

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CONNECTIVITÉ CHEZ LES AMPHIPODES MARINS ARCTIQUES À  
DIFFÉRENTES ÉCHELLES SPATIALES ET TEMPORELLES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

ASTRID TEMPESTINI

DÉCEMBRE 2016



# UNIVERSITÉ DU QUÉBEC À RIMOUSKI

Service de la bibliothèque

## Avertissement

La diffusion de ce mémoire ou de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire « *Autorisation de reproduire et de diffuser un rapport, un mémoire ou une thèse* ». En signant ce formulaire, l'auteur concède à l'Université du Québec à Rimouski une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de son travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, l'auteur autorise l'Université du Québec à Rimouski à reproduire, diffuser, prêter, distribuer ou vendre des copies de son travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de la part de l'auteur à ses droits moraux ni à ses droits de propriété intellectuelle. Sauf entente contraire, l'auteur conserve la liberté de diffuser et de commercialiser ou non ce travail dont il possède un exemplaire.



## REMERCIEMENTS

Ce projet est l'aboutissement de 5 années de travail, de milliers d'heures de laboratoire, d'analyses, de lecture. Toutefois, je n'aurais pas pu accomplir ce travail et surmonter cette épreuve sans l'aide de plusieurs personnes que je tiens à remercier.

Merci aux membres mon de jury de thèse: Alisson Derry, Nicolas Bierne, Christian Nozais et France Dufresne pour leur évaluation et leurs commentaires sur ce projet.

Merci à ma directrice France Dufresne pour m'avoir donné l'opportunité et la liberté de travailler sur ce projet, mais aussi d'enseigner au premier et deuxième cycle, ce qui fut une révélation pour moi.

Merci également à mes collègues de laboratoire pour leurs encouragements, les discussions et leur bonne humeur sans lesquelles je n'aurais jamais pu avancer.

Merci à Soren Rysgaard et à l'Institut des ressources naturelles de Nuuk (Groenland) pour m'avoir accueilli chaleureusement.

Merci à Louis Fortier pour m'avoir permis de partir sur l'Amundsen et de tester mon manteau d'hiver pour la première fois.

Merci au capitaine et à l'équipage de l'Amundsen pour m'avoir ramené à bon port.

Merci à mes différents évaluateurs qui par leurs commentaires m'ont aidé tout au long de mon cheminement au doctorat.

Merci au personnel de soutien du Département de Biologie de l'UQAR : Danie, Diane, Lyne, Pauline, Louise, Annie et Mélanie pour leur précieuse aide administrative et leur bonne humeur. Merci à Jonathan pour le matériel et la discipline...Un merci particulier à Pierre Rioux pour son humour et son sarrau.

Merci à mes amis Kim, Matthieu, Hacène, Bertrand, Roland, Adriana, Mathilde, Blandine, Joannie, Pascale, Irène pour vos conseils, votre aide, votre soutien et votre compréhension. Un gros merci à mes amies de longues dates Caro et Micaela pour être toujours là quand j'en ai besoin.

Merci à mes amis du groupe d'Italien de Rimouski. I pranzi del venerdì sono sempre un momento di gioia e di scopertà.

Merci à ma famille et mes parents Nicole et Serge en particulier pour m'avoir toujours soutenu dans mes choix.

Enfin, un merci tout spécial à mon futur mari Kaven, pour son soutien indéfectible tout au long de ce doctorat.

*À ma Grand-mère Geneviève,*

## AVANT PROPOS

La présente thèse de doctorat est présentée sous la forme de six chapitres. Le premier et le sixième chapitre correspondent respectivement à l'introduction générale et à la conclusion générale de la thèse. Les chapitres II, III, IV et V sont rédigés en anglais et sont écrits sous la forme de manuscrits en préparation à soumettre à des revues scientifiques avec comités de lecture. Le chapitre II est en voie d'être soumis pour publication à la revue *PLoS One*, le chapitre III est en voie d'être soumis pour publication à la revue *Marine Biology*, le chapitre IV est en voie d'être soumis pour publication à la revue *Molecular Ecology* alors que le chapitre V est en voie d'être soumis pour publication à la revue *Polar Biology*.

Ce projet a été financé par le Conseil de recherches en sciences naturelles et génie du Canada (CRSNG) grâce à une subvention du réseau canadien sur la santé des océans octroyée à France Dufresne. A. Tempestini a également reçu deux bourses de recherche interdisciplinaire de la part du programme de formation du CRSNG en sciences environnementales nordiques EnviroNord ainsi qu'une bourse de voyage de la part du réseau canadien de la santé des océans (CHONe).



## TABLE DES MATIÈRES

LISTE DES FIGURES.....	X
LISTE DES TABLEAUX.....	XVI
LISTE DES ABRÉVIATIONS, ACRONYMES ET SIGLES.....	XVIII
RÉSUMÉ .....	XX
INTRODUCTION	
CHAPITRE I	
1.1 Définition et intérêt.....	1
1.2 Connectivité en milieu marin.....	2
1.3 Les facteurs influençant la connectivité .....	2
1.3.1 Les facteurs intrinsèques .....	3
1.3.2 Les facteurs extrinsèques .....	4
1.3.3 La vicariance .....	5
1.4 Mesures de la connectivité.....	5
1.4.1 Les méthodes directes .....	6
1.4.2 Approches indirectes .....	6
1.5.1 Le polymorphisme nucléotidique.....	13
1.5.2 Code-barres ADN.....	16
1.6 Le domaine d'étude : l'Arctique.....	18
1.7 Les Amphipodes .....	25
1.8 Les objectifs de thèse.....	25
1.7.1 Objectif 1(Chapitre II) : Connectivité à l'échelle tri-océanique chez les amphipodes.....	27
1.7.2 Objectif 2 (Chapitre III) : Phylogénie d'un genre pélagique d'amphipode .....	29
1.7.3 Objectif 3 (Chapitre IV): Structure des populations au niveau de l'archipel canadien .....	30
1.7.4 Objectif 4 (Chapitre V): Structure des populations au niveau du détroit de Béring.....	32
CHAPITRE II	
2.1 Résumé .....	34

2.2	Abstract.....	35
2.3	Introduction.....	36
2.4	Material and Methods .....	39
	2.4.1 Samples .....	39
	2.4.2 COI amplification.....	39
	2.4.3 BOLD Sequences: .....	39
	2.4.4 Trimming and Multiple alignments.....	40
	2.4.5 Distance based methods .....	40
2.5	Results .....	44
	2.5.1 Nucleotide diversity .....	44
	2.5.2 3% as species threshold.....	44
	2.5.3 16% threshold.....	47
	2.5.4 Barcode gap detection with ABGD.....	50
	2.5.5 Comparison among thresholds.....	56
	2.5.6 Species threshold.....	56
2.6	Discussion.....	59
	2.6.1 Arctic diversity.....	59
	2.6.2 Single threshold.....	61
2.7	Conclusions .....	62
2.8	Acknowledgements.....	63
<b>CHAPITRE III</b>		
3.1	Résumé .....	65
3.2	Abstract.....	66
3.3	Introduction.....	67
3.4	Material and Methods .....	71
	3.4.1 Samples collection.....	71
	3.4.2 DNA extractions.....	73
	3.4.3 PCR amplifications and alignments .....	74
	3.4.4 Phylogenetic analyses .....	76
3.5	Results .....	79

3.5.1	The Hyperiididae phylogeny.....	79
3.5.2	Amino acids phylogeny.....	80
3.5.3	Single gene phylogenies.....	85
3.5.4	Concatenated dataset.....	97
3.5.5	Species tree.....	97
3.6	Discussion.....	101
3.6.1	The Hyperiididae family.....	101
3.6.2	Cryptic <i>Themisto</i> species.....	102
3.6.3	Limit of the techniques.....	105
3.7	Conclusions.....	105
3.8	Acknowledgements.....	106
CHAPITRE IV		
4.1	Résumé.....	108
4.2	Abstract.....	109
4.3	Introduction.....	110
4.4	Material and Methods.....	113
4.4.1	Samples.....	113
4.4.2	Mitochondrial analyses.....	114
4.4.3	Rad-tags.....	115
4.5	Results.....	120
4.5.1	Mitochondrial analysis.....	120
4.5.2	Rad-tag analysis.....	132
4.6.1	Mitochondrial diversity.....	142
4.6.2	Patterns of mitochondrial genetic structure.....	143
4.6.3	Rad-tag.....	144
4.7	Conclusion.....	148
4.8	Acknowledgement.....	149
CHAPITRE V		
5.1	Résumé.....	151
5.3	Introduction.....	153

5.4	Material and Methods.....	155
5.4.1	Samples and DNA extraction.....	155
5.4.2	Mitochondrial analysis.....	155
5.4.3	RAD-seq.....	157
5.5	Results.....	161
5.5.1	Mitochondrial analysis.....	161
5.5.2	Rad-tag analysis.....	168
5.6	Discussion.....	177
5.6.1	Cryptic speciation.....	177
5.6.2	Genetic structure.....	178
5.6.3	Limit of the techniques.....	180
5.7	Conclusion.....	180
5.8	Acknowledgement.....	181
CHAPITRE VI		
CONCLUSION GÉNÉRALE.....		182
6.1	Connectivité à l'échelle tri-océanique (Chapitre II).....	184
6.1.1	Échange tri-océanique.....	184
6.1.2	Seuil pour définir une espèce.....	186
6.2	Phylogénie (Chapitre III).....	187
6.2.1	Connectivité limitée chez les espèces pélagiques.....	188
6.2.2	Limites de la phylogénie Hyperiidæ.....	190
6.3	Connectivité à l'échelle des populations (Chapitre IV et V).....	190
6.3.1	Refuge en profondeur versus zone contact.....	191
6.3.2	Les Rad-tags.....	193
6.3.3	Limites des SNPs et de la génomique.....	197
6.4	Perspectives.....	198
6.4.1	Comparaison de la connectivité chez d'autres espèces pélagique en Arctique.....	199
6.4.2	Comparaison de la connectivité entre les deux pôles.....	199
BIBLIOGRAPHIE.....		201

APPENDICE A	
.....	240
APPENDICE B	
.....	243

## LISTE DES FIGURES

Figure	Page
Figure 1.1 Modèles de structurations des populations (modifié d'après Frankham <i>et al.</i> , 2009) : (a) le modèle source-puits; (b) le modèle en îles; (c) le modèle pas-à-pas ou « stepping-stone »; (d) le modèle pas-à-pas à deux dimensions et (e) le modèle de métapopulations, où les populations éteintes sont en blanc, les populations nouvellement fondées sont en gris. Les flèches symbolisent la migration entre les populations. Les cercles noirs symbolisent les populations. ....	8
Figure 1.2 Différents scénarios de taux de migration et de différenciation (modifié de Hellberg <i>et al.</i> , 2002 et de Gagnaire <i>et al.</i> , 2015). Le taux de migration $N_e m$ dépend de la migration ( $m$ ) et de la taille effective ( $N_e$ ). Le dégradé de couleurs correspond au niveau de $F_{ST}$ attendu selon les différents valeurs de $m$ et de $N_e$ .....	9
Figure 1.3 Étapes du séquençage RAD, de la digestion du génome à la découverte des SNPs (tiré de Emerson <i>et al.</i> , 2010). (A) Le génome est digéré par une enzyme de restriction (StbfI), les sites de coupure de l'enzyme sont matérialisés en rouge. Sur chaque fragment coupé, un code-barres (en noir) est rajouté après un adaptateur. (B) Après le séquençage, tous les reads identiques vont être regroupés ensemble pour former des piles. (C) Chaque pile va être comparée deux à deux pour identifier les SNPs. (D) Les piles qui ont un nucléotide de différence, si celui-ci est retrouvé en suffisamment grande fréquence sur l'ensemble des échantillons, sera considéré comme un SNPs. (E) Sinon, il ne sera pas retenu. ....	15
Figure 1.4 Carte des différentes zones refuges et zones de contacts secondaires en Arctique (modifié d'après Hardy <i>et al.</i> , 2011). Les étoiles représentent les zones refuges. Les triangles représentent les zones de contacts secondaires.	21
Figure 1.5 Exemples de modèles de réseaux d'haplotypes et de distribution des fréquences alléliques en fonction d'un gradient Nord (N)- Sud (S), sous différents scénarios de refuges et recolonisations (modifié de Maggs <i>et al.</i> , 2008). Pour chaque modèle, un réseau de 5 ou 6 haplotypes, numérotés de 1 à 5-6, est représenté en haut, et leur fréquence dans quatre populations (a-d).	

- Dans tous les cas, l'allèle 6 correspond à un allèle privé restreint à une région particulière. (A) Le modèle nul dans lequel aucun changement dans les fréquences n'est observé en fonction de la latitude. (B) Le modèle représente un gradient de la richesse allélique qui décline de la zone refuge au Sud vers le Nord. (C) Le modèle représente le cas d'un contact secondaire entre deux populations isolées dans des refuges distincts. .... 23
- Figure 1.6 Diagramme récapitulatif de la démarche de la thèse. OTU= operational taxonomic unit; MtDNA= ADN mitochondrial; Rad-seq= Restriction site associated DNA sequencing. .... 26
- Figure 2.1 Map of the sampling location of each sequence used in this study. .... 43
- Figure 2.2 Distribution of OTUs by family and ocean. OTUs are defined based on 3% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue. .... 46
- Figure 2.3 Venn diagram of shared OTUs among oceans. OTUs are defined with a 3% threshold. .... 47
- Figure 2.4 Distribution of OTUs by family and ocean. OTUs are defined based on 16% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue. .... 49
- Figure 2.5 Venn diagram of shared OTUs among ocean. OTUs were defined with a 16% threshold. .... 50
- Figure 2.6 Distribution of the K2P distances for the entire dataset. .... 53
- Figure 2.7 Distribution of OTUs by family and ocean. OTUs are defined based on 7% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue. .... 54
- Figure 2.8 Venn diagram of shared OTUs among ocean. OTUs are defined with a 7% threshold. .... 55
- Figure 2.9 Number of OTUs identified to the family according to the species threshold used (3% light grey; 7% grey and 16% dark). .... 57
- Figure 2.10 Boxplot distribution of the COI K2P distances for the 10 selected families representative of the Amphipoda order at the intraspecies (S–white), intragenus (G–grey) and intrafamily (S–dark grey) levels. Median (central bar), position of the upper and lower quartiles (central box) and extremes of the data (dots) are represented. Black lines represent different thresholds used for discriminating species: long dash=0.16 substitution/site (Lefébure *et al.*,

- 2006); dashed line=7% of divergence (this study) and solid line=3% of divergence (Hebert *et al.*, 2003b). ..... 58
- Figure 3.1 Global distribution of *Themisto* species (modified from Bowman, 1960; Wing, 1976 and Yamada *et al.*, 2004). Each color represents the distribution of a species: in green *T. gaudichaudii*, in orange *T. japonica*, in pink *T. pacifica*, in purple *T. abyssorum*, in yellow *T. compressa* and in solid line *T. libellula*. ..... 70
- Figure 3.2 Saturation of the cytochrome oxidase I gene estimated for the family phylogeny. s and v represent transition and transversion respectively. .... 81
- Figure 3.3 Bayesian phylogeny of the cytochrome oxidase I DNA sequences under the HKY +G+I model. Numbers represent posterior probability support. .... 82
- Figure 3.4 Maximum likelihood tree of hyperiids estimated from the cytochrome oxidase I DNA sequences, under the HKY model with gamma distribution and invariant sites considered. aLTR values are shown. .... 83
- Figure 3.5 Bayesian phylogeny of hyperiids performed with amino acids sequences of the cytochrome oxidase I gene under the general reversible mitochondrial model with a gamma distribution. Numbers represent posterior probabilities. .... 84
- Figure 3.6 Saturation plots for COI (a), H3 (b), 18S (c) and 28S (d). The green triangles represent the transversion and the blue crosses represent the transition among distances. .... 88
- Figure 3.7 Maximum Likelihood phylogeny (Log-likelihood : -2398.84107) for the cytochrome oxidase I gene. aLTR supports are shown under the nodes. .... 89
- Figure 3.8 Maximum Likelihood phylogeny (Log-likelihood: -634.27815) for the Histone 3 gene. aLTR supports are shown under the node. .... 90
- Figure 3.9 Maximum Likelihood phylogeny (Log-likelihood: -3453.95) for the 18S gene. aLTR supports are shown under the nodes. .... 91
- Figure 3.10 Maximum Likelihood tree (Log likelihood=-4688.49522) for the 28S gene. aLTR supports are shown under the nodes. .... 92
- Figure 3.11 Bayesian phylogeny for the COI dataset. Posterior probabilities are shown under the nodes. .... 93
- Figure 3.12 Bayesian phylogeny for the Histone 3 dataset. Posterior probabilities are shown under the nodes. .... 94



- Figure 3.13 Bayesian phylogeny for the 18S dataset. Posterior probabilities are shown under the nodes..... 95
- Figure 3.14 Bayesian phylogeny for the 28S dataset. Posterior probabilities are shown under the nodes..... 96
- Figure 3.15 Maximum Likelihood phylogeny (Log-likelihood : -11621.65725) of the concatenated dataset. aLTR support are shown under the nodes..... 98
- Figure 3.16 Bayesian phylogeny of the concatenated dataset. Posterior probabilities are shown under the nodes..... 99
- Figure 3.17 Species-tree estimates accounting deep coalescence events. Numbers represent posterior density. .... 100
- Figure 4.1 Sampling locations of individuals used for this study. .... 119
- Figure 4.2 Haplotype network based on 491 bp of mitochondrial cytochrome c oxidase I sequences of *T. abyssorum*. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to sampling sites (Figure 4.1). Black dots correspond to missing (non-sampled) haplotypes. .... 122
- Figure 4.3 Haplotype network based on 576 bp of mitochondrial cytochrome c oxidase I sequences obtained for *T. libellula*. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype. The number of traits on a line corresponds to the number of mutations between haplotypes. .... 123
- Figure 4.4 Spatial distribution of twenty haplotypes found in *T. libellula* among sampling sites. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype. .... 124
- Figure 4.5 Bayesian Skyline Plot for *T. abyssorum* estimated under the HKY+G model (BIC=10661.94). The blackline represents the mean and the blue lines represent the confidence interval. The y-axis represents the effective population size estimate and the x-axis represents the time in years..... 130
- Figure 4.6 Bayesian skyline plot for *T. libellula* estimated under the HKY+G model (BIC=4566.09). The blackline represents the mean and the blue lines represent the confidence interval. The y-axis represent the effective population size estimate and the x-axis represents the time in years..... 131
- Figure 4.7 Bayescan test for selection on individual SNPs between the two populations of *T. abyssorum*. Red number represents SNPs potentially under

- balancing selection and green number represents SNPs potentially under diversifying selection. .... 133
- Figure 4.8 BayeScan test for selection on individual SNPs among the three populations of *T. libellula*. Red number represents SNPs potentially under balancing selection and green number represents SNPs potentially under diversifying selection. .... 134
- Figure 4.9 Discriminant analysis of Principal Component for *T. abyssorum* performed with all 442 neutral loci. .... 135
- Figure 4.10 Discriminant analysis of Principal Component for *T. libellula* performed with all 2144 neutral loci. Each color represents a cluster. .... 136
- Figure 4.11 Structure plot using the SNP dataset with two clusters for *T. abyssorum*, under the admixture model. Distinct colors represent estimated clusters, sampling sites are separated by a black line. .... 139
- Figure 4.12 Structure plot using the SNP dataset with two clusters for *T. libellula*, under the admixture model. Each color represents an estimated cluster; sampling sites are separated by a black line. .... 140
- Figure 5.1 Map of the study area of *Themisto sp.* Black circles represent sampling locations: Prince William Sound (PWS), South East Bering (SEB), North East Bering (NEB) and Chukchi Sea (CHU). .... 160
- Figure 5.2 Haplotype network based on 627 bp of mitochondrial cytochrome c oxidase I sequences obtained for *Themisto sp.* Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype. The number of traits on a line corresponds to the number of mutations between sequences. .... 163
- Figure 5.3 Spatial distribution of haplotypes among sampling sites for *Themisto sp.* Circle sizes are proportional to the number of individuals observed for that haplotype and colour codes refer to a haplotype. .... 166
- Figure 5.4 Bayesian Skyline Plots of effective population size through time obtained under the F81 model (BIC= 2161.04). Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year. .... 167
- Figure 5.5 Discriminant analysis based on the 893 loci found in both *Themisto libellula* and *Themisto sp.* .... 170

- Figure 5.6 Structure plot based on 893 loci found in *Themisto libellula* and *Themisto sp.*, under the admixture model. Distinct colors represent estimated cluster. .... 171
- Figure 5.7 BayeScan test for selection on individual SNPs among the two populations of *Themisto sp.* implemented in the BayeScan program. The red number represents SNPs potentially under balancing selection and the green number represents SNPs potentially under diversifying selection. .... 173
- Figure 5.8 Discriminant analysis of Principal Component for *Themisto sp.* performed with the 4805 loci. .... 174
- Figure 5.9 Structure plot using the SNP dataset with two clusters for *Themisto sp.*, under the admixture model. Distinct colors represent estimated clusters; sampling sites are separated by a black line. .... 175
- Figure 6.1 Diagramme récapitulatif de la thèse. OTU= operational taxonomic unit; MtDNA= ADN mitochondrial; Rad-seq= Restriction site associated DNA sequencing. .... 183
- Figure A.1 Rarefaction curve showing the number of OTUs found in each ocean according the the number of sequence sampled. (A) at 3% threshold. .... 240
- Figure A.2 Rarefaction curve showing the number of OTUs found in each ocean according the the number of sequence sampled (B) at 7% threshold. .... 241
- Figure A.3 Rarefaction curve showing the number of OTUs found in each ocean according the the number of sequence sampled (C) at 16% threshold. .... 242
- Figure B.1 Bayesian Skyline Plots of effective population size through time obtained using a 27%/My rate for *T. abyssorum*. Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year. .... 243
- Figure B.2 Bayesian Skyline Plots of effective population size through time obtained using a 27%/My rate for *T. libellula*. Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year. .... 244

## LISTE DES TABLEAUX

Tableau	Page
Tableau 2.1 Species found in the three oceans. Species are defined with a 3% threshold.....	45
Tableau 2.2 Species found in the three oceans. Species are defined with a 16% threshold.....	48
Tableau 2.3 Species found in the three oceans, Species are defined with a 7% threshold.....	52
Tableau 3.1 List of species used for this study. Coordinates are expressed in decimal degree (DD).....	72
Tableau 3.2 List of PCR settings used. ....	75
Tableau 3.3 Estimate of divergence between <i>Themisto</i> species based on cytochrome oxidase I DNA sequences. Percentages of divergence among <i>Themisto</i> species are shown.....	87
Tableau 4.1 Number of individuals and their geographic locations used for this study. N= number of individuals used for this study, mt= mitochondrial. Geographic coordinates are in decimal degree. In parenthesis number of individuals used for the analysis. ....	113
Tableau 4.2 Results of the mitochondrial analysis of the cytochrome c oxidase I (COI) for each <i>Themisto</i> species. N= number of individuals, H= number of haplotypes, h= Haplotypic diversity, $\pi$ =nucleotide diversity. Standard deviation (sd) are in parenthesis. ....	121
Tableau 4.3 Analysis of Molecular Variance AMOVA for each <i>Themisto</i> species. * significant at $p < 0.05$ . ....	126
Tableau 4.4 Pairwise $\phi_{ST}$ between sampling locations for <i>T. abyssorum</i> . * significant at $p < 0.05$ ....	127
Tableau 4.5 Pairwise $\phi_{ST}$ between sampling location for <i>T. libellula</i> . * significant at $p < 0.05$ . ....	128

Tableau 4.6	Results of Tajima's D, Fu's Fs neutrality tests. * significant at $p < 0.05$ . 129	
Tableau 4.7	Number of putative loci retained following each filtering step. ....	132
Tableau 4.8	Membership of each individual according the DAPC.....	135
Tableau 4.9	Membership of individual after the DAPC. ....	136
Tableau 4.10	Evanno table for each <i>Themisto</i> species. In bold: the best K according to the Evanno method. ....	138
Tableau 4.11	Analysis of Molecular Variance (AMOVA) for each <i>Themisto</i> species. **= significant value $p < 0.05$ .....	141
Tableau 5.1	Coordinates of the sampling sites, and number of individuals used in this study for each site. Geographic coordinates are in decimal degree. ....	156
Tableau 5.2	Results of the mitochondrial analysis of cytochrome oxidase I (COI) for <i>Themisto sp.</i> N= number of individuals, H= number of haplotypes, h= haplotypic diversity, $\pi$ =nucleotide diversity. Standard deviations (sd) are in parentheses.....	162
Tableau 5.3	Pairwise $\phi_{ST}$ between sampling locations for each <i>Themisto sp.</i> ** significant at $p < 0.001$ . ....	164
Tableau 5.4	Analysis of Molecular Variance on mitochondrial dataset of <i>Themisto</i> <i>sp.</i> ** significant at $p < 0.001$ .....	164
Tableau 5.5	Tajima's D, Fu's Fs neutrality tests calculated on the 627 bp of COI. * significant at $p < 0.05$ . ....	165
Tableau 5.6	Number of putative loci retained following each filtering step. ....	168
Tableau 5.7	Evanno table for <i>Themisto sp.</i> In bold: the best K according to the Evanno method. ....	172
Tableau 5.8	Analysis of Molecular Variance (AMOVA) for <i>Themisto sp.</i> **= significant value $p < 0.05$ .....	176

## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ADN: acide désoxyribonucléique

aLTR : approximate likelihood ratio

BI: bayesian inference

bp : Base pair

CHU: Chukchi Sea

COI: cytochrome oxydase I

DAPC: discriminant analysis of Principal Components

DLP: durée pélagique larvaire

H3: histone 3

Ma : million d'années

ML: maximum likelihood

mtDNA: ADN mitochondrial

My; million years

NEB: Northeast Bering

OTU: operationnal taxonomic unit

PP: posterior probability

PWS: Prince William Sound

rDNA: acide désoxyribonucléique ribosomal

SEB: Southeast Bering

SNP: single nucleotide polymorphism





## RÉSUMÉ

La connectivité, ou les échanges d'individus entre les populations, est un thème important en écologie marine, car elle permet aux espèces de se maintenir et d'évoluer. Les études sur la connectivité marine montrent des inégalités dans les zones d'études, dans les échelles spatiales et les organismes. La biodiversité de l'océan Arctique est encore mal connue. Cette biodiversité marine arctique est le résultat de l'histoire complexe de ce milieu et des échanges avec les océans Pacifique et Atlantique limitrophes. Les outils génétiques ont souvent permis de caractériser non seulement une biodiversité cryptique mais aussi les échanges en milieux marins. Depuis quelques années, l'apparition des nouvelles techniques de séquençage a permis de caractériser le polymorphisme nucléotidique (SNP) réparti sur l'ensemble du génome chez les organismes non-modèles. Ces nouvelles techniques de séquençage tel que le séquençage RAD, ont permis de mettre en évidence des structures de population qui n'étaient pas détectées avec d'autres marqueurs. L'objectif général de ma thèse est de documenter la connectivité en milieu marin arctique à différentes échelles spatiales et temporelles afin de détecter la présence éventuelle de barrière à la dispersion, en utilisant comme modèles les amphipodes. Afin de répondre à cet objectif, chacun des quatre chapitres documente un aspect de la connectivité. Tout d'abord le chapitre II rapporte la contribution de chacun des océans limitrophes de l'Arctique dans la composition de la biodiversité marine. Comme les espèces d'amphipodes sont difficiles à caractériser, nous avons dans premier temps utilisé les 2557 séquences de code-barres ADN provenant de notre échantillonnage et d'une base de données que nous avons regroupés en espèces potentielles. Puis dans un second temps, nous avons comparé les compositions en espèces de chaque océan Arctique, Atlantique et Pacifique. Nous avons trouvé que les océans Arctique et Atlantique partageaient plus d'espèces tandis que l'océan Pacifique semblait plus distinct; suggérant une connectivité plus limitée entre l'océan Pacifique et l'Arctique. Ensuite, le chapitre III s'est intéressé aux relations phylogénétiques sur un genre particulier des amphipodes : les *Themisto*. La famille des Hyperiididae dont les *Themisto* font partie, sont une famille exclusivement pélagique. Leur habitat dans la colonne d'eau suggère de très fort potentiel de dispersion, pourtant la présence de nombreuses espèces dans cette famille indique des évènements d'isolement. En utilisant quatre gènes mitochondriaux et nucléaires de plusieurs espèces, nous avons reconstruit par les méthodes bayésiennes et de maximum de vraisemblance la phylogénie du genre des *Themisto* de l'hémisphère Nord. Nous avons mis en évidence au niveau du détroit de Béring la présence d'une espèce cryptique *Themisto sp.*, confondus auparavant avec l'espèce Arctique *T.*

*libellula* et celle du Pacifique *T. pacifica*. Enfin les chapitres IV et V ont documenté la connectivité à l'échelle des populations chez trois espèces de *Themisto* de l'Arctique. Nous avons pour cela caractérisé la diversité génétique avec les marqueurs mitochondriaux chez 70, 123 et 33 individus de *Themisto abyssorum*, *T. libellula* et *Themisto sp.* respectivement. L'étude des réseaux mitochondriaux suggère la présence d'un potentiel refuge en profondeur chez *T.abyssorum* et la présence de deux groupes distincts chez *Themisto sp.* L'utilisation du séquençage RAD sur 158, 204 et 33 individus de *T. abyssorum*, *T.libellula* et *Themisto sp.* nous a permis de caractériser plusieurs centaines de SNPs pour chacune de ces espèces. L'analyse des SNP chez chacune de ces espèces a mis en évidence la présence de deux groupes distincts. Chez *T. abyssorum* chacun de ces deux groupes semblait suggérer la présence d'une structuration verticale. A l'inverse, chez *T. libellula* les deux groupes semblaient correspondre à une séparation écologique. Enfin, chez *Themisto sp.* les deux groupes correspondaient aux groupes identifiés avec les marqueurs mitochondriaux, et indiquent la présence d'échanges très restreints entre le Golfe de l'Alaska et le détroit de Béring. Dans l'ensemble, nos résultats montrent que la connectivité chez les amphipodes marins arctiques est limitée à la fois entre les océans mais aussi à l'intérieur d'un océan même chez des espèces qui passent toute leur vie dans la colonne d'eau. Cela suggère que la connectivité chez les amphipodes marins résulte de l'interaction entre des processus historiques mais aussi leur adaptation à l'environnement qui limite les échanges à l'intérieur de leur population.

**Mots-clés : Connectivité; Marin; Arctique; Amphipodes; Génomique; *Themisto*; SNP; Phylogénie; Population.**

## INTRODUCTION

### CHAPITRE I

#### 1.1 Définition et intérêt

À l'intérieur de l'aire de répartition d'une espèce, les individus ne peuvent pas toujours s'unir aléatoirement du fait de contraintes géographiques ou environnementales. Les unités à l'intérieur desquelles les individus s'unissent le plus souvent sont appelées des populations (Waples et Gaggiotti, 2006). Au cours du temps, ces populations vont évoluer différemment et vont progressivement se différencier génétiquement les unes des autres (Hey et Machado, 2003), conduisant à une structuration spatiale des populations. Le degré par lequel les populations au travers de l'aire de répartition d'une espèce échangent des individus, jeunes ou adultes, est appelée connectivité (Palumbi, 2003; Cowen et Sponaugle, 2009; Kool *et al.*, 2013). La connectivité englobe donc un aspect démographique et un aspect génétique. La dispersion des individus aura un impact sur la population de départ (diminution du nombre d'individus) et aussi sur celle d'arrivée (augmentation du nombre d'individus) affectant ainsi la démographie des populations. En d'autres termes, la connectivité démographique quantifie l'impact des mouvements des individus sur la croissance démographique de la population (Lowe et Allendorf, 2010). Elle permet d'assurer la stabilité des populations en fournissant des immigrants qui compensent les pertes locales ou en favorisant la colonisation des parcelles inoccupées (Cowen et Sponaugle, 2009; Lowe et Allendorf, 2010). Quand les individus qui dispersent se reproduisent dans une population, on parle de connectivité génétique. La connectivité génétique se réfère donc au degré d'influence du flux génétique sur les processus

évolutifs au sein des populations (Hedgecok *et al.*, 2007; Thomas et Bell, 2010). Comprendre la connectivité est importante car elle permet de déterminer la résilience d'une espèce et son évolution en prévenant les extinctions locales et en favorisant les échanges génétiques (Kinlan et Gaines, 2003).

## 1.2 Connectivité en milieu marin

En milieu marin, la connectivité a longtemps été perçue comme très importante du fait de l'absence de barrière évidente à l'intérieur des océans qui limite la dispersion des individus et la présence de larves (Palumbi, 1994; Levin, 2006; Bohonak, 1999; Hellberg, 2009). Certains exemples de connectivité très fortes entre les pôles ont été rapportés chez des espèces de foraminifères (Darling *et al.*, 2000) ou au large du Japon et de la côte ouest africaine (Tsuchiya *et al.*, 2009), ce qui suggère la présence de forts échanges sur de grandes distances. Plusieurs études ont démontré que des processus plus complexes étaient à l'œuvre tels que la présence de discontinuités dans l'habitat ou des facteurs intrinsèques pouvant engendrer une connectivité restreinte en milieu océanique (revues dans Peijnenburg et Goetze, 2013). Il est donc important de mieux connaître la connectivité et la dispersion en milieu marin pour améliorer l'élaboration des aires marines protégées (Palumbi, 2003), et plus généralement la conservation et le maintien de la biodiversité (Weersing et Toonen, 2007). L'estimation de la connectivité marine s'avère complexe de par la présence d'un stade larvaire chez plusieurs organismes marins (Levin, 2006; Weersing et Toonen, 2007), la difficulté de délimiter des populations (Kool *et al.*, 2013), l'échelle spatiale à laquelle la connectivité se passe (Cowen *et al.*, 2006; 2007; Selkoe et Toonen, 2011).

## 1.3 Les facteurs influençant la connectivité

Les organismes peuvent disperser de manière active ou passive lorsqu'ils dépendent d'un agent externe pour le transport. Chez les disperseurs actifs, la dispersion va dépendre surtout de leurs propres limites à se disperser (De Bie *et al.*, 2012). À

l'inverse chez les disperseurs passifs, la capacité de dispersion va dépendre du vecteur de dispersion et de l'environnement dans lequel ils évoluent (De Bie *et al.*, 2012). Beaucoup d'espèces qui dispersent passivement vont utiliser des vecteurs comme les plastiques, des animaux ou les eaux de ballasts ce qui rend leur dispersion difficile à estimer (Thiel et Haye, 2006; Winston, 2012; Chan *et al.*, 2013). Plus de 70% des organismes marins possèdent une phase larvaire planctonique surtout dédiée à la dispersion (Pinet, 2009). De par leurs petites tailles et leur faible capacité de nage, ces larves vont se disperser passivement en étant transportées par l'intermédiaire des courants, la distance de leur dispersion étant déterminée par la durée pélagique de leur stade larvaire (DPL- Cowen *et al.*, 2007). Il existe une corrélation entre la DPL et la distance de dispersion. Plus la DPL est petite (de l'ordre de quelques heures à quelques jours) plus la distance de dispersion est faible (Shanks *et al.*, 2003; Bradbury *et al.*, 2008). Par exemple, les échinodermes qui ont en moyenne une DPL de plus de 40 jours ont une plus faible structuration des populations et donc une plus grande capacité de dispersion que les bryozoaires dont la DPL est inférieure à 20 jours (Bradbury *et al.*, 2008). Cette tendance est toutefois à moduler. Shanks *et al.*, (2003; 2009) soulignent que la DPL peut être un mauvais indicateur de la dispersion d'un organisme car d'autres facteurs comme le comportement peuvent aussi l'influencer. En milieu marin, beaucoup d'études sur la connectivité ont été effectuées sur des espèces à développement larvaire et peu d'études existent sur les espèces à développement direct aplaniques (Goetze, 2005; 2010; Peijnenburg *et al.*, 2006; Werning et Toonen, 2009; Blanco-Bercial *et al.*, 2011; Winston, 2012; Blanco-Bercial et Bucklin, 2016).

### 1.3.1 Les facteurs intrinsèques

Le comportement des larves influence aussi la dispersion des organismes marins (Folt et Burns, 1999; Woodson et McManus, 2007). Au moins quatre comportements ont

été impliqués dans la structuration spatiale des organismes marins (Folt et Burns, 1999) : la migration verticale, l'évitement des prédateurs ou la recherche de nourriture (Woodson et McManus, 2007). Par exemple, les larves du crabe *Carcinus maenas* sont capables de rester à proximité des côtes car leur migration verticale leur permet de se maintenir dans les courants qui les ramènent vers celles-ci (Marta-Almeida *et al.*, 2006). Souvent, les juvéniles d'espèces d'invertébrés marins se retrouvent proches de la surface pour éviter le cannibalisme par les adultes et sont aussi associés à des niche différentes de ces derniers (Węśławski *et al.*, 2010), contribuant à la structure génétique complexe souvent observée chez les espèces marines.

### 1.3.2 Les facteurs extrinsèques

Les facteurs abiotiques jouent un rôle important dans l'isolement et l'orientation des échanges en milieu marin (Weersing et Toonen, 2009). La présence de courants, de gyres, ou de thermoclines va structurer spatialement les populations de manière moins linéaire que ce que l'on pouvait supposer (Norris, 2000; White *et al.*, 2010). Même à des échelles spatiales petites de l'ordre de quelques dizaines de kilomètres et en fonction des saisons, les courants et les tourbillons marins peuvent moduler la connectivité des populations par exemple en empêchant la mixité entre différents pools de larves de différentes localités, contribuant ainsi à complexifier la structure génétique des populations (Mitarai *et al.*, 2009; Selkoe *et al.*, 2010). Sur des échelles spatiales plus grandes (>300km), des corrélations entre la présence d'une forte structuration génétique c'est-à-dire une forte différenciation génétique entre des populations et des facteurs océanographiques ou environnementaux ont été trouvées chez de nombreuses espèces marines (Goetze, 2005; Selkoe *et al.*, 2010; Havermans *et al.*, 2013; Blanco-Bercial et Bucklin, 2016). Par exemple, les courants au large de Point Conception en Californie, sont probablement à l'origine de la structuration de plusieurs espèces possédant un stade planctonique (Pelc *et al.*, 2009). Néanmoins, les études manquent à une échelle spatiale plus petite (<300km) (Selkoe *et al.*, 2010).

### 1.3.3 La vicariance

Les processus de vicariance (phénomènes qui divisent une population en sous-groupes à cause de phénomènes géologiques tel que la tectonique ou la diminution du niveau de la mer) ont aussi influencé la structuration spatiale de beaucoup d'espèces marines (Norris, 2000; Heads, 2005). Par exemple, les premières fermetures du détroit de Panama vers la fin du Mésozoïque (66 Ma) ou au début du Cénozoïque (65 Ma) ont contribué à l'isolement des populations marines du Pacifique et de l'Atlantique, conduisant à leur spéciation (Knowlton *et al.*, 1993; revu dans Heads, 2005). Plus récemment, les oscillations climatiques du Quaternaire ont aussi contribué à l'isolement et à la divergence entre les populations marines. En particulier, le dernier maximum glaciaire survenu il y a 12000-24000 ans, pendant lequel la couverture de glace formait une barrière physique entre les populations marines, a contribué à la diminution du niveau de la mer de 135 m en dessous du niveau actuel (Mix *et al.*, 2001). La présence de cette couverture de glace a rendu beaucoup d'habitats inappropriés pour plusieurs espèces intertidales et du littoral, aussi bien dans les régions tempérés que polaires, forçant leur distribution à se déplacer plus au sud (Hewitt, 1996; Grosberg et Cunningham 2001). Plusieurs études ont documenté les conséquences de la dernière glaciation, survenue 10-12000 ans auparavant, sur les populations marines (Marko, 2004; Maggs *et al.*, 2008; Hardy et al. 2011; Coyer *et al.*, 2011). La présence de refuges où des populations ont pu être isolées et survivre a contribué à augmenter la structure génétique des populations (Hewitt, 1996; Hardy *et al.*, 2011).

## 1.4 Mesures de la connectivité

#### 1.4.1 Les méthodes directes

Les méthodes directes de capture-marquage-recapture peuvent être utilisées pour suivre les déplacements des organismes marins (Kool *et al.*, 2013). Ces méthodes consistent à marquer des individus juvéniles, les relâcher puis les re-capturer à un endroit particulier, mais sont difficiles à mettre en œuvre sur de petits organismes ou sur de très grandes échelles géographiques (Berry *et al.*, 2004; Broque et Petit, 2009). Le recours aux méthodes génétiques se veut une alternative incontournable pour l'étude de la dispersion en utilisant les flux de gènes qui sont la conséquence de la dispersion des individus qui ont réussi à se reproduire dans leur population d'accueil. (Hedgecock *et al.*, 2007; Lowe et Allendorf 2010). L'estimation des flux de gène peut se faire par deux grandes approches statistiques : les approches indirectes qui utilisent la distribution spatiale de la variation génétique entre les populations par rapport à un modèle théorique et les approches directes, plus récentes, comme les assignations qui vont permettre la détection des événements de migration (Berry *et al.*, 2004; Hedgecock *et al.*, 2007).

#### 1.4.2 Approches indirectes

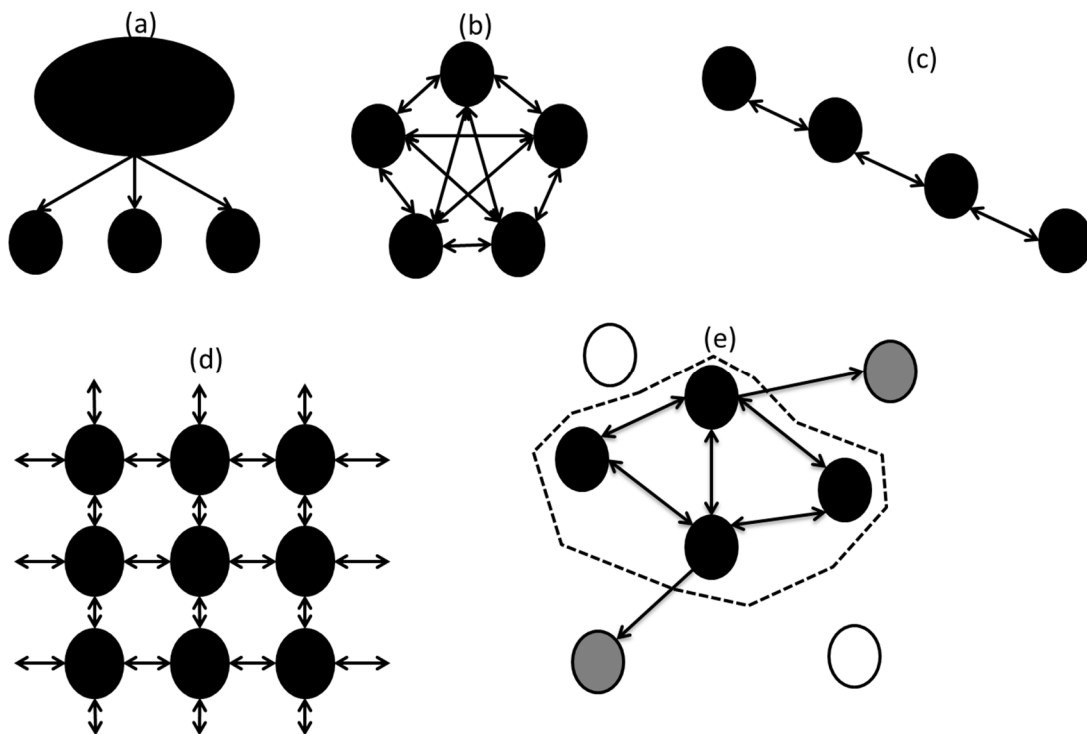
La différenciation génétique entre des populations sert généralement à estimer les flux de gènes en se basant sur un modèle théorique (Bohonak et Roderick, 2001). Lorsqu'un équilibre est atteint entre la perte de la variation à cause de la dérive et l'arrivée de nouveaux allèles par la migration, selon un modèle théorique, l'indice de différenciation  $F_{ST}$  peut permettre d'estimer le taux de migration  $N_e m$  par la relation suivante:

$$N_e m = (1 - F_{ST}) / (4 F_{ST})$$

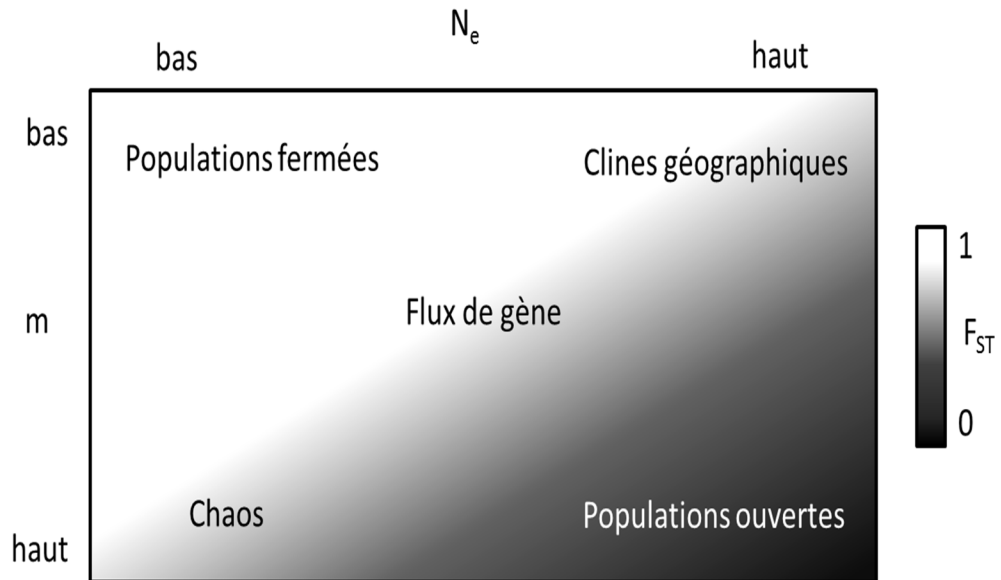
où  $N_e$  est la taille effective de la population locale et  $m$  est la proportion de migrants entrant dans la population à chaque génération. La valeur du  $F_{ST}$  augmente du minimum théorique de 0 (synonyme de forts flux de gènes et d'une absence de



structure génétique détectable) à un maximum de 1 (impliquant une absence de flux de gènes) (Grosberg et Cunningham, 2001; Hedgecock *et al.*, 2007). Cette relation suppose que les populations se conforment à un modèle théorique en îles où les populations de même taille ont la même probabilité d'échanger des migrants ce qui est assez peu réaliste (Whitlock et McCauley, 1999; Grosberg et Cunningham, 2001; Neigel, 2002; Hedgecock *et al.*, 2007; Lowe et Allendorf, 2010). Malgré cela, le  $F_{ST}$  reste utilisé comme une mesure de divergence relative entre les populations (Bohonak, et Roderick, 2001; Neigel, 2002; Marko et Hart, 2011). D'autres modèles de migration ont été développés (Figure 1.1). Le modèle source-puits, dans lequel la population source fournit les migrants aux populations puits (Figure 1.1a). Le modèle de migration pas à pas (« stepping-stone »), dans lequel des populations adjacentes vont échanger plus de migrants que des populations éloignées (Hellberg *et al.*, 2002) (Figure 1.1c et 1.1d). Ce modèle semble refléter davantage la réalité de beaucoup d'organismes marins côtiers (Hellberg *et al.*, 2002). Le modèle de métapopulation, quant à lui, permet d'intégrer une dynamique plus complexe des populations (Hanski et Gaggiotti, 2004) (Figure 1.1e). En effet, la taille des populations peut varier au cours du temps, allant même jusqu'à disparaître, et des migrants provenant de populations voisines peuvent recréer une nouvelle population en occupant un territoire ainsi libéré. Le principal avantage d'utiliser le  $F_{ST}$  pour estimer la dispersion est qu'elle ne nécessite que relativement peu d'individus (Gagnaire *et al.*, 2015).



**Figure 1.1** Modèles de structurations des populations (modifié d'après Frankham *et al.*, 2009) : (a) le modèle source-puits; (b) le modèle en îles; (c) le modèle pas-à-pas ou « stepping-stone »; (d) le modèle pas-à-pas à deux dimensions et (e) le modèle de métapopulations, où les populations éteintes sont en blanc, les populations nouvellement fondées sont en gris. Les flèches symbolisent la migration entre les populations. Les cercles noirs symbolisent les populations.



**Figure 1.2** Différents scénarios de taux de migration et de différenciation (modifié de Hellberg *et al.*, 2002 et de Gagnaire *et al.*, 2015). Le taux de migration  $N_e m$  dépend de la migration ( $m$ ) et de la taille effective ( $N_e$ ). Le dégradé de couleurs correspond au niveau de  $F_{ST}$  attendu selon les différents valeurs de  $m$  et de  $N_e$ .

