of specimens (97.9%) could be assigned to known species. Those specimens that could not be assigned belonged to one genus, *Amblypneustes*, known to include morphologically and genetically similar species. Additionally, a few cases of potential cryptic species were recorded.

Smaller groups are also targeted in barcoding studies. For instance, sea spiders (Pycnogonida) were recently sampled as part of a marine inventory of the Ross Sea, Antarctica, and 25 species were identified based on morphological and molecular data (18S, 12S, 16S, COI) (Nielsen, Lavery and Lorz, 2009). Although statistics related to the level of genetic divergence were not provided by this study, a general concordance between barcode clusters and morphospecies was reported (only one case of misidentification or potential cryptic species) and no new species was revealed during the survey. However, with a larger geographic sampling for an abundant and circumpolar species, Krabbe *et al.* (2010) found multiple cryptic mitochondrial lineages, geographically restricted, within one nominal species. A much smaller group than sea spiders (see Table 2.1 in Bouchet, 2006), chaetognaths are mostly planktonic invertebrates with simple morphology but complex roles in the pelagic realm together with large distribution areas at the global scale. Successful identification can be performed with standard COI barcodes, even though the level of intraspecific variation is slightly higher than in other marine groups (Table 1.2) (Jennings, Bucklin and Pierrot-Bults, 2010).

A large and morphologically difficult group, therefore with underestimated diversity, but with potential roles as indicators of anthropogenic impact on marine systems, nematodes could greatly benefit from DNA barcoding (Table 1.1). So far, the 18S gene was found to amplify across many taxa and with 97% identification success (Bhadury *et al.*, 2006).

Parasites are very often excluded from marine faunal inventories. However, they are very common and play important roles in marine ecosystems by affecting population dynamics of their hosts. Therefore, a reliable identification system would be of great utility in community ecology (e.g., identifying all life cycles in different hosts) as well as for public health (e.g., human parasites). In the marine realm, a recent attempt to barcode parasites of intertidal species from New Zealand targeted a group of trematode species, all of which could be distinguished based on DNA sequences (Leung *et al.*, 2009). Although the authors chose to amplify a short DNA fragment downstream of the "Folmer" region, while the standard 5'end can generally be amplified in this group (Locke *et al.*, 2010), the study provided important ecological data on the trematode species analyzed with notes on new host-parasite interactions in intertidal mudflats.

Fishes are among the most studied marine groups and are currently barcoded within two global campaigns, FISH-BOL (http://www.fishbol.org) and SHARK-BOL (http://www.sharkbol.org) (Ward, Hanner and Hebert, 2009). One of the early studies on barcoding marine life looked at 207 fish species from Australia and showed that all could be discriminated based on their COI sequence, including five species of *Squalus* previously described but not formally named (Ward *et al.*, 2005). Other studies found barcoding to be useful in identifying fishes from Pacific Canada (Steinke *et al.*, 2009), North Atlantic (Ward *et al.*, 2008) or fish larvae from the Great Barrier Reef (Pegg *et al.*, 2006). When including shared species between distant geographical areas, DNA barcodes could be useful to test the relationship between distance and intraspecific variation. For instance, Ward *et al.* (2008) found only two

out of 15 species shared between North Atlantic and Australasia with deep intraspecific divergence (2.75% and 7.44%). On the other hand, Zemlak *et al.* (2009) showed that populations of commercial fish with inshore distribution in South Africa and Australia have high levels of genetic divergence (mean within species: 5.10%) and estimated that one third of the 1,000 shared species between these two regions are cryptic taxa. As a general remark, DNA barcodes were shown to be a powerful tool in discriminating marine fishes (98% success). Rare cases of incongruence were due to potential cryptic species or species complexes (deeply divergent intraspecific clusters), or to cases of hybrids, recent radiation, taxonomic over-splitting or morphological misidentification (shared haplotypes) (Ward *et al.*, 2009).

Sea turtles are represented by only seven species worldwide but are threatened across their entire distribution range, therefore DNA barcodes could be very useful in species conservation and wildlife forensics by identifying turtle meat and eggs illegally traded or carcasses stranded on beaches (Vargas, Araujo and Santos, 2009). Although sea turtles represent an ancient group with slow mutation rate, all species were successfully identified and no cryptic species was revealed based on genetic distances and character-based methods (Naro-Maciel *et al.*, 2009). Two recently radiated species showed the only interspecific distance below the threshold of 2–3% but even so, there was no overlap between intra- and interspecific values. Other marine reptiles, such as snakes, will be barcoded within a large iBOL project targeting all vertebrates (A. Borisenko, pers. comm.), while birds connected to the marine environment are already being barcoded within "All Birds Barcoding Initiative" (http://www.barcodingbirds.org).

The most studied and charismatic marine vertebrates (whales, dolphins and the other cetaceans), lack a comprehensive library of DNA barcodes. However, a newly established campaign, Mammalia Barcode of Life (http://www.mammaliabol.org), has as goal to provide DNA barcodes for all mammals by 2015, marine species as well.

DNA barcoding is a tool for species identification and discovery (by flagging divergent clusters) and modern taxonomy and systematics is increasingly incorporating COI sequences as additional data into their fields (Järnegren *et al.*, 2007; Krug *et al.*, 2007; Derycke *et al.*, 2008; Cardenas *et al.*, 2009; De Wit, Rota and Erseus, 2009). DNA barcodes might become a standard character to be included with species description and low sequencing prices will soon make this tool widely available to researchers from economically poor but biodiversity rich countries. Although we saw a multitude of cases arguing for potential cryptic species ("taxon-splitting"), there will definitely be cases of "taxon-lumping" revealed with a DNA-based approach. For instance, two lumpsucker species with different morphology were found to have identical sequences for multiple genes and to actually represent one sexually dimorphic species (Byrkjedal, Rees and Willassen, 2007). Moreover, DNA barcodes could be incorporated into large phylogenies (Kappner and Bieler, 2006; Larsson, Ahmadzadeh and Jondelius, 2008) or used for inferring preliminary phylogeographic patterns (Costa *et al.*, 2009).

1.7 Current status

1.7.1 How many marine barcodes?

We attempted to make a synopsis of marine groups that have been targeted by DNA barcoding by focusing on published data. Some of the papers reviewed here were contributions to the Marine Barcode of Life Project (MarBOL, http://www.marinebarcoding.org), a joint effort of CBOL and Census of Marine Life (CoML; http://www.coml.org) to provide 50,000 barcodes for marine species by mid-2010. Since the project is still in progress, only preliminary results are available at this moment. However, with more than 37,000 barcodes produced (MarBOL website, February 2010), the project is moving fast forward confirming the usefulness of such an approach for marine systems (Figure 1.3).

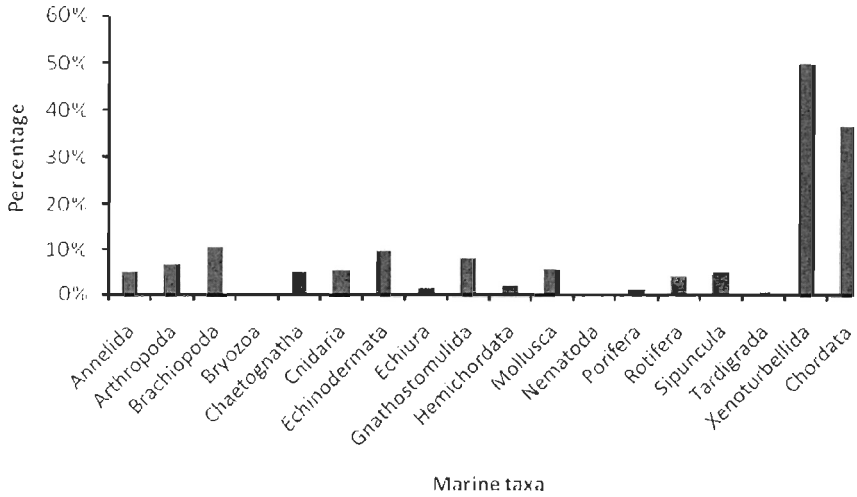


Figure 1.3 Proportion of barcoded species across marine animal taxa. (Data provided by D. Steinke, MarBOL coordinator)

There is a wealth of on-going case-studies in the marine realm that will be published in the near future (http://www.bolinfonet.org/casestudy; Taxonomy Browser in BOLD). Whether taxon-oriented (FISH-BOL, SharkBOL, Sponge Barcoding Project), nationwide (Canada, Australia, Norway, India) or locally focused on entire biota (Churchill, Moorea), targeting ecosystems (ReefBOL), ecoregions (Polar Barcode of Life) or multiple taxa from the entire marine environment (MarBOL), large-scale barcoding campaigns will provide a vast amount of information in need for accurate treatment and analysis.

A first glimpse at the Canadian case-study might suggest that marine biodiversity has been severely underestimated even in a marine non-hotspot area. First, there is an enormous amount of marine species, mostly invertebrates, collected in the past and still awaiting formal description and naming (only 48% of marine species classified; Mosquin *et al.*, 1995; Archambault *et al.*, 2010). Second, the opening of the Northwest Passage due to climate change will lead to new Arctic explorations, most likely ending with new faunal discoveries, especially in less-

known groups (e.g., polychaetes). Third, DNA barcodes indicate that cryptic speciation might take place even in well-known marine taxa (though to less extent) and geographical areas. For instance, DNA barcodes showed that one quarter of polychaete identified morphospecies actually consists of potential cryptic species when considering a nationwide scale with all three oceans, Atlantic, Arctic and Pacific, included (C. Carr, pers. comm.). Based on this result and knowing that there are at least 673 infaunal polychaetes for the three oceans (Archambault *et al.*, 2010), this would mean that around 840 species of polychaetes are present in Canadian waters alone. Cryptic speciation seems to be common in different groups of marine algae (G. Saunders, pers. comm.) but less so in fish (Steinke *et al.*, 2009) or marine crustaceans (Radulovici, Sainte-Marie and Dufresne, 2009). However, marine crustaceans include a wide variety of groups with different potential for dispersal (hence different potential to speciate) and once a nationwide scale is included and taxonomic input provided, crustaceans might likely exhibit various extents of cryptic speciation (Radulovici *et al.*, unpublished).

1.7.2 Special issues with marine taxa

Where are we now? Recent developments provide non-invasive DNA extraction with total voucher recovery (Porco *et al.*, 2010), as well as extraction of DNA leaked into the aquatic environment (Ficetola *et al.*, 2008) or ethanol (Shokralla, Singer and Hajibabaei, 2010). Primers are being developed for various taxa and additional markers or larger COI fragments used in cases of slow mutation rate (e.g., sponges, cnidarians). The BIO high-throughput facilities provide around 250,000 barcodes per year and that amount will double in the future (G. Singer, pers. comm.). We have the technological capacity to barcode the entire life, yet marine barcoding lags far behind the terrestrial counterpart (Figure 1.4). Why? The long-standing tradition of preserving marine material by using formalin, which prevents DNA amplification, represents a serious impediment in using museum specimens for DNA barcoding, in contrast to terrestrial taxa. Therefore, fresh material stored in

ethanol must be collected during sampling cruises, which are very expensive and usually focused on one or a few particular groups of marine organisms. These specimens have to be identified by trained taxonomists who are drastically decreasing in number. Moreover, most marine groups do not benefit from the help of amateurs, in contrast to terrestrial groups such as birds or butterflies. Consequently, a greater effort is inevitable when barcoding marine taxa.

1.7.3 Taxonomy and barcoding

At the moment we are unable to assess the impact of DNA barcoding on species diversity in terms of number of new species described as a result of this approach. The reason is simple: barcoding studies have the role to screen large sample sizes and flag cases of intraspecific deep divergence ("cryptic species"). However, the task of investigating further the extent of this phenomenon (additional genetic, ecological, behavioral data) culminating in a new species description does not belong to a barcoder but to a taxonomist. And since the number of taxonomists is rapidly decreasing (Packer *et al.*, 2009b) while marine barcodes are rapidly accumulating, the majority of flagged cases stop at the level of "potential cryptic species". Without a larger interest and involvement of highly trained taxonomists in marine barcoding studies, the advancement of the understanding of marine speciation will not be very rapid, potentially leading to another "tale of stupidity" (Boero, 2010).

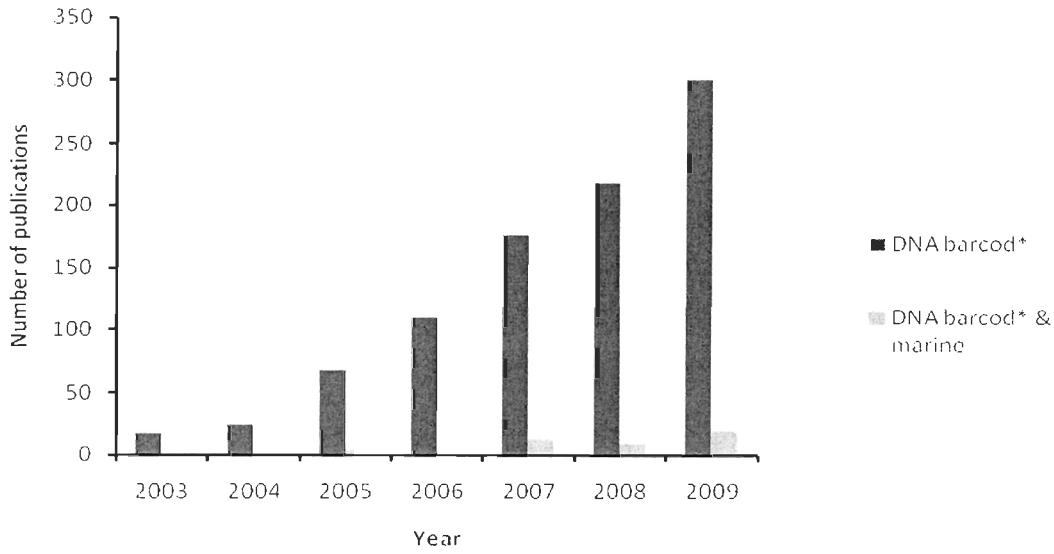


Figure 1.4 The amount of barcoding studies targeting marine systems ("DNA barcod*" and "marine" as keywords in Web of Science) relative to barcoding studies in general ("DNA barcod*").

1.7.4 Future directions

Most of the studies reviewed here did not flag a high amount of cryptic speciation but this discovery is contingent upon the scale of the studies. An increased geographic scale and the inclusion of groups with lower potential for dispersal will surely bring interesting results. Since a few cases of deep divergence have been found in fishes, the most popular marine group for barcoding, surveys of similar scales in understudied groups will be promising for species discovery.

New methods for sampling the deep-sea will lead to the discovery of many new species. Sampling expeditions with on-board laboratories might become commonplace. While most barcoding studies are still taxon-oriented, there are a few others opening new directions by targeting marine communities (e.g., zooplankton, Machida *et al.*, 2009; Bucklin *et al.*, 2010). DNA microarrays ("chips") will be developed for certain marine groups (Kochzius *et al.*, 2008), allowing reliable

identification of known species. Once reference libraries are completed, next generation sequencing will allow reliable identification of environmental samples (e.g., water, sediment) or species diet, with reverberations for studying the ecosystem level of biodiversity.

1.7.5 Species as currency for biodiversity

This review looked at reliable methods for biological identifications. But do we need species names? The idea that species might not represent equal parts of the global diversity ("some animals are more equal than others"; Warwick and Somerfield, 2008), resulted in alternative approaches for biodiversity assessments, for instance including the diversity of higher taxa (e.g., taxonomic distinctness rather than species diversity; Warwick and Clarke, 1995). Moreover, in functional ecology species names are not important but just the functional group (e.g., predator, prey). In this case, one might argue that barcodes are useless because they do not offer any functional information, while morphological characters (e.g., mouthparts in crustaceans) could be an indication of specimens' functional group and their role in ecosystems. Alternatively, at the genetic level of biodiversity, species names are not crucial. Clusters of DNA barcodes might be used in biodiversity surveys by using a phylogenetic diversity analysis (Faith, 1994; Faith and Baker, 2006). Therefore, we should take advantage of various methods, including classical taxonomy, for a holistic approach to biodiversity.

1.8 Conclusions

DNA barcoding is a unique concept with many innovative attributes undergoing continuous improvement. It is not the goal but the tool to be used in order to improve our understanding of the surrounding world. It is a fast, reliable and cheap method for species identification and discovery. It provides permanent tags unchanged

during taxonomic revisions. It will have multiple applications for marine life: identification of larvae, invasive species, cryptic species, new species, illegal trade of protected species, stock management, biodiversity assessments, ecosystem monitoring, revisions of certain taxa, inference of phylogenetic relationships, and phylogeographic and speciation patterns. Most of the studies reviewed here were published within the last two-three years and there was no sign that traditional taxonomy is being replaced by DNA barcoding, as once feared, but that they are complementary approaches. Not only that species are not seen as merely strings of nucleotides, but we are witnessing a renaissance of taxonomy due to the need (and curiosity) to understand how and why divergent barcode clusters are (if really) morphologically identical. As seen above, the apparent "failure" of DNA barcoding to identify species is mainly due to a lack of comprehensive reference libraries and taxonomists will play a vital role in completing such a global database. Millions of barcodes will soon be generated and new species revealed, in need for proper taxonomic description. Furthermore, as marine inventories are not carried out by taxonomist experts at museums but by trained personnel at university or governmental institutions, there is a pressing need to make a concordance between taxonomy and DNA barcoding. Therefore, taxonomy is far from being extinct.

Whether DNA barcoding with the plethora of global and local campaigns will succeed in meeting close deadlines (500,000 species by 2015) or not, remains an open question. During the last ten years, CoML had the objective to assess and explain the diversity, distribution, and abundance of marine life, contributing significantly to an understanding of the marine environment and the inhabitants of the global oceans. However, even with the amount of new information generated by CoML, it is only the beginning. DNA barcoding might be of great help in this direction, leading to a shift in our view of marine biodiversity, patterns and processes included. But above all, DNA barcoding provides data freely accessible to everyone. And even if computers and Internet access, needed to browse data in BOLD, are not yet a commodity in many countries, DNA barcoding represents the largest

experiment of open-access data sharing which could help decision making to preserve and protect marine biodiversity now and into the future.

Acknowledgements

We are grateful to the following people for sharing information with us: Dirk Steinke (BIO) provided data on the global progress of marine barcoding, graphically represented in Figure 1.3; David Porco (BIO) and Robert Jennings (UMB) sent us their papers ahead of print; Christina Carr (BIO) shared her interesting results on barcoding Canadian polychaetes and Gary Saunders (UNB) provided an update on barcoding marine algae; Alex Borisenko and Greg Singer (BIO) kindly provided details on new iBOL projects. We thank two anonymous reviewers for useful comments upon the manuscript. This work is a contribution to the Canadian Barcode of Life Network, as well as to the Canadian Healthy Oceans Network, both funded through the Natural Sciences and Engineering Research Council.

CHAPTER II

DNA BARCODING OF MARINE CRUSTACEANS FROM THE ESTUARY AND GULF OF ST LAWRENCE: A REGIONAL-SCALE APPROACH

Molecular Ecology Resources 2009, 9 (Suppl. 1), 181–187.

Adriana E. RADULOVICI[1], Bernard SAINTE-MARIE[2], France DUFRESNE[1]

[1] Département de Biologie, Université du Québec à Rimouski, Québec G5L 3A1, Canada
[2] Institut Maurice-Lamontagne, Pêches et Océans Canada, CP 1000, Mont-Joli, Québec, Canada G5H 3Z4

Author contribution
Concept: AER, BSM, FD
Sampling: AER, BSM
Laboratory operations: AER
Data analysis: AER
Manuscript writing: AER
Comments: BSM, FD
Funding: FD, BSM

2.1 Résumé

Les crustacés marins représentent un groupe taxonomique qui se caractérise par une forte diversité morphologique et écologique. Ils sont difficiles à identifier par les approches traditionnelles et nécessitent généralement l'aide de taxonomistes hautement qualifiés pour une identification certaine. La méthode rapide d'identification par le code-barres d'ADN, s'est révélée un outil très efficace pour l'identification des espèces, notamment pour de nombreux groupes de Métazoaires, y compris certains groupes de crustacés. Notre travail consiste ici à élargir la base de données d'ADN barcode par l'étude de 80 espèces de malacostracés provenant de l'estuaire et du golfe du Saint-Laurent. Les séquences d'ADN pour 460 spécimens ont été regroupées en groupements correspondant à des espèces morphologiquement connues dans 95% des cas. Les distances génétiques entre les espèces étaient en moyenne 25 fois plus élevées qu'au sein de chaque espèce. Une divergence intraspécifique élevée (de 3,78 à 13,6%) a été observée chez des spécimens appartenant à quatre espèces morphologiques, suggérant la présence d'espèces cryptiques. Par ailleurs, nous avons révélé la présence d'une espèce envahissante d'amphipode présente dans l'estuaire du Saint-Laurent. Cette étude confirme l'utilité de l'ADN barcode pour l'identification des crustacés marins.

Mots-clés: Crustacea; code-barres d'ADN; Golfe du St. Laurent; diversité d'espèces

2.2 Abstract

Marine crustaceans are known as a group with a high level of morphological and ecological diversity but are difficult to identify by traditional approaches and usually require the help of highly trained taxonomists. A faster identification method, DNA barcoding, was found to be an effective tool for species identification in many metazoan groups including some crustaceans. Here we expand the DNA barcode database with a case study involving 80 malacostracan species from the Estuary and Gulf of St Lawrence. DNA sequences for 460 specimens grouped into clusters corresponding to known morphological species in 95% of cases. Genetic distances between species were on average 25 times higher than within species. Intraspecific divergence was high (3.78–13.6%) in specimens belonging to four morphological species, suggesting the occurrence of cryptic species. Moreover, we detected the presence of an invasive amphipod species in the St. Lawrence Estuary. This study reconfirms the usefulness of DNA barcoding for the identification of marine crustaceans.

Keywords: Crustacea; DNA barcoding; Gulf of St. Lawrence; species diversity

2.3 Introduction

A biodiversity crisis has emerged in the last decades and we are confronted with the highest extinction rates since the formation of human society (Pimm *et al.*, 1995). Mitigation measures are needed but difficulties arise due to the unknown extent of biodiversity and spatial distribution of species assemblages. At the species level, the most investigated of biodiversity levels, it is generally agreed that only a small fraction of all species has been formally described, between 1.5–1.8 million out of an estimated 10 million (Wilson 2003). In the face of dwindling numbers of trained taxonomists, a fast identification method was needed to assist in species inventories. In this context, Hebert *et al.* (2003) proposed the use of a small fragment of mitochondrial DNA from the 5′-end of the cytochrome *c* oxidase subunit 1 (COI) gene as a reliable, quick and cost-effective identification system for the whole animal kingdom. Although the method faces strong criticism (Will and Rubinoff, 2004; Ebach and Holdrege, 2005; Will, Mishler and Wheeler, 2005), it has nonetheless proven effective in a variety of animal groups in both terrestrial and aquatic environments (Hebert *et al.*, 2004; Hajibabaei *et al.*, 2006; Clare *et al.*, 2007; Hubert *et al.*, 2008). However, the proposed threshold value of 3% COI sequence divergence for species delineation (Hebert *et al.*, 2003) may be problematic in some cases (Barber and Boyce, 2006; Burns *et al.*, 2008).

Diversity in the sea includes about 300 000 described species, a much smaller number than documented for the terrestrial realm (Gray 1997). However, marine faunal inventories fail to identify about one-third of specimens to the species level (Schander and Willassen, 2005) and the existence of cryptic species (Knowlton 1993, 2000; Etter *et al.*, 1999) creates another difficulty for biodiversity assessments. Crustaceans are an interesting target for DNA barcoding because they represent one of the most diverse metazoan groups from a morphological and ecological point of view. The subphylum Crustacea includes 52 000 described species divided into 849 families, 48 orders and six classes, but their estimated number is much higher

(Martin and Davis, 2001). There is no general agreement on crustacean systematics at higher classification levels (e.g. class) (Boxshall, 2007), and recent molecular phylogenies have challenged systematics at the family and genus levels (Englisch, Coleman and Wagele, 2003; Browne, Haddock and Martindale, 2007; Hou, Fu and Li, 2007). Morphological identification of crustaceans can be difficult, time-consuming and very often requires highly trained taxonomists. Previous work on crustaceans found DNA barcoding to be a useful tool for specimen identification in both marine and freshwater species (Bucklin *et al.*, 2007; Costa *et al.*, 2007).

This study builds on previous barcoding work on crustaceans by focusing on marine species from the Estuary and Gulf of the St. Lawrence River. This geographic region of Atlantic Canada is known for its complexity, having such a wide range of physiographic, oceanographic and bathymetric characteristics that Brunel, Bossé and Lamarche (1998) divided it into 20 biogeographical zones. Although some 770 crustacean species are known from the Estuary and Gulf (Brunel, Bossé and Lamarche 1998), we chose to focus mainly on amphipods and decapods. The former represents the most speciose crustacean order and is an important component of marine food webs. The latter includes species (lobster, shrimp and crabs) that are important economically in providing large harvests and high income to Atlantic Canada, and ecologically as top predators in the marine benthic ecosystem. Our study adds to existing databases a large number of specimens sampled across a vast geographical area for a better representation of intraspecific variation. DNA barcodes reported in this study represent permanent species tags that will not change during taxonomic revisions.

2.4 Material and methods

2.4.1 Samples

We used 507 crustacean specimens collected in the Estuary and Gulf of the St. Lawrence River in 2000 (7 specimens) and between 2005 and 2008 (500 specimens) (Figure 2.1). The specimens represented 87 described species in 60 genera, 39 families and 5 orders (Amphipoda, Decapoda, Euphausiacea, Isopoda, Mysida) of a single class (Malacostraca). Deep-water specimens were collected during trawl surveys conducted by Fisheries and Oceans Canada (DFO), while littoral specimens were collected at low tide using dip nets and baited traps. Samples were stored in 100% ethanol (2005–2008) or in 70% ethanol (2000). Morphological identifications were done by experts or followed available keys for North Atlantic amphipods (Bousfield 1973), decapods (Squires, 1990), isopods (Schultz, 1969), mysids (Brunel, 1960) and euphausiids (Mauchline, 1971). Scientific names followed the Integrated Taxonomic Information System (www.itis.gov) and the list of McLaughlin *et al.* (2005). In most cases, the whole specimen was stored as a morphological voucher for future reference. For a few large decapod species, we obtained only tissue (legs or abdominal muscle) for barcoding and we stored these samples as tissue vouchers. However, additional specimens of each of these decapod species have been stored as proper morphological vouchers. In a few juvenile amphipods and crab larvae, no voucher could be preserved due to their very small body size, but photographs were taken prior to DNA extraction. All details regarding taxonomy, vouchers and collection sites with geographical coordinates can be found on the Barcode of Life Data System website (BOLD, www.barcodinglife.org) under the "Crustaceans of the St. Lawrence Gulf" project (WWGSL) by following "View all records" – "Specimen Page" (Ratnasingham and Hebert, 2007). In order to ensure geographical coverage for DNA barcodes, when possible, we included multiple specimens (at least two per site) from different geographical areas of the Gulf of St Lawrence (e.g. North Shore vs. Southern Gulf).

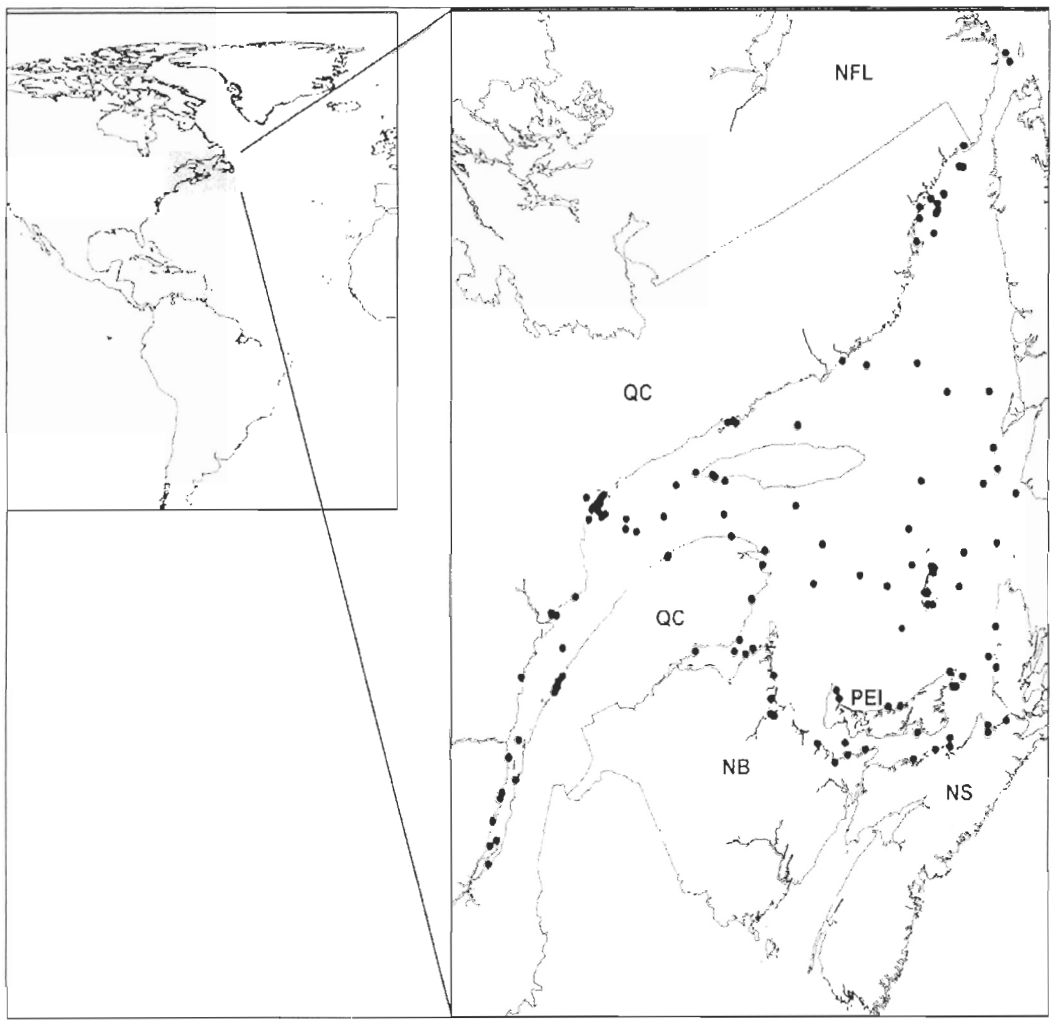


Figure 2.1 Distribution map for all sampling sites within the Estuary and Gulf of the St. Lawrence River. Canadian provinces surrounding the study area: Québec (QC), New Brunswick (NB), Nova Scotia (NS), Prince Edward Island (PEI), Newfoundland and Labrador (NFL).

2.4.2 DNA extraction, amplification, sequencing

Laboratory operations were carried out at the Canadian Centre for DNA Barcoding (CCDB), University of Guelph. Total genomic DNA was extracted from small amounts of tissue (1-mm^3 muscle tissue or whole legs for small specimens) by using an automated silica-based protocol with glass fibre filtration plates (Ivanova, Dewaard and Hebert, 2006). The barcode region was amplified with alternative sets of primers depending on the reaction success: LCO1490/HCO2198 (Folmer *et al.*, 1994) with M13 tails, CrustDF1′ (5-GGTCWACAAA YCATAAAGAYATTGG-3′) – CrustDR1 (5′-TAAACYTC AGGRTGACCRAARAAYCA-3′) (D. Steinke, University of Guelph, in preparation) and CrustF1/HCO (Costa *et al.*, 2007). All primer sequences can be found on the BOLD website within the project WWGSL ("View all records" – "Sequence Page" for each specimen). The polymerase chain reaction (PCR) was performed in 12.5 µL volume containing 2 µL H_2O, 6.25 µL 10% trehalose, 1.25 µL 10× PCR buffer, 0.625µL MgCl2 (50 mm), 0.0625 µL d NTPs (10 mm), 0.06 µL Platinum Taq polymerase (Invitrogen), 0.125 µL of each primer (10 µm) and 2 µL DNA template. PCR thermal conditions included: 1 min at 94°C, five cycles of 94°C for 40 s, 45°C for 40 s and 72°C for 1 min, followed by 35 cycles of 94°C for 40 s, 51°C for 40 s and 72°C for 1 min, and a final step of 72°C for 5 min. PCR products were visualized on 96-well precast 2% agarose gels (Invitrogen E-Gel 96 system) and bidirectionally sequenced with BigDye version 3.1 on an ABI 3730xl DNA Analyser (Applied Biosystems). Primers used for sequencing depended on those used for amplification, namely M13F/M13R, CrustDF1/CrustDR1 or CrustF1/HCO. Additional details about laboratory protocols for each step are available from the CCDB website (www.dnabarcoding.ca).

2.4.3 Data analysis

DNA sequences were aligned with SeqScape version 2.1.1 (Applied Biosystems) and manually checked for ambiguities. DNA sequences, as well as trace files, are available on the BOLD website within the project WWGSL ("View all records" – "Sequence Page" for each specimen) and on GenBank (Accession nos FJ581463 – FJ581922). A BLAST search including one sequence per species was performed on GenBank (megablast algorithm). The Kimura 2-parameter (K2P) model for base substitution (Kimura, 1980) was used in analyses on the BOLD website to obtain pairwise genetic distances. A neighbor-joining tree (NJ) based on K2P distances was also built in BOLD for a graphic representation of intraspecific distances. MEGA 4 (Tamura *et al.*, 2007) was used to test the NJ tree by bootstrap analysis with 1000 replications. Genetic distances between specimens were calculated for each taxonomic level with the "Distance Summary" command implemented by BOLD. Cases of intraspecific divergence higher than 3% were considered as potential cryptic species.

2.5 Results

Amplification failed in the seven specimens stored in 70% ethanol, representing the amphipods *Dyopedos monacanthus* (*n* = 1), *Gammarellus homari* (*n* = 1), *Gammarus fasciatus* (*n* = 1), *Gammarus lacustris* (*n* = 2), and *Jassa marmorata* (*n* = 2). Consequently, successful amplification of the barcode region was not obtained for five of the 87 species studied here. The remaining 500 specimens yielded a positive amplification of COI. Short or low-quality sequences (double peaks, background noise) obtained from 36 specimens and possibly representing pseudogenes were discarded. Only 25% of our sequences had matches in GenBank due to the fact that most species in our study had not been COI-sequenced before. Additionally, the amphipod *Stegocephalus inflatus* (*n* = 2) and the isopod *Calathura*

brachiata (*n* = 2) did not match crustacean COI sequences, possibly due to contamination. One discrepancy appeared between our morphological identifications and GenBank: COI sequences of amphipod specimens in poor condition that we morphologically identified as *Marinogammarus obtusatus* matched those of the invasive species *Echinogammarus ischnus* from GenBank.

The database resulting from this study includes DNA sequences for 460 specimens belonging to 80 species in 56 genera. The number of COI sequences per species varied between 1 and 29 with a mean of 5.75. The 658-bp COI fragment had 432 variable sites and 226 conserved sites, while 419 sites were parsimony-informative. Ambiguities were present in a few cases but they did not change the final result. The mean intraspecific divergence was 0.91% while the maximum reached 13.6% (Appendix A). By contrast, the minimum interspecific distance was 2.81%, between *Hyas araneus* and *H. coarctatus* (Apendix 2.1). The two levels of variation, namely within and between species, showed a small overlap (Figure 2.2). Morphological species were represented by individual clusters containing highly similar sequences in 95% of cases. However, four cases of deep intraspecific divergence, greater than 3%, were observed and the respective clusters were considered to be potential cryptic species (Table 2.1; Figure 2.3). With these clusters removed, the mean intraspecific divergence is 0.51%. The crab larvae sequenced in this study matched *Chionoecetes opilio* sequences, a result confirmed by rearing a few larvae in the laboratory.

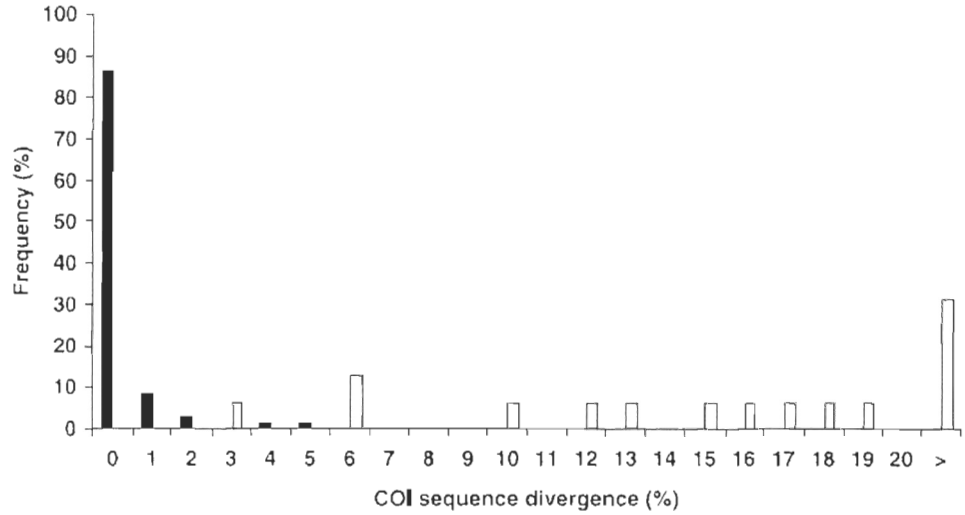


Figure 2.2 Frequency distribution of mean divergences for COI sequences for 80 species of malacostracan crustaceans from the Gulf of St. Lawrence. Two taxonomic levels are represented: species (solid bars) and genus (shaded bars). For maximum intraspecific divergences higher than 3%, see Table 2.1.

Table 2.1 Crustacean species with maximum intraspecific COI sequence divergences higher than 3%

Species name	Maximum intraspecific divergence (%)	Putative number of cryptic lineages	Maximum intra-lineage divergence (%)	Bootstrap support for each cryptic lineage
Ampelisca eschrichtii	13.6	2	0; 0.61	99; 99
Ischyrocerus anguipes	4.24	2	1.39; 2.17	94; 99
Neomysis americana	3.78	2	0; 0.45	99; 99
*Spirontocaris spinus**	6.91	3	0.5; 1.07; —	98; 99; —

*This species has three lineages, one represented by a single specimen (therefore, no pairwise comparison and no bootstrap support).

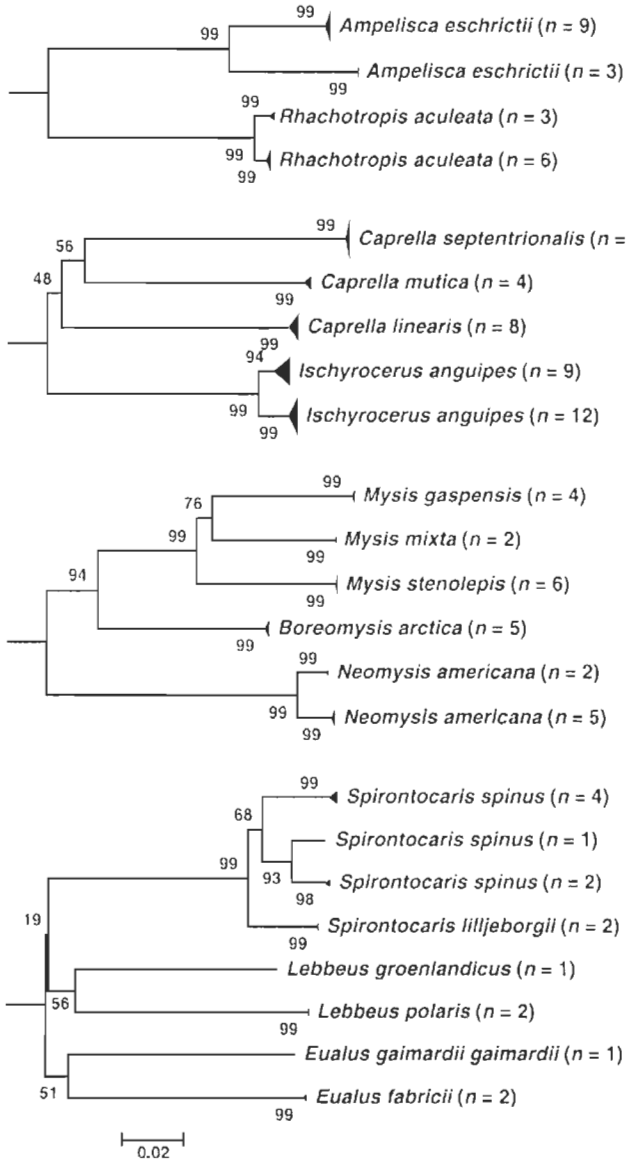


Figure 2.3 Branches of the neighbor-joining tree highlighting the four species complexes (and related species) found for malacostracan crustaceans from the St. Lawrence Gulf. Bootstrap values based on 1000 replications are included.

2.6 Discussion

This study further supports the validity of DNA barcoding for species identification in marine crustaceans. The ratio of interspecific to intraspecific variation (25×) was much higher than the threshold (10×) proposed by Hebert *et al.* (2004) as a species boundary. Therefore, assigning specimens to species was usually straightforward with no overlap between intra- and interspecific distances (95% of cases).

In four morphological species COI sequences grouped into 2–3 clusters that diverged by at least 3% (Table 2.1; Figure 2.3), suggesting either the presence of cryptic species or nuclear mitochondrial pseudogenes (numts). A growing concern regarding numts and DNA barcoding is that, if undetected, numts might lead to an overestimation of species richness (Song *et al.*, 2008). In crustaceans, numts have been found to diverge from the COI gene by up to 18.8% (Williams and Knowlton, 2001). To investigate the possibility of having amplified numts, we used a few steps suggested by Song *et al.* (2008). We found no stop-codons (quality control tool on BOLD) or indels, the sequences were of high quality, had the expected length (658 bp), matched COI sequences in GenBank, and the proportion of adenine-thymine did not differ strikingly among lineages. Moreover, intraspecific clusters were not related to geography. Consequently, we suggest that the amphipods *Ampelisca eschrichtii* and *Ischyrocerus anguipes*, the mysid *Neomysis americana* and the decapod *Spirontocaris spinus* represent species complexes. Classical taxonomy has already inferred the existence of species complexes in North American *Ampelisca* spp. and *I. anguipes* based on the existence of size morphs or subtle differences in morphology (Kaim-Malka, 2000; King and Holmes, 2004; references therein). Additional taxonomic, ecological and molecular work is required to investigate the full extent of cryptic speciation in crustaceans from the Gulf of St. Lawrence, as DNA barcoding can only serve to flag such cases.

The smallest divergence between species was 2.81% in *Hyas araneus* and *H. coarctatus*, two species that are morphologically distinct from the larval stages to adulthood but genetically close (Hultgren and Stachowicz, 2008). This finding is in agreement with other cases of DNA barcoding difficulties for arthropod identification (Barber and Boyce, 2006; Burns *et al.*, 2008), suggesting once more that the 3% cut-off in sequence divergence is not always applicable and that caution must be exercised in cases of incomplete lineage sorting.

Practical applications of DNA barcoding of crustaceans include detection of invasive species, substitution in processed seafood and estimation of stock size of harvested species based on larval abundances (Costa *et al.*, 2007). We report here the presence of an invasive amphipod, *Echinogammarus ischnus*, in the St. Lawrence Estuary near Berthier-sur-Mer. This species has spread from its native Ponto-Caspian region into Western Europe and the Great Lakes of North America. In Canada, it has been previously reported along the St. Lawrence River upstream from Montréal (Palmer and Ricciardi, 2004) and the present study confirms its north-eastern expansion. This species was identified as the morphologically similar *Marinogammarus obtusatus* based on specimens in poor condition, but all sequences matched those of *E. ischnus* determined in a previous phylogeographical study (Cristescu *et al.*, 2004). Without these reference sequences, our error might have gone unnoticed, thus emphasizing the importance of classical taxonomy to barcoding. Reciprocally, this example also stresses the success of DNA barcoding in rapidly detecting invasive species.

The 80 species sequenced in the present study represent only 20% of about 400 species inventoried within the Estuary and Gulf of the St Lawrence River (Brunel, Bossé and Lamarche, 1998) for the five malacostracan orders represented here. Some 20 other amphipod species were not included due to uncertain morphological identifications. Full taxonomic coverage of the known crustacean species from the Estuary and Gulf is hampered by sampling difficulties. Indeed,

except for decapods of economic importance (60% sequenced), other malacostracan species are not targeted by regular sampling surveys and seldom show up as by-catch. Moreover, for some taxa (e.g. amphipods), the use of dip nets, baited traps or bottom trawls will lead to a sampling bias towards highly mobile species. Therefore, the fraction of species diversity representing the most common (Brunel, Bossé and Lamarche, 1998) and most mobile (Sainte-Marie and Brunel, 1985) forms was explored in this study. There are two avenues to create a comprehensive database for the Gulf crustaceans in the future: fund research cruises targeting rarer crustaceans; and/or technological advances for high-throughput DNA extraction from formalin-preserved crustaceans. Exploiting museum collections, one of the goals of DNA barcoding, is a difficult task when working with crustaceans due to the traditional use of formalin which negatively affects DNA recovery. Consequently, barcoding studies are most successful when performed on groups that can make use of museum "dry" collections (e.g. insects, birds, mammals). There is no global campaign yet to barcode all crustacean species (or at least Malacostraca) as exists for other animal groups (e.g. fish, birds, lepidopterans); however, building regional databases throughout the world will bring us closer to understanding crustacean diversity. In summary, DNA barcoding is a very useful tool for the identification of malacostracan crustaceans by assigning unknown specimens to known species, insofar as species assignations in GenBank are reliable. DNA barcoding may lead to species discovery by flagging cryptic species, although more data than COI sequences are necessary for describing new species. However, based on DNA barcoding of the most common species at the regional scale of the Estuary and Gulf of St. Lawrence, cryptic species do not appear to be very common.

Acknowledgements

This research was supported through funding to the Canadian Barcode of Life Network from NSERC, Genome Canada (through the Ontario Genomics Institute) and other sponsors listed at www.bolnet.ca. Financial support for fieldwork in the Magdalen Islands was provided by Centre de recherche sur les milieux insulaires et maritimes. We are indebted to the following people for providing specimens: Mikio Moriyasu and Marcel Hébert (DFO Gulf Region), Diane Archambault and the crews on "Teleost" and "Calanus", Valérie Bélanger and Gesche Winkler (Institut des sciences de la mer de Rimouski), Annick Drouin (Laval University), the Saunders group (University of New Brunswick, Fredericton), and also to Pierre Brunel (Université de Montréal), Robert Chabot (Université du Quebec à Rimouski) and David Wildish (DFO Maritimes Region) for providing taxonomic support. We thank Traian Brad for valuable field assistance. We are grateful to the CCDB team in Guelph, especially to Janet Topan for laboratory assistance, Natalia Ivanova for help with laboratory protocols and SeqScape, the IT support team, Dirk Steinke for providing primers, Mehrdad Hajibabaei for access to the CCDB facilities and Paul Hebert for his interest in this project. We acknowledge three anonymous reviewers for providing helpful comments which improved the quality of this manuscript.

CHAPTER III

CRYPTIC DIVERSITY AND EVOLUTIONARY INSIGHTS IN NORTH ATLANTIC TALITRIDS AS INFERRED FROM DNA BARCODES

Adriana E. RADULOVICI[1], David J. WILDISH[2], Marcos PÉREZ-LOSADA[3], Sara E. LECROY[4], France DUFRESNE[1]

[1] *Département de Biologie, Université du Québec à Rimouski, Rimouski G5L 3A1 (QC), Canada*
[2] *Atlantic Reference Centre, Huntsman Marine Science Centre, St. Andrews E5B 2L7 (NB), Canada*
[3] *Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal*
[4] *University of Southern Mississippi, Gulf Coast Research Laboratory, Ocean Springs 39564 (MS), U.S.A.*

(Article in preparation)

Author contribution

Concept: AER, DJW
Sampling: AER, DJW, SEL
Laboratory operations: AER
Data analysis: AER, MP-L
Manuscript writing: AER
Comments: DJW, MP-L
Funding: FD, DJW, SEL

3.1 Résumé

Les Talitridae représentent une grande famille d'amphipodes (plus de 200 espèces décrites et plus de 2,000 espèces estimées), répartis le long des rives de la plupart des continents. Ils appartiennent à la seule famille d'amphipodes qui ait colonisé la terre. Ils sont artificiellement divisés en plusieurs groupes écologiques sans aucun soutien phylogénétique (espèces qui s'enfouissent sous du varech échoué, espèces palustres, espèces qui s'enfouissent sous le sable, espèces terrestres). Nous élaborons dans ce chapitre, une bibliothèque de référence de code-barres d'ADN (séquences de cytochrome *c* oxydase 1, COI) pour les Talitridae de l'Atlantique du Nord, principalement présents le long de la côte nord-américaine (Golfe du Mexique et Golfe du Saint-Laurent) ainsi que pour quelques espèces européennes. L'analyse phylogénétique (inférence bayésienne, maximum de vraisemblance) de 218 séquences d'ADN a révélé la présence d'espèces cryptiques nord-américaines. De plus, certains genres semblent non monophylétiques et le caractère polyphylétique des groupes écologiques est renforcé. La liste initiale de 15 espèces morphologiquement définies a été étendue à 24 espèces supposées, principalement par la découverte de trois complexes d'espèces (*Platorchestia platensis, Orchestia grillus, Tethorchestia* sp. B). La spéciation cryptique suit essentiellement un modèle allopatrique (sauf pour *O. grillus*) et certains de ces groupes sont soutenus comme 'espèces nouvelles' par des preuves morphologiques. Des recherches à venir devront inclure du matériel des localités types, afin de clarifier la position phylogénétique des "vraies" espèces morphologiques. Le séquençage d'autres gènes mitochondriaux et nucléaires ainsi que l'ajout de taxons supplémentaires seront nécessaires pour une analyse complète des relations phylogénétiques au sein des Talitridae.

Mots-clés: diversité cryptique; code-barres d'ADN; évolution; Atlantique du Nord; Talitridae

3.2 Abstract

Talitridae represents a large family of amphipods (>200 species described, >2,000 species estimated) distributed along the shores of most continents and the only amphipod family that has colonized the land. They are artificially divided into ecological groups (wrack, palustral, sand-burrower, land-hopper) with no phylogenetic support. Here we build a reference library of DNA barcodes (cytochrome *c* oxidase 1, COI) for talitrids from the North Atlantic, mainly from the North American coast (Gulf of Mexico, open-Atlantic coast, Gulf of St. Lawrence), together with a few species from Europe. A total of 218 DNA sequences indicated the presence of cryptic species in North American talitrids while phylogenetic analyses (Bayesian inference, maximum likelihood) showed some genera to be non-monophyletic and reinforced the polyphyletic character of the ecological groups. The initial list of 15 morphologically defined species was extended to 24 putative species mainly by discovering three species complexes (*Platorchestia platensis*, *Orchestia grillus*, *Tethorchestia* sp. B). Cryptic species were geographically separated (except for *O. grillus*) and some of these clusters were supported as new species by morphological evidence. Further directions should include barcoding of material from type localities, in order to clarify which cluster is the "real" morphological species. Additional mitochondrial and nuclear genes, as well as more taxa, are needed for in-depth analysis of phylogenetic relationships within Talitridae.

Keywords: cryptic diversity; DNA barcoding; evolution; North Atlantic; Talitridae

3.3 Introduction

Amphipods represent a highly diverse crustacean order but only one family, the Talitridae Rafinesque, 1815, was successful in colonizing the terrestrial environment. Talitrids are distributed worldwide along coastlines, in freshwater, brackish and marine habitats, as well as on land, in grassland and wet forests at low and high altitudes. Due to their body modifications for hopping, they are commonly known as "hoppers". Presumably evolving from aquatic ancestors during the Cretaceous (Bousfield, 1984, but see Conceição, Bishop and Thorpe, 1998), talitrid diversification was probably stimulated by the appearance of long coastlines following the break-up of the supercontinents, and it was more accentuated in tropical and temperate regions of Tethyan and Gondwanan origin than in Laurasian successors.

Current amphipod taxonomy at higher levels (e.g., subfamily, family, and superfamily) is unsatisfactory and without a phylogenetic basis. Families are usually presented alphabetically and higher taxonomic levels are under continuous revision with talitrid systematics following the same dynamic trend. Currently, taxonomic levels above Talitridae include superfamily Talitroidea (consisting of four families: Talitridae, Hyalidae, Chiltoniidae and Dogielinotidae), infraorder Talitrida and suborder Gammaridea (Serejo, 2004). Within talitrids, genera are frequently being split into multiple taxa and currently amount to 52 genera (Serejo and Lowry, 2008). As a proxy for talitrid classification, Bousfield (1984) proposed the use of four "systematic-ecological (polyphyletic and overlapping, but pragmatically useful) units": 1) palustral species (marsh-hoppers) with primitive morphology, semi-aquatic in marine, estuarine (salt marshes, swamps) and freshwater habitats in tropical and temperate regions; 2) beach fleas (beach-hoppers) with more advanced morphology but no capability to engineer their substrate, mostly semi-terrestrial, supralittoral in rocky and sedimentary habitats but terrestrial in coastal rain forests, from tropical to

boreal shores; 3) sand-hoppers with highly specialized morphology for substrate engineering (i.e., burrowing in sand), semi-terrestrial, supralittoral on sandy beaches, from tropical to boreal shores; 4) land-hoppers with advanced morphology for terrestrial life, that usually do not engineer substrate, present in forest leaf litter of coastal and high-altitude rain forests, in tropical and temperate regions. Although widely used in the literature, these lumped non-monophyletic groups create difficulties in inferring talitrid evolutionary history. For this reason we have followed the strictly ecological classification in Wildish (1988). Believed to have undergone strong adaptive radiation due to their high species richness and endemicity (Serejo, 2004), land-hoppers are of unknown origin and several scenarios have been proposed, including evolution from primitive beach fleas (Bousfield, 1984) to palustral ancestors (Lindeman, 1991). Moreover, the process of colonizing land probably included multiple events (Wildish, 1988). In addition to the previous morphology-based phylogenies, restricted in taxa and geographic coverage, limited effort has been put into resolving genetic relationships of talitrids. Focused on European fauna, mainly UK and the Mediterranean basin, genetic investigations have been conducted on a total of nine species using both allozymes (Conceição, Bishop and Thorpe, 1998; De Matthaeis, Davolos and Cobolli, 1998) and DNA sequences (Tafani *et al.*, 2004; Davolos and Maclean, 2005).

Aquatic talitrids play important ecological roles by decomposing plant material cast up on shores and provide a food source for other invertebrates (e.g., insects, spiders, crabs) and vertebrates (e.g., shore birds, mammals), although their narrow zonal distribution might limit their role at the ecosystem scale (review in Wildish, 1988). Many species reach high densities and biomass, representing a dominant component of wrack in the supralittoral zone, the ecotone between marine and terrestrial environments. Moreover, some species have been proposed as biological indicators for heavy metal contamination (Ugolini *et al.*, 2004) and quality of sandy beaches suffering from anthropogenic activities such as tourism (Ketmaier, Scapini and De Matthaeis, 2003). Amphipods (as all peracarid crustaceans) are direct developers, with eggs hatching into juveniles inside the brood pouch of females.

Lacking larval stages (which are usually highly dispersive) and inhabiting a narrow supralittoral zone with patches of suitable and unsuitable habitat, talitrids are believed to use passive dispersal (rafting, phoretic associations with birds or mammals, and human-mediated transport in ballast water) more than active methods (e.g., hopping, swimming) for dispersing at various spatial scales (Wildish, 1988).

The total number of extant talitrids is unknown although estimates as high as 1,000 land-hopper species have been mentioned (Bousfield, 1984). If the estimated number of land-hoppers represents half of all talitrid species, as the current ratio for described species suggests, it follows that around 2,000 talitrid species might exist on Earth, a number an order of magnitude higher than all currently described species (>200; Bousfield, 1984; Serejo and Lowry, 2008). Regardless of the precision of these estimates, there is definitely a large number of talitrid species still to be discovered and described.

As with many other amphipods, talitrids are difficult to identify based on morphological characters, especially in juvenile and immature stages. In the field, some hoppers can be identified based on their epidermal pigment pattern, considered to be less variable within than among species, although parasitic infestation can affect the color (LeCroy, 2010). Moreover, the pigmentation is not preserved in some storage liquids (e.g., alcohol). Such a diverse and taxonomically difficult group would benefit greatly from DNA barcoding. This method has been proposed as a fast, reliable and cost-effective method for animal species identification by using a fragment of the mitochondrial gene cytochrome *c* oxidase 1 (COI) (Hebert *et al.*, 2003). DNA barcoding has been successfully tested in a variety of marine groups (reviewed in Radulovici, Archambault and Dufresne, 2010), including crustaceans (Costa *et al.*, 2007; Radulovici, Sainte-Marie and Dufresne, 2009; da Silva *et al.*, 2011), and it has been used together with morphological characters to detect and describe a new talitrid species from Taiwan (Cheng *et al.*, 2011). However, taxonomic or regional inventories of talitrids based on DNA

barcoding are still lacking.

Here we focus on talitrid fauna from the East Coast of North America, from the Gulf of St. Lawrence (GSL) to the Gulf of Mexico (GOM), including also selected species from the Caribbean and Europe (UK), spanning various biogeographical provinces, each with a different glaciation history. There is an increase in species and genus richness from North to South, with three genera and five species currently recognized for Atlantic Canada (Brunel, Bossé and Lamarche, 1998; Bousfield, 1973), increasing to six genera and eight species in GOM (LeCroy *et al.*, 2009), while many more species occur in the Caribbean, some of which have been collected decades ago and still await formal description (Bousfield, 1984). In addition, a decrease in body size from north to south has been documented in two talitrid species (Wildish *et al.*, 2011). The known North Atlantic talitrid checklist includes generalists and specialists, endemic and cosmopolitan species, good and poor dispersers, species with various habitat requirements and salinity tolerances. Because there are no native land-hoppers in North America (Bousfield, 1982), this group has been excluded from our study, while the term "talitrid hoppers" is employed as a general label for the other three ecological groups (wrack, sand-burrowing and palustral species). Our goals were to: i) establish a barcode reference library for North-Western Atlantic (NWA) talitrid hoppers; ii) uncover the level of cryptic diversity; and iii) infer phylogenetic relationships among North Atlantic talitrids.

3.4 Material and methods

3.4.1 Sample collection

Talitrids were collected on various sandy beaches, rocky shores and marshes along the East Coast of North America, from GSL to GOM and the Caribbean (Table

3.1). A few species from Europe (UK) were included as well in order to infer phylogenetic relationships within the genus *Orchestia* Leach, 1814 (Figure 3.1). As outgroup to talitrids, we chose *Parhyale fascigera* Stebbing, 1897 from the closely related family Hyalidae Bulycheva, 1957. Samples were collected with pitfall traps or by hand and immediately fixed and then stored in 95% ethanol. Whenever possible, we tried to collect multiple specimens per species (to reveal the intra- versus interspecific genetic variation) and from multiple geographic localities (to gain insights into the geographic variation of DNA barcodes).

Regarding the salinity tolerance and ecological habitats occupied by talitrids (Table 3.2), we included species with a wide range of salinity tolerance from freshwater to brackish and marine species, and species representing three out of four "systematic-ecological" groups of Bousfield (1984). All specimens were identified to the species level based on morphological characters according to available taxonomic keys for North America (LeCroy, 2010; Bousfield, 1973) and the nomenclature followed the "World Amphipoda Database" (Lowry, 2010) available in the World Register of Marine Species (WoRMS; http://www.marinespecies.org). Voucher specimens were stored for future reference. Details regarding collection, including geographic coordinates, taxonomic identification and images can be found in the Barcode of Life data Systems (BOLD, Ratnasingham and Hebert, 2007) within two projects: WWTAL ("Barcoding Amphipoda – Talitridae") and WWGSL ("Crustaceans of the St. Lawrence Gulf") under the "Specimen Page".

Table 3.1 Species name, taxonomic authority, sample size (N) and geographic origin of taxa included in this study

Species name	N	Geographic region*
Americorchestia heardi Bousfield, 1991	5	USA (FL)
Americorchestia longicornis (Say, 1818)	15	Canada (NS, QC)
Americorchestia megalophthalma (Bate, 1862)	16	Canada (NB, NS, PEI, QC)
Chelorchestia forceps Smith & Heard, 2001	2	USA (FL)
Orchestia aestuariensis Wildish, 1987	5	UK
Orchestia cavimana Heller, 1865	5	UK
Orchestia gammarellus (Pallas, 1766)	25	UK, Canada (NL)
Orchestia grillus (Bosc, 1802)	35	USA (FL, ME, MS, SC), Canada (NB)
Orchestia mediterranea Costa, 1853	9	UK
Platorchestia platensis (Krøyer, 1845)	39	Canada (NB), USA (FL, MS)
Talitrus saltator (Montagu, 1808)	9	UK
Tethorchestia antillensis Bousfield, 1984	8	USA (FL)
Tethorchestia sp. B Bousfield, 1984	30	USA (FL), Belize, Mexico (QR)
Uhlorchestia uhleri (Shoemaker, 1930)	2	USA (MS)
Parhyale fascigera Stebbing, 1897	13	Mexico (QR), USA (FL)
Total	218	

* Abbreviations: FL – Florida, ME – Maine, MS – Mississippi, NB – New Brunswick, NS – Nova Scotia, NL – Newfoundland and Labrador, PEI – Prince Edward Island, QC – Quebec, QR – Quintana Roo, SC – South Carolina.

Table 3.2 Salinity tolerance and ecological habitat (Wildish, 1988) of talitrids included in this study

Species name	Salinity tolerance	Ecological habitat
Americorchestia heardi	Brackish	Sand-burrower
Americorchestia longicornis	Marine	Sand-burrower
Americorchestia megalophthalma	Marine	Sand-burrower
Chelorchestia forceps	Marine	Palustral
Orchestia aestuariensis	Estuarine	Wrack
Orchestia cavimana	Freshwater, brackish	Wrack
Orchestia gammarellus	Marine, brackish	Wrack
Orchestia grillus	Marine, brackish	Wrack
Orchestia mediterranea	Marine, brackish	Wrack
Platorchestia platensis	Marine	Wrack
Talitrus saltator	Marine	Sand-burrower
Tethorchestia antillensis	Marine	Wrack
Tethorchestia sp. B	Marine	Wrack
Uhlorchestia uhleri	Marine	Palustral

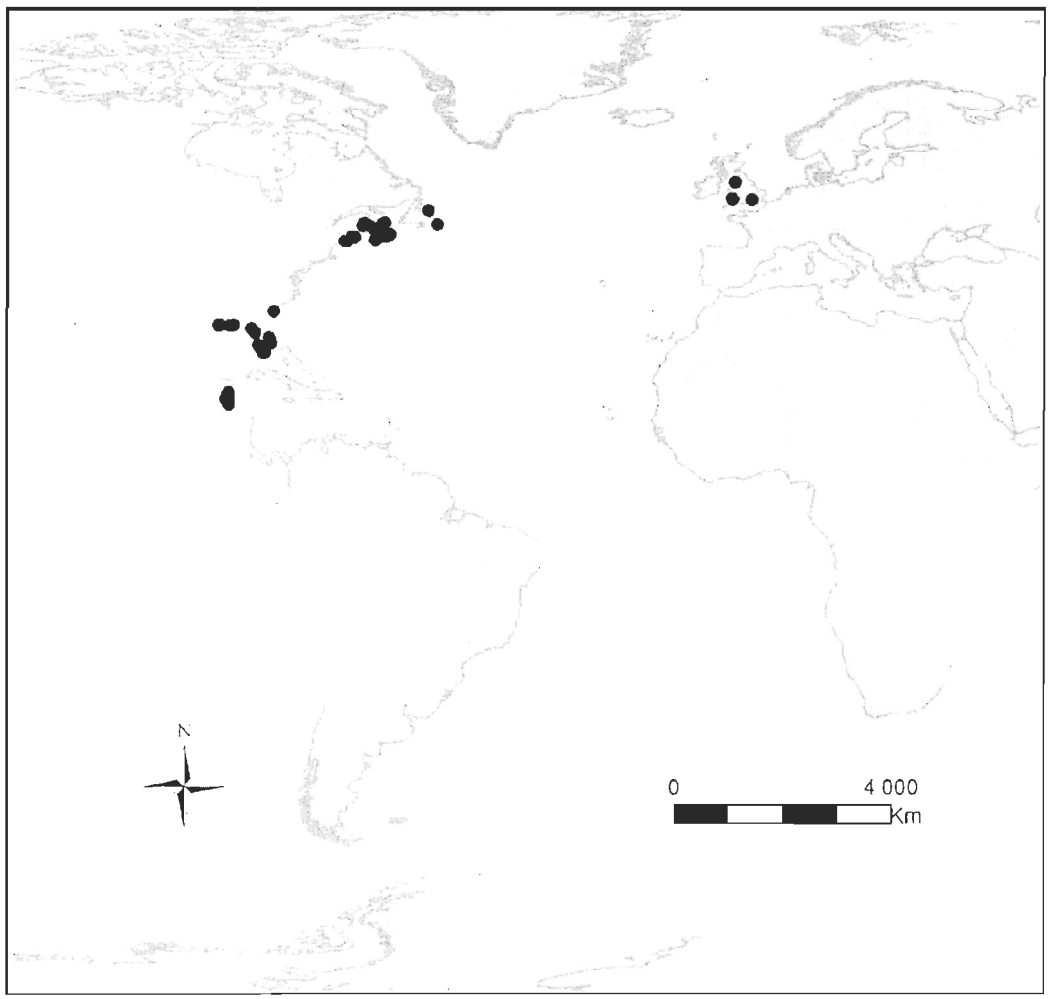


Figure 3.1 Map with sampling localities for supralittoral amphipod species barcoded in this study.

3.4.2 DNA extraction, amplification, sequencing

Genomic DNA was extracted from a small amount of muscle tissue, usually from one pereopod, preserving the rest of the organism as voucher. However, small-size specimens were used entirely for lysis and the exoskeleton recovered afterwards and stored in ethanol for future reference. A glass fibre protocol was used for extraction (Ivanova, Dewaard and Hebert, 2006) and the barcode region, a 658bp fragment at the 5'-end of the COI gene, was amplified and sequenced with standard protocols (Radulovici, Sainte-Marie and Dufresne, 2009). Two alternative sets of primers were used and their sequences are available in BOLD: LCO1490 – HCO2198 (Folmer *et al.*, 1994) with M13 tails and CrustDF1 – CrustDR1 (Steinke, unpublished).

3.4.3 Data analysis: genetic diversity and phylogenetic relationships

DNA sequences were manually edited in Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI) and aligned with the MUSCLE algorithm and default settings in MEGA 5 (Tamura *et al.*, 2011). Trace files and edited sequences are available in BOLD within WWTAL and WWGSL projects under the "Sequence Page". As a routine test for detecting pseudogenes, we checked the quality of COI sequences, their length, and the presence of STOP codons or indels in the reading frame. Sequences were also checked for contamination by using BLAST searches in GenBank at the National Center for Biotechnology Information website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Pairwise genetic distances within and among species were based on Kimura 2-parameter (K2P) model for base substitution (Kimura, 1980) and performed in MEGA 5. Graphical representation of genetic distances (maximum intraspecific – minimum interspecific) was performed in R 2.13.1 (R Development Core Team, 2010). When morphological species were split into barcode clusters diverging by more than 10× the mean value for intraspecific variation, they were considered potential cryptic species and treated as separate

Molecular Operational Taxonomic Units (MOTU; Blaxter, 2004) in further analyses. In these cases, median-joining haplotype networks were constructed in Network 4.6 (Bandelt, Forster and Rohl, 1999), to visualize relationships between haplotypes.

The final dataset used for phylogenetic analyses included all of our provisional talitrid species, an outgroup species *P. fascigera*, together with 26 COI sequences downloaded from GenBank and belonging to the following species: *Platorchestia paludosus* (HQ010305-06, HQ010311, HQ010322, HQ010325, HQ010329-30, HQ010333-36; Cheng *et al.*, 2011), *P. japonica* (EF570353; Hou, Fu and Li, 2007; HQ010337-39; Cheng *et al.*, 2011), *Orchestia cavimana* (EU276197; Browne, Haddock and Martindale, 2007) and *O. gammarellus* (EU276190-EU276199; Henzler and Ingolfsson, 2008). The haplotype dataset was used in jModelTest 1.0.1 (Posada, 2008) to find the appropriate model of sequence evolution under the Akaike Information Criterion (AIC) (Posada and Buckley, 2004). A maximum-likelihood (ML) phylogeny was built in RAxML 7.2.8 (Stamatakis, Hoover and Rougemont, 2008) as web-server application through Vital IT unit of the Swiss Institute of Bioinformatics (http://phylobench.vital-it.ch/raxml-bb). In addition, a Bayesian inference (BI) phylogeny was built in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) by using the General Time Reversible (GTR) model with gamma distribution (+G) and a proportion of invariable sites (+I). Two simultaneous analyses, each consisting of four chains, were run for 10 million generations, sampling every 1,000 generations. The initial 25% of samples were discarded as burn-in and the final consensus tree was rooted and edited in FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree).

3.5 Results

3.5.1 DNA barcodes

Almost all talitrid species were successfully amplified with the available set of primers. Exceptions were *Americorchestia salomani* Bousfield, 1991 (U.S.A.) and *Tethorchestia* sp. (Bahamas). Because these samples yielded positive results for the 18S gene (data not included here), the failure of COI amplification is probably due to mutations in one of the primer binding sites and not to DNA degradation.

A total of 218 talitrids belonging to 14 morphologically defined species from seven genera and one hyalid species (outgroup) were barcoded in this study (Table 3.1). Most COI sequences recovered the full length of the barcode region (658 bp) while a few sequences were shorter due to low-quality extremities. To have a uniform dataset, we trimmed all sequences to a length of 629 bp. A BLAST search in GenBank returned positive matches for only four species, three of which were barcoded in a previous study on the GSL crustacean fauna (Radulovici, Sainte-Marie and Dufresne, 2009). The 629 bp DNA fragment included 287 variable sites and it was translated into 209 amino acids. No contaminations or pseudogenes were detected.

3.5.2 Genetic distances

DNA barcoding of North Atlantic talitrids showed that morphological species usually correspond to clusters of highly similar sequences, reciprocally monophyletic in a phylogenetic tree (Figure 3.2).

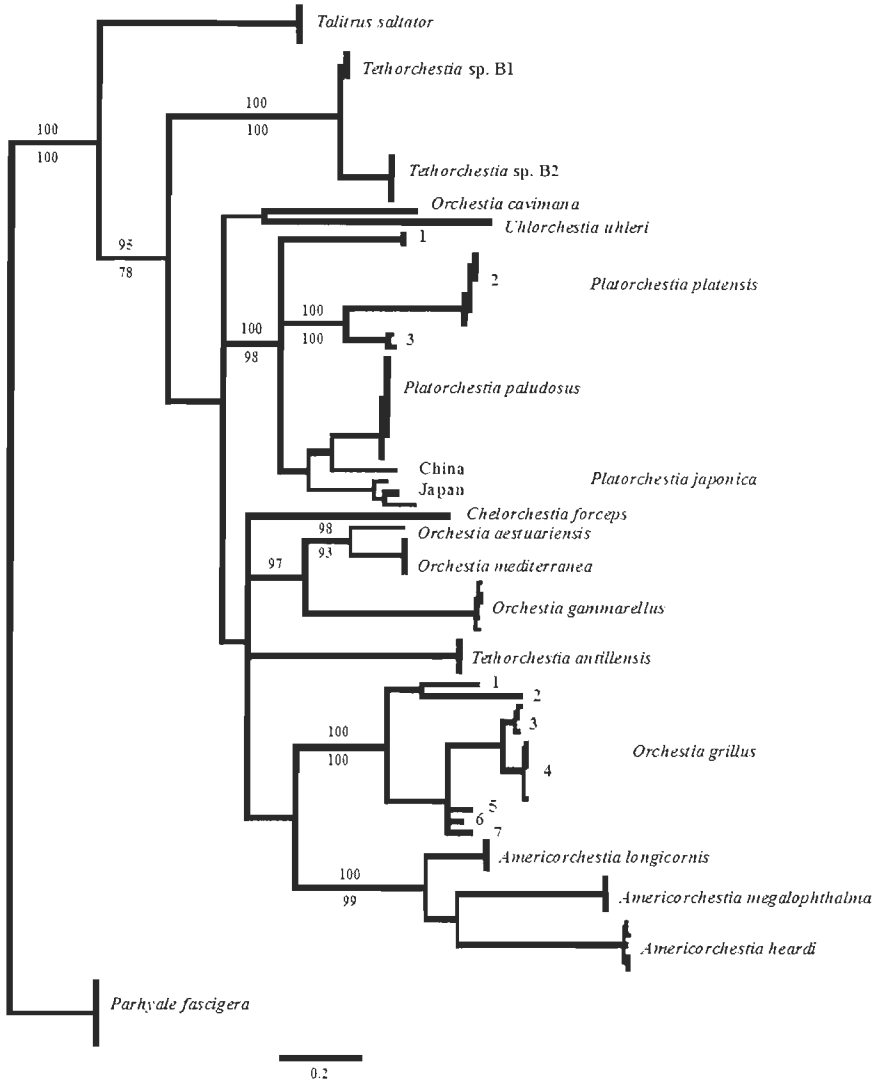


Figure 3.2 Phylogenetic tree of talitrid species based on COI sequences. Tree has been rooted with the hyalid *P. fascigera*. Sequences for *P. japonica* and *P. paludosus* are public sequences from GenBank. Numbers on branches represent posterior probability >95% for the Bayesian (BI) tree (above) and bootstrap support >70% for the maximum-likelihood (ML) tree (below).

However, exceptions were observed in three morphological species (21% of cases) resulting in multiple intraspecific barcode clusters. *Platorchestia platensis* reached a maximum of 19% pairwise genetic distance and was divided into three clusters each with sample sizes between 3-27 individuals and separated by mean distances of about 15% (Table 3.3). *Orchestia grillus* had a maximum intraspecific divergence of ~18% and was split into seven clusters separated by mean values between 4-12.9% (Table 3.3). In the latter case, barcode clusters included between one and 17 individuals, three clusters being represented by only one individual, the only singletons in our dataset. *Tethorchestia* sp. B reached a maximum of 6.5% and was split into two clusters. Following these results, the 12 barcode clusters were considered as potential cryptic species and treated as separate MOTU for further analyses. Consequently a total of 24 talitroidean MOTU's (including one hyalid species) were generated during this study. Sample size varied between 1 and 27 with an average of 9 individuals per MOTU. Mean distance was 0.3% (±0.01) within an MOTU and 13.7% between MOTU's. Maximum divergence within an MOTU reached 1.6% in *Tethorchestia* sp. B1 while the closest MOTU's were separated by 4% (*O. grillus* 3 and 4) as opposed to 9.6% in the closest morphologically defined species (Table 3.3). Regardless of the unit employed (MOTU versus species), the barcoding gap (intra- versus inter-) was clear and assigning unknown specimens to MOTU's was straightforward (Figure 3.3).

Table 3.3 Genetic diversity of talitroid species included in this study: number of haplotypes, mean K2P distances within and between species

MOTU	COI sequences	Haplotypes	Mean intrasp. (±SE)	Mean intersp. (±SE)
A. heardi	5	5	0.009 (0.002)	0.193 (0.019)
A. longicornis	15	3	0.003 (0.002)	0.193 (0.019)
A. megalophthalma	16	4	0.002 (0.001)	0.196 (0.019)
C. forceps	2	1	0	0.181 (0.019)
O. aestuariensis	5	1	0	0.096 (0.013)
O. cavimana	5	1	0	0.205 (0.020)
O. gammarellus	25	4	0.005 (0.002)	0.207 (0.021)
O. grillus 1	1	1	n/a	0.129 (0.016)
O. grillus 2	1	1	n/a	0.129 (0.016)
O. grillus 3	5	3	0.005 (0.002)	0.040 (0.007)
O. grillus 4	8	6	0.004 (0.002)	0.040 (0.007)
O. grillus 5	2	1	0	0.056 (0.009)
O. grillus 6	1	1	n/a	0.056 (0.009)
O. grillus 7	17	1	0	0.056 (0.009)
O. mediterranea	9	4	0.002 (0.001)	0.096 (0.013)
P. platensis 1	3	2	0.002 (0.002)	0.155 (0.018)
P. platensis 2	27	7	0.006 (0.002)	0.143 (0.016)
P. platensis 3	9	2	0.013 (0.005)	0.143 (0.016)
T. saltator	9	4	0.003 (0.002)	0.194 (0.020)
T. antillensis	8	4	0.002 (0.001)	0.264 (0.024)
T. sp B1	11	9	0.008 (0.002)	0.058 (0.010)
T. sp B2	19	5	0.003 (0.001)	0.058 (0.010)
U. uhleri	2	1	0	0.201 (0.020)
P. fascigera	13	7	0.003 (0.001)	0.205 (0.021)
Total	218	78		

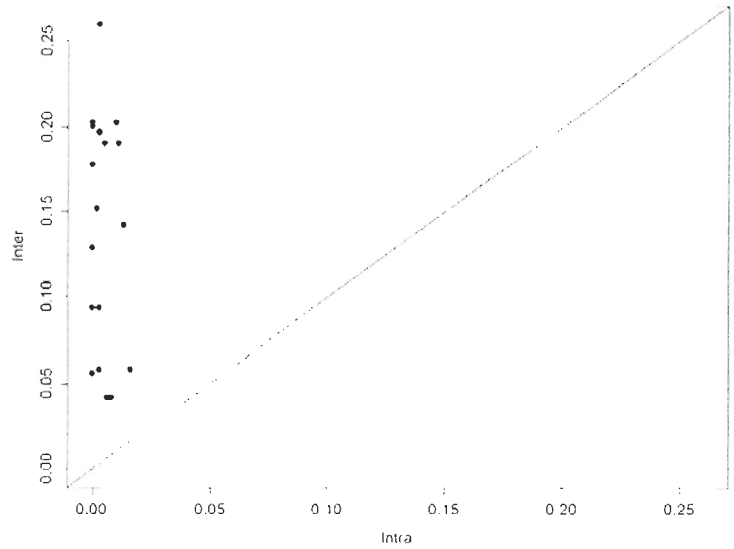


Figure 3.3 Distance graph for minimum interspecific distances related to maximum intraspecific variation (based on K2P distances). All values are above the line, hence no overlap between these two categories.