Chez les organismes marins, cinq patrons de structures sont observés (Hellberg *et al.*, 2002)(Figure 1.2): allant de complètement absents (forte connectivité-  $F_{ST}$  nul) jusqu'à complètement isolés (connectivité nulle et  $F_{ST}$  fort) en passant par des structures intermédiaires avec des niveaux de  $F_{ST}$  intermédiaires. Toutefois, il est difficile de généraliser les mesures de  $F_{ST}$  car il existe beaucoup de variations dans les données génétiques utilisées pour les générer et les valeurs de  $F_{ST}$  obtenues dépendent du type de marqueurs utilisés (Hedgecock *et al.*, 2007). La corrélation entre les paires de  $F_{ST}$  et la distance géographique peut servir à détecter l'isolation par la distance, c'est-à-dire la corrélation entre les flux de gènes et la distance géographique (Levin, 2006; Grosberg et Cunningham, 2001). En fait, la structuration spatiale des espèces

marines est rarement concordante avec ces simples modèles de migration, et la différenciation observée entre les populations est peu associée avec une tendance géographique ou une stabilité temporelle (e.g. le chaos génétique- Johnson et Black, 1982; Selkoe *et al.*, 2010). La valeur du  $F_{ST}$  reflète l'ensemble de la migration passée et contemporaine, rendant les deux impossibles à distinguer par cet indice (Benzie, 1999). Cette approche a souvent été utilisée pour estimer la connectivité en milieu marin chez les algues (Coyer *et al.*, 2011), les crustacés (Thomas et Bell, 2010). Cependant, utiliser les approches indirectes pour estimer la connectivité n'est pas sans inconvénients. Les modèles utilisés dans l'estimation du  $F_{ST}$  repose sur certaines prémisses qui sont souvent peu respectées. Premièrement, l'inférence des flux de gènes suppose la neutralité des marqueurs utilisés. La sélection tend à influencer les  $F_{ST}$  et si elle n'est pas prise en compte peut biaiser les taux de migration (Whitelock et McCauley, 1999). Deuxièmement, ces modèles supposent que les populations sont à l'équilibre mutation/ dérive ce qui est très rarement le cas chez les populations marines (Whitelock et McCauley, 1999, Gagnaire *et al.*, 2015). Chez les populations avec de très larges effectifs comme les populations de nombreux organismes marins, la dérive génétique n'a que peu d'influence comparativement à la sélection ce qui peut conduire au non respect de cette prémisse (Peijnenburg et Goetze, 2013, Gagnaire *et al.*, 2015). Les modèles impliquent également des échanges symétriques entre populations, ce qui n'est pas souvent le cas chez les espèces marines (Hellberg, 2009). De plus, la grande taille des populations marines résulte souvent en de très petits estimés de  $F_{ST}$  avec de très grands intervalles de confiance (Waples *et al.*, 2008), d'où la difficulté de faire ressortir des différences significatives chez ces espèces marines lorsque le nombre de marqueurs moléculaires est peu élevé.

### 1.4.3 Approches directes

Contrairement aux approches basées sur le  $F_{ST}$ , la détection d'événements de migration récent à travers des analyses de parenté ou d'assignation ont l'avantage de ne pas reposer sur un modèle théorique (revu dans Manel *et al.*, 2005; Hedgecock *et al.*, 2007; Marko et Hart, 2011). Ces méthodes permettent de visualiser les mouvements des individus qui peuvent ou non se reproduire plutôt que les flux de gènes qui eux impliquent nécessairement la reproduction de l'individu migrant dans la population d'accueil (Gagnaire *et al.*, 2015). Les approches directes ou méthodes d'assignation peuvent se subdiviser en deux catégories : les méthodes de classification et les méthodes de regroupement (Manel *et al.*, 2005).

#### 1.4.3.1 Méthodes de classification

Dans les méthodes de classification qui contiennent les analyses de parenté ou les analyses d'admixture, les individus sont assignés à des catégories prédéfinies. Chez un espèce de corail *Seriatopora hystrix*, les analyses d'assignation ont permis de mettre en évidence que la migration de plusieurs individus provenant d'un site non perturbé permettaient de recoloniser un site perturbé (Underwood *et al.*, 2007). En appliquant ces techniques au poisson clown (*Amphiprion polymnus*) de Papouasie Nouvelle Guinée, les auteurs ont pu identifier les juvéniles à leur population d'origine (Saenz-Agudelo *et al.*, 2009). Par contre, pour être efficaces ces approches nécessitent beaucoup de données, une bonne connaissance de la répartition de l'espèce et un bon échantillonnage pour pouvoir faire un bon estimé (Waples et Gaggiotti, 2006; Gagnaire *et al.*, 2015). Pour les analyses de parenté, une contrainte supplémentaire se rajoute à savoir un échantillonnage exhaustif afin de pouvoir retracer les parents potentiels. Ces approches nécessitent donc un très grand échantillonnage à une échelle géographique pertinente limitant ainsi leur utilisation dans le milieu marin aux organismes dont les distributions sont déjà relativement bien connues (Gagnaire *et al.*, 2015). Les plupart des espèces marines ont des tailles

effectives très grandes de l'ordre de centaines de milliers voire de millions d'individus (Hayes *et al.*, 2005; Peijnenburg et Goetze, 2013), rendant ces méthodes difficilement applicables (Gagnaire *et al.*, 2015). Même si les méthodes génétiques présentent certains avantages dans la rapidité de mise en œuvre (Berry *et al.*, 2004), les méthodes d'assignation requièrent un grand nombre de marqueurs par individu qui peuvent être parfois difficile à acquérir, notamment chez des organismes non-modèles (Gagnaire *et al.*, 2015).

#### 1.4.3.2 Les méthodes de regroupement

Les méthodes de regroupement peuvent détecter des migrants sans nécessairement avoir besoin d'avoir leur population d'origine, demandant donc moins d'effort d'échantillonnage. La première étape avant de caractériser les échanges entre les populations est de délimiter celles-ci (Waples et Gaggiotti, 2006). Les méthodes de regroupement vont estimer les populations mais aussi les migrants (Pritchard *et al.*, 2000). Ces méthodes d'assignation bayésienne se basent sur les génotypes des individus pour tout d'abord estimer le nombre de population,  $k$ , qui correspond le mieux aux données, puis assigner chaque individu à sa population potentielle (Pearse et Crandall, 2004; Berry *et al.*, 2004; Manel *et al.*, 2005). Ainsi, ces méthodes ne nécessitent pas de connaître au préalable le nombre de populations, puisqu'une « simple lecture » du génotype d'un individu permet de connaître le nombre de sous-populations. Elles permettent d'identifier et de tester la présence de panmixie, c'est-à-dire si tous les échantillons proviennent d'une seule et même population, ou, dans le cas contraire, si une structuration existe (Manel *et al.*, 2005). Berry *et al.*, (2004) ont comparé l'efficacité de mesurer la dispersion avec cette approche génétique versus une approche directe de capture-marquage-recapture chez un lézard. Très peu de différences dans les estimés de dispersion ont été mis en évidence, démontrant que cette approche génétique était tout aussi fiable que les estimés obtenus par une mesure directe. Toutefois, leur efficacité à détecter des migrants diminue avec la différenciation génétique et la présence de structuration

spatiale (Waples et Gaggiotti, 2006). Ces méthodes sont de plus en plus utilisées dans l'estimation de la connectivité marine comme chez le maquereau de l'Atlantique *Scomber scombrus* L. (Rodriguez-Ezpelata *et al.*, 2016), le homard américain *Homarus americanus* (Benestan *et al.*, 2015) ou le copépode *Centropages typicus* (Blanco-Bercial et Bucklin, 2016).

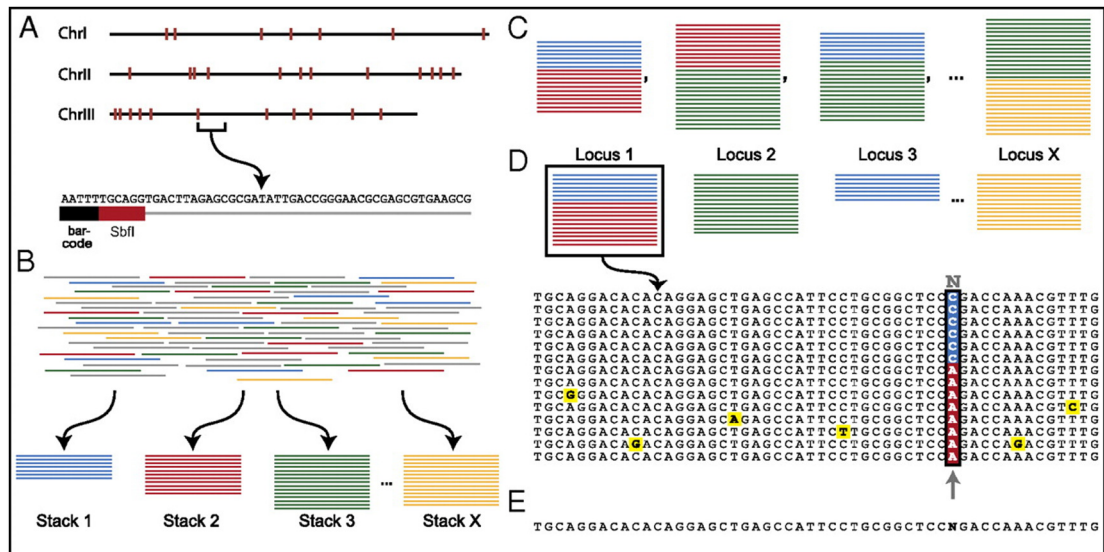
## 1.5 Estimation du flux génique à l'aide de marqueurs moléculaires

### 1.5.1 Le polymorphisme nucléotidique

L'ADN mitochondrial (ADNmt) a longtemps été un marqueur de choix pour l'estimation du flux génique (Avisé *et al.*, 1987; Beheregaray, 2008; Hickerson *et al.*, 2010). Ses particularités, comme l'absence de recombinaison, sa supposée neutralité, et sa plus petite taille effective comparée à celle du génome nucléaire permettent d'étudier les facteurs qui influencent la connectivité sur une grande échelle de temps, allant de la structure des populations à la phylogénie où les populations sont devenues des espèces à part entière (Ballard et Whitelock, 2004). Jusqu'à tout récemment, les études se basant sur des marqueurs nucléaires se limitaient à une dizaine de locus répartis dans le génome (Reitzel *et al.*, 2013). Ces dernières années, le séquençage de génomes entiers a permis l'identification du polymorphisme nucléotidique (Single Nucleotide Polymorphisme (SNP)) permettant de comparer les individus à des milliers de positions génomiques (Vignal *et al.*, 2002; Morin *et al.*, 2004; Kool *et al.*, 2013). Un SNP représente un seul changement de base dans une séquence d'ADN, avec habituellement deux alternatives possibles à un locus (Vignal *et al.*, 2002). Jusqu'à récemment, caractériser les SNPs représentait un important défi (Brumfield *et al.*, 2003; Morin *et al.*, 2004). Avec l'apparition et l'avancée des techniques de séquençage haut-débit (Metzker, 2010), l'analyse approfondie du polymorphisme nucléotidique devient de plus en plus abordable même pour les organismes non-modèles (Helyar *et al.*, 2011). La méthode des RAD-tags, ou séquençage RAD, utilise les enzymes de restriction associées au séquençage, ce qui permet en même

temps d'identifier et de génotyper plusieurs milliers de marqueurs distribués au travers du génome (Davey et Blaxter, 2011; Davey *et al.*, 2011). Brièvement, cette technique repose sur l'utilisation d'enzymes de restriction, qui vont couper le génome à différents endroits (Figure 1.3). Chacun de ces fragments va subir une transformation consistant en l'ajout d'un identifiant propre à chaque individu et un adaptateur spécifique de la plateforme de séquençage. Ensuite, ces fragments modifiés de chacun des individus vont être mélangés et coupés; seuls les fragments d'une certaine taille seront retenus pour le séquençage (Davey et Blaxter, 2011; Davey *et al.*, 2011). Une fois le séquençage effectué, s'en suit un travail d'identification des SNPs *in silico*, à partir de l'ensemble des fragments ainsi obtenus (appelés reads) pour tous les individus. L'identification des SNPs ne nécessite pas nécessairement d'avoir un génome de référence, c'est-à-dire un génome annoté d'une espèce proche, mais peut se faire *de novo*. Les reads identiques sont assemblés en séquences uniques, puis celles-ci sont comparées les unes par rapport aux autres. En autorisant seulement un petit nombre de différences entre elles, les SNPs peuvent être mis en évidence (Davey et Blaxter, 2011).





**Figure 1.3** Étapes du séquençage RAD, de la digestion du génome à la découverte des SNPs (tiré de Emerson *et al.*, 2010). (A) Le génome est digéré par une enzyme de restriction (Stbfl), les sites de coupe de l'enzyme sont matérialisés en rouge. Sur chaque fragment coupé, un code-barres (en noir) est rajouté après un adaptateur. (B) Après le séquençage, tous les reads identiques vont être regroupés ensemble pour former des piles. (C) Chaque pile va être comparée deux à deux pour identifier les SNPs. (D) Les piles qui ont un nucléotide de différence, si celui-ci est retrouvé en suffisamment grande fréquence sur l'ensemble des échantillons, sera considéré comme un SNPs. (E) Sinon, il ne sera pas retenu.

Le séquençage RAD est devenu de plus en plus populaire depuis 2011. Cette technique a permis d'identifier et de génotyper plusieurs milliers de SNPs chez plusieurs organismes marins comme l'épinoche *Gasterosteus aculeatus* (Hohenlohe *et al.*, 2010); le homard américain *Homarus americanus* (Benestan *et al.*, 2015); le mysid *Euphausia superba* (Deagle *et al.*, 2015) et dernièrement chez le copépode *Centropages typicus* (Blanco-Bercial et Bucklin, 2016). L'avantage de cette technique réside dans le grand nombre de marqueurs utilisés, ce qui permet de révéler des structures de populations invisibles avec d'autres types de marqueurs. Chez le

copépode *Centropages typicus*, les RAD-tags ont permis de mettre en évidence la présence d'une structuration au niveau de l'océan Atlantique nord là où d'autres marqueurs n'avaient rien révélés (Blanco-Barcial et Bucklin, 2016). En milieu marin, les études utilisant les SNPs sont encore très peu nombreuses (Unal et Bucklin, 2010; Benestan *et al.*, 2015; Deagle *et al.*, 2015; Blanco-Bercial et Bucklin, 2016). Chez certaines espèces à très grands génomes telles que *Euphasia superba*, l'identification de SNP demeure problématique puisque de nombreux SNPs se trouvent dans des endroits répétés du génome et qu'aucun génome de référence n'existe pour faciliter leur identification (Deagle *et al.*, 2015). L'utilisation de différents outils génétiques combinés apportent différentes perspectives sur la structuration des populations et leur connectivité (Hewitt, 2004; Hickerson *et al.*, 2010).

### 1.5.2 Code-barres ADN

L'utilisation de données moléculaires a permis de révéler l'existence de complexes d'espèces cryptiques en milieu marin (espèces morphologiquement similaires mais génétiquement distinctes –Bickford *et al.*, 2007) dans ce qui était autrefois considéré comme une seule espèce cosmopolite sur la base de caractères morphologiques (Knowlton, 1993; Grosberg et Cunningham, 2001; Beheregaray et Caccone, 2007). La technique du code-barre ADN (Hebert *et al.*, 2003a) a permis de standardiser la caractérisation des espèces cryptiques, en utilisant un fragment d'environ 658bp d'ADN mitochondrial du gène du cytochrome oxydase c I (COI) pour discriminer les espèces animales. Cette région du COI permet de discriminer les espèces sur la base d'une différence entre la divergence intra et interspécifique. En effet, la divergence observée plus faible à l'intérieure des espèces de métazoaires (typiquement <3%) qu'entre les espèces (10 à 25%; Hebert *et al.* 2003b). Pfenninger et Schwenk (2007) ont démontré que la présence d'espèces cryptiques n'est pas restreinte à des taxa particuliers ou propres à certaines régions géographiques. En milieu marin, la présence d'espèces cryptiques est même plutôt commune (Radulovici *et al.*, 2010; Bucklin *et al.*, 2011; Carr *et al.*, 2011; Lörz *et al.*, 2012; Appeltans *et al.*, 2012;

Havermans *et al.*, 2013). De récentes études estiment qu'il y a entre 0,7 et 2,2 millions d'espèces marines (Mora *et al.*, 2011; Appeltans *et al.*, 2012), dont une grande majorité ne sont pas encore connues (Pimm *et al.*, 2014). La présence d'espèces cryptiques suggère que de nombreuses espèces marines ont en fait des distributions plus limitées et que les processus de spéciation en milieu marin peuvent être beaucoup plus complexes que ce que l'on pensait (Hellberg *et al.*, 2002; Peijnenburg et Goetze, 2013). La présence de ces espèces cryptiques soulève la question de la spéciation en milieu marin. La spéciation, c'est-à-dire le processus par lequel les populations deviennent génétiquement et reproductivement isolées, est complexe et peut se faire de façon différente selon les espèces (Palumbi, 1992; Norris et Hull, 2011). Dans le milieu marin, deux types de mécanismes sont impliqués pour expliquer l'apparition de nouvelles espèces: les mécanismes basés sur l'absence de flux de gène et les mécanismes en présence de flux de gène (Norris, 2000). L'isolement géographique conduit à la séparation physique d'une population en deux sous-populations qui vont diverger pour devenir éventuellement au cours du temps des espèces distinctes (Palumbi, 1994), c'est la spéciation allopatrique (Coyne et Orr, 2004). Par exemple, la fermeture de l'isthme du Panama il y a environ 3 millions d'années a contribué à la formation de plusieurs espèces sœurs en isolant physiquement les populations (Knowlton, 1993; Knowlton et Weigt, 1998, Lessios *et al.*, 1998). Comme il existe peu de barrières dans les océans, la spéciation allopatrique est assez peu commune (Palumbi, 1994). D'autres modèles de spéciation ont été développés pour expliquer l'apparition d'espèces en présence de flux de gènes (revu dans Norris, 2000 et Coyne et Orr, 2004). La spéciation peut être la conséquence d'un effet fondateur, c'est-à-dire d'un groupe de migrants qui colonise un nouvel habitat ou bien un groupe d'individus qui s'isole de la population principale (i.e. spéciation péripatrique – Coyne et Orr, 2004). La spéciation peut survenir en périphérie de l'aire de distribution d'une espèce (i.e. spéciation parapatrique). La spéciation sympatrique où les populations divergent alors qu'elles partagent la même aire géographique a particulièrement intrigué les chercheurs.

Plusieurs critères sont nécessaires pour définir une spéciation sympatrique comme le chevauchement des aires de répartition ou une absence d'introgression c'est-à-dire d'échange de gène entre des espèces génétiquement proches (Coyne et Orr, 2004; revus dans Bolnik et Fitzpatrick, 2007). Le milieu marin semble se prêter beaucoup plus à la spéciation sympatrique ou parapatricque qu'allopatricque et quelques études ont essayé d'identifier des cas de spéciation marine (e.g. Crow *et al.*, 2010), mais ce modèle reste encore controversé (Fitzpatrick *et al.*, 2009).

#### 1.6 Le domaine d'étude : l'Arctique

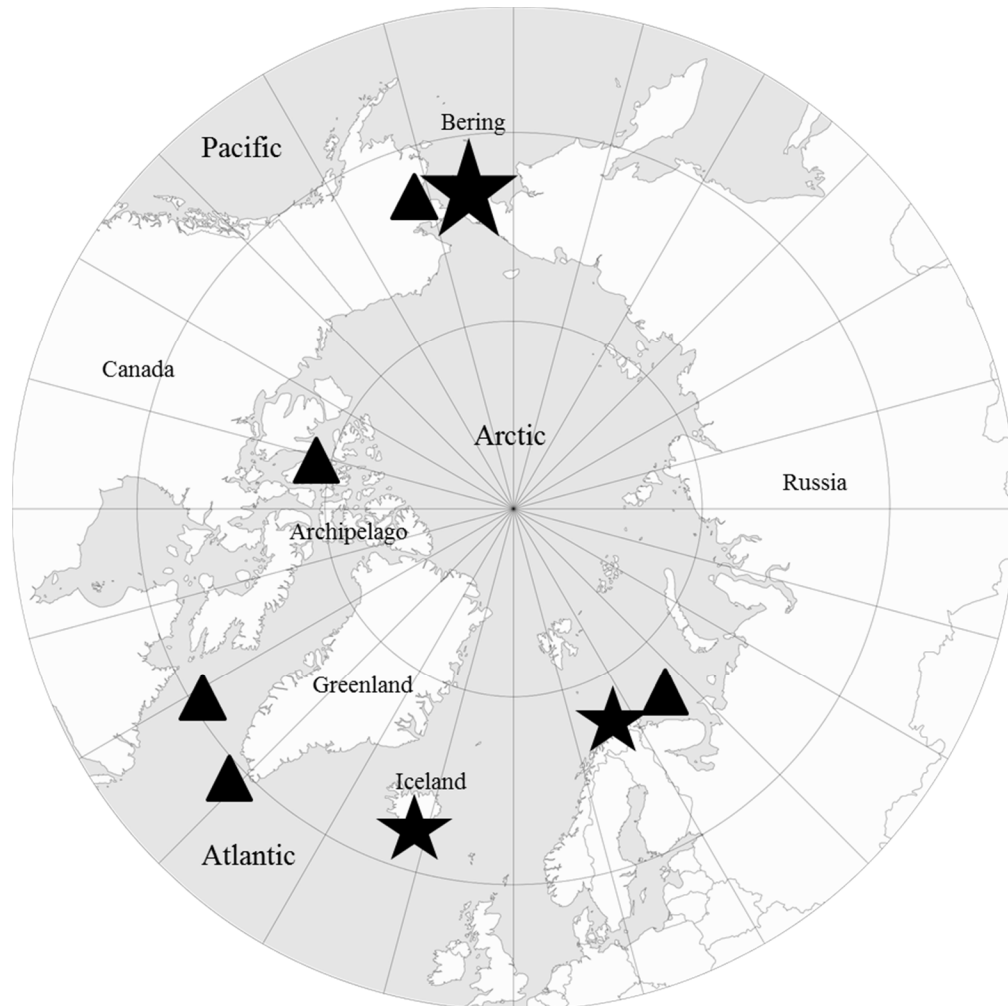
L'océan Arctique est un océan enclavé ayant de faibles connexions avec les deux océans qui l'entourent, il est aussi bordé de mers peu profondes et d'un Archipel côté canadien. L'océan Arctique est connecté avec l'océan Pacifique par le détroit de Béring et avec l'Atlantique par le détroit de Fram. Tandis que les eaux de l'Atlantique se répandent dans l'ensemble de l'océan Arctique via le détroit de Fram, les mers de Barents et de Kara; les eaux du Pacifique longent les côtes canadiennes de l'archipel pour en ressortir au niveau de la baie de Baffin et des côtes du Labrador (Wassman *et al.*, 2015). Chaque masse d'eau transporte dans l'Arctique non seulement des nutriments mais aussi des individus provenant des océans Atlantique et Pacifique (revu dans Wassman *et al.*, 2015; Bluhm *et al.*, 2015). Néanmoins, l'établissement de ces espèces dépendra de leur capacité à survivre et se reproduire dans les conditions arctiques (Wassman *et al.*, 2015).

L'océan Arctique est l'un des océans les moins bien connus (Wassman *et al.*, 2011). De ce fait, les connaissances sur la biodiversité marine arctique sont encore limitées (Archambault *et al.*, 2010; Hardy *et al.*, 2011; Bluhm *et al.*, 2011). Même si peu d'espèces sont endémiques de l'Arctique (Josefson et Moviesky, 2013), la biodiversité marine arctique en général semble être comparable à celle observée au niveau des côtes du Pacifique et de l'Atlantique canadien (Darnis *et al.*, 2012).

Quelques études ont été réalisées pour répertorier les espèces marines présentes dans l'océan Arctique (Sirenko, 2001; Bucklin *et al.*, 2010a;b; Kosobokova et Hopcroft, 2011; Bucklin *et al.*, 2011; Bluhm *et al.*, 2011), démontrant que la diversité dans l'océan Arctique est encore sous-estimée. De plus, la mise en place d'un projet de recensement de la biodiversité polaire (Polar Barcode of Life) à l'aide de la technique du code-barres ADN, a permis d'améliorer ces connaissances, même si certaines zones sont encore méconnues (Archambault *et al.*, 2010; Bluhm *et al.*, 2011). Cette biodiversité marine arctique est le résultat de l'histoire complexe de ce milieu (Hardy *et al.*, 2011) et des échanges avec les océans Pacifique et Atlantique limitrophes. L'océan Arctique se compose de deux parties : la partie canadienne, ancienne et datant du Crétacé (65 Ma), qui est isolée d'une partie eurasienne plus récente (55 Ma) par les dorsales du Lomonosov et d'Alpha-Mendelejev (Clarke, 1988). À l'opposé de ces deux bassins profonds, les mers adjacentes de l'Arctique forment un des écosystèmes les plus changeants au monde, caractérisé par d'importantes variations hydrographiques, saisonnières et environnementales (Carmack et Wassmann, 2006). De plus, les multiples ouvertures du Détroit de Béring ont permis l'invasion de taxons du Pacifique dans l'Atlantique, conduisant à une homogénéisation de la composition de la faune au niveau de l'Arctique (Vermeij, 1991; Hardy *et al.*, 2011). Plusieurs études ont montré que les océans Pacifique et Atlantique ont contribué à la colonisation de l'Arctique (Vermeij, 1991; Addison et Hart, 2005; Coyer *et al.*, 2011; Hardy *et al.*, 2011; Carr *et al.*, 2011; Palumbi et Kessing, 1991; Dodson *et al.*, 2007). D'après la plupart de ces études, il semble que l'océan Pacifique ait joué un rôle important dans la recolonisation suite aux dernières glaciations (Vermeij, 1991). Toutefois, l'Atlantique semble aussi avoir eu un rôle plus important dans la recolonisation de l'Arctique chez certaines espèces comme les polychètes (Carr *et al.*, 2011).

Les variations du niveau des mers il y a 2,4 millions d'années durant le Pléistocène ont entraîné une fermeture périodique des connexions entre les océans. Cela a isolé

l'archipel canadien et l'océan Arctique de l'océan Pacifique (Matthiessen *et al.*, 2009) et favorisé la divergence entre les populations par des évènements de colonisations/extinctions (Hewitt, 1996; Cunningham et Collins, 1998), ou de renforcement (Coyne et Orr, 2004). Le détroit de Béring représente la seule ouverture entre l'océan Arctique et l'océan Pacifique et les multiples ouvertures et fermetures de ce détroit au cours du temps ont eu une grande influence sur la faune arctique (Vermeij, 1991). Ainsi, au moins six évènements d'ouverture du détroit de Béring ont été documentés (Harris, 2005), mais la dernière submersion il y a environ 15 000 ans a permis l'ouverture définitive du détroit pour les organismes marins. Les premiers échanges transarctiques sont estimés aux alentours de 3,5 Ma (Vermeij, 1991); même si les premières ouvertures du détroit de Béring sont datées autour de 5,5 Ma (Gladenkov *et al.* 2002) ou même antérieurement (17Ma- Sher, 1999). Plusieurs études ont documenté les échanges transarctiques chez les échinodermes (Palumbi et Kessing 1991; Hardy *et al.* 2011), les bivalves (Vainola 2003; Nikula *et al.* 2007), les algues (Van Oppen *et al.* 1995 ; Coyer *et al.* 2011), le capelan (Dodson *et al.* 2007) ou les polychètes (Carr *et al.* 2011), montrant l'influence de la dynamique de ce détroit dans la biodiversité arctique.

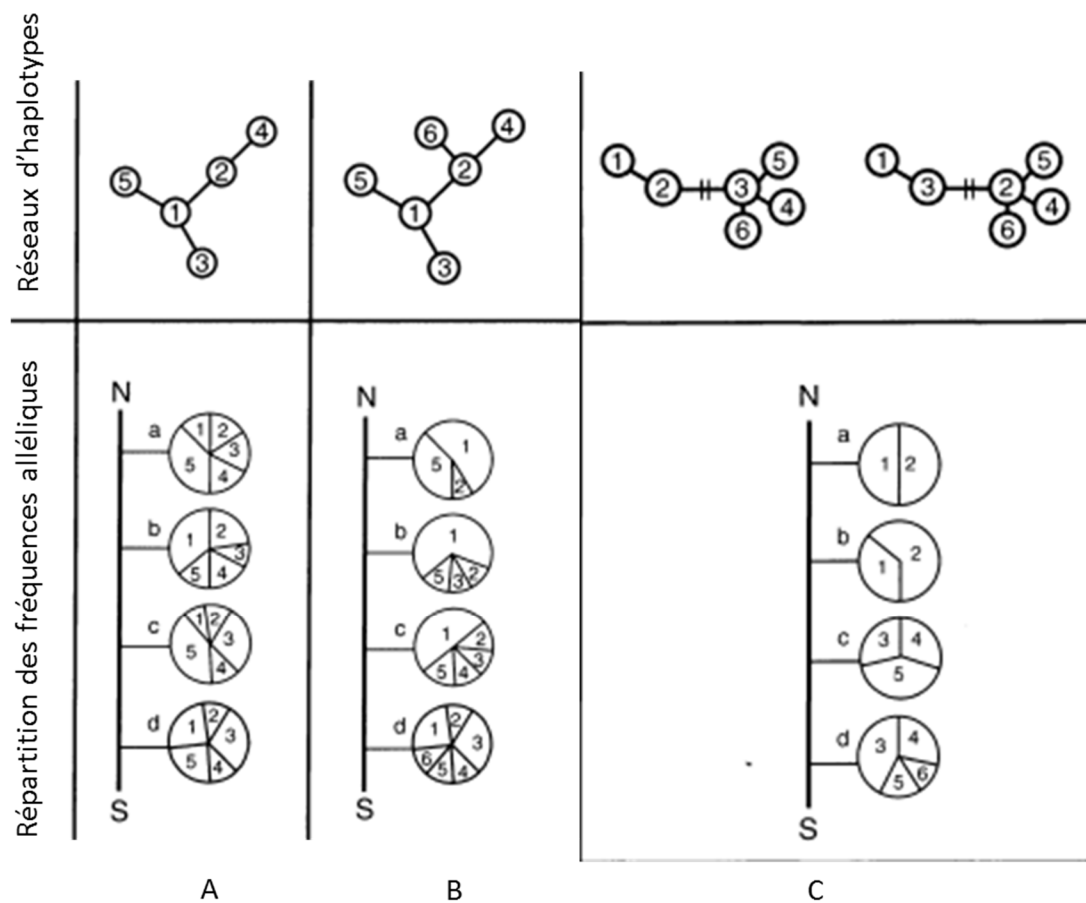


**Figure 1.4** Carte des différentes zones refuges et zones de contacts secondaires en Arctique (modifié d'après Hardy *et al.*, 2011). Les étoiles représentent les zones refuges. Les triangles représentent les zones de contacts secondaires.

L'océan Arctique est un environnement particulier de par la présence de glace. La couverture de glace en Arctique a varié au cours du temps (revu dans Polyak *et al.*, 2010) et les oscillations climatiques passées ont eu aussi un impact sur la structuration de la biodiversité marine arctique (Hardy *et al.*, 2011; Bucklin *et al.*,

2010a). Durant le Pliocène (~3Ma), une couverture de glace était déjà présente en Arctique (Matthiessen *et al.* 2009), réorganisant fortement le système Arctique (Polyak *et al.*, 2012). Toutefois, ce sont surtout les cycles de glaciations du Quaternaire, notamment le dernier maximum glaciaire aux alentours de 15 000 ans, qui ont eu énormément de conséquences sur les populations marines (Hewitt, 1996; Hardy *et al.*, 2011; Maggs *et al.*, 2008). Les cycles de glaciations ont recouvert l'Arctique de glace, contribuant à l'expansion des populations des espèces arctiques qui se rétractaient en périodes interglaciaires (Provan et Bennett, 2008). Ces fluctuations démographiques ont influencé aussi la diversité génétique de ces populations (Hewitt, 1996; 2004; revu dans Maggs *et al.*, 2008, Allcock et Strugnell, 2012). La présence de zones refuges a permis à certaines populations de survivre mais aussi de diverger pendant les périodes glaciaires, conduisant à la présence d'allèles endémiques de ces zones refuges (Slatkin, 1995- allèle privé). Dans l'océan Arctique, quelques zones refuges ont été identifiées (revu dans Hardy *et al.*, 2011; Figure 1.4), notamment en Béringie, contribuant à maintenir une certaine diversité dans les populations d'espèces arctiques. Des zones de contacts secondaires ont aussi été mises en évidence en Arctique. Ces zones correspondent à la rencontre de différentes populations provenant de différents refuges. Comme les zones refuges, les zones de contacts secondaires se caractérisent par une forte diversité génétique mais un réseau d'haplotypes différents (Figure 1.5).





**Figure 1.5** Exemples de modèles de réseaux d'haplotypes et de distribution des fréquences alléliques en fonction d'un gradient Nord (N)- Sud (S), sous différents scénarios de refuges et recolonisations (modifié de Maggs *et al.*, 2008). Pour chaque modèle, un réseau de 5 ou 6 haplotypes, numérotés de 1 à 5-6, est représenté en haut, et leur fréquence dans quatre populations (a-d). Dans tous les cas, l'allèle 6 correspond à un allèle privé restreint à une région particulière. (A) Le modèle nul dans lequel aucun changement dans les fréquences n'est observé en fonction de la latitude. (B) Le modèle représente un gradient de la richesse allélique qui décline de la zone refuge au Sud vers le Nord. (C) Le modèle représente le cas d'un contact secondaire entre deux populations isolées dans des refuges distincts.

Le milieu Arctique change plus rapidement que n'importe quelle autre région (Wassman *et al.*, 2011; Wassman, 2011; Wassman *et al.*, 2015; Cronin et Cronin, 2015). Depuis plusieurs décennies, la couverture de glace de l'océan Arctique diminue pour atteindre des minimums records (Vihma, 2014). Cette diminution de la couverture de glace a plusieurs conséquences sur la faune mais aussi sur les activités anthropiques dans les hautes latitudes. Premièrement, la diminution de la couverture de glace saisonnière libère de nouvelles routes navigables, entraînant des conséquences écologiques et économiques importantes (Corbett *et al.*, 2010). Avec l'augmentation du trafic maritime, les risques de catastrophe ou d'invasion d'espèces non-indigènes par le relargage des eaux de ballasts sont accrus (Chan *et al.*, 2013). Deuxièmement, la perte de glace en Arctique est synonyme de perte d'habitat pour de nombreuses espèces (Brierley et Kingsford, 2009). Plusieurs espèces sont dépendantes de la glace pour survivre, comme l'ours polaire (*Ursus maritimus*). La diminution de la glace diminue l'accès des ours à leurs proies et affecte leur survie (Rode *et al.*, 2012). Troisièmement, la perte de glace peut avoir un impact sur la production primaire (Arrigo *et al.*, 2008), et sur la migration verticale du zooplancton (Berge *et al.*, 2009). Les changements globaux qui surviennent actuellement entraînent un réchauffement des eaux arctiques qui vont aussi avoir des conséquences sur la faune marine. Un réchauffement de moins de 4°C va être souvent « bénéfique » pour la plupart des espèces en accélérant leur croissance et en diminuant la durée de leur phase larvaire (Byrne, 2012). Par contre, le réchauffement des eaux permet aussi aux espèces boréales d'étendre leur zone de distribution plus au nord. Plusieurs études ont documenté un changement dans les aires de répartition de certaines espèces marines qui remontent vers de plus hautes latitudes du fait du réchauffement de l'eau (Kuzmin *et al.* 1998 ; Anderson et Piatt, 1999; Berge *et al.* 2005; Grebmeier *et al.* 2006 ; Mueter et Litzow, 2008; Agnalt *et al.* 2011; Wassman *et al.* 2011; Kraft *et al.*, 2013b). De plus, la présence de la diatomée *Neodenticula seminae* du Pacifique dans la mer du Labrador après 800 000 ans d'absence coïncide avec les récents changements observés en Arctique (Reid *et al.*, 2007; Wassman *et al.*, 2011; 2015). Il

est probable que, ces prochaines années, ces phénomènes de migration chez d'autres espèces marines deviennent de plus en plus courants et de plus en plus fréquents (Wassman *et al.*, 2015). En plus de modifier la distribution des espèces, le réchauffement des océans modifie aussi l'environnement marin dans lequel elles évoluent.

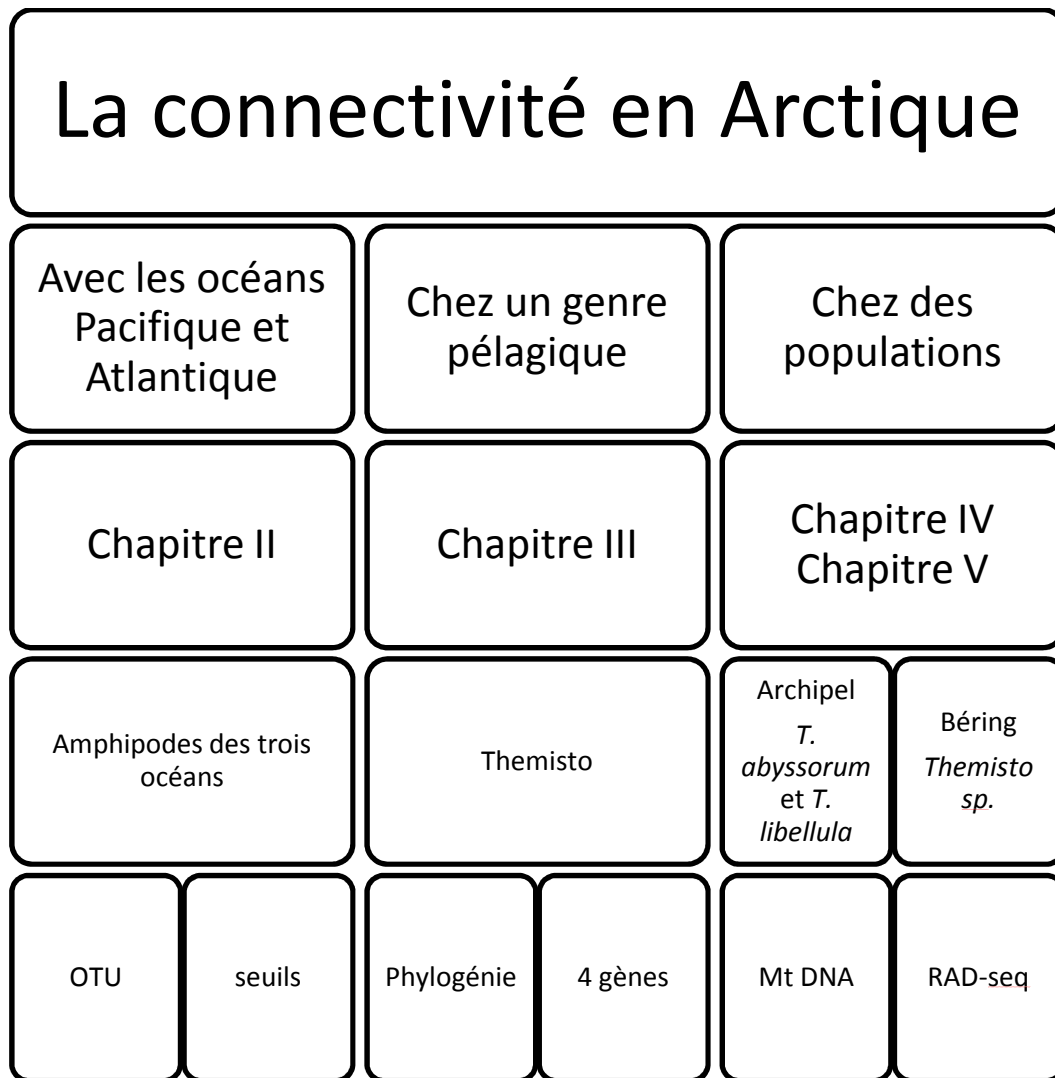
### 1.7 Les Amphipodes

Les crustacés de l'ordre des amphipodes sont retrouvés aussi bien dans les milieux aquatiques que terrestres à toutes les latitudes. Ils possèdent une grande diversité en espèce ainsi qu'au niveau morphologique rendant leur identification parfois difficile (Browne *et al.*, 2007). Ils sont caractérisés par un développement direct c'est-à-dire qu'ils ne produisent pas de larves ainsi qu'une absence d'organe natatoire les limitant dans leur dispersion sur de grandes distances (Väinölä *et al.*, 2008; Radulovici, 2012). On connaît que très peu de chose sur la diversité génétique des amphipodes. Aucune étude à ce jour n'a été réalisée sur les échanges génétiques entre les populations d'amphipodes marins. Pourtant, en Arctique, les amphipodes domine la macrofaune en terme d'abondance et de biomasse (Josefson et Mokievsky, 2013), et comme tous les autres organismes arctiques, les changements climatiques vont aussi affecter la diversité des amphipodes ainsi que leurs populations.

### 1.8 Les objectifs de thèse

Les conditions environnementales extrêmes, les importants apports en eaux douces par les fleuves, ainsi que l'histoire complexe de l'océan Arctique ont influencé les échanges avec les océans limitrophes mais aussi les échanges entre les populations arctiques. Pourtant, les études sur la connectivité marine montrent des inégalités dans les zones d'études, dans les échelles spatiales et les organismes étudiés (Bouchet, 2006; Tittensor *et al.*, 2010; Archambault *et al.*, 2010). Notamment, le milieu arctique comporte encore certaines zones d'ombre, en particulier l'Archipel canadien. L'objectif général de ma thèse est de documenter la connectivité en milieu marin

arctique à différentes échelles spatiales et taxonomiques afin de détecter la présence éventuelle de barrière à la dispersion, en utilisant comme modèles les amphipodes (Figure 1.6).



**Figure 1.6** Diagramme récapitulatif de la démarche de la thèse. OTU= operational taxonomic unit; MtDNA= ADN mitochondrial; Rad-seq= Restriction site associated DNA sequencing.

### 1.7.1 Objectif 1(Chapitre II) : Connectivité à l'échelle tri-océanique chez les amphipodes

Les glaciations du Pléistocène ont joué un rôle important sur la biodiversité marine arctique (Hardy *et al.*, 2011), plus particulièrement dans la subdivision des populations de nombreux organismes aquatiques canadiens (Vermeij, 1991; Carr *et al.*, 2011; Layton *et al.*, 2014). De plus, le sens des courants remontant du Pacifique vers l'Arctique permet l'arrivée de taxa du Pacifique dans l'Arctique, mais empêche les taxa de l'Atlantique de coloniser le Pacifique (Vermeij, 1991). Récemment, une étude sur la biodiversité des polychètes au niveau des trois océans canadiens a mis en évidence que l'océan Pacifique jouait un rôle mineur dans la recolonisation de l'Arctique (Carr *et al.*, 2011).

Plusieurs espèces d'amphipodes sont endémiques de l'Arctique. Sirenko (2001) en dénombre plus de 550 espèces, dont une majorité appartient à la famille des Gammaridae. Néanmoins, l'identification morphologique de ces espèces peut être souvent problématique à cause de la présence de caractères homoplasiques (Radulovici *et al.*, 2010 ; Machida *et al.*, 2009); pour cette raison, l'utilisation du code-barres ADN pour identifier les espèces peut représenter un certain avantage (Lefébure *et al.*, 2006, Costa *et al.*, 2007, Matzen da Silva *et al.*, 2011). La technique du code-barres ADN permet de s'affranchir de certaines difficultés taxonomiques en identifiant les espèces à un niveau moléculaire. La généralisation de cette technique pour l'identification et la création d'une plate-forme de code-barres ADN (BOLD : Barcode Of Life Database), permet de réaliser des méta-analyses utiles pour l'évaluation de barrières à la dispersion, et l'estimation de la connectivité sur de grandes échelles géographiques (Carr *et al.* 2011).

L'utilisation de données génétiques pour délimiter les espèces représente un outil attractif pour décrire la biodiversité plus rapidement, surtout dans le contexte actuel de crise environnementale. En se basant sur les séquences de nucléotides, les espèces

peuvent être définies, sur la base de leur divergence, en OTU (Operational Taxonomic Unit) qui représente une espèce potentielle (Vogler et Monaghan, 2007; Goldstein et DeSalle, 2011). Il existe plusieurs méthodes pour délimiter les espèces en utilisant les données moléculaires qui se regroupent en deux grandes catégories : les méthodes qui utilisent les arbres phylogénétiques et celles qui utilisent les distances génétiques (revues dans Goldstein et DeSalle, 2011; Delicour et Flot, 2015). Les méthodes basées sur les arbres vont définir les espèces sur la base de leur monophylie. Les méthodes basées sur les distances vont définir les espèces en se basant sur un seuil au-delà duquel deux séquences seront considérées comme appartenant à des espèces distinctes. Initialement, un seuil de 3% a été utilisé pour délimiter les espèces (Hebert et al. 2003b). Ce seuil est encore souvent utilisé pour définir les espèces même si des critiques existent. En effet, dans certaines études, il a été démontré que cette valeur avait tendance à surestimer le nombre d'espèces notamment chez les crustacés (Lefébure *et al.*, 2006). Un autre seuil de 16% a été proposé pour délimiter les crustacés y compris les amphipodes (Lefébure *et al.*, 2006). Néanmoins, ce seuil n'a été estimé que sur une seule famille d'amphipodes : les Gammaridae, mais à ce jour aucun seuil n'est défini pour délimiter l'ensemble des espèces d'amphipodes.

Le premier chapitre de cette thèse a pour objectif de décrire et d'examiner les patrons de connectivité des amphipodes entre les océans Arctique, Pacifique et Atlantique en utilisant la technique du code-barres ADN. En se basant sur l'orientation des courants arrivant du Pacifique vers l'Arctique, l'Archipel canadien doit échanger plus avec le Pacifique qu'avec l'océan Atlantique.

Hypothèse 1: Le Pacifique a joué un rôle majeur dans la colonisation de la faune marine arctique chez les amphipodes.

Prédiction : L'océan Arctique aura davantage d'OTUs en commun avec l'océan Pacifique qu'avec l'océan Atlantique.

Pour répondre à cet objectif, nous avons échantillonné des amphipodes provenant de l'Archipel canadien, dont nous avons obtenu les codes-barres. En parallèle, nous avons récupéré sur la base de données BOLD les séquences COI de tous les amphipodes des océans Atlantique, Arctique et Pacifique. À partir de cet ensemble de données, nous avons utilisé différents seuils pour définir les OTUs présentes dans chacun des océans. Nous avons aussi estimé un seuil à l'aide d'une méthode bayésienne. Nous avons comparé les performances de chacun des seuils dans l'identification et la délimitation des espèces d'amphipodes.

#### 1.7.2 Objectif 2 (Chapitre III) : Phylogénie d'un genre pélagique d'amphipode

Par définition, les espèces holozooplanktoniques passent leur vie entière dans la colonne d'eau. Cette caractéristique implique un potentiel de dispersion très grand médié par les courants. Pourtant, plusieurs études génétiques ont révélé la présence d'espèces cryptiques chez les taxa pélagiques (Goetze, 2005; 2010; Lörz *et al.*, 2012; Havermans *et al.*, 2013), suggérant la présence d'une isolation et une spécialisation environnementale plus importante que ce qui était attendu traditionnellement chez ces espèces (Norris, 2000; Goetze, 2005).

Parmi les amphipodes, dont la plupart des espèces sont à majorité benthiques, le sous-ordre des Hyperiidæ se distingue par son mode de vie pélagique (Browne *et al.*, 2007). Au niveau des régions polaires, le genre *Themisto* est important en termes de biomasse et d'abondance (Bowman 1960). En Arctique, deux espèces vivants en sympatrie sont retrouvées : *Themisto libellula* et *Themisto abyssorum*. Cependant, leur teneur en lipides semble indiquer des préférences écologiques différentes *T. libellula* semblant être plus proche de la surface et dépendant de la glace alors *T. abyssorum* semble être d'origine sub-arctique et associée à des eaux plus profondes (Dunbar, 1957; Koszteyn *et al.* 1995; Auel *et al.* 2002). Étudier les relations

phylogénétiques entre les différentes espèces de *Themisto* permettrait de mieux comprendre les processus de spéciation en milieu marin.

Afin de mieux comprendre les relations entre les différentes espèces de *Themisto*, nous avons utilisé différents marqueurs mitochondriaux et nucléaires pour reconstruire une phylogénie du genre à l'échelle de l'hémisphère Nord. Pour cela nous avons obtenu plusieurs spécimens provenant des océans Pacifique, Atlantique et Arctique auxquels nous avons ajouté les données d'un *Themisto* du Pacifique (*T. pacifica*) provenant d'une autre étude (Hurt *et al.*, 2013). Nous avons reconstruit les phylogénies par maximum de vraisemblance et par inférence bayésienne de chacun des gènes, puis après vérification, nous avons reconstruit une phylogénie du genre.