3.5.3 Phylogenetic relationships

The dataset used for phylogenetic analyses included 244 sequences (218 generated in this study and 26 sequences from GenBank) which were collapsed to 91 haplotypes (78 from our dataset and 13 from GenBank). Both trees had identical topology and showed good support for the same clusters (Figure 3.2). Regardless of cryptic species harboured by some taxa, two genera were monophyletic (*Platorchestia* and *Americorchestia)* and well-supported while two others proved to be polyphyletic (*Orchestia* and *Tethorchestia*). Within-genus relationships were less-resolved and except for sister-species such as *O. aestuariensis – O. mediterranea* (morphologically, ecologically and genetically close), no other clear inference about

the relatedness between congeneric species could be made. None of the three "systematic-ecological" groups was monophyletic.

3.6 Discussion

3.6.1 DNA barcodes for species identification

Our dataset included all the presently known and described talitrid genera from the East Coast of North America except *Talitroides* Bonnier, 1898. This genus is represented by two widespread species, *T. topitotum* (Burt, 1934) and *T. alluaudi* (Chevreux, 1896), living inland in leaf litter. Being land hoppers and exotic species, introduced to North America together with greenhouse plants (LeCroy, 2010), these species were not included in our analysis. In addition, two native, but less common species, could not be collected: *Americorchestia barbarae* Bousfield, 1991 and *Uhlorchestia spartinophila* Bousfield & Heard, 1986.

This study adds support for the use of DNA barcoding for species identification among Crustacea by showing the importance of this molecular tool for talitrid taxonomy and it adds to similar studies targeting other amphipod groups (Witt, Threloff and Hebert, 2006; Costa *et al.*, 2007; Radulovici, Sainte-Marie and Dufresne, 2009). Beginning with 15 morphologically defined species (14 talitrids and one hyalid), DNA barcoding suggested the existence of 24 genetic clusters representing putative species. This increase was due to cryptic speciation in three morphological species, *P. platensis*, *O. grillus* and *Tethorchestia* sp. B, detected only as a result of including a geographical component when sampling for DNA barcoding. Consequently, 21% of species showed cryptic diversity, compared to only 5% found in a previous study restricted to GSL crustaceans (Radulovici, Sainte-Marie and Dufresne, 2009), and this difference is probably due to the longer

geological time available in GOM (millions of years) versus the brief period (~10,000 years) of re-colonization after the last Ice Age in GSL.

3.6.2 Genetic diversity

Limited sampling does not allow for detailed analyses on genetic diversity in relation to the dispersal potential, ecological type or correlation with specific environmental factors. However, some interesting insights can be gained if compared to previous studies.

In our study European talitrid hoppers were represented by four *Orchestia* spp. (*O. cavimana*, *O. gammarellus*, *O. aestuariensis*, *O. mediterranea*), which occur in the Medway Estuary (UK) as well as many other locations on Atlantic and Mediterranean coasts of Europe. In the Medway Estuary only the two first and the two last species overlap in distribution. The potential for interspecific matings between them has been tested in the laboratory. Hybridization does not occur between *cavimana* × *gammarellus*, but the cross male *aestuariensis* × female *mediterranea* yielded hybrids, although the reciprocal cross (with male *mediterranea*) was sterile (Wildish, 1970). Naturally occurring hybrids were subsequently found in the Tamar Estuary, UK (Wildish, 1987).

O. cavimana Heller, 1865 is a freshwater talitrid found in wrack on river banks or lake shores but also extending into dilute brackish water in estuaries, and with a disjunct distribution in Northern Europe and the Mediterranean basin (Wildish, 1969). Our specimens from the Medway Estuary (UK) shared the same COI haplotype with a specimen from the inland Tegeler See (Germany; Browne, Haddock and Martindale, 2007). Apparently, the same haplotype is reaching Northern Italy (Lake Garda) and this wide distribution might be the consequence of recent expansion (Ketmaier and De Matthaeis, 2010), as this species is currently increasing its range on the Estonian coasts (Herkul, Kotta and Kotta, 2006). However, two cryptic divergent clusters (19% COI distance) have recently been found in *O. cavimana*

(Ketmaier and De Matthaeis, 2010). As one of them includes the type locality, which is a freshwater spring on Mount Olympus, Cyprus, and this haplotype appears to be restricted to the Eastern Mediterranean and Black Seas (Cyprus, Turkey), it follows that the rest of *O. cavimana* specimens (including ours) belong to a different species, still to be formally described.

O. aestuariensis (estuarine range 6-10‰) and *O. mediterranea* (marine and estuarine, range >10‰) may occur together in lowland European estuaries and are considered as sister species. While fertile hybrids occur, their sex ratio is skewed towards males, limiting the chances of genetic exchange between the two species (Wildish,1988). Very close morphologically, these two species are also the closest genetically (9.6%) in our dataset, with *O. mediterranea* showing higher diversity (four haplotypes from the same sampling locality as opposed to only one haplotype in *O. aestuariensis*).

O. gammarellus is the type species of *Orchestia* Leach, 1814, a marine talitrid with amphi-Atlantic distribution, in Northern Europe, the Mediterranean and Newfoundland (Canada) to Maine (U.S.A) (Bousfield, 1973). Our four haplotypes included three singletons (Chittick Beach and Fogo Island, Canada) and one very common haplotype distributed in the UK (Medway, Ogmore and Duddon Estuaries) and Canada (Fogo Island, Witless Bay) (data not shown). This low genetic variation and lack of genetic structure on both sides of the Atlantic is consistent with the hypothesis of Henzler and Ingolfsson (2008) that *O. gammarellus* recently colonized North America from Europe via northern islands as stepping-stones.

3.6.3 Cryptic diversity in a wrack generalist: *Platorchestia platensis* (Krøyer, 1845)

Platorchestia platensis is a very common wrack generalist with wide distribution across continents both in warm-temperate and tropical regions. The type species was described from Montevideo (Uruguay), and has since been recorded from the shores of all continents, except Antarctica. However, morphological work has highlighted various forms, some of them with full species status at present: a closely related form in mid-Atlantic islands (*P. monodi* Stock, 1996) and *P. paraplatensis* Serejo & Lowry, 2008 in Australia. To facilitate the classification, Miyamoto and Morino (2004) proposed the use of sexually dimorphic characters to divide the genus into three groups, with *P. platensis* in group 1 and *P. monodi* in group 2.

Already mentioned as a species complex (Bousfield, 1984; Bousfield and Poinar, 1995; Serejo and Lowry, 2008) due to multiple closely related variants, *P. platensis* is an interesting model to study speciation. Because our sampling was limited to only one continental coast (NWA and GOM), our genetic data shows the existence of only three divergent clusters separated by ~15% COI distance (Figure 3.4). Considering the distance between the closest pair of morphological species in our dataset (9.6% for *O. aestuariensis* – *O. mediterranea*), these three clusters should be considered as separate species and the common view of one cosmopolitan *P. platensis* should be discarded. Due to the development of molecular tools, and especially with the recent popularity of DNA barcoding, the existence of cosmopolitan distributions has been challenged and results often showed the presence of complexes of cryptic species in many widely-distributed taxa including marine invertebrates (Gómez *et al.*, 2007). In addition, the European *P. platensis*, believed to be an introduced species with an expanding range, exhibits morphological variation compared to material from the type locality (Serejo and Lowry, 2008). Molecular investigation might reveal yet another cluster of this species complex in Europe. By sequencing specimens from the type locality, some light will

be shed on the "real" *P. platensis* and all its "relatives" that should be treated as separate species.

Our three putative species are distributed as follows: one in NWA (including GSL) (subject of a separate phylogeographic study) and the other two inside GOM (one group exclusively in Florida, the other only in Mississippi). Previous genetic studies on marine species have regarded Florida as a sharp phylogeographic break between the open-Atlantic coast and the Gulf (review in Neigel, 2009), which might explain our results, cluster 1 being more distant to the other two (15.5%, Table 3.3, Figures 3.2 and 3.4). It is more difficult to interpret the existence of two genetically divergent but morphologically similar clusters, situated relatively close in a space with no obvious barrier (coastlines of Florida and Mississippi). Moreover, *P. platensis* is considered a wrack generalist, very abundant, highly tolerant to environmental variations, good competitor with other talitrids and using rafting in wrack to disperse, successfully invading Europe and spreading along its coastlines in the last 150 years (Persson, 2001 and references therein). The difficulty of morphologically discriminating these two groups based on light microscopy would explain why taxonomists have not recognized the three haplotypes (but see LeCroy, 2010). A similar case of cryptic speciation has been found in *P. japonica* from Taiwan and Japan (Cheng *et al.*, 2011) with three clusters separated by lower values (10.4-14.3%) than the *P. platensis* groups. The genetic cluster from Taiwan (Japan having the type locality) was named (*P. paludosus*) and described based on fine morphological differences (type of setae) revealed only by scanning electron microscopy (SEM). Therefore, our puzzling pattern might be solved by future SEM investigations.

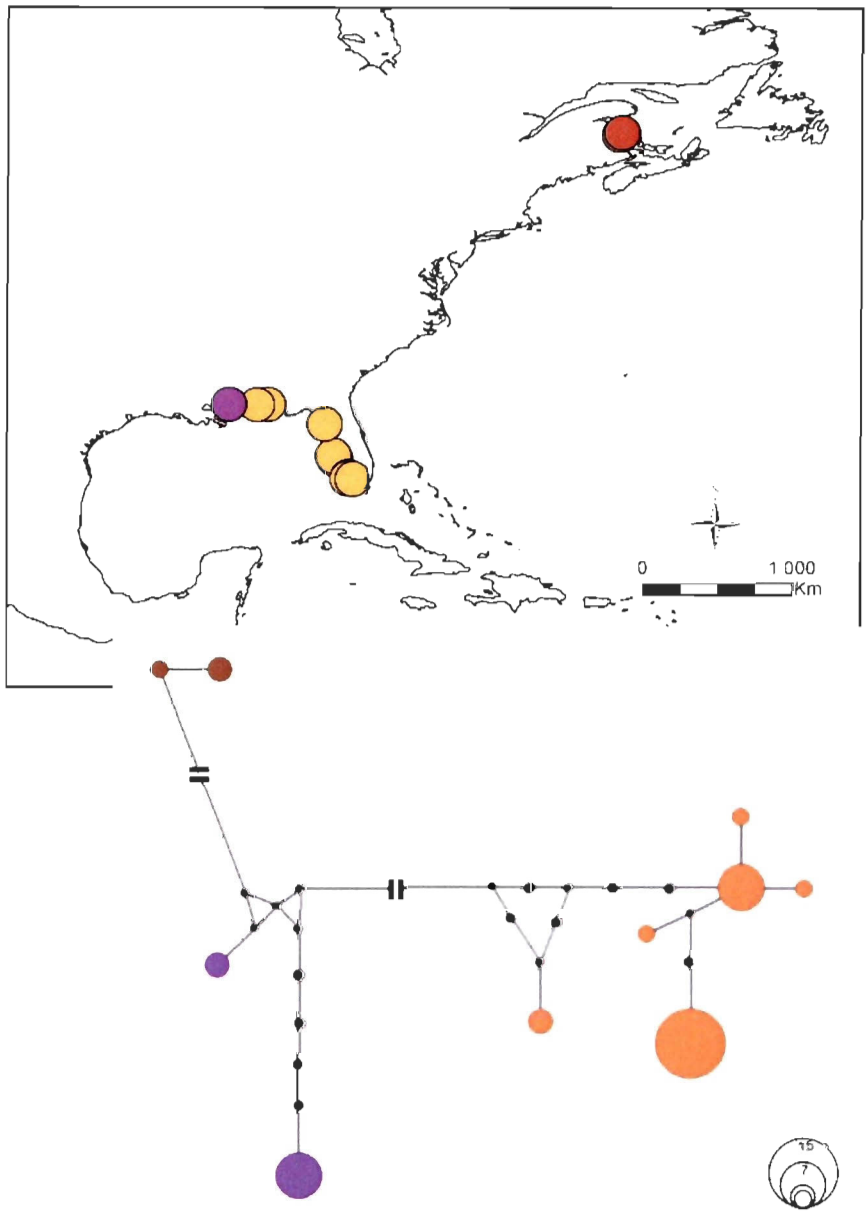


Figure 3.4 Haplotype network and geographical distribution of haplotypes for *Platorchestia platensis* species complex. Each putative species (clusters 1 to 3) has a different colour: 1 – brown, 2 – orange, 3 – violet). Interrupted lines represent deep divergences separating MOTU.

3.6.4 Cryptic diversity in a palustral specialist: *Orchestia grillus* (Bosc, 1802)

Orchestia grillus is a salt marsh specialist, nestling among roots of *Spartina* and other marsh grasses where it feeds on wrack and marsh debris from GSL to GOM (Bousfield, 1973). Specific habitat requirements (salt marshes which are typically separated by variable geographic distances) should result in limited dispersal between populations (depending on dispersal distance), therefore reduced gene flow among populations translating into strong genetic structure culminating with speciation (Bohonak, 1999). This species complex consisted of seven clusters with COI divergences ranging from 4% to ~13% (Table 3.3.). Geographically, clusters 3 and 4 were situated in South Carolina and Maine-New Brunswick, respectively, while the other five groups were all distributed in GOM, one in Mississippi and four in Florida (Figure 3.5). The existence of three singletons might indicate the amplification of pseudogenes, especially in clusters 1 and 2, the most divergent from the rest (12.9% distance; Figure 3.2). In our analysis, we could not detect any obvious sign of pseudogenes (e.g., STOP codons, indels, double-peaks) but this does not discard the possibility of having them in this dataset (Buhay, 2009). Regardless of this possibility, believed to overestimate species richness if undetected (Song *et al.*, 2008), and considering the most conservative measure for our dataset (three possible "untrue" clusters), there is enough proof for cryptic speciation in *O. grillus*. While additional morphological (SEM), ecological and genetic work is required in order to clarify the extent of this species complex, future biodiversity assessments of NWA and GOM should be aware of hidden diversity in this and other talitrid species.

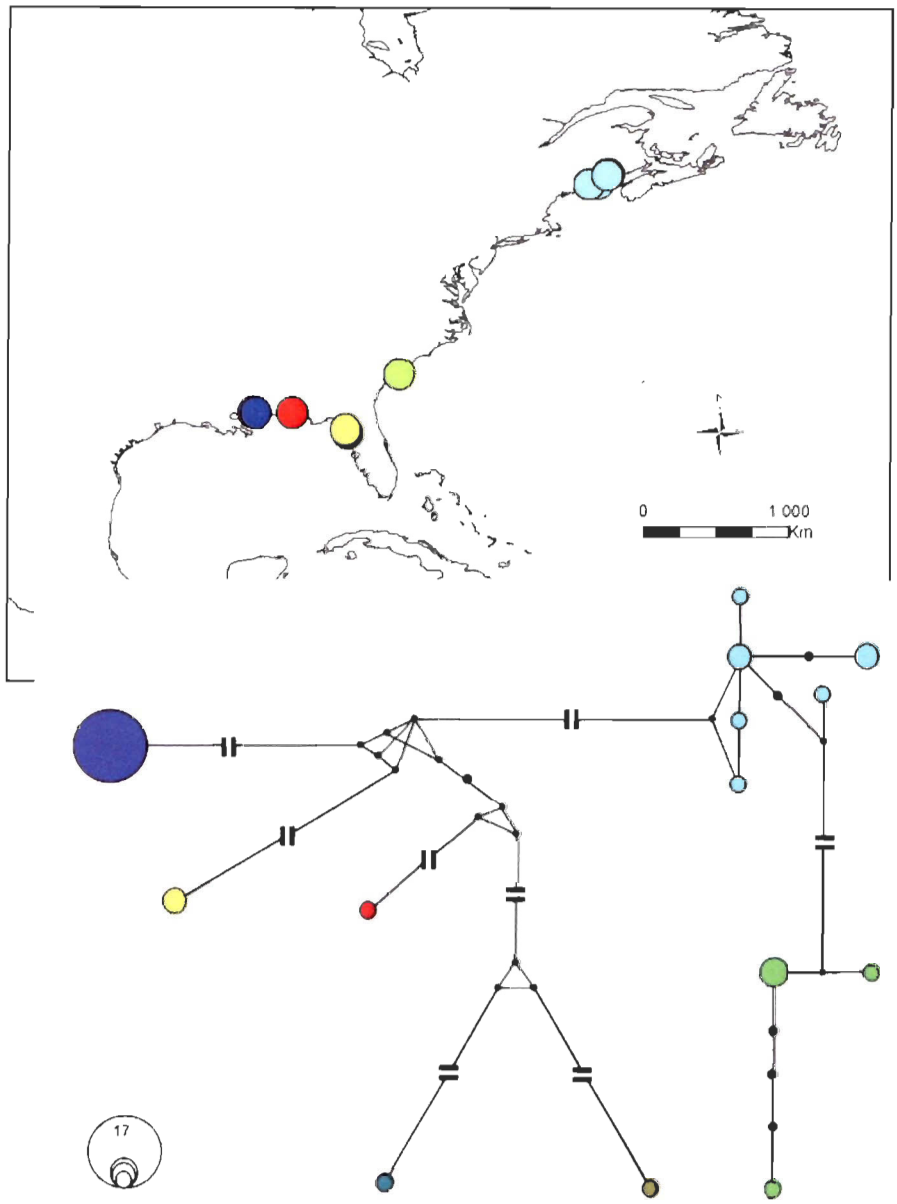


Figure 3.5 Haplotype network and geographical distribution of haplotypes for *Orchestia grillus* species complex. Each putative species (clusters 1 to 7) has a different colour: 1 – brown, 2 – dark green, 3 – light green, 4 – light blue, 5 – yellow, 6 – red, 7 – dark blue. Interrupted lines represent deep divergences separating MOTU.

3.6.5 Cryptic diversity in *Tethorchestia* sp. B Bousfield, 1984

When Bousfield (1984) erected a new genus of beach fleas from the Caribbean with the type species *Tethorchestia antillensis* Bousfield, 1984, he gave a brief description of the new genus and the new species with no drawings and mentioned the existence of six additional species to be described later (spp. B through G). Subsequently he provided illustrations for the type species (Bousfield and Poinar, 1995) but never for the other *Tethorchestia* spp. The only undescribed species from Florida was sp. B, hence our use of the name for those Floridian specimens that did not belong to the type species. This study showed that "*Tethorchestia*" sp. B should be separated in a different genus (see Section 3.4.6 and Figure 3.2), and the genetic finding is reinforced by morphological differences between the type species and our specimens (LeCroy, 2010). By extending our sampling to the Caribbean (Mexico, Belize), we discovered two genetic clusters distanced by 5.8% (Table 3.3), reciprocally monophyletic and geographically separated, in Florida and the Caribbean, respectively (Figure 3.6). Although 5.8% is lower than the minimal distance (9.6%) between morphological sister-species studied here, it is still above thresholds used for delimiting putative amphipod species (3.75% in *Hyalella*; Witt, Threloff and Hebert, 2006) and well above the minimal interspecific value for other crustaceans such as decapods (2.8% in *Hyas*; Radulovici, Sainte-Marie and Dufresne, 2009). More importantly, obtaining an interspecific value for two known sister-species when the overall talitrid diversity is largely still unknown, does not support its use as a universal threshold for the entire family or for amphipods in general. In addition to genetic data, there are morphological differences that support the split of sp. B into two species (Wildish and LeCroy, in preparation).

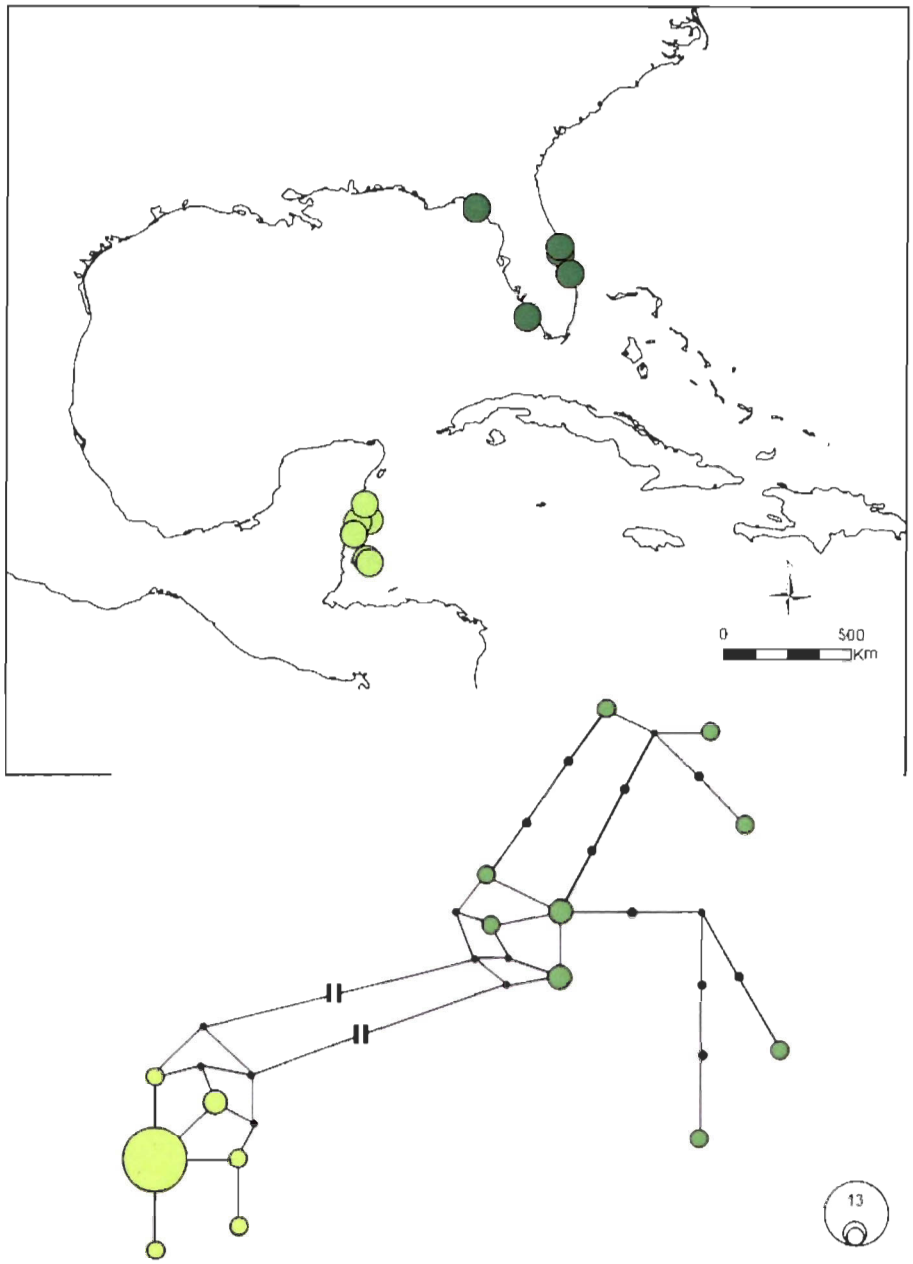


Figure 3.6 Haplotype network and geographical distribution of haplotypes for *Tethorchestia* sp. B complex. Each putative species (clusters 1 and 2) has a different color: 1 – dark green, 2 – light green. Interrupted lines represent deep divergences separating MOTU's.

3.6.6 Phylogeny

Our inference of phylogenetic relationships between talitrid taxa is limited to one gene, COI, known to be a good marker at the species level, hence its use in DNA barcoding. Although higher taxonomic levels are usually clarified by using slower evolving and multiple genes, our study is still the largest phylogenetic endeavour to date and gives an interesting glimpse at evolutionary relationships among talitrids.

Few genera formed well-supported monophyletic clusters. *Platorchestia* spp. formed a monophyletic cluster containing all six putative species from the two species complexes, *P. platensis* and *P. japonica* (Figure 3.2). While the two *P. platensis* clusters from GOM (2 and 3) are sister species, it is unresolved if cluster 1 from NWA is more closely related to them or to the other species complex, *P. japonica* from Asia. In the latter complex, *P. paludosus* seemed to be sister species with Chinese collections assigned to *P. japonica* and less related to collections from Japan, but none of these branches were supported. The obvious conclusion is that the two cryptic *P. japonica* clusters should be considered distinguished species, and named and described, as well as *P. platensis* 1, 2, 3. Moreover, by including additional genes and more *Platorchestia* spp., it will become evident if the various species complexes are at least monophyletic with regional cryptic species or if the entire classification within this genus has to be revised and the morphological groups proposed by Miyamoto and Morino (2004) should be discarded. *Platorchestia* Bousfield, 1982 has been traditionally considered within *Orchestia* Leach, 1814 (type genus for Talitridae) and our data support this split and shows *Platorchestia* to be sister group to various *Orchestia* spp., as well as other genera, although with low support in both analyses (Figure 3.2).

On the other hand, *Orchestia* is definitely non-monophyletic. A well-supported cluster included sister species *O. aestuariensis* and *O. mediterranea* together with *O. gammarellus,* all of which are morphologically and ecologically similar. Another well-

supported cluster consisted of the *O. grillus* species complex (seven clusters), sister group to *Americorchestia* spp., while O. *cavimana* was separated together with *U. uhleri* in the same weakly supported group (Figure 3.2). More detailed investigations into the morphology of *Orchestia* might bring support for a division into multiple genera. If a revision proves to be necessary, as seems to be the case here, *O. grillus* would be placed in a new genus, as *O. gammarellus* is the type species for *Orchestia*. *O. grillus* is a marsh specialist with habitat requirements and life history traits quite different from those of most *Orchestia* spp. (see Section 3.6.4).

The North American genus *Americorchestia* formed a monophyletic cluster, well-supported in both BI and ML analyses (Figure 3.2). This genus consists of five species divided into two morphological/ecological groups, named after the species inhabiting sandy beaches on the open-Atlantic coast: *megalophthalma* (including also GOM species, *A. salomani* and *A. barbarae*) and *longicornis* (with the GOM counterpart, *A. heardi*). We included only three species, which showed clear separation from one another (~20%) but no support for a geographical (NWA versus GOM) or "systematic" (towards genus splitting) differentiation. Therefore, within-genus relationships will be clarified once the two missing species (*A. barbarae* and *A. salomani*) are collected and sequenced. The same observation is valid for genus level patterns. In the present phylogeny, the sand-burrower *Americorchestia* was close to the wrack generalist *Orchestia*. However, the former was erected as a separate genus (Bousfield, 1991) from a large group of sand-burrowing talitrids, *Talorchestia sensu lato*, which is frequently being split into additional genera (see WoRMS for updated taxonomy). It was also mentioned to be more similar to taxa from the NW Pacific (*Talorchestia sensu lato*) than to other sand-burrowers (*Talorchestia sensu stricto*, *Megalorchestia*, *Talitrus*) and our phylogeny confirms its distinctness from *Talitrus saltator* (Montagu, 1808). A similar result has been found for the European sand-burrowers (*Talorchestia*, *Talitrus*, *Orchestia*) (Conceição, Bishop and Thorpe, 1998) (but see Davolos and Maclean, 2005 for an opposite result). Only by increasing sampling to sand-burrowing taxa morphologically and ecologically closer or more distant to *Americorchestia*, will phylogenetic relationships

among sand-burrower taxa and their closest relatives from the "systematic-ecological" groups be resolved.

Some *Tethorchestia* spp. were mentioned as having overlapping morphological characters with *Orchestia* spp., hence a hypothetical evolutionary connection between genera throughout North Atlantic and the marsh specialist *O. grillus* (Bousfield, 1984). In our phylogeny, only *T. antillensis* is close to *Orchestia* spp., being nested in the *Orchestia – Americorchestia* cluster, but with low support. By contrast, *Tethorchestia* sp. B1 and B2 were very distant and sister group to all other talitrid species (except for *T. saltator*), with good support in both analyses. This strongly indicates that B1 and B2 do not belong to *Tethorchestia* and morphological characters further support the COI phylogeny and the need for a taxonomic revision to erect a new genus for these taxa (see Section 3.6.5). As no formal description has been provided for the extra six *Tethorchestia* spp. collected in the Caribbean (Bousfield, 1984), this genus presently consists of only two extant, *T. antillensis* and *T. karukarae*, and one extinct species, *T. palaeorchestes*. As a result the phylogeny within this genus is more difficult to investigate.

Other palustral species (besides *O. grillus*) are less widely distributed in our sampling area and we were able to include only two species from GOM. *Uhlorchestia uhleri* (Shoemaker, 1930), is an American species distributed from Maine (U.S.A.) to Southern GOM (Bousfield, 1973; Velasco, Sanchez and Florido, 2005). Together with its sister species, *U. spartinophila* (not collected for this study), they are the only representatives of this genus. *Chelorchestia forceps* Smith & Heard 2001 is known only from GOM. Both palustral species grouped with wrack generalists in weakly supported clusters, showing polyphyly for this group, the same pattern as for the other two groups (wrack generalists, sand-burrowers). Our phylogeny cannot be easily compared with Bousfield's well-known morphological phylogeny, conducted at the genus level and restricted mostly to the North Pacific (Bousfield, 1982), because of differences in the taxa sampled. However, both phylogenies found that "systematic-ecological" groups are polyphyletic and they

should therefore be used with caution. In the absence of a better system, this classification can be useful to some extent although it is hampered by morphological convergence. The barcoding initiative is constantly growing and many talitrid taxa will be collected in the global attempt to catalogue the world's biodiversity (International Barcode of Life, http://ibol.org/). Although complete phylogenies cannot be based solely on COI, the amount of information generated through iBOL will shed some light on talitrid systematics and will stimulate subsequent genetic studies including multiple genes.

3.6.7 Biogeography

By sampling a large coastline at the continental scale, it was inevitable to include multiple marine biogeographic provinces (Arctic, Cold-temperate NWA, Warm-temperate NWA, Tropical NWA and Northern European Seas) with various marine ecoregions: Southern Labrador, GSL – Eastern Scotian Shelf, Scotian Shelf, Gulf of Maine – Bay of Fundy, Carolinian, Floridian, Northern GOM, Western Caribbean, North Sea and the Celtic Sea (Spalding *et al.*, 2007). Talitrid distribution is fairly well known on European shores and along the American open-Atlantic coast but is still incomplete in the GOM, especially in the southern region. With present knowledge, our dataset consisted of taxa ranging from "cosmopolitan" (see Section 3.6.3) to amphi-Atlantic and endemic on either of the two coasts. There were four endemic genera (*Americorchestia*, *Tethorchestia*, *Chelorchestia*, *and Uhlorchestia*) to the American Atlantic and Caribbean regions (Bousfield, 1984). While the knowledge of GOM talitrids is limited (but still believed to reach 30% endemic species), talitrids are mainly known from the northern GOM and currently include eight species, four being endemic (LeCroy *et al.*, 2009). The strongest barrier to talitrid distribution is believed to be the Mississippi Delta, which separates sand-burrowing hoppers into an eastern (*A. salomani* and *A. heardi*) and a western group (*A. barbarae* and unknown counterpart species for *A. heardi*) (Bousfield, 1991).

The primary goal of DNA barcoding studies is species identification, with limited sample size per species. Limited sampling does limit our ability to apply our data to phylogeographic studies. However, some patterns such as deep intraspecific divergences at large spatial scales (i.e., potential cryptic species) can be easily detected and follow-up investigations should be conducted in order to explain the observed patterns. In this study, we identified that a new morphologically defined species ("*Tethorchestia* sp. B") actually belonged to a new genus, presented new distribution records (to be included in future distribution maps) and discovered multiple divergent clusters inclusive of putative new species among taxa with supposedly continuous distribution (*P. platensis* and *O. grillus*). At a large spatial scale (NWA), the major break was between the open Atlantic coast and GOM, in agreement with previous studies (reviewed in Neigel, 2009). At smaller scales (northern GOM), the Mississippi Delta may act as a dispersal barrier for talitrid species distribution (see above) with consequences for genetic structure (Neigel, 2009). This boundary does not explain our findings for *P. platensis* and *O. grillus*. Considering the historical biogeography of the entire North Atlantic, with tectonic movements (e.g., Central American Isthmus) and glacial cycles, the large scale patterns observed here may reflect the impact of glacial cycles at the genetic level (Hewitt, 2000). However, their present-day maintenance might involve some physical (oceanographic), ecological (microhabitat preference) or biological (behavioral) barriers and should be the focus of more detailed investigations. By extending sampling to the western GOM (west of the Mississippi boundary) and the tropical southern GOM, new intraspecific clusters might be revealed. Various glacial cycles had different impacts on GSL (completely covered by ice sheets) compared to GOM (sea-level drop), and this might explain the higher diversity (as number of haplotypes) seen in *A. heardi* (GOM) as opposed to *A. longicornis* and *A. megalophthalma* (GSL and open-Atlantic Canadian coast). Moreover, southern populations have multiple generations per year and are active all-year round while northern populations have only one generation per year (Wildish *et al.*, 2011) and are inactive in the sediment during long winters (Wildish, 1988). These life-history

traits might affect the mutation rate resulting in higher diversity in the south over the evolutionary time scale.

3.7 Conclusions

This study has shown the importance of molecular tools for taxonomic studies and their potential evolutionary implications. Talitrids are already a species-rich group and many more species (possibly hundreds to thousands) await discovery, mainly in Indo-Pacific tropical areas. In addition, many undescribed taxa have been collected over the previous decades and remain in museum collections awaiting formal description. In this context, DNA barcoding comes as a tremendously useful tool to identify, classify and discover new taxa, a "professional organizer" for the plethora of synonyms, similar forms, unknown distributions (native or introduced) and mysterious taxa known as sp. A, B, C, D as in Bousfield (1984).

Starting with 15 morphologically defined species we increased the diversity list to 24 putative species mainly by discovering three species complexes (*P. platensis*, *O. grillus*, *Tethorchestia* sp. B). However, the clarification of the "real" (i.e., type) species requires molecular work on material from the type locality, to be collected and analyzed in the future. Cryptic species showed an allopatric distribution (except for clusters 1 and 2 in *O. grillus*) and further SEM investigation might highlight fine morphological differences, while ecological studies might reveal microhabitat variation. As many cryptic species are detected worldwide in all taxa, there is a need to formally recognize new species and genera in order to ease the backlog of unnamed functional units of biodiversity (i.e., putative species revealed through DNA barcoding).

We conducted the largest phylogenetic study (based on DNA sequences) on talitrids to date. Although the tips of the tree (i.e., putative species), as well as some clusters (e.g., *Platorchestia*, *Americorchestia*) were well-supported, phylogenetic

relationships within and between genera were less resolved. With the selected taxa and only one gene, there is evidence for polyphyly in some genera (e.g., *Tethorchestia*, *Orchestia*) and in all ecological groups included (sand-burrowers, wrack generalist and palustral hoppers). There is an obvious need to increase sampling of various taxa worldwide and to include additional mitochondrial and nuclear genes, together with morphological characters, in order to have a better picture of talitrid evolutionary history and the link between "systematic-ecological" and true phylogenetic groups. We believe that the phylogenetic tree provided in this study (Figure 3.2), when updated with more of the world's talitrid fauna and additional genes, might provide a satisfactory higher level classification of this group.

Acknowledgements

We are grateful to Rachael King and Joseph Hunt for collecting some of the talitrids and to Traian Brad, Fred French and Cortney Wyatt for valuable assistance during fieldwork. Sampling in the Magdalen Islands (Québec) was funded by CERMIM while talitrids from the Biosphere Reserve 'Banco Chinchorro' (Mexico) were collected with funding from MarBOL and local support from ECOSUR (Manuel Elias Gutierrez, Martha Valdez-Moreno). We thank Jim Lowry for sharing ideas on talitrid diversity with us. Servet Çizmeli helped with analyses in R and Arnaud Tarroux helped with map creation in ArcGIS. This work is a contribution to the Canadian Barcode of Life Network, as well as to the Canadian Healthy Oceans Network, both funded through the Natural Sciences and Engineering Research Council.

CHAPTER IV

DNA BARCODING BEYOND SPECIES IDENTIFICATION: PHYLOGEOGRAPHIC PATTERNS IN *GAMMARUS OCEANICUS* SEGERSTRÅLE, 1947

Adriana E. RADULOVICI[1], Bernard SAINTE-MARIE[2], Dirk STEINKE[3], Guglielmo TITA[4], Filipe O. COSTA[5], France DUFRESNE[1]

[1]*Département de Biologie, Université du Québec à Rimouski, Rimouski G5L 3A1 (QC), Canada*
[2]*Maurice-Lamontagne Institute, Fisheries and Oceans, Mont-Joli (QC), Canada*
[3]*Biodiversity Institute of Ontario, University of Guelph, Guelph, (ON), Canada*
[4]*Centre de recherche sur les milieux insulaires et maritimes, Université du Québec à Rimouski, Havre-aux-Maison, (QC), Canada*
[5]*Centro de Biologia Molecular e Ambiental, Universidade do Minho, Braga, Portugal*

(Article in preparation)

Author contribution

Concept: AER, BSM
Sampling: AER, DS, GT, FOC
Laboratory operations: AER, DS, FOC
Data analysis: AER, DS
Manuscript writing: AER
Comments: BSM
Funding: FD, BSM, DS, GT, FOC

4.1 Résumé

Les communautés en eau peu profonde actuellement distribuées dans l'Atlantique Nord ont été façonnées par des événements historiques tels que l'ouverture du détroit de Béring, un échange transarctique, et plus récemment par les glaciations du Pléistocène. Au cours du dernier maximum glaciaire, des calottes glaciaires massives ont recouvert les rives rocheuses américaines, menant vraisemblablement à l'extinction des espèces intertidales et à une recolonisation à partir de l'Europe après le retrait des glaciers. Dans ce chapitre, nous étudions un amphipode intertidal, *Gammarus oceanicus*, habitant les rivages rocheux des deux côtés de l'Atlantique. *Les* séquences ADN du cytochrome *c* oxydase 1 *(COI)* ont été utilisées pour étudier les modèles phylogéographiques de cette espèce amphi-atlantique. Un total de 273 séquences provenant de 87 sites d'échantillonnage a montré l'existence de deux groupes séparés par une distance génétique conséquente moyenne (2,4% de divergence) et par des étendues géographiques importantes (milliers de km). Aucun haplotype n'est partagé entre les groupes. Un groupe est distribué en Europe et dans l'Arctique canadien (Baie d'Hudson) tandis que l'autre est limité au Canada atlantique (Golfe du Saint-Laurent et l'ouverture de la côte atlantique). Par ailleurs, une analyse AMOVA a montré un certain niveau de structuration génétique dans ce dernier groupe. Nos résultats indiquent la présence des deux côtés de l'Atlantique de refuges glaciaires tels que les Grands Bancs, le banc Georges (Amérique du Nord), la Manche et la mer d'Irlande (Europe). Ce modèle est cohérent avec les résultats précédents observés chez les algues marines (ex. *Ascophyllum nodosum*) utilisées comme source de nourriture et d'abris par *G. oceanicus.* La dispersion à plus petite échelle (comme au Canada atlantique) semble être entravée par des caractéristiques océanographiques. Bien que la distance génétique entre les groupes ne soit pas très élevée, il pourrait cependant indiquer un phénomène de spéciation en cours. L'intérêt croissant pour le code-barres moléculaire permettra l'utilisation des mêmes séquences d'ADN pour réaliser des études supplémentaires. Par exemple, des analyses en phylogéographie comparative de taxons co-distribués permettront de comprendre l'impact de différentes influences sur l'actuelle structure génétique des organismes du littoral.