#### 1.7.3 Objectif 3 (Chapitre IV): Structure des populations au niveau de l'archipel canadien

Dans l'hémisphère nord, il a été avancé que la diversité génétique diminue quand la latitude augmente (Hellberg *et al.*, 2001; 2009), étayant une colonisation du pôle après la période glaciaire du Pléistocène (Hellberg, 2009). Chez les espèces marines arctiques, la présence de refuges où les espèces auraient pu survivre a pu contribuer à maintenir un certain niveau de diversité génétique. Quelques études (Fedorov et Stenseth, 2002; Hardy *et al.*, 2011) ont montré que les populations de l'archipel canadien résultaient d'un contact secondaire entre des populations provenant de différents refuges. Néanmoins, ces études ne combinaient pas plusieurs marqueurs génétiques. En fait, très peu d'études sur la structuration des populations marines ont utilisé une combinaison de différents marqueurs. Moins d'une quarantaine d'études à ce jour, ont été réalisées sur les espèces holozooplanktoniques et aucune n'a encore été conduite sur une espèce provenant de l'Arctique en utilisant le séquençage à haut débit (Beheregaray, 2008; revu dans Peijnenburg et Goetze, 2013).



L'archipel canadien est caractérisé par un ensemble de chenaux plus ou moins profond, traversé par un courant principal unidirectionnel provenant de la mer de Beaufort en direction de la baie de Baffin. Le sens des courants au niveau de l'Archipel devrait orienter les échanges entre les individus rendant les populations relativement homogènes. L'objectif de ce chapitre est de caractériser la diversité génétique de deux espèces de *Themisto* au niveau de l'Archipel canadien.

Hypothèse: L'Archipel canadien est une zone récemment recolonisée à partir de zones refuges.

Prédiction : Pour chaque espèce, les réseaux d'haplotypes mitochondriaux devraient porter la signature d'un contact secondaire.

Hypothèse : Les *Themisto* sont des espèces pélagiques qui dispersent sur de grandes distances.

Prédiction : En l'absence de barrière à la dispersion au niveau de l'Archipel, une faible ou aucune structuration spatiale au niveau nucléaire ne sera détectée.

Pour tester ces hypothèses, nous avons échantillonné des *Themisto* au niveau de l'Archipel canadien et du Détroit de Béring. Dans un premier temps, nous avons vérifié la présence de structuration à l'aide de marqueurs mitochondriaux. Le lien entre la distance génétique et géographique a été vérifié à l'aide d'un test de Mantel. Dans un deuxième temps, à l'aide du génotypage par séquençage RAD-seq, nous avons estimé la présence de structuration spatiale plus fine. L'analyse discriminante par composante et une méthode d'assignation bayésienne ont aidé à détecter la présence de groupes génétiques distincts.

#### 1.7.4 Objectif 4 (Chapitre V): Structure des populations au niveau du détroit de Béring

L'objectif 2 a permis de mettre au jour la présence en Arctique au niveau du détroit de Béring d'une espèce cryptique dénommée provisoirement *Themisto sp.* Comme aucune information n'est disponible sur cette espèce, l'objectif de ce chapitre est de caractériser la structure génétique de cette espèce. Au niveau du détroit de Béring, les courants proviennent du Pacifique et vont vers l'Arctique. Chez le copépode *Calanus glacialis*, il a été montré que ces courants favorisaient l'arrivée des copépodes en Arctique mais que l'environnement arctique limitait encore leur établissement (Nelson *et al.*, 2009). La Béringie est délimitée au sud par les îles Aleutiennes. Au niveau de ces îles, chez la morue du Pacifique, des isolements entre les populations du Pacifique Nord et de la Béringie ont été caractérisés (Spies, 2012), démontrant que malgré la présence de ces courants, d'autres facteurs limitent la connectivité entre les populations de part et d'autres de ces îles.

Hypothèse: *Themisto sp.* forme une seule grande population en Béringie.

Prédiction: Comme les courants au niveau de la Béringie proviennent du Pacifique et vont vers l'Arctique, on s'attend à n'observer aucune structuration pour cette espèce au niveau de la Béringie.

Pour répondre à cet objectif, nous avons obtenu des *Themisto sp.* provenant de quatre sites répartis du Golfe de l'Alaska à la mer de Tchouktche. Nous avons vérifié la présence d'une structuration spatiale à l'aide des marqueurs mitochondriaux et des SNPs obtenus par séquençage RAD. Les analyses de diversité ainsi que les tests d'assignation ont permis de vérifier la présence éventuelle d'une panmixie.

## CHAPITRE II

### SPECIES IDENTIFICATION AND CONNECTIVITY OF MARINE AMPHIPODS IN CANADA'S THREE OCEANS.

Astrid Tempestini<sup>1</sup>, Soren Rysgaard<sup>2</sup> et France Dufresne<sup>1</sup>

Manuscrit à soumettre à *PLoS One*

<sup>1</sup> Université du Québec à Rimouski, Département de biologie et Centre d'études nordiques, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1

<sup>2</sup> University of Manitoba, 584, Wallace Building, Winnipeg, MB, Canada R3T 2N2

Contribution des auteurs (initiales)

Mise en place des protocoles expérimentaux : AT

Récolte des données : AT

Analyse des données : AT

Interprétation des résultats : AT

Rédaction du manuscrit : AT

Commentaires et correction du manuscrit : AT, FD, SR

Financement: FD

## 2.1 Résumé

Comprendre la distribution de la biodiversité marine est essentielle pour mieux évaluer l'impact des changements globaux. La faune marine Arctique est dominée par les amphipodes. Même si les amphipodes représentent un ordre important des crustacés, leurs nombres dans les océans canadiens est actuellement inconnu. Les espèces de cet ordre sont connues pour être difficile à distinguer sur une base morphologique à cause de lacune dans leur description taxonomique ou la présence d'homoplasie dans les caractères. Le code-barres ADN est un outil qui permet d'utiliser une séquence ADN pour identifier et délimiter les espèces. L'objectif de cette étude est d'estimer la biodiversité des amphipodes marins au niveau des trois océans canadiens en utilisant le code-barres ADN. Pour ce faire, nous avons produit une base de données à partir des séquences publiques d'amphipodes provenant de la base des codes-barres ADN BOLD, auxquelles nous avons ajouté 374 nouvelles séquences provenant de l'Archipel canadien. Dans un premier temps nous avons délimité les espèces en utilisant divers seuils proposés dans la littérature et un seuil estimé par une méthode bayésienne ABGD. Nous avons par la suite comparé la composition en amphipodes de chacun des océans afin de mettre en évidence l'influence des océans Atlantiques et Pacifiques sur la biodiversité Arctiques. L'analyse de 2557 séquences représentant 243 espèces potentielles révèlent une forte connectivité entre les océans Arctique et Atlantique. Nos résultats suggèrent qu'un seul seuil pour délimiter les espèces n'est probablement pas le plus adapté chez les amphipodes. Cette étude met en lumière le défi de délimiter les espèces sur la base d'un code-barres ADN et l'urgence d'augmenter les inventaires de code-barres chez les invertébrés marins.

**Mots-clés:** Code-barres; Amphipode; Arctique; Biodiversité; Marin; Trans-Arctique

## 2.2 Abstract

Monitoring the distribution of marine biodiversity is a crucial step to better assess the impacts of global changes. Arctic marine fauna is dominated by amphipods in terms of abundance and biomass. Amphipods represent an important marine order of crustaceans but the number of species found in the different Canadian oceans is currently unknown. This group represents a taxonomic challenge due to poor taxonomic descriptions and morphological homoplasy. The aim of this study was to assess the biodiversity of marine amphipods in the three Canadian oceans using DNA barcodes. To do so, we produced a database of DNA barcodes of amphipods from the three Canadian Oceans publicly available from the BOLD website to which we added 374 new sequences from the Canadian Arctic Archipelago. We first delimited amphipod species based on various thresholds and estimated the appropriate barcode gap with the ABGD method. We then compared the composition of amphipods among the three oceans in order to assess the influence of past transarctic exchanges on Arctic diversity. Analysis of 2557 sequences representing 243 provisional species revealed high connectivity between the Atlantic and Arctic Oceans. Our results also suggest that a single threshold to delimitate species is not suitable for amphipods. This study highlights the challenges involved in species delimitation and the need to obtain complete barcoding inventories in marine invertebrates.

**Key-words: DNA barcoding; Amphipods; Arctic; Biodiversity; Marine; Trans-arctic**

### 2.3 Introduction

Oceans cover most of the planet, but are still poorly known in terms of biological composition and species richness (Angel, 1993; Webb *et al.*, 2010; Archambault *et al.*, 2010, Appeltans *et al.*, 2012). As climate changes and human pressures are growing, understanding the distribution of marine biodiversity is a crucial step towards an effective monitoring of marine ecosystems (Webb *et al.*, 2010). Numerous species of marine invertebrates are still awaiting taxonomic description (Appeltans *et al.*, 2012). Marine crustaceans are notoriously difficult to identify to the species level by traditional approaches due to their enormous morphological diversity and because morphological stasis is frequent in this group (Knowlton, 1993; Radulovici *et al.*, 2010). DNA barcoding proposed by Hebert *et al.*, (2003a) is a method for identifying unknown specimens based on mitochondrial DNA (mtDNA) sequences. This technique has proven to be a very important tool to flag cryptic species in crustaceans (Radulovici *et al.*, 2009). Despite a decade of barcoding, only 7000 crustacean species have been barcoded out of a total of 67,000 described species and it is estimated that there could be as many as 150,000 species (Raupach and Radulovici, 2015).

Identification and estimation of species diversity based on a single genetic locus often appears the best option available for groups for which taxonomy is poor or inexistent (Pulliandre *et al.*, 2012; Carstens *et al.*, 2013). DNA barcode can be used to assign unknown individuals to taxonomic entities based on sequence similarity (Vogler and Monaghan, 2007, Goldstein and DeSalle, 2011). Inexact matches are either grouped with taxa present in the database or identified as new to the database based on whether they fall within a threshold of sequence similarity (Vogler and Monaghan, 2007). Therefore delimiting species based on barcode requires setting a specific threshold beyond which two sequences will belong to different putative species referred as operational taxonomic units (OTUs). Traditionally, this threshold corresponds to the gap (i.e. barcode gap) observed among intraspecific and

interspecific distance distributions. Clustering methods are the most popular methods to delimitate species based on genetic distance (Kekkonen *et al.*, 2014). Diverse methods exist to estimate the threshold used to group sequence into provisional OTU. The standard distance cut-off of 10 times the distance within species proposed by Hebert *et al.*, (2004), was too conservative and failed to detect some new species (Hickerson *et al.*, 2006; Knowles and Carstens, 2007). The method proposed by Lefébure *et al.*, (2006) aims to test the correlation between taxonomic ranks and molecular divergences and determine the appropriate molecular thresholds to help taxonomic decision. More recently, an alternative approach based on Bayesian methods, automatically finds the distance where the barcode gap is (ABGD- Pulliandre *et al.*, 2012). All these methods take the advantage of the widely used cytochrome c oxidase (COI) mitochondrial gene for delimitating species. In crustaceans, barcode gap detection as proposed by Lefébure *et al.*, (2006) has proven to be an efficient tool to discriminate crustaceans' species like decapods (da Silva *et al.*, 2011) and amphipods (Costa *et al.*, 2009).

Despite the fact that Arctic regions are already impacted by global warming, the Arctic Ocean is one of the less understood region in the world and its marine biodiversity is one of the least characterized (Wassman, 2011; Bluhm *et al.*, 2011; Josefson and Mokievsky, 2013; Wassman *et al.*, 2015). Arctic marine biodiversity has been shaped by the complex history and environment of this region. At least six openings of the Bering Strait since 5.5 million years ago have allowed trans-Arctic exchanges and invasions of the North Atlantic region by the North Pacific species (Vermeij, 1991; Gladenkov *et al.*, 2002; Heads, 2005). Moreover, Quaternary glaciations have pushed some taxa out of the Arctic and recolonisation of the Arctic occurred from neighboring oceans during interglacial periods (Hardy *et al.*, 2011). Several studies have documented these transarctic exchanges in molluscs (Vermeij, 1991; Layton *et al.*, 2014); in algae (Coyer *et al.*, 2011); in fishes (Dodson *et al.*, 2007; Pálsson *et al.*, 2009) and in polychaetes (Carr *et al.*, 2011). So, over millions of

years, marine species have dispersed through the Arctic several times leading to a complex pattern of biodiversity (Hardy *et al.*, 2011).

Arctic marine fauna is dominated in terms of abundance and biomass by amphipods (Josefson and Mokievsky, 2013). Amphipods are small crustaceans characterized by a direct development and weak active dispersal capabilities (Väinölä *et al.*, 2008). These characteristics favor cryptic speciation and endemism, and prevent widespread distribution as has been demonstrated for isopods (Raupach *et al.*, 2007). Despite their key role in the arctic food web (Darnis *et al.*, 2012), little is known about the biodiversity of Arctic marine amphipods.

The aim of this study was to investigate the biodiversity of marine arctic amphipod using DNA barcodes. In order to achieve this objective, we reviewed the public sequences of all amphipods in the BOLD database to which we added new sequences of amphipods from the Canadian Arctic Archipelago. We first aimed to delimit amphipods species in the three Canadian oceans using two approaches: species threshold from the literature and the ABGD method. Second, we compared species composition of each ocean to see large-scale connectivity. Our results are interpreted in the context of biodiversity estimates and molecular taxonomy.



## 2.4 Material and Methods

### 2.4.1 Samples

Samples of Arctic amphipods were collected in 2011 in the Canadian Arctic Archipelago during the NCGS Amundsen expedition. Samples from Greenland were provided by the Greenlandic natural resources institute of Nuuk, in 2012. All samples were collected using vertical and/or horizontal nets with 200, 500 and 750  $\mu\text{m}$  mesh size. All individuals were preserved in 95% ethanol.

### 2.4.2 COI amplification

The DNA of 374 samples was extracted using the E.Z.N.A tissue extraction kit (OmegaBiotek), or the QuickExtract kit (OmegaBiotek) following the manufacturer's protocols. Polymerase chain reaction (PCR) was conducted as described in Dionne *et al.*, 2011. A 658bp fragment of the cytochrome c oxidase I was amplified using the primer pair LCO1490/HCO2198 (Folmer *et al.*, 1994). PCR products were visualized on a 1.5% agarose gel, and then sequenced at Genome Quebec (McGill University, Canada). Over the 374 individuals used, 202 individuals were successfully sequenced and chromatograms manually checked on MEGA5 (Tamura *et al.*, 2011) and the presence of pseudogene was assessed by translating sequences into amino acids. All sequences were deposited in the GenBank database under the accession numbers XXXX.

### 2.4.3 BOLD Sequences:

All amphipod sequences and their corresponding collection information that were publicly available in the Barcode of Life Data Systems (May 2015; BOLD, [www.boldsystem.org](http://www.boldsystem.org), "Amphipoda") were downloaded. A total of 2808 sequences with their geographical coordinates, corresponding to the Canadian Pacific, to the Arctic Ocean and to the North East Atlantic, were retained for further analyses

(supplementary material). In order to avoid the presence of pseudogenes, sequences were translated in amino acids to verify the presence of stop codons.

#### 2.4.4 Trimming and Multiple alignments

A total of a 3010 sequences were first aligned with MAFFT (Kato and Standley 2013) version 7 web server (<http://mafft.cbrc.jp/alignment/server>) with default parameters. Based on this first alignment, sequences with a length under 500 bp were removed. A second multiple alignment was performed on MAFFT (Kato and Standley, 2013) and sequences were trimmed to the same length of 528 bp. As methods for delimiting species are sensitive to missing data, we removed sequences with too many missing values. The final dataset consists of 2557 sequences of 532 bp long.

#### 2.4.5 Distance based methods

##### 2.4.5.1 Threshold from the literature

We used two thresholds to cluster our sequences into provisional species (referred as operational taxonomic unit–OTU), a threshold of 3% commonly used to define species and the 16% threshold of Lefébure *et al.*, (2006). Uncorrected pairwise distances were first computed for both threshold values and sequences were clustered into OTU using the nearest neighbor method. Briefly, this method allows the clustering of sequences in the same OTU if they are at most X% distant from the most similar sequence in the OTU. All analyses were conducted in MOTHUR (Shloss *et al.*, 2009).

##### 2.4.5.2 Automatic Barcode Gap Discovery (ABGD)

Since intra-specific divergences are smaller than inter-specific ones (Harvemans *et al.* 2013), a gap in the distribution of all pairwise distances can be identified using the Automatic Barcode Gap Discovery method available at [wwwabi.snv.jussieu.fr/public/abgd](http://wwwabi.snv.jussieu.fr/public/abgd). This method is described in detail in Puillandre

(*et al.*, 2012). Briefly, the data was partitioned into a number of groups (i.e. species) such that the distance between two sequences taken from distinct groups was always larger than a given threshold distance (i.e. barcode gap- Pulliandre *et al.*, 2012). Kimura 2 parameters (K2P) and default parameters were used to identify the barcode gap. After the barcode gap discovery, sequences were clustered into OTU based on the estimated threshold.

#### 2.4.5.3 Species threshold identification

As the threshold proposed by Lefébure *et al.*, (2006) was based on a single family of amphipods, we used the same method to identify a species threshold for all Amphipoda. To do so, we collected all amphipod sequences from BOLD (January, 2016). Among the 15516 records, all sequences without a taxonomic identification to the species level, sequences with less than 500bp and species associated with pseudogenes were discarded. A total of 8471 sequences corresponding to 89 families were examined further. Families with less than 30 sequences or containing less 3 species were also discarded from this dataset. The diversity assessments for the amphipods and for the most represented families were analysed from the data set with 3879 sequences from 272 species, 70 genera, and 10 families. After performing a first alignment with MAFFT, all sequences were trimmed to the same length of 501bp. After this step, Kimura 2 parameters pairwise distances were computed at each taxonomic level intrafamily (F), intragenus (G) and intraspecies (S) in MEGA5 (Tamura *et al.*, 2011) and plotted by family using the boxplot representation available in R and described in Lefébure *et al.*, 2006. The three ranges of thresholds were plotted to see which one best discriminates the different amphipod species.

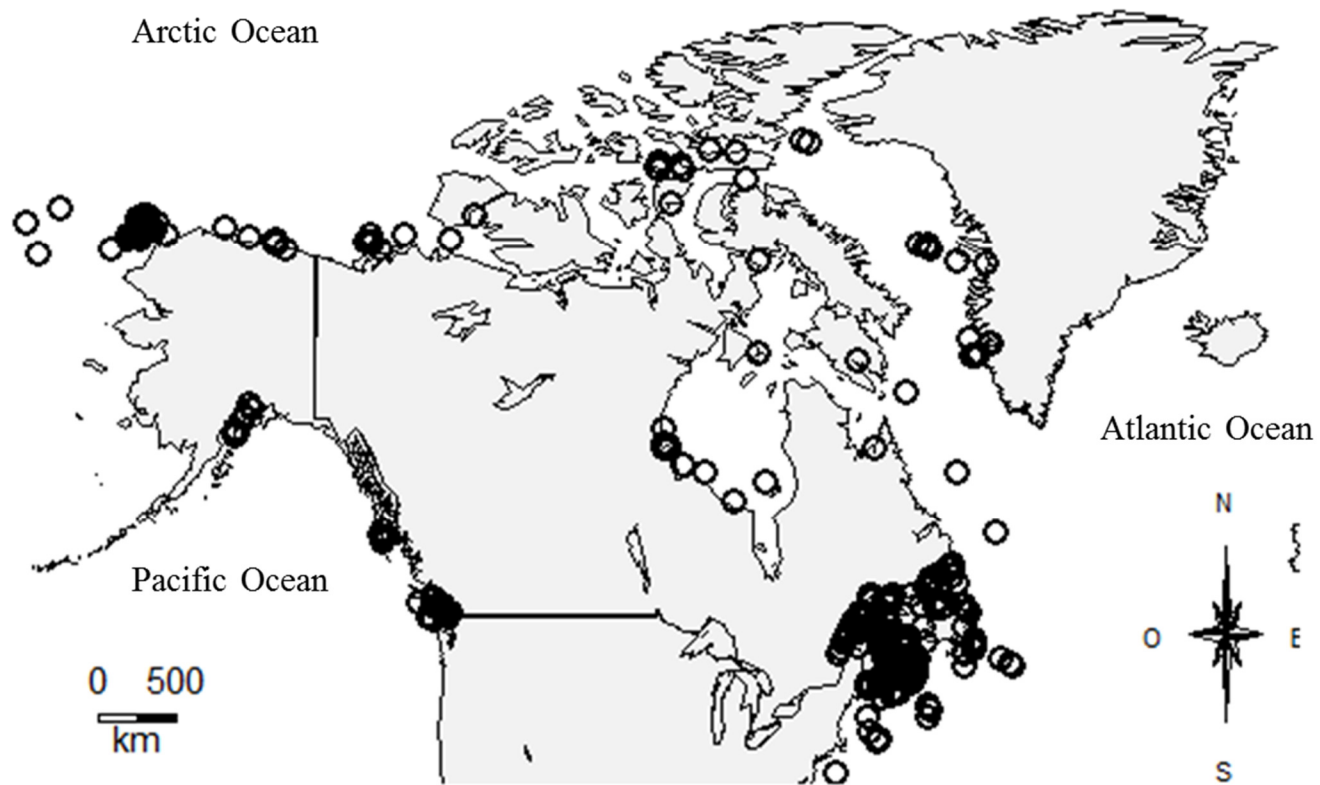
#### 2.4.5.4 Taxonomic amphipod reference database

In order to associate taxonomy to the OTU we identified, we created a molecular taxonomic database containing all amphipod identified to the species level and its barcode sequence. To do so, we selected all amphipod sequences from BOLD with

the following criteria: taxonomic identification at least to the genus level and a COI sequence longer than 500bp. All sequences were aligned with MAFFT version 7 web server and trimmed to 501 bp. After the construction step, each OTU was aligned to the reference database with a confidence threshold of 90%. In order to assess the accuracy of our molecular identification, we aligned 305 sequences of 43 taxonomic identified species to our taxonomic database. As no discrepancy was observed, we confirmed the validity for OTU identification. All analyses were performed with MOTHUR (Shloss *et al.*, 2009).

#### 2.4.5.5 Analysis among locations

To provide an overview of the similarity in OTU composition between the three oceans, a Venn diagram was obtained for each threshold.



**Figure 2.1** Map of the sampling location of each sequence used in this study.

## 2.5 Results

### 2.5.1 Nucleotide diversity

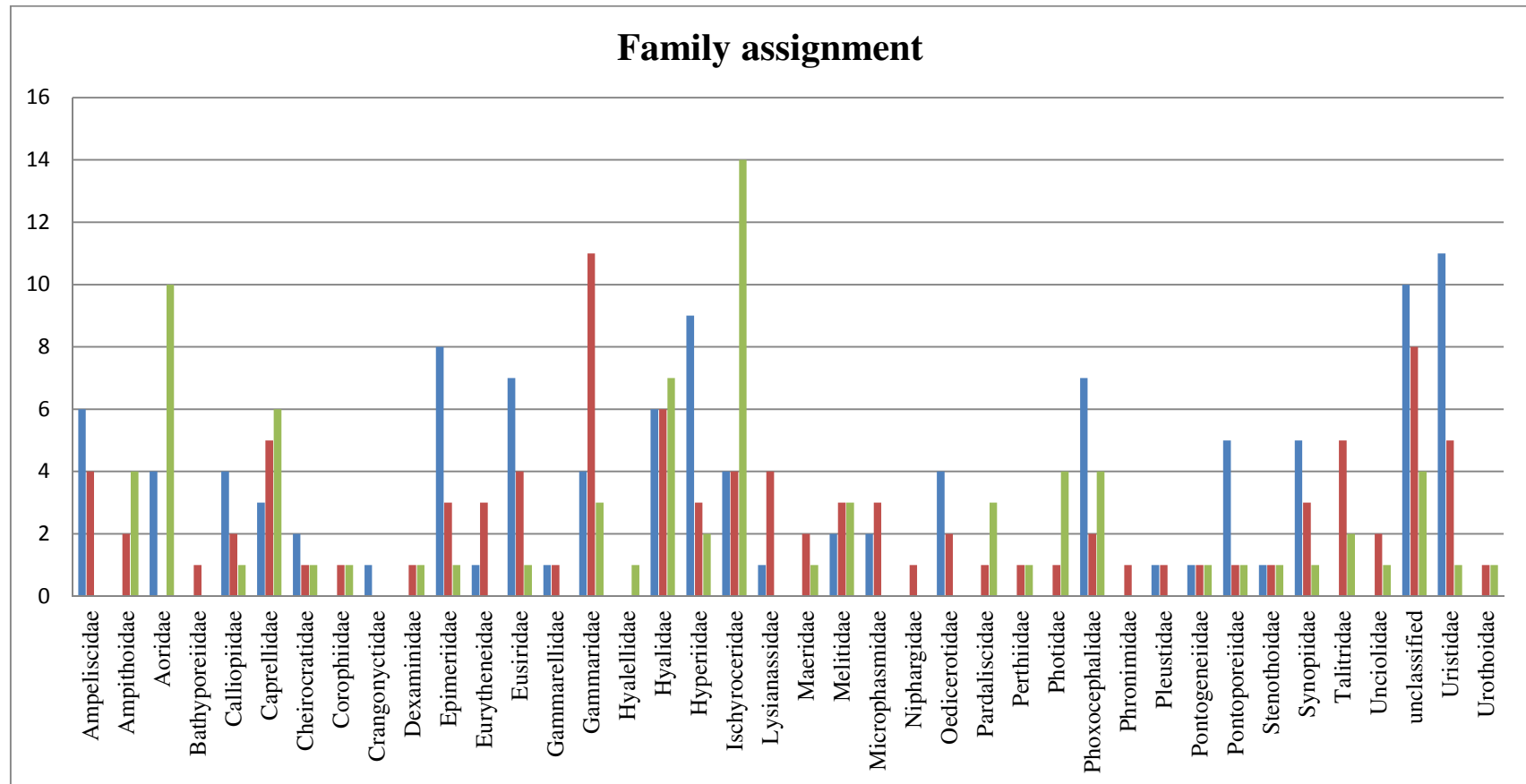
Out of a total of 3010 sequences, 2738 had a length superior of 500 bp. After trimming and cleaning steps, a 532 bp COI fragment was used for this analysis for 2557 specimens of Amphipoda. There were 405 sequences from locations within the Pacific Ocean, 1003 sequences from the Arctic region and 1149 from the Atlantic (Figure 2.1). Within the final alignment, 116 conserved sites and 405 parsimony informative sites were detected.

### 2.5.2 3% as species threshold

The 3% threshold proposed by Hebert *et al.*, (2003b) allowed the identification of 270 OTUs, for which 111 were from the Arctic, 102 from the Atlantic, and 83 from the Pacific. A large proportion of OTUs belonged to the Gammaridae family (Figure 2.2). We were unable to identify 24 OTUs to taxonomic level. At the species level, only eight species were found in the three oceans: *Parhyale hawaiensis*, *Ischyrocerus anguipes*, *Ischyrocerus latipes*, *Jassa sp.*, *Elasmopus rapax*, *Pontoginia inermis*, *Stenothoe freccanda*, and *Onisimus glacialis* (Tableau 2.1). However, no shared OTUs were found between the three oceans (Figure 2.3) suggesting that these species consist of distinct OTUs.

**Tableau 2.1** Species found in the three oceans. Species are defined with a 3% threshold.

<b>taxon</b>	<b>Arctic</b>	<b>Atlantic</b>	<b>Pacific</b>
<i>Parhyale hawaiiensis</i>	1	1	2
<i>Ischyrocerus anguipes</i>	2	2	1
<i>Ischyrocerus latipes</i>	1	1	5
<i>Jassa sp.</i>	1	1	4
<i>Elasmopus rapax</i>	1	1	1
<i>Pontogeneia inermis</i>	1	1	1
<i>Stenothoe freccanda</i>	1	1	1
<i>Onisimus glacialis</i>	3	1	2



**Figure 2.2** Distribution of OTUs by family and ocean. OTUs are defined based on 3% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue.





**Figure 2.3** Venn diagram of shared OTUs among oceans. OTUs are defined with a 3% threshold.

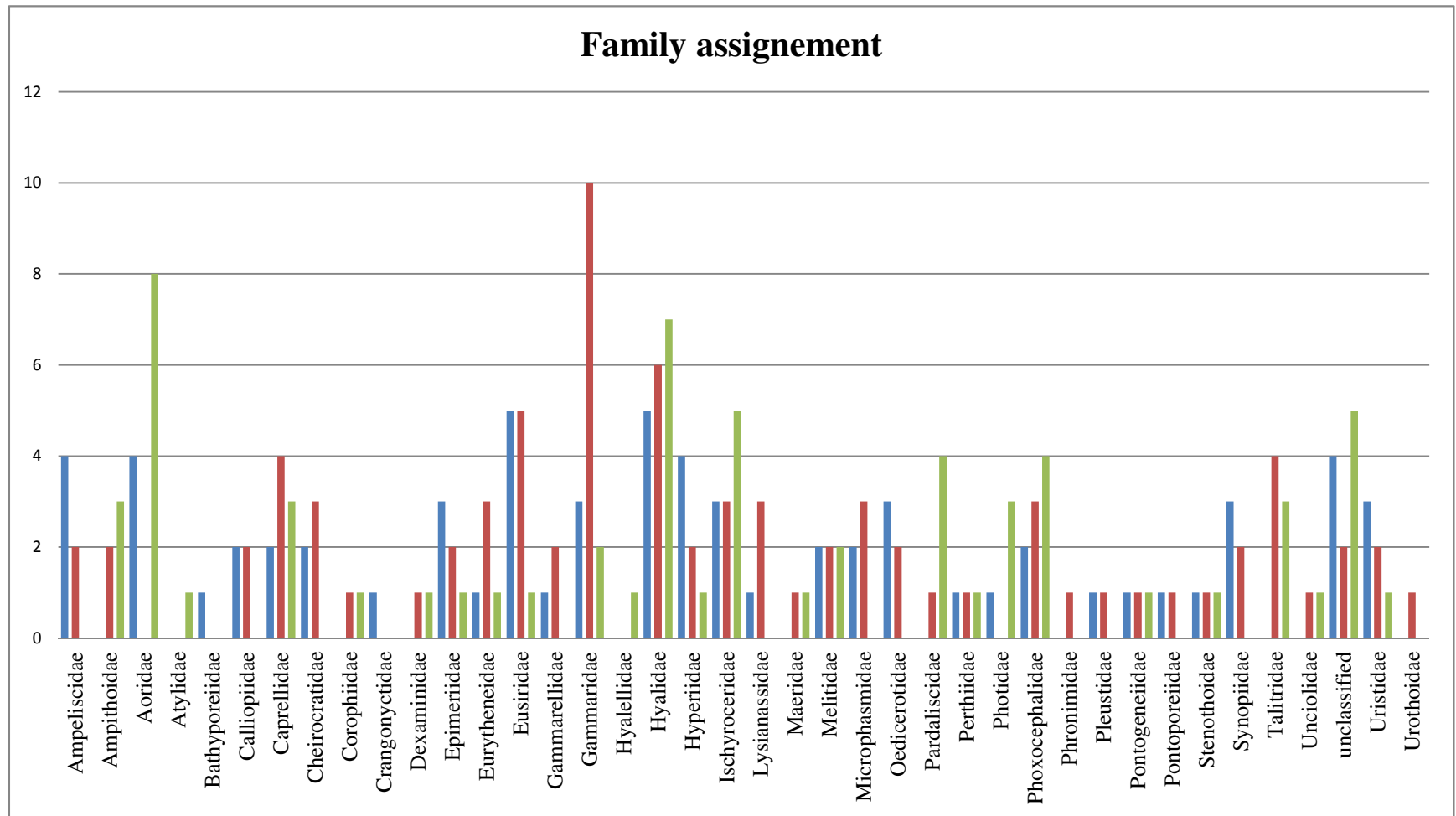
### 2.5.3 16% threshold

Under the 16% threshold of Lefébure (*et al.*, 2006) 183 OTUs were found for which 67 were in the Arctic, 83 in the Atlantic and 63 in the Pacific (Figure 2.4). Only 27 OTUs were not assigned to a taxonomic level. At the species level, only four species were common to the three oceans: *Ischyrocerus anguipes*, *Ischyrocerus latipes*,

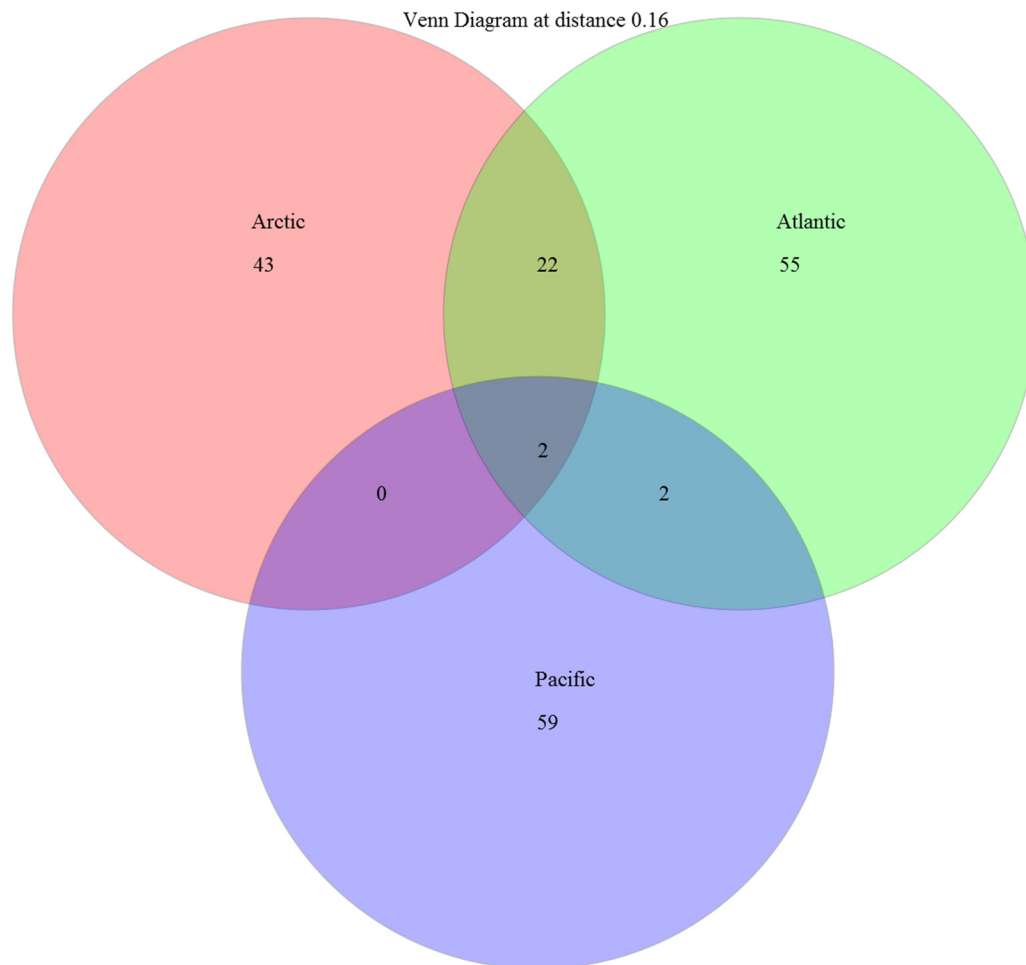
*Pontogeneia inermis* and *Stenothoe frecanda* (Tableau 2.2). Two OTUs were common to the three oceans (Figure 2.5) unlike with the 3% threshold.

**Tableau 2.2** Species found in the three oceans. Species are defined with a 16% threshold.

<b>taxon</b>	<b>Arctic</b>	<b>Atlantic</b>	<b>Pacific</b>
<i>Ischyrocerus anguipes</i>	1	1	1
<i>Pontogeneia inermis</i>	1	1	1
<i>Stenothoe frecanda</i>	1	1	1



**Figure 2.4** Distribution of OTUs by family and ocean. OTUs are defined based on 16% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue.



**Figure 2.5** Venn diagram of shared OTUs among ocean. OTUs were defined with a 16% threshold.

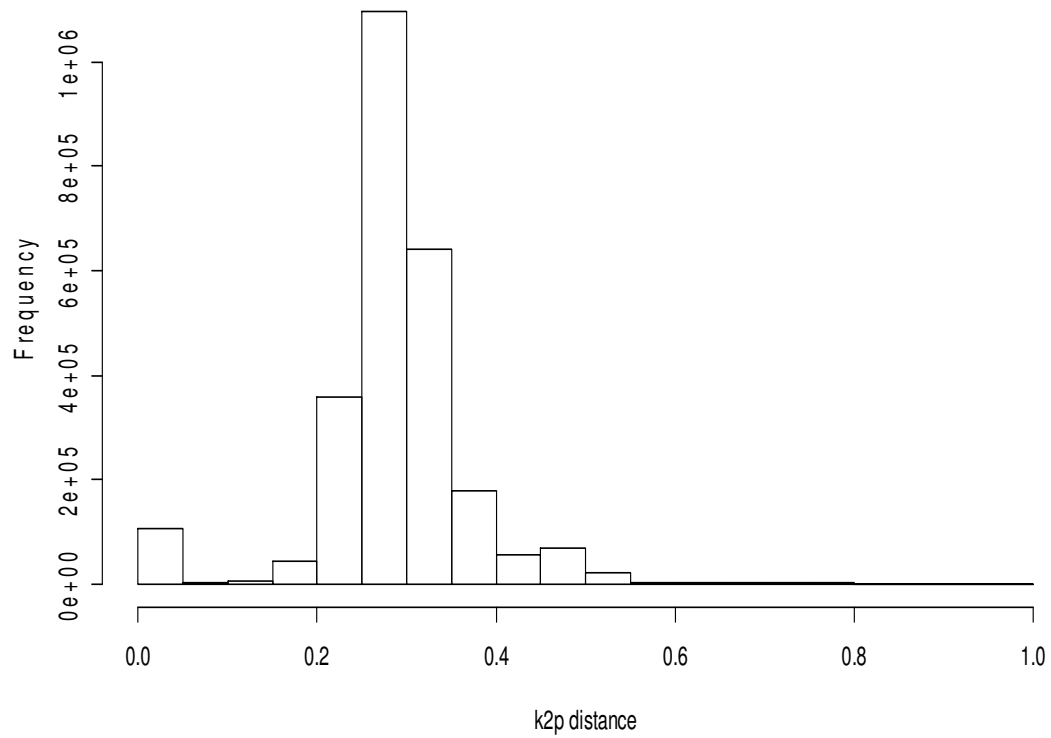
#### 2.5.4 Barcode gap detection with ABGD

The distribution of K2P genetic distances displayed two modes separated by a gap ('barcode gap') between 0.05 and 0.1 (exact value= 0.07 ABGD; Figure 2.6). Based on this value, we successfully clustered 2557 sequences into 243 OTUs, for which 102 were from the Arctic, 96 from the Atlantic, and 72 from the Pacific. A first look

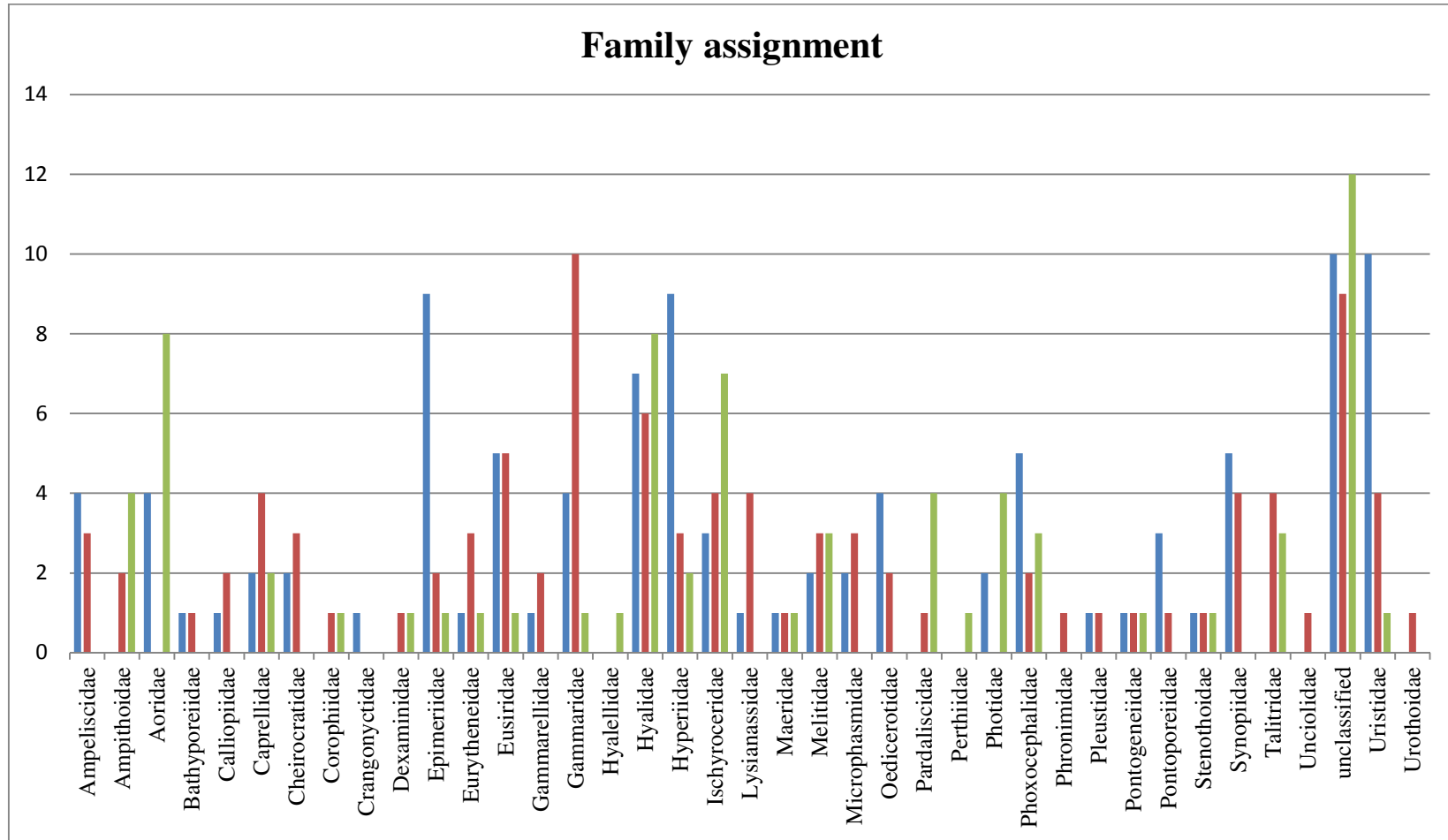
at the family identification showed that not all the locations and the families were well sampled across the three Canadian oceans (Figure 2.7). Five families (Crangonyctidae, Hyalellidae, Phronimidae, Uniciolidae and Urothidae) were represented by single OTUs in one location only. Hyalidae, Gammaridae, Aoridae and Ischyroceridae included the most OTUs with 23, 15, 13 and 13 respectively. Eleven OTUs could not be identified taxonomically to the species level. At the species level, only seven species were found in the three oceans: *Ischyrocerus anguipes*, *Ischyrocerus latipes*, *Jassa sp.*, *Elasmopus rapax*, *Pontoginia inermis*, *Stenothoe freccanda*, and *Onisimus glacialis* (Tableau 2.3) but no shared OTUs among the three oceans were found (Figure 2.8) suggesting that these species were composed by distinct OTUs.

**Tableau 2.3** Species found in the three oceans, Species are defined with a 7% threshold.

<b>Taxa</b>	<b>Arctic</b>	<b>Atlantic</b>	<b>Pacific</b>
<i>Ischyrocerus anguipes</i>	1	1	1
<i>Ischyrocerus latipes</i>	1	1	3
<i>Jassa sp.</i>	1	1	2
<i>Elasmopus rapax</i>	1	1	1
<i>Pontogeneia inermis</i>	1	1	1
<i>Stenothoe freccanda</i>	1	1	1
<i>Onisimus glacialis</i>	3	1	1



**Figure 2.6** Distribution of the K2P distances for the entire dataset.



**Figure 2.7** Distribution of OTUs by family and ocean. OTUs are defined based on 7% threshold. Atlantic OTUs are represented in red, Pacific OTUs in green and Arctic OTUs are in blue.





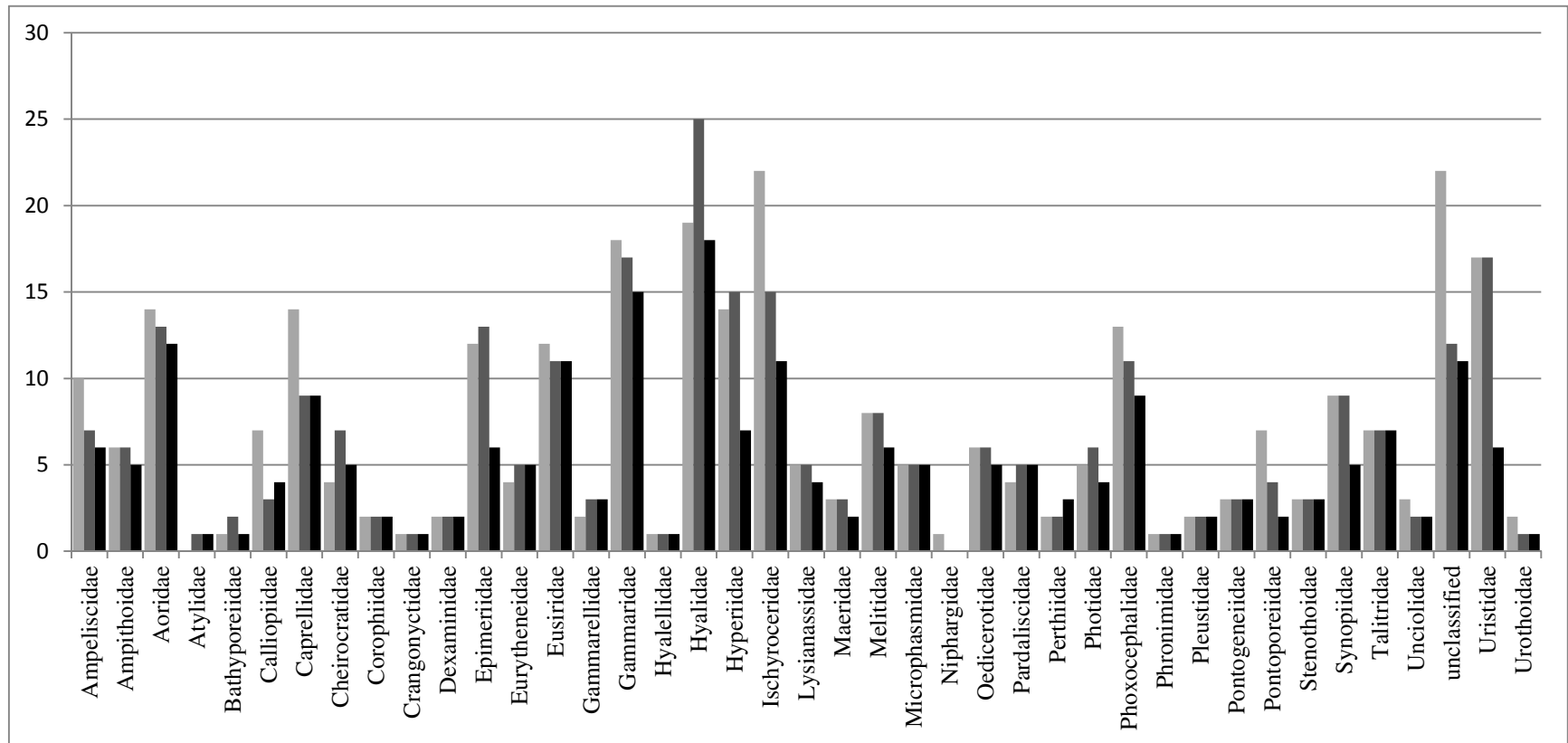
**Figure 2.8** Venn diagram of shared OTUs among ocean. OTUs are defined with a 7% threshold.