Mots-clés: espèces cryptiques; code-barres d'ADN; *Gammarus oceanicus*; glaciations; Atlantique du Nord; phylogéographie

4.2 Abstract

Shallow-water communities currently distributed in North Atlantic have been shaped by historical events such as the opening of the Bering Strait, followed by a trans-Arctic interchange, and more recently by the Pleistocene glaciations. During the last glacial maximum, massive ice sheets covered the American rocky shores presumably leading to extirpation of intertidal species and re-colonization from Europe after the retreat of the glaciers. Here we investigated an intertidal amphipod, *Gammarus oceanicus*, inhabiting rocky shores on both Atlantic coasts. DNA sequences belonging to the barcode region, cytochrome *c* oxidase 1 (COI), have been used to investigate phylogeographic patterns in this amphi-Atlantic species. A total of 273 sequences from 87 sampling sites showed the existence of two clusters separated by medium genetic (2.4% divergence) and large geographic (thousands of kilometres) distances. No haplotype was shared between clusters. One group was distributed in Europe and Arctic Canada (Hudson Bay) while the other was restricted in Atlantic Canada (Gulf of St. Lawrence and the open Atlantic coast). Moreover, the AMOVA analysis showed a certain level of genetic structuring in the latter group. Our results indicate persistence on both Atlantic coasts, in glacial refugia such as Grand Banks, Georges Bank (North America), the English Channel and the Irish Sea (Europe). This pattern is concordant with previous findings in marine seaweeds (e.g., *Ascophyllum nodosum*) which are used as food source and shelter by *G. oceanicus*. Dispersal at smaller scales (e.g., Atlantic Canada) seems to be hampered by oceanographic characteristics. Although the gap between clusters is not very high, it might be indicative of ongoing speciation. With the growing efforts for DNA barcoding of various groups, there will soon be an extraordinary opportunity to use the same DNA sequences for additional studies such as comparative phylogeography of co-distributed taxa in order to unravel the impact of various forces on the present-day genetic structure of coastal organisms.

Keywords: cryptic species; DNA barcoding; *Gammarus oceanicus*; glaciations; North Atlantic; phylogeography

4.3 Introduction

North Atlantic communities have been shaped during various steps of the geologic and climatic history of the Northern Hemisphere. Originating from the initial break-up of Pangaea during the Jurassic, the North Atlantic Ocean was largely influenced by climatic oscillations. Rapid cooling in late Eocene (from subtropical to temperate and cold) resulted in the emergence of new biotopes to and through which marine life adapted and diversified, therefore Atlantic species adapted to primitive climate (i.e., subtropical and warm-temperate) might be considered phylogenetically older than species distributed in new biotopes (i.e., cold-temperate and arctic) (Golikov and Tzvetkova, 1972). Besides the "local" North Atlantic radiation, a large input of species resulted from the opening of the Bering Strait in the early Pliocene, around 3.5 million years ago (MYA), followed by trans-Arctic interchange and an invasion of the North Atlantic by North Pacific species (Vermeij, 1991). The latest historical events with great impact on marine, freshwater and terrestrial communities were the Pleistocene glaciations leading to contractions and expansions of species ranges during glacial and interglacial phases, respectively. The peak of the last glacial cycle, known as the Last Glacial Maximum (LGM), occurred ~24KYA (thousands of years ago, calibrated years), when massive ice sheets covered large areas of North America and Europe, including coastal habitats, and low-stand values for sea level reached -130 m (Mix, Bard and Schneider, 2001), exposing the continental shelves. In this context, it was considered that cold-temperate and arctic species survived in large refugia south of the unfavourable habitat (ice and permafrost) and re-colonized the northern habitats once the glaciers began to retreat. Biogeographic data (e.g., endemic species, disjunct distributions) raised the issue of periglacial refugia (i.e., small ice-free patches in the north) where small populations could have survived during LGM but proof for such refugia is usually scarce (review in Brochmann *et al.*, 2003) and never supported by multiple types of data (geomorphology, radio-carbon dating, palynological data, fossils, climate reconstruction). For coastal species, reliable information on their persistence is

difficult to gather as their hypothetical refugia are presently below sea level. Therefore indirect evidence for the impact of glacial history is harnessed from molecular data (Hewitt, 2000, 2004) and various genetic patterns can be considered a signature of LGM (Figure 3 in Maggs *et al.*, 2008).

In the Northwest Atlantic (NWA) the Laurentide Ice Sheet extended south to Long Island Sound completely covering rocky shores, a habitat type lacking south of this boundary (Ingolfsson, 1992 and references therein). Therefore it has been considered that intertidal and subtidal communities associated with rocky shores went extinct during LGM, such that the present-day structure is the result of post-glacial colonization from Europe (Ingolfsson, 1992). Some studies found genetic evidence in support of this hypothesis among a few intertidal invertebrates (Wares and Cunningham, 2001; Breton *et al.*, 2003; Ilves *et al.*, 2010), including amphipods (Henzler and Ingolfsson, 2008). However, other coastal species were found to have a long history on both sides of the Atlantic (Wares and Cunningham, 2001; Ilves *et al.*, 2010; Olsen *et al.*, 2010; Panova *et al.*, 2011). The controversial NWA periglacial refugia have been recently supported by reconstructions of the ice advance and retreat (Shaw, 2006; Charbit *et al.*, 2007), although ice-free areas do not necessarily imply favourable environmental conditions to support viable populations (Brochmann *et al.*, 2003).

Species with a disjunct distribution are good models for studying the role of vicariance and dispersal on the present-day genetic architecture (Avise, 2000) and a large amount of genetic data generated within phylogeographic studies of single species is rapidly accumulating (review in Maggs *et al.*, 2008). Data usually consist of DNA sequences belonging to mitochondrial and chloroplast genes, less often to nuclear genes. Moreover, the recent development of DNA barcoding greatly expands the DNA database and its potential use. DNA barcoding is a molecular tool for species identification, which uses DNA sequences to assign unidentified specimens to known species (Hebert *et al.*, 2003). In animals, both phylogeographic and DNA

barcoding studies use the same gene fragment, the 5'end of cytochrome *c* oxidase 1 (COI). Although these two types of studies tackle different levels of biodiversity (genetic variation in phylogeography and species richness in DNA barcoding), recent large-scale initiatives involving DNA barcoding (International Barcode of Life Project, iBOL, www.ibol.org) will generate extensive datasets (objective of 5 million barcodes by 2015) that could benefit phylogeographic studies focused on single species, co-distributed species or entire communities.

Here we investigate the phylogeographic structure of a North Atlantic intertidal invertebrate using COI sequences generated during DNA barcoding. We chose the amphipod *Gammarus oceanicus* Segerstråle, 1947, to investigate the impact of glacial history on the present-day genetic structure (i.e., survival on one coast with subsequent colonization of the other coast versus long-term persistence on both coasts) because of its amphi-Atlantic distribution and preference for rocky shores. This is one of the most common and abundant coastal invertebrate species, living under stones or among algae in the intertidal and subtidal zones (0-25 m, Segerstråle, 1947), on sheltered bays and rocky shores. Its disjunct distribution includes the NWA coast from Foxe Basin and Baffin Island (north) to Long Island Sound (south), and the European coast from Franz Joseph Land (north) to northern France (south), but also the mid-Atlantic Greenland, Iceland and Faroe Islands (Steele and Steele, 1972; Bousfield, 1973). *Gammarus oceanicus* is euryhaline and omnivorous, grazing on seaweed but feeding on other invertebrates too, including crustaceans (e.g., mysids, A. E. Radulovici, pers. obs.). Although adults are food-flexible, juveniles need seaweeds for food and shelter, hence an intrinsic relation between *G. oceanicus* and various intertidal seaweeds (*Ascophyllum nodosum*, *Fucus* spp.). This species plays an important role in intertidal food webs, as prey for fish, birds and marine mammals.

In North America, the entire present-day distribution range of *G. oceanicus* was covered by the Laurentide Ice Sheet, therefore European populations were

considered to be the source of existing NWA populations (Ingolfsson, 1992 but see genetic patterns in Henzler, 2006; Costa *et al.*, 2009). The objectives of this study are two-fold: i) reveal phylogeographic patterns at the amphi-Atlantic scale by including data from a large geographic area covering most of the species range); ii) reveal genetic structure in Atlantic Canada in order to assess potential survival during LGM and post-glacial colonization routes in NWA.

4.4 Material and methods

4.4.1 Sample collection

Sampling was conducted between 2006-2010 along the shores of Eastern Canada (Gulf of St. Lawrence – GSL, open Atlantic coast) and in a few sites in northern Canada and Norway (Figure 4.1). Amphipods were collected at low tide with dip nets and immediately stored in 95% ethanol. Morphological identifications followed available keys for NWA (Bousfield, 1973). Specimens were stored as vouchers for future reference. Details regarding collection, geographic coordinates, taxonomy, vouchers and images can be found in Barcode of Life Data System (BOLD, Ratnasingham and Hebert, 2007), within the project GAMOC ("Phylogeography of *Gammarus oceanicus*") under the "Specimen Page". In order to increase our geographic coverage for this taxon, we included published data for Iceland, Poland and Canada (Costa *et al.*, 2009, project code: FCGA; Radulovici, Sainte-Marie and Dufresne, 2009, project code: WWGSL), as well as sequences of additional Canadian specimens provided by Dr. Paul Hebert (University of Guelph).

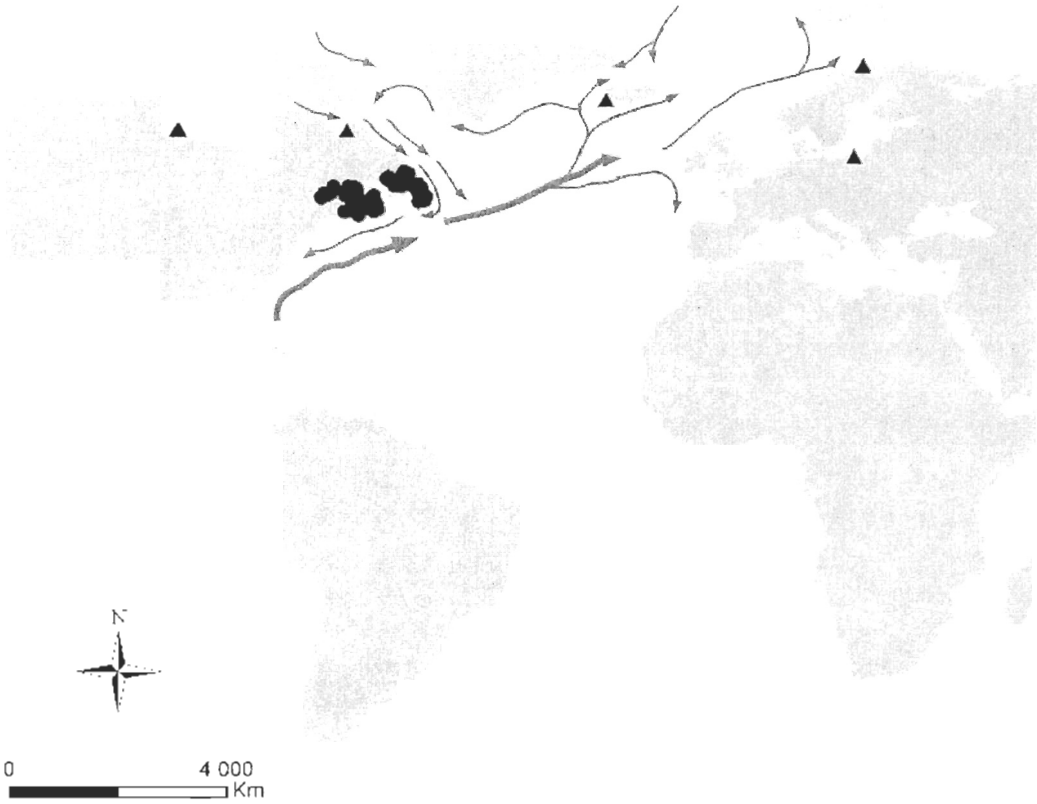


Figure 4.1 Collection sites for *Gammarus oceanicus* and the general circulation patterns for North Atlantic. Circles represent sites from Atlantic Canada (Southern cluster) while triangles are used for sites from the Arctic Canada and Europe (Northern cluster).

4.4.2 DNA extraction, amplification, sequencing

Genomic DNA was extracted from small amounts of muscle tissue, usually from one pereopod, preserving the rest of the organism as a voucher. The amplification and sequencing of the barcode region, a 658 bp fragment at the 5'-end of the COI gene, followed previously described protocols (Radulovici, Sainte-Marie and Dufresne, 2009). Two alternative sets of primers were used and their sequences

are available in BOLD: LCO1490 – HCO2198 (Folmer *et al.*, 1994) with M13 tails and CrustDF1 – CrustDR1 (Steinke, unpublished).

4.4.3 Data analysis: genetic diversity, structure and demographic history

DNA sequences were manually edited in Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI) and aligned in MUSCLE with the default settings in MEGA 5 (Tamura *et al.*, 2011). COI sequences were translated into amino acids in MEGA 5, to verify the reading frame and to assess the possibility of having amplified pseudogenes. Details regarding DNA sequences, trace files and amino acid translation can be found in BOLD within the projects GAMOC, WWGSL and FCGA, under the "Sequence Page". Pairwise genetic distances between COI haplotypes used the Kimura 2-parameter (K2P) evolutionary correction (Kimura, 1980) and were generated in MEGA 5.

Population structure was assessed with a two-step approach. First, COI haplotypes were used for Bayesian inference (BI) and maximum likelihood (ML) phylogenies. A closely related species, *Gammarus duebeni*, was used as outgroup. The most appropriate model of sequence evolution was chosen by running the dataset in jModelTest 1.0.1 (Posada, 2008) under the Akaike Information Criterion (AIC) (Posada and Buckley, 2004). The General Time Reversible (GTR) model with a proportion of invariable sites (+I) was used in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) to run two independent analyses, including 10 million generations and sampling every 1,000 generations. The initial 25% of samples were discarded as burn-in and the final consensus tree was rooted and edited in FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). An ML tree was built in RAxML 7.2.8 (Stamatakis, Hoover and Rougemont, 2008) a web-server application available through Vital IT unit of the Swiss Institute of Bioinformatics (http://phylobench.vital-it.ch/raxml-bb). Following the phylogenetic results, which revealed the existence of two divergent clusters, the next step included basic genetic structure analyses

performed on separate (e.g., each cluster) and combined datasets. Haplotype networks for each cluster were constructed in Network 4.6 (Bandelt, Forster and Rohl, 1999), which uses a median-joining algorithm to build parsimony networks. Atlantic Canada was sampled thoroughly in terms of geographic coverage (except for the Labrador coast) but with a small sample size per site, therefore we pooled sites into larger groups. Multiple approaches were used for finding genetically and geographically cohesive groups: spatial analysis of molecular variance, SAMOVA (Dupanloup, Schneider and Excoffier, 2002), discriminant analysis of principal components, DAPC (Jombart, Devillard and Balloux, 2010), Bayesian analysis of population structure, BAPS (Corander *et al.*, 2008). Since none of these methods gave a clear result, our final division was loosely based on the biogeographical zones of GSL (Brunel, Bossé and Lamarche, 1998) resulting in 10 groups (Table 4.1 and Figure 4.2). These artificial groups were treated as "populations" in subsequent analyses. Molecular diversity indices such as haplotype diversity (*Hd*) and nucleotide diversity (π) were calculated for each population in Arlequin 3.5 (Excoffier and Lischer, 2010) based on K2P distances. Geographic structure was tested by hierarchical analysis of molecular variance (AMOVA) with K2P distances and 10,000 permutations and Φ-statistics. The first AMOVA investigated the existence of genetic differentiation between the two clusters, while the second AMOVA tested for further potential subdivision in Atlantic Canada. For the latter, populations were grouped into three regions: northern GSL, southern GSL and open Atlantic coast. Pairwise Φ_{ST} population comparisons were calculated with haplotype frequencies and 10,000 permutations, taking into account K2P distances.

Table 4.1 Genetic diversity for pooled collection sites: sample size (N), number of haplotypes per population (H), haplotype diversity (*Hd*) and nucleotide diversity (π) with standard deviation between brackets

	Code	Populations	N	H	*Hd (SD)*	*π (SD)*
	Atlantic Canada					
1	ESL	Estuary	32	8	0.8165 (0.0358)	0.0029 (0.0019)
2	GAP	Gaspe Peninsula	16	8	0.8750 (0.0591)	0.0058 (0.0034)
3	SGF	Southern Gulf	27	10	0.8462 (0.0427)	0.0030 (0.0019)
4	PEI	Prince Edward Island	21	8	0.7952 (0.0677)	0.0032 (0.0020)
5	MIS	Magdalen Islands	37	4	0.2508 (0.0909)	0.0005 (0.0006)
6	NSH	North Shore	11	4	0.6727 (0.1232)	0.0035 (0.0023)
7	WNF	Western Newfoundland	8	6	0.8929 (0.1113)	0.0029 (0.0021)
8	ENF	Eastern Newfoundland	29	12	0.8744 (0.0380)	0.0050 (0.0029)
9	NSC	Nova Scotia	14	8	0.8242 (0.0977)	0.0033 (0.0022)
10	FYB	Fundy Bay	33	13	0.8958 (0.0295)	0.0076 (0.0042)
		South cluster	228	67	0.9507 (0.006)	0.0075 (0.0041)
	Arctic Canada					
11	CHU	Churchill	28	2	0.0714 (0.0652)	0.0001 (0.0002)
12	NQC	Northern Quebec	2	1	NA	NA
	Europe					
13	NOR	Norway	6	3	0.6000 (0.2152)	0.0021 (0.0017)
14	ICE	Iceland	7	1	NA	NA
15	POL	Poland	2	2	1.0000 (0.5000)	0.0032 (0.0038)
		North cluster	45	7	0.3576 (0.089)	0.0011 (0.0009)
		Total *G. oceanicus*	**273**	**74**	**0.9485 (0.006)**	**0.0155 (0.0079)**

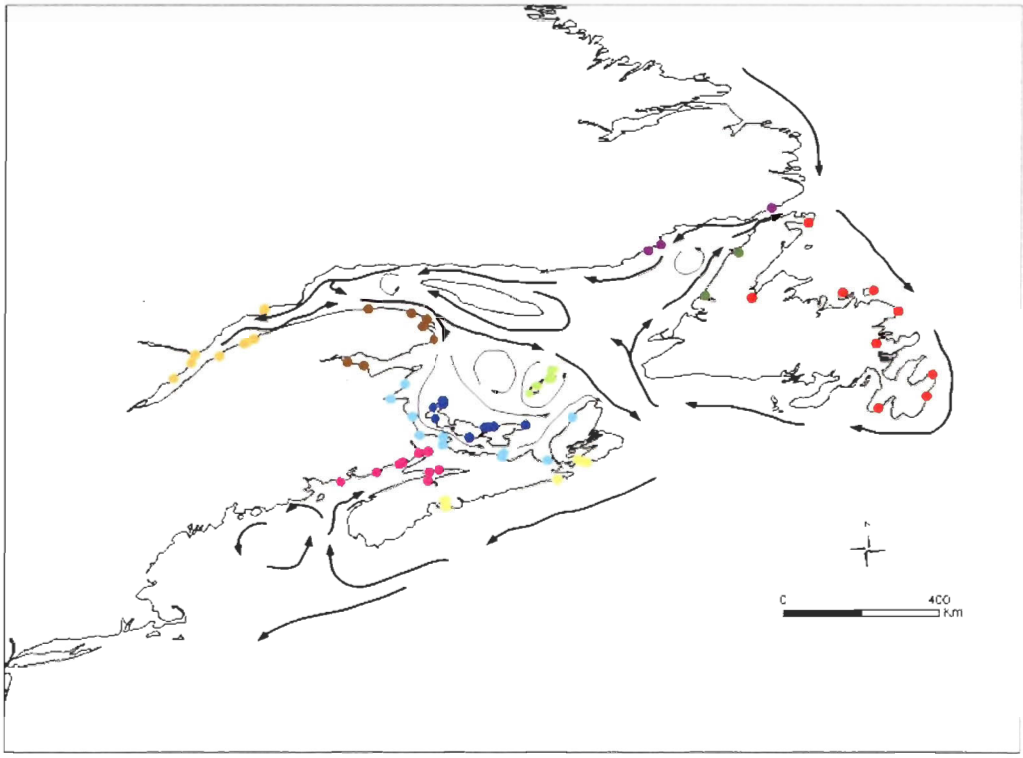


Figure 4.2 The Gulf of St. Lawrence with colored sites corresponding to our populations: ESL – orange, GAP – brown, SGF – light blue, PEI – dark blue, MIS – light green, WNF – dark green, NSH – violet, ENF – red, NSC – yellow, FYB – pink. Oceanography includes main currents (thick arrows), secondary currents (thin arrows) and gyres (circle arrows), according to DFO data.

Present-day genetic diversity can be influenced by demographic history. Mutation-drift equilibrium was tested by three analyses for both the combined and separate datasets: mismatch distribution, two neutrality tests, Tajima's D (Tajima, 1989) and Fu's F's (Fu, 1997), in Arlequin 3.5. The analysis of mismatch distribution of pairwise differences between COI sequences took into consideration the observed values compared against simulated values under a demographic expansion model and a spatial expansion model. Goodness-of-fit between observed and simulated data was tested by the sum of squared deviations (SSD) and the raggedness index (r) based on 10,000 permutations. The neutrality tests were performed in order to

infer recent population changes and their significance was tested with 10,000 permutations.

4.5 Results

4.5.1 COI – genetic diversity

A total of 246 amphipods were successfully sequenced (BOLD project code: GAMOC). Twenty-seven additional COI sequences of *G. oceanicus* were included in our analyses: ten sequences from Canada (provided by P. Hebert), three sequences generated in a previous study from Canada (Radulovici, Sainte-Marie and Dufresne, 2009, BOLD project code: WWGSL) and 14 published sequences from Canada, Iceland and Poland (Costa *et al.*, 2009; BOLD project code: FCGA). Although the last study included 33 *G. oceanicus* sequences, we chose only those with trace files in BOLD, good quality and longer than 620 bp. The final dataset consisted of 273 amphipods from 87 sampling sites with a range of 1-11 specimens per site. At the regional level there were 15 amphipods from four European sites, 30 from seven Arctic Canadian sites and 228 collected in 76 sites along the shores of cold-temperate Atlantic Canada. For the last region, the 10 populations (Figure 4.2) had a sample size varying between eight and 37 specimens (Table 4.1). Considering also published sequences (Costa *et al.*, 2009) from Maine (U.S.A), which were not included in our analyses for reasons mentioned above but compared to our dataset (data not shown), we fully covered the southern range of the species distribution in North America.

The majority of COI sequences spanned the full barcode length – 658 bp. However, the presence of a few shorter sequences resulted in a final trim to a uniform length of 621 bp. The alignment included 66 polymorphic sites and no indels or stop codons (indication of pseudogenes) were detected. A total of 69 mutations

formed 74 haplotypes (H1-H74). Most of these mutations were silent, occurring at the third position of codons. However, five mutations at the first codon position led to changes in the string of 207 amino acids. These non-synonymous mutations involved four transitions and one transversion, all in samples from Atlantic Canada. One transition occurred at codon 106 leading to a change from Valine (GTC) to Isoleucine (ATC) within H29, H33, H65 and H69. One transition (ATT – GTT) at codon 42 resulted in the change of Isoleucine with Valine within H52, while a transversion (TTA – GTA) in codon 87 changed the coded amino acid from Leucine to Valine within H73. The last two transitions involved a change from Glycine to Serine at codon 27 within H28 (GGA – AGA) and at codon 112 within H64 (GGT – AGT). K2P distances between haplotypes varied from 0.2±0.2% to a maximum of 3.0±0.7% and had a mean value of 1.0±0.2%.

4.5.2 Population structure

Both BI and ML trees had similar topologies and mainly showed a split of samples into two clusters: one distributed exclusively in Europe (Poland, Iceland, and Norway) and Arctic Canada (subsequently referred to as "Northern cluster" although Poland and GSL share similar latitude) and one distributed exclusively in Atlantic Canada (subsequently referred to as "Southern cluster") (Figure 4.1 and 4.3). Mean distances within clusters were 0.4±0.2% (north) and 0.8±0.2% (south), while between groups it reached 2.4±0.5%. The two groups consisted of 45 sequences and seven haplotypes in the north and 228 sequences and 67 haplotypes in the south, and no haplotype was shared between the two clusters.

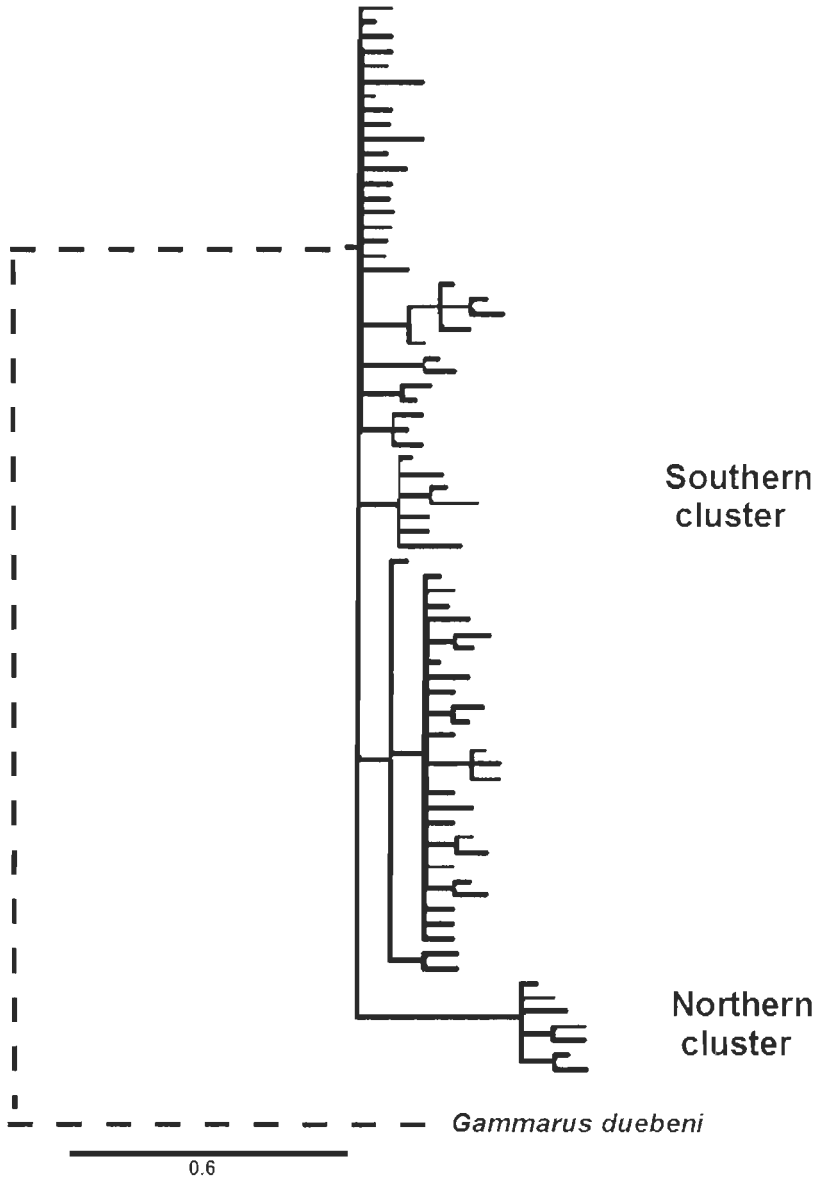


Figure 4.3 Phylogenetic tree for *G. oceanicus* based on COI haplotypes. The dashed line to the outgroup, *G. duebeni,* is not illustrated to scale. Although both BI and ML trees identified the same two divergent clusters (Northern and Southern), both of them had weak support.

The Northern haplotype network (Figure 4.4), while based on only a few haplotypes, showed more variability in Europe compared to Arctic Canada: two haplotypes from two specimens in Poland, three haplotypes from six specimens in Norway, while all six Icelandic specimens shared the same haplotype with almost all samples from Arctic Canada (both Hudson Strait and Hudson Bay). One common haplotype (H6) accounted for 80% of specimens, while 11% were singletons (i.e., haplotypes represented by one specimen) (Appendix B). The Southern haplotype network (Figure 4.5) showed a highly diverse group and a star-like phylogeny with two central haplotypes (H10 and H19), separated by two mutational steps and dividing the network into a group corresponding to the southern GSL and one corresponding to the northern GSL, Estuary and the open Atlantic coast. The two central haplotypes accounted for ~18% of specimens, another 18% were represented by singletons, while 14% shared the most abundant haplotype (H37, N=32) which was restricted to Magdalen Islands. Most haplotypes were connected by one mutational step, however many missing haplotypes were needed to connect all haplotypes and a few cases of homoplasy appeared as reticulation in the network.

Overall, haplotypic diversity *Hd* was high (0.9485±0.0060) and the nucleotide diversity π was moderate (0.0155±0.0079). At the regional level, *Hd* was high in the south (0.9507±0.0060) but low in the north (0.3576±0.0890), while π was moderate in the south (0.0075±0.0041) and low in the north (0.0011±0.0009) (Table 4.1).

The AMOVA analysis conducted for Atlantic Canada showed that around half of the variation occurs within populations, and the rest is shared at higher levels (within and between groups) (Table 4.2). Pairwise Φ_{ST} between populations showed high levels of genetic differentiation between populations (Table 4.3).

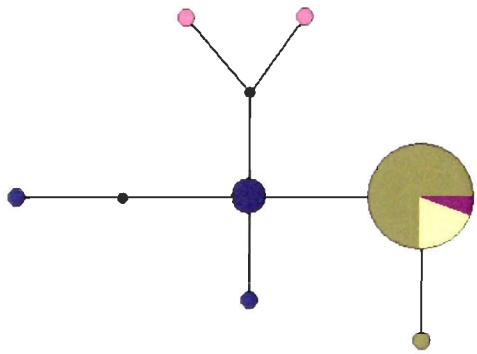


Figure 4.4 Median-joining haplotype network for the Northern cluster. Circles represent haplotypes, their size being proportional to the subset of samples exhibiting the particular haplotype, black dots are missing haplotypes, and lines represent mutational steps. Colors represent different populations: NOR – dark blue, POL – light pink, ICE – yellow, NQC – purple, CHU – brown.

Table 4.2 Hierarchical analysis of molecular variance (AMOVA) for *G. oceanicus*

Structure tested	Source of variation	% Variance	Φ-statistics[a]
Canada+Europe	Among groups	74.88	0.74884
	Among populations within groups	11.46	0.45638
	Within populations	13.65	0.86346
Atlantic Canada	Among groups[b]	24.25	0.24255
	Among populations within groups	23.07	0.30460
	Within populations	52.67	0.47327

[a]All fixation indices were significant at P<0.01

[b]Groups: Southern Gulf (SGF, PEI, MIS), Northern Gulf (EST, GAP, NSH, WNF), open Atlantic coast (ENF, NSC, FYB).

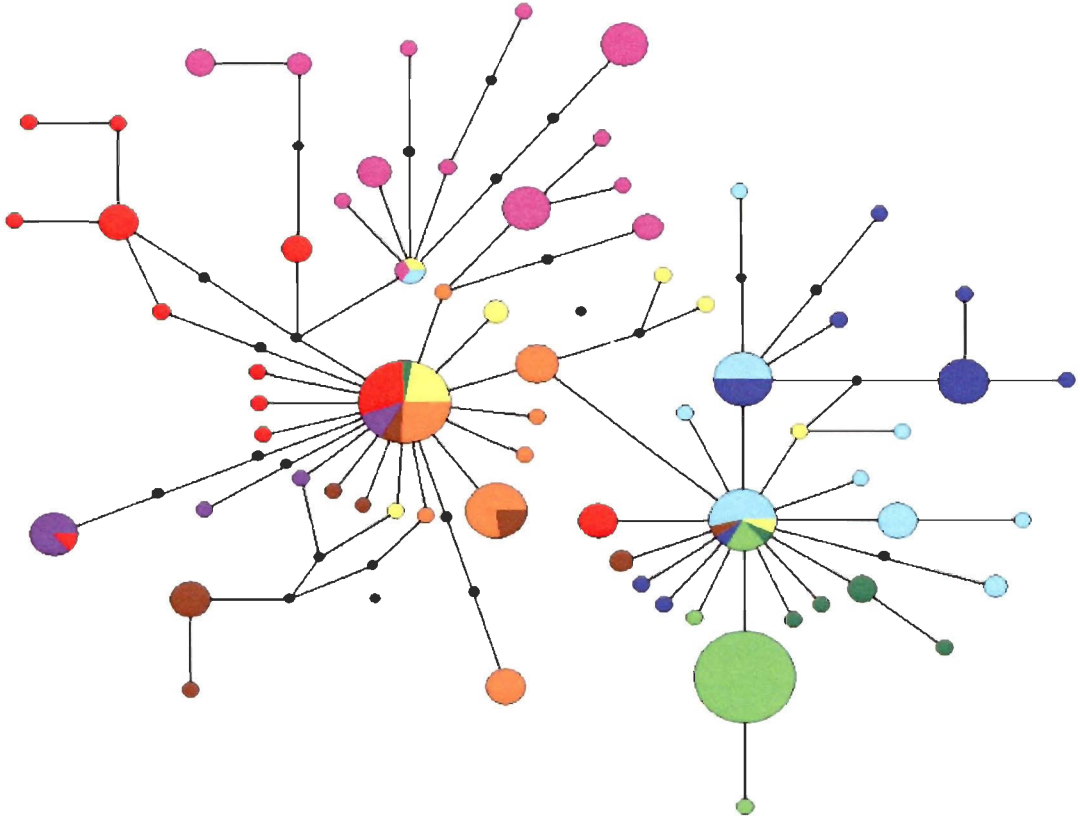


Figure 4.5 Median-joining haplotype network for the Southern cluster. Each circle represents one haplotype (the size corresponds to the number of individuals sharing that haplotype), each line represents one mutation step, and black dots are missing haplotypes. Colors represent different populations, identical to the ones used in Figure 4.2.

Table 4.3 Pairwise Φ_{ST} values between populations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	0.195													
3	0.492	0.426												
4	0.609	0.528	0.290											
5	0.720	0.677	0.488	0.686										
6	0.346	0.303	0.589	0.661	0.837									
7	0.436	0.344	0.142	0.414	0.662	0.554								
8	0.182	0.213	0.380	0.496	0.624	0.278	0.342							
9	0.095	0.144	0.388	0.538	0.731	0.309	0.354	0.091						
10	0.228	0.230	0.440	0.517	0.602	0.272	0.363	0.176	0.161					
11	0.861	0.787	0.878	0.875	0.969	0.857	0.883	0.778	0.854	0.686				
12	0.866	0.778	0.882	0.877	0.977	0.855	0.871	0.781	0.854	0.684	0.439			
13	0.887	0.829	0.901	0.902	0.982	0.903	0.932	0.809	0.895	0.718	0.627	0.894		
14	0.923	0.905	0.936	0.941	0.986	0.952	0.969	0.872	0.944	0.800	0.802	0.945	-0.072*	
15	0.868	0.778	0.883	0.879	0.980	0.861	0.891	0.774	0.858	0.669	0.400*	0.667*	0.000*	-0.332*

* Non-significant P-values (P>0.05) after 10,000 permutations.
Populations: 1 – EST, 2 – GAP, 3 – SGF, 4 – PEI, 5 – MIS, 6 – NSH, 7 – WNF, 8 – ENF, 9 – NSC, 10 – FYB, 11 – NOR, 12 – POL, 13 – ICE, 14 – NQC, 15 – CHU.

4.5.3 Demographic history

Graphic representations for mismatch distributions showed unimodal distributions, more accentuated in the Southern cluster where the sample size was higher (Figure 4.6). The goodness-of-fit tests (SSD and r) had non-significant *P*-values, therefore we cannot reject a spatial expansion model in both clusters. Both neutrality tests, Tajima's D and Fu's F's, had negative values and were significant, indicating recent mutations due to demographic expansion or selective sweeps in both clusters (Table 4.4).

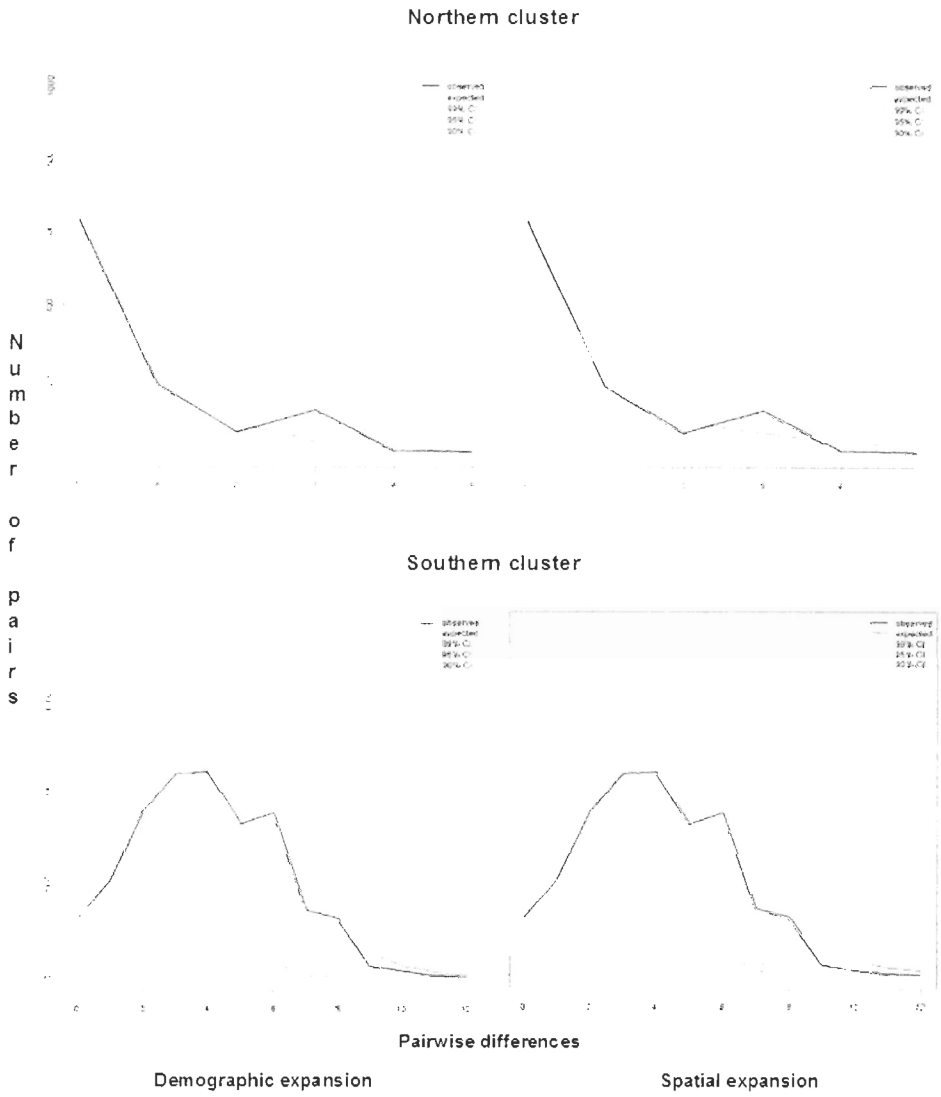


Figure 4.6 Mismatch distribution for each cluster of *G. oceanicus*. Black lines: observed values, grey lines: expected values. Colored dashed lines represent confidence intervals (90%, 95%, 99%).