### 2.5.5 Comparison among thresholds

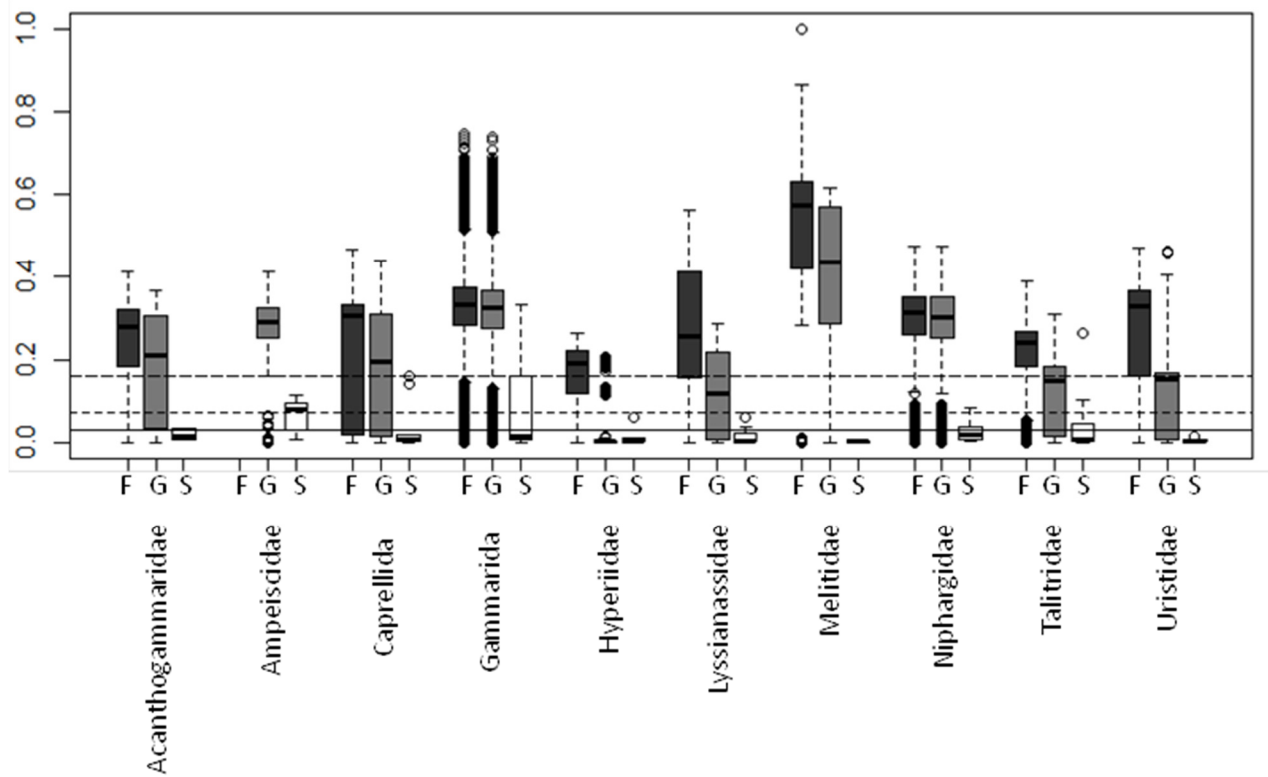
As expected, the number of OTUs decreased with an increase of the threshold value used for all families except four (Bathyporeiidae, Cheirocratidae, Epimeriidae, Hyalidae) where more OTUs were recovered with the 7% threshold than at 3% (Figure 2.9). Ten families were equally recovered (Corophiidae, Crangonyctidae, Dexaminidae, Hyalellidae, Microphasmidae, Phronimidae, Pleustidae, Pontogeneiidae, Sthenotoidea, Talitridae) regardless of the threshold.

### 2.5.6 Species threshold

Divergences were estimated for the validated dataset from 3879 sequences representing 272 species, 79 genera, and 10 families. As expected, genetic divergence increased with taxonomic rank: a higher divergence was observed at the Family level (K2P from 0 to 0.9), than at the Genus level (K2P from 0 to 0.7), and the Species level (K2P from 0 to 0.3). However, the threshold used to delimit species varied among families (Figure 2.10). The threshold proposed by Lefébure (0.16 substitution/site) discriminated well the Gammaridae species but not the other amphipods species from different genera. The lowest species divergence was observed in the Melitidae (0.001), the lowest genus and family divergences were observed in the Hyperiididae (0.26). The highest species (0.3) and genus (0.7) divergence was observed within the Gammaridae and the highest family divergence (0.9) was observed within the Melitidae.



**Figure 2.9** Number of OTUs identified to the family according to the species threshold used (3% light grey; 7% grey and 16% dark).



**Figure 2.10** Boxplot distribution of the COI K2P distances for the 10 selected families representative of the Amphipoda order at the intraspecies (S–white), intragenus (G–grey) and intrafamily (S–dark grey) levels. Median (central bar), position of the upper and lower quartiles (central box) and extremes of the data (dots) are represented. Black lines represent different thresholds used for discriminating species: long dash=0.16 substitution/site (Lefébure *et al.*, 2006); dashed line=7% of divergence (this study) and solid line=3% of divergence (Hebert *et al.*, 2003b).

## 2.6 Discussion

Inventories of marine biodiversity are much needed in the context of global changes. A decade of DNA barcoding has resulted in the acceleration of species discovery and has helped to provide partial or complete inventories lists for selected taxa (Buckin *et al.*, 2011; Raupach and Radulovici, 2015). The ability to assign species identities to DNA sequences depends on the availability of comprehensive DNA reference libraries such as BOLD (Trebitz *et al.*, 2015). Our study provides a useful example of how one can retrieve information in BOLD database to produce an exhaustive first step inventory of specific animal groups or specific regions that are otherwise lost among the different publicly available projects. We showed that widespread species are in fact composed of different OTUs suggesting the presence of cryptic species. This implies that a careful look on taxonomic keys is needed for amphipods. Moreover, we found that using a single threshold for species delimitation is not always accurate. Here we discuss our results, the limits of our approach, and future research perspectives for the molecular taxonomy in a context of biodiversity change.

### 2.6.1 Arctic diversity

Our study, based on the analysis of more than 2550 sequences distributed throughout the three Canadian oceans, indicates the presence of 243 provisional amphipod species. We recovered 102 putative Arctic species representing circa 85% of the known amphipods inventory in the Arctic (from the Chukchi Sea) (Sirenko, 2001). We recovered 96 potential species in the Atlantic and less in the Pacific (72 putative species). In addition, our results indicate a higher similarity between Arctic and Atlantic oceans (22 OTUs shared) than between Arctic and Pacific (no OTUs shared). This suggests that the Atlantic Ocean contributed significantly to the recolonization of the Arctic. Marine arctic fauna is mostly derived from recent and repeated colonisations from both Pacific and Atlantic species after Pleistocene glaciation events or multiple Bering Strait openings (reviewed in Vermeij, 1991 and in Hardy *et*

*al.*, 2011). Similarity between Arctic and Atlantic fauna has also been reported in polychaetes (Carr *et al.*, 2011). The Pacific harbors the highest number of OTUs compared to the number of sequences available (405), suggesting a higher diversity compared to other oceans, which is congruent with other studies (Vermeij, 1991; Carr *et al.*, 2011). The absence of shared OTU between Arctic and Pacific oceans suggests the presence of a barrier restricting exchanges between these oceans. Moreover, the fact that the retrieved Pacific OTUs were not found elsewhere confirms the isolation of Pacific taxa from colder Arctic waters. Ocean currents and temperature differences between these two regions might act as barrier to the establishment of Pacific species into the Arctic Ocean. For example, in the copepod *Calanus glacialis*, Nelson *et al.*, (2009) have shown isolation between Pacific and Arctic populations. Further sampling of the Pacific region is needed to confirm this isolation as our results might be biased by differential sampling efforts between the Pacific and the Atlantic oceans.

We found a relatively higher proportion of benthic species belonging to the intertidal or infralittoral family of Ischyroceridae, Hyalidae and Gammaridae families (McLaughlin *et al.*, 2005; Serejo and Sittrop, 2009) than pelagic species. In the unique pelagic amphipod family (e.g. Hyperiididae), we recovered 15 OTUs of which the majority were Arctic. In the Eurasian arctic waters inventory, Sirenko (2001) referenced the presence of eight Hyperiididae species. Here, we found nine potential species (9 OTUs) in the Arctic waters. This result suggests that pelagic diversity is not well known maybe due to sampling difficulties and further efforts are needed to better characterized ocean diversity. As sampling in the Arctic is quite challenging, it is most likely that rare taxa were not included in our analyses (Appendice A). Although, our study does not include depth information, it will be also interesting to investigate diversity of OTUs among ocean according to depth to get a more precise picture of the marine biodiversity.

We were not able to identify all OTUs to the species level. Several explanations can be considered. First, during the taxonomic reference creation, we discarded more than a half of the sequences due to pseudogenes or short sequences. Secondly, the 15 516 sequences of amphipods available in BOLD, correspond to 1 514 species which is under the number of amphipods estimates (Appeltans *et al.*, 2012). Despite these, using barcoding techniques has provided useful information on the amphipod biodiversity in the three Canadian oceans. This approach can benefit the study of oceanic Arctic region which is one of the least studied (Archambault *et al.*, 2010; Raupach and Radulovici, 2015; Wassman *et al.*, 2015). Moreover, we also showed that widespread marine amphipods species are composed of different OTUs, suggesting an underestimated diversity. More and more studies have also showed that widespread marine species were in fact cryptic species complex (Goetze, 2005; 2010; Carr *et al.*, 2011; Lörz *et al.*, 2012; Havermans *et al.*, 2013), suggesting the utility of combining different types of data (e.g. molecular and morphological) to identify species.

### 2.6.2 Single threshold

Delimiting species relies on a threshold over which species belong to the same or to two different species. The 3% of divergence (Hebert *et al.*, 2003b) is a commonly used threshold in the literature. Based on the interspecific distance distribution among amphipod families, we found this threshold can discriminate most amphipod species, but is not suitable for all. Two authors have proposed the use of a higher threshold (16% -Costa *et al.*, 2009; Lefébure *et al.*, 2006) for crustaceans. By including ten amphipod families, we have showed that the use of 16% is appropriate for the discrimination of species of Gammaridae but is too high to discriminate species from other families. Based on the automatic barcode gap detection, we estimated a threshold of 7% for discriminating amphipod species, which is intermediate between the previous thresholds proposed. This threshold seems appropriate for the majority of amphipod species but again based on diverse family distances distribution, this

threshold will underestimate Gammaridae or Talitridae species. Instead, we found that for discriminating amphipod species, a threshold specific to each family will be more appropriate. However, we should emphasize that not all amphipod families are equally studied and less than 10 sequences are not enough to estimate intraspecific distances (Meier and Zhang, 2008; Zhang *et al.*, 2010). For example Gammaridae is one of the best studied amphipod family for which a large amount of molecular data is available but other amphipod families such as Hyperiididae have not received the same attention. Furthermore, by restricting our distance distributions on better sequenced families, we focused on 10 amphipod families that represent less than a quarter of all amphipod families (Martin and Davis, 2001). Nevertheless, our analysis provides the first molecular attempt to determine species threshold in amphipods. Increasing the DNA barcoding effort on the other amphipod families will help to refine our species threshold.

We conducted our species clustering analyses on distance based methods instead of tree-based methods like General mixed Yule model (GMYC). Several authors emphasize the importance of combining several methods in species delimitation (Carstens *et al.*, 2013). Comparison between ABGD and GMYC show mostly congruency with a higher OTU content for the GMYC methods (Kekkonen and Hebert, 2014 and references therein). Tree-based methods are sensitive to error in the tree provided (Kekkonen and Hebert, 2014). Phylogenetic reconstruction of Amphipoda is not fully resolved yet (Browne *et al.*, 2007) and might be problematic with the GMYC method.

## 2.7 Conclusions

Our analyses have contributed to the assessment of marine arctic amphipod biodiversity in revealing potential cryptic species and in showing the sharing of OTUs between Arctic and Atlantic amphipods. Moreover, thanks to the increasing barcode data available in amphipods, we were able to show that threshold value for



species identification in amphipods needs to be estimated for each family. It has become evident that species definition should not be restricted to a COI sequence but should include other information such as ecological niche. Moreover, with global warming, Arctic waters are undergoing changes with recent reports of southern amphipod species moving up north (Kraft *et al.*, 2013b). Pursuing arctic amphipods studies with DNA barcodes will ultimately lead to a better understanding of biodiversity and the mechanisms of speciation in marine environments.

## 2.8 Acknowledgements

We wish to thank the officers and crew of the CCGS Amundsen during sampling efforts in 2011. We are also grateful to the Natural Resources Institute of Nuuk, Greenland for providing samples. We thank Dr. A. Comeau for MOTHR presentation and Dr. K. Dionne for comments on a previous version of this manuscript.

## CHAPITRE III

### PHYLOGENY OF THE GENUS *THEMISTO* IN THE NORTHERN HEMISPHERE

Astrid Tempestini<sup>1</sup>, Louis Fortier<sup>2</sup>, Alexei Pinchuck<sup>3</sup> et France Dufresne<sup>1</sup>

Manuscrit à soumettre à *Marine Biology*

<sup>1</sup> Université du Québec à Rimouski, Département de biologie et Centre d'études nordiques, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1

<sup>2</sup> Université Laval, Département de Biologie, Pavillon Vachon., 1045, avenue de la Médecine, Québec, QC Canada G1V 0A6

<sup>3</sup> School of Fisheries and Ocean Sciences, University of Alaska, 17101 Point Lena Loop Rd, Juneau, AK 99801, USA

Contribution des auteurs (initiales)

Mise en place des protocoles expérimentaux : AT

Récolte des données : AT

Analyse des données : AT

Interprétation des résultats : AT,FD

Rédaction du manuscrit : AT

Commentaires et correction du manuscrit : AT, FD, LF, AP

Financement: FD, LF

### 3.1 Résumé

Les Amphipodes sont un ordre des crustacés caractérisés par un développement direct. Parmi eux, la famille des Hyperiidæ est le seul groupe pélagique et parasitaire. Dans les hautes latitudes, deux espèces d'hyperiidæ appartenant au genre *Themisto* sont des acteurs importants de l'écosystème. Malgré leur abondance, peu d'information existe sur leurs liens phylogénétiques. Dans cette étude, nous présentons la première phylogénie moléculaire basée sur plusieurs loci du genre *Themisto* de l'hémisphère Nord. Nous avons réalisé des phylogénies moléculaires en utilisant des méthodes bayésiennes et de maximum de vraisemblance sur sept espèces de *Themisto* provenant des océans Atlantique, Arctique et Pacifique. Pour ce faire, nous avons utilisé quatre loci : un fragment du gène mitochondrial du cytochrome oxydase I (COI), trois fragments de gènes nucléaires Histone 3, et deux fragments des gènes ribosomiaux 18S ADNr et 28S ADNr. Nous proposons aussi une mise à jour de la phylogénie moléculaire de la famille Hyperiidæ. Nos résultats sont commentés sous la perspective de la spéciation.

**Mots-clés : Phylogénie moléculaire; Hyperiidæ; *Themisto*; Spéciation; Marin; Holozooplankton**

### 3.2 Abstract

Amphipods are a highly speciose order of crustaceans with a life cycle characterized by direct development and no larval stage, making them interesting models for marine speciation studies. The Hyperiididae family is a strictly pelagic parasitic group of Amphipoda. In northern latitudes, two hyperiid species belonging to the genus *Themisto* are important components of ecosystems in term of abundance and biomass but little is known about their genetic relationships. Here, we present the first multi-loci molecular phylogenetic assessment of the *Themisto* genus in the Northern Hemisphere. We performed Bayesian and Maximum Likelihood reconstructions based on three nuclear loci (18S rDNA, 28S rDNA, and Histone 3) and mitochondrial cytochrome oxidase I data on seven specimens of *Themisto* from the Pacific, Arctic and Atlantic Oceans. In addition, we also provide an updated molecular phylogeny of the Hyperiididae family. Our results are discussed in the light of marine speciation.

**Key-words:** Molecular phylogeny; Hyperiididae; *Themisto*; Speciation; Holozooplankton; Marine

### 3.3 Introduction

Marine crustaceans are notoriously difficult to identify to the species level by traditional approaches due to their enormous morphological diversity (Radulovici *et al.*, 2009), the frequent presence of morphological homoplasy (Knowlton, 1993; Hurt *et al.*, 2013), and also because precise taxonomic description is lacking for many groups (Knowlton, 1993). The use of cytochrome oxidase I DNA-barcoding (COI) Hebert *et al.*, 2003a) has proven an efficient tool to flag cryptic species in crustaceans (Radulovici *et al.*, 2009; Costa *et al.*, 2007; Witt *et al.*, 2006; Lörz *et al.*, 2012) and has also been instrumental in resolving local phylogenies (Costa *et al.*, 2007). However, the reconstruction of a phylogeny based on a single mitochondrial gene is not always accurate with the species phylogeny because of the presence of mitochondrial introgression (Zhang and Hewitt, 1996; Song *et al.*, 2008) and due to the uniparental inheritance of mitochondria (Ballard and Whitelock, 2004; Holzer *et al.*, 1997; Nichols, 2001). Hence, using data from both mitochondrial and nuclear genes to reconstruct species phylogeny is an invaluable tool to increase our understanding of species boundaries and speciation in this important group.

Amphipods are a highly speciose order of crustaceans of the class Malacostraca, belonging to the superorder Peracarida, whose life cycle is characterized by direct development (Vainölä *et al.*, 2008), which limits their capabilities of long dispersal and increases the opportunity for speciation in this group (Radulovici, 2012). The order Amphipoda is further subdivided into three large benthic sub-orders: Gammaridea, Caprellidea and Ingolfiellidea, and one strictly pelagic taxon: Hyperiidia, (Vinagrov *et al.*, 1996; Martin and Davis, 2001). Hyperiid amphipods (suborder Hyperiidia) are a substantial component of zooplankton communities throughout all oceans (Bowman and Grüner, 1973; Auel *et al.*, 2002; Yamada *et al.*, 2004), often having close associations with gelatinous zooplankton (Laval, 1980; Gasca *et al.*, 2004). Currently, there are about 250 species of hyperiids described, of

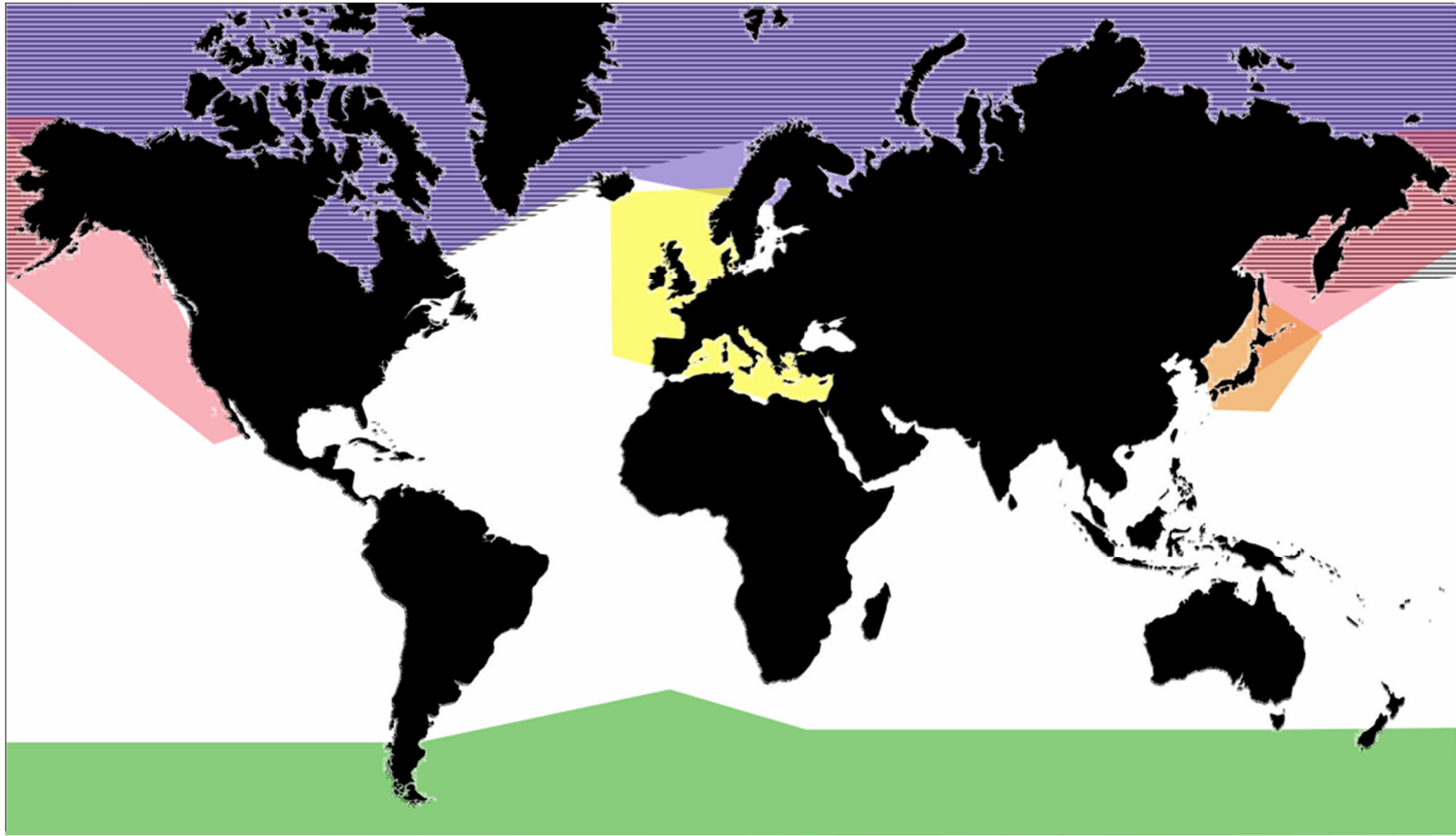
which only nine species are reported from the Arctic Ocean (Vinogradov and Semenova, 2009).

In higher latitudes of both hemispheres, species belonging to the hyperiid genus *Themisto* are important in terms of abundance and biomass (Bowman, 1960; Auel *et al.*, 2002). They are mainly carnivorous, and thus play key roles in pelagic ecosystems of polar oceans, linking lower and higher trophic levels (Auel *et al.*, 2002; Yamada *et al.*, 2004; Watts and Tarling, 2012). The genus *Themisto* is thought to include seven species: *Themisto abyssorum*, *T. australis*, *T. compressa*, *T. gaudichaudii*, *T. japonica*, *T. libellula* and *T. pacifica* (Bowman, 1960; Zeidler, 2004). *Themisto* species show morphological, ecological (Wing, 1976; Percy, 1993; Dalpadado *et al.*, 2001; Auel *et al.*, 2002; Yamada *et al.*, 2004; Marion *et al.*, 2008; Prokopowicz *et al.*, 2009; Kraft, 2013a), and geographical differences (Figure 3.1; Bowman, 1960; Wing, 1976; Yamada *et al.*, 2004). Since the first description of this genus in the 19<sup>th</sup> Century by Guérin, the taxonomy of the genus *Themisto* has undergone a series of changes (reviewed in Schneppenheim and Weigmann-Haas, 1986; Zeidler, 2004). Three generic names (*Parathemisto*, *Euthemisto* or *Themisto*) have been used in the past by taxonomists to classify the species of the genus. These changes involved splitting species previously belonging to the same genus into different genera. *Themisto* species share similar morphological characteristics (Yamada *et al.*, 2004), making taxonomic identification often difficult.

The first molecular phylogeny of *Themisto* used allozymes and morphological data (Schneppenheim and Weigmann-Haas, 1986). This study revealed the presence of four Atlantic Ocean *Themisto* species: *T. abyssorum*, *T. compressa*, and *T. libellula* in the Northern Hemisphere, and *T. gaudichaudii* in Antarctic waters. However, not all the northern *Themisto* species were included in their phylogeny, so nothing is known about the position of *Themisto japonica* and *T. pacifica* within this genus. More recently, Hurt *et al.*, (2013) have reconstructed the phylogeny of the Hyperiididae

family using COI as well as three nuclear genes but have included only two *Themisto* species.

In this paper, we provide the first molecular phylogeny of the genus *Themisto* using COI and three nuclear genes (Histone 3, 18S rDNA, 28SrDNA). We test for the presence of cryptic species in the Northern hemisphere by sampling *Themisto* individuals from several locations in the Arctic Ocean. Our study also aimed to assess the relationship between the three genera *Hyperia*, *Hyperoche*, and *Themisto* using cytochrome oxidase I (COI) gene and update the phylogeny of Browne *et al.*, (2007). The resulting phylogenies (gene-trees and species-tree) are interpreted in the context of marine speciation and cryptic diversity.



**Figure 3.1** Global distribution of *Themisto* species (modified from Bowman, 1960; Wing, 1976 and Yamada *et al.*, 2004). Each color represents the distribution of a species: in green *T. gaudichaudii*, in orange *T. japonica*, in pink *T. pacifica*, in purple *T. abyssorum*, in yellow *T. compressa* and in solid line *T. libellula*.



### 3.4 Material and Methods

#### 3.4.1 Samples collection

Hyperiid individuals were collected with 500 and 750  $\mu\text{m}$  mesh size horizontal or vertical nets from various locations in the Arctic Canadian Archipelago during the ArcticNet expedition in 2011(Figure3.1) *Themisto* from the Bering area were collected in 2012 and 2013 and provided by A. Pinchuk. Specimens from Japan were collected in 2012 from the Oyashio and Kuroshio region and provided by Hiroomi Myiamoto from the University of Tokyo, Atmosphere and Ocean Research Institute.. All amphipods were preserved in 95% ethanol until analyses. Individuals that were not identified to the species or genus level were also included in our phylogeny and labelled as Unknown. Other hyperiid sequences from the phylogeny of Hurt *et al.*, (2013) were mined on Genbank (Tableau 3.1). An outgroup species belonging to another genus (*Cystisoma pellucida*) was used in order to root the hyperiidae phylogeny. For the *Themisto* phylogeny, *Hyperoche capucinus* and *Phrosina semilunata* were used as outgroups.

**Tableau 3.1** List of species used for this study. Coordinates are expressed in decimal degree (DD).

Species	Genbank accession number	Latitude (DD)	Longitude (DD)	Reference
<i>Themisto japonica</i> (1)	-	37.8	142.13	This study
<i>Themisto japonica</i> (2)	-	37.8	142.13	
<i>Themisto libellula</i>	-	71.42	-101.41	
<i>Themisto abyssorum</i>	-	74.06	-83.22	
<i>Themisto sp. (PWS)</i>	-	60.73	-147.10	
<i>Themisto sp. (SEB)</i>	-	54.40	-166.67	
<i>Themisto compressa</i>	-	69.81	19.86	
<i>Unknown (bp113inc1)</i>	-	70.43	-133.38	
<i>Unknown (30Ihy1)</i>	-	74.06	-83.22	
<i>Unknown (314inc8)</i>	-	68.59	-106.36	
<i>Unknown (323inc1)</i>	-	74.09	-80.34	
<i>Cystisoma pellucida</i>	EF989676.1	36.43	124.07	
<i>Hyperia macrocephala</i>	KC754383.1	-60.93	-52.99	
<i>Phrosina semilunata</i>	EF989671.1	21.27	158.14	
<i>Hyperietta parviceps</i>	EF989686.1	19.35	156.00	
<i>Hyperioides longipes</i>	EF989685.1	27.34	79.54	
<i>Phronimopsis spinifera</i>	EF989683.1	27.34	79.54	
<i>Lestrigonus schizogeneios</i>	EF989684.1	19.35	156.00	
<i>Hyperoche capucinus</i>	EF989665.1	-60.93	-52.99	
<i>Hyperoche martinezi</i>	EF989668.1	36.61	-122.16	
<i>Hyperoche medusarum</i>	EF989667.1	35.99	-122.99	
<i>Iulopis loveni</i>	EF989669.1	27.34	79.54	
<i>Themisto pacifica</i>	EF989664.1	36.71	-124.11	

**Tableau 3.1** Continued

Species	Genbank accession number	Latitude (DD)	Longitude (DD)	Reference
<i>Hyperioides sibaginis</i>	HM053497.1	NA	NA	Genbank
<i>Hyperietta vosseleri</i>	HM053498.1	NA	NA	
<i>Lestrigonus bengalensis</i>	HM053499.1	NA	NA	
<i>Hyperiella dilatata</i>	KC754384.1	NA	NA	
<i>Hyperiella antarctica</i>	KC754382.1	NA	NA	
<i>Themisto gaudichaudii</i> (1) <b>MaHyp000</b>	DQ889133.1	NA	NA	Costa <i>et al.</i> , 2009
<i>Hyperia galba</i> (1) <b>MaHyp002</b>	DQ889153.1	NA	NA	
<i>Hyperia galba</i> (2) <b>MT00515</b>	KT209327.1	NA	NA	Raupach <i>et al.</i> , 2015
<i>Themisto gaudichaudii</i> (2)	KC754380.1	NA	NA	Chen <i>et al.</i> , 2013

#### 3.4.2 DNA extractions

Morphological identifications were performed using Bowman and Gruner (1973), Vinogradov *et al.*, (1996), and Zeilder (2000) taxonomic keys. Some individuals were only identified to the genus level because of taxonomic uncertainty. Pereopods or complete individuals were used for DNA extractions using EZNA Tissue kit (Omega Bio-tek) following the manufacturer protocol. Quality of DNA was evaluated on a Nanodrop spectrophotometer 8000 UV-Vis (Thermo scientific).

### 3.4.3 PCR amplifications and alignments

We chose one mitochondrial gene (cytochrome c oxidase I-COI), one nuclear protein-coding gene (Histone 3- H3) and two nuclear ribosomal genes (18S rDNA and 28S rDNA) because they have been used in previous studies for resolving phylogenetic history at different taxonomic levels (Giribet *et al.*, 2001), and also because they have been successfully employed to investigate the Hyperiid family phylogeny (Hurt *et al.*, 2013). A fragment of 658 bp of COI was amplified for all species using the LCO1490/HCO2198 primers (Folmer *et al.*, 1994), according to the protocol of Dionne *et al.* (2011). For the nuclear genes, the reaction mix contained 1X PCR buffer without MgCl<sub>2</sub>, 2.2- 4mM MgCl<sub>2</sub> (Table 3.2), 0.5 mM dNTPs, 0.4 µM of each primer, 1.5 U of Taq DNA polymerase (Life technologies), DNA template (around 40–80 ng) and water for a final volume of 25 µl. PCR settings for each gene (Table 3.2) consisted of an initial denaturation of 3 min at 94°C, followed by 40 cycles of 40 s at 94°C, 40 s at the annealing temperature and 60 s at 72°C, and a final elongation of 5 min at 72°C. All PCR products were verified on a 1.5% agarose gel and then were direct sequenced by Génome Québec (McGill University, Canada). All sequences have been deposited in GenBank under the accession numbers xxx xxx. All sequences were checked manually on MEGA 5 (Tamura *et al.*, 2011). Multiple sequences alignment were performed with MUSCLE (Edgar, 2004) and verified by eye. Gblock 0.91 (Castresana, 2000) was used due to the presence of poorly aligned regions in the 18S rDNA and 28S rDNA datasets.

**Tableau 3.2** List of PCR settings used.

Gene	Primers	Sequence (5'-3')	Annealing Temperature	[MgCl <sub>2</sub> ]	Reference
COI	LCO1490	GGTCAACAAATCATAAAGATATTGG	46°- 51°	2.2 mM	Folmer <i>et al.</i> , 1994
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA			
Histone 3	H3aF	ATGGCTCGTACCAAGCAGACCGC	54°	4 mM	Colgan <i>et al.</i> , 1998
	H3aR	ATATCCTTAGGCATAATAGTGAC			
18S rDNA	E4f	CTGGTTGATTCTGCCAGT	48°	2.4 mM	van Hannen <i>et al.</i> , 1999
	E1628R	CGACGGGCGGT GTGTA			
28S rDNA	28sf-hyp	TGAGTAGTGGCGAGCGAAAAGGG	54°	2.2 mM	This study
	28sr-hyp	GACGTAAGTCTCACGGTCA			
	28sf1120	AGGCCCATTCATCAGTCAAC			
	28sr2950	AGGCTGCAAACCTATGAGACC		2.2 mM	

### 3.4.4 Phylogenetic analyses

#### 3.4.4.1 The Hyperiidæ family

Phylogeny was performed based on two methods: Maximum Likelihood (ML) and Bayesian approaches (BI) on individual from Table 3.1. ML method evaluates several trees and parameters by selecting those that maximize the likelihood, meaning those that render the observed data the most plausible, whereas in Bayesian reconstruction, these parameters are considered as random variables with statistical distributions (Zhang and Rannala 2012). To estimate the branch support for the ML trees, an approximate likelihood ratio test (aLRT) like Shimodaira-Hasegawa was calculated as described in Anisimova and Gascuel (2006) and implemented in PhyML 3.0 (Guindon *et al.*, 2010) available at <http://phylogeny.lirmm.fr/> (Dereeper *et al.*, 2010). Bayesian analyses for each dataset was performed on two runs for 5 000 000 generations; trees were sampled every 1000 generations under the appropriate evolutionary model. The first 25 % of sampled tree were discarded as burn-in. Bayesian reconstructions were performed using MrBayes 3.2.5 (Ronquist *et al.*, 2012). The posterior probabilities (PP) were calculated with the 50 % majority-rule consensus tree. Values of  $PP \geq 0.95$  were considered as strong support for the clade whereas PP values comprised between 0.85 and 0.94 were seen as intermediate support. Evolutionary models for phylogenetic analyses were selected independently for each gene using TOPALi v2.5 (Milne *et al.*, 2004) under the Bayesian Information Criterion (BIC).

Saturation at third codon positions can decrease the phylogenetic signal contained in a sequence and can lead to inaccurate phylogeny reconstruction (Xia, 2009). Hence, the presence of saturation within each gene was tested with the Xia test (Xia *et al.*, 2003) available in DAMBE 5.3.27 (Xia and Xie, 2001). Briefly, this method tests if the saturation present within the sequences can be problematic for the phylogeny reconstruction (Xia, 2009). As saturation was detected for the COI gene at the family

level and because distantly related species were compared, a Bayesian phylogeny based on amino acids was also generated under the General reversible mitochondrial protein with gamma rates model (BIC=4811.07) with MrBayes 3.2.5 (Ronquist *et al.*, 2012). To set the correct codon data-frame, we used the complete amino acid sequence of the COI gene from the species *Onisimus nanseni* (Genbank accession number: NC\_013819).

#### 3.4.4.1 *Themisto* phylogenies

Divergence at COI between each *Themisto* species was estimated using the Kimura two parameters model on all *Themisto* sequences. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA 5 (Tamura *et al.*, 2011).

ML and BI phylogenies were performed on *Themisto* from Table 3.1 for each gene with PhyML 3.0 (Guindon *et al.*, 2010) and MrBayes 3.2.5 (Ronquist *et al.*, 2012) respectively using the same parameters as described in section 3.3.4.1. Evolutionary models for phylogenetic analyses were selected independently for each gene using TOPALi v2.5 (Milne *et al.*, 2004) under the Bayesian Information Criterion (BIC).

Congruencies between mitochondrial and nuclear genes were tested using the congruence among distance matrices approach (CADM; Legendre and Lapointe, 2004; Campbell *et al.*, 2011) available in the package ape (Paradis *et al.*, 2004) in R. Briefly, congruency tests examine if there is incongruence among all distance matrices. If the null hypothesis is rejected, the data are considered congruent. The significance of the CADM test was performed with a permutation test with 1000 permutations. Kimura two parameters was chosen as a distance measure and calculated in MEGA5 (Tamura *et al.*, 2011). As no incongruence was detected between the four genes, all the sequences were concatenated and ML and BI phylogenetic reconstruction were performed on this dataset. For the Bayesian

reconstruction of concatenated datasets, separate partitions were assigned to each gene and rates were allowed to vary across partitions.

Since concatenated genes phylogeny reconstruction could lead to poor species tree estimation due to the incomplete lineage sorting (Heled and Drummond, 2010 and references therein), we used a Bayesian method to reconstruct species tree of related organisms using multiple genes and accounting for the deep coalescence process as implemented in BEST (Liu, 2008). This method is based on the assumption that genes are independent and that there is no introgression between species. It allows the estimation of time since divergence and population parameters based on unlinked genes. We used a partitioned dataset and set up all the appropriate models of evolution as determined with TOPALi v2 (Milne *et al.*, 2004).



## 3.5 Results

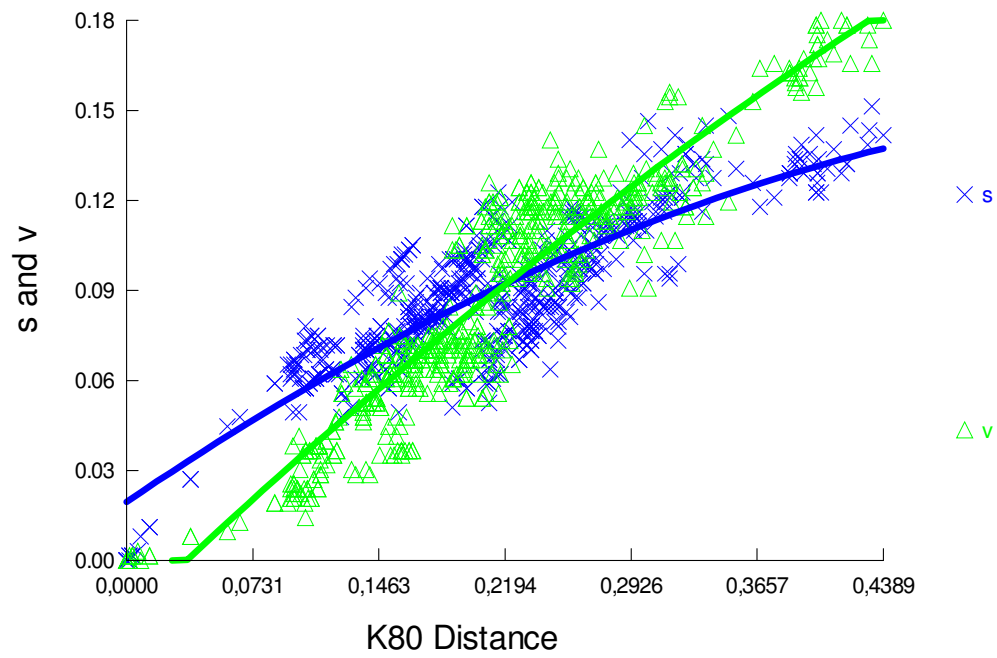
### 3.5.1 The Hyperiididae phylogeny

No indels occurred in the COI alignment of the 32 hyperiididae species. Over the 593 mutations observed, we observed 265 invariable, 284 parsimony informative sites and 170 synonymous changes. To assess the possibility of running a molecular clock, we tested the saturation among all the 32 taxa used (Figure 3.2). As the presence of a plateau was observed, indicating the loss of phylogenetic signal, the use of a molecular clock was inappropriate.

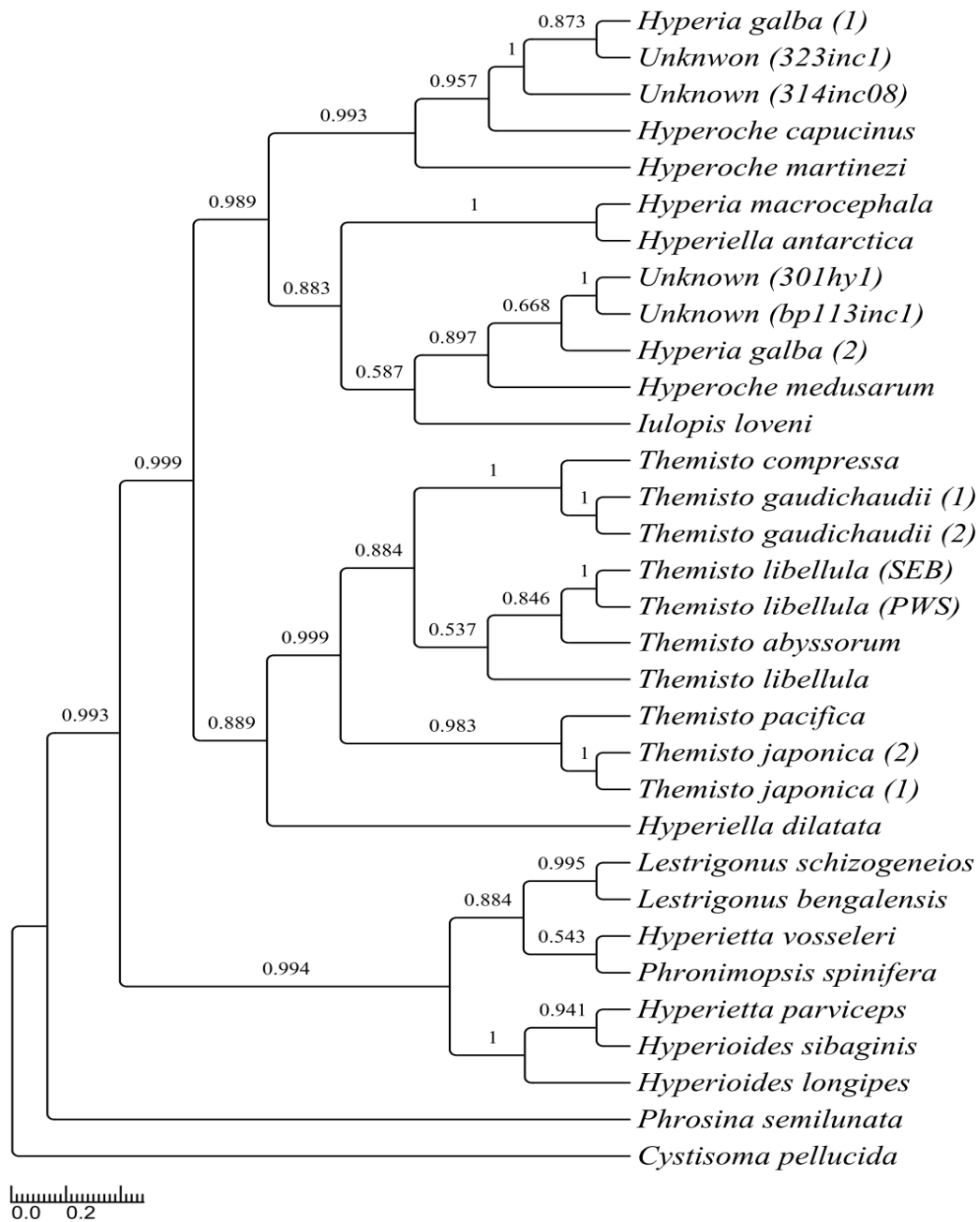
The Bayesian and the Maximum likelihood (log likelihood= -6783.77) inferences under the HKY+I+G model (BIC= 13948.88) produced highly resolved trees, with most of the clade separations having a support > 0.70 (Figure 3.4 and 3.5). Both trees have a similar topology and differ in the placement of some species and the support of some nodes. In both phylogenies, the *Themisto* genus is placed at the base of the Hyperiididae with high support (PP=1, aLTR=0.91). Within the *Themisto* genus, we observed the presence of a new species (*T. libellula* (SEB) and (PWS)) distinct from *T. libellula* provisionally named *Themisto* sp.. Both phylogenies differ within the genus *Themisto*; *Themistolibellula* (SEB) and (PWS) are placed with *T. libellula* (aLTR=0.53) in ML phylogeny rather than with *T. abyssorum* as observed in the Bayesian phylogeny (PP=0.91). In opposition with the ML phylogeny, *Hyperoche capucinus* is branching before *Hyperoche martinezi* in the BI phylogeny. Some species positions were not resolved in both phylogenies. It is the case, in particular, of the placement of *Iulopis loveni* (PP=0.6; aLTR=0) and the links among the *Hyperia* and *Hyperoche* species. There was incongruence in the placement in both phylogenies of two individuals morphologically identified as *Hyperia galba*, one is placed with an Unknown (323inc1) specimen (PP=0.88; aLTR= 0.99) and the other is placed with *Hyperoche capucinus* (PP=0.91; aLTR=0.83).

### 3.5.2 Amino acids phylogeny

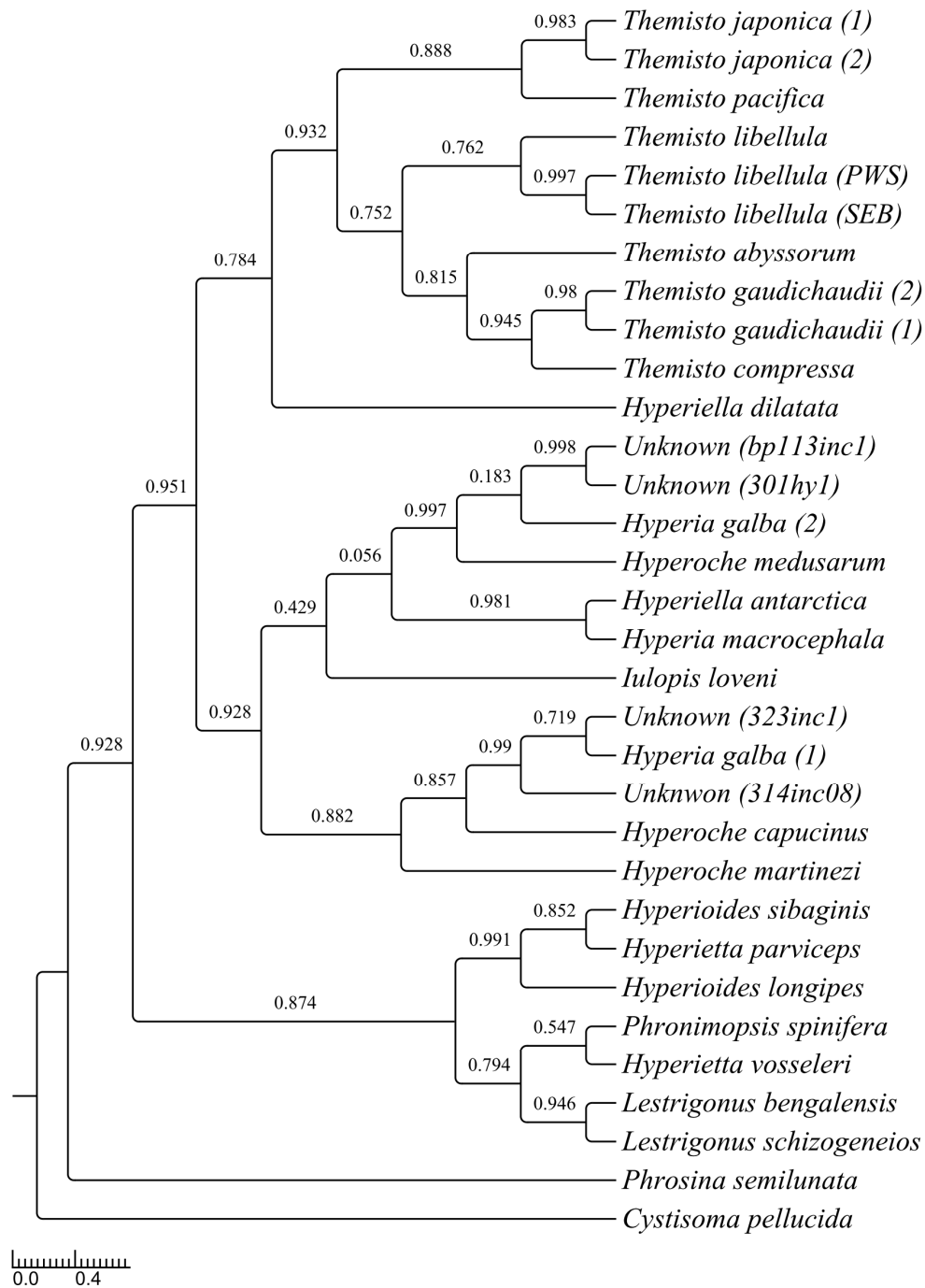
Over the 203 amino acids from the COI, 88 sites were variables and 61 were informative. Most of this variation occurred at the third codon position. The Bayesian phylogeny obtained under the Mitochondrial Reverse and gamma sites model (BIC=4290.16) is shown in Figure 3.5. The branching order of the amino- acid based tree was very similar to those based on the nucleotide sequences; the only differences between these trees were the position of some species and the unresolved phylogenetic links among *Themisto* species. *Phronimopsis spinifera* is branched with the *Lestrigonus* specimens (PP=0.7) instead of being placed with *Hyperietta vosseleri* as observed in the nucleotide phylogenies (Figure 3.3 and 3.4). The placement of *Hyperiella dilatata* was not fully resolved either in the amino acid phylogeny or in the nucleotide phylogenies. In the amino acid phylogeny, a polytomy was also observed within the genus *Themisto*, suggesting recently diverged species.



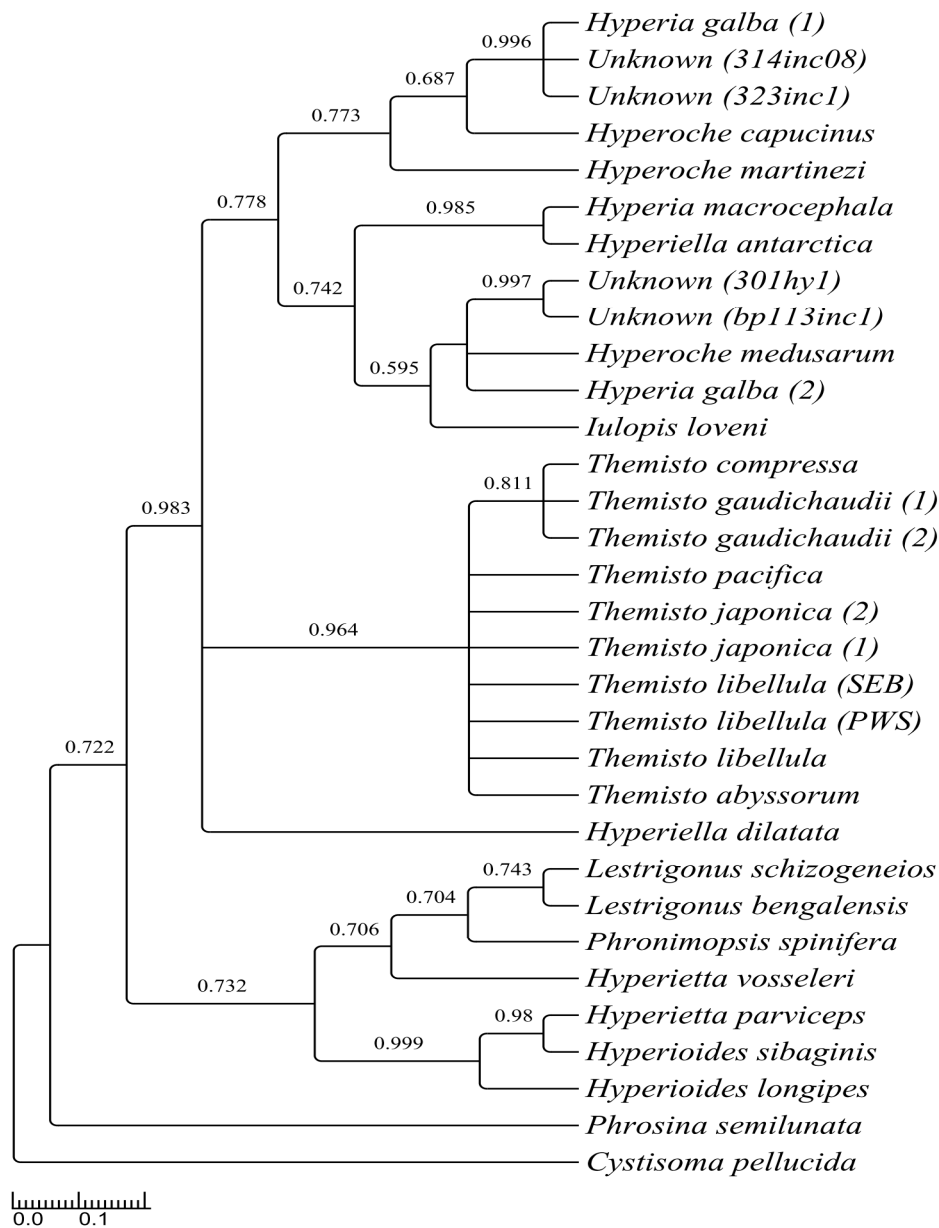
**Figure 3.2** Saturation of the cytochrome oxidase I gene estimated for the family phylogeny. s and v represent transition and transversion respectively.



**Figure 3.3** Bayesian phylogeny of the cytochrome oxidase I DNA sequences under the HKY +G+I model. Numbers represent posterior probability support.



**Figure 3.4** Maximum likelihood tree of hyperiids estimated from the cytochrome oxidase I DNA sequences, under the HKY model with gamma distribution and invariant sites considered. aLTR values are shown.



**Figure 3.5** Bayesian phylogeny of hyperiids performed with amino acids sequences of the cytochrome oxidase I gene under the general reversible mitochondrial model with a gamma distribution. Numbers represent posterior probabilities.

### 3.5.3 Single gene phylogenies

The final concatenated dataset for the *Themisto* phylogeny was 4040 bp long, including 611 bp of COI, 283 bp of H3, 1434 bp of 18S and 1712 bp of 28S characters. No indels were detected for the COI and the H3 alignments. In both 18S and 28S alignments, indels were observed and ranged from 1 to 14 bp and from 1 to 49 bp, respectively. Saturation was observed for the mtDNA COI gene but not for the three nuclear genes (Figure 3.6), indicating a loss of the phylogenetic signal for the COI dataset. The divergence estimates among *Themisto* species were around 10 %, except between *T. compressa* and *T. gaudichaudii* (3.83 %; Table 3.3). The highest divergence was observed between *T. abyssorum* and *T. gaudichaudii*, and the lowest divergence was observed between the two specimens of *T. libellula* (SEB) and *T. libellula* (PWS) (0.3%) and the two *T. japonica* (0.66%).

The ML and the BI trees obtained under the HKY+G (BIC=4965.36) for the COI gene are shown in Figures 3.7 and 3.11 along with posterior probabilities and likelihood ratios supports. Maximum likelihood and Bayesian trees produced very similar topologies, differing only in the placement of some specimens. In the ML tree, *T. abyssorum* was placed with the outgroup *Hyperoche capucincus* but the support was weak (aLTR=0.15). Instead, in the BI tree, *T. abyssorum* was regrouped with the other *T. libellula* (SEB) and (PWS) (PP=0.55). In both phylogenies, *T. libellula* was grouped with *T. japonica* and *T. pacifica* (aLTR=0.21; PP=0.52).

Multiple alignments of the Histone 3 gene showed no indels and eleven parsimony informative sites. The ML and Bayesian trees for Histone 3 obtained under the TrNef + G model (BIC=1377.94) are shown in Figures 3.8 and 3.12. Polytoomy was observed among *Themisto* in both ML and BI trees. The main difference between both trees was the placement of *T. compressa*, which in the ML tree was close to the outgroup *Hyperoche capucinus*, but the support was extremely weak (aLTR=0.01).

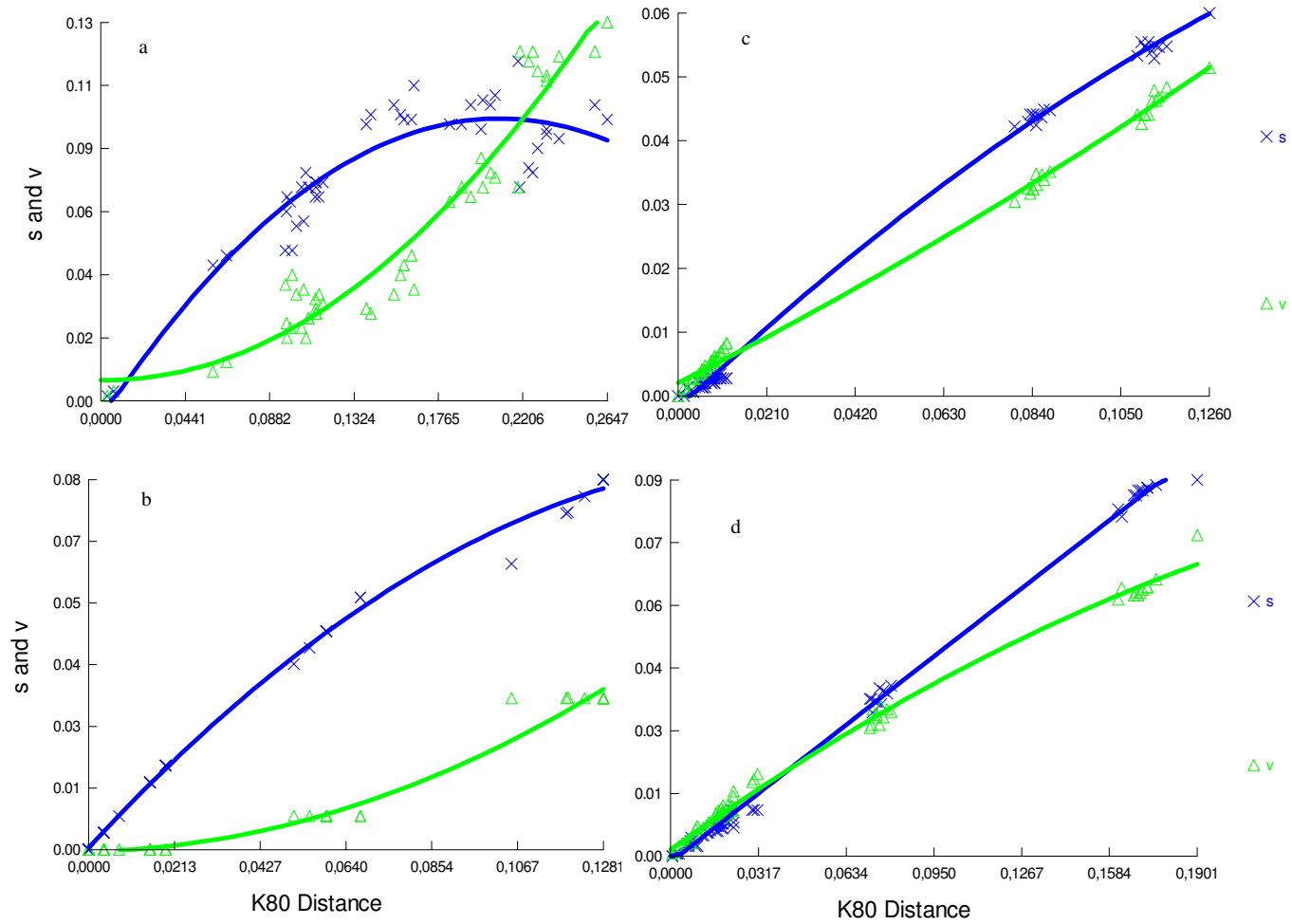
The ML and BI trees constructed under the TrNef +G model (BIC= 7063.58) for the 18s rDNA are shown in Figures 3.9 and 3.13. ML and BI produced similar tree topologies and differed only in the strength support of the nodes. aLTR supports ranged from 0.74 to 0.98 and PP ranged from 0.93 to 1. Specimens of *T. japonica* and *T. pacifica* were not grouped together in both ML and BI phylogeny.

The ML and BI trees under the K80+G model (BIC=7375.88) for the 28s rDNA is shown in Figures 3.10 and 3.14. ML and BI differed in some branching patterns. In the ML tree, one outgroup was branched with *Themisto compressa* but, as the support was extremely weak (aLTR=0.05) no consideration could be taken. Both phylogenies placed *T. libellula* with *T. abyssorum* and *T. compressa* (aLTR=0.86; PP=0.97).

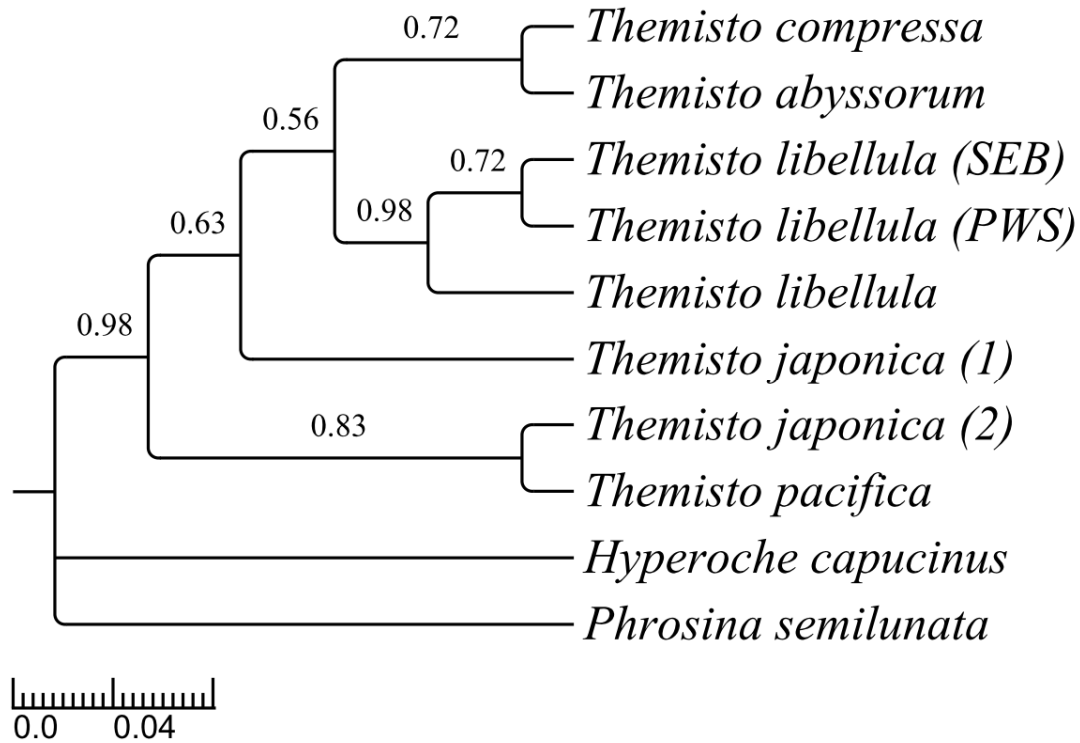


**Tableau 3.3** Estimate of divergence between *Themisto* species based on cytochrome oxidase I DNA sequences. Percentages of divergence among *Themisto* species are shown.

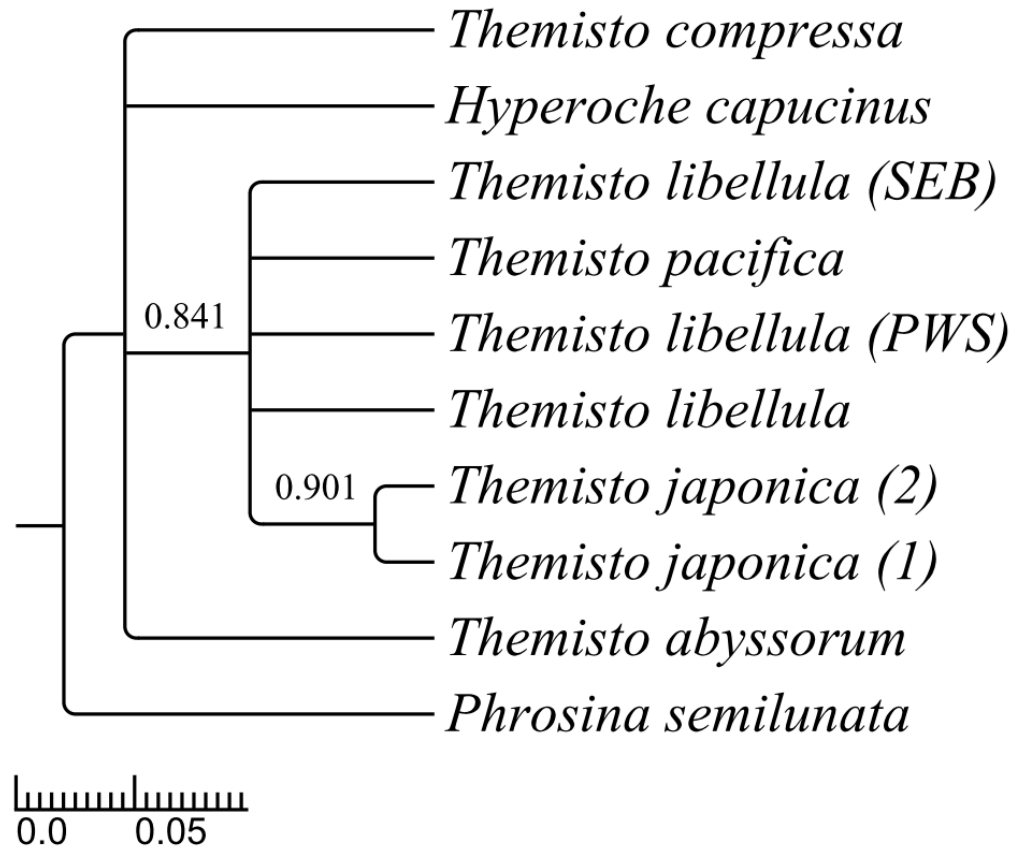
	<i>T. compressa</i>	<i>T. pacifica</i>	<i>T. japonica</i> (2)	<i>T. libellula</i>	<i>T. abyssorum</i>	<i>T. japonica</i> (1)	<i>T. libellula</i> (SEB)	<i>T. libellula</i> (PWS)	<i>T. gaudichaudii</i>
<i>T. compressa</i>	0.00								
<i>T. pacifica</i>	11.0	0.00							
<i>T. japonica</i> (2)	10.3	6.15	0.00						
<i>T. libellula</i>	11.4	11.73	10.55	0.00					
<i>T. abyssorum</i>	17.3	17.74	17.94	18.74	0.00				
<i>T. japonica</i> (1)	10.7	6.93	0.66	11.43	18.42	0.00			
<i>T. libellula</i> (SEB)	12.2	11.85	11.85	10.47	15.57	12.27	0.00		
<i>T. libellula</i> (PWS)	12.4	12.30	12.30	10.73	15.88	12.73	0.33	0.00	
<i>T. gaudichaudii</i>	3.83	11.24	12.60	12.15	19.03	13.04	13.16	13.38	0.00



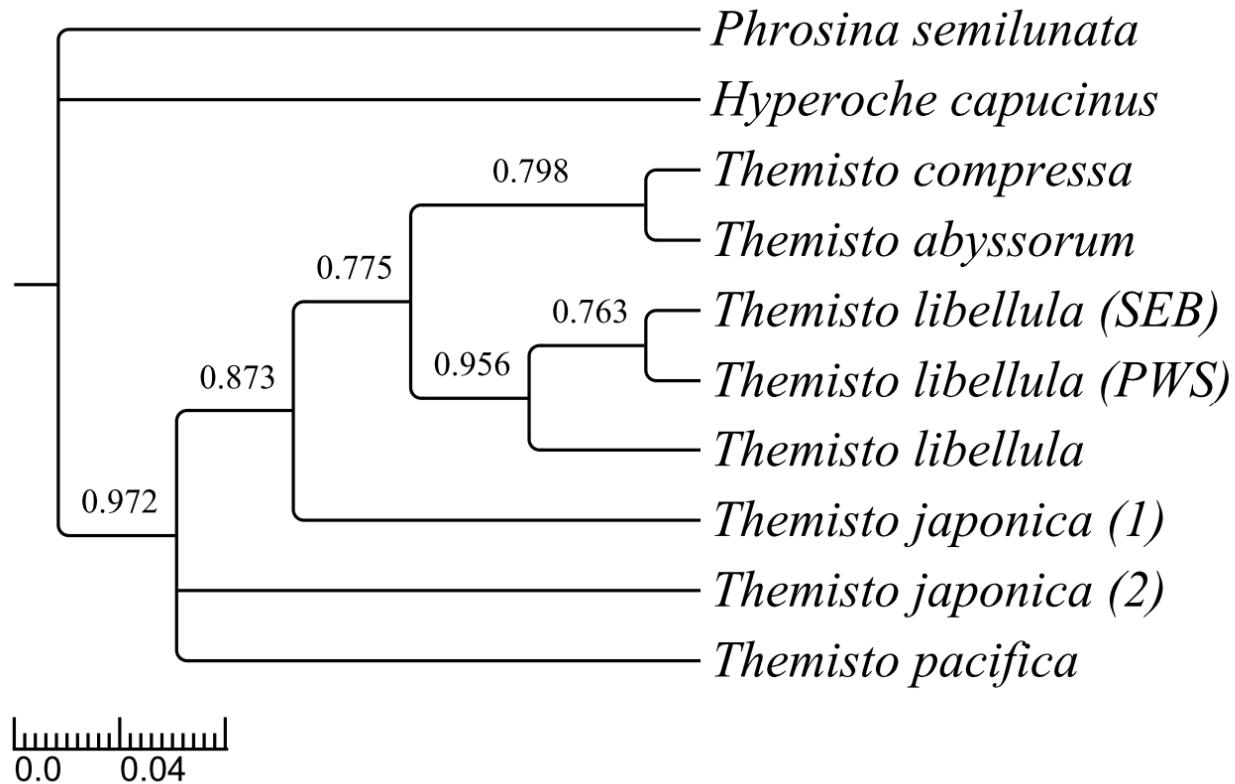
**Figure 3.6** Saturation plots for COI (a), H3 (b), 18S (c) and 28S (d). The green triangles represent the transversion and the blue crosses represent the transition among distances.



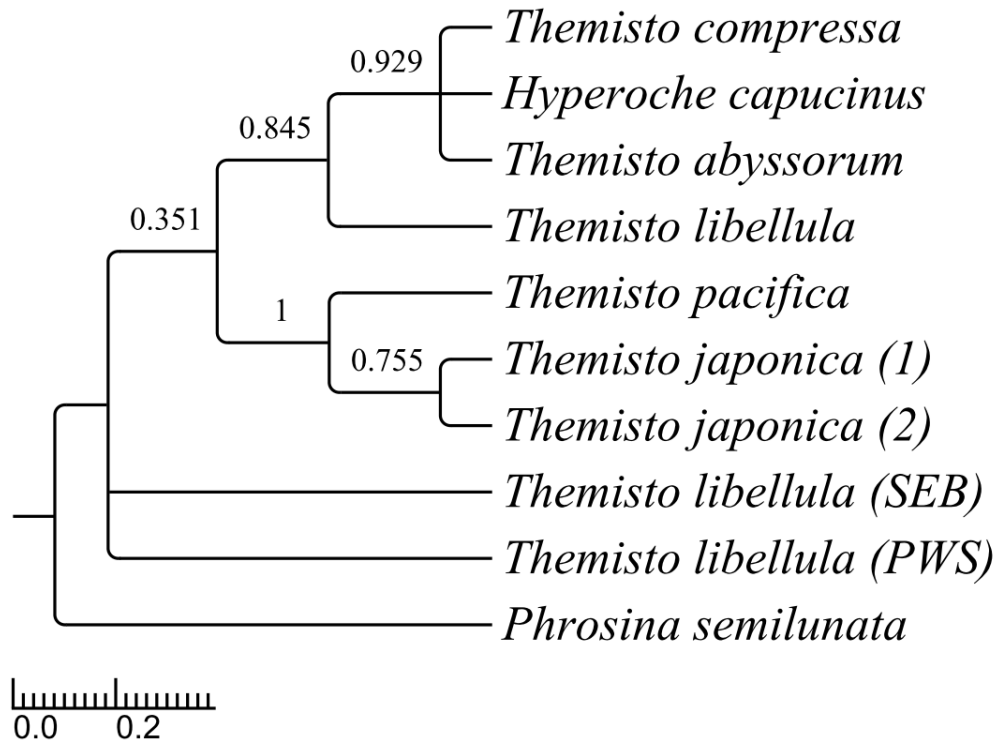
**Figure 3.7** Maximum Likelihood phylogeny (Log-likelihood : -2398.84107) for the cytochrome oxidase I gene. aLTR supports are shown under the nodes.



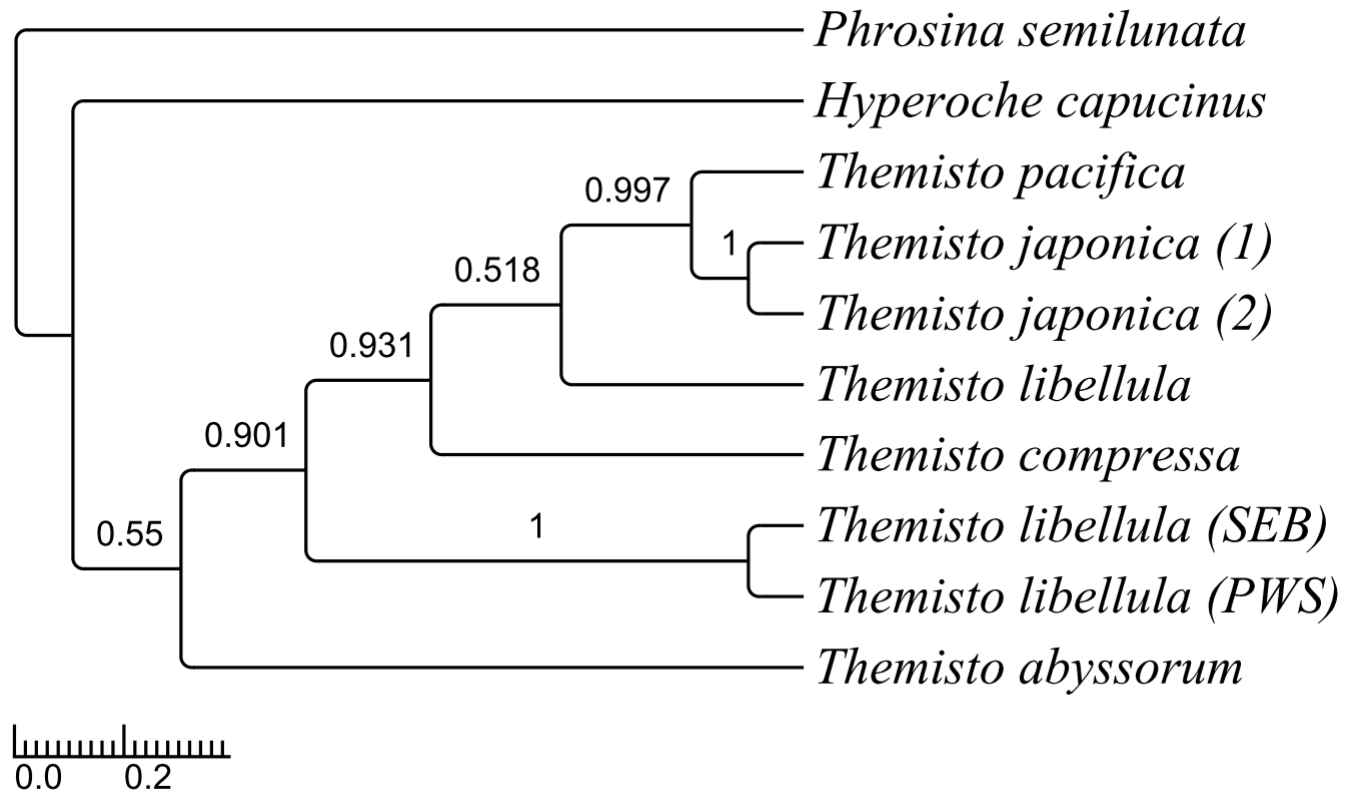
**Figure 3.8** Maximum Likelihood phylogeny (Log-likelihood: -634.27815) for the Histone 3 gene. aLTR supports are shown under the node.



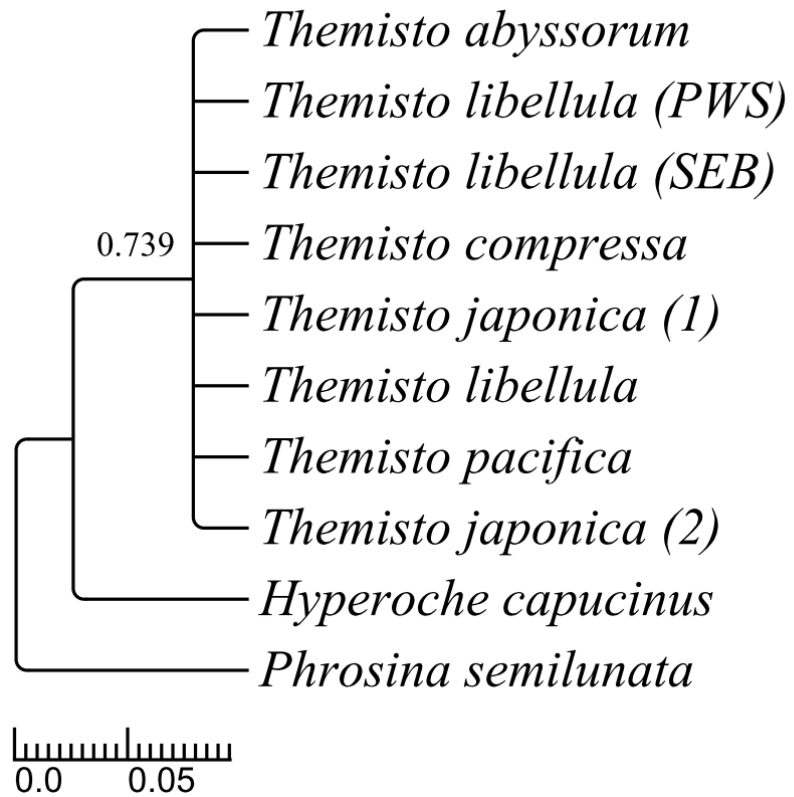
**Figure 3.9** Maximum Likelihood phylogeny (Log-likelihood: -3453.95) for the 18S gene. aLTR supports are shown under the nodes.



**Figure 3.10** Maximum Likelihood tree (Log likelihood=-4688.49522) for the 28S gene. aLTR supports are shown under the nodes.

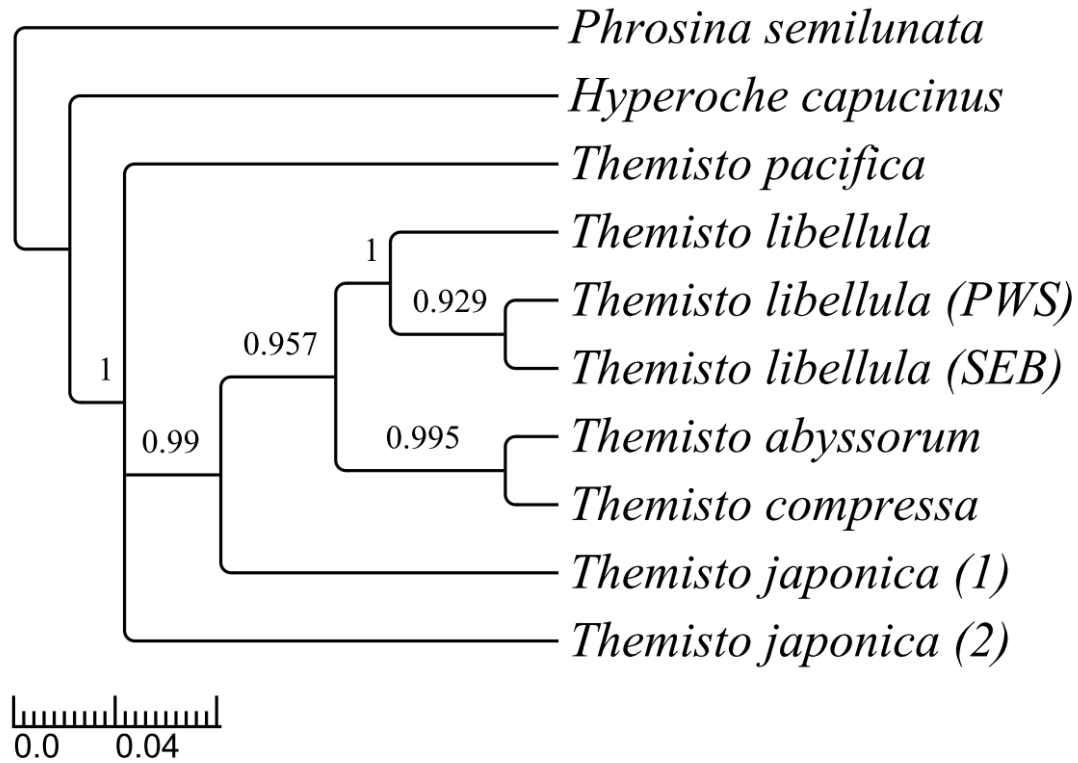


**Figure 3.11** Bayesian phylogeny for the COI dataset. Posterior probabilities are shown under the nodes.

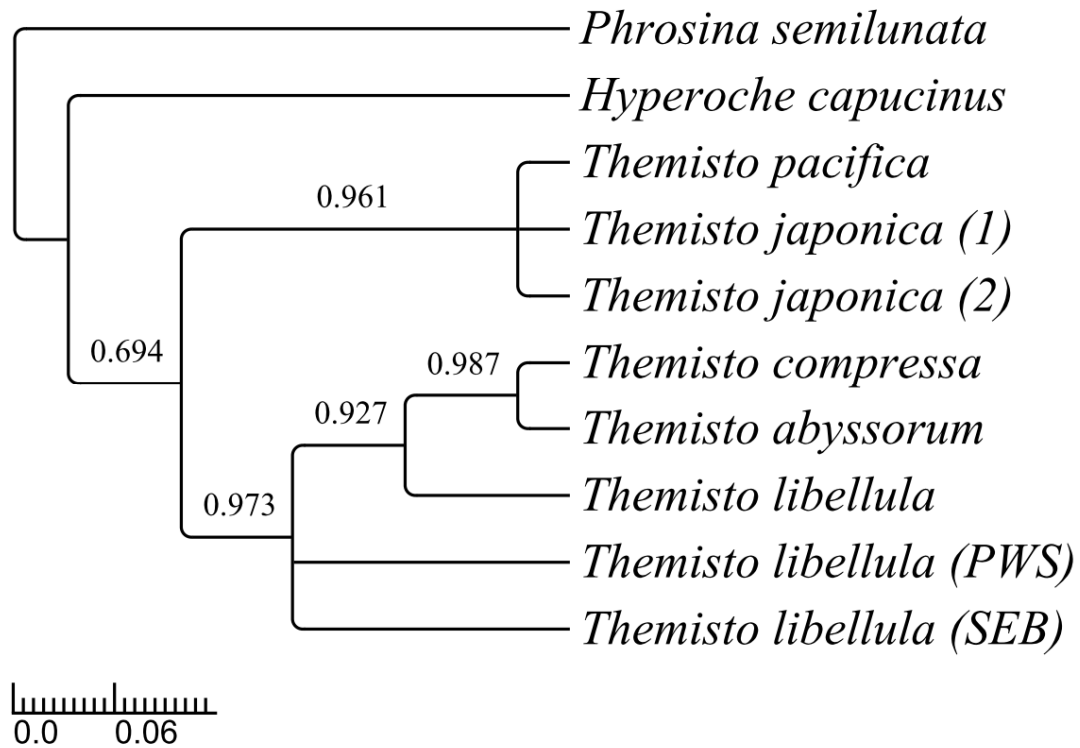


**Figure 3.12** Bayesian phylogeny for the Histone 3 dataset. Posterior probabilities are shown under the nodes.





**Figure 3.13** Bayesian phylogeny for the 18S dataset. Posterior probabilities are shown under the nodes.



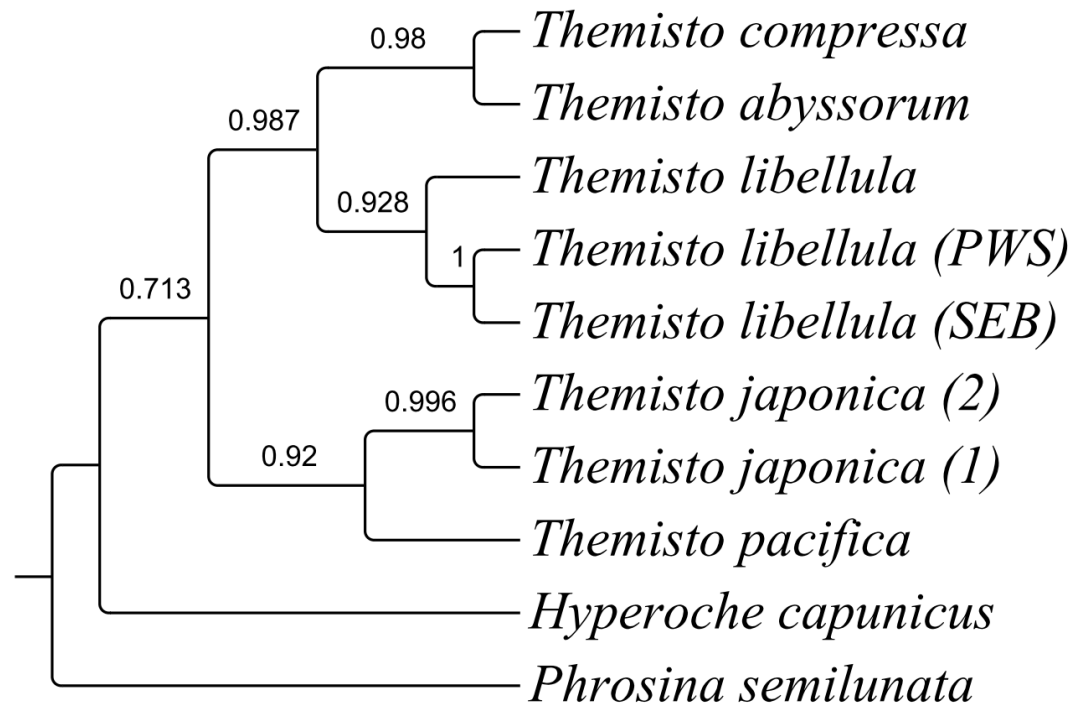
**Figure 3.14** Bayesian phylogeny for the 28S dataset. Posterior probabilities are shown under the nodes.

#### 3.5.4 Concatenated dataset

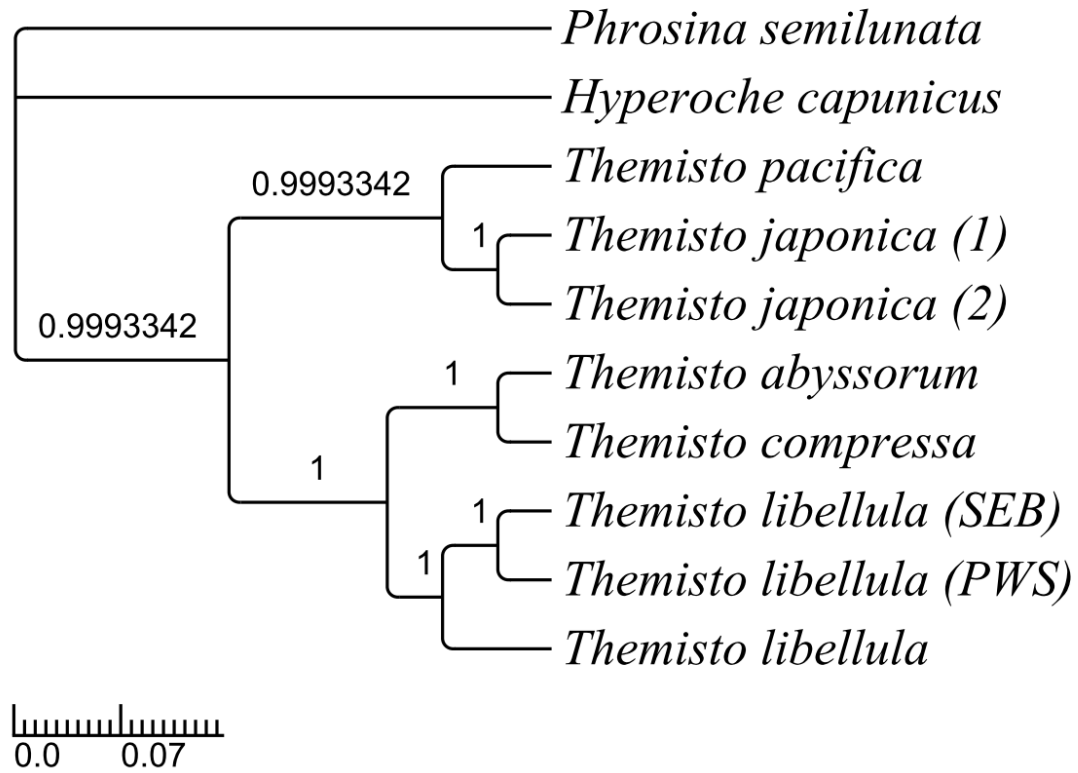
The CADM test showed congruency between the four matrices ( $p=0.0001$ ); therefore, all genes have been concatenated for further analyses. The ML tree constructed under the HKY+G model (BIC= 23432.13) and the BI tree constructed with partitioned data are shown with their node supports in Figures 3.15 and 3.16, respectively. All *Themisto* species were grouped in a strong monophyletic clade (PP= 1). Three supported groups were found: a first group contained the two Atlantic species *T. abyssorum* and *T. compressa* (aLTR=0.97; PP=0.99); a second group included the Arctic species *T. libellula* and *T. libellula* (SEB) and (PWS) (aLTR=0.91; PP= 1); and a last group included Pacific species *T. japonica* and *T. pacifica* (aLTR=0.92, PP= 0.99).

#### 3.5.5 Species tree

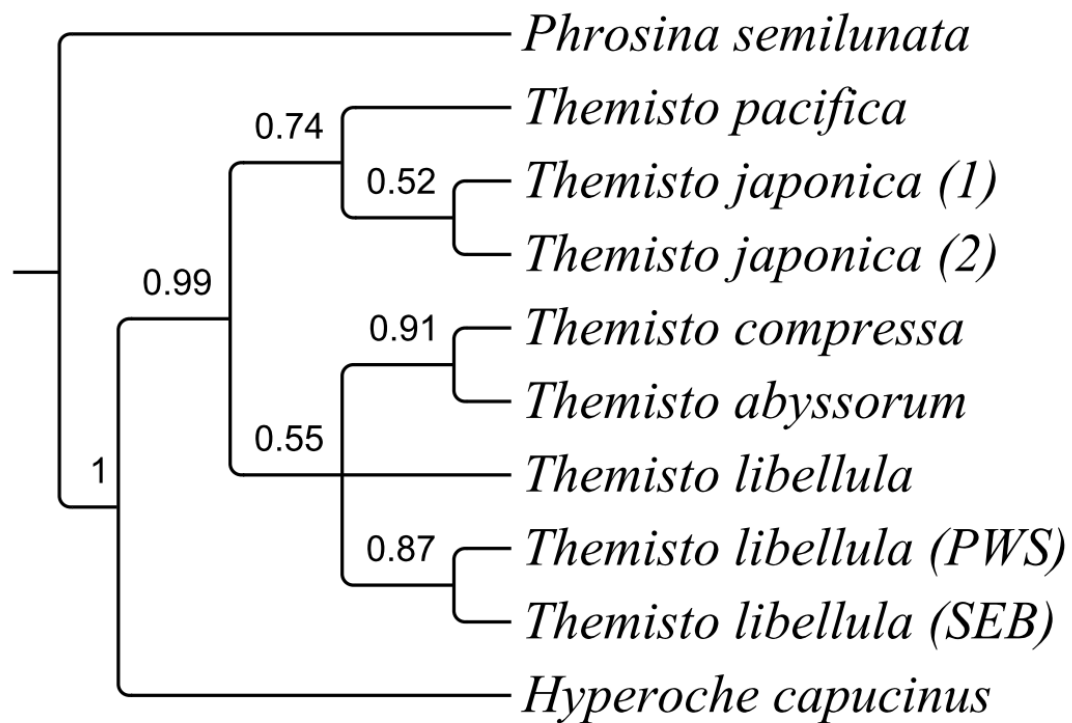
Results from our species reconstruction based on multilocus dataset recovered the same pattern observed in the concatenated gene trees (Figure 3.17). All *Themisto* species were grouped in a strong monophyletic clade (PP= 1). A pacific group was identified, including *T. japonica* and *T. pacifica* species (PP=0.99). The Atlantic group composed by *T. abyssorum* and *T. compressa* (PP=0.91) was also recovered. The main difference of the species tree was observed with the position of *T. libellula*. It was impossible to conclude if *T. libellula* belonged to the Atlantic clade or to the Arctic clade (PP=0.55).



**Figure 3.15** Maximum Likelihood phylogeny (Log-likelihood : -11621.65725) of the concatenated dataset. aLTR support are shown under the nodes.



**Figure 3.16** Bayesian phylogeny of the concatenated dataset. Posterior probabilities are shown under the nodes.



**Figure 3.17** Species-tree estimates accounting deep coalescence events. Numbers represent posterior density.

### 3.6 Discussion

This study provides the most detailed and up-to-date phylogeny of the *Themisto* genus in the northern hemisphere. It differs from previous studies by including for the first time Atlantic, Arctic, and Pacific *Themisto* species. Our results showed the utility of using genetic data in complement of regular taxonomic identification for discovering unprecedented species diversity.

#### 3.6.1 The Hyperiidæ family

Our results differed mostly from those of Hurt *et al.*, (2013) regarding the position of species from the *Themisto* genus. In their BI and ML phylogenies based on COI, the grouping of *Themisto* branched sometimes with *Hyperoche* and sometimes at the base of the Hyperiidæ family. In our case, all phylogenies based on DNA and amino acids placed *Themisto* at the base of the Hyperiidæ family. In addition, our Hyperiidæ phylogeny revealed that species classified in the same genus were not grouped together, as observed for *Hyperiella dilatata* and *Hyperiella antarctica* and for *Hyperietta parviceps* and *Hyperietta vosseleri*. We cannot exclude erroneous morphological identification of these species or that the COI sequences are too saturated to provide a clear resolution at the genus level. As Browne *et al.*, (2007) and Hurt *et al.*, (2013), we were not able to see a clear phylogenetic relationship between the genera *Hyperoche* and *Hyperia*. In particular, two species identified as *Hyperia galba* were placed in two different clades. It will be useful to investigate in more details the molecular relationships within these two genera in the light of newly available taxonomic descriptions (e.g. Zeilder, 2015). Increasing the sampling effort for this important group will help also to get a better image of the phylogenetic link within this family.

Both DNA and amino acid phylogenies agreed on the branching order of most species. Small incongruences resided in the branching of *Hyperiella dilatata* and *Hyperoche martinezii*. The former branched out of the *Themisto* clade in the DNA phylogeny but not in the amino acid one. Some authors have advanced that nucleic acid phylogenies can outperform amino acid phylogenies due to their faster evolving potential (Simmons *et al.*, 2002). The presence of a polytomy for the *Themisto* genus observed in the amino acid phylogeny but not in the DNA phylogeny seems to support this idea. It is also important to remember that our family phylogeny is produced with a single mitochondrial gene. Several studies have demonstrated the relevance of COI when inferring phylogeny in amphipods (Browne *et al.*, 2007; Lörz and Held, 2004). However, molecular phylogenies based on a single gene are not the same as the true species phylogeny (Nichols, 2001; Maddison, 1997). Unfortunately, at the moment, COI is the only gene that has been sequenced for several Hyperiididae species. Even if nuclear loci are available for some Hyperiididae, we noted some inconsistencies in their sequences. Future work on nuclear genes is therefore needed to improve the resolution of the clades that are presently poorly supported.

### 3.6.2 Cryptic *Themisto* species

Both concatenated gene and species trees revealed the presence of a new *Themisto* species. This new lineage that inhabits the Bering area showed more than 12 % divergence at COI with all other *Themisto* species. Traditionally, *T. libellula* has been known to inhabit Arctic regions (Dunbar, 1946; 1957) and to coexist in the Bering area with *T. pacifica* (Wing, 1976). Here, we showed that another species inhabits these regions and is different from the *T. pacifica* species found near Californian coasts. Our study reports cryptic species diversity within this genus and suggests the need for a taxonomic revision. In this study, we focused mostly on the northern hemisphere, and future work should integrate southern specimens in order to get a full picture of species diversity and species genetic relationships within this genus.



The smallest divergence for COI was observed between *T. compressa* and *T. gaudichaudii*. Initially, *T. gaudichaudii* was thought to have a bipolar distribution (Bowman, 1960), but based on allozymic data, this bipolar distribution was due to the presence of two cryptic species: *T. compressa* in Northern hemisphere and *T. gaudichaudii* in the southern waters (Schneppenheim and Wagman-Hass, 1986). Nonetheless, the small divergence observed between these two species is indicative of a putatively more recent speciation event compared to the other *Themisto* species. A small genetic divergence was also observed between Arctic and Antarctic clades of the abyssal amphipod *Eurythenes gryllus* or in three planktonic foraminifera morphospecies, suggesting more connections among poles than previously suspected (Darling *et al.*, 2000; Havermans *et al.*, 2013).

Based on the concatenated gene or the species trees, we retrieved two distinct groups: one strictly Pacific group, composed by *T. japonica* and *T. pacifica*, and another Arctic-Atlantic group that comprised all other *Themisto* species. Surprisingly, *Themisto sp.* did not belong to the Pacific clade but seemed to be closer to the Arctic *T. libellula*. Due to the northward current from the Pacific into the Bering and the strong influence of the Pacific Ocean in the colonisation of the Bering area, we could expect that *Themisto sp.* would be closer to the Pacific *Themisto* species. Previous studies conducted in the Bering area have found a discontinuity among Pacific and Beringian populations (Addison and Hart, 2005; Coyer *et al.*, 2011), suggesting that Alaskan populations have a complex history. In particular, the impact of the Pleistocene glaciations where Bering was a potential marine ice-free region (Hardy *et al.*, 2011) has contributed to the isolation and divergence between Pacific and Arctic populations. Nevertheless, as saturation was detected at the mitochondrial level and in absence of fossil for the Hyperiid, we were not able to calibrate and date the time when these species diverged. Thus, inferring the cause of speciation among *Themisto* congeners in absence of time and ecological data is not yet possible.

In parallel to the recent work on holozooplanktonic species (copepods: Lee and Frost, 2002; Goetze, 2003; 2010; amphipods: Lörz *et al.*, 2012), we have revealed an unsuspected species diversity at the global scale, leading to the question of how marine speciation arises. In the absence of physical barriers in the oceans, it is quite difficult to understand how populations can be isolated enough to become new species. Ecological preferences or vicariant events can be at the origin of speciation process in marine environment (Norris, 2000; Palumbi, 1994). Oceanographic habitat preferences are considered an important factor in the differentiation of marine population (Norris, 2000; Norris and Hull 2011). In the *Eucalanus* copepods, these oceanographic preferences are considered the main determinants of their speciation (Goetze, 2005). Ecological habitat preferences are known among *Themisto* species (Bowman, 1960; Wing, 1976; Koszteyn *et al.*, 1995; Eiane and Daase, 2002; Yamada *et al.*, 2004), each species being associated with a range of temperatures indicating a preference for specific water masses (Koszteyn *et al.*, 1995). Thus, it will be interesting to see if oceanographic preferences are associated with each *Themisto* species, in particular for *Themisto sp.* In addition, Hyperiididae are known to be a parasitic family of gelatinous zooplankton (Laval, 1980). Even if no preferences in host have been previously established for any *Themisto* species, host preferences have been reported among genera. Host selection represents an alternative to select suitable habitats and therefore lead also to adaptation of the parasite (Paterson and Banks, 2001; Hyuse *et al.*, 2005). In hyperiids family, it appears that their association with gelatinous zooplankton have led to the modification of their appendages to better feed on or better attach to their specific host (Browne *et al.*, 2007 and references therein). Therefore, in the near future, it will be useful to compare the phylogeny of gelatinous plankton with the phylogeny obtained for the Hyperiididae family so an eventual pattern of coevolution could be highlighted.