Table 4.4 Historical demography parameters with 95% confidence intervals ranges and P-values between brackets

Model	Parameter	North cluster	South cluster	Total
Sudden expansion				
	Tau	3.000 (0.082- 4.250)	3.871 (2.010-6.422)	1.883 (0.305-13.619)
	Theta 0	0 (0-0.009)	0.519 (0-1.935)	5.258 (0-13.039)
	Theta 1	0.545 (0-inf)	25.527 (11.823-inf)	42.461 (11.636-inf)
	SSD	0.010 (P=0.43)	0.002 (P=0.5)	0.014 (P=0.29)
	r	0.242 (P=0.52)	0.016 (P=0.57)	0.019 (P=0.17)
Spatial expansion				
	Tau	2.140 (0-12.964)	3.230 (1.801-5.442)	1.795 (0.578-14.298)
	Theta	0.324 (0.001-0.968)	1.017 (0.001-4.193)	5.170 (0.001-12.447)
	M	0.295 (0-inf)	31.941 (12.904-inf)	38.393 (0.964-inf)
	SSD	0.006 (P=0.55)	0.002 (P=0.5)	0.014 (P=0.27)
	r	0.242 (P=0.62)	0.016 (P=0.66)	0.019 (P=0.19)
	Tajima's D	-1.604 (P=0.03)	-1.863 (P=0)	-1.130 (P=0.11)
	Fu's F	-3.741 (P=0.01)	-25.569 (P=0)	-24.591 (P=0.0003)

4.6 Discussion

4.6.1 Origin and genetic divergence: amphi-Atlantic structure

G. oceanicus is a species adapted to a cold climate (Steele and Steele, 1972) and it might have evolved during changing conditions of late Pliocene (Golikov and Tzvetkova, 1972). During the Quaternary climate change with its sea level and ice cover fluctuations, the northern *G. oceanicus* populations followed an extinction-recolonization (from the south) pattern which finally led to the present-day distribution of this species. The phylogenetic analysis (Figure 4.3) identified two clusters that are segregated latitudinally (Europe and Arctic Canada versus Atlantic Canada) (Figure 4.1). Populations belonging to these groups showed high Φ_{ST} values (0.66-0.98) (Table 4.2) and differentiation in the AMOVA analysis (74% of variation occurred between the two groups) (Table 4.2). Moreover, the clusters were separated by mean pairwise distances of 2.4% for COI, a value which might be indicative of species boundary (see 2.8% in *Hyas* spp.; Radulovici, Sainte-Marie and Dufresne, 2009). By using the sequence divergence and a molecular clock commonly used in crustaceans (~2% per MY; Raupach *et al.*, 2010a), the separation time can be estimated at roughly 600KYA, during the Pleistocene which was an epoch of rapid radiation for *Gammarus* spp. in general (Steele and Steele, 1972). The use of a molecular clock assumes uniform mutation rates along lineages and it is still a debated issue (Emerson, 2007). Amphipods lack good fossil records and no molecular clock has been calibrated, hence our use of the common crustacean clock of ~2%/MY. However, some crustacean groups might evolve faster than others while northern and southern populations of the same species might have a different number of generations per year translated into variation in the intraspecific evolutionary rate (Thomas *et al.*, 2010). Nonetheless, in *G. oceanicus* even by using faster clocks, as proposed in other amphipods (9.6%/MY; Henzler, 2006) or in

mysids (27%/MY, Audzijonyte and Vainola, 2006), the separation time between the two groups is placed before LGM (125 KYA and 44 KYA respectively,).

Consequently, the two intraspecific clusters originated sometime in the Pleistocene when the coastal habitat became unsuitable for *G. oceanicus* and populations receded to ice-free areas in Europe as well as in North America, from which they recolonized northern areas during the following interglacial period. After LGM, both clusters expanded their ranges as shown by the analysis of mismatch distribution and the neutrality tests (Table 4.4, Figure 4.6). The lack of shared haplotypes indicates a clear genetic isolation of clusters even after the retreat of the ice sheets.

The Northern cluster included fewer samples, fewer haplotypes, lower haplotype and nucleotide diversity compared to the Southern cluster, a genetic pattern that might indicate recent expansion from a small refugium or a post-glacial bottleneck. Moreover, the most common haplotype (H6) was shared by CHU, NQC and ICE, even though these populations are separated by thousands of kilometers of deep water, land masses and strong currents (Figure 4.1), indicating a European source for Canadian Arctic populations. Recent and rapid colonization of large territories is usually explained by species dispersal capacity. As with all amphipods, *G. oceanicus* is a direct developer (i.e., eggs hatch into juveniles) and lacks a pelagic dispersive larval phase, which may favour large-scale dispersal via currents. Although a good swimmer, *G. oceanicus* is restricted to active dispersal only at small scales during high tide, in shallow infralittoral waters (Ingolfsson and Agnarsson, 2003), but is incapable of surviving and dispersing along the deeper ocean bottoms. Therefore, this species must have succeeded in rapidly colonizing Arctic Canada from Europe via Iceland by means of passive dispersal. Amphipods can be transported between sites by rafting in clumps of detached algae (Ingolfsson, 1995), phoretic associations with other animals (birds, aquatic mammals) or by humans (e.g., shipping). While rafting is considered a common way by which invertebrates,

including amphipods (Henzler, 2006), colonize new habitats (review in Thiel and Gutow, 2005), its success is highly dependent on physical (currents, winds, temperature), as well as biological (food source, competition and predation while rafting) factors (Vandendriessche, Vincx and Degraer, 2007). Moreover, successful rafting (i.e., reaching a new site) does not necessarily imply successful colonization (i.e., reproduction and propagation in the new site). In *G. oceanicus*, rafting from Iceland to Labrador and Newfoundland might be possible, especially in the early post-glacial period when the current system had an opposite pattern (from east to west); this scenario should lead to introgression of northern haplotypes into the Southern cluster which was not identified in this study. However, rafting from Iceland to Hudson Bay seems less probable (Figure 4.1) and shipping activities are too recent in the Arctic to explain colonization. Therefore, we consider passive dispersal through seabirds to be the most probable mechanism of dispersal between ICE and CHU. Seabirds have yearly migrations between Europe and Canada with southwest Greenland being an important wintering ground for many species breeding on both continents (Boertmann *et al.*, 2004). Birds feed on intertidal invertebrates, including amphipods, hence a possibility for external (on feathers) or internal (digestive tract) transportation. These mechanisms have not been investigated in marine intertidal amphipods yet. However, studies have found some freshwater amphipods, including *Gammarus* spp., to be transported over land in bird feathers (Swanson, 1984 and references therein). Other crustacean species (copepods, branchiopods) were able to survive as eggs inside the digestive tract and hatch afterwards (review in Figuerola and Green, 2002), while recent findings have shown adult snails giving birth to juveniles after surviving through a bird's gut (Wada, Kawakami and Chiba, 2011). In amphipods, internal transport as eggs is improbable due to parental care (i.e., females carrying eggs and then juveniles in the brood pouch). Beavers, muskrats or dogs have been mentioned as carrying amphipods in their fur (Swanson, 1984 and references therein). In *G. oceanicus*, the most probable candidates for smaller-scale dispersal in northern habitats (e.g., within Hudson Bay) would be polar bears which feed in the intertidal area where amphipods might attach

or penetrate their fur. Although speculative at this point, dispersal of *G. oceanicus* through phoretic associations should not be discarded. After all, there is a lack of knowledge on the potential for amphipod colonization through infrequent transport by birds carrying berried females or potential amphipod mates.

4.6.2 Glacial refugia in North Atlantic

Traditional views on LGM considered that massive thick ice sheets covered extensive parts of the northern continents (Figure 1 in Hewitt, 2000), therefore northern coastal species disappeared or receded to southern refugia on both coasts. However, there is geologic evidence for ice-free areas in the north (not indicative of biological survival though; see Brochmann *et al.*, 2003), and multiple and controversial locations for coastal glacial refugia have been proposed.

Coastal refugia were believed to have been common in GSL (around Gaspe Peninsula, west coast of Newfoundland, Pielou, 1991; Magdalen Islands, Prest *et al.*, 1976) but the latest reconstructions of the ice sheet give alternative refugia on the Atlantic continental shelves (Grand Banks, Georges Bank and the Flemish Cap) while the Gulf seems to have been completely covered by ice (Shaw, 2006). The former two banks became coastal plains during LGM due to a low sea level and acted as potential refugia for entire coastal and terrestrial communities. By contrast, the Flemish Cap was still below the sea level (-10 m) and although *G. oceanicus* is a marine species living in shallow waters (-25 m), it is associated with intertidal seaweeds at least in the juvenile stage. Therefore, the Flemish Cap is less likely to have been a valid refugium for this species. In this context, when ice began to break up and melt, GSL was probably rapidly re-colonized by active or passive dispersal through the Cabot Strait (as Belle-Isle was still blocked by ice), starting with the southern regions and ending with the Estuary, the last region to be deglaciated (Shaw *et al.*, 2006).

Proposed marine refugia in Europe include Iceland, northern Norway, southwest Ireland, the English Channel (Hurd Deep), the Mediterranean, Iberian Peninsula and the Azores (review in Maggs *et al.*, 2008). While Iceland has been proposed based on genetic patterns found in an isopod species (*Idotea balthica*; Wares and Cunningham, 2001) and would be consistent with our data on *G. oceanicus*, this possibility has been dismissed for coastal species based on geologic evidence (Ingolfsson, 2009), although groundwater amphipods may have survived there (Kornobis *et al.*, 2010). Consequently, it is believed that both Iceland and Canada were rapidly colonized by coastal species surviving in other European refugia, which took advantage of the post-glacial sea-current system to move from east to west (Ingolfsson, 1992). A boreal species adapted to live in shallow water with seaweeds, *G. oceanicus* probably survived in suitable habitats in the English Channel and the Irish coast.

The long-term persistence of *G. oceanicus* in NWA is indirectly supported by phylogeographic patterns of seaweeds, its main food source or habitat. Olsen *et al.* (2010) showed genetic patterns consistent with amphi-Atlantic survival of *Ascophyllum nodosum*, while other seaweeds apparently survived in southern European refugia and only recently colonized NWA (*Fucus vesiculosus*, Muhlin and Brawley, 2009; *Chondrus crispus,* Hu *et al.*, 2010), probably facilitating dispersal and colonization of its associated fauna capable of rafting.

4.6.3 Genetic structure in Atlantic Canada

The Southern cluster was restricted to Atlantic Canada. DNA sequences belonging to *G. oceanicus* from Maine (data not included here; Costa *et al.*, 2009) were found to share the same haplotype with some FYB samples, therefore we are confident that we covered the southern distribution range of this species and no European haplotypes occur in NWA.

High genetic diversity, the star-like phylogeny with many rare haplotypes and demographic analyses are indicating rapid post-glacial expansion while the presence of two central haplotypes might indicate secondary contact between populations surviving in separate refugia. As Long Island Sound is the southern limit for *G. oceanicus* distribution and also for the Laurentide Ice Sheet, our results are indirect evidence for survival in some NWA periglacial refugia. Pairwise Φ_{ST} (Table 4.3) and the AMOVA analysis (Table 4.2) showed a high level of present-day genetic structuring in NWA, although the phylogenetic tree lacked resolution in finding genetically differentiated clusters (both clusters had weak support) (Figure 4.3). Genetic structure is the consequence of limited gene flow between populations, despite the potential for rafting, phoretic associations and human-mediated transport at the scale of Atlantic Canada. Previous studies on marine invertebrates found various genetic patterns in NWA (especially Atlantic Canada and Gulf of Maine) from lack of structure in the sea cucumber, *Cucumaria frondosa* (So *et al.*, 2011), snow crab, *Chionoecetes opilio* (Puebla *et al.*, 2008), sea urchin, *Strongylocentrotus droebachiensis* (Addison and Hart, 2004) to certain levels of genetic differentiation in lobster, *Homarus americanus* (Kenchington *et al.*, 2009) and barnacles, *Semibalanus balanoides* (Dufresne, Bourget and Bernatchez, 2002). However, none of the previous studies had a thorough sampling of GSL and the adjacent Atlantic coast and the targeted species had a pelagic developmental phase. This study indicates that an intertidal species with direct development survived during LGM in NWA although species with a larval phase were considered favorites for escaping harsh conditions in the north due to their potential for large-scale dispersal (Faurby *et al.*, 2011). It also shows genetic structure at small-scale which is concordant with limited active dispersal (as direct developer) but in contrast with the potential for large-scale passive dispersal. However, fast evolving nuclear markers (e.g., microsatellites) are needed in order to assess the level of population connectivity at fine spatial scale in NWA.

4.6.4 Magdalen Islands

The isolated Magdalen archipelago lies in the center of the GSL, being separated by ~90 km from the closest landmass (PEI) (Figure 4.2). A thorough sampling of most of the largest islands revealed the existence of three haplotypes restricted to the archipelago, two singletons (H38, H39) and one common haplotype (H37) found at every sampling site (Appendix B). A fourth haplotype (H19) had a central position in the Southern network (Figure 4.5) and its presence demonstrates the close genetic relatedness (one mutation) between the present-day insular population and one of the surviving ancestral populations. There are two alternative explanations for this pattern: a population surviving in one of the glacial refugia outside GSL started to colonize southern GSL immediately after deglaciation (see sections above) and one haplotype founded the future insular population (the founder effect) or a glacial refugium was actually situated in the Magdalen archipelago and the colonization of the southern GSL started from there when the ice broke-up and melted. The hypothesis of an unglaciated Magdalen archipelago during the Pleistocene (Prest *et al.*, 1976) has led to its consideration as a glacial refugium for insects (Hamilton, 2002) or small mammals (Youngman, 1967). This might be considered a valid argument when designing scenarios for rapid colonization of the Gulf following deglaciation (faster colonization from inside GSL rather than from outside, the heavy flow of outgoing ice-melt and icebergs making it difficult to move upstream into the Cabot Strait (Figure 3 in Shaw *et al.*, 2006). However, there is no geologic evidence in coastal sediments to show that large areas of GSL remained ice-free (Bernard Hétu, UQAR, pers. comm.). On the other hand, there is no information on the minimum space required for the survival of a minimum viable population of *G. oceanicus* and GSL has not been thoroughly investigated for geological evidence of glacial refugias. Consequently, although we follow the most accepted scenario of a completely ice-covered GSL and glacial refugia only in the Maritimes (Grand Banks, Georges Bank), we do not exclude the possibility of a refugium inside GSL.

Regardless of the location of the glacial refugium and the exact origin for the insular amphipod fauna, the present-day genetic structure of *G. oceanicus* indicates genetic isolation of the geographically isolated Magdalenian population from the other populations in Atlantic Canada. Therefore the potential dispersal mechanism through rafting at the GSL scale might be hindered by oceanographic features such as local currents and gyres (Figure 4.2) or by some biological features (e.g., food limitation affecting survival during rafting). Moreover, based on currents and wind patterns, there should be migration through rafting especially from the tip of the Gaspe Peninsula or from northeast PEI/Cape Breton, less likely from WNF or NSH, but we found no evidence for such migration patterns. The fourth haplotype in MIS (H19) is shared between Old Harry Harbour/Brion Island (MIS) and the Northumberland Strait (southwest PEI, SGF), Chaleur Bay, WNF and NSC, the least likely sources of potential migrants for the Magdalen fauna. Therefore, the Magdalen population might be completely isolated from the surrounding populations, being in the slow process of differentiation and speciation and should be investigated further with fast-evolving markers.

4.6.5 Present-day barriers to dispersal

Glaciations divided the ancient distribution range of *G. oceanicus* and the present-day genetic structure can be easily interpreted as the result of vicariance. However, the lack of mitochondrial introgression between groups (across the Atlantic) remains a puzzle. With numerous species re-colonizing NWA via passive dispersal from Europe through Mid-Atlantic islands as stepping stones (Ingolfsson, 1992; Wares and Cunningham, 2001; Henzler, 2006; Ilves *et al.*, 2010), there is no obvious explanation for the European *G. oceanicus* colonizing Arctic but not Atlantic Canada. Seabird-mediated dispersal between GSL and Hudson Bay is hampered by the feeding behavior of birds during the migration, namely stop-over at inland lakes, which does not permit the viable transport of intertidal marine amphipods. Human-mediated dispersal led to successful recent invasions of species belonging to

Gammarus sensu lato into the non-native Atlantic coast (*G. tigrinus* in Europe, *Echinogammarus ischnus* in NWA). One possible explanation for the lack of introgression between *G. oceanicus* clusters would be ecological divergence of the amphipods' ability to use food sources (e.g., seaweed species). Indeed, recent work showed such local adaptation in another marine grazer, *Idotea balthica*, in which local populations adapt to their host (*Fucus* versus *Zostera*) resulting in parallel divergence (Vesakoski *et al.*, 2009). As our study was mainly focused on Atlantic Canada, our sampling was quite intensive towards the species southern range limit but scattered in the northern areas. Therefore we lack samples from the contact area between clusters, namely the Labrador coast (Figure 4.1). Samples from this region would add valuable information regarding a genetic contact zone and the processes maintaining it.

4.7 Conclusions

This study showed the existence of two divergent intraspecific clusters for the common intertidal amphipod *Gammarus oceanicus* with amphi-Atlantic distribution. These two clusters did not correspond to a European – North American separation but rather to a certain latitudinal segregation between north and south. COI sequences showed high divergence (2.4%) and no shared haplotypes between clusters, an indication of potential cryptic species. Ancestors of the Northern cluster probably survived in glacial refugia in Europe and began a colonization process into Arctic Canada soon after deglaciation, possibly by multiple means of dispersal (e.g., birds, rafting by algae) and via stepping stones in the North Atlantic (Norway, Iceland, Greenland). Ancestors of the Southern cluster probably survived in two glacial refugia in or south of the Canadian Maritimes and colonized from there and separately the southern GSL and the northern GSL together with the Estuary. The missing link between the two clusters is the remote coast of Labrador. Therefore, sampling along this coast is vital in order to clarify the distribution range (overlapping

or not) of the two clusters and the possibility of hybridization (e.g., mating trials). In addition, fast evolving genetic markers (such as microsatellites) would help clarify the genetic structure at small spatial scales.

DNA barcoding can reveal the deep splits within morphological species, indicative of cryptic (incipient) speciation. While it is a tool for species identification and not for population studies, the large number of DNA barcode data being generated at the global level (>1.3 million in BOLD, August 2011) will have major implications for other types of research such as comparative phylogeography of co-distributed species. By its large-scale approach, DNA barcoding has an unprecedented role in generating exploratory data on which general hypotheses on genetic diversity will be formulated and subsequently tested with "confirmatory approaches" (Jaeger and Halliday, 1998).

Acknowledgements

We are grateful to Sean Locke for collecting amphipods in Ungava Bay, to Lyne Morissette and Silje Ramsvatn for Norwegian samples, to Traian Brad, Fred French and Robert Chabot for valuable assistance during fieldwork, to Pierre Brunel for taxonomic help and to Paul Hebert for sharing some unpublished sequences. Staff at the Biodiversity Institute of Ontario helped during all the steps of this project. Fieldwork in the Magdalen Islands (QC) was funded by CERMIM. This work is a contribution to the Canadian Barcode of Life Network, as well as to the Canadian Healthy Oceans Network, both funded through the Natural Sciences and Engineering Research Council.

GENERAL CONCLUSIONS

Overview: context and originality

In the context of a "biodiversity crisis" combined with the "taxonomic impediment", there is a need for a fast inventory of global diversity in order to design viable conservation actions. DNA barcoding is such an inventory tool, providing fast, reliable and cost-effective species identification. Libraries built through barcoding projects are rapidly accumulating at the global level and the iBOL project has as objective to provide 5 million barcodes for 500,000 species by 2015.

The general goal of my PhD thesis was to use molecular methods (specifically DNA barcoding generating COI sequences) as a means to assess biodiversity in the marine environment. A specific goal consisted in testing the efficacy of DNA barcoding in marine crustaceans from the North Atlantic with the implicit result of providing a reference library of COI sequences. As with every barcoding study, it included an inherent test of species hypothesis (i.e., does every traditional species consist of only one cluster of highly similar sequences?). At the species level of biodiversity, this study focused on detecting the existence of potential cryptic species in five crustacean orders (Amphipoda, Isopoda, Mysida, Decapoda, Euphausiacea), as well as assessing the monophyletic/polyphyletic nature of genera within one amphipod family (Talitridae). Implications of such tests concern biodiversity indices such as species richness and taxonomic distinctness. At the genetic diversity level, the goal of this study was to reveal patterns of genetic structure in the common intertidal amphipod, *Gammarus oceanicus*, with an amphi-Atlantic distribution. This is a study investigating genetic patterns of biodiversity and not the processes responsible for creating various patterns, which are more difficult to be inferred due to confounding factors.

This is the first DNA barcoding study for marine crustaceans from the NWA. It is also the most comprehensive study on crustacean diversity (i.e., species richness) based on molecular methods. It involves the most thorough geographic sampling in NWA both for species (>200 sites for the entire project) and at the genetic level (87 sites for *G. oceanicus*) investigations. Also, the taxon sampling is very diverse, including 92 species encompassing five orders. The chapters presented here have each an original side. Chapter I is the first comprehensive review on the role of DNA barcoding for marine biodiversity. Chapter II is the first study on barcoding marine crustaceans in the NWA, specifically from one geographic area, namely the St. Lawrence estuarine and marine system. Chapter III presents the most comprehensive phylogenetic analysis for Talitridae based on molecular data, in addition to providing a barcoding library for this family. Chapter IV is one of the first studies to use DNA barcodes (i.e., COI sequences generated during large-scale barcoding studies) beyond species identification in marine crustaceans, by providing a phylogeographic analysis for one of the most common and abundant intertidal amphipods and with amphi-Atlantic distribution.

Overview: main findings at two biodiversity levels

Invasive species

One of the unexpected findings of this study was the detection of an invasive species in ESL, *Echinogammarus ischnus* (Figure 13 D). It is an amphipod native to the Ponto-Caspian basin which expanded its distribution range to western Europe and North America through shipping activities during the past century (Cristescu *et al.*, 2004). While in the native range this species shows genetic differentiation concordant with geographic isolation between basins (e.g., Black and Caspian Seas) and limited dispersal capability (as amphipods are direct developers), the invaded range includes very similar populations at the genetic level indicating colonization

from a small source of individuals originating in the northern Black Sea (Cristescu *et al.*, 2004). In North America it has been previously reported from the Great Lakes, Detroit River and St. Lawrence River, near Montreal (Witt, Hebert and Morton, 1997; Cristescu *et al.*, 2004; Palmer and Ricciardi, 2004) and this study identified a new locality, Berthier-sur-Mer, a few hundred km downstream from Montreal, thus a range expansion in this species. Previous studies have shown that *E. ischnus* competes for resources with the native *Gammarus fasciatus*, replacing the latter in some areas (Palmer and Ricciardi, 2004) but the overall effect on the local food webs is unknown.

A routine barcoding study revealed a case of range expansion for an invasive species (but see Chapter II for details on how this positive match was possible), hence the practicality of barcoding. The ability to identify invasive species, especially in the initial phases of settlement, will be of great help in decision-making related to limiting the spread of non-native guests. Invasive species are considered to be the second greatest threat to biodiversity after habitat destruction (www.iucn.org), but are they really so negative? Invasion is a natural process that has occurred since life appeared on Earth, shaping the present-day distribution and genetic make-up of many species. For example, at least 12 marine interchanges took place during the last 25 million years and the one caused by the opening of the Bering Strait resulted in a large invasion of the North Atlantic by North Pacific taxa with an interesting evolutionary consequence: ~47% of Atlantic species with Pacific origin are now distinct from their ancestors (Vermeij, 2005). A large part of our food, livelihood and aesthetic life is based on introduced species and some of these unpopular life forms actually have positive effects on native diversity (Davis *et al.*, 2011) or both positive and negative effects on different native taxa (Briggs, 2007 and references therein), therefore the human bias that "non-native" (alien, exotic, invasive, introduced) equals "harmful" does not always hold true.

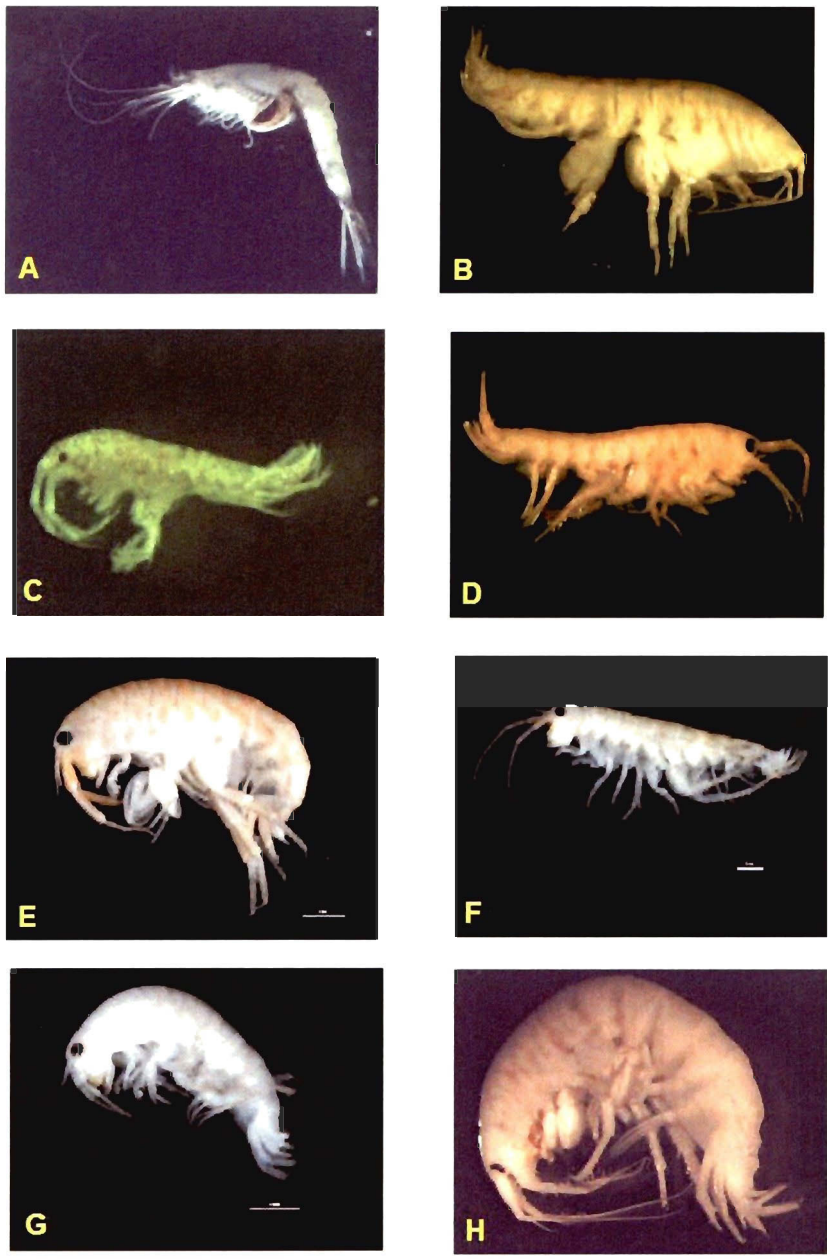


Figure 13 Some of the species mentioned in Chapters II, III and IV. A: *Neomysis americana*; B: *Ampelisca eschrichtii*; C: *Ischyrocerus anguipes*; D: *Echinogammarus ischnus*; E: *Platorchestia platensis*; F: *Orchestia grillus*; G: *Tethorchestia* sp. B; H: *Gammarus oceanicus*.

The "invasion" scale is global (Figure 14) and considered to have increased so drastically in the last century that we might ask ourselves not "What species is invasive?" but rather "What species is actually native?" (Carlton, 1989). With sailing activities between seas since the oldest historical times and a lack of archaeological species checklists, it might be difficult to identify the true origin of species in some cases. Davis *et al.* (2011) recently proposed a more practical view on invasion by considering the environmental impact rather than the origin of a species. This new approach is embedded in the general view that communities (natural and cultural) are continuously evolving, with a mix of long-term and new residents having an impact on each other and building together new forms (ecosystems, cultures), therefore it might be useless (in terms of funding and outcome) to try to recreate some previous "rightful" state if the newcomers are not harmful to the locals (Davis *et al.*, 2011).

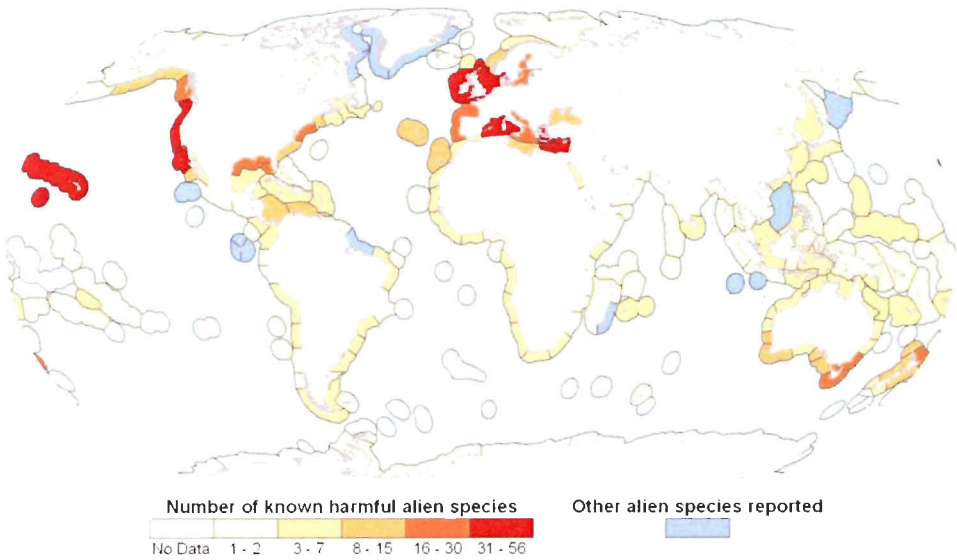


Figure 14 Distribution map for invasive species according to marine ecoregions. Darker red shades indicate higher number of harmful species and greater impact on native communities while dark blue indicates ecoregions with less harmful species. (Source: Molnar *et al.*, 1998)

Underestimation of species richness

Chapters II-III showed that DNA barcoding is a useful tool for species identification in crustaceans because morphological species usually correspond to clusters of similar COI sequences separated by large genetic distances ("barcoding gaps") from other species.

The 460 specimens barcoded in Chapter II belonged to 80 species, 56 genera, 36 families and five malacostracan orders (Amphipoda, Isopoda, Mysida, Euphausiacea, Decapoda) and they represented only 20% of about 400 crustacean species inventoried within ESL and GSL (Brunel, Bossé and Lamarche, 1998). They also represent the most common (Brunel, Bossé and Lamarche, 1998) and most mobile (Sainte-Marie and Brunel, 1985) species. Four species showed intraspecific clusters with divergences greater than 3% (the proposed threshold for species delineation; Hebert *et al.*, 2003) or the 10× mean intraspecific value (Hebert *et al.*, 2004). These special cases (5% of all species analyzed) included two amphipods (*Ampelisca eschrichtii*, *Ischyrocerus anguipes*), one mysid (*Neomysis americana*) (Figure 13, A-C) and one decapod species (*Spirontocaris spinus*). Except for *A. eschrichtii*, all other species seem to present sympatric intraspecific clusters. However, such allegations are difficult to make when working with marine species, some of them collected by trawling, baited traps, plankton nets, hence lacking precise details on microhabitat. Therefore, habitat specialization leading to sympatric diversification is difficult to test in the sea.

The 218 specimens barcoded in Chapter III belonged to 15 species, 8 genera and two families, one of them (Hyalidae) used as outgroup in phylogenetic analyses. The main family investigated here was Talitridae, the only amphipod family with both aquatic and terrestrial distribution. Talitrids are a species-rich group and many more species (hundreds to thousands) await discovery, while many undescribed taxa have been collected decades ago and remain in museum collections awaiting formal

description. In this group, three species (20% of the total) showed divergent intraspecific clusters indicating potential cryptic species. *Platorchestia platensis* and *Tethorchestia* sp. B (Figure 13, E, G) showed intraspecific allopatric segregation, while *Orchestia grillus* Figure 13, F) encompassed seven clusters of which two were apparently sympatric (note: as they were represented by single individuals and very divergent from the other clusters, they might be pseudogenes). The 273 specimens included in Chapter IV and used for phylogeographic analyses in *Gammarus oceanicus* revealed the existence of two intraspecific clusters (see the next sub-section).

DNA barcoding detected potential cryptic speciation occurring in eight morphological species encompassing a total of 23 clusters, hence a total of 15 unknown clusters (i.e., potential new species) for science. Among the morphological species, there were six amphipod and one mysid species, all peracarids with direct development as opposed to only one decapod species (larval development) showing cryptic speciation. This finding might be explained by a different potential for dispersal related to the developmental mode, leading to different speciation rates. However, all these species complexes have an unclear status in the present, waiting to be investigated and, hopefully validated, by taxonomists.

Although cryptic speciation might be considered infrequent in marine crustaceans from the NWA (8.7% species complexes), it is still a measure of biodiversity underestimation at the species level. Moreover, the phylogenetic analyses performed in Talitridae showed polyphyly in some genera (e.g., *Tethorchestia*, *Orchestia*) with potential taxonomic revision towards genus splitting, hence a higher richness in high-taxa (e.g., genus) revealed by DNA barcoding and that was previously overlooked.

In this study, DNA barcoding proved its usefulness beyond species identification, complementing molecular phylogenetics, phylogeography and taxonomy (Hajibabaei *et al.*, 2007).

Linking genetic and species level

The phylogeographic study on the amphi-Atlantic *G. oceanicus* (Figure 13H), one of the most common and abundant intertidal species, had the largest geographic coverage in terms of number of sites and marine ecoregions (Spalding *et al.*, 2007) sampled. A previous study (Henzler, 2006) included more sequences (326 COI) and better European coverage (but still largely incomplete for this species range). However, it had less coverage in Atlantic Canada (only 20 sites compared to our 73 sites) and, most importantly, no specimens from Arctic Canada, thus ignoring an important part of the post-glacial colonization in *G. oceanicus*. Presently, genetic data only partially cover this species' distribution range with no data from the southern European range, hence an incomplete picture of species history persists despite the large amount of sequence data currently available.

DNA barcoding followed by phylogeographic analyses revealed the existence of two divergent intraspecific *G. oceanicus* clusters (2.4% COI distance) (Figure 4.3), which did not correspond to a European – North American separation but to a partial latitudinal segregation between north and south. Two phylogeographic patterns have been revealed: i) large genetic gaps with two major lineages allopatric as in Atlantic Canada and Europe (type I in Avise *et al.* 1987; Figure 15); ii) small (or inexistent) genetic gaps with lineages allopatric as in Arctic Canada and Europe (type III, Figure 15). Such patterns could be easily explained according to the vicariance and dispersal hypotheses, respectively (Figure 16). The species distribution range became fragmented due to Pleistocene glaciations which forced populations to migrate south in refugial areas where genetic differentiation (due to genetic drift and natural selection acting on mutations) began the process of speciation. The present-

day vicariant clusters are genetically and geographically separated and if the reproductive isolation is proven, it will be a classic example of allopatric speciation. On the other hand, some of the European survivors managed to colonize mid-Atlantic islands (e.g., Iceland) and Arctic Canada in the postglacial era and probably through a rapid colonization process (seen the genetic identity of specimens from Iceland and Hudson Bay; Figure 4.4). Although a recent colonization process, the northern cluster still includes geographically isolated populations and there is no data on the extent of gene flow currently occurring between these populations living on different shores, islands and continents. Given enough time to develop reproductive barriers, the northern cluster may become a classic example of allopatric speciation through dispersal and colonization. The probable contact area between clusters, namely the Labrador coast, could not be sampled for this study. However, a secondary contact between clusters, if existent, should occur in this geographic area, therefore the Labrador coast is the missing link towards clarifying any incipient speciation (i.e., how far on the speciation path are these two clusters) in *G. oceanicus*.

As a note, classic speciation models (in particular allopatry) might not be very accurate when applied to the marine environment. Species presently distributed on both coasts of the Atlantic (amphi-Atlantic) are considered to have a disjunct distribution and genetic differentiation would occur through vicariance. However, the vicariant hypothesis includes initial contiguous distribution occupied by an ancestral population which splits into two (or more) clusters after the formation of barriers (Futuyma, 1998). In this regard, could marine distributions of coastal species (including islands and continents) ever be considered as contiguous? The North Atlantic is a "young" ocean but the present-day configuration is multi-million years old preceding the appearance of *G. oceanicus* and many other extant marine species with disjunct distributions.