### 3.6.3 Limit of the techniques

Our single gene phylogenies are mostly congruent. The main difference we noted between ML and Bayesian inferences was on the support of the nodes. Lower values of support were associated in ML, as it is often reported in the literature (Anisimova and Gascuel, 2006). Besides, the placement of one outgroup (*Hyperoche capucinus*) failed in two maximum likelihood phylogenies (Histone 3 and 28S). Less phylogenetic resolution was observed with the Histone 3; in particular one of the two outgroups, falling into the same clade. This gene is known to be highly conserved in eukaryotic genome (Rooney *et al.*, 2002) and it may be of limited help to reconstruct phylogeny among recently diverged species. Similarly, the nucleotide sequences of 18S rDNA and 28S rDNA translated into some ambiguities (Chu *et al.*, 2009). Indeed, the large gaps observed in the sequences of these genes for different genera might have complicated alignments and hence phylogenetic resolution.

It is difficult to reconstruct the history of a genus based only on molecular data because gene trees and species trees might not be the same (Maddison, 1997). Incomplete lineage sorting predominates in shallow divergent species, leading to uninformative phylogenetic histories (Moore, 1995; Degnan and Rosberg, 2009). Adding more individuals from different populations would probably help to improve the phylogeny of this genus (Lecointre *et al.*, 1999; Maddison and Knowles, 2006).

### 3.7 Conclusions

Phylogenetic relationships among the Hyperiididae are far from being resolved, but increasing sampling through oceans should contribute to resolve problematic taxa grouping. In addition, results presented here demonstrated that our knowledge on marine biodiversity remains underestimated. The presence of cryptic species was highlighted thanks to the combination of different loci. Our results have clearly resolved the placement of *Themisto* genus at the edge of the Hyperiididae family. Even

if our phylogeny is only a partial picture of the phylogenetic link of this genus, it provides useful information on marine speciation at higher latitudes. In particular, the presence of a cryptic species in the Bering area, distinct from the Pacific and Atlantic species, suggests the presence of barriers to dispersion among these regions. Since the genus *Themisto* is bipolar (Dunbar, 1957; Bowman, 1973; Schneppenheim and Weissman-Hass, 1986), integrating specimens from the Southern Hemisphere will provide a complete picture of the genus and will give some clue to their speciation history.

### 3.8 Acknowledgements

We wish to thank the officers and crew of the CCGS Amundsen for the sampling efforts in 2011. We are grateful to Miyamoto Hiroomi of the University of Tokyo, Atmosphere and Ocean Research Institute for providing samples. We also acknowledge Dr. Kaven Dionne for comments on an earlier version of this manuscript.

## CHAPITRE IV

### GENETIC STRUCTURE ACROSS THE CANADIAN ARCTIC ARCHIPELAGO: A TALE OF TWO MARKERS IN SYMPATRIC AMPHIPOD SPECIES

Astrid Tempestini<sup>1</sup>, Louis Fortier<sup>2</sup>, Soren Rysgaard<sup>3</sup> et France Dufresne<sup>1</sup>

Manuscrit à soumettre à *Molecular Ecology*

<sup>1</sup> Université du Québec à Rimouski, Département de biologie et Centre d'études nordiques, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1

<sup>2</sup> Université Laval, Département de Biologie, Pavillon Vachon., 1045, avenue de la Médecine, Québec, QC Canada G1V 0A6

<sup>3</sup> University of Manitoba, Department of Geological Science, 584, Wallace Building, Winnipeg, MB, R3T 2N2

Contribution des auteurs (initiales)

Mise en place des protocoles expérimentaux : AT

Récolte des données : AT

Analyse des données : AT

Interprétation des résultats : AT, FD

Rédaction du manuscrit : AT

Commentaires et correction du manuscrit : AT, FD, LF, SR

Financement: FD, LF, SR

#### 4.1 Résumé

Les espèces marines holozooplanktoniques ont en théorie des populations très faiblement structurées sur de vastes distances à cause de l'absence de barrière dans l'océan. En comparant les structures génétiques d'espèces proches occupant des niches écologiques différentes, les effets des facteurs historiques sur les populations peuvent être observés. L'objectif de la présente étude était d'observer la structuration génétique chez deux espèces d'amphipodes holozooplanktoniques: *Themisto abyssorum* et *T. libellula*, retrouvés en sympatrie au niveau de l'Archipel Arctique canadien. Nous avons utilisé un fragment du gène mitochondrial du cytochrome oxydase I (COI) et la technique du séquençage RAD. *T. abyssorum*, associé aux grandes profondeurs, révèle une très forte diversité mitochondriale lorsque comparé à *T. libellula*, qui est associé à des eaux moins profondes. Les analyses mitochondriales ont aussi révélé des différences entre les populations groenlandaises et canadiennes de *T. libellula*. L'examen de plusieurs centaines (442), voire de milliers (2144) loci a révélé la présence d'une structuration chez ces deux espèces. Nos résultats suggèrent des échanges limités chez des espèces holozooplanktoniques à une échelle spatiale relativement petite.

**Mots-clés : Arctique; Génomique; Population; *Themisto*, Séquençage RAD; SNP**

## 4.2 Abstract

Holozooplanktonic species are expected to show small amount of genetic structure over great distances due to the absence of barriers in the ocean. Comparing patterns of population structure in closely related species with different ecological niches can provide useful information on the influence of extrinsic factors on genetic connectivity in marine populations. To do so we use mitochondrial cytochrome oxidase I (COI) and Rad sequencing to investigate genetic structuration in two sympatric marine arctic crustaceans *Themisto abyssorum* and *T.libellula* over the Canadian Arctic Archipelago. The deep-sea species *T. abyssorum* exhibits a higher level of mitochondrial haplotypic diversity compared to the shallow waters species *T. libellula*. Mitochondrial analyses reveal genetic difference between Greenlandic and Canadian populations in *T. libellula*. The survey of hundred (884) to thousands (4288) of single nucleotide polymorphism (SNP) uncovered the presence of unsuspected genetic structure in both species. Our results suggest more limited exchanges even for these species with a high dispersal capability at a relatively small spatial scale.

**Key-words: Arctic; Genomic; Population; *Themisto*; RAD sequencing; SNP**

### 4.3 Introduction

Investigating marine connectivity is essential to understand how marine biodiversity arises (Palumbi, 2003; 2004). Three decades of study using molecular markers have revealed more isolation between populations despite no obvious barrier in open marine environments (e.g. Bucklin *et al.*, 1987; Goetze, 2005; Benestan *et al.*, 2015; Waters and Roy, 2004; Blanco-Bercial and Bucklin, 2016; Peijnenburg and Goetze, 2013; Levin, 2006). Limited exchanges observed between marine populations are influenced by both extrinsic factors such as hydrographic regimes (Rocha *et al.*, 2007); currents (Mitarai *et al.*, 2009; White *et al.*, 2010); or physical barriers (Knowlton *et al.*, 1993; Knowlton, 1993) and intrinsic factors such as larval production and pelagic larval duration (Pelc *et al.*, 2009; Bradbury *et al.*, 2008; Selkoe *et al.*, 2006). In general, broadcast spawning species will show less genetic structure than species with no larval stage (Pelc *et al.*, 2009; Bradbury *et al.*, 2008 but see Selkoe and Toolken, 2011). In addition, more the time spent in the water column is important the more the dispersion is large and the less genetic structure is observed (Bradbury *et al.*, 2008). Therefore for holozooplanktonic species which spend their entire life in the water column should have no genetic differentiation over great distances. However, recent studies (Goetze, 2005; Peijnenburg and Goetze, 2013; Blanco-Bercial and Bucklin, 2016) have demonstrated the presence of genetic differentiation in holozooplanktonic species like copepods suggesting the presence of barriers in the open ocean.

Marine genetic population studies have relied on the use of mitochondrial gene and on a limited number of nuclear markers such as microsatellites (Reitzel *et al.*, 2013). In the last decade, single nucleotide polymorphism (SNPs) has appeared as markers of choice in population genetic (Morin *et al.*, 2004). Their characterization in non-model organism has become easier with the advance of next generation sequencing techniques (Davey *et al.*, 2011; Reitzel *et al.*, 2013). To date, several studies have



shown their power to detect marine population structure in crustaceans (Benestan *et al.*, 2015; Blanco-Bercial and Bucklin, 2016) or in cnidarian (Reitzel *et al.*, 2013) where mitochondrial markers have failed to reveal it. However, most of these studies have focused on the North Atlantic Ocean (Benestan *et al.*, 2015; Blanco-Bercial and Bucklin, 2016) or along the Californian coast (Reitzel *et al.*, 2013) but no studies have focused on the Arctic Ocean.

Biodiversity of the Arctic Ocean is not fully characterized as compared to other oceans (Archambault *et al.*, 2010). Projected changes in sea ice coverage, hydrographic regimes, and sea water temperatures due to global warming will have great consequence on endemic marine arctic species (Wassman, 2011; Wassman *et al.*, 2011). The distribution of many marine species will likely shift as a result of changing dispersal pathways. Gaps in our knowledge limit the extent to which we can detect and document changes in species distribution as a result of environmental changes (Carr *et al.*, 2011). The Canadian Arctic Archipelago forms a network of shallow channels that connect the central Arctic region with Baffin Bay characterized by weak south and east currents (Archambault *et al.*, 2010). Marine populations from this area are influenced by complex geological and climatic history of the Arctic (Hewitt, 2004; Bucklin *et al.*, 2010a; Hardy *et al.*, 2011) but information on population genetics are scarce. To our knowledge, a single study has documented the genetic diversity of a marine polychaete in this area (Hardy *et al.*, 2011) suggesting the presence of a secondary contact zone in this area.

Amphipods belonging to the *Themisto* genus are one of the most abundant arctic crustaceans (Dunbar, 1957; Bowman, 1960; Wing, 1976; Auel *et al.*, 2002). They also play an important role in the marine arctic food-web by linking lower and higher trophic levels (Darnis *et al.*, 2012; Prokovick *et al.*, 2013). Two *Themisto* species are found sympatrically in the Canadian Archipelago: *Themisto abyssorum* (Boeck 1870) and *Themisto libellula* (Lichtenstein, 1822). Each species is associated with specific

water masses. *T. abyssorum* is an arctic-boreal species that occurs at depths of 0 to 1000 m and prefers temperatures between 0° and +6°C (Wencki, 2000; Koszteyn *et al.*, 1995; Dalpadado, 2002). *T. libellula* is found between 0 to 100 m, prefers temperature below 4°C, and is a good indicator of Arctic water masses (Wencki, 2000; Pinchuk *et al.*, 2013). Their holozooplanktonic life (i.e. organisms that spend their entire life in the water column) implies a high potential for dispersal over great distances and an absence of genetic structure but this has not been tested yet for these species.

Comparison of the population genetic structure of sympatric closely related species with different ecological niches can be a powerful approach to investigate how present-day and historical forces have influence current genetic diversity (Peijnenburg *et al.*, 2005; Zink 2002; Peijnenburg and Goetze, 2013; Pelc *et al.*, 2009; Moritz and Faith, 1998; Bernatchez and Wilson, 1998). In particular, we aim to test the presence of a secondary contact zone in the Canadian Arctic Archipelago as proposed by Hardy *et al.*, (2011) as well as assess the influence of currents on genetic exchanges across the Canadian Archipelago. We show here contrasting patterns of population genetic structure in these sympatric species inhabiting the Canadian Arctic Archipelago at both mitochondrial cytochrome oxidase I (COI) gene sequences and nuclear RAD sequencing data. We discuss the utility of this approach in the study of non-model marine organisms and its contribution to the survey of arctic biodiversity.

## 4.4 Material and Methods

### 4.4.1 Samples

Samples were collected from six stations (Figure 4.1, Table 4.1) during the summer-fall ArcticNet expedition 2011 through the Canadian Archipelago with horizontal or vertical net of mesh size varying between 200 and 750  $\mu\text{m}$  and preserved in ethanol 95%. Samples from Greenland were provided by the Greenlandic institute of natural resources. Samples were identified through the species level with appropriate taxonomic keys (Vinagrov *et al.*, 1996; Martins and Davis, 2001).

**Tableau 4.1** Number of individuals and their geographic locations used for this study. N= number of individuals used for this study, mt= mitochondrial. Geographic coordinates are in decimal degree. In parenthesis number of individuals used for the analysis.

Station	<i>T. abyssorum</i>			<i>T. libellula</i>		Geographic coordinates	
	mt N	N Rad- tag		mt N	N Rad- tag	Latitude (DD)	Longitude (DD)
<b>115</b>	14	38		21	49 (47)	76.1982	-71.122
<b>301</b>	17	24		20	41	74.0649	-83.2299
<b>323</b>	7	36		12	15	74.0928	-80.3414
<b>312</b>	-	-		20	36 (35)	69.1086	-100.4503
<b>1216</b>	7	9		-	-	70.748	-127.913
<b>310M</b>	11	20 (18)		-	-	71.4202	-101.4109
<b>Gibbs</b>	-	-		16	40	70.4643	-72.1479
<b>St22 (Greenland)</b>	-	-		2	-	71.345	-59.865
<b>St14 (Greenland)</b>	-	-		7	-	70.73	-55.53
<b>472</b>	-	-		18	26	72.1424	-130.5412
<b>460</b>	14	33		-	-	71.0453	-130.3377
<b>pa2.15 (Greenland)</b>	-	-		6	-	71.109	-55.528
<b>total</b>	70	160 (158)		122	204(207)		

#### 4.4.2 Mitochondrial analyses

##### 4.4.2.1 Mitochondrial amplification

DNA extraction was performed on individuals of *T. abyssorum*, *T. libellula* with the E.Z.N.A (Omega-bioteck) or QuickExtract (epicentre) kits according to the manufacturer instructions. Genetic variation was screened at the cytochrome oxidase c I mtDNA (COI) gene. DNA was amplified with polymerase chain reaction (PCR) for 5-10 individuals per sampling site (Table 4.1) using the protocol from Dionne et al. (2011) and the universal primers LCO1490 and HCO2198 from Folmer et al. (1994). PCR products were checked on 1.5 % agarose gel and unidirectionally sequenced by Genome Québec (McGill University, CANADA) with the same primers. DNA sequences were verified by eyes and aligned with MUSCLE (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011).

##### 4.4.2.2 Genetic diversity

Number of haplotypes (H), haplotype diversity (h) and nucleotide diversity ( $\pi$ ) were calculated for mtDNA COI sequences for each location using DNAsp v5 (Librado and Rozas, 2009). Genetic variability within and among sample locations was investigated using two different approaches. First, pairwise  $\phi_{ST}$  values between samples were estimated, with a significance level at the 0.05 determined by 10,000 permutations. Second, a global analysis of hierarchical molecular variance (AMOVA) was performed to test the distribution of the variation at different levels. All these analyses were conducted in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). For each species, haplotype networks were constructed to evaluate intraspecific relationships among all haplotypes using the Median-Joining (MJ) algorithm implemented in Popart (<http://popart.otago.ac.nz>). We tested for a correlation between geographic and genetic distances using a Mantel test. Matrices of geographic distances between

sampling locations were calculated with the geographic distance matrix generator software (Ersts, 2016). Linear Slatkin genetic distances were generated in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). Statistical significance of the correlation was assessed with a 1000 times permutation test.

#### 4.4.2.3 Demographic history

For each species, demographic history was inferred using neutrality tests, and Bayesian Skyline Plot. Tajima's D test (Tajima, 1989), Fu and Li F\* and Fu and Li D\* tests (Fu and Li, 1993), which allow to test a sign of population expansion or contraction, were performed on the entire mitochondrial dataset of each *Themisto* species. Time changes in population size were estimated for each species using a Bayesian skyline method implemented in BEAST 1.8.2 (Drummond *et al.*, 2012). The DNA substitution model was identified according to the best Bayesian information criterion (BIC) using Topaliv2 (Milne *et al.*, 2004). Chains were run for 10 million steps, and convergence was assessed when effective sample sizes (ESS) reached at least 200. For the time scale, a strict molecular clock was used with a rate of 2 % / My has proposed for crustaceans by Knowlton and Weigt (1998). All other parameters were automatically adjusted. Outputs were analyses in Tracer 1.5 (Rambault *et al.*, 2014).

#### 4.4.3 Rad-tags

##### 4.4.3.1 DNA extraction

Genomic extractions were performed on 160xx individuals of *T. abyssorum* and 207 individuals of *T. libellula* with an E.Z.N.A kit (omega-bioteck) according to the manufacturer's instructions. Each DNA extraction was treated with 5 µL of RNase-A solution (10mg/mL, sigma Aldrich). Follow libraries preparation, pooled genomic DNA was cut by the PstI and MstI restriction enzymes and 100 bp paired ends were sequenced on an Illumina HiSeq platform. In order to avoid bias in base calling,

individuals from each sampling location were randomly put on plates. Samples were subsequently pooled according to species. Libraries and sequencing were performed at the Institut de Biologie Intégrative et des Systèmes (*IBIS, Laval University, Quebec, Canada*).

#### 4.4.3.2 SNP identification

Hi-seq Illumina output for each species was used to genotype individuals according to the UNEAK pipeline available in TASSEL 3.0 (Bradbury et al, 2007) as described in White et al. (2014). Briefly, the UNEAK pipeline first trimmed all the reads to 64 bp long. Then, identical reads were grouped into tags. All tags were compared to find candidate SNPs with 1 bp mismatch between paired tags. Finally, a network filter step was used to discard complicated networks, which are usually a mixture of repeats, paralogs and error tags (Lu et al, 2013). The reciprocal pairs of tags retained after this step were considered as valid SNPs. Following SNP identification, individuals were scored for their genotype, allowing a sequencing error rate of 3%. Loci that were called in at least 70% of individuals were retained using TASSEL 5.0 (Glaubitz et al, 2014). In order to distinguish SNPs from sequencing errors (Glaubitz et al. 2014), we set a minimum allele frequency to 0.05. As paralogous sequences can inflate heterozygosity (Pan et al, 2015) and therefore bias analyses, we also removed loci with  $H_e > 0.5$  and individuals with more than 30 % of missing value. Loci that showed deviations from the Hardy-Weinberg equilibrium were also removed to avoid genotyping errors (Hoskin et al, 2004). Hardy-Weinberg equilibrium tests were conducted on each loci for each *Themisto* species for all and within each collection site with the adegenet package (Jombart, 2008; Jombart and Ahmel, 2011) available in R. SNPs with a p-value  $< 0.0001$  in any site, or  $< 0.001$  in multiple collection sites, were discarded (p-value computed using 999 simulations).

#### 4.4.3.3 Detection of outliers

As phylogeographic parameter estimates often rely on the assumption of neutral markers, we identified outlier SNPs in order to remove them from the analyses, as suggested by Beaumont and Nichols (1996). Briefly, using a Bayesian approach, outlier loci are loci deviating from the neutral expectation under a certain model of population. Loci under selection were identified with BayeScan 2.1 (Foll and Gaggiotti, 2008) as described in Glover *et al.*, (2012; Supporting information) with a prior odds of 100000 and default parameters. Graphical outputs were visualized in R.

#### 4.4.3.4 Population structure and assignment tests

The initial analysis of population structure was conducted with an individual-based discriminant analysis of principal component (DAPC) using the adegenet package (Jombart, 2008; Jombart and Ahmel, 2011) available in R. We also inferred population structuring with a Bayesian method as implemented in STRUCTURE v2.3.2 (Pritchard *et al.*, 2000). Briefly, individuals are assigned to putative populations (K) based on their genotype without prior geographic information. The admixture model was used with uncorrelated allele frequencies. The Markov Chain Monte Carlo was run for 10000 repetitions with a burn-in period of 10000. Values for K were evaluated from 1 to 20 (5 replicates each). Populations were then defined based on results from DAPC and STRUCTURE. STRUCTURE results were analyzed using the online version of the software Structure HARVESTER (Earl and vonHoldt, 2012; <http://taylor0.biology.ucla.edu/structureHarvester>), the best number of clusters was chosen according to the method of Evanno *et al.*, (2005). Results were visualized with the web application STRUCTURE PLOT (Ramasamy *et al.*, 2014; <http://btismysore.in/strplot/index.php>).

#### 4.4.3.5 Genetic diversity and $F_{ST}$

Pairwise  $\phi_{ST}$  were calculated for each population and their significance was assessed with a 10000 times permutation test in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). In order to investigate the distribution of genetic diversity among and within populations, we performed an analysis of molecular variance (AMOVA).

#### 4.4.3.6 Isolation by distance

We tested for the presence of a correlation between geographic and genetic distances using a Mantel test. Matrices of geographic distances between sampling locations were calculated with the geographic distance matrix generator software (Ersnst, 2016). Linear Slatkin genetic distances were generated in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). Statistical significance of the correlation was assessed with a 1000 times permutation test.





**Figure 4.1** Sampling locations of individuals used for this study.

## 4.5 Results

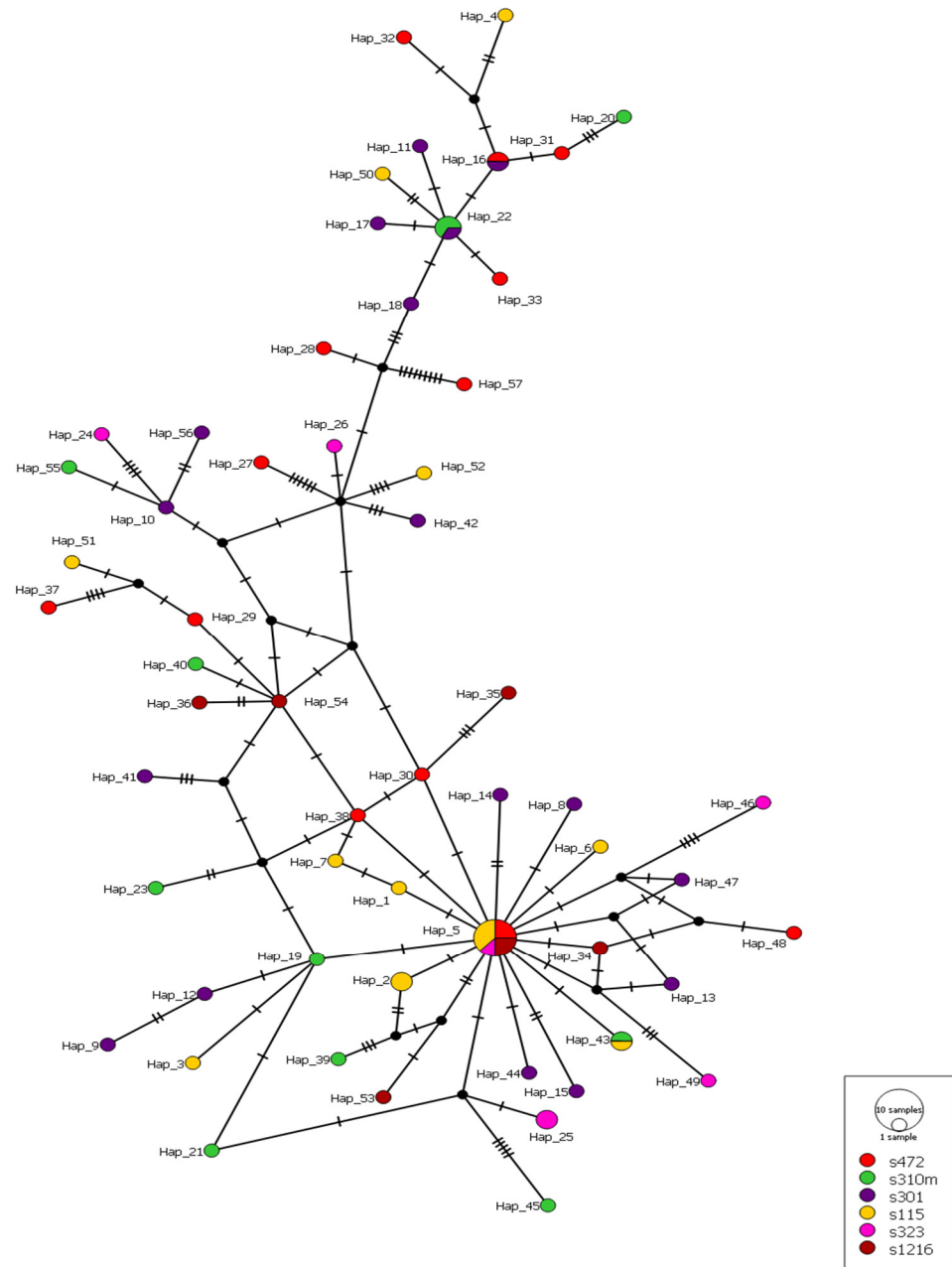
### 4.5.1 Mitochondrial analysis

We successfully sequenced 491bp and 576 bp of the COI mitochondrial gene for 70 *T. abyssorum*, 122 *T. libellula* respectively. We detected 72 variable sites and 54 synonymous changes over the 491 bp, for *T. abyssorum*. By contrast, we observed only 12 variable sites and 6 synonymous changes in *T. libellula*. No indels were detected in both datasets. The divergence within species ranged from 0 to 4.1% for *T. abyssorum* and from 0 to 1.24 % for *T. libellula*. A higher number of haplotypes were detected in *T. abyssorum* than in *T. libellula* (57 vs 20; Table 4.2). *T. abyssorum* had the highest GC content (0.37), haplotypic diversity ( $h=57$ ), and nucleotide diversity ( $\pi= 0.01388$ ) (are in parenthesis). *T. libellula* had lower haplotypic diversity ( $h=20$ ) and nucleotide diversity ( $\pi=0.00241$ ) even though more individuals were sequenced in this species. Sampling sites harboured similar levels of haplotypic diversity in *T. abyssorum*. By contrast, in *T. libellula*, stations had varied levels of haplotypic diversity (e.g.  $0.167\pm 0.0134$  for station 323 and  $0.714\pm 0.00074$  in station 14).

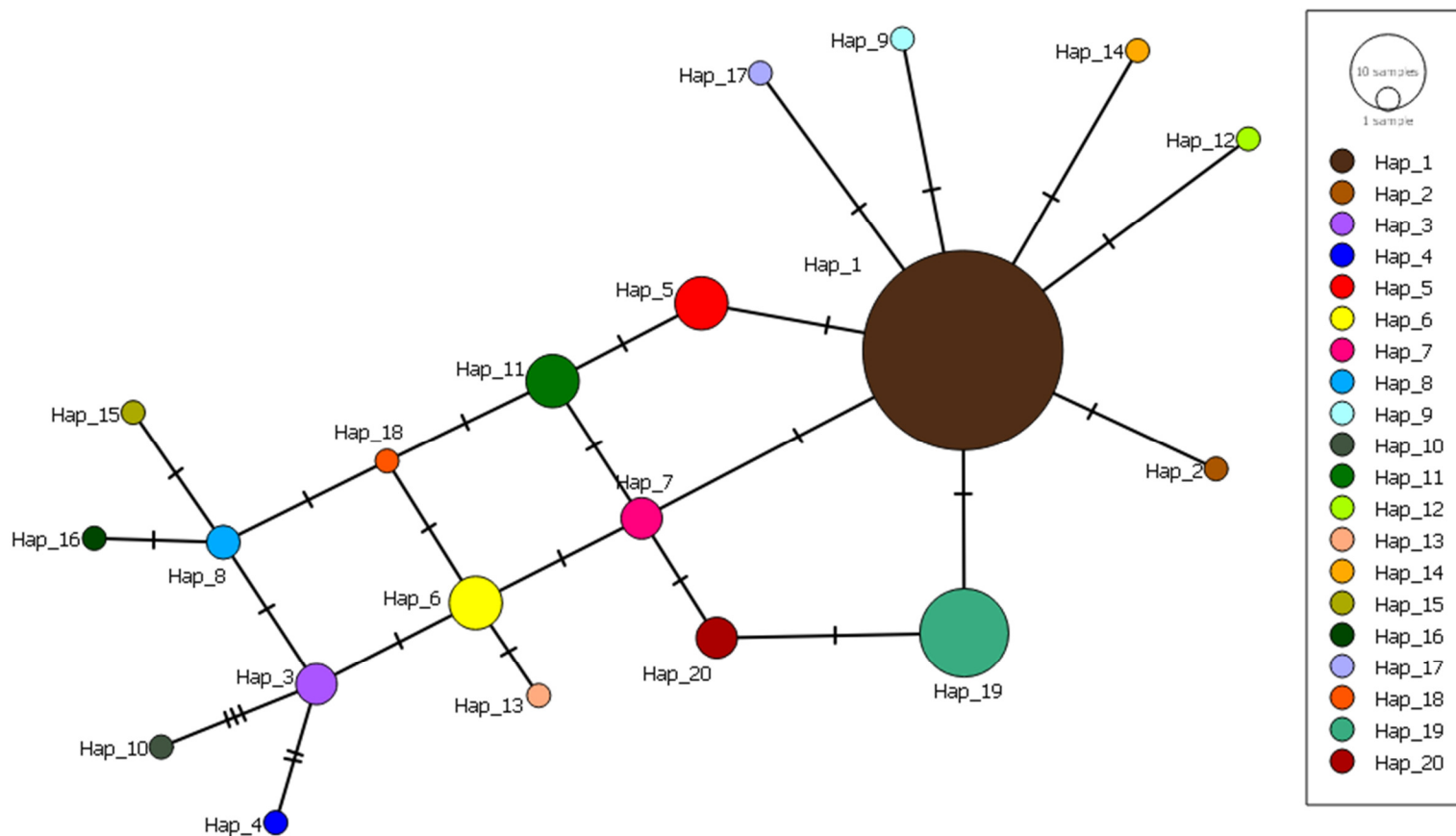
Star-like haplotype networks were observed with one central haplotype in both species (Figures 4.2 and 4.3). In *T. abyssorum*, 51 unique haplotypes were characterized and haplotype 5 was found in four stations (st460, st1216, st115 and st323) throughout the Archipelago. In *T. libellula*, 11 haplotypes were unique and haplotype 1 was common and widely distributed throughout the Archipelago whereas haplotype 19 was restricted to Greenland and Gibbs Fjord (Figure 4.4). Likewise, haplotype networks did not show clear patterns of spatial structure in both species. The lack of relationship between genetic and geographical distances was confirmed by non-significant Mantel tests in *T. abyssorum* ( $r=-0.09$ ,  $p>0.5$ ) and *T. libellula* ( $r= 0.25$ ,  $p>0.05$ ).

**Tableau 4.2** Results of the mitochondrial analysis of the cytochrome c oxidase I (COI) for each *Themisto* species. N= number of individuals, H= number of haplotypes, h= Haplotypic diversity,  $\pi$ =nucleotide diversity. Standard deviation (sd) are in parenthesis.

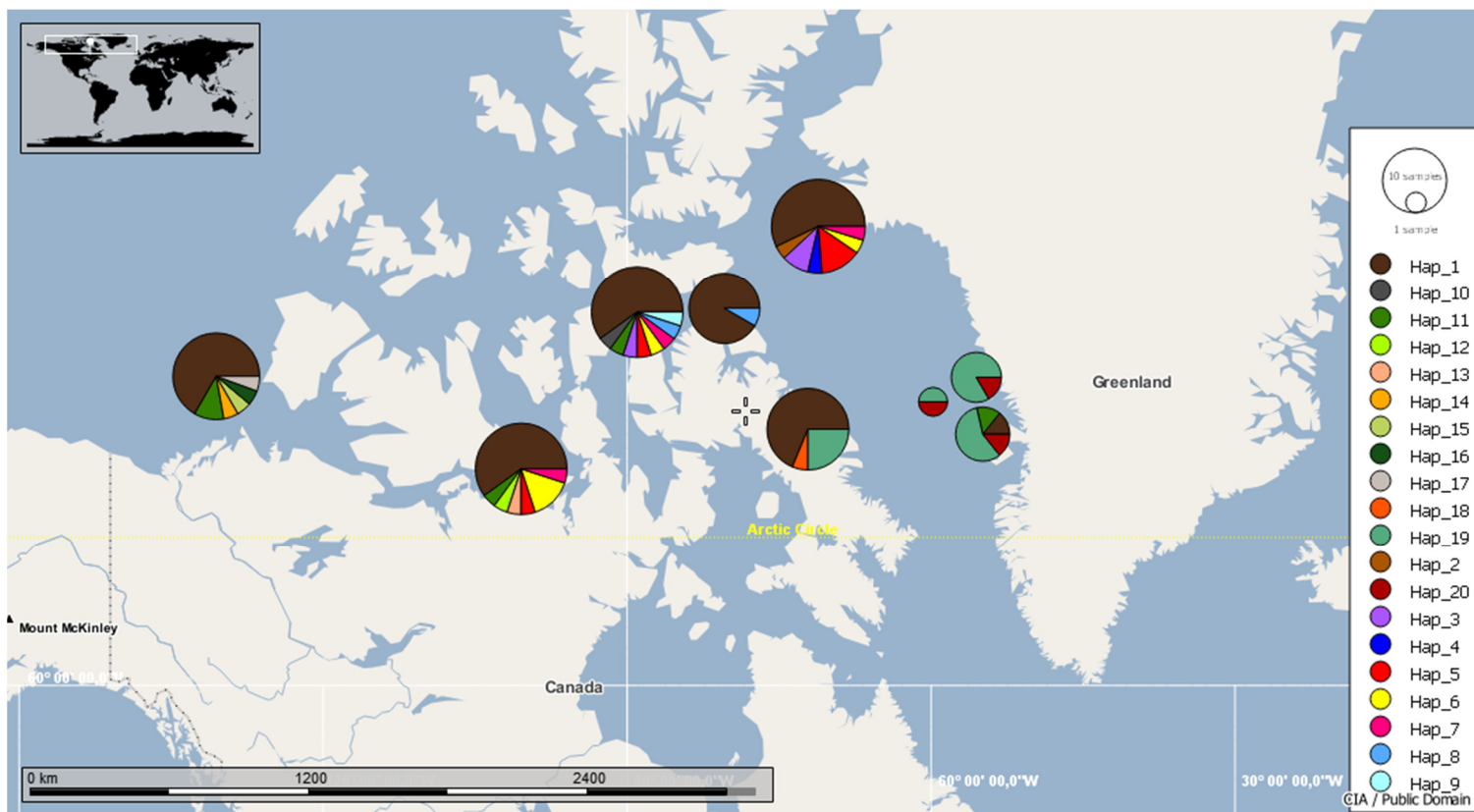
	N	H	h ( $\pm$ sd)	$\pi$ ( $\pm$ (sd)	G+C content
<i>Themisto abyssorum</i>	70	57	0.986 ( $\pm$ 0.008)	0.01388 ( $\pm$ 0.00097)	0.370
<b>St460</b>	14	13	0.989 ( $\pm$ 0.031)	0.01705 ( $\pm$ 0.0023)	0.369
<b>St1216</b>	7	6	0.952 ( $\pm$ 0.096)	0.00737 ( $\pm$ 0.00158)	0.371
<b>St310m</b>	11	10	0.982 ( $\pm$ 0.046)	0.01663 ( $\pm$ 0.00217)	0.369
<b>St323</b>	7	6	0.952 ( $\pm$ 0.096)	0.01348 ( $\pm$ 0.00277)	0.371
<b>St301</b>	17	17	1 ( $\pm$ 0.02)	0.01448 ( $\pm$ 0.00095)	0.370
<b>St115</b>	14	11	0.956 ( $\pm$ 0.0045)	0.00991 ( $\pm$ 0.0022)	0.370
<i>Themisto libellula</i>	122	20	0.646 ( $\pm$ 0.048)	0.00241 ( $\pm$ 0.0003)	0.317
<b>St472</b>	18	6	0.562 ( $\pm$ 0.134)	0.00277 ( $\pm$ 0.00093)	0.317
<b>StGibbs</b>	16	3	0.492 ( $\pm$ 0.117)	0.00135 ( $\pm$ 0.00056)	0.318
<b>St312</b>	20	7	0.637 ( $\pm$ 0.116)	0.000203( $\pm$ 0.00045)	0.317
<b>St323</b>	12	2	0.167 ( $\pm$ 0.134)	0.000116 ( $\pm$ 0.00093)	0.317
<b>St301</b>	20	9	0.653( $\pm$ 0.122)	0.00298 ( $\pm$ 0.00086)	0.317
<b>St115</b>	21	7	0.667 ( $\pm$ 0.107)	0.00261 ( $\pm$ 0.0007)	0.317
<b>st22</b>	2	1	1 ( $\pm$ 0.5)	0.00174 ( $\pm$ 0.00087)	0.319
<b>pa2.15</b>	6	2	0.333 ( $\pm$ 0.215)	0.00058 ( $\pm$ 0.00037)	0.319
<b>st14</b>	7	4	0.714 ( $\pm$ 0.00076)	0.00215 ( $\pm$ 0.00076)	0.319



**Figure 4.2** Haplotype network based on 491 bp of mitochondrial cytochrome c oxidase I sequences of *T. abyssorum*. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to sampling sites (Figure 4.1). Black dots correspond to missing (non-sampled) haplotypes.



**Figure 4.3** Haplotype network based on 576 bp of mitochondrial cytochrome c oxidase I sequences obtained for *T. libellula*. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype. The number of traits on a line corresponds to the number of mutations between haplotypes.



**Figure 4.4** Spatial distribution of twenty haplotypes found in *T. libellula* among sampling sites. Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype.

AMOVA analysis conducted on *each Themisto species* showed that the main variation occurred within sampling sites (Table 4.3). For *T. abyssorum* only a small fraction (2%,  $\phi_{ST} = 0.02$ ;  $p < 0.001$ ) occurred among sampling sites. The other two AMOVA performed on *T. libellula* showed the same tendency except that a higher proportion of genetic variation was found among sampling sites. When we included a comparison by region, 35% ( $\phi_{SC} = -0.008$ ;  $p < 0.001$ ) of the genetic variation occurred between Greenlandic and Canadian group of sampling sites, suggesting a potential separation among these regions.

Pairwise  $\phi_{ST}$  values between sampling sites ranged from -0.007 to 0.09 in *T. abyssorum* (Table 4.4). Most of them were not significant ( $p > 0.05$ ) except between the station 460 and 1216. For *T. libellula*, Pairwise  $\phi_{ST}$  values between sampling sites ranged from -0.009 to 0.62 (Table 4.5). Almost all  $\phi_{ST}$  between Greenlandic and Canadian sites were significant ( $p < 0.05$ ), except between Station 301, and Gibbs Fjord with station Pa 2.15, and between Gibbs Fjord and St14.

**Tableau 4.3** Analysis of Molecular Variance AMOVA for each *Themisto* species. \* significant at  $p < 0.05$ .

	Source of variation	Percentage of variation
<i>T. abyssorum</i>	Among Sampling location	2.02
	Within Sampling location	97.98
$\phi_{ST} = 0.02$ *	Among Sampling location	11.31
		88.69
<i>T. libellula</i>	Within Sampling location	
$\phi_{ST} = 0.113$ *	Among Greenland site and Canadian Archipelago	35.44
	Among sampling locations within groups	-0.56
$\phi_{SC} : - 0.008$ *	Within sampling location	65.13
$\phi_{CT} : 0.35432$ *		



**Tableau 4.4** Pairwise  $\phi_{ST}$  between sampling locations for *T. abyssorum*. \* significant at  $p < 0.05$

	<b>St115</b>	<b>St301</b>	<b>St 323</b>	<b>St 310m</b>	<b>St460</b>	<b>St 1216</b>
<b>St 115</b>	0					
<b>St 301</b>	0.03074	0				
<b>St 323</b>	0.02133	0.03883	0			
<b>St 310m</b>	-0.00236	-0.03829	0.02436	0		
<b>St 460</b>	0.05554	-0.00585	0.05998	-0.03046	0	
<b>St 1216</b>	-0.00799	0.07258	0.02399	0.03250	0.09639*	0

**Tableau 4.5** Pairwise  $\phi_{ST}$  between sampling location for *T. libellula*. \* significant at  $p < 0.05$ .

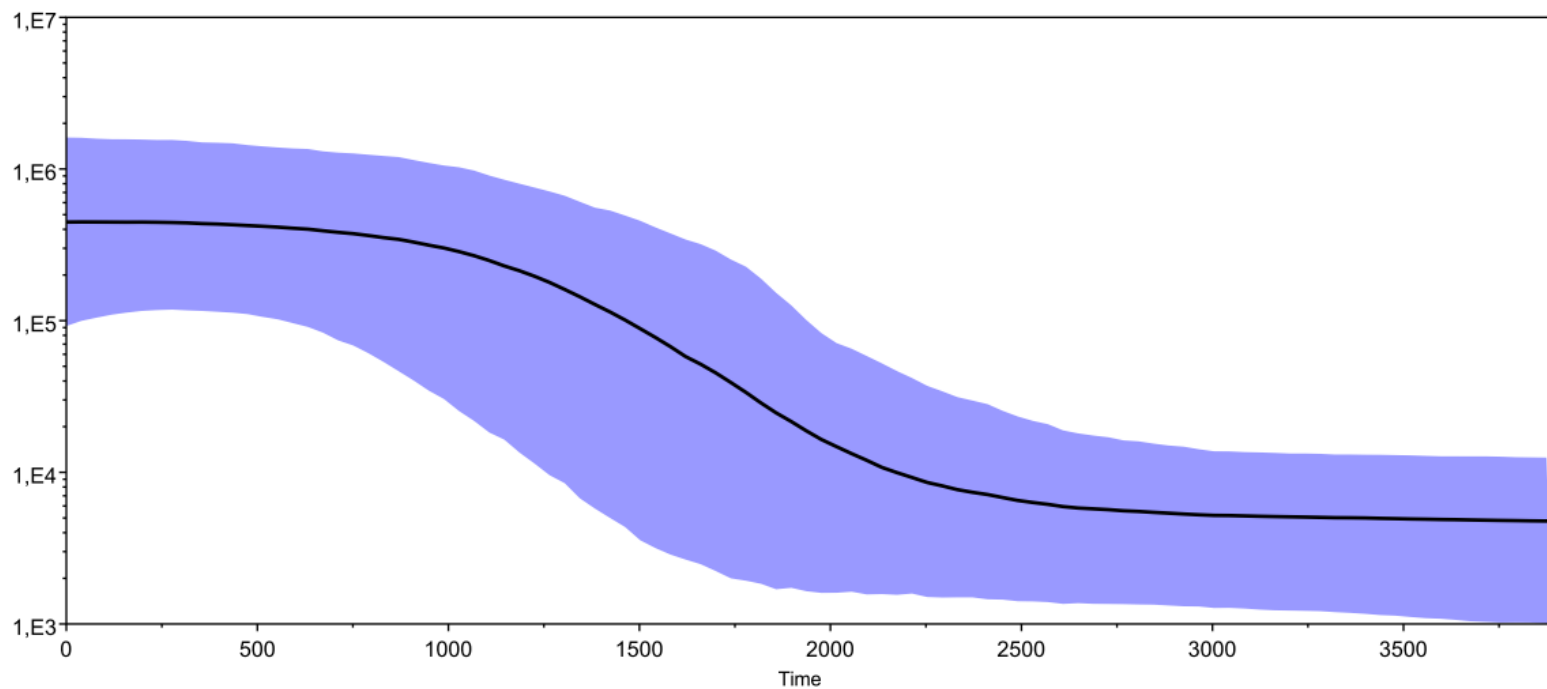
	<b>St115</b>	<b>St301</b>	<b>St312</b>	<b>St323</b>	<b>St460</b>	<b>Stgibb</b>	<b>St22</b>	<b>Stpa2.15</b>	<b>St14</b>
<b>St115</b>	0.00000								
<b>St301</b>	-0.03721	0.00000							
<b>St312</b>	-0.02100	-0.02277	0.00000						
<b>St323</b>	-0.02576	-0.00724	0.00678	0.00000					
<b>St460</b>	-0.03145	-0.02761	-0.01453	-0.02767	0.00000				
<b>Stgibb</b>	0.06101	0.07458	0.08136	0.02333	0.05405	0.00000			
<b>St22</b>	0.42799*	0.40077*	0.49230*	0.62895*	0.41846*	0.43144*	0.00000		
<b>Stpa2.15</b>	0.33459*	0.28488	0.40057*	0.58699*	0.31742*	0.40584	-0.09091	0.00000	
<b>St14</b>	0.23260*	0.21064*	0.27187*	0.35047*	0.21334*	0.16720	-0.01435	-0.24793	0.00000

**Tableau 4.6** Results of Tajima's D, Fu's Fs neutrality tests. \* significant at  $p < 0.05$ .

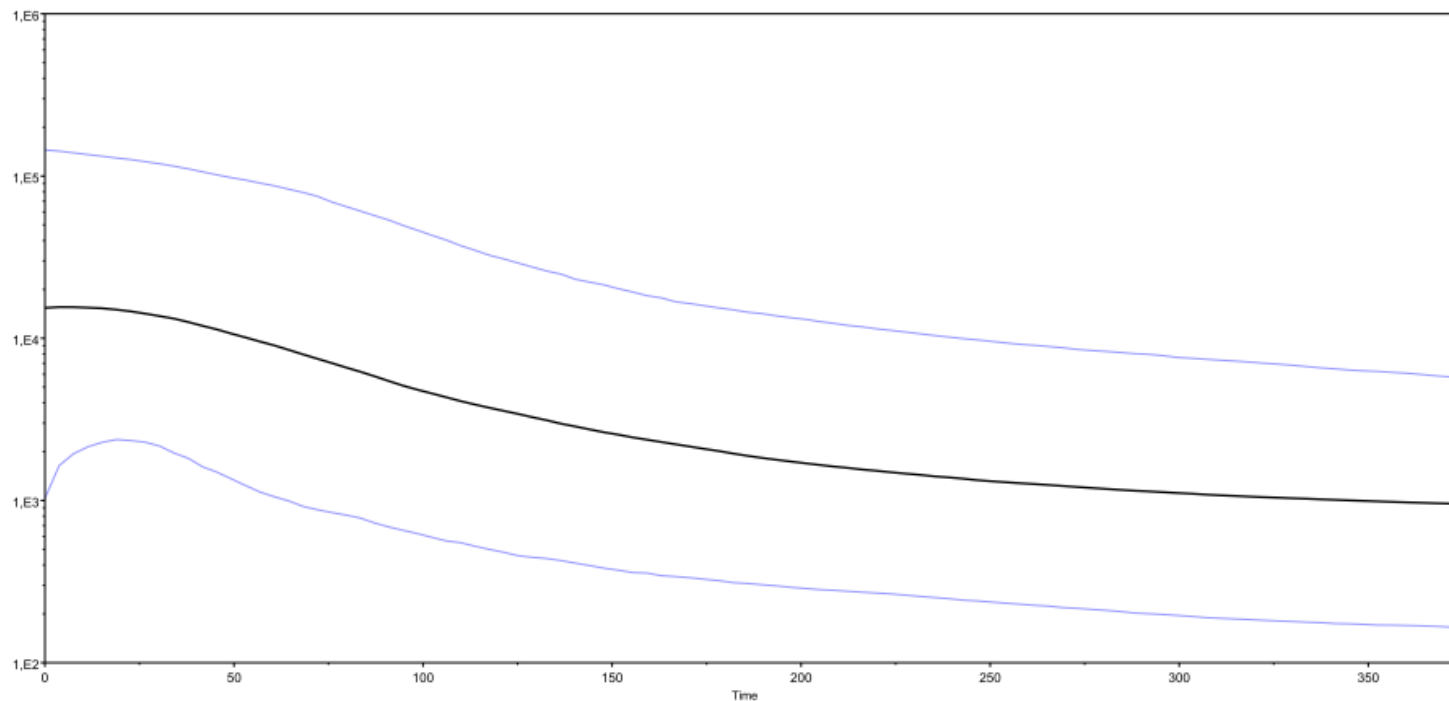
	<b>Tajima's D</b>	<b>Fu and Li D*</b>	<b>Fu and Li F*</b>
<i>T.abysorum</i>	<b>-1.997*</b>	<b>-3.693*</b>	<b>-3.624*</b>
<i>T. libellula</i>	-1.5467	<b>-4.05212*</b>	<b>-3.72122*</b>
<i>T.libellula</i> (Canada)	<b>-1.67948*</b>	<b>-3.90435*</b>	<b>-3.67619*</b>
<i>T.libellula</i> (Greenland)	-0.39538	-0.01225	-0.012852

The results of Tajima's D and Fu and Li D\* and F\* tests to infer the demographic history of each *Themisto* species is presented in Table 4.6. Significant negative values were observed for *T. abyssorum* and the group composed by the Canadian individuals of *T. libellula*. A non-significant negative value of Tajima's D associated with significant negative value of Fu and Li tests was also detected in the entire dataset of *T. libellula*. Negative values were also estimated in the Greenlandic dataset of *T. libellula* but were not significant.

Bayesian skyline plots confirmed the previous signal of expansion we detected in both species (Figures 4.5 and 4.6). An earlier time since expansion in *T. abyssorum* was estimated (~1000 years ago) compared to the more recent time estimate in *T. libellula* (less than 200 years). Moreover, *T. abyssorum* has a higher effective population size estimate ( $1 \times 10^6$ ) compared to *T. libellula* ( $1 \times 10^5$ ).



**Figure 4.5** Bayesian Skyline Plot for *T. abyssorum* estimated under the HKY+G model (BIC=10661.94). The blackline represents the mean and the blue lines represent the confidence interval. The y-axis represents the effective population size estimate and the x-axis represents the time in years.



**Figure 4.6** Bayesian skyline plot for *T. libellula* estimated under the HKY+G model (BIC=4566.09). The blackline represents the mean and the blue lines represent the confidence interval. The y-axis represent the effective population size estimate and the x-axis represents the time in years.

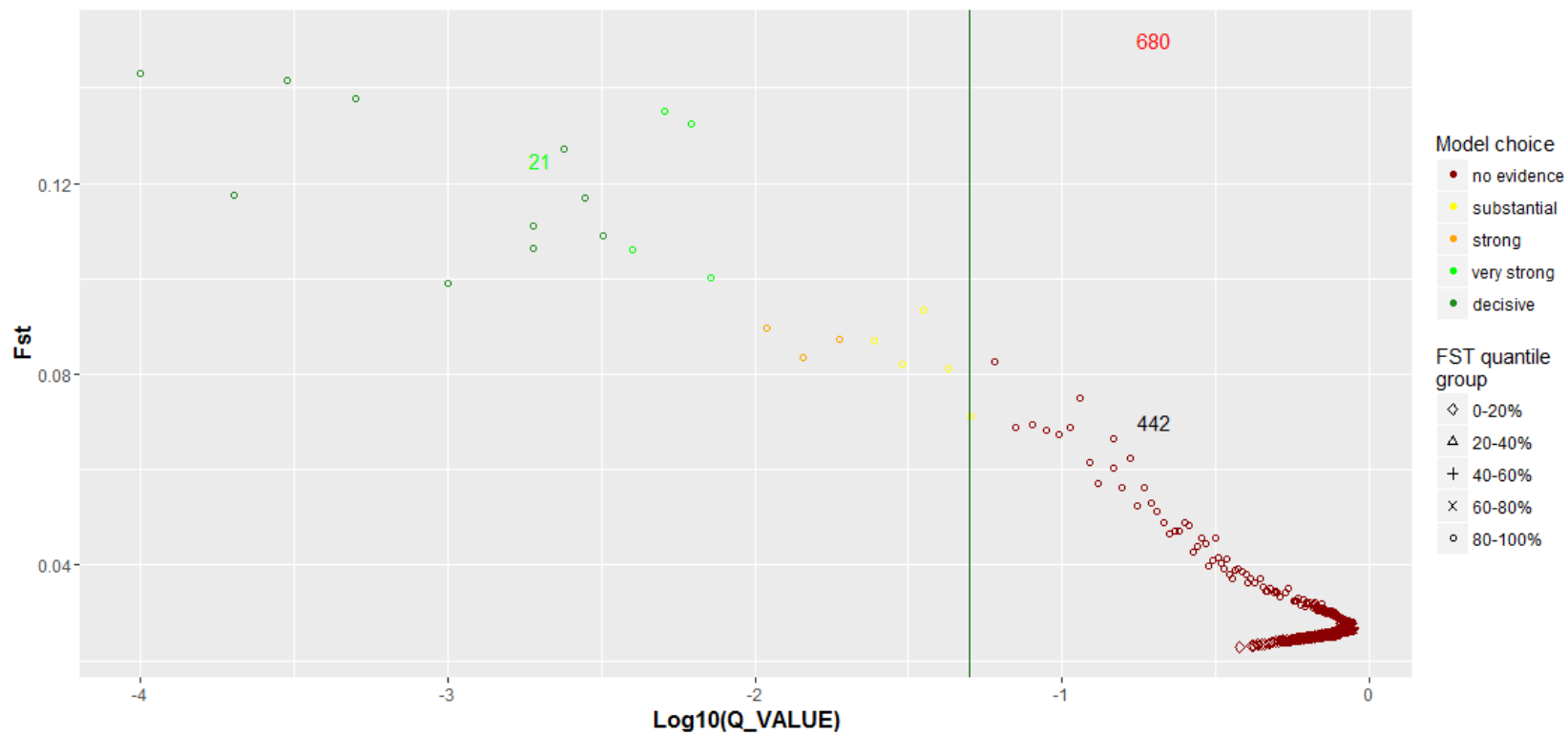
#### 4.5.2 Rad-tag analysis

RAD sequencing produced 368 017 190 reads for *T. abyssorum* and 449 183 454 reads for *T. libellula*. As the genome size of *T. libellula* was available (C= 3,97pg-Rees *et al.*, 2007), we estimated that the sequencing depth for each *Themisto* dataset was close to 10X. The number of reads for each individual ranged from 69 609 to 4 555 343 for *T.abysorum* and from 31 995 to 4 487 248 for *T. libellula*, probably due the variable quality of DNA among individuals. A total of 91040 loci for *T. abyssorum* and 83116 loci for *T. libellula* have been identified as potential loci for this analysis prior to filtering. After filtering, the final dataset consisted of 4967 loci for *T. abyssorum*, 6167 loci for *T. libellula* (Table 4.7). After removing loci not in Hardy-Weinberg equilibrium, 2755 loci for *T. abyssorum* and 2194 loci for *T. libellula* were used for further analysis.

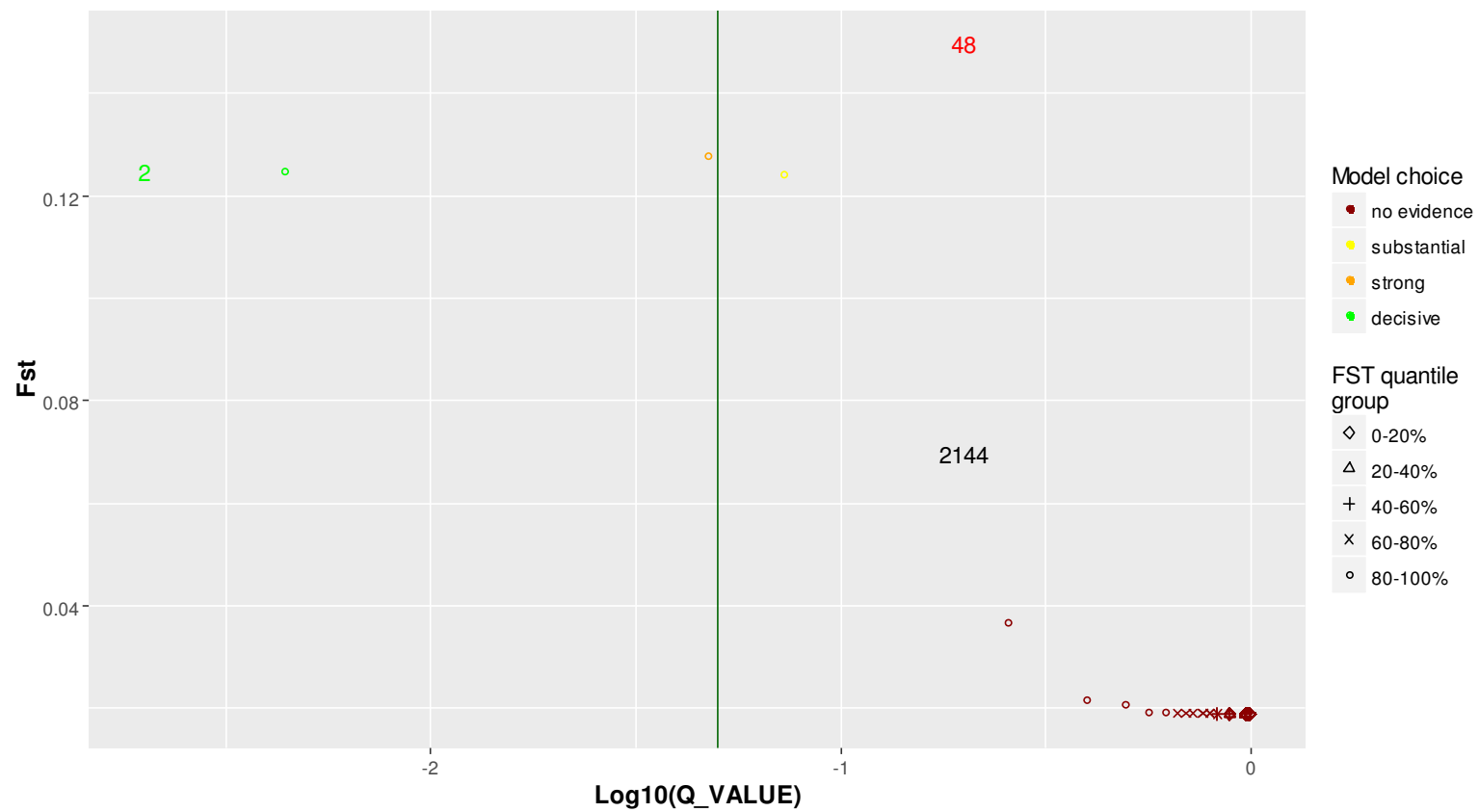
**Tableau 4.7** Number of putative loci retained following each filtering step.

Filter	<i>T.abysorum</i>	<i>T. libellula</i>
initial	91040	83116
Minor allele frequency >0.05 >85% of population	4967	6167
Hardy-Weinberg all population	2755	5032
each population	1143	2194

Of the 1143 loci used for *T. abyssorum*, we detected 442 neutral loci, 680 loci under balancing selection, and 21 loci under diversifying selection (Figure 4.7). For *T. libellula*, we detected 2144 neutral loci, 48 loci under balancing selection and 2 loci under diversifying selection over the 2194 loci retained for this analysis (Figure 4.8).

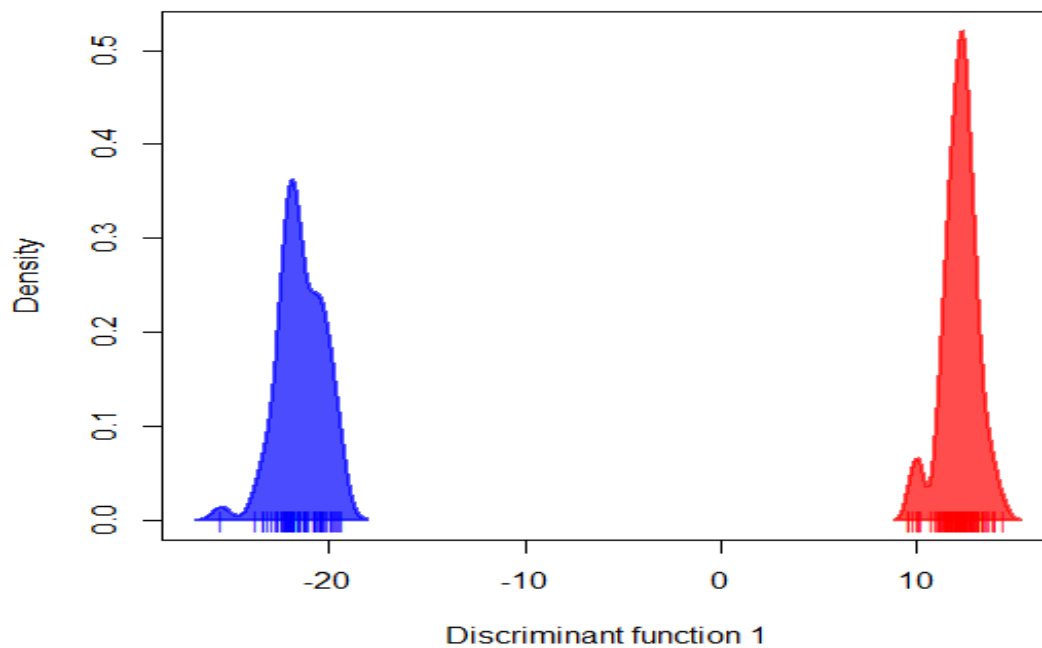


**Figure 4.7** Bayescan test for selection on individual SNPs between the two populations of *T. abyssorum*. Red number represents SNPs potentially under balancing selection and green number represents SNPs potentially under diversifying selection.



**Figure 4.8** BayeScan test for selection on individual SNPs among the three populations of *T. libellula*. Red number represents SNPs potentially under balancing selection and green number represents SNPs potentially under diversifying selection.

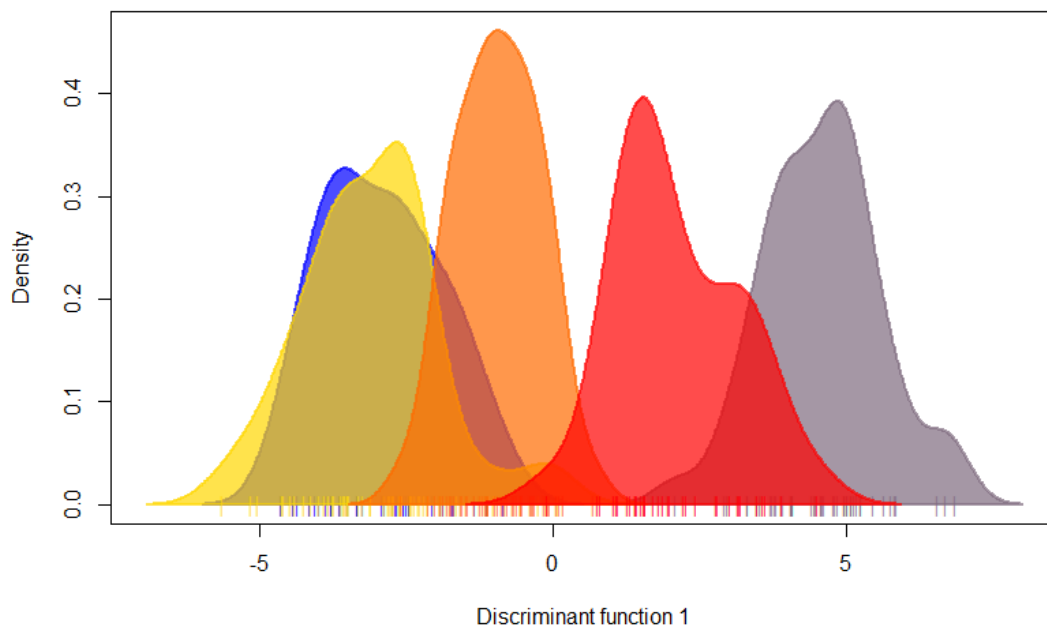




**Figure 4.9** Discriminant analysis of Principal Component for *T. abyssorum* performed with all 442 neutral loci.

**Tableau 4.8** Membership of each individual according the DAPC.

	<b>Cluster 1</b>	<b>Cluster 2</b>
<b>St115</b>	25	13
<b>St323</b>	24	12
<b>st301</b>	16	8
<b>St310M</b>	13	5
<b>St460</b>	16	17
<b>St1216</b>	4	5



**Figure 4.10** Discriminant analysis of Principal Component for *T. libellula* performed with all 2144 neutral loci. Each color represents a cluster.

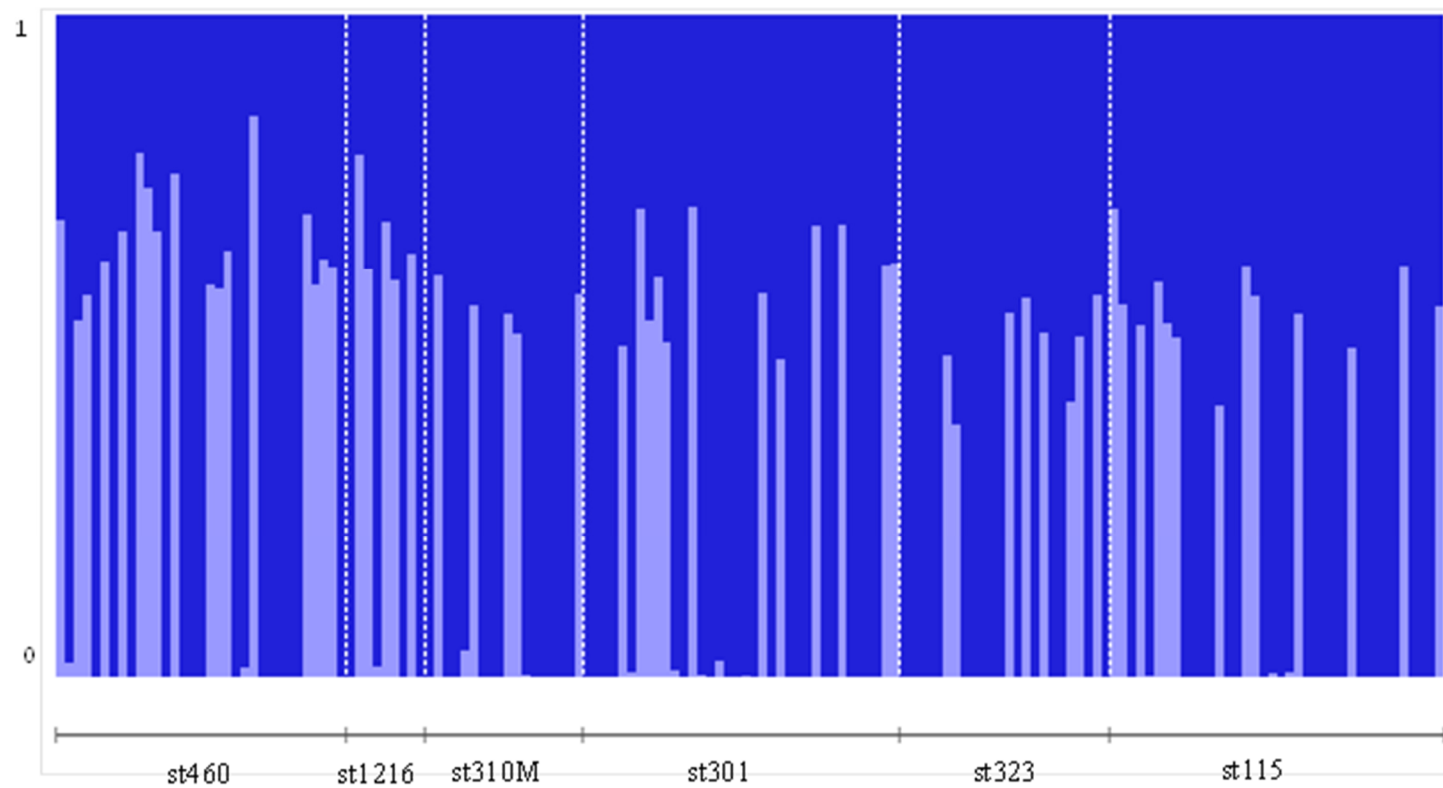
**Tableau 4.9** Membership of individual after the DAPC.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
<b>St115</b>	14	3	9	17	4
<b>St301</b>	13	2	10	11	5
<b>St323</b>	4	0	5	6	0
<b>chu</b>	0	1	0	0	0
<b>Gibbs</b>	15	5	8	10	2
<b>St312</b>	1	17	2	1	14
<b>St472</b>	0	16	0	1	8

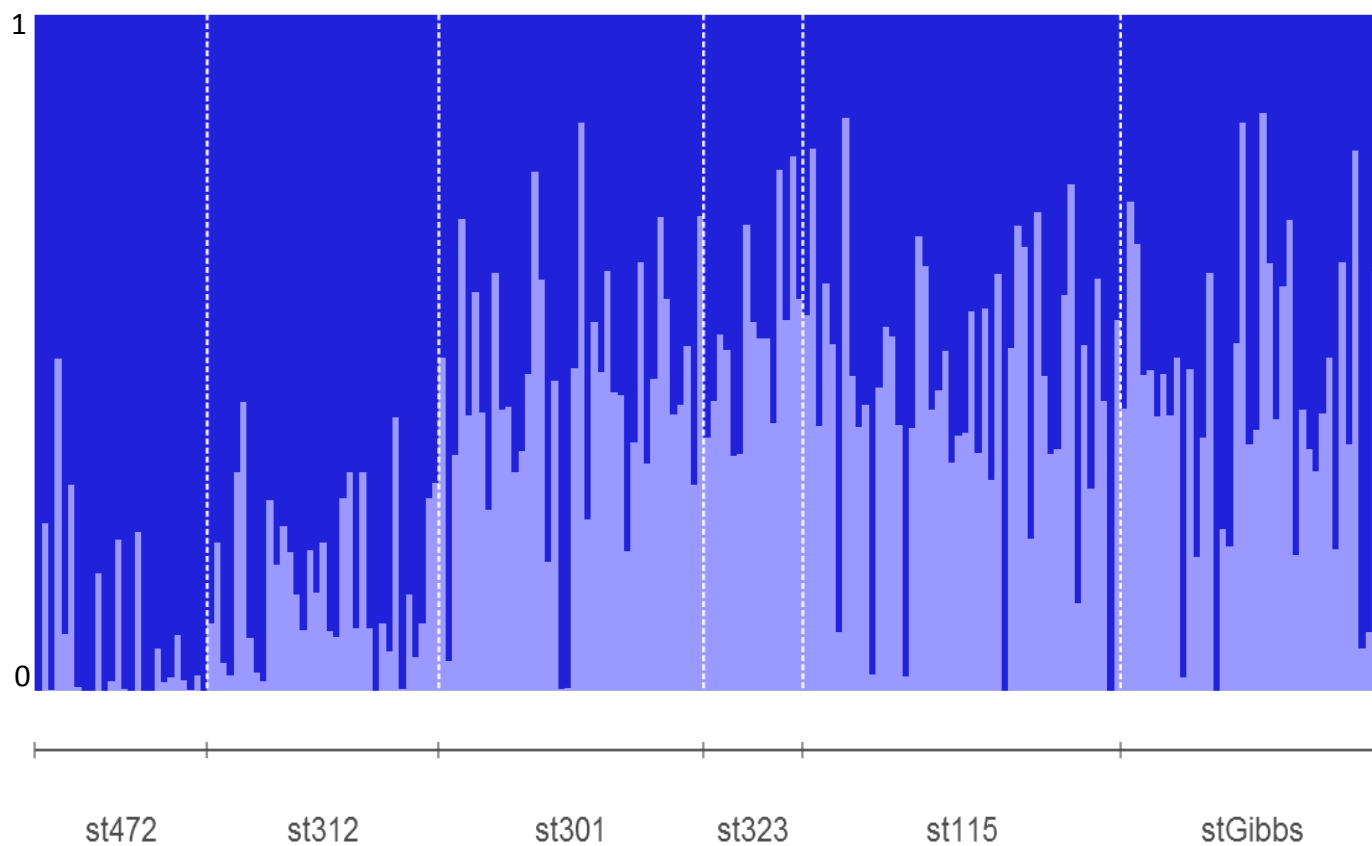
DAPC and STRUCTURE analyses detected two clusters in *T. abyssorum* (Figure 4.9 and Figure 4.11). The first cluster contains more individuals from all stations, whereas cluster 2 was composed by fewer individuals from all the stations (Table 4.8). Both analysis detected 102 individuals belonged to the first cluster (cluster 1; in orange Figure 4.11) and 56 individuals belong to the second cluster (cluster 2; in yellow Figure 4.11). Based on discriminant analysis, five clusters were identified for *T. libellula* (Figure 4.10 and Figure 4.12). Clusters 2 and 5 were composed mostly of all the individuals from Stations 472 and 312 whereas clusters 1, 3, and 4 included individuals from the other station (115, 301, 323 and Gibbs Fjord) included individuals from all other stations (Table 4.9). STRUCTURE analysis suggested K=2 (Table 4.10). The first cluster (in yellow) was mostly found in stations 312 and station 472 whereas the second cluster was composed exclusively by individuals from stations 115, 301, 323, and Gibbs Fjords.

**Tableau 4.10** Evanno table for each *Themisto* species. In bold: the best K according to the Evanno method.

	K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
<i>T. abyssorum</i>	1	5	-57833.18	0.6611	NA	NA	NA
	<b>2</b>	<b>5</b>	<b>-56055.84</b>	<b>14.5929</b>	<b>1777.34</b>	<b>3793.14</b>	<b>259.93035</b>
	3	5	-58071.64	2981.9664	-2015.8	4112.94	1.379271
	4	5	-64200.38	4835.9287	-6128.74	8390.84	1.735104
	5	5	-78719.96	5254.4141	-14519.58	10673.36	2.031313
	6	5	-82566.18	11766.078	-3846.22	164.26	0.01396
	7	5	-86576.66	5872.6016	-4010.48	14190.78	2.416438
	8	5	-76396.36	3208.472	10180.3	19066.6	5.94258
	9	5	-85282.66	12951.028	-8886.3	14093.72	1.088232
	10	5	-80075.24	18198.796	5207.42	NA	NA
<i>T. libellula</i>	1	5	-381650.8	4.36	NA	NA	NA
	<b>2</b>	<b>5</b>	<b>-375596.34</b>	<b>49.6835</b>	<b>6054.46</b>	<b>2214.46</b>	<b>44.571311</b>
	3	5	-371756.34	456.4738	3840	340.84	0.74668
	4	5	-367575.5	97.6145	4180.84	4182.78	42.849977
	5	5	-367577.44	178.8346	-1.94	130.72	0.730955
	6	5	-367448.66	102.9091	128.78	225.04	2.186783
	7	5	-367544.92	400.2207	-96.26	5.64	0.014092
	8	5	-367635.54	409.6047	-90.62	198.68	0.485053
	9	5	-367527.48	228.6037	108.06	456.3	1.99603
	10	5	-367875.72	738.2274	-348.24	877.62	1.188821



**Figure 4.11** Structure plot using the SNP dataset with two clusters for *T. abyssorum*, under the admixture model. Distinct colors represent estimated clusters, sampling sites are separated by a black line.



**Figure 4.12** Structure plot using the SNP dataset with two clusters for *T. libellula*, under the admixture model. Each color represents an estimated cluster; sampling sites are separated by a black line.

As clustering analysis revealed no clear geographic separation between groups in *T. abyssorum*, AMOVA analyses were performed on a single group. Results revealed a small but significant proportion (1.54%,  $\phi_{ST}=0.015$ ) of the variation found among sampling sites (Table 4.11). By contrast, for *T. libellula*, as two geographic groups were detected from STRUCTURE analyses. Hence, hierarchical genetic population analysis was performed using two regions (one corresponding to the Beaufort/Amundsen Gulf) and the other corresponding to the Lancaster Sound (Table 4.11). A small but significant proportion of the variation occurs between the two groups (2.16%;  $\phi_{CT}=0.0215$ ) and the main variation occur within sampling sites (97.65%;  $\phi_{ST}=0.023$ ). The isolation by distance analysis was not significant for *T. abyssorum* ( $r=0.22$ ;  $p=0.16$ ) nor for *T. libellula* ( $r=0.017$ ,  $p=0.45$ ).

**Tableau 4.11** Analysis of Molecular Variance (AMOVA) for each *Themisto* species.  
\*\*= significant value  $p<0.05$

	Source of variation	Percentage of variation
<i>T. abyssorum</i>	Among sampling stations	1.54 $\phi_{ST}=\mathbf{0.015^{**}}$
	Within sampling stations	98.46
<i>T. libellula</i>	Among groups	2.16 $\phi_{CT}=\mathbf{0.0215^{**}}$
	Among populations within groups	0.19 $\phi_{SC}=\mathbf{0.0019^{**}}$
	Within populations	97.65 $\phi_{ST}=\mathbf{0.023^{**}}$

## 4.6 Discussion

We report contrasted patterns of genetic structure in two widely distributed and abundant species of amphipods from the *Themisto* genus that are an important component of the diets of many northern marine fish, birds and mammals (Darnis *et al.*, 2012). Our analyses, based on COI sequences and nuclear SNPs, identified unsuspected genetic divergence in these high latitudes pelagic invertebrates. While our results provide evidence of limited genetic exchanges in the pelagic environment, different complex demographic histories appear to be at the origin of these different patterns. Here, we discuss how our results contribute to our understanding of seascape genetic and arctic marine diversity.

### 4.6.1 Mitochondrial diversity

*T.abysorum* displayed very high haplotypic diversity as often the case in pelagic invertebrates such as the krill *Euphausia superba* (Goodall-Copestake *et al.*, 2010; Deagle *et al.*, 2015), marine amphipods of the genus *Caprella* (Ashton *et al.*, 2008; Pilar-Cabezas *et al.*, 2013), and the sea cucumber *Holothuria nobilis* (Uthicke and Benzie, 2003). In general, high levels of haplotypic diversity are associated with the very high effective population size seen in marine species (Raupach *et al.*, 2010). Our effective population size estimates were indeed high for both species:  $10^6$  and  $10^5$  for *T. abyssorum* and *T. libellula* respectively, and in the range of population estimates for other pelagic organisms (Goetze, 2005; Peijnenburg *et al.*, 2005). The effective population size estimates for *T.libellula* differ from that of *T. abyssorum* by a factor of ten, which could explain the differences in mitochondrial gene diversity between these species. In addition, both species vary in generation time (1 year for *T.*



*abyssorum* and 1-3 years in *T. libellula*) (Dunbar, 1957; Noyon *et al.*, 2009, Dalpadado, 2002) which may also affect genetic diversity levels.

#### 4.6.2 Patterns of mitochondrial genetic structure

In *T. libellula*, populations from Canadian and Greenlandic sites differed by 3% at the COI gene. Even if the sample sizes from Greenland sites were small, these stations harboured a different composition of haplotypes. In addition, in Gibbs Fjord we have recorded the presence of Greenlandic haplotypes. Distinction have also be found between Greenlandic and Canadian waters in the Capelin (Dodson *et al.*, 2007) and in the halibut fish *Reinhardtius hippoglossoides* (Knusten *et al.*, 2007), suggesting a complex history of this region. An increased sampling among both sites of Davis Strait will help to better assess genetic boundaries between Atlantic and Arctic waters and especially the impact of the currents on Canadian and Greenlandic populations.

Higher levels of intraspecific divergence were found in *T. abyssorum* (up to 4%) but these differences were not associated with differences between specific sampling locations. This high intraspecific divergence is equivalent to the divergence found between two well-recognised species of *Themisto*: *Themisto compressa* and *Themisto gaudichaudii* (3.89%) in another study (Chapter III). This leads to the problem of defining a species based on the percentage of divergence of a mitochondrial gene only. Several studies have proposed a 16% threshold to discriminate crustaceans' species (Lefébure *et al.*, 2006; Matzen da Silva *et al.*, 2011). A more recent estimate of threshold for species discrimination in amphipods was estimated at 7% (Chapter II). Since the analysis of SNPs did not identify two separate groups corresponding to the most divergent COI sequences, it is unlikely that the two mitochondrial groups belong to distinct species.

During the last Pleistocene glaciations, Arctic waters were mostly covered by ice, forcing marine species to move to southern places or in ice free refugia (Hewitt, 2004; Hardy *et al.*, 2011). Differences in genetic diversity levels observed between *T. abyssorum* and *T. libellula* could stem from the fact that the former species is found in deeper waters (Wencki, 2000) and hence may have survived the glaciations in a deep-sea refugium. Species associated with depth did not experience the same level of demographic contraction compared to shallow water species (Allcock and Strungell, 2012). To our knowledge, the presence of deep-sea refugium in the Arctic is not known but recent works suggest the presence of non-glaciated area during the Pleistocene in the Canadian Archipelago (England and Furtz, 2008; England *et al.*, 2009).

The estimated time since expansion (1000-50 years ago) was not concordant with an expansion after the last glaciation (around 10000 years) in both species. Nevertheless, we cannot exclude that an inappropriate mutation rate could bias the dating of the expansion event (Grant *et al.*, 2012). In the absence of fossils, the calibration of a molecular clock is often impossible, and a proxy mutation rate is used instead (Marko, 2002). Here, we used the general estimate proposed by Knowlton and Weigt (1998) for crustaceans because the mutation rate in COI for the *Themisto* genus is unknown. We cannot exclude that this estimate is not appropriate for this group because it has been reported that some crustacean groups might evolve faster than others (Thomas *et al.*, 2010). Therefore, cautions must be taken regarding our estimates of the time since the expansions. Other substitution rates proposed for other crustaceans (amphipod: 9.6%/My; Henzler, 2006 or in mysids 27%/My, Audzijonyte and Väinölä, 2006) were also tested for the Bayesian Skyline plot, resulting in a more recent timing in population change (Appendice B).

#### 4.6.3 Rad-tag

As no obvious barrier is characterized in the High Arctic oceanic region, we expected that the populations of these pelagic amphipods would be panmictic, as observed in other pelagic invertebrate in the Antarctic region (Deagle *et al.*, 2015). However, our analyses on neutral SNPs have clearly revealed the presence of subtle genetic differentiation in holozooplanktonic species at scale of the Archipelago. These results contrast with those obtained for a strictly pelagic mysidacea species *Euphausia superba* where no genetic differentiation was found around the Antarctic using SNPs markers (Deagle *et al.*, 2015). Discriminant analysis and Bayesian clustering method (e.g. STRUCTURE) results were largely congruent in the number of clusters found. Although, the number of *k* must be interpreted with caution especially from a biological point of view (Pritchard *et al.*, 2000; Meirmans, 2015), our results suggest two genetically distinct clusters for both species with different biological interpretation. No obvious ecological separation was attached to the genetic clusters identified for *T. abyssorum*. Moreover, as no geographic correlation was detected between these clusters, a potential vertical structure by depth can be considered. Vertical distribution is not uncommon in the ocean (Kosobokova and Hopcroft, 2010), but it is often missed in genetic studies. Several studies have demonstrated a vertical structuring distribution of genotypes in marine species like in the eurybathic amphipod *Eurythenes gryllus* (France and Kocher, 1996). More recently, Weiner and colleagues (2012) have documented the vertical distribution of three distinct genotypes of the foraminifer *Hastigerina pelagica* corresponding to vertical niches. For abyssal species like *T. abyssorum*, which is often associated with high depth (Wenki, 2000), a vertical distribution is plausible. A sampling design adapted to vertical stratification would be needed to test the hypothesis of genetic structuring with depth.

For the other Canadian arctic *T. libellula*, the discriminant analysis suggested five groups instead of the two clusters proposed by the STRUCTURE analysis. This

discrepancy might be caused by the different approaches employed by these two clustering methods. DAPC is thought to be more efficient in detecting genetic clustering than STRUCTURE, especially in weakly differentiated populations (Benestan *et al.*, 2015; Kanno *et al.*, 2011; Jombart *et al.*, 2010). Based on the DAPC, two main groups can also be identified, and are congruent with those identified with STRUCTURE. These two genetic clusters indicate a separation between the Amundsen Gulf and the Lancaster Sound corresponding to the ecological zones observed in the Canadian Archipelago, one being oligotrophic and the latter being eutrophic (Ardyna *et al.*, 2011). However, at the mitochondrial level, we were unable to detect this structure. Several reasons can explain this mito-nuclear discordance (review in Toews and Brelsford, 2012). A geographic isolation and then secondary contact can be considered. In fact, the Canadian Archipelago was advanced as a potential secondary contact zone in polychaetes (Hardy *et al.*, 2011). Secondary contact zones are characterized by a high mitochondrial diversity with an increase of allelic richness (Hardy *et al.*, 2011). Another explanation for the mito-nuclear discordance we observed might be sex biased dispersal. Sex bias asymmetry can be caused by male-biased dispersal, mating behaviour or sex-biased offspring production (Toews and Brelsford, 2012). In the Hyperiididae, a parasitic behavior has been described in some species (Laval, 1980; Gasca and Haddock, 2004), but it is less documented for the genus *Themisto* (Laval, 1980). It appears that the adult female is responsible for the dissemination of her limited progeny, seeking the correct hosts and releasing its marsupium on them (Gasca and Haddock, 2004). Medusae and salps are the main hosts for *Themisto* progeny (Laval, 1980). It would be interesting in the future to investigate their host connectivity pattern and genetic population structure in order to compare with those of their parasites. Differences in dispersal capability among hosts would affect the connectivity scales of their parasites. The migration of gelatinous zooplankton is not well documented.

The genetic clusters identified for *T. abyssorum* and *T. libellula* are not completely isolated. The low values of  $\phi_{ST}$  ( $\phi_{ST}=0.015$  for *T. abyssorum* and  $\phi_{SC}=0.02$  for *T. libellula*) we have estimated these two species suggest a high gene flow between populations. Low  $F_{ST}$  values are common in marine species (Knusten *et al.*, 2011). Several studies have also reported similar low  $F_{ST}$  values in species with planktonic larval stages such as oceanic threespine stickleback populations (Hohenhole *et al.*, 2010), in the Pacific lamprey (Hess *et al.*, 2013) or in the American lobster (Benestan *et al.*, 2015). An absence of significant  $F_{ST}$  was also reported in the holozooplankton Antarctic mysid (Deagle *et al.*, 2015), suggesting unlimited gene flow around the Antarctic.

We found a higher proportion of loci under balancing selection than under diversifying selection; this confirms the idea that balancing selection is more recurrent in the genome (Nielsen, 2005; Charlesworth, 2006). Genomic reference is lacking for non-model organisms and information relative to the localization of selected loci is often unknown (Helyar *et al.*, 2011). Therefore, more efforts are needed to characterize these regions probably involved in local adaptation. Several studies have demonstrated the presence of local adaptations in marine organisms with high dispersal (reviewed in Sandford and Kelly, 2011, Peijnenburg and Goetze, 2013), in particular because oceanic waters are neither homogenous in space nor in depth (Norris, 2000). It has been advanced that *Themisto* species might be associated to specific water masses (Dunbar, 1957; Auel *et al.*, 2002) or water temperatures (Wing, 1976). We suppose that this potential difference in water niche will also be visible at the genomic level. Therefore, linking ecological features to genomic resources among these arctic species will help to understand mechanisms that lead to speciation in oceanic environments, in particular for marine zooplankton, where selection is expected to be the main evolutive force acting (Peijnenburg and Goetze 2013).

The number of SNPs used for this analysis is less than the dozen of thousands SNPs used in other studies: (Benestan *et al.*, 2015; Deagle *et al.*, 2015). Since the capacity of detecting population structure depends on the total number of alleles examined (Helyar *et al.*, 2011), it leads to the question of how many markers are needed. It appears that several hundred SNPs are sufficient to detect genetic structure in non-model organisms (Morin *et al.*, 2009; Coates *et al.*, 2009, Turakulov and Easteal, 2003; Hess *et al.*, 2011; Andrews *et al.*, 2016) even when the sampling size is small (Willing *et al.*, 2012). It has recently been advanced that the use of more individuals (>100) will be more appropriate for the study of populations with subtle genetic differentiation ( $F_{ST} < 0.01$ ) (Benestan *et al.*, 2016). Despite the limited number of individuals per sampling sites, we have detected a more pronounced genetic structure in both species. Recent studies have also highlighted some drawbacks in SNP calling (Pan *et al.*, 2015; Deagle *et al.*, 2015; Andrews *et al.*, 2016; Rodriguez-Ezpeleta *et al.*, 2016). In particular, repetitive region are known to be a source of error in SNP calling (Ilut *et al.*, 2014; Deagle *et al.*, 2015). Recently, Rodriguez-Ezpeleta *et al.*, (2016) have showed that STRUCTURE analyses can be sensitive to the parameters used during SNP calling but not PCA; therefore showing the utility of combining the two analyses.

#### 4.7 Conclusion

The various analyses we carried out on mitochondrial and nuclear datasets indicated contrasting genetic structure patterns and more limited connectivity among *Themisto* species despite their similar dispersal abilities. The discordant population genetic patterns harbored by these species suggest different responses to extrinsic factors (Palumbi, 1996; Norris and de Vargas, 2000) as observed also among sea urchins species (Addison and Hart, 2005). To gain insight into dispersal patterns of holozooplanktonic species, it will be useful in the future to combine ecological

factors, temporal and vertical sampling information. Today, marine transport in the Arctic Ocean is increasing (Corbett *et al.*, 2010) as the risk of marine invaders (Molnar *et al.*, 2008; Chan *et al.*, 2013). Using genomic techniques to improve our knowledge of marine arctic populations will help maintain and monitor Arctic marine biodiversity.

#### 4.8 Acknowledgement

We wish to thank the officers and crew of the CCGS Amundsen during sampling efforts in 2011. We are grateful to the Natural Resources Institute of Nuuk, Greenland for providing samples. We are also grateful to the Institut de Biologie Intégrative et des Systèmes for their technical support. We thank Dr. Jérôme Laroche for his training on UNEAK and Rad-seq analysis.

## CHAPITRE V

### SPATIAL STRUCTURING IN A CRYPTIC AMPHIPOD FROM THE BERING AREA

Astrid Tempestini<sup>1</sup>, Alexei Pinchuck<sup>2</sup> et France Dufresne<sup>1</sup>

Manuscrit à soumettre à *Polar Biology*

<sup>1</sup> Université du Québec à Rimouski, Département de biologie et Centre d'études nordiques, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1

<sup>2</sup> School of Fisheries and Ocean Sciences, University of Alaska, 17101 Point Lena Loop Rd, Juneau, AK 99801, USA

Contribution des auteurs (initiales)

Mise en place des protocoles expérimentaux : AT

Récolte des données : AT

Analyse des données : AT

Interprétation des résultats : AT

Rédaction du manuscrit : AT

Commentaires et correction du manuscrit : AT, FD, AP

Financement: FD, AP



## 5.1 Résumé

Les populations d'espèces pélagiques sont souvent perçues comme génétiquement homogènes sur de très grandes échelles géographiques. Dans ce chapitre, nous examinons la structure génétique de l'espèce cryptique de l'Arctique nouvellement découverte: *Themisto sp.* Nos analyses basées sur 33 individus, provenant de quatre sites d'échantillonnage répartis au travers du Déroit de Béring et de l'Alaska, ont montré la présence d'une très forte structuration génétique. Les analyses des données mitochondriales et de Rad-tags ont montré la présence d'une forte structuration entre le Golf de l'Alaska et la mer de Béring. L'estimation du coefficient de différenciation ( $F_{ST}$ ) basé sur 4805 SNP entre les populations du Pacifique et de la Béringie suggère des échanges extrêmement limités ( $\phi_{CT}=0.26$ ) entre ces deux régions. Nos résultats indiquent la présence de barrière à la migration au large de l'Alaska, favorisant la spéciation marine des espèces pélagiques.

**Mots-clés:** Arctique; Génomique; Population; *Themisto*; Séquençage RAD; SNP; Béring

## 5.2 Abstract

Pelagic species are often perceived as genetically homogenous over large geographic scales. Here, we examine the genetic structure of the newly discovered cryptic arctic holozooplankton *Themisto sp.*. Our analyses of 33 individuals from four sites throughout the Alaskan waters show that zooplankton species can be highly genetically structured. Mitochondrial and Rad-sequencing analyses revealed strong spatial structuring between the Gulf of Alaska and the Bering Sea.  $F_{ST}$  based on 4805 SNPs between Pacific and Bering Sea populations indicated very limited exchanges ( $\phi_{CT}=0.26$ ) maybe due to the Pacific Rim acting as a barrier to gene flow. Our results suggest the presence of strong opportunities for marine speciation in open-ocean zooplankton in this region.

**Key-words: Arctic; Genomic; Population; *Themisto*; RAD sequencing; SNP , Bering**

### 5.3 Introduction

The lack of obvious barriers to dispersal in the marine environment has led to the paradigm of widespread and genetically homogenous marine populations (Palumbi, 1994). A wealth of studies using molecular markers have revealed the presence of genetic structure in pelagic species, suggesting that barriers to gene flow is more common in open-oceans (Palumbi, 1994; Bucklin *et al.*, 1997; Waters and Roy, 2004; Goetze, 2005; Levin, 2006). Marine genetic population studies have traditionally relied on the use of mitochondrial gene sequences and on a limited number of nuclear markers such as microsatellites (Reitzel *et al.*, 2013). The development of genotype-by-sequencing techniques has enabled the scrutiny of thousands of nuclear markers throughout the genome but in a more limited number of individuals than with traditional markers due to sequencing costs (Holenhole *et al.*, 2010; Davey *et al.*, 2011; Reitzel *et al.*, 2013; Benestan *et al.*, 2015; Deagle *et al.*, 2015, Andrews *et al.*, 2016). This technique offers the possibility to reveal the presence of subtle genetic differentiation that was not observed with other types of markers. Recently, Blanco-Bercial and Bucklin (2016) have discovered using rad-seq, an unsuspected genetic structure in the pelagic copepod *Centropages typicus* over the North Atlantic Ocean. To date, this technique has not been used to investigate genetic structure in the Arctic Ocean.

*Themisto* species are the most abundant crustaceans in the Arctic Ocean (Dunbar, 1957; Bowman, 1960; Wing, 1976; Auel *et al.*, 2002) and play an important role in marine arctic food webs (Darnis *et al.*, 2012). *Themisto* species are holozooplankton (i.e. organisms that spend their entire life in the water column), which implies a high potential for dispersal over great distances (Grosberg and Cunningham, 2001). The holoplanktonic lifestyle of these amphipods suggests that their populations could be genetically homogenous throughout the Arctic Ocean but it has not been verified. A recent molecular phylogeny (Chapter III) has shown the presence of a cryptic

*Themisto* species in the Bering, referred as *Themisto sp.* that was genetically distinct from *T. libellula* (18%) and from *T. pacifica* (11%) at the cytochrome I (COI). Since this species was always confounded with its congeners, no information on its genetic structure is known. Recently, southern species have expanded their distribution into the Arctic Ocean due to global warming (Kraft *et al.*, 2013b), increasing the emergency to better characterize the current arctic biodiversity.

The Bering Strait is the only connection between the Pacific and Arctic Oceans, allowing northward pacific input of water into the Arctic Ocean. Moreover, the complex history of the region has influenced marine arctic species assemblages and populations (Hewitt, 2004; Coyer *et al.*, 2009; Hardy *et al.*, 2011). At least six openings of the Bering Strait have been documented (reviewed in Vermeij, 1991; Cunnigham and Collins, 1998; Heads, 2005), allowing Pacific invasion into the Arctic Ocean. In addition, during the Quaternary glaciation, the Bering Strait was a refugium for numerous marine species (Hardy *et al.*, 2011). All these factors have led to the complex genetic diversity of marine species in this area, contributing to divergence and the isolation between pacific and arctic populations (Dodson *et al.*, 2007; Hardy *et al.*, 2011), as it has been described for the isolated Bering and Arctic populations of the copepod *Calanus glacialis* (Nelson *et al.*, 2009).

The aim of this study was to characterize the population structure and connectivity of *Themisto sp.* populations from the Chukchi Sea, northern and southern Bering Sea and from Prince William Sound, Alaska, combining both the mitochondrial cytochrome oxidase I (COI) gene and nuclear RAD sequencing (rad-seq). As well, we tested for genetic admixture between *Themisto sp.* and its congener *T. libellula* using SNPs obtained from rad-seq. We discuss our results in the light of cryptic speciation and genetic structure in the open-ocean.

## 5.4 Material and Methods

### 5.4.1 Samples and DNA extraction

Samples from the Bering area were provided by A. Pinchuck. The 204 samples of *Themisto libellula* were collected with 500 and 750 µm mesh size horizontal or vertical nets from various locations in the Arctic Canadian Archipelago during the ArcticNet expedition in 2011 as described in Chapter IV. Samples were identified through the species level with appropriate taxonomic keys (Vinagrov *et al.*, 1996; Martins et Davis, 2001). Genomic extractions were performed on 33 individuals of *Themisto sp.* (Figure 5.1) with an E.Z.N.A kit (omega-biotek) according to the manufacturer's instructions. Each DNA extraction was treated with 5 µL of RNase-A solution (10 mg/mL, sigma Aldrich).

### 5.4.2 Mitochondrial analysis

#### 5.4.2.1 Mitochondrial amplification

Genetic variation of the 33 individuals was screened at one mitochondrial DNA (mtDNA) region. The cytochrome oxidase c I mtDNA (COI) was amplified by polymerase chain reaction (PCR) for 6-12 individuals per sampling site (Table 5.1) using the protocol from Dionne *et al.*, (2011) and the universal primers LCO1490 and HCO2198 from Folmer *et al.*, (1994). PCR products were checked on 1.5 % agarose gel and unidirectionally sequenced by Genome Québec (McGill University, Canada). DNA sequences were verified by eyes and aligned with MUSCLE (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011).

**Tableau 5.1** Coordinates of the sampling sites, and number of individuals used in this study for each site. Geographic coordinates are in decimal degree.

	<b>Latitude</b>	<b>Longitude</b>	<b>N</b>	<b>Accession</b>
<b>Chukchi (CHU)</b>	69.91225	-171.988202	8	XX
<b>North East Bering (NEB)</b>	62.312308	-173.809573	7	XX
<b>South East Bering (SEB)</b>	54.407991	-166.675937	6	XX
<b>Prince William Sound (PWS)</b>	60.731934	-147.102874	12	XX

#### 5.4.2.2 Genetic diversity

Number of haplotypes (H), haplotype diversity (h) and nucleotide diversity ( $\pi$ ) were calculated for mtDNA COI sequences from each location using DNAsp v5 (Librado and Rozas, 2012). Genetic variability within and among sample locations was investigated using two different approaches. First, pairwise  $\phi_{ST}$  values between samples were estimated with a significance level of 0.05 determined by 10,000 permutations. Secondly, a global analysis of hierarchical molecular variance (AMOVA) was performed to test the distribution of the variation at different levels. All these analyses were conducted in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). For each species, haplotype networks were constructed to evaluate intraspecific relationships among all haplotypes using the Median-Joining (MJ) algorithm implemented in Popart (<http://popart.otago.ac.nz/index.shtml>). We tested for the presence of a correlation between geographic and genetic distances using a Mantel

test. Matrices of geographic distances between sampling locations were calculated with the geographic distance matrix generator software (Ersts, 2016). Linear Slatkin genetic distances were generated in Arlequin v.3.5.2 (Excoffier *et al.*, 2005). Statistical significance of the correlation was assessed with a 1000 times permutation test.