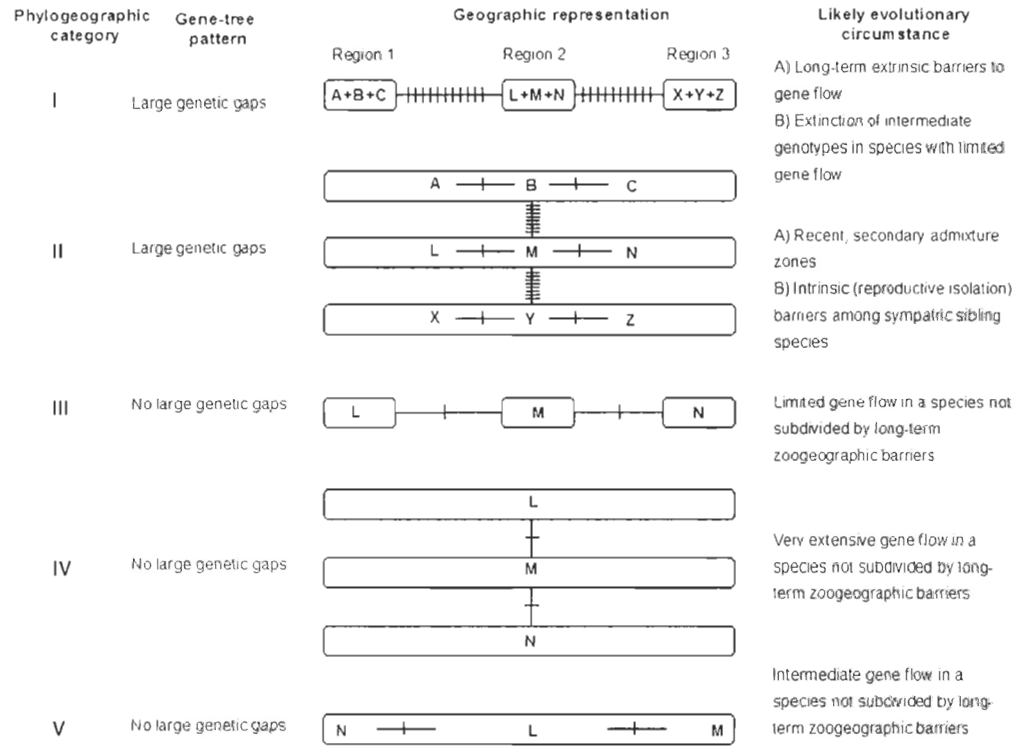


Figure 15 Phylogeographic patterns with the evolutionary circumstances involved. Rectangles represent various mtDNA haplotypes (denoted by letters) or groups of closely related haplotypes with their geographic distribution. Haplotypes are connected in networks with dashes indicating the number of mutational steps involved within specific pathways. (Modified from Avise *et al.*, 1987)

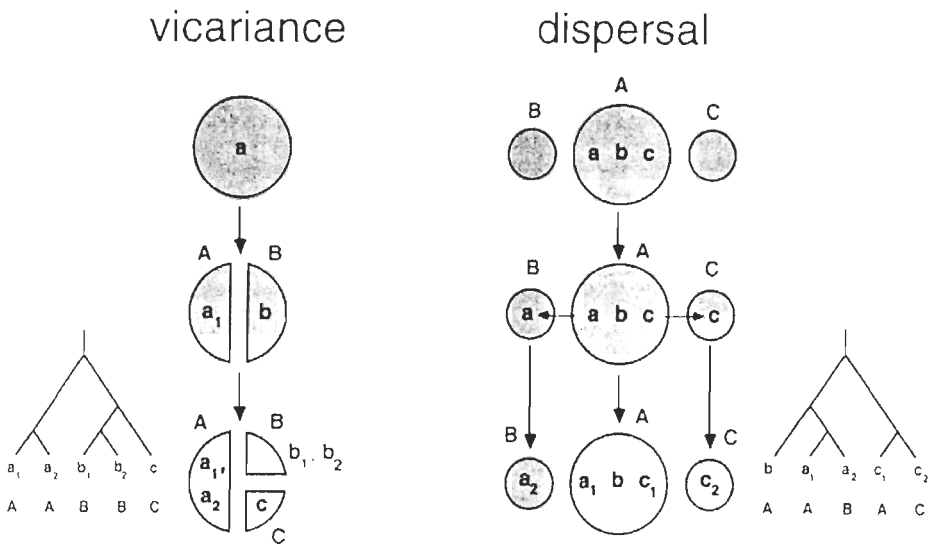


Figure 16 Phylogenetic relationships between populations or species inhabiting geographically separated areas, under vicariance and dispersal. Lowercase letters represent taxa, uppercase letters represent geographic areas. (Modified from Avise, 2000)

Limitations of DNA barcoding

Life is too complex to be easily and fully resolved by DNA barcoding. Limitations, as well as advantages, of employing DNA barcoding are mainly related to the characteristics of mtDNA (see Chapter I). Being a haploid marker, it cannot detect *hybrids* unless they are already differentiated from their parent species. The importance of this loss for biodiversity molecular assessments has to be addressed on the basis of hybridization frequency. Other difficulties are caused by: mitochondrial *introgression*, *incomplete lineage sorting*, *heteroplasmy*, *intracellular endosymbionts* (review in Frézal and Leblois, 2008), *contamination* or *taxonomic misidentifications* leading to the attachment of DNA barcodes to erroneous species (e.g., *E. ischnus* initially misidentified due to specimens in poor-shape; see Chapter II). *Thresholds* have to be considered more as rough indications rather than cut-off

values with subsequent careless splitting or lumping of morphological taxa. This is especially true for species complexes, recently diverged species, slow evolving groups (e.g., cnidarians; Hebert, Ratnasingham and deWaard, 2003) or in cases of incomplete taxon sampling (Meyer and Paulay, 2005).

Pseudogenes (or nuclear mitochondrial pseudogenes, numts) are mtDNA sequences which are duplicated during cell division and migrate into the nucleus. Although non-functional, pseudogenes can still be amplified with universal primers, thus blurring the variation patterns of orthologous mtDNA. This is a general situation occurring for all mt genes but it is of special concern for DNA barcoding due to the use of COI sequences for species identification and discovery (Song *et al.*, 2008). If undetected, pseudogenes would lead to large overestimates of diversity indices. Divergence values above 3% (threshold for species delimitation, but see above) will generate inflation of the species richness index (see *Orchestia* spp.; Chapter III) with great impact on barcoding studies (Song *et al.*, 2008; Buhay, 2009). Lower divergences (<3%) will inflate the genetic diversity indices (e.g., haplotype diversity) with great impact on phylogeographic studies (Bertheau *et al.*, 2011). A series of steps have been proposed when dealing with the nightmare of pseudogenes (Figure 17). The routine in barcoding studies includes some quality control measures: check DNA sequences for indels, STOP codons, double peaks, background noise, and length variation. Cloning is not a routine step due to the long time required for this process and cost. Additional markers can be used when detecting deep clusters. However, other mt genes might also turn out to be pseudogenes. Nuclear genes have been proposed as a second marker to give strength to the process of species discovery (Raupach *et al.*, 2010b). However, nDNA and mtDNA would probably give opposite results for recently diverged species due to their different evolutionary rates. Amplifying mtDNA from isolated mitochondria or from tissues rich in mitochondria might help in "getting" the right gene but these procedures might be expensive or time consuming. DNA barcoding involves standardization (across protocols and research groups), rapid processing and low cost. In this context,

probably the easiest way to ensure quality control and bring support for cryptic speciation would be a second round of operations starting with tissue. As the amplification of pseudogenes, usually in large copies, is a random process, a second sequencing should result in a different sequence (if pseudogenes are involved) or an identical sequence (if the orthologous gene has been amplified).

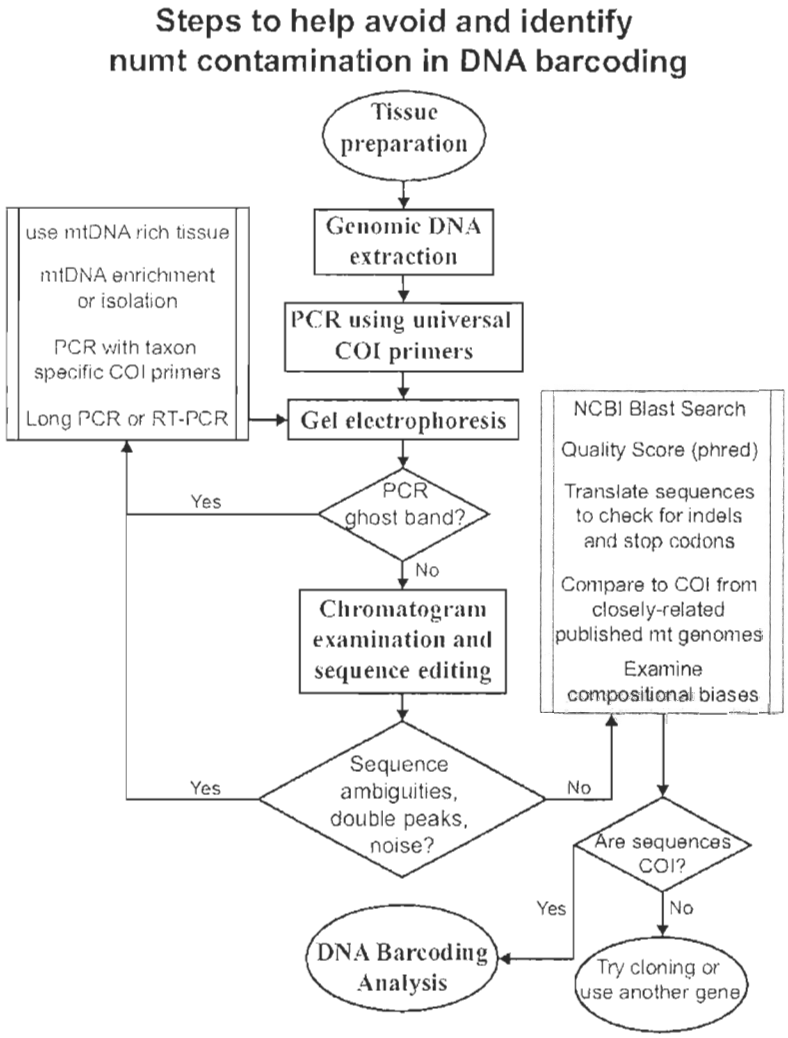


Figure 17 Measures to limit the amplification of pseudogenes in DNA barcoding studies. (Source: Song *et al.*, 2008)

Genetic studies based on mtDNA have assumed *neutrality* for this marker when inferring effective population size or demographic history. However, the mt genome has a vital role in cellular functioning by coding proteins involved in the respiratory chain (Figure 7). Mutations in the mtDNA caused by oxidative impact through ROS, for instance, would lead to a deficient functioning of the respiratory chain due to mismatches between mt and nDNA, both involved in this cellular function, therefore they would be incompatible with life (Ballard and Whitlock, 2004; Gershoni, Templeton and Mishmar, 2009). With deleterious mutations being removed through purifying selection, it follows that the genetic variation inferred through genetic analyses would mirror neutral processes (review in Galtier *et al.*, 2009). However, recent investigations have challenged the general assumption regarding neutrality of mtDNA, making this issue one of the most controversial aspects in genetic studies. According to Bazin *et al.* (2006), mtDNA exhibits low diversity values at the intraspecific level across all animal groups investigated due to recurrent selective sweeps (adaptive evolution) rather than to population size or ecology. Selective sweeps imply positive selection of certain haplotypes with subsequent drops ("sweeps") in overall mtDNA diversity due to lack of recombination of the mt genome. Such sweeps might be caused by selection of beneficial haplotypes (e.g., more efficient energetic metabolism according to temperature), selection of "selfish" mutations (e.g., higher replication rate regardless of the effect on the fitness) or by genetic hitchhiking (e.g., maternally inherited symbionts, such as *Wolbachia*, affecting the host mt genome in order to spread across host's distribution range) (Ballard and Rand, 2005; review in Galtier *et al.*, 2009).

The entire DNA barcoding approach is based on the existence of "barcoding gaps" between genetic variation within and among species. Low intraspecific diversity observed in some groups could be a result of recurrent selective sweeps or an artifact of small sample size used in barcoding studies. However, recent analyses of large COI datasets in birds have found no evidence for positive selection (selective sweeps) (Kerr, 2011). DNA barcodes did not fit the neutrality predictions

either and the implication was that the evolution of COI is largely driven by purifying selection. As large barcoding datasets are currently being built at the global scale, it will soon be possible to test neutrality predictions across taxa and the level of correlation between intra- and interspecific levels of diversity.

Only a few of the above-mentioned limitations were encountered during this study and the concerned data were not included in further analyses. Cases of cross-contamination between taxa during laboratory operations and of pseudogene amplification (Chapters II-III) did occur but with low frequency, therefore pseudogenes do not seem to affect the success of DNA barcoding in marine crustaceans. A bigger limitation for this project was the low sequencing success (65%) with "universal" primers, hence the need for better barcoding protocols (i.e., primer design). Above all, the lack of taxonomists to validate the species complexes detected in this study (and most barcoding studies) is probably the most stringent problem currently encountered in the barcoding world.

The species... issue

Species are considered the unit of biodiversity and, yet, there is no definition for this fundamental component of the living world. The species concept is probably the most controversial issue in biology, partially due to a semantic shift by which methods for species identification were raised to the rank of "concept" (Hey, 2006). Consequently, more than 25 concepts were proposed (Coyne and Orr, 2004) with the biological species concept (BSC) being the most accepted one (Table 2).

Table 2 Various species concepts (SC) (Modified from Futuyma, 1998). The closest concept to DNA barcoding, the Phylogenetic SC, and the most popular concept, Biological SC, are highlighted

BIOLOGICAL SC	A species is a group of individuals fully fertile *inter se*, but barred from interbreeding with other similar groups by its physiological properties (producing either incompatibility of parents, or sterility of the hybrids, or both). Species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups.
PHYLOGENETIC SC	A phylogenetic species is an irreducible (basal) cluster of organisms that is diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent. A species is the smallest monophyletic group of common ancestry.
EVOLUTIONARY SC	A species is a single lineage (an ancestral-descendant sequence) of populations or organisms that maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate.
RECOGNITION SC	A species is the most inclusive population of individual biparental organisms that share a common fertilization system.
COHESION SC	A species is the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms.
ECOLOGICAL SC	A species is a lineage (or a closely related set of lineages) that occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range.
INTERNODAL SC	Individual organisms are conspecific by virtue of their common membership in a part of the genealogical network between two permanent splitting events or between a permanent split and an extinction event.

What is a species and why is it important? Species are a virtual tool that we need in order to organize the diversity of life into categories that our mind can understand. Organizing diversity started with classifying organisms into "species" based on their phenotype and on reproductive compatibility. With the advent of

molecular methods, came the importance of the genotype for species delineation as well. Linnaeus started to organize life in a systematic way by giving unique binomial names to species and this approach has been in place for the past 250 years, leading to the description of a fraction of global biodiversity. DNA barcoding is an additional tool to classify life but based on DNA sequences rather than morphological characters. It provides a molecular tag linked to existing binomial species names. However, it also uncovers new categories (i.e., cryptic species), unknown to the scientific community, therefore challenging traditional views on diversity. The role of DNA barcoding in species discovery as opposed to species identification (much easier to agree upon) is still a debated issue (Rubinoff, 2006; Ebach and de Carvalho, 2010). Indeed, species validation should not rely solely on one marker and on small sample sizes (although many morphological species are known from single specimens and/or localities; Stork, 1997).

DNA barcoding does not validate species *per se* but detects interesting cases for further investigation. The barcode clusters identified, and which are usually reciprocally monophyletic, would correspond to potential cryptic species according to PSC (Table 2). In some cases, these divergent clusters correspond to reproductively isolated groups, generating a close link with BSC (Gómez *et al.*, 2007). However, most specimens used for DNA barcoding are not kept alive and they cannot be used for mating trials. The current situation includes a large body of genetic information (>1.3 million barcodes in BOLD, August 2011), with many of these sequences lacking scientific names (due to the taxonomic impediment). The molecular work is rapidly advancing with the obvious result of generating millions of barcode clusters with no validation either way (nominal species approved or discarded). One solution for assigning names to sequences consists of sequencing already identified museum material (e.g., type specimens). Unfortunately, protocols are available mostly for "dry" material (e.g., insects, birds, mammals, plants) and les so for "wet" collections, such as crustaceans, which included formalin during preservation. Another solution might consider a shift in our view on biodiversity by using number-tags (i.e., barcode

clusters identified by unique numbers) rather than names (at least temporarily). In this way, barcode clusters could be considered as the functional units of biodiversity. Although challenging for our mind, number-tags are as good as names in some cases (e.g., finding streets in a city) and, in the biodiversity world, they might act as proxies for estimating diversity.

No matter what will be the future of biodiversity classification, it is important to keep in mind that "species" are dynamic rather than amorphous things in named boxes. The elusive "species issue" is not a problem or a failure (Hey, 2006) but an interesting puzzle. By focusing too much on defining an indefinable concept, we cannot see the forest for the trees ("it is somewhat depressing that evolutionary biologists continue to spend so much time arguing about what constitutes a species when the debate cannot be resolved by normal scientific methods"; Coyne and Orr, 2004). Molecular methods in general, and DNA barcoding in particular, have challenged the practicality of using species in biodiversity inventories or environmental monitoring due to the amount of cryptic species being detected. The end-users of species lists (conservation biologists, macroecologists) have difficulties due to taxonomic inflation. However, they have to acknowledge that life is more complex than clean species lists and advance their investigation methods accordingly (e.g., phylogenetic diversity; Faith, 1994) .

Below-species investigations: limitations of COI

Mitochondrial DNA, and especially the COI gene, has been widely employed as a useful marker for studies at the intraspecific level (Avise *et al.*, 1987; Avise, 2000). COI has the power to identify phylogeographic clusters, therefore there is a bonus for using this gene in barcoding studies: while building reference libraries for species identification, the same data can be used in phylogeographic studies (provided an appropriate sample size). Chapter IV used that bonus and the analyses

revealed deep phylogeographic patterns with allopatric divergent clusters. However, a more detailed analysis of the genetic structure at smaller spatial scales was not well-supported. The baseline in any analysis (i.e., population assignment) could not be accurately identified, hence the use of a proxy for population subdivision. Sampling sites were grouped into "populations" according to biogeographical zones of GSL (Brunel, Bossé and Lamarche, 1998). This is definitely an artificial measure with no biological support. Unfortunately, every method employed for finding genetically and geographically cohesive groups (SAMOVA, DAPC, BAPS), has failed. Two remarks can be made about this outcome. First, the short COI barcodes do not have enough resolution for population structure investigations. While it is a good marker at the species level and coarse intraspecific level (i.e., large scale patterns) due to its slow mutation rate, it cannot offer good resolution at fine scale, where highly variable markers (e.g., microsatellites) should be used instead. Second, the level of population connectivity in *G. oceanicus*, as in many other marine invertebrates, is largely unknown. Direct estimation is almost impossible to gain unless micro-tagging devices are developed. Indirect estimation can be based on genetic data (e.g., hypervariable markers). However, additional methods such as modeling analyses based on biological and environmental factors should be included in order to validate patterns inferred from genetic data. It is intriguing that one of the basic aspects in biology, namely population size, is largely ignored when it comes to marine species (even the common intertidal ones). But then, "population", just as "species", is still a puzzle for scientists (Table 3).

Table 3 Definitions of "population" according to various criteria (Modified from Waples and Gaggiotti, 2006)

Ecological paradigm	A group of organisms of the same species occupying a particular space at a particular time
	A group of individuals of the same species that live together in an area of sufficient size that all requirements for reproduction, survival and migration can be met
	A group of organisms occupying a specific geographical area
	A set of individuals that live in the same habitat patch and therefore interact with each other
	A group of individuals sufficiently isolated that immigration does not substantially affect the population dynamics or extinction risk over a 100-year time frame
Evolutionary paradigm	A community of individuals of a sexually reproducing species within which matings take place
	A major part of the environment in which selection takes place
	A group of interbreeding individuals that exist together in time and space
	A group of conspecific organisms that occupy a more or less well-defined geographical region and exhibit reproductive continuity from generation to generation
	A group of individuals of the same species living close enough together that any member of the group can potentially mate with any other member
Statistical paradigm	An aggregate about which we want to draw inference by sampling
	The totality of individual observations about which inferences are to be made, existing within a specified sampling area limited in space and time
Variations	*Stock*: a species, group, or population (of fish) that maintains and sustains itself over time in a definable area
	Demes: separate evolutionary units
	Natural population: bounded by natural ecological or genetic barriers
	Local population: (i) individuals have a chance to interact ecologically and reproductively with other members of the group, and (ii) some members are likely to migrate between local groups

In a nutshell: findings and extrapolations

This study built a reference library for 92 morphological species from five orders of malacostracan crustaceans from the North Atlantic and revealed a few interesting results (Figure 18). One invasive species was detected in ESL (*E. ischnus*) and its impact on the estuarine food web is unknown.

At the species level, DNA barcoding identified a total of eight species complexes consisting of 23 clusters. If validated by taxonomists, these barcode clusters would lead to the description of 15 new species. Three species complexes harboring 12 clusters (nine clusters unknown to science) were identified within Talitridae, the only amphipod family to have colonized the land. Although taxonomic revision is needed for species validation, these results clearly indicate an underestimation of crustacean diversity in the North Atlantic.

Above the species level, DNA barcoding revealed polyphyly for two genera indicating the need for taxonomic revision. These two genera will probably be split resulting in an overall increase for this limited dataset (seven genera) with two new genera (Wildish and LeCroy, in prep.). Therefore employing a taxonomic distinctness index at the genus level based on the current classification would lead to slightly erroneous results, underestimating diversity and distinctness. The phylogenetic analysis also showed that all three ecological-systematic groups used to classify talitrids (sand-burrowers, wrack generalists and palustral hoppers) are polyphyletic and a large revision at the family level should be conducted.

Below the species level, DNA barcoding and phylogeographic analyses showed a certain level of genetic structure in *G. oceanicus* in Atlantic Canada culminating with a phylogeographic pattern type I (Figure 15). Two clusters separated genetically by 2.4% COI distance and geographically by thousands of km

might be indicative of cryptic speciation. In terms of genetic variation, crustaceans seem to be more diverse at the intraspecific level (0.5% mean variation when removing species complexes) compared to other groups (Chapters I-II; Costa *et al.*, 2007) reflecting the age of the group and/or faster evolutionary rate compared to vertebrates and other invertebrate groups.

Among the two main groups barcoded here, there were seven species complexes in peracarids and only one in eucarids. This finding might be explained by a different potential for dispersal in peracarids (direct developers) compared to eucarids (larval development), leading to different speciation rates (Figure 4). However, extensive studies have to be made to test the dispersal-cryptic speciation hypothesis and the geographical pattern involved (allopatric/sympatric).

A number of 8.7% cases of cryptic speciation in this study does not allow for extrapolations on the frequency of cryptic species in crustaceans. What might seem a low value overall, might be influenced by various factors, biological (dispersal potential) and human (taxonomic accuracy for various groups). In addition, extrapolation regarding threshold for crustaceans should be considered very carefully. An interspecific value of 2.8% between two morphological species of crabs (*Hyas* spp.) does not generalize this value as a universal threshold for crustaceans. All crustaceans are not evolutionary equal and some groups might accumulate mutations faster than other groups (hence the problematic use of "universal" molecular clocks).

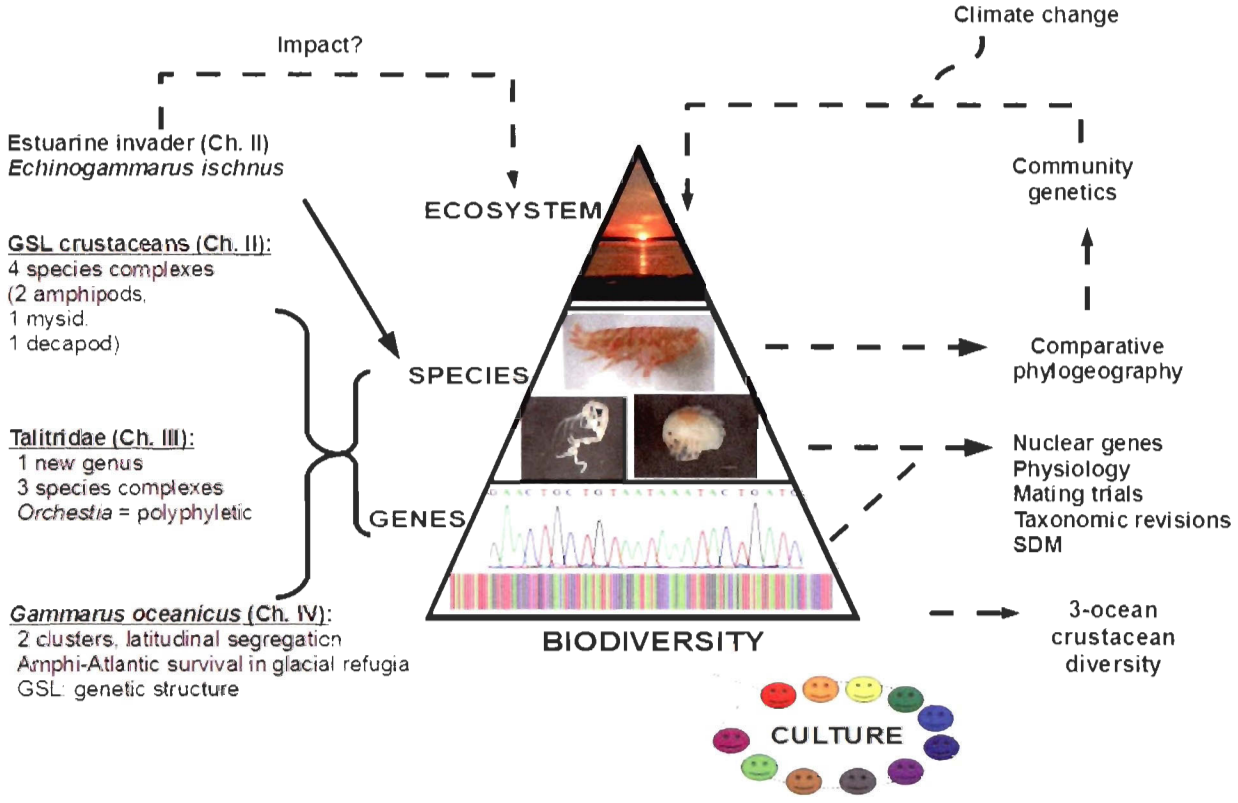


Figure 18 Schematic representation of the main findings of this project. The pyramid represents biodiversity with its main levels, including cultural diversity. Results for each chapter are mentioned on the left side, while future directions are on the right side of the pyramid. Arrows connect the results with the biodiversity levels tackled (full lines for results, dashed lines for future investigations).

Future directions

COI sequences analyzed for this thesis (N=907) represent about half of all sequences (N>2,000) generated during the barcoding project of marine crustaceans from NWA. In turn, this total amount represents ~65% of all specimens tested (successful DNA extraction, failed COI amplification or sequencing). Such a medium success is not a failure of DNA barcoding, just an indirect indication of the complexity of Malacostraca. Within a highly diversified class with various groups (e.g., amphipods, isopods, decapods etc.) probably having different mutation rates, the existing "universal" primers fail to amplify COI across all taxa. Consequently, a lot of effort has to be put in developing new primers, probably at the family level (especially in amphipods).

Extending research from this study

There are many loose ends after the completion of this study. DNA barcoding has the role to screen large sample sizes and identify cases of discordance between morphology and genetics. However, it cannot bring answers to all questions, thus the multitude of directions to be taken further (Figure 18).

One invasive species has been identified in ESL (Chapter II) but there is no additional information in this case. How abundant is this species in the estuary? What role does it play in the food web? What impact does it have on local estuarine fauna? What measures (if any) should be taken?

Eight species complexes have been detected (Chapter II-IV) but there is a stringent need to continue investigations on these groups. Are all the 23 clusters valid species? Are there some pseudogenes that might blur our inference on cryptic

speciation? Are these putative species different at the morphological, ecological, behavioral or physiological level? If they are valid species, will they ever be described according to the current standards? The barcoding analyses were based solely on genetic distances. Therefore other methods (character-based or coalescent inference) should be taken into account for comparison (Pons *et al.*, 2006; Rach *et al.*, 2008).

Phylogenetic and phylogeographic analyses (Chapters III-IV) should be based on multiple markers, therefore there is a need to include nuclear genes into these analyses. The field of phylogeography is rapidly evolving in terms of analyzing and visualizing data. The methods applied here belong largely to descriptive phylogeography, in which genetic patterns are believed to be the result of biogeographic processes (vicariance and dispersal) (Avise *et al.*, 1987; Avise, 2000). The next step, statistical phylogeography, will include coalescent models to estimate parameters and testing of phylogeographic hypotheses (Hickerson *et al.*, 2010 and references therein). For *G. oceanicus*, the missing link between the two clusters is the remote coast of Labrador. Therefore, sampling along this coast (as well as in Europe) is vital in order to clarify the distribution range of the two clusters and the possibility of hybridization or sympatric speciation due to local adaptation to food source. Mating trials between specimens from the two clusters and the application of fast evolving genetic markers (e.g., microsatellites) would bring more clarity on the issue of cryptic speciation in *G. oceanicus*. For talitirids, SEM investigations might reveal fine morphological characters to distinguish between cryptic species in *Platorchestia* and *Orchestia*. A new genus and two new species (former *Tethorchestia* sp. B) are currently being described, their discovery being driven by the findings of DNA barcoding (Wildish and LeCroy, in prep.). Multiple nuclear genes, a matrix of morphological characters and many additional taxa should be included in the phylogenetic analysis of Talitridae. Sampling entire distribution ranges of species is highly desirable to investigate the geography of genetic variation.

Genetic diversity across marine regions

Only half of the sequences generated while barcoding Canadian crustaceans have been included in this study. The complete database (>2,000 DNA sequences) spans hundreds of species across the five malacostracan orders and across Canada's three oceans. A large-scale analysis is currently envisaged for testing the hypothesis of speciation rates being correlated with dispersal capacities (developmental mode) (Figure 4). Another goal is finding spatial scales for diversity patterns (e.g., oceanographic areas more genetically diverse than other areas). This type of analysis has been conducted in polychaetes from Canadian oceans (Carr *et al.*, 2011), therefore an interesting comparison between two invertebrate groups (crustaceans and polychaetes) could be done.

Complete crustacean (Malacostraca) inventory for NW Atlantic

Only the most common species have been included in this study. Full taxonomic coverage of the known crustacean species from Atlantic Canada is hampered by sampling difficulties. Indeed, except for decapods of economic importance, other malacostracan species are not targeted by regular sampling surveys and seldom show up as by-catch. Moreover, for some taxa (e.g., amphipods), the use of dip nets, baited traps or bottom trawls will lead to a sampling bias towards highly mobile species. There are two possibilities to create a comprehensive database for crustaceans in the future: research cruises targeting rarer crustaceans or technological advances for high-throughput DNA extraction from formalin-preserved crustaceans, neither of them very probable to occur in the near future. At the global level, many small-scale studies are targeting crustaceans. By combining these datasets, a global database of crustacean barcodes will eventually emerge. New directions in biodiversity studies involving barcodes, such as environmental barcoding (a special Working Group in iBOL) and quantifying food webs (Smith *et al.*, 2011), require reference libraries of high quality (i.e., validated by

taxonomists), thus the need for a close collaboration between barcoders and taxonomists.

Genomics

The rapid development of next-generation sequencing technologies will generate an enormous amount of DNA sequences and even entire genomes, processed fast and cheap. These data will allow for better resolution in species delimitation (congruence of nuclear and mitochondrial markers), population assignment (e.g., by using microsatellites) and inferring genetic patterns at various spatial scales (e.g., microsatellites, mitochondrial and nuclear genes). Moreover, data will be helpful in identifying genes with potential role in speciation (Miglietta, Faucci and Santini, 2011), allowing us to understand the mechanisms driving the formation and the extinction of species as part of global biodiversity.

Comparative phylogeography

Future studies on comparative phylogeography of co-distributed taxa will have important implications due to their strength in inferring patterns (i.e., repeated patterns, in many taxa, provide support for historical hypotheses). Such studies will identify geographical areas where communities exhibit unique evolutionary histories. These areas should be prioritized in conservation plans, thus ensuring not only the preservation of present-day diversity but also of the processes generating this diversity (Moritz and Faith, 1998). Understanding the past (i.e., evolutionary history) will help scientists predict the future. There is a need to make good predictions about the impact of climate change on biodiversity at various spatial scales and levels. Comparative phylogeography will certainly be a part of the research fields involved in modeling the response of communities to a changing environment although more powerful coalescent model-based methods have to be created (Hickerson *et al.*,

2010). In addition to phylogeography, large datasets of DNA sequences might be used in landscape genetics (concerned with a smaller temporal scale than phylogeography), although some debate still exists about choosing appropriate markers for various temporal scales (Bohonak and Vandergast, 2011; Wang, 2011).

Interdisciplinarity

DNA barcodes (and genetic data in general) provide useful but limited information. For a complete picture on any given aspect, there is a need to work across disciplines. In the case of marine crustaceans from NWA, some interdisciplinary links have been mentioned above. Another important direction is to link genetic biodiversity with ecosystem functioning by investigating the functional role of cryptic species. Spatial distribution of (cryptic) species can be tackled with ecological niche modeling (ENM; synonym with spatial distribution modeling, SDM). This type of analysis can bring support (or not) to genetic studies on past distributions (e.g., survival or extinction due to glacial cycles) and can be used to predict future range shifts due to climate change, for instance.

This study had a four-fold focus: i) biodiversity (two levels); ii) molecular methods; iii) marine crustaceans; and iv) North Atlantic. However, all four keywords were only partially addressed and by no means will this thesis shed light on marine biodiversity (except that it is underestimated). Specifics of this project: i) species level – only species identification (no measure of species diversity etc.); genetic level – mainly phylogeographic patterns investigated; ii) only one molecular method used (DNA barcoding) with only one marker (COI); iii) only selected species of Malacostraca targeted, mainly shallow-water and benthic taxa; and iv) only coastal areas of North Atlantic (and Arctic Canada in Chapter IV) were sampled by opportunistic methods. When it comes to marine biodiversity and speciation, most studies are conducted in coastal areas and on relatively well-known taxa, a lot fewer studies occur in deep-sea or open-water and in poorly known groups such as algae,

meiofauna, microbes (Miglietta, Faucci and Santini, 2011) or parasites. Therefore the gaps in our knowledge are very large and we will probably never fill them completely but just start the process and try to work not only on "How many species are out there?" but also on "What do we know about the species that already have names? What role do they play in marine ecosystems? How will communities evolve in the context of global change?"

Overall advantages of DNA barcoding

DNA barcoding arises as an exceptional tool and some of its advantages have been mentioned throughout this thesis. It is a tool developed for *species identification and discovery* but with implications at the genetic (detect phylogeographic clusters) and ecosystem levels (identify the make-up of functional groups). It is a revolutionary method by which anything carrying DNA could be easily identified, hence many practical applications (food traceability, specimen trading, detection of disease vectors, pests, invasive species, etc.). One of the main advantages of DNA barcoding is the capacity to perform large-scale screenings of diversity and pick up those cases (e.g., morphological species in disagreement with barcode clusters) in need for detailed investigation at the morphological, ecological, physiological or behavioral levels. All data related to DNA barcoding are maintained in curated databases online while vouchers are stored in public institutions for future reference. By using a standard gene fragment, comparisons across taxa and geographic regions are greatly improved.

Another great advantage of DNA barcodes regards their function as permanent species tags, a crucial role in a world where up to 20% of species names might be synonyms due to multiple descriptions of the same species (Stork, 1997) and where species are continuously being split or lumped following taxonomic revisions. DNA barcoding provides a temporal and spatial snapshot on genetic

diversity: at time t, there were x barcode clusters from y morphological species in a given area. As the barcoding initiative is constantly growing, many new distribution records for various species will be added to the global database on biodiversity. Moreover, the barcode database will act as a baseline for species and genetic diversity estimates to be compared with future shifts caused by global change. The large amount of data generated by DNA barcoding will act as exploratory research and will likely serve to formulate new hypotheses about genetic diversity in space and across taxa (e.g., molecular evolution of various groups) subsequently tested as part of confirmatory research (Jaeger and Halliday, 1998).

Only history will tell if DNA barcoding succeeds in advancing and improving research on biodiversity, in fostering close collaborations between barcoders and taxonomists and in changing mentalities about sharing scientific results. At the moment, DNA barcoding (through iBOL) is the largest biodiversity genomics project ("natural history re-loaded") and the largest experiment of open-access data sharing, involving non-scientists in creating a bio-literate world.

Preserving global biodiversity

"In the end, we will conserve only what we love, we will love only what we understand and we will understand only what we are taught." (Baba Dioum, Senegalese environmentalist).

Humans are an intrinsic part of the living world and our existence is dependent on natural resources (Figure 1), thus the need to preserve biodiversity. Failure to wisely manage natural resources might lead to the collapse of human societies (e.g., the Rapa Nui culture on Easter Island that might have declined as a consequence of deforestation; Diamond, 2005). While it is easy to agree on the importance of biodiversity, it is more difficult to find a consensus for taking action and finding

precise conservation measures, in the context of one species having an immense impact on the remaining many millions of species as well as on the general climate. Finding sustainable solutions was beyond the scope of this thesis. The overall goal here was to use molecular methods (specifically DNA barcoding) to tackle two biodiversity levels. By no means did the results fill gaps in our knowledge on marine life, rather it added a few drops of information (and information does not equal knowledge; Boero, 2010). Further studies will go into more details regarding the crustacean diversity in the North Atlantic. However, the most important question rising from this and all the other studies on biodiversity is: "So what? Would knowing all the species living out there and their genetic make-up solve the biodiversity crisis?" I argue it is a timid but vital step in solving a complex situation as we have to start by knowing what lives where. However, scientific information has to be put into practice and the first step consists of a closer dialogue between scientists and the rest of the world. A big step forward in sustaining life on Earth will be made by including humans and cultural diversity within global biodiversity. Saving endangered cultures *in situ* (not by translocation in reserves) will imply saving local biodiversity and knowledge, as most of these indigenous cultures are usually linked to hotspots of biodiversity. Most importantly is to keep in mind that species and cultures are continuously evolving both in their native habitat and beyond, and to design conservation plans accordingly.

"- What's the use of their having names, the Gnat said, if they won't answer to them?
- No use to them, said Alice, but it's useful to the people who name them, I suppose. If not, why do things have names at all?"