#### 5.4.2.3 Demographic history

For each species, demographic history was inferred using neutrality tests, mismatch distribution and Bayesian Skyline Plot. Tajima's D test (Tajima, 1989), Fu and Li F\* and Fu and Li D\* tests (Fu and Li, 1993), which allow to test a sign of population expansion or contraction, were performed on the entire mitochondrial dataset of each *Themisto* species. Time changes in population size were estimated for each species using a Bayesian skyline method implemented in BEAST 1.8.2 (Drummond *et al.*, 2012). The DNA substitution model was identified according to the best Bayesian information criterion (BIC) using Topaliv2 (Milne *et al.*, 2004). Chains were run for 10 million steps, and convergence was assessed when effective sample sizes (ESS) reached at least 200. For the time scale, a strict molecular clock was used with a rate of 2 % / My as proposed for crustaceans by Knowlton and Weigt (1998). All other parameters were automatically adjusted. Outputs were analysed in Tracer 1.5 (Rambault *et al.*, 2014).

#### 5.4.3 RAD-seq

##### 5.4.3.1 SNP identification

Following libraries preparation, samples of *Themisto sp.* were pooled with *T. libellula* and randomly put on plates in order to avoid bias in base calling. Pooled genomic DNA was cut by the PstI and MstI restriction enzymes and 100 bp paired ends were sequenced on an Illumina HiSeq platform at the Institut de Biologie Intégrative et des Systèmes (IBIS, Laval University, Quebec, Canada). Hi-seq Illumina output was used

to genotype individuals according to the UNEAK pipeline available in TASSEL 3.0 (Bradbury *et al.*, 2007) as described in White *et al.*, (2013). Briefly, the UNEAK pipeline first trimmed all the reads to 64 bp long. Then, identical reads were grouped into tags. All tags were compared to find candidate SNPs with 1 bp mismatch between paired tags. Finally, a network filter step was used to discard complicated networks, which are usually a mixture of repeats, paralogs and error tags (Lu *et al.*, 2013). The reciprocal pairs of tags retained after this step were considered as valid SNPs. Following SNP identification, individuals were scored for their genotype, allowing a sequencing error rate of 3%. In TASSEL 5.0 (Glaubitz *et al.*, 2014), loci that were called in at least 70% of individuals were retained. In order to remove SNPs from sequencing errors (Glaubitz *et al.*, 2014), we set a minimum allele frequency to 0.05. As paralogous sequences can inflate heterozygosity (Pan *et al.*, 2015) and therefore can bias analyses, we also removed loci with  $H_e > 0.5$  and individuals with more than 30 % of missing value. Because genotyping errors can lead to false conclusions, we removed loci that deviated from the Hardy-Weinberg equilibrium (Hoskin *et al.*, 2004). Hardy-Weinberg equilibrium tests were conducted on each locus for all and within each collection site with the adegenet package (Jombart, 2008; Jombart and Amel, 2011) available in R. SNPs with a p-value  $< 0.0001$  in any site, or  $< 0.001$  in multiple collection sites, were discarded (p-value computed using 999 simulations).

#### 5.4.3.2 Detection of outliers

As phylogeographic parameters estimates often rely on the assumption of neutral markers, we detected outliers SNPs to remove them from the analysis, as suggested by Beaumont and Nichols (1996). Using a Bayesian approach, we identified loci under selection with BayeScan 2.1 (Foll and Gaggiotti, 2008) as described in Glover *et al.*, (2012) with a prior odd of 100000 and default parameters. Graphical outputs were visualized in R.

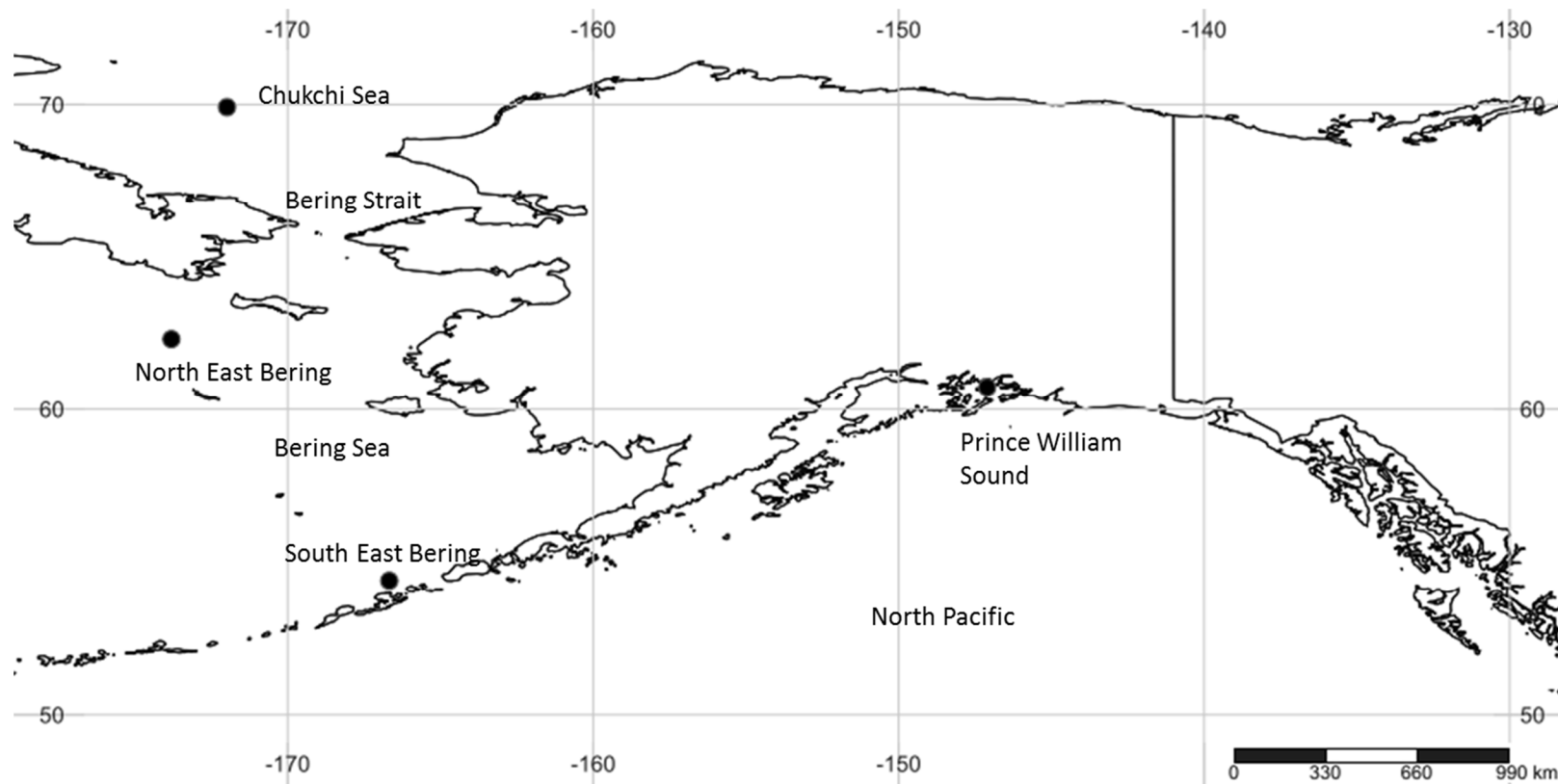


#### 5.4.3.3 Population structure and assignment tests

We first verified the presence of hybrids between *T. libellula* and *Themisto sp.* using discriminant analysis of principal component (DAPC) available in the adegenet (Jombart, 2008; Jombart and Ahmel, 2011) package of R, and the Bayesian assignment method implemented in STRUCTURE v2.3.2 (Pritchard *et al.*, 2000). STRUCTURE aimed to assign individuals to K putative populations based on their genotype. For this purpose, we assumed an admixture model, uncorrelated allele frequencies and we did not use population priors. We performed this analysis by varying from K=2 to K=10 and conducted 5 runs to check the consistency of the results with an initial burn-in of 10 000 steps followed by 10 000 of Markov Chain Monte Carlo repetitions. We also conducted the same analysis on *Themisto sp.* only in order to identify potential populations. STRUCTURE results were analyzed using the online version of STRUCTURE HARVESTER (Earl and vonHoldt, 2012; <http://taylor0.biology.ucla.edu/structureHarvester>). The best number of clusters was chosen according to the method of Evanno *et al.*, (2005). Results were visualized with the web application STRUCTURE PLOT (Ramasamy *et al.*, 2014; <http://btismysore.in/strplot/index.php>).

#### 5.4.3.4 Genetic diversity and $F_{ST}$

Pairwise  $F_{ST}$  were calculated and significance was assessed with a 10000 times permutation test. Analyses of molecular variance (AMOVA) were performed in order to assess the distribution of genetic diversity among and within populations. All these analyses were performed in Arlequin v3.5.2 (Excoffier *et al.*, 2005).



**Figure 5.1** Map of the study area of *Themisto* sp. Black circles represent sampling locations: Prince William Sound (PWS), South East Bering (SEB), North East Bering (NEB) and Chukchi Sea (CHU).

## 5.5 Results

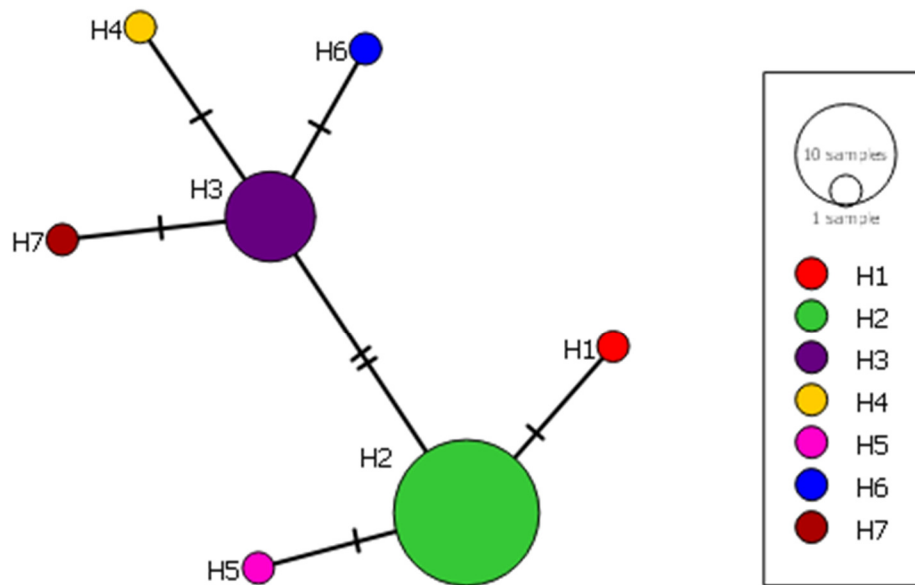
### 5.5.1 Mitochondrial analysis

The 627 bp of the COI mitochondrial gene sequences for 33 *Themisto sp.* revealed 7 variable sites. The uncorrected pairwise distances between sequences ranged from 0 to 0.6%. Prince William Sound harboured the highest haplotypic ( $0.576 \pm 0.16$ ) and nucleotide diversity ( $0.00133 \pm 0.0005$ ) compared to the other sampling sites (Table 5.1). Southeast Bering represented the least diverse region, with a single haplotype found over the six individuals (Table 5.2).

Mitochondrial networks consisted of 7 haplotypes for *Themisto sp.* with one haplotype (H2) spread throughout the sampled stations (Figures 5.2 and 5.3). Haplotype 1 and haplotype 6 were limited to Chukchi and Northeast Bering seas. All haplotypes were separated by only one or two mutations. Two groups of haplotypes were found: one including haplotypes 3, 4, 6, and 7 mostly restricted to Prince William Sound and the remaining haplotypes composing the other group.

**Tableau 5.2** Results of the mitochondrial analysis of cytochrome oxidase I (COI) for *Themisto sp.* N= number of individuals, H= number of haplotypes, h= haplotypic diversity,  $\pi$ =nucleotide diversity. Standard deviations (sd) are in parentheses.

<i>Themisto sp.</i>	N	H	h ( $\pm$ sd)	$\pi$ ( $\pm$ sd)	G+C content
<b>All sampling sites</b>	33	7	0.587 ( $\pm$ 0.081)	0.00195 ( $\pm$ 0.00029)	0.324
<b>Chukchi sea (CHU)</b>	8	2	0.25 ( $\pm$ 0.18)	0.0004 ( $\pm$ 0.00029)	0.324
<b>North East Bering (NEB)</b>	7	2	0.286 ( $\pm$ 0.196)	0.00046 ( $\pm$ 0.00031)	0.324
<b>South East Bering (SEB)</b>	6	1	0	0	0.324
<b>Prince William Sound (PWS)</b>	12	5	0.576 ( $\pm$ 0.16)	0.00133 ( $\pm$ 0.0005)	0.324



**Figure 5.2** Haplotype network based on 627 bp of mitochondrial cytochrome c oxidase I sequences obtained for *Themisto sp.* Circle sizes are proportional to number of individuals observed for that haplotype and colour codes refer to a haplotype. The number of traits on a line corresponds to the number of mutations between sequences.

**Tableau 5.3** Pairwise  $\phi_{ST}$  between sampling locations for each *Themisto sp.*  
 \*\* significant at  $p < 0.001$ .

	<b>CHU</b>	<b>NEB</b>	<b>PWS</b>	<b>SEB</b>
<b>CHU</b>	0			
<b>NEB</b>	0	0		
<b>PWS</b>	0.73168**	0.73168**	0	
<b>SEB</b>	-0.04025	-0.04025	0.73990**	0

Pairwise  $\phi_{ST}$  between sampling sites were null except with Prince William Sound (Table 5.3, where all  $\phi_{ST}$  were high (0.73) and significant. No significant correlation between geographic and genetic distance was detected with the Mantel test ( $r = 0.37$ ,  $p > 0.1$ ). AMOVA analysis performed on one group of *Themisto sp.* showed that the majority of the genetic variation (70%) occurred among sampling sites and less variation occurred within sampling sites (Table 5.4).

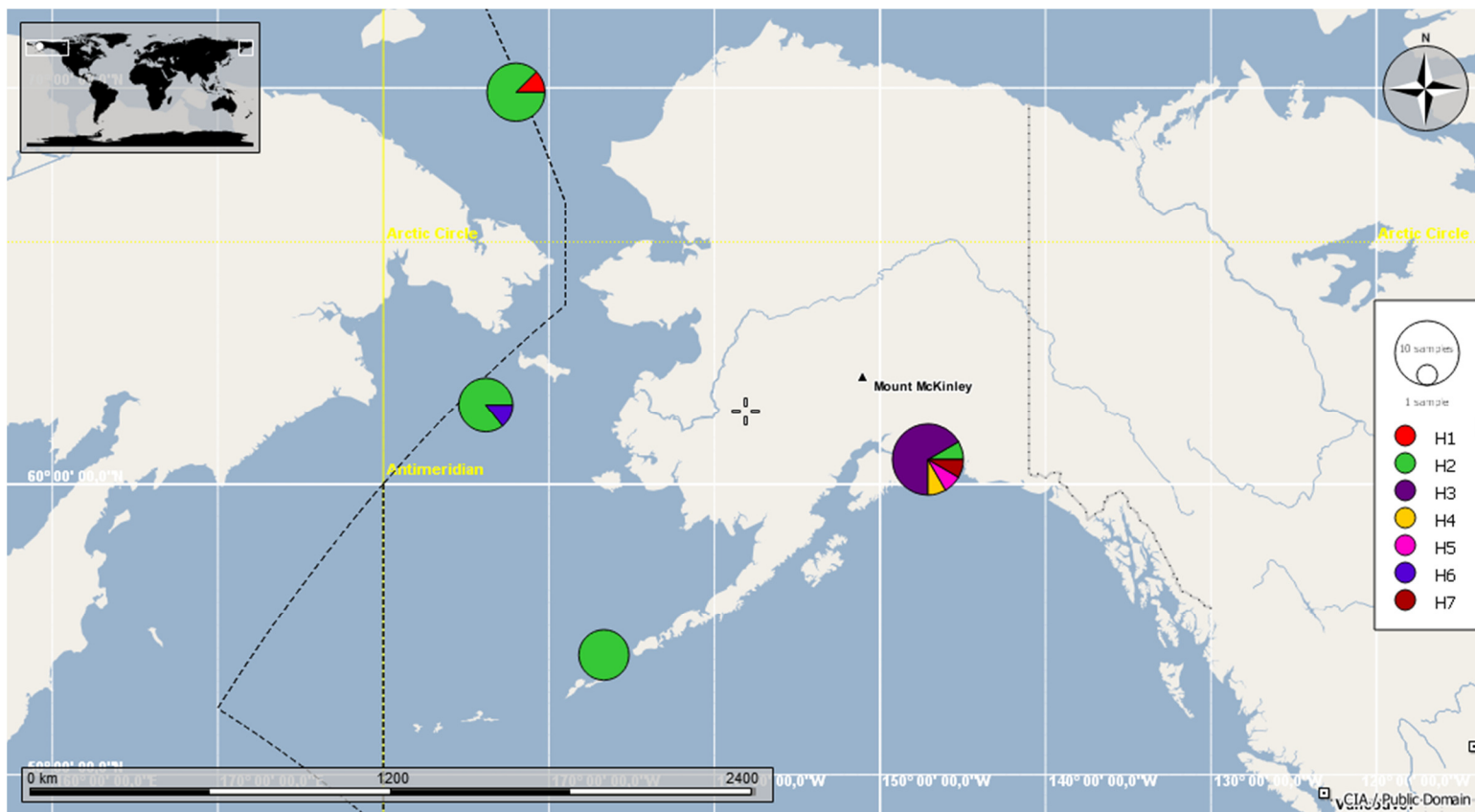
**Tableau 5.4** Analysis of Molecular Variance on mitochondrial dataset of *Themisto sp.*. \*\* significant at  $p < 0.001$ .

	<b>Source of variation</b>	<b>Percentage of variation</b>
<i>Themisto sp.</i>	Among sampling location	70.8
$\phi_{ST} = 0.7$ **	Within sampling location	29.2

**Tableau 5.5** Tajima's D, Fu's Fs neutrality tests calculated on the 627 bp of COI. \* significant at  $p < 0.05$ .

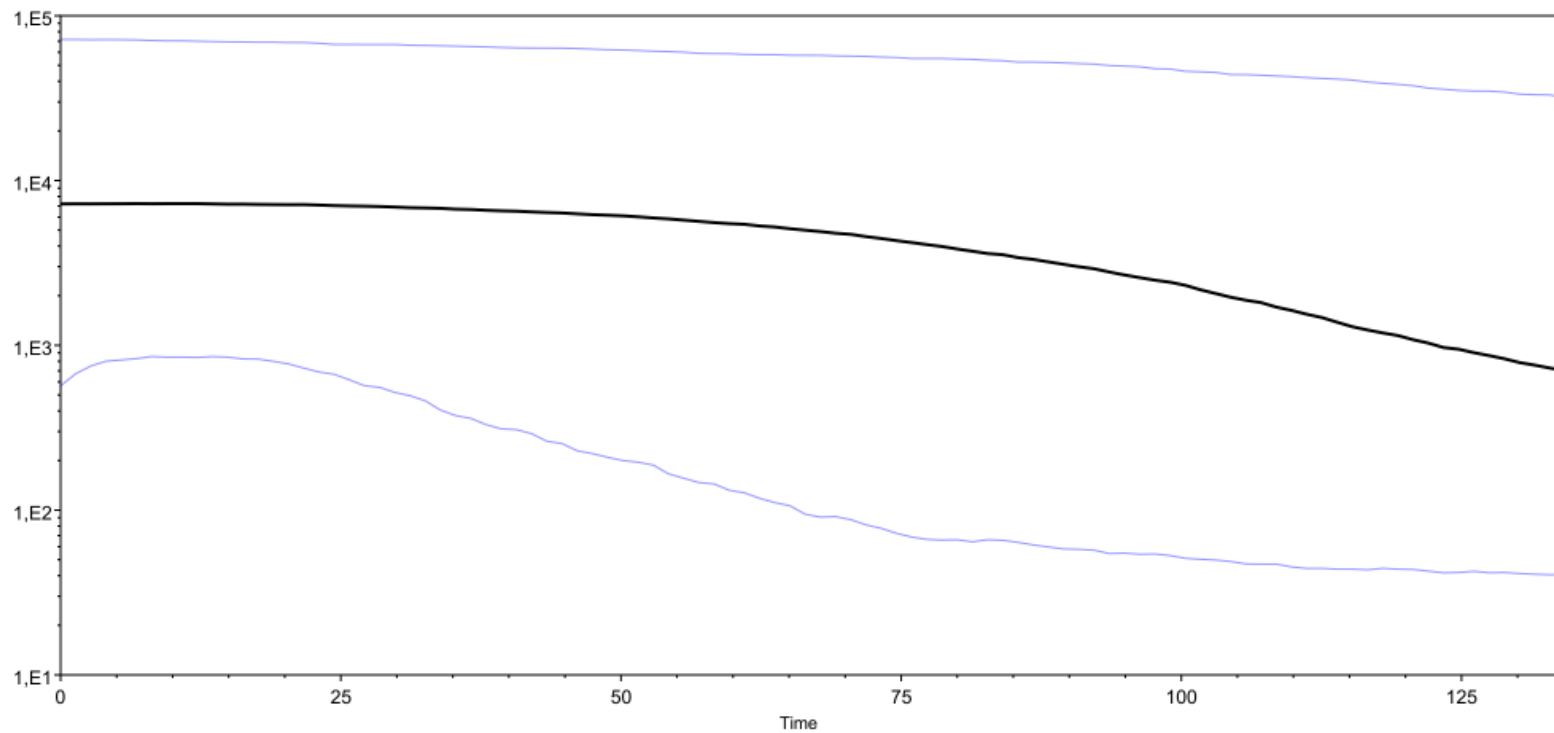
	<b>Tajima's D</b>	<b>Fu and Li D*</b>	<b>Fu and Li F*</b>
<i>Themisto sp.</i>	-0.84609	-2.29569	-2.16495
<b>Region PWS</b>	<b>-1.83094*</b>	<b>-2.22902*</b>	<b>-2.40959*</b>
Region (NEB+SEB+CHU)	-1.51414	-2.08204	-2.21420

The results of Tajima's D and Fu and Li D\* and F\* tests to infer the demographic history are shown in Table 5.5. Significant negative values were only observed for the regional group of Prince William Sound, suggesting a population expansion as confirmed by the Bayesian skyline analysis (Figure 5.4). Effective population size of *Themisto sp.* is estimated around  $10^4$  individuals and the time since the expansion is estimated around 50 years.



**Figure 5.3** Spatial distribution of haplotypes among sampling sites for *Themisto sp.* Circle sizes are proportional to the number of individuals observed for that haplotype and colour codes refer to a haplotype.





**Figure 5.4** Bayesian Skyline Plots of effective population size through time obtained under the F81 model (BIC= 2161.04). Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year.

### 5.5.2 Rad-tag analysis

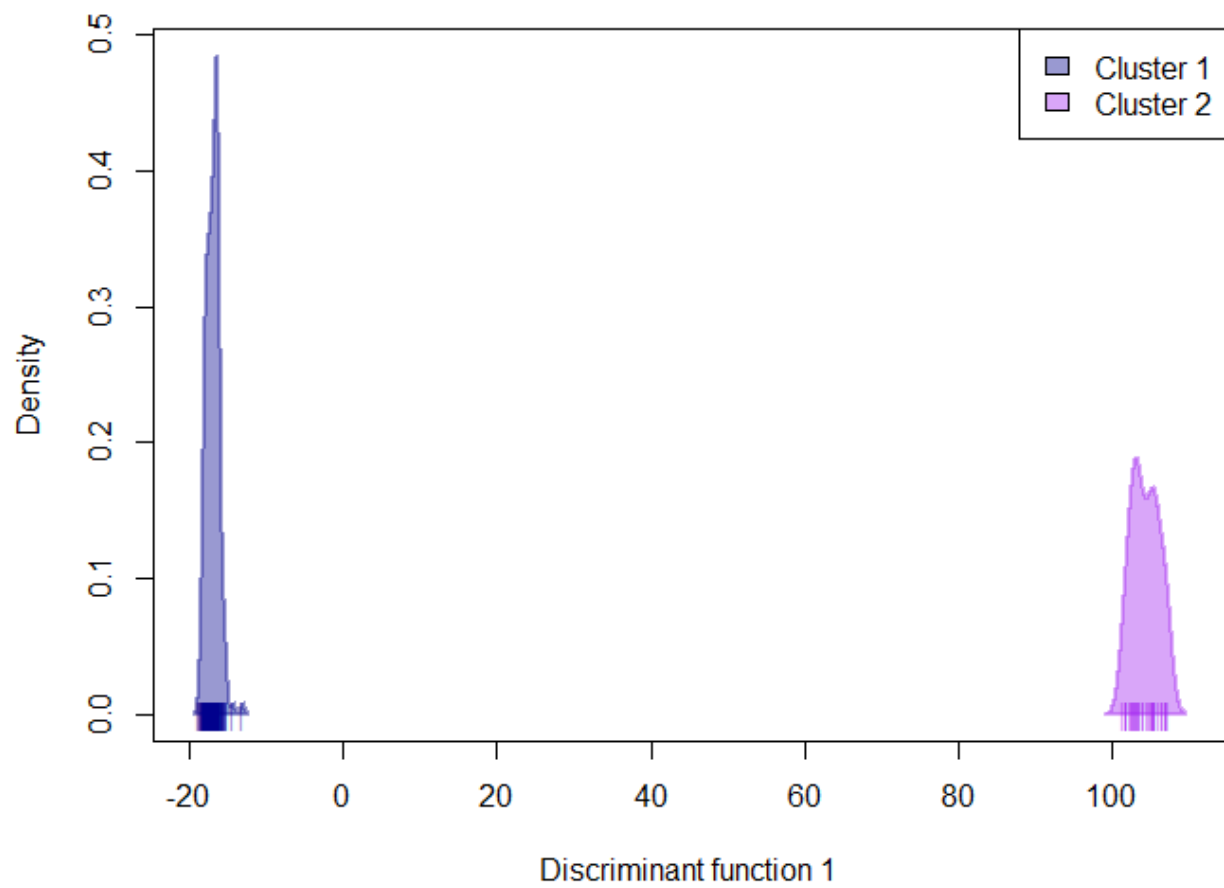
RAD sequencing produced 449 183 454 reads for *T. libellula* / *Themisto sp.* As the genome size of *T. libellula* is known (C= 3,97pg, Rees *et al.*, 2007), we estimated the sequencing depth for each *Themisto* dataset to 10X. Number of reads for each individual ranged from 31 995 to 4 487 248, probably due the variable quality of DNA among individuals. Prior to filtering, a total of 83116 loci for *T.libellula/Themisto sp.* were identified as potential loci for this analysis. After filtering, the final dataset consisted of 3114 SNP common to *T. libellula* and *Themisto sp.* and 6130 loci for *Themisto sp.* (Table 5.6). After removing loci not in Hardy-Weinberg equilibrium, 893 loci for *T. libellula* and 4908 for *Themisto sp.* were used for further analyses.

**Tableau 5.6** Number of putative loci retained following each filtering step.

Step filter	<i>T.libellula/Themisto sp.</i>	<i>Themisto sp.</i>
Initial	83 116	
Minor allele frequency >0.05	6787	6130
>85% of population		
Hardy-Weinberg		
All	1216	5032
Population	893	4908

#### 5.5.2.1 Hybrids between *Themisto libellula* and *Themisto sp.*

Two distinct clusters were identified based on DAPC and STRUCTURE (Figures 5.5 and 5.6). Each cluster corresponded to a single species, ruling out the possibility of hybridization between *Themisto libellula* and *Themisto sp.*



**Figure 5.5** Discriminant analysis based on the 893 loci found in both *Themisto libellula* and *Themisto sp.*



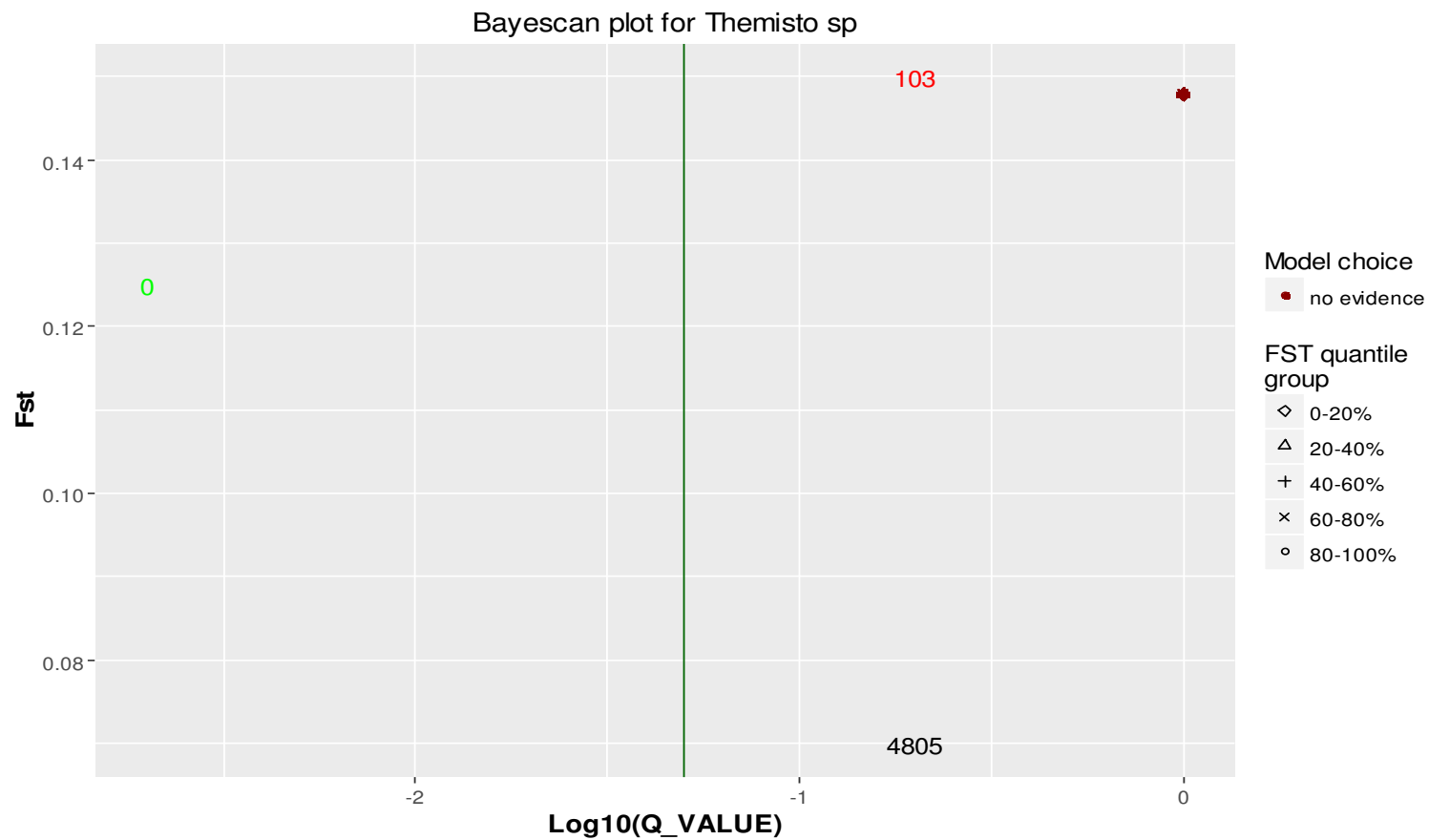
**Figure 5.6** Structure plot based on 893 loci found in *Themisto libellula* and *Themisto sp.*, under the admixture model. Distinct colors represent estimated cluster.

### 5.5.2.2 Population genetic of *Themisto sp.*

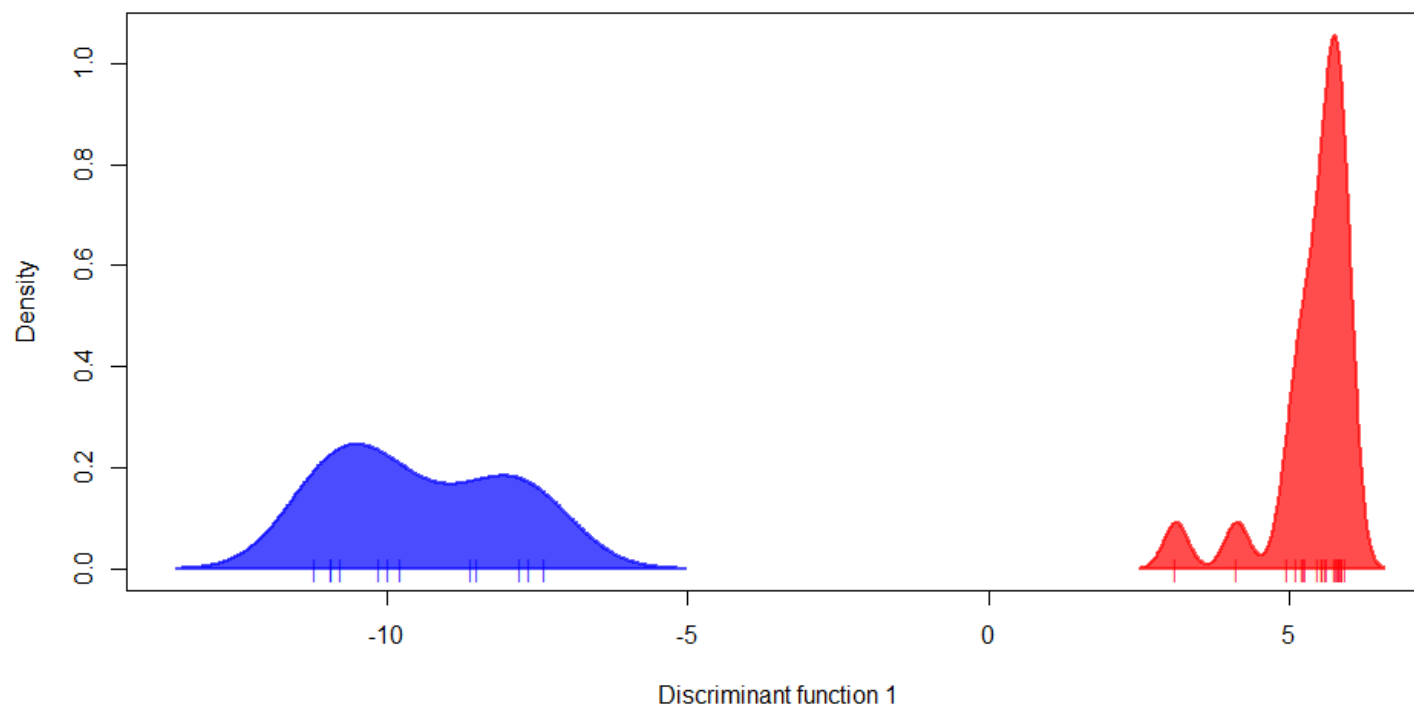
Putative loci under selection were detected using BayeScan and a false discovery rate set to 0.05 (Figure 5.7) for each *Themisto* species. We detected over 4908 loci, of which 4805 were neutral and 103 loci were under balancing selection. Based on discriminant analysis, *Themisto sp.* showed two distinct clusters (Figure 5.8). Analysis of the STRUCTURE suggested K=2 for *Themisto sp.* as the best values for the number of populations (Table 5.7). All the results were congruent with the DAPC. *Themisto sp.* harboured two clearly distinct groups (Figure 5.9): one from the Prince William Sound (in yellow) and another from the Bering area (in orange). Assignment analyses suggested limited exchanges between Prince William Sound and the Bering area as confirmed by the  $F_{ST}$  estimate between these two groups ( $\phi_{sc}=0.26$ ;  $p<0.05$ ; Table 5.8).

**Tableau 5.7** Evanno table for *Themisto sp.* In bold: the best K according to the Evanno method.

	K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
<i>Themisto sp.</i>	1	5	-140828.6	40.2457	NA	NA	NA
	<b>2</b>	<b>5</b>	<b>-118726.3</b>	<b>112.1178</b>	<b>22102.24</b>	<b>30112.84</b>	<b>268.58204</b>
	3	5	-126736.9	12917.899	-8010.6	15887.9	1.229914
	4	5	-118859.6	172.6046	7877.3	7942.08	46.013131
	5	5	-118924.4	430.6058	-64.78	2.62	0.006084
	6	5	-118991.8	659.4125	-67.4	129.96	0.197085
	7	5	-119189.2	831.4603	-197.36	460.28	0.55358
	8	5	-118926.2	511.2698	262.92	225.04	0.440159
	9	5	-118888.4	86.5021	37.88	3021.24	34.926783
	10	5	-121871.7	6866.5638	-2983.36	NA	NA

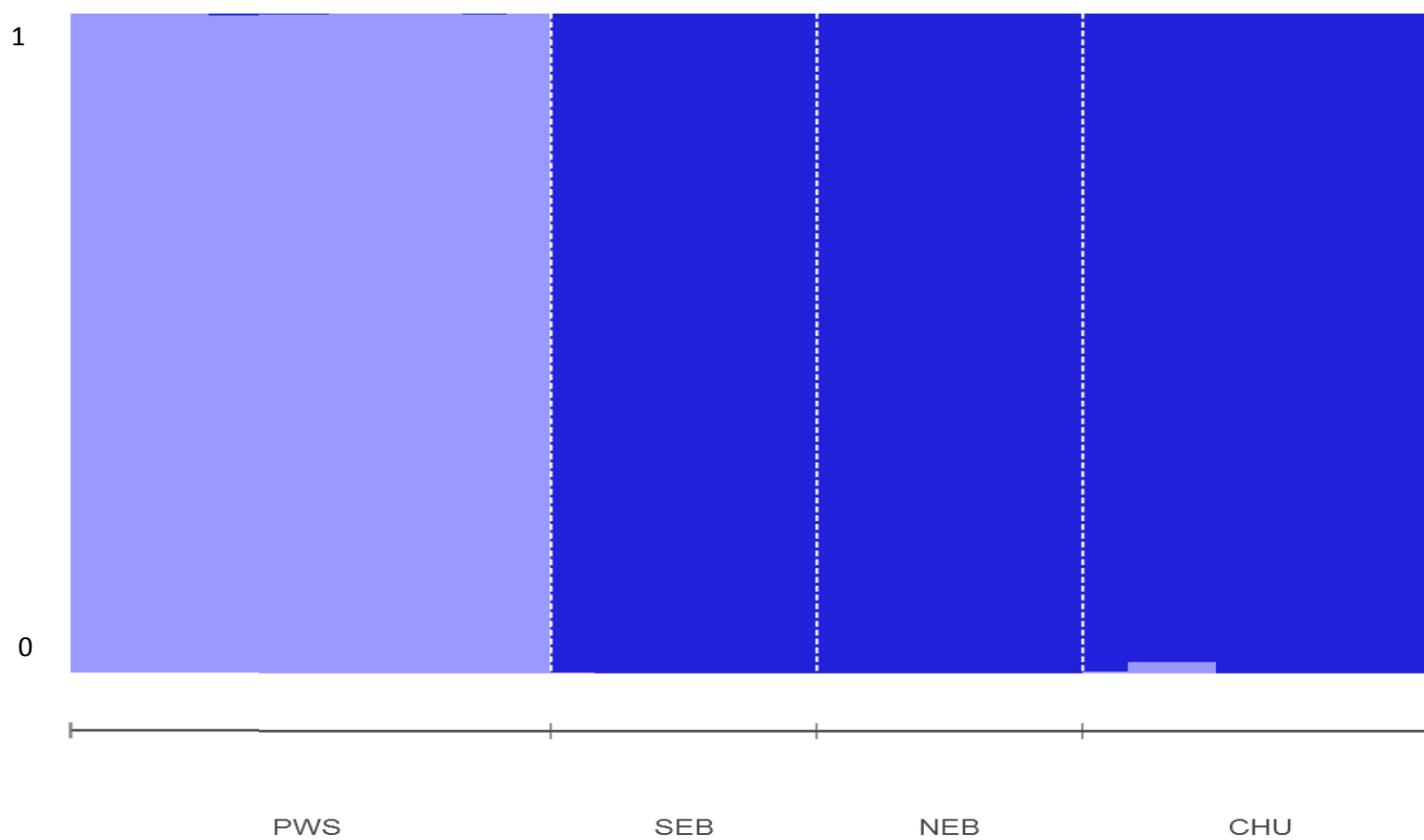


**Figure 5.7** BayeScan test for selection on individual SNPs among the two populations of *Themisto sp.* implemented in the BayeScan program. The red number represents SNPs potentially under balancing selection and the green number represents SNPs potentially under diversifying selection.



**Figure 5.8** Discriminant analysis of Principal Component for *Themisto sp.* performed with the 4805 loci.





**Figure 5.9** Structure plot using the SNP dataset with two clusters for *Themisto sp.*, under the admixture model. Distinct colors represent estimated clusters; sampling sites are separated by a black line.

As the mitochondrial and the clustering methods indicated clearly two groups, we also performed the AMOVA using this grouping for the two regions (Table 5.8). Very strong structuring was revealed between the Prince William Sound and the Bering areas (26.7%;  $\phi_{CT}$ = 0.26;  $p<0.05$ ). Small but significant differences among Bering populations was also observed (3%,  $\phi_{SC}$ =0.00035;  $p<0.05$ ).

**Tableau 5.8** Analysis of Molecular Variance (AMOVA) for *Themisto sp.* \*\*= significant value  $p<0.05$

Source of variation	Percentage of variation
Among groups	26.64 $\phi_{CT} = 0.26670^{**}$
Among populations within groups	0.03 $\phi_{SC} = 0.00035^{**}$
Within populations	73.33 $\phi_{ST} = 0.26645$

## 5.6 Discussion

The present study revealed the genetic distinctiveness of *Themisto sp.* at the nuclear level from its arctic congener *T. libellula*. In addition, deep genetic structure was found at ocean basins scale at both the COI gene and nuclear SNPs in this holozooplanktonic invertebrate. Our first SNPs results in this newly discovered species suggest homogeneous populations in the Bering area. Here, we discuss how our results benefit to our understanding of seascape genetic and arctic diversity.

### 5.6.1 Cryptic speciation

Our SNPs results suggest that *Themisto sp.* and its arctic congener *T. libellula*, do not hybridize and are distinct at the nuclear level. As no reference genome is available for *Themisto* or closely related species, we were not able to identify genomic regions potentially involved in speciation. Genomic reference is lacking for a vast number of non-model organisms and information relative to the localization of loci is often unknown (Helyar *et al.*, 2011). Several studies have shown the presence of local adaptations in marine organisms with high dispersal (Goetze, 2005; reviewed in Sandford and Kelly, 2011, Peijnenburg and Goetze, 2013). Ecological differences in temperature and diet preferences have been identified among *Themisto* species (Dunbar, 1957; Wing, 1976; Auel *et al.*, 2002; Dalpadado *et al.*, 2008). Therefore, linking ecological features to genomic data among these arctic species will help to understand mechanisms that lead to speciation in oceanic environments, in particular for marine zooplankton, where selection is expected to be the main evolutive force acting (Goetze, 2005; Peijnenburg and Goetze, 2013). In the near future, it will be interesting to characterize the location of these SNPs to see if these regions are linked to the thermal niches of *Themisto* species.

### 5.6.2 Genetic structure

Both SNPs and mitochondrial sequences revealed deep separation between populations from Prince William Sound and the Bering Sea. Global  $\phi_{ST}$  was very high (mt  $\phi_{ST}=0.7$ , p-value<0.001; SNP  $\phi_{CT}=0.26$ , p-value<0.05), suggesting very restricted genetic exchanges. Low  $F_{ST}$  values are common in marine species (Knusten *et al.*, 2011), and to our knowledge, no  $F_{ST}$  values as high have been reported in holozooplankton species based on the use of thousand SNPs. Another study based on microsatellites found a weaker but significant differentiation between these two regions in the walleye pollock (O'reilly *et al.*, 2004). Even in the close relatives *T. abyssorum* and *T. libellula*, in the Canadian Arctic Archipelago, we have not detected such a high level of differentiation (Chapter IV). Other studies have reported limited exchanges between these regions in the Giant Pacific Octopus *Enteroctopus dofleinitopus* (Toussaint *et al.*, 2012), in the pacific halibut *Hippoglossus stenolepis* (Nielsen *et al.*, 2010), in the walleye Pollock *Theragra chalcogramma* (Mulligan *et al.*, 1992) or in the Pacific cod *Gadus macrocephalus* (Spies, 2012). These results suggest that, despite constant northward inflow from the Pacific, populations are kept isolated. Two hypotheses can be considered: isolation due to historical events like glaciations and ecological differentiation.

Pleistocene glaciations are known to be one of the last climatic changes to have influenced the genetic structure of many marine species (Hewitt, 1996). During the Pleistocene glacial maxima, the Bering Sea was mostly covered by ice, forcing species to move south into refugia (Hewitt, 1996). At the end of this cooler period, populations that were previously isolated were in contact again. The Bering area was proposed as a refuge for several marine species like the capelin (Dodson *et al.*, 2007) and polychaete (Hardy *et al.*, 2011). Other studies have proposed the presence of more southern refugia along the Alaskan coast (Coyer *et al.*, 2011). The presence of

private haplotype we observed in Prince William Sound could be a sign of a refugium. However, as our sample size is limited, we have probably not captured all the genetic diversity present in the entire area. Pursuing sampling effort in these regions will help to better understand the history of the region. In addition, the time since population expansion (50 years) for *Themisto sp.* does not support this hypothesis. Nonetheless, Bayesian skyline reconstructions are sensitive to the mutation rate chosen (Grant *et al.*, 2012). In the absence of fossils, we used a general molecular estimate proposed by Knowlton and Weigt (1998) for crustaceans. Using a more appropriate molecular clock will undoubtedly refine this time estimate.

In marine environments, currents can promote or can impede genetic exchanges among populations (Mitarai *et al.*, 2009; White *et al.*, 2010). As currents flow northward from the Pacific to the Arctic, we expected more admixture from the Prince William Sound into the Bering area but our results suggest extremely limited exchanges between these areas. In the Pacific cod *Gadus microcephalus*, ecological differences among different spawning groups spread from either side of the Aleutian Archipelago have been documented, highlighting the influence of physical factors on the structure of marine populations in this area (Spies, 2012). In addition, future work should characterize the northern distribution limit of *Themisto sp.* to determine if the Chukchi shelf might act as a barrier that limits the entry of Bering individuals into the Arctic, as observed in the copepod *Calanus glacialis* (Nelson *et al.*, 2009). No differences were found between Bering and Chukchi *Themisto*, but this might be due to our limited sampling. Better understanding the spatial structure of this species will help us to better understand mechanisms that lead to speciation in marine environment.

Ecological preferences are known to be a source of divergence between marine population (Goetze, 2005; Peijnenburg and Goetze, 2005). We have identified 105

SNPs under selection without any information on their genomic location. De Wit and Palumbi (2013) have identified in the red abalone population several genes containing outliers SNPs. All of these genes were involved in the calcification process, which is perturbed with oceanic acidification due to climatic changes. Thus, identifying the gene of the selected SNP we detected in *Themisto sp.* will help to understand how this species will respond to the future oceans perturbations. *Themisto pacifica* is used as an indicator of subarctic waters (Lavaniegos, 2014) and *T. libellula* for tracking arctic waters (Pinchuck *et al.*, 2013). We do not know anything about the ecology of *Themisto sp.*, but it will be useful to describe it in order to see to which water mass it is associated.

### 5.6.3 Limit of the techniques

Our DAPC and STRUCTURE analyses were congruent with mitochondrial analyses even if we have not tested the impact of various parameter values in SNP calling. Our results, based on 33 individuals from four sites, certainly did not capture all the diversity present in this area. It is highly probable that our  $F_{ST}$  estimates and effective population size estimates were not accurate due to our small sample size. Grant (2015) has demonstrated that Bayesian skyline plot often fails to recover when expansion has occurred with less than 50 individuals. Thus, adding more individuals would help to have a better picture of the genetic structure in the Alaskan waters of this cryptic species.

## 5.7 Conclusion

Widespread genetic exchanges are often assumed in holozooplanktonic species (Palumbi, 1994). The various analyses we carried out on mitochondrial and nuclear datasets indicated a strong spatial structuration of *Themisto sp.* despite our limited sample size. The high differentiation between Prince William Sound and the Bering

region suggested very restricted exchanges despite a continuous northward current. To gain insight into dispersal patterns of holozooplanktonic species, it will be useful in the future to include ecological factors and temporal sampling information. Using genomic techniques to improve our knowledge on marine arctic populations will benefit to maintain and monitor Arctic marine biodiversity.

### 5.8 Acknowledgement

We thank Pamela Pharand for her help in the laboratory. This study was supported by Canadian Healthy Ocean Network. We acknowledge the Centre de Bioinformatique, Université Laval and Dr. Jérôme Laroche for assistance in SNP identification. We thank Dr. Kaven Dionne for advice during this project and for providing constructive criticisms of earlier drafts of this paper.

## CHAPITRE VI

### CONCLUSION GÉNÉRALE

L'objectif général de cette thèse était d'étudier la connectivité des amphipodes marins de l'Arctique à différentes échelles spatiales et temporelles. Les résultats obtenus dans cette thèse montrent qu'en milieu marin, les échanges sont plus limités et complexes que ce qui était supposé traditionnellement. Nous avons