Lewis Carroll, *Through the Looking Glass*

APPENDIX A

SPECIES LIST WITH DETAILS ABOUT SAMPLE SIZE (N), MEAN AND MAXIMUM INTRASPECIFIC DIVERGENCE (%), AND NEAREST NEIGHBOR DISTANCE (NN %)

Order	Family	Species	N	Mean %	Max %	NN %
Amphipoda	Ampeliscidae	*Ampelisca eschrictii*	12	5,5	13,6	27,25
	Ampithoidae	*Ampithoe longimana*	3	0,2	0,3	21,34
		Ampithoe rubricata	5	0,06	0,15	21,34
	Calliopiidae	*Calliopius laeviusculus*	29	0,89	2,32	26,58
		Halirages fulvocinctus	2	0	0	24,29
	Caprellidae	*Caprella linearis*	8	0,57	1,08	21,26
		Caprella mutica	4	0,38	0,61	21,9
		Caprella septentrionalis	12	0,34	0,92	24,42
	Epimeriidae	*Paramphithoe hystrix*	1	N/A	N/A	24,29
	Eusiridae	*Eusirus cuspidatus*	2	0	0	22,7
		Pontogeneia inermis	3	0,2	0,3	25,43
		Rhachotropis aculeata	9	0,95	2,01	23,92
	Gammarellidae	*Gammarellus angulosus*	1	N/A	N/A	19,48
	Gammaridae	*Echinogammarus ischnus*	4	0	0	24,8
		Gammarus duebeni	6	0	0	23,34
		Gammarus lawrencianus	21	0,5	1,54	24,37
		Gammarus mucronatus	9	0,47	0,92	26,03
		Gammarus oceanicus	3	0,72	1,07	19,48
		Gammarus tigrinus	13	1,33	2,48	23,67
	Hyalidae	*Hyale prevostii*	1	N/A	N/A	21,89
	Hyperiidae	*Themisto libellula*	5	0,06	0,15	31,09
	Ischyroceridae	*Ischyrocerus anguipes*	21	2,23	4,24	21,26
	Lysianassidae	*Orchomenella minuta*	2	0,15	0,15	10,71
		Orchomenella pinguis	11	0,05	0,15	10,71
		Psammonyx nobilis	4	0,08	0,15	18,25

		Psammonyx terranovae	8	0,08	0,3	18,25
	Melitidae	*Melita dentata*	2	0	0	30,7
		Melita formosa	5	0,06	0,15	30,7
	Oedicerotidae	*Monoculodes intermedius*	2	0	0	22,64
		Oediceros saginatus	5	0	0	22,64
	Pleustidae	*Pleustes panoplus*	3	0,92	1,38	27,51
	Pontoporeiidae	*Monoporeia sp.*	2	0	0	25,84
	Talitridae	*Americorchestia longicornis*	10	0,03	0,15	19,7
		Americorchestia megalophthalma	4	0	0	19,7
		Platorchestia platensis	3	0,3	0,46	22,24
	Uristidae	*Anonyx makarovi*	10	0,91	1,86	12,43
		Anonyx sarsi	13	0,28	0,92	12,43
		Onisimus litoralis	4	0,1	0,15	25,24
Decapoda	Cancridae	*Cancer irroratus*	8	0,02	0,31	21,27
	Crangonidae	*Argis dentata*	5	0,74	1,85	17,86
		Crangon septemspinosa	13	0,54	0,92	20,46
		Pontophilus norvegicus	2	0,31	0,31	23,09
		Sabinea sarsii	2	0,15	0,15	5,75
	Crangonidae	*Sabinea septemcarinata*	5	0,31	0,61	5,75
		Sclerocrangon boreas	7	0,16	0,3	17,86
	Galatheidae	*Munidopsis curvirostra*	3	0	0	21,86
	Hippolytidae	*Eualus fabricii*	2	0,15	0,15	16,78
		Eualus gaimardii	1	N/A	N/A	17,32
		Eualus macilentus	6	0,1	0,3	20,84
		Lebbeus groenlandicus	1	N/A	N/A	16,78
		Lebbeus polaris	2	0	0	17,47

		Spirontocaris lilljeborgii	2	0	0	5,4
		Spirontocaris spinus	7	3,92	6,91	5,4
	Lithodidae	*Lithodes maja*	7	0,44	0,92	14,57
	Nephropidae	*Homarus americanus*	5	0,21	0,46	20,53
	Oregoniidae	*Chionoecetes opilio*	6	0,13	0,3	11,96
		Hyas araneus	8	0,44	0,77	2,81
		Hyas coarctatus	7	0,26	0,46	2,81
	Paguridae	*Pagurus acadianus*	4	0,23	0,46	10,68
		Pagurus arcuatus	3	1,02	1,23	12,46
		Pagurus longicarpus	8	0,62	1,08	17,61
		Pagurus pubescens	3	0,61	0,76	10,68
	Palaemonidae	*Palaemonetes vulgaris*	6	0,14	0,3	25,16
	Pandalidae	*Pandalus borealis*	4	0,44	0,93	12,08
		Pandalus montagui	12	0,29	1,54	12,08
	Panopeidae	*Dyspanopeus sayi*	3	0	0	18,67
	Pasiphaeidae	*Pasiphaea multidentata*	3	0	0	23,93
	Portunidae	*Carcinus maenas*	6	0,31	0,92	21,29
Euphausiacea	Euphausiidae	*Meganyctiphanes norvegica*	10	0,37	0,92	15,72
		Thysanoessa raschii	2	0	0	15,72
Isopoda	Aegidae	*Aega psora*	1	N/A	N/A	30,69
		Syscenus infelix	5	0,61	1,23	34,23
	Idoteidae	*Edotia triloba*	3	0,2	0,31	32,63
		Idotea balthica	1	N/A	N/A	25,2
	Janiridae	*Jaera albifrons*	1	N/A	N/A	27,46
Mysida	Mysidae	*Boreomysis arctica*	5	0,24	0,46	21,82
		Mysis gaspensis	4	0,23	0,3	15,29

	Mysis mixta	2	0	0	15,29
	Mysis stenolepis	6	0,1	0,3	15,42
	Neomysis americana	7	1,85	3,78	24,86

APPENDIX B

LIST OF HAPLOTYPES AND THEIR FREQUENCY IN EACH POPULATION. THE TWO CENTRAL HAPLOYTPES (H10, H19) ARE IN BOLD LETTERS

Population	Code	N	Haplotype number	Haplotype frequency
Norway	NOR	6	H1	4
			H2	1
			H3	1
Poland	POL	2	H4	1
			H5	1
Iceland	ICE	7	H6	7
Churchill	CHU	28	H6	27
			H7	1
Northern Quebec	NQC	2	H6	2
Estuary	EST	32	H8	6
			H9	10
			H10	7
			H11	1
			H12	1
			H13	1
			H14	5
			H15	1
Gaspe Peninsula	GAP	16	H9	3
			H10	2
			H16	1
			H17	5
			H18	1
			H19	1
			H20	2
			H74	1
Southern Gulf	SGF	27	**H19**	8
			H21	1
			H22	2
			H23	1
			H24	5
			H25	6
			H26	1
			H27	1
			H28	1
			H29	1
Prince Edward Island	PEI	21	**H19**	1
			H25	6
			H30	1
			H31	8

Prince Edward Island	PEI		H32	1
			H33	1
			H34	1
			H35	1
			H36	1
Magdalen Islands	MIS	37	**H19**	3
			H37	32
			H38	1
			H39	1
North Shore	NSH	11	**H10**	3
			H40	6
			H41	1
			H42	1
Western Newfoundland	WNF	8	**H10**	1
			H19	1
			H43	3
			H44	1
			H45	1
			H46	1
Eastern Newfoundland	ENF	29	**H10**	8
			H40	1
			H47	1
			H48	5
			H49	1
			H50	1
			H51	3
			H52	1
			H53	1
			H54	5
			H55	1
			H56	1
Nova Scotia	NSC	14	**H10**	6
			H19	1
			H27	1
			H57	2
			H58	1
			H59	1
			H60	1
			H61	1
Fundy Bay	FBY	33	H27	1
			H62	3
			H63	3
			H64	1
			H65	1
			H66	4

Fundy Bay	FBY	H67	1
		H68	2
		H69	1
		H70	7
		H71	1
		H72	1
		H73	7

REFERENCES

Addison, J. A., and M. W. Hart. 2004. «Analysis of population genetic structure of the green sea urchin (*Strongylocentrotus droebachiensis*) using microsatellites». *Marine Biology*, vol. 144, no 2, p. 243-251.

Archambault, P., P. V. R. Snelgrove, J. A. D. Fisher, J. M. Gagnon, D. J. Garbary, M. Harvey, E. L. Kenchington, V. Lesage, M. Levesque, C. Lovejoy, D. L. Mackas, C. W. McKindsey, J. R. Nelson, P. Pepin, L. Piche and M. Poulin. 2010. «From Sea to Sea: Canada's Three Oceans of Biodiversity». *PLoS ONE*, vol. 5, no 8, e12182.

Audzijonyte, A., and R. Vainola. 2006. «Phylogeographic analyses of a circumarctic coastal and a boreal lacustrine mysid crustacean, and evidence of fast postglacial mtDNA rates». *Molecular Ecology*, vol. 15, no 11, p. 3287-3301.

Avise, J. C. 1975. «Systematic value of electrophoretic data». *Systematic Zoology*, vol. 23, p. 465-481.

--------. 2000. *Phylogeography: the history and formation of species.* Cambridge, MA: Harvard University Press.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, J. E. Neigel, C. A. Reeb and N. C. Saunders. 1987. «Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics». *Annual Review of Ecology and Systematics*, vol. 18, p. 489-522.

Ballard, J. W. O., and D. M. Rand. 2005. «The population biology of mitochondrial DNA and its phylogenetic implications». *Annual Review of Ecology Evolution and Systematics*, vol. 36, p. 621-642.

Ballard, J. W. O., and M. C. Whitlock. 2004. «The incomplete natural history of mitochondria». *Molecular Ecology*, vol. 13, no 4, p. 729-744.

Bandelt, H. J., P. Forster and A. Rohl. 1999. «Median-joining networks for inferring intraspecific phylogenies». *Molecular Biology and Evolution*, vol. 16, no 1, p. 37-48.

Barber, P., and S. L. Boyce. 2006. «Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae». *Proceedings of the Royal Society B – Biological Sciences*, vol. 273, no 1597, p. 2053-2061.

Barbuto, M., A. Galimberti, E. Ferri, M. Labra, R. Malandra, P. Galli and M. Casiraghi. 2010. «DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (*Mustelus* spp.)». *Food Research International*, vol. 43, no 1, p. 376-381.

Bartlett, S. E., and W. S. Davidson. 1991. «Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes». *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 48, no 2, p. 309-317.

Bayon, R., and M. Jenkins. 2010. «The business of biodiversity». *Nature*, vol. 466, no 7303, p. 184-185.

Bazin, E., S. Glémin and N. Galtier. 2006. «Population size does not influence mitochondrial genetic diversity in animals». *Science*, vol. 312, no 5773, p. 570-572.

Bertheau, C., H. Schuler, S. Krumbock, W. Arthofer and C. Stauffer. 2011. «Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs». *Molecular Ecology Resources*, vol. 11, no 6, p. 1056-1059.

Bhadury, P., M. C. Austen, D. T. Bilton, P. J. D. Lambshead, A. D. Rogers and G. R. Smerdon. 2006. «Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes». *Marine Ecology – Progress Series*, vol. 320, p. 1-9.

Blaxter, M. L. 2004. «The promise of a DNA taxonomy». *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, vol. 359, no 1444, p. 669-679.

Block, B. A., I. D. Jonsen, S. J. Jorgensen, A. J. Winship, S. A. Shaffer, S. J. Bograd, E. L. Hazen, D. G. Foley, G. A. Breed, A. L. Harrison, J. E. Ganong, A. Swithenbank, M. Castleton, H. Dewar, B. R. Mate, G. L. Shillinger, K. M. Schaefer, S. R. Benson, M. J. Weise, R. W. Henry and D. P. Costa. 2011. «Tracking apex marine predator movements in a dynamic ocean». *Nature*, vol. 475, no 7354, p. 86-90.

Boero, F. 2009. «Recent innovations in marine biology». *Marine Ecology – an Evolutionary Perspective*, vol. 30, p. 1-12.

--------. 2010. «The study of species in the era of biodiversity: a tale of stupidity». *Diversity*, vol. 2, no 1, p. 115-126.

Boertmann, D., P. Lyngs, F. R. Merkel and A. Mosbech. 2004. «The significance of Southwest Greenland as winter quarters for seabirds». *Bird Conservation International*, vol. 14, no 2, p. 87-112.

Bohonak, A. J. 1999. «Dispersal, gene flow, and population structure». *Quarterly Review of Biology*, vol. 74, no 1, p. 21-45.

Bohonak, A. J., and A. G. Vandergast. 2011. «The value of DNA sequence data for studying landscape genetics». *Molecular Ecology*, vol. 20, no 12, p. 2477-2479.

Bouchet, P. 2006. «The magnitude of marine biodiversity». In *The Exploration of Marine Biodiversity: Scientific and Technological Challenges,* C.M. Duarte, p. 31-64. Bilbao: Fundacion BBVA.

Bousfield, E. L. 1973. *Shallow-water Gammaridean Amphipoda of New England*. Ithaca, NY: Corell University Press.

--------. 1982. «The amphipod superfamily Talitroidea in the Northeastern Pacific region. I. Family Talitridae: systematics and distributional ecology». *Publications in Biological Oceanography*, vol. 11, p. 1-75.

Bousfield, E. L. 1984. «Recent advances in the systematics and biogeography of land hoppers (Amphipoda:Talitridae) of the Indo-Pacific region». *Bishop Museum Special Publications*, vol. 72, p. 171-210.

Bousfield, E. L. 1991. «New sandhoppers (Crustacea: Amphipoda) from the Gulf coast of the United States». *Gulf Research Reports*, vol. 8, no 3, p. 271-283.

Bousfield, E. L., and G. O. Poinar. 1995. «New terrestrial amphipod from Tertiary amber deposits of the Dominican Republic». *Journal of Crustacean Biology*, vol. 15, no 4, p. 746-755.

Boxshall, G. A. 2007. «Crustacean classification: on-going controversies and unresolved problems». *Zootaxa*, no 1668, p. 313-325.

Breton, S., H. D. Beaupre, D. T. Stewart, W. R. Hoeh and P. U. Blier. 2007. «The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough?». *Trends in Genetics*, vol. 23, no 9, p. 465-474.

Breton, S., F. Dufresne, G. Desrosiers and P. U. Blier. 2003. «Population structure of two northern hemisphere polychaetes, *Neanthes virens* and *Hediste diversicolor* (Nereididae), with different life-history traits». *Marine Biology*, vol. 142, no 4, p. 707-715.

Briggs, J. C. 1994. «Species diversity – land and sea compared». *Systematic Biology*, vol. 43, no 1, p. 130-135.

--------. 2007. «Marine biogeography and ecology: invasions and introductions». *Journal of Biogeography*, vol. 34, no 2, p. 193-198.

Brochmann, C., T. M. Gabrielsen, I. Nordal, J. Y. Landvik and R. Elven. 2003. «Glacial survival or *tabula rasa*? The history of North Atlantic biota revisited». *Taxon*, vol. 52, no 3, p. 417-450.

Browne, W. E., S. H. D. Haddock and M. Q. Martindale. 2007. «Phylogenetic analysis of lineage relationships among hyperiid amphipods as revealed by examination of the mitochondrial gene, cytochrome oxidase I (COI)». *Integrative and Comparative Biology*, vol. 47, no 6, p. 815-830.

Brunel, P. 1960. «Artificial key to the Mysidacea of the Canadian Atlantic continental shelf». *Canadian Journal of Zoology*, vol. 38, p. 851-855.

--------. 2005. «Visages de la biodiversite marine». *VertigO*, vol. 6, no 1, p. 1-13.

Brunel, P., L. Bossé and G. Lamarche. 1998. *Catalogue of the marine invertebrates of the Estuary and Gulf of St. Lawrence*. Ottawa: Canadian Special Publication of Fisheries and Aquatic Sciences 126, National Research Council of Canada.

Bucklin, A., M. Guarnieri, R. S. Hill, A. M. Bentley and S. Kaartvedt. 1999. «Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR». *Hydrobiologia*, vol. 401, p. 239-254.

Bucklin, A., R. R. Hopcroft, K. N. Kosobokova, L. M. Nigro, B. D. Ortman, R. M. Jennings and C. J. Sweetman. 2010. «DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition». *Deep Sea Research Part II – Topical Studies in Oceanography*, vol. 57, no 1-2, p. 40-48.

Bucklin, A., P. H. Wiebe, S. B. Smolenack, N. J. Copley, J. G. Beaudet, K. G. Bonner, J. Farber-Lorda and J. J. Pierson. 2007. «DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea)». *Journal of Plankton Research*, vol. 29, no 6, p. 483-493.

Buhay, J. E. 2009. «"COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies». *Journal of Crustacean Biology*, vol. 29, no 1, p. 96-110.

Burns, J. M., D. H. Janzen, M. Hajibabaei, W. Hallwachs and P. D. N. Hebert. 2008. «DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no 17, p. 6350-6355.

Byrkjedal, I., D. J. Rees and E. Willassen. 2007. «Lumping lumpsuckers: molecular and morphological insights into the taxonomic status of *Eumicrotremus spinosus* (Fabricius, 1776) and *Eumicrotremus eggvinii* Koefoed, 1956 (Teleostei : Cyclopteridae)». *Journal of Fish Biology*, vol. 71, p. 111-131.

Cardenas, P., C. Menegola, H. T. Rapp and M. C. Diaz. 2009. «Morphological description and DNA barcodes of shallow-water *Tetractinellida* (Porifera: Demospongiae) from Bocas del Toro, Panama, with description of a new species». *Zootaxa*, no 2276, p. 1-39.

Carlton, J. T. 1989. «Man's role in changing the face of the ocean – biological invasions and implications for conservation of near-shore environments». *Conservation Biology*, vol. 3, no 3, p. 265-273.

Carr, C. M., S. M. Hardy, T. M. Brown, T. A. Macdonald and P. D. N. Hebert. 2011. «A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes». *PLoS ONE*, vol. 6, no 7, e22232.

CBD. 1992. *Convention on Biological Diversity*. www.cbd.int.

Chapman, A. D. 2009. *Numbers of living species in Australia and the world.* Canberra: Report for the Australian Biological Resources Study.

Charbit, S., C. Ritz, G. Philippon, V. Peyaud and M. Kageyama. 2007. «Numerical reconstructions of the Northern Hemisphere ice sheets through the last glacial-interglacial cycle». *Climate of the Past*, vol. 3, no 1, p. 15-37.

Cheng, Y.-T., K. Nakazono, Y. K. Lin and B. K. K. Chan. 2011. «Cryptic diversity of the semi-terrestrial amphipod *Platorchestia japonica* (Tattersall, 1922) (Amphipoda: Talitrida: Talitridae) in Japan and Taiwan, with description of a new species». *Zootaxa*, no 2787, p. 1-18.

Clare, E. L., B. K. Lim, M. D. Engstrom, J. L. Eger and P. D. N. Hebert. 2007. «DNA barcoding of Neotropical bats: species identification and discovery within Guyana». *Molecular Ecology Notes*, vol. 7, no 2, p. 184-190.

Cohen, N. J., J. R. Deeds, E. S. Wong, R. H. Hanner, H. F. Yancy, K. D. White, T. M. Thompson, M. Wahl, T. D. Pham, F. M. Guichard, I. Huh, C. Austin, G. Dizikes and S. I. Gerber. 2009. «Public health response to puffer fish (Tetrodotoxin) poisoning from mislabeled product». *Journal of Food Protection*, vol. 72, no 4, p. 810-817.

Collin, R. 2001. «The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Calyptraeidae)». *Molecular Ecology*, vol. 10, no 9, p. 2249-2262.

CoML. 2010. *First Census of Marine Life: Highlights of a decade of discovery*. Washington DC: Consortium for Ocean Leadership.

Conceição, M. B., J. D. D. Bishop and J. P. Thorpe. 1998. «Genetic relationships between ecologically divergent species of talitrid amphipod (Crustacea)». *Marine Ecology – Progress Series*, vol. 165, p. 225-233.

Conklin, K. Y., A. Kurihara and A. R. Sherwood. 2009. «A molecular method for identification of the morphologically plastic invasive algal genera *Eucheuma* and *Kappaphycus* (Rhodophyta, Gigartinales) in Hawaii». *Journal of Applied Phycology*, vol. 21, no 6, p. 691-699.

Corander, J., P. Marttinen, J. Siren and J. Tang. 2008. «Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations». *BMC Bioinformatics*, vol. 9, e559.

Costa, F. O., J. R. deWaard, J. Boutillier, S. Ratnasingham, R. T. Dooh, M. Hajibabaei and P. D. N. Hebert. 2007. «Biological identifications through DNA barcodes: the case of the Crustacea». *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 64, no 2, p. 272-295.

Costa, F. O., C. M. Henzler, D. H. Lunt, N. M. Whiteley and J. Rock. 2009. «Probing marine *Gammarus* (Amphipoda) taxonomy with DNA barcodes». *Systematics and Biodiversity*, vol. 7, no 4, p. 365-379.

Costa, F. O., T. Neuparth, C. W. Theodorakis, M. H. Costa and L. R. Shugart. 2004. «RAPD analysis of southern populations of *Gammarus locusta*: comparison with allozyme data and ecological inferences». *Marine Ecology – Progress Series*, vol. 277, p. 197-207.

Costanza, R., R. dArge, R. deGroot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, R. V. Oneill, J. Paruelo, R. G. Raskin, P. Sutton and M. vandenBelt. 1997. «The value of the world's ecosystem services and natural capital». *Nature*, vol. 387, no 6630, p. 253-260.

Costello, M. J. 1998. «To know, research, manage and conserve marine biodiversity». *Oceanis – Concepts and methods for studying marine biodiversity, from gene to ecosystem*, vol. 24, no 4, p. 25-49.

Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace and R. K. Wayne. 2000. «Considering evolutionary processes in conservation biology». *Trends in Ecology and Evolution*, vol. 15, no 7, p. 290-295.

Cristescu, M. E. A., J. D. S. Witt, I. A. Grigorovich, P. D. N. Hebert and H. J. MacIssac. 2004. «Dispersal of the Ponto-Caspian amphipod *Echinogammarus ischnus*: invasion waves from the Pleistocene to the present». *Heredity*, vol. 92, no 3, p. 197-203.

Crutsinger, G. M., M. W. Cadotte and N. J. Sanders. 2009. «Plant genetics shapes inquiline community structure across spatial scales». *Ecology Letters*, vol. 12, no 4, p. 285-292.

Crutsinger, G. M., M. D. Collins, J. A. Fordyce, Z. Gompert, C. C. Nice and N. J. Sanders. 2006. «Plant genotypic diversity predicts community structure and governs an ecosystem process». *Science*, vol. 313, no 5789, p. 966-968.

da Silva, J. M., S. Creer, A. dos Santos, A. C. Costa, M. R. Cunha, F. O. Costa and G. R. Carvalho. 2011. «Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca)». *PLoS ONE*, vol. 6, no 5, e19449.

Davis, M., M. K. Chew, R. J. Hobbs, A. E. Lugo, J. J. Ewel, G. J. Vermeij, J. H. Brown, M. L. Rosenzweig, M. R. Gardener, S. P. Carroll, K. Thompson, S. T. A. Pickett, J. C. Stromberg, P. Del Tredici, K. N. Suding, J. G. Ehrenfeld, J. P. Grime, J. Mascaro and J. C. Briggs. 2011. «Don't judge species on their origins». *Nature*, vol. 474, no 7350, p. 153-154.

Davis, W. 2010. «Last of their kind». *Scientific American*, vol. 303, p. 60-67.

Davolos, D., and N. Maclean. 2005. «Mitochondrial COI-NC-COII sequences in talitrid amphipods (Crustacea)». *Heredity*, vol. 94, no 1, p. 81-86.

Dawson, M. N. 2001. «Phylogeography in coastal marine animals: a solution from California?». *Journal of Biogeography*, vol. 28, no 6, p. 723-736.

De Matthaeis, E., D. Davolos and M. Cobolli. 1998. «Genetic divergence between populations and species of talitrids from Aegean Islands». *Journal of Heredity*, vol. 89, no 1, p. 37-43.

De Wit, P., E. Rota and C. Erseus. 2009. «*Grania* (Annelida: Clitellata: Enchytraeidae) of the Great Barrier Reef, Australia, including four new species and a re-description of *Grania trichaeta* Jamieson, 1977». *Zootaxa*, no 2165, p. 16-38.

Deagle, B. E., A. Chiaradia, J. McInnes and S. N. Jarman. 2010. «Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out?». *Conservation Genetics*, vol. 11, no 5, p. 2039-2048.

Derycke, S., G. Fonseca, A. Vierstraete, J. Vanfleteren, M. Vincx and T. Moens. 2008. «Disentangling taxonomy within the *Rhabditis* (*Pellioditis*) *marina* (Nematoda, Rhabditidae) species complex using molecular and morhological tools». *Zoological Journal of the Linnean Society*, vol. 152, no 1, p. 1-15.

Diamond, J. 2005. *Collapse: how societies choose to fail or succeed*. New York, NY: Viking Press.

Duffy, J. E., and J. J. Stachowicz. 2006. «Why biodiversity is important to oceanography: potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes». *Marine Ecology – Progress Series*, vol. 311, p. 179-189.

Dufresne, F., E. Bourget and L. Bernatchez. 2002. «Differential patterns of spatial divergence in microsatellite and allozyme alleles: further evidence for locus-specific selection in the acorn barnacle, *Semibalanus balanoides*?». *Molecular Ecology*, vol. 11, no 1, p. 113-123.

Dupanloup, I., S. Schneider and L. Excoffier. 2002. «A simulated annealing approach to define the genetic structure of populations». *Molecular Ecology*, vol. 11, no 12, p. 2571-2581.

Ebach, M. C., and M. R. de Carvalho. 2010. «Anti-intellectualism in the DNA Barcoding Enterprise». *Zoologia*, vol. 27, no 2, p. 165-178.

Ebach, M. C., and C. Holdrege. 2005. «More taxonomy, not DNA barcoding». *Bioscience*, vol. 55, no 10, p. 822-823.

Ekrem, T., E. Willassen and E. Stur. 2007. «A comprehensive DNA sequence library is essential for identification with DNA barcodes». *Molecular Phylogenetics and Evolution*, vol. 43, no 2, p. 530-542.

Emerson, B. C. 2007. «Alarm bells for the molecular clock? No support for Ho *et al.*'s model of time-dependent molecular rate estimates». *Systematic Biology*, vol. 56, no 2, p. 337-345.

Englisch, U., C. O. Coleman and J. W. Wagele. 2003. «First observations on the phylogeny of the families Gammaridae, Crangonyctidae, Melitidae, Niphargidae, Megaluropidae and Oedicerotidae (Amphipoda, Crustacea), using small subunit rDNA gene sequences». *Journal of Natural History*, vol. 37, no 20, p. 2461-2486.

Erpenbeck, D., J. N. A. Hooper and G. Worheide. 2006. «CO1 phylogenies in diploblasts and the 'Barcoding of Life' – are we sequencing a suboptimal partition?». *Molecular Ecology Notes*, vol. 6, no 2, p. 550-553.

Etter, R. J., M. A. Rex, M. C. Chase and J. M. Quattro. 1999. «A genetic dimension to deep-sea biodiversity». *Deep-Sea Research Part I – Oceanographic Research Papers*, vol. 46, no 6, p. 1095-1099.

Excoffier, L., and H. E. L. Lischer. 2010. «Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows». *Molecular Ecology Resources*, vol. 10, no 3, p. 564-567.

Faith, D. P. 1994. «Phylogenetic pattern and the quantification of organismal biodiversity». *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, vol. 345, no 1311, p. 45-58.

Faith, D. P., and A. M. Baker. 2006. «Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges». *Evolutionary Bioinformatics*, vol. 2, p. 121-128.

Faurby, S., A. Jørgensen, R. M. Kristensen and P. Funch. 2011. «Phylogeography of North Atlantic intertidal tardigrades: refugia, cryptic speciation and the history of the Mid-Atlantic Islands». *Journal of Biogeography*, vol. 38, no 8, p. 1613-1624.

Ficetola, G. F., C. Miaud, F. Pompanon and P. Taberlet. 2008. «Species detection using environmental DNA from water samples». *Biology Letters*, vol. 4, no 4, p. 423-425.

Figuerola, J., and A. J. Green. 2002. «Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies». *Freshwater Biology*, vol. 47, no 3, p. 483-494.

Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek. 1994. «DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates». *Molecular Marine Biology and Biotechnology*, vol. 3, no 5, p. 294-299.

Frankham, R. 1996. «Relationship of genetic variation to population size in wildlife». *Conservation Biology*, vol. 10, no 6, p. 1500-1508.

Freire, J., and E. Gonzalez-Gurriaran. 1998. «New approaches to the behavioural ecology of decapod crustaceans using telemetry and electronic tags». *Hydrobiologia*, vol. 372, p. 123-132.

Frère, C. H., E. Krzyszczyk, E. M. Patterson, S. Hunter, A. Ginsburg and J. Mann. 2010. «Thar She Blows! A novel method for DNA collection from cetacean blow». *PLoS ONE*, vol. 5, no 8, e12299.

Frézal, L., and R. Leblois. 2008. «Four years of DNA barcoding: Current advances and prospects». *Infection Genetics and Evolution*, vol. 8, no 5, p. 727-736.

Fu, Y. X. 1997. «Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection». *Genetics*, vol. 147, no 2, p. 915-925.

Futuyma, D. J. 1998. *Evolutionary biology*. Sunderland, MA: Sinauer Associates.

Galtier, N., B. Nabholz, S. Glémin and G. D. D. Hurst. 2009. «Mitochondrial DNA as a marker of molecular diversity: a reappraisal». *Molecular Ecology*, vol. 18, no 22, p. 4541-4550.

Gaston, K. J. 2000. «Global patterns in biodiversity». *Nature*, vol. 405, no 6783, p. 220-227.

Gershoni, M., A. R. Templeton and D. Mishmar. 2009. «Mitochondrial bioenergetics as a major motive force of speciation». *Bioessays*, vol. 31, no 6, p. 642-650.

Gilly, W. F., U. Markaida, C. H. Baxter, B. A. Block, A. Boustany, L. Zeidberg, K. Reisenbichler, B. Robison, G. Bazzino and C. Salinas. 2006. «Vertical and horizontal migrations by the jumbo squid *Dosidicus gigas* revealed by electronic tagging». *Marine Ecology – Progress Series*, vol. 324, p. 1-17.

Golikov, A. N., and N. L. Tzvetkova. 1972. «The ecological principle of evolutionary reconstruction as illustrated by marine animals». *Marine Biology*, vol. 14, no 1, p. 1-9.

Gómez, A., P. J. Wright, D. H. Lunt, J. M. Cancino, G. R. Carvalho and R. N. Hughes. 2007. «Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon». *Proceedings of the Royal Society B – Biological Sciences*, vol. 274, no 1607, p. 199-207.

Grassle, J. F., and N. J. Maciolek. 1992. «Deep-sea species richness – regional and local diversity estimates from quantitative bottom samples». *American Naturalist*, vol. 139, no 2, p. 313-341.

Gray, J. S. 1997. «Marine biodiversity: patterns, threats and conservation needs». *Biodiversity and Conservation*, vol. 6, no 1, p. 153-175.

Grosberg, R. K., and C. W. Cunningham. 2000. «Genetic structure in the sea: from populations to communities». In *Marine Community Ecology*, M. D. Bertness, S. D. Gaines and M. E. Hay, p. 61-84. Sunderland, MA: Sinauer Associates.

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs and P. D. N. Hebert. 2006. «DNA barcodes distinguish species of tropical Lepidoptera». *Proceedings of*

the National Academy of Sciences of the United States of America, vol. 103, no 4, p. 968-971.

Hajibabaei, M., G. A. C. Singer, P. D. N. Hebert and D. A. Hickey. 2007. «DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics». *Trends in Genetics*, vol. 23, no 4, p. 167-172.

Halpern, B. S., S. Walbridge, K. A. Selkoe, C. V. Kappel, F. Micheli, C. D'Agrosa, J. F. Bruno, K. S. Casey, C. Ebert, H. E. Fox, R. Fujita, D. Heinemann, H. S. Lenihan, E. M. P. Madin, M. T. Perry, E. R. Selig, M. Spalding, R. Steneck and R. Watson. 2008. «A global map of human impact on marine ecosystems». *Science*, vol. 319, no 5865, p. 948-952.

Hamilton, K. G. A. 2002. «Îles-de-la-Madeleine (Magdalen Is.): a glacial refugium for short-horned bugs (Homoptera: Auchenorrhyncha)?». *Le Naturaliste Canadien*, vol. 26, no 1, p. 25-40.

Harris, D. J. 2003. «Can you bank on GenBank?». *Trends in Ecology and Evolution*, vol. 18, no 7, p. 317-319.

Hebert, P. D. N., A. Cywinska, S. L. Ball and J. R. DeWaard. 2003. «Biological identifications through DNA barcodes». *Proceedings of the Royal Society of London Series B – Biological Sciences*, vol. 270, no 1512, p. 313-321.

Hebert, P. D. N., S. Ratnasingham and J. R. deWaard. 2003. «Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species». *Proceedings of the Royal Society of London Series B – Biological Sciences*, vol. 270, p. 96-S99.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak and C. M. Francis. 2004. «Identification of birds through DNA barcodes». *PLoS Biology*, vol. 2, no 10, p. 1657-1663.

Hedgecock, D. 1986. «Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates?». *Bulletin of Marine Science*, vol. 39, no 2, p. 550-564.

Henzler, C. M. 2006. «Rafting and refugia in North Atlantic amphipods: understanding the effects of Pleistocene climate change on low-dispersal marine species». PhD thesis, Durham (NC), Duke University.

Henzler, C. M., and A. Ingolfsson. 2008. «The biogeography of the beachflea, *Orchestia gammarellus* (Crustacea, Amphipoda, Talitridae), in the North Atlantic with special reference to Iceland: a morphometric and genetic study». *Zoologica Scripta*, vol. 37, no 1, p. 57-70.

Herkul, K., J. Kotta and I. Kotta. 2006. «Distribution and population characteristics of the alien talitrid amphipod *Orchestia cavimana* in relation to environmental conditions in the Northeastern Baltic Sea». *Helgoland Marine Research*, vol. 60, no 2, p. 121-126.

Hewitt, G. 2000. «The genetic legacy of the Quaternary ice ages». *Nature*, vol. 405, no 6789, p. 907-913.

Hewitt, G. M. 2004. «Genetic consequences of climatic oscillations in the Quaternary». *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, vol. 359, no 1442, p. 183-195.

Hey, J. 2006. «On the failure of modern species concepts». *Trends in Ecology and Evolution*, vol. 21, no 8, p. 447-450.

Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B. Johnson, L. Rissler, P. F. Victoriano and A. D. Yoder. 2010. «Phylogeography's past, present, and future: 10 years after Avise, 2000». *Molecular Phylogenetics and Evolution*, vol. 54, no 1, p. 291-301.

Hohenlohe, P. A. 2004. «Limits to gene flow in marine animals with planktonic larvae: models of *Littorina* species around Point Conception, California». *Biological Journal of the Linnean Society*, vol. 82, no 2, p. 169-187.

Hollingsworth, P. M., L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, R. S. Cowan, D. L. Erickson, A. J. Fazekas, S. W. Graham, K. E. James, K. J. Kim, W. J. Kress, H. Schneider, J. van AlphenStahl, S. C. H. Barrett, C. van den Berg, D. Bogarin, K. S. Burgess, K. M. Cameron, M. Carine, J. Chacon, A. Clark, J. J. Clarkson, F. Conrad, D. S. Devey, C. S. Ford, T. A. J. Hedderson, M. L. Hollingsworth, B. C. Husband, L. J. Kelly, P. R. Kesanakurti, J. S. Kim, Y. D. Kim, R. Lahaye, H. L. Lee, D. G. Long, S. Madrinan, O. Maurin, I. Meusnier, S. G. Newmaster, C. W. Park, D. M. Percy, G. Petersen, J. E. Richardson, G. A. Salazar, V. Savolainen, O. Seberg, M. J. Wilkinson, D. K. Yi, D. P. Little and C. P. W. Grp. 2009. «A DNA barcode for land plants». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no 31, p. 12794-12797.

Holmes, B. H., D. Steinke and R. D. Ward. 2009. «Identification of shark and ray fins using DNA barcoding». *Fisheries Research*, vol. 95, no 2-3, p. 280-288.

Hou, Z. G., J. H. Fu and S. Q. Li. 2007. «A molecular phylogeny of the genus *Gammarus* (Crustacea : Amphipoda) based on mitochondrial and nuclear gene sequences». *Molecular Phylogenetics and Evolution*, vol. 45, no 2, p. 596-611.

Hu, Y. P., R. A. Lutz and R. C. Vrijenhoek. 1992. «Electrophoretic identification and genetic analysis of bivalve larvae». *Marine Biology*, vol. 113, no 2, p. 227-230.

Hu, Z., M. D. Guiry, A. T. Critchley and D. Duan. 2010. «Phylogeographic patterns indicate transatlantic migration from Europe to North America in the red seaweed *Chondrus crispus* (Gigartinales, Rhodophyta)». *Journal of Phycology*, vol. 46, no 5, p. 889-900.

Hubert, N., R. Hanner, E. Holm, N. E. Mandrak, E. Taylor, M. Burridge, D. Watkinson, P. Dumont, A. Curry, P. Bentzen, J. Zhang, J. April and L. Bernatchez. 2008. «Identifying Canadian freshwater fishes through DNA barcodes». *PLoS ONE*, vol. 3, no 6, p. 1-8.

Hughes, A. R., and J. J. Stachowicz. 2004. «Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no 24, p. 8998-9002.

Hultgren, K. M., and J. J. Stachowicz. 2008. «Molecular phylogeny of the brachyuran crab superfamily Majoidea indicates close congruence with trees based on larval morphology». *Molecular Phylogenetics and Evolution*, vol. 48, no 3, p. 986-996.

Ilves, K. L., W. Huang, J. P. Wares and M. J. Hickerson. 2010. «Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation». *Molecular Ecology*, vol. 19, no 20, p. 4505-4519.

Ingolfsson, A. 1992. «The origin of the rocky shore fauna of Iceland and the Canadian Maritimes». *Journal of Biogeography*, vol. 19, no 6, p. 705-712.

--------. 1995. «Floating clumps of seaweed around Iceland – natural microcosms and a means of dispersal for shore fauna». *Marine Biology*, vol. 122, no 1, p. 13-21.

--------. 2009. «A marine refugium in Iceland during the Last Glacial Maximum: fact or fiction?». *Zoologica Scripta*, vol. 38, no 6, p. 663-665.

Ingolfsson, A., and I. Agnarsson. 2003. «Amphipods and isopods in the rocky intertidal: dispersal and movements during high tide». *Marine Biology*, vol. 143, no 5, p. 859-866.

Ivanova, N. V., A. V. Borisenko and P. D. N. Hebert. 2009. «Express barcodes: racing from specimen to identification». *Molecular Ecology Resources*, vol. 9, p. 35-41.

Ivanova, N. V., J. R. Dewaard and P. D. N. Hebert. 2006. «An inexpensive, automation-friendly protocol for recovering high-quality DNA». *Molecular Ecology Notes*, vol. 6, no 4, p. 998-1002.

Jaeger, R. G., and T. R. Halliday. 1998. «On confirmatory versus exploratory research». *Herpetologica*, vol. 54, no 2, p. S64-S66.

Järnegren, J., C. Schander, J. A. Sneli, V. Rønningen and C. M. Young. 2007. «Four genes, morphology and ecology: distinguishing a new species of *Acesta* (Mollusca; Bivalvia) from the Gulf of Mexico». *Marine Biology*, vol. 152, no 1, p. 43-55.

Jennings, R. M., A. Bucklin, H. Ossenbrugger and R. R. Hopcroft. 2010. «Species diversity of planktonic gastropods (Pteropoda and Heteropoda) from six ocean regions based on DNA barcode analysis». *Deep Sea Research Part II – Topical Studies in Oceanography*, vol. 57, no 24-26, p. 2199-2210.

Jennings, R. M., A. Bucklin and A. Pierrot-Bults. 2010. «Barcoding of Arrow Worms (Phylum Chaetognatha) from three oceans: genetic diversity and evolution within an enigmatic phylum». *PLoS ONE*, vol. 5, no 4, e9949.

Johnson, M. S., and R. Black. 2006. «Effects of mode of reproduction on genetic divergence over large spatial and temporal scales in intertidal snails of the genus *Bembicium* Philippi (Gastropoda : Littorinidae)». *Biological Journal of the Linnean Society*, vol. 89, no 4, p. 689-704.

Jombart, T., S. Devillard and F. Balloux. 2010. «Discriminant analysis of principal components: a new method for the analysis of genetically structured populations». *BMC Genetics*, vol. 11, e94.

Kaim-Malka, R. A. 2000. «Elevation of two eastern Atlantic varieties of *Ampelisca brevicornis* (Costa, 1853) (Crustacea, Amphipoda) to full specific rank with redescription of the species». *Journal of Natural History*, vol. 34, no 10, p. 1939-1966.

Kappner, I., and R. Bieler. 2006. «Phylogeny of venus clams (Bivalvia : Venerinae) as inferred from nuclear and mitochondrial gene sequences». *Molecular Phylogenetics and Evolution*, vol. 40, no 2, p. 317-331.

Kelly, D. W., H. J. MacIsaac and D. D. Heath. 2006. «Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate». *Evolution*, vol. 60, no 2, p. 257-267.

Kemppainen, P., M. Panova, J. Hollander and K. Johannesson. 2009. «Complete lack of mitochondrial divergence between two species of NE Atlantic marine

intertidal gastropods». *Journal of Evolutionary Biology*, vol. 22, no 10, p. 2000-2011.

Kenchington, E. L., G. C. Harding, M. W. Jones and P. A. Prodohl. 2009. «Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*». *Molecular Ecology*, vol. 18, no 8, p. 1654-1667.

Kerr, K. C. R. 2011. «Searching for evidence of selection in avian DNA barcodes». *Molecular Ecology Resources*, vol. 11, no 6, p. 1045-1056.

Ketmaier, V., and E. De Matthaeis. 2010. «Allozymes and mtDNA reveal two divergent lineages in *Orchestia cavimana* (Amphipoda: Talitridae)». *Journal of Crustacean Biology*, vol. 30, no 2, p. 307-311.

Ketmaier, V., F. Scapini and E. De Matthaeis. 2003. «Exploratory analysis of talitrid population genetics as an indicator of the quality of sandy beaches». *Estuarine, Coastal and Shelf Science*, vol. 58, no 1, p. 159-167.

Ki, J.-S., H. Hop, S.-J. Kim, I.-C. Kim, H. G. Park and J.-S. Lee. 2010. «Complete mitochondrial genome sequence of the Arctic gammarid, *Onisimus nanseni* (Crustacea; Amphipoda): novel gene structures and unusual control region features». *Comparative Biochemistry and Physiology D – Genomics and Proteomics*, vol. 5, no 2, p. 105-115.

Kimura, M. 1980. «A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences». *Journal of Molecular Evolution*, vol. 16, no 2, p. 111-120.

King, R., and J. M. C. Holmes. 2004. «A new species of *Ischyrocerus* (Crustacea: Amphipoda) from Ireland, with a review of *Ischyrocerus anguipes* and *Ischyrocerus minutus* from the North-East Atlantic». *Journal of Natural History*, vol. 38, no 14, p. 1757-1772.

Knowlton, N. 1993. «Sibling species in the sea». *Annual Review of Ecology and Systematics*, vol. 24, p. 189-216.

--------. 2000. «Molecular genetic analyses of species boundaries in the sea». *Hydrobiologia*, vol. 420, p. 73-90.

Kochzius, M., M. Nolte, H. Weber, N. Silkenbeumer, S. Hjorleifsdottir, G. O. Hreggvidsson, V. Marteinsson, K. Kappel, S. Planes, F. Tinti, A. Magoulas, E. G. Vazquez, C. Turan, C. Hervet, D. C. Falgueras, A. Antoniou, M. Landi and D. Blohm. 2008. «DNA microarrays for identifying fishes». *Marine Biotechnology*, vol. 10, no 2, p. 207-217.

Kornobis, E., S. Palsson, B. K. Kristjansson and J. Svavarsson. 2010. «Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland». *Molecular Ecology*, vol. 19, no 12, p. 2516-2530.

Krabbe, K., F. Leese, C. Mayer, R. Tollrian and C. Held. 2010. «Cryptic mitochondrial lineages in the widespread pycnogonid *Colossendeis megalonyx* Hoek, 1881 from Antarctic and Subantarctic waters». *Polar Biology*, vol. 33, no 4, p. 575-575.

Krug, P. J., R. A. Ellingson, R. Burton and A. Valdes. 2007. «A new *Poecilogonous* species of sea slug (Opisthobranchia: Sacoglossa) from California: Comparison with the planktotrophic congener *Alderia modesta* (Loven, 1844)». *Journal of Molluscan Studies*, vol. 73, p. 29-38.

Kyle, C. J., and E. G. Boulding. 2000. «Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal». *Marine Biology*, vol. 137, no 5-6, p. 835-845.

Lankau, R. A. 2011. «Conflicts in maintaining biodiversity at multiple scales». *Molecular Ecology*, vol. 20, no 10, p. 2035-2037.

Larsson, K., A. Ahmadzadeh and U. Jondelius. 2008. «DNA taxonomy of Swedish *Catenulida* (Platyhelminthes) and a phylogenetic framework for catenulid classification». *Organisms Diversity and Evolution*, vol. 8, no 5, p. 399-412.

LeCroy, S. E. 2010. *An illustrated identification guide to the nearshore marine and estuarine gammaridean Amphipoda of Florida. Volume 5. Family Talitridae*. Tallahassee: Florida Department of Environmental Protection.

LeCroy, S. E., R. Gasca, I. Winfield, M. Ortiz and E. Escobar-Briones. 2009. «Amphipoda (Crustacea) of the Gulf of Mexico». In *Gulf of Mexico: Origin, Waters, and Biota. Volume 1, Biodiversity,* D.L. Felder, Camp, D.K., p. 941-972. Corpus Christi, TX: Texas A&M University Press.

Lefébure, T., C. J. Douady, M. Gouy and J. Gibert. 2006. «Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation». *Molecular Phylogenetics and Evolution*, vol. 40, no 2, p. 435-447.

Lemire, B. 2005. «Mitochondrial genetics ». In *WormBook,* The *C. elegans* Research Community (http://www.wormbook.org).

Leung, T. L. F., K. M. Donald, D. B. Keeney, A. V. Koehler, R. C. Peoples and R. Poulin. 2009. «Trematode parasites of Otago Harbour (New Zealand) soft-sediment intertidal ecosystems: life cycles, ecological roles and DNA

barcodes». *New Zealand Journal of Marine and Freshwater Research*, vol. 43, no 4, p. 857-865.

Lindeman, D. 1991. «Phylogeny and zoogeography of the New-World terrestrial amphipods (landhoppers) (Crustacea, Amphipoda, Talitridae)». *Canadian Journal of Zoology*, vol. 69, no 4, p. 1104-1116.

Locke, S. A., J. D. McLaughlin, S. Dayanandan and D. J. Marcogliese. 2010. «Diversity and specificity in *Diplostomum* spp. metacercariae in freshwater fishes revealed by cytochrome *c* oxidase I and internal transcribed spacer sequences». *International Journal for Parasitology*, vol. 40, no 3, p. 333-343.

Lowry, J. K. (2010). World Amphipoda Database, World Register of Marine Species at http://www.marinespecies.org

Luttikhuizen, P. C., J. Campos, J. van Bleijswijk, K. Peijnenburg and H. W. van der Veer. 2008. «Phylogeography of the common shrimp, *Crangon crangon* (L.) across its distribution range». *Molecular Phylogenetics and Evolution*, vol. 46, no 3, p. 1015-1030.

MA. 2005. *Millennium Ecosystem Assessment: Ecosystems and Human Well-being: Synthesis*. Washington DC: Island Press.

Machida, R. J., Y. Hashiguchi, M. Nishida and S. Nishida. 2009. «Zooplankton diversity analysis through single-gene sequencing of a community sample». *BMC Genomics*, vol. 10, e438.

Maggs, C. A., R. Castilho, D. Foltz, C. Henzler, M. T. Jolly, J. Kelly, J. Olsen, K. E. Perez, W. Stam, R. Vainola, F. Viard and J. Wares. 2008. «Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa». *Ecology*, vol. 89, no 11, p. 108-122.

Malay, M. C. D., and G. Paulay. 2010. «Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*)». *Evolution*, vol. 64, no 3, p. 634-662.

Marko, P. B., S. C. Lee, A. M. Rice, J. M. Gramling, T. M. Fitzhenry, J. S. McAlister, G. R. Harper and A. L. Moran. 2004. «Mislabelling of a depleted reef fish». *Nature*, vol. 430, no 6997, p. 309-310.

Martin, J. W., and G. E. Davis. 2001. *An updated classification of the recent Crustacea*. Collection «Contributions in Science». Los Angeles, CA: Natural History Museum of Los Angeles.

Martinez, I., M. Aschan, T. Skjerdal and S. M. Aljanabi. 2006. «The genetic structure of *Pandalus borealis* in the Northeast Atlantic determined by RAPD analysis». *ICES Journal of Marine Science*, vol. 63, no 5, p. 840-850.

Mauchline, J. 1971. «Euphausiacea – Adults». *Conseil International pour l'Exploration de la Mer*, vol. 200 Plankton Sheet, no 134, p. 1-8.

McDevit, D. C., and G. W. Saunders. 2009. «On the utility of DNA barcoding for species differentiation among brown macroalgae (Phaeophyceae) including a novel extraction protocol». *Phycological Research*, vol. 57, no 2, p. 131-141.

McLaughlin, P. A., D. K. Camp, M. V. Angel, E. L. Bousfield, P. Brunel, R. C. Brusca, D. B. Cadien, A. C. Cohen, K. Conlan, L. G. Eldredge, D. L. Felder, J. W. Goy, T. A. Haney, B. Hann, R. W. Heard, E. A. Hendrycks, H. H. I. Hobbs, J. R. Holsinger, B. Kensley, D. R. Laubitz, S. E. LeCroy, R. Lemaitre, R. F. Maddocks, J. W. Martin, P. Mikkelsen, E. Nelson, W. A. Newman, R. M. Overstreet, W. J. Poly, W. W. Price, J. W. Reid, A. Robertson, D. C. Rogers, A. Ross, M. Schotte, F. R. Schram, C.-T. Shih, L. Watling and G. D. F. Wilson. 2005. *Common and scientific names of aquatic invertebrates from the United States and Canada: Crustaceans*. Bethesda, MD: American Fisheries Society.

Medeiros-Bergen, D. E., R. R. Olson, J. A. Conroy and T. D. Kocher. 1995. «Distribution of holothurian larvae determined with species-specific genetic probes». *Limnology and Oceanography*, vol. 40, no 7, p. 1225-1235.

Meyer, C. P., and G. Paulay. 2005. «DNA barcoding: error rates based on comprehensive sampling». *PLoS Biology*, vol. 3, no 12, p. 2229-2238.

Miglietta, M. P., A. Faucci and F. Santini. 2011. «Speciation in the sea: overview of the symposium and discussion of future directions». *Integrative and Comparative Biology*, vol. 51, no 3, p. 449-455.

Mikkelsen, N. T., C. Schander and E. Willassen. 2007. «Local scale DNA barcoding of bivalves (Mollusca): a case study». *Zoologica Scripta*, vol. 36, no 5, p. 455-463.

Mitchell, A. 2008. «DNA barcoding demystified». *Australian Journal of Entomology*, vol. 47, p. 169-173.

Mix, A. C., E. Bard and R. Schneider. 2001. «Environmental processes of the ice age: land, oceans, glaciers (EPILOG)». *Quaternary Science Reviews*, vol. 20, no 4, p. 627-657.

Miyamoto, H., and H. Morino. 2004. «Taxonomic studies on the Talitridae (Crustacea, Amphipoda) from Taiwan. II. The genus *Platorchestia*». *Publications of the Seto Marine Biological Laboratory*, vol. 40, p. 67-96.

Molnar, J. L., R. L. Gamboa, C. Revenga and M. D. Spalding. 2008. «Assessing the global threat of invasive species to marine biodiversity». *Frontiers in Ecology and the Environment*, vol. 6, no 9, p. 485-492.

Moniz, M. B. J., and I. Kaczmarska. 2010. «Barcoding of diatoms: nuclear encoded ITS revisited». *Protist*, vol. 161, no 1, p. 7-34.

Mooney, H. A. 2010. «The ecosystem-service chain and the biological diversity crisis». *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, vol. 365, p. 31-39.

Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson and B. Worm. 2011. «How many species are there on Earth and in the ocean?». *PLoS Biol*, vol. 9, no 8, e1001127.

Moritz, C., and D. P. Faith. 1998. «Comparative phylogeography and the identification of genetically divergent areas for conservation». *Molecular Ecology*, vol. 7, no 4, p. 419-429.

Mosquin, T., P. G. Whiting and D. E. McAllister. 1995. *Canada's biodiversity: the variety of life, its status, economic benefits, conservation costs and unmeet needs*. Ottawa: Canadian Museum of Nature.

Moura, C. J., D. J. Harris, M. R. Cunha and A. D. Rogers. 2008. «DNA barcoding reveals cryptic diversity in marine hydroids (Cnidaria, Hydrozoa) from coastal and deep-sea environments». *Zoologica Scripta*, vol. 37, no 1, p. 93-108.

Muhlin, J. F., and S. H. Brawley. 2009. «Recent versus relic: discerning the genetic signature of *Fucus vesiculosus* (Heterokontophyta; Phaeophyceae) in the northwestern Atlantic». *Journal of Phycology*, vol. 45, no 4, p. 828-837.

Naeem, S., D. E. Bunker, A. Hector, M. Loreau and C. Perrings. 2009. «Introduction: the ecological and social implications of changing biodiversity. An overview of a decade of biodiversity and ecosystem functioning research». In *Biodiversity, ecosystem functioning, and human wellbeing: An ecological and economic perspective,* D. Bunker S. Naeem, A. Hector, M. Loreau and C. Perrings. Oxford: Oxford University Press.

Naro-Maciel, E., B. Reid, N. N. Fitzsimmons, M. Le, R. DeSalle and G. Amato. 2009. «DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity». *Molecular Ecology Resources*, vol. 10, no 2, p. 252-263.

Neigel, J. E. 2009. «Population genetics and biogeography of the Gulf of Mexico». In *Gulf of Mexico: Origin, Waters, and Biota. Volume 1, Biodiversity,* D.L. Felder, Camp, D.K., p. 1353-1369. Corpus Christi, TX: Texas A&M University Press.

Nielsen, J. F., S. Lavery and A. N. Lorz. 2009. «Synopsis of a new collection of sea spiders (Arthropoda: Pycnogonida) from the Ross Sea, Antarctica». *Polar Biology*, vol. 32, no 8, p. 1147-1155.

Nielsen, R., and M. Matz. 2006. «Statistical approaches for DNA barcoding». *Systematic Biology*, vol. 55, no 1, p. 162-169.

O'Connor, N. E., and T. P. Crowe. 2005. «Biodiversity loss and ecosystem functioning: distinguishing between number and identity of species». *Ecology*, vol. 86, no 7, p. 1783-1796.

Olsen, J. L., F. W. Zechman, G. Hoarau, J. A. Coyer, W. T. Stam, M. Valero and P. Aberg. 2010. «The phylogeographic architecture of the fucoid seaweed *Ascophyllum nodosum*: an intertidal 'marine tree' and survivor of more than one glacial-interglacial cycle». *Journal of Biogeography*, vol. 37, no 5, p. 842-856.

Packer, L., J. Gibbs, C. Sheffield and R. Hanner. 2009a. «DNA barcoding and the mediocrity of morphology». *Molecular Ecology Resources*, vol. 9, p. 42-50.

Packer, L., J. C. Grixti, R. E. Roughley and R. Hanner. 2009b. «The status of taxonomy in Canada and the impact of DNA barcoding». *Canadian Journal of Zoology – Revue Canadienne de zoologie*, vol. 87, no 12, p. 1097-1110.

Palmer, M. E., and A. Ricciardi. 2004. «Physical factors affecting the relative abundance of native and invasive amphipods in the St. Lawrence River». *Canadian Journal of Zoology – Revue Canadienne De Zoologie*, vol. 82, no 12, p. 1886-1893.

Palumbi, S. R. 1994. «Genetic divergence, reproductive isolation, and marine speciation». *Annual Review of Ecology and Systematics*, vol. 25, p. 547-572.

--------. 2003. «Population genetics, demographic connectivity, and the design of marine reserves». *Ecological Applications*, vol. 13, no 1, p. S146-S158.

--------. 2004. «Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management». *Annual Review of Environment and Resources*, vol. 29, p. 31-68.

Panova, M., A. M. H. Blakeslee, A. W. Miller, T. Mäkinen, G. M. Ruiz, K. Johannesson and C. André. 2011. «Glacial history of the North Atlantic

marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages». *PLoS ONE*, vol. 6, no 3, e17511.

Pegg, G. G., B. Sinclair, L. Briskey and W. J. Aspden. 2006. «MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia». *Scientia Marina*, vol. 70, p. 7-12.

Persson, L. E. 2001. «Dispersal of *Platorchestia platensis* (Krøyer) (Amphipoda : Talitridae) along Swedish coasts: a slow but successful process». *Estuarine Coastal and Shelf Science*, vol. 52, no 2, p. 201-210.

Pielou, E. C. 1991. *After the Ice Age: the return of life to glaciated North America*. Chicago, IL: The University of Chicago Press.

Pimm, S. L., and P. Raven. 2000. «Biodiversity – Extinction by numbers». *Nature*, vol. 403, no 6772, p. 843-845.

Pimm, S. L., G. J. Russell, J. L. Gittleman and T. M. Brooks. 1995. «The future of biodiversity». *Science*, vol. 269, no 5222, p. 347-350.

Pinsky, M. L., O. P. Jensen, D. Ricard and S. R. Palumbi. 2011. «Unexpected patterns of fisheries collapse in the world's oceans». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no 20, p. 8317-8322.

Plaisance, L., N. Knowlton, G. Paulay and C. Meyer. 2009. «Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding». *Coral Reefs*, vol. 28, no 4, p. 977-986.

Pleijel, F., U. Jondelius, E. Norlinder, A. Nygren, B. Oxelman, C. Schander, P. Sundberg and M. Thollesson. 2008. «Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies». *Molecular Phylogenetics and Evolution*, vol. 48, no 1, p. 369-371.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin and A. P. Vogler. 2006. «Sequence-based species delimitation for the DNA taxonomy of undescribed insects». *Systematic Biology*, vol. 55, no 4, p. 595-609.

Porco, D., R. Rougerie, L. Deharveng and P. D. N. Hebert. 2010. «Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied arthropods in a high-throughput context: the example of Collembola». *Molecular Ecology Resources*, vol. 10, p. 942-945.

Posada, D. 2008. «jModelTest: phylogenetic model averaging». *Molecular Biology and Evolution*, vol. 25, no 7, p. 1253-1256.

Posada, D., and T. R. Buckley. 2004. «Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests». *Systematic Biology*, vol. 53, no 5, p. 793-808.

Prest, V. K., J. Terasmae, J. V. Matthews and S. Lichti-Federovich. 1976. «Late-Quaternary history of Magdalen Islands, Quebec». *Maritime sediments*, vol. 12, no 2, p. 39-59.

Puebla, O., J.-M. Sévigny, B. Sainte-Marie, J.-C. Brêthes, A. Burmeister, E. G. Dawe and M. Moriyasu. 2008. «Population genetic structure of the snow crab (*Chionoecetes opilio*) at the Northwest Atlantic scale». *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 65, no 3, p. 425-436.

Puillandre, N., E. E. Strong, P. Bouchet, M. C. Boisselier, A. Couloux and S. Samadi. 2009. «Identifying gastropod spawn from DNA barcodes: possible but not yet practicable». *Molecular Ecology Resources*, vol. 9, no 5, p. 1311-1321.

R Development Core Team. 2010. *R: a language and environment for statistical computing. version 2.13.1. http://www.r-project.org/.*

Rach, J., R. DeSalle, I. N. Sarkar, B. Schierwater and H. Hadrys. 2008. «Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata». *Proceedings of the Royal Society B – Biological Sciences*, vol. 275, no 1632, p. 237-247.

Radulovici, A. E., P. Archambault and F. Dufresne. 2010. «DNA barcodes for marine biodiversity: moving fast forward?». *Diversity*, vol. 2, no 4, p. 450-472.

Radulovici, A. E., B. Sainte-Marie and F. Dufresne. 2009. «DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: a regional-scale approach». *Molecular Ecology Resources*, vol. 9, p. 181-187.

Ratnasingham, S., and P. D. N. Hebert. 2007. «BOLD: The Barcode of Life Data System (www.barcodinglife.org)». *Molecular Ecology Notes*, vol. 7, no 3, p. 355-364.

Rauch, E. M., and Y. Bar-Yam. 2004. «Theory predicts the uneven distribution of genetic diversity within species». *Nature*, vol. 431, p. 449-452.

Raupach, M. J., S. Thatje, J. Dambach, P. Rehm, B. Misof and F. Leese. 2010a. «Genetic homogeneity and circum-Antarctic distribution of two benthic shrimp species of the Southern Ocean, *Chorismus antarcticus* and *Nematocarcinus lanceopes*». *Marine Biology*, vol. 157, no 8, p. 1783-1797.

Raupach, M., J. Astrin, K. Hannig, M. Peters, M. Stoeckle and J.-W. Wagele. 2010b. «Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes». *Frontiers in Zoology*, vol. 7, no 1, p. 26.

Regier, J. C., J. W. Shultz, A. Zwick, A. Hussey, B. Ball, R. Wetzer, J. W. Martin and C. W. Cunningham. 2010. «Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences». *Nature*, vol. 463, no 7284, p. 1079-1083.

Reusch, T. B. H., A. Ehlers, A. Hammerli and B. Worm. 2005. «Ecosystem recovery after climatic extremes enhanced by genotypic diversity». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no 8, p. 2826-2831.

Richards, V. P., J. D. Thomas, M. J. Stanhope and M. S. Shivji. 2007. «Genetic connectivity in the Florida reef system: comparative phylogeography of commensal invertebrates with contrasting reproductive strategies». *Molecular Ecology*, vol. 16, no 1, p. 139-157.

Ronquist, F., and J. P. Huelsenbeck. 2003. «MrBayes 3: Bayesian phylogenetic inference under mixed models». *Bioinformatics*, vol. 19, no 12, p. 1572-1574.

Rubinoff, D. 2006. «Utility of mitochondrial DNA barcodes in species conservation». *Conservation Biology*, vol. 20, no 4, p. 1026-1033.

Rubinoff, D., S. Cameron and K. Will. 2006. «A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification». *Journal of Heredity*, vol. 97, no 6, p. 581-594.

Sainte-Marie, B., and P. Brunel. 1985. «Suprabenthic gradients of swimming activity by cold-water gammaridean amphipod Crustacea over a muddy shelf in the Gulf of Saint Lawrence ». *Marine Ecology – Progress Series*, vol. 23, p. 57-69.

Sala, E., and N. Knowlton. 2006. «Global marine biodiversity trends». *Annual Review of Environment and Resources*, vol. 31, p. 93-122.

Saunders, G. W. 2005. «Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications». *Philosophical Transactions of the Royal Society B-Biological Sciences*, vol. 360, no 1462, p. 1879-1888.

--------. 2008. «A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose

Dilsea-Neodilsea complex and *Weeksia*». *Botany – Botanique*, vol. 86, no 7, p. 773-789.

--------. 2009. «Routine DNA barcoding of Canadian Gracilariales (Rhodophyta) reveals the invasive species *Gracilaria vermiculophylla* in British Columbia». *Molecular Ecology Resources*, vol. 9, p. 140-150.

Schander, C., and E. Willassen. 2005. «What can biological barcoding do for marine biology?». *Marine Biology Research*, vol. 1, no 1, p. 79-83.

Scheltema, R. S. 1986. «On dispersal and planktonic larvae of benthic invertebrates: an eclectic overview and summary of problems». *Bulletin of Marine Science*, vol. 39, no 2, p. 290-322.

Schmaltz, G., C. M. Somers, P. Sharma and J. S. Quinn. 2006. «Non-destructive sampling of maternal DNA from the external shell of bird eggs». *Conservation Genetics*, vol. 7, no 4, p. 543-549.

Schultz, G. A. 1969. *How to know the marine isopods*. Dubuque, IA: WMC Brown Co. Publications.

Segerstråle, S. V. 1947. «New observations on the distribution and morphology of the amphipod, *Gammarus zaddachi* Sexton, with notes on related species». *Journal of the Marine Biological Association of the United Kingdom*, vol. 27, p. 219-244.

Seifert, K. A., R. A. Samson, J. R. Dewaard, J. Houbraken, C. A. Levesque, J. M. Moncalvo, G. Louis-Seize and P. D. N. Hebert. 2007. «Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no 10, p. 3901-3906.

Serejo, C. S. 2004. «Cladistic revision of talitroidean amphipods (Crustacea, Gammaridea), with a proposal of a new classification». *Zoologica Scripta*, vol. 33, no 6, p. 551-586.

Serejo, C. S., and J. K. Lowry. 2008. «The coastal Talitridae (Amphipoda: Talitroidea) of Southern and Western Australia, with comments on *Platorchestia platensis* (Krøyer, 1845)». *Records of the Australian Museum*, vol. 60, no 2-3, p. 161-206.

Sévigny, J. M., I. A. McLaren and B. W. Frost. 1989. «Discrimination among and variation within species of *Pseudocalanus* based on the GPI locus». *Marine Biology*, vol. 102, no 3, p. 321-327.

Sévigny, J. M., L. Savard and D. G. Parsons. 2000. «Genetic characterization of the northern shrimp, *Pandalus borealis*, in the Northwest Atlantic using electrophoresis of enzymes». *Journal of Northwest Atlantic Fishery Science*, vol. 27, p. 161-175.

Shaw, J. 2006. «Palaeogeography of Atlantic Canadian continental shelves from the Last Glacial Maximum to the present, with an emphasis on Flemish Cap». *Journal of Northwest Atlantic Fishery Science*, vol. 37, p. 119-126.

Shaw, J., D. J. W. Piper, G. B. J. Fader, E. L. King, B. J. Todd, T. Bell, M. J. Batterson and D. G. E. Liverman. 2006. «A conceptual model of the deglaciation of Atlantic Canada». *Quaternary Science Reviews*, vol. 25, no 17-18, p. 2059-2081.

Shearer, T. L., and M. A. Coffroth. 2008. «Barcoding corals: limited by interspecific divergence, not intraspecific variation». *Molecular Ecology Resources*, vol. 8, no 2, p. 247-255.

Shokralla, S., G. A. C. Singer and M. Hajibabaei. 2010. «Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol». *BioTechniques*, vol. 48, no 3, p. 305-306.

Siddall, M. E., F. M. Fontanella, S. C. Watson, S. Kvist and C. Erseus. 2009. «Barcoding bamboozled by Bacteria: convergence to metazoan mitochondrial primer targets by marine microbes». *Systematic Biology*, vol. 58, no 4, p. 445-451.

Smith, M. A., E. S. Eveleigh, K. S. McCann, M. T. Merilo, P. C. McCarthy and K. I. Van Rooyen. 2011. «Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses». *PLoS ONE*, vol. 6, no 7, e14424.

Smith, P. J., S. M. McVeagh and D. Steinke. 2008. «DNA barcoding for the identification of smoked fish products». *Journal of Fish Biology*, vol. 72, no 2, p. 464-471.

So, J. J., S. Uthicke, J.-F. Hamel and A. Mercier. 2011. «Genetic population structure in a commercial marine invertebrate with long-lived lecithotrophic larvae: *Cucumaria frondosa* (Echinodermata: Holothuroidea)». *Marine Biology*, vol. 158, no 4, p. 859-870.

Song, H., J. E. Buhay, M. F. Whiting and K. A. Crandall. 2008. «Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no 36, p. 13486-13491.

Spalding, M. D., H. E. Fox, B. S. Halpern, M. A. McManus, J. Molnar, G. R. Allen, N. Davidson, Z. A. Jorge, A. L. Lombana, S. A. Lourie, K. D. Martin, E. McManus, C. A. Recchia and J. Robertson. 2007. «Marine ecoregions of the world: a bioregionalization of coastal and shelf areas». *Bioscience*, vol. 57, no 7, p. 573-583.

Squires, H. J. 1990. «Decapod Crustacea of the Atlantic coast of Canada». *Canadian Bulletin of Fisheries and Aquatic Sciences*, vol. 221, p. 1-532.

Stamatakis, A., P. Hoover and J. Rougemont. 2008. «A rapid bootstrap algorithm for the RAxML web servers». *Systematic Biology*, vol. 57, no 5, p. 758-771.

Stech, M., E. Kolvoort, M. J. J. E. Loonen, K. Vrieling and J. D. Kruijer. 2011. «Bryophyte DNA sequences from faeces of an arctic herbivore, barnacle goose (*Branta leucopsis*)». *Molecular Ecology Resources*, vol. 11, no 2, p. 404-408.

Steele, V. J., and D. H. Steele. 1972. «The biology of *Gammarus* (Crustacea, Amphipoda) in the northwestern Atlantic. V. *Gammarus oceanicus* Segerstråle». *Canadian Journal of Zoology – Revue Canadienne De Zoologie*, vol. 50, p. 801-813.

Steinke, D., T. S. Zemlak, J. A. Boutillier and P. D. N. Hebert. 2009. «DNA barcoding of Pacific Canada's fishes». *Marine Biology*, vol. 156, no 12, p. 2641-2647.

Steinke, D., T. S. Zemlak and P. D. N. Hebert. 2009. «Barcoding Nemo: DNA-based identifications for the ornamental fish trade». *PLoS ONE*, vol. 4, no 7, e6300.

Stork, N. E. 1997. «Measuring global biodiversity and its decline». In *Biodiversity II: Understanding and protecting our biological resources,* M. L. Reaka-Kudla, D. E. Wilson and E. O. Wilson, p. 41-68. Washington DC: Joseph Henry Press.

Swanson, G. A. 1984. «Dissemination of amphipods by waterfowl». *Journal of Wildlife Management*, vol. 48, no 3, p. 988-991.

Tafani, B., A. Ugolini, M. Bazzicalupe, A. Mengoni and S. Ruffo. 2004. «Phylogenetic relationships among Mediterranean sandhoppers». *Journal of Natural History*, vol. 38, p. 499-508.

Tajima, F. 1989. «Statistical method for testing the neutral mutation hypothesis by DNA polymorphism». *Genetics*, vol. 123, no 3, p. 585-595.

Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. «MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0». *Molecular Biology and Evolution*, vol. 24, no 8, p. 1596-1599.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. «MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods». *Molecular Biology and Evolution*, vol. 28, no 10, p. 2731-2739.

Tautz, D., P. Arctander, A. Minelli, R. H. Thomas and A. P. Vogler. 2003. «A plea for DNA taxonomy». *Trends in Ecology & Evolution*, vol. 18, no 2, p. 70-74.

TEEB. 2010. *Mainstreaming the economics of nature: a synthesis of the approach, conclusions and recommendations of TEEB*. www.teebweb.org.

Teske, P. R., I. Papadopoulos, G. I. Zardi, C. D. McQuaid, M. T. Edkins, C. L. Griffiths and N. P. Barker. 2007. «Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: planktonic, abbreviated and direct development». *Marine Biology,* vol. 152, no 3, p. 697-711.

Thiel, M., and L. Gutow. 2005. «The ecology of rafting in the marine environment. II. The rafting organisms and community». *Oceanography and Marine Biology – an Annual Review,* vol. 43, p. 279-418.

Thomas, J. A., J. J. Welch, R. Lanfear and L. Bromham. 2010. «A generation time effect on the rate of molecular evolution in invertebrates». *Molecular Biology and Evolution*, vol. 27, no 5, p. 1173-1180.

Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. Vanden Berghe and B. Worm. 2010. «Global patterns and predictors of marine biodiversity across taxa». *Nature*, vol. 466, no 7310, p. 1098-1107.

Ugolini, A., F. Borghini, P. Calosi, M. Bazzicalupo, G. Chelazzi and S. Focardi. 2004. «Mediterranean *Talitrus saltator* (Crustacea, Amphipoda) as a biomonitor of heavy metal contamination». *Marine Pollution Bulletin*, vol. 48, p. 526-532.

UNEP. 2010. *Marine Biodiversity Assessment and Outlook Series: Global Synthesis.* United Nations Environment Programme. www.unep.org.

Valentine, J. W., and D. Jablonski. 2010. «Origins of marine patterns of biodiversity: some correlates and applications». *Palaeontology*, vol. 53, p. 1203-1210.

Vandendriessche, S., M. Vincx and S. Degraer. 2007. «Floating seaweed and the influences of temperature, grazing and clump size on raft longevity – a microcosm study». *Journal of Experimental Marine Biology and Ecology*, vol. 343, no 1, p. 64-73.

Vargas, S. M., F. C. F. Araujo and F. R. Santos. 2009. «DNA barcoding of Brazilian sea turtles (Testudines)». *Genetics and Molecular Biology*, vol. 32, no 3, p. 608-612.

Velasco, C., A. J. Sanchez and R. Florido. 2005. «A new record of *Uhlorchestia uhleri* (Shoemaker, 1930) (Peracarida, Amphipoda, Talitridae) from the southern Gulf of Mexico». *Crustaceana*, vol. 78, p. 499-502.

Veliz, D., P. Duchesne, E. Bourget and L. Bernatchez. 2006. «Genetic evidence for kin aggregation in the intertidal acorn barnacle (*Semibalanus balanoides*)». *Molecular Ecology*, vol. 15, no 13, p. 4193-4202.

Vermeij, G. J. 1991. «Anatomy of an invasion: the trans-Arctic interchange». *Paleobiology,* vol. 17, no 3, p. 281-307S.

--------. 2005. «Invasions as expectation: a historical fact of life». In *Species invasions: insights into ecology, evolution and biogeography,* D. F. Sax, J. J. Stachowicz and S. D. Gaines, p. 315-339. Sunderland, MA: Sinauer Associates.

Vermeij, G. J., and R. K. Grosberg. 2010. «The Great Divergence: when did diversity on land exceed that in the sea?». *Integrative and Comparative Biology*, vol. 50, no 4, p. 675-682.

Vesakoski, O., J. Rautanen, V. Jormalainen and T. Ramsay. 2009. «Divergence in host use ability of a marine herbivore from two habitat types». *Journal of Evolutionary Biology*, vol. 22, no 7, p. 1545-1555.

Vrijenhoek, R. C. 2009. «Cryptic species, phenotypic plasticity, and complex life histories: Assessing deep-sea faunal diversity with molecular markers». *Deep-Sea Research Part II – Topical Studies in Oceanography*, vol. 56, no 19-20, p. 1713-1723.

Wada, S., K. Kawakami and S. Chiba. 2012. «Snails can survive passage through a bird's digestive system». *Journal of Biogeography*, vol. 39, no 1, p. 69-73.

Wang, I. J. 2011. «Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses». *Molecular Ecology*, vol. 20, no 12, p. 2480-2482.

Waples, R. S., and 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*, vol. 15, no 6, p. 1419-1439.

Ward, R. D., F. O. Costa, B. H. Holmes and D. Steinke. 2008. «DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species». *Aquatic Biology*, vol. 3, no 1, p. 71-78.

Ward, R. D., R. Hanner and P. D. N. Hebert. 2009. «The campaign to DNA barcode all fishes, FISH-BOL». *Journal of Fish Biology*, vol. 74, no 2, p. 329-356.

Ward, R. D., B. H. Holmes and T. D. O'Hara. 2008. «DNA barcoding discriminates echinoderm species». *Molecular Ecology Resources*, vol. 8, no 6, p. 1202-1211.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last and P. D. N. Hebert. 2005. «DNA barcoding Australia's fish species». *Philosophical Transactions of the Royal Society B – Biological Sciences*, vol. 360, no 1462, p. 1847-1857.

Wares, J. P., and C. W. Cunningham. 2001. «Phylogeography and historical ecology of the North Atlantic intertidal». *Evolution*, vol. 55, no 12, p. 2455-2469.

Warwick, R. M., and K. R. Clarke. 1995. «New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress». *Marine Ecology – Progress Series*, vol. 129, no 1-3, p. 301-305.

Warwick, R. M., and P. J. Somerfield. 2008. «All animals are equal, but some animals are more equal than others». *Journal of Experimental Marine Biology and Ecology*, vol. 366, no 1-2, p. 184-186.

Watts, P. C., and J. P. Thorpe. 2006. «Influence of contrasting larval developmental types upon the population genetic structure of cheilostome bryozoans». *Marine Biology*, vol. 149, no 5, p. 1093-1101.

Webb, K. E., D. K. A. Barnes, M. S. Clark and D. A. Bowden. 2006. «DNA barcoding: A molecular tool to identify Antarctic marine larvae». *Deep-Sea Research Part II – Topical Studies in Oceanography*, vol. 53, no 8-10, p. 1053-1060.

Weetman, D., A. Ruggiero, S. Mariani, P. W. Shaw, A. R. Lawler and L. Hauser. 2007. «Hierarchical population genetic structure in the commercially exploited shrimp *Crangon crangon* identified by AFLP analysis». *Marine Biology*, vol. 151, no 2, p. 565-575.

Wildish, D. J. 1969. «Studies on taxonomy, ecology and behaviour of *Orchestia gammarella* (Pallas), *O. mediterranea* A. Costa and *O. cavimana* Heller in the Medway estuary». PhD thesis, London, University of London.

--------. 1970. «Polymorphism in *Orchestia mediterranea* A. Costa». *Crustaceana*, vol. 19, p. 113-118.

--------. 1987. «Estuarine species of *Orchestia* (Crustacea: Amphipoda: Talitroidea) from Britain». *Journal of the Marine Biological Association of the United Kingdom*, vol. 67, p. 571-583.

--------. 1988. «Ecology and natural history of aquatic Talitroidea». *Canadian Journal of Zoology*, vol. 66, p. 2340-2359.

Wildish, D. J., A. Kraft, G. W. Pohle and S. E. LeCroy. 2011. «Variation in body size of two species of supralittoral talitrids (Amphipoda, Talitroidea) from the Gulf of Mexico and Atlantic coasts of North America ». *Crustaceana*, vol. 84, no. 10, p. 1243-1250.

Will, K. W., B. D. Mishler and Q. D. Wheeler. 2005. «The perils of DNA barcoding and the need for integrative taxonomy». *Systematic Biology*, vol. 54, p. 844-851.

Will, K. W., and D. Rubinoff. 2004. «Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification». *Cladistics – the International Journal of the Willi Hennig Society*, vol. 20, no 1, p. 47-55.

Williams, S. T., and N. Knowlton. 2001. «Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*». *Molecular Biology and Evolution*, vol. 18, no 8, p. 1484-1493.

Wilson, E. O. 1988. *Biodiversity*. Washington DC: National Academy Press.

Wilson, E. O. 2003. «The encyclopedia of life». *Trends in Ecology and Evolution*, vol. 18, no 2, p. 77-80.

Witt, J. D. S., P. D. N. Hebert and W. B. Morton. 1997. «*Echinogammarus ischnus*: another crustacean invader in the Laurentian Great Lakes basin». *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 54, no 2, p. 264-268.

Witt, J. D. S., D. L. Threloff and P. D. N. Hebert. 2006. «DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation». *Molecular Ecology*, vol. 15, no 10, p. 3073-3082.

Wong, E. H. K., and R. H. Hanner. 2008. «DNA barcoding detects market substitution in North American seafood». *Food Research International*, vol. 41, no 8, p. 828-837.

Worheide, G., and D. Erpenbeck. 2007. «DNA taxonomy of sponges – progress and perspectives». *Journal of the Marine Biological Association of the United Kingdom*, vol. 87, no 6, p. 1629-1633.

Worm, B., E. B. Barbier, N. Beaumont, J. E. Duffy, C. Folke, B. S. Halpern, J. B. C. Jackson, H. K. Lotze, F. Micheli, S. R. Palumbi, E. Sala, K. A. Selkoe, J. J. Stachowicz and R. Watson. 2006. «Impacts of biodiversity loss on ocean ecosystem services». *Science*, vol. 314, no 5800, p. 787-790.

Youngman, P. M. 1967. «Insular populations of the meadow vole, *Microtus pennsylvanicus*, from Northeastern North America, with descriptions of two new subspecies». *Journal of Mammalogy*, vol. 48, no 4, p. 579-588.

Zeale, M. R. K., R. K. Butlin, G. L. A. Barker, D. C. Lees and G. Jones. 2011. «Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces». *Molecular Ecology Resources*, vol. 11, no 2, p. 236-244.

Zemlak, T. S., R. D. Ward, A. D. Connell, B. H. Holmes and P. D. N. Hebert. 2009. «DNA barcoding reveals overlooked marine fishes». *Molecular Ecology Resources*, vol. 9, p. 237-242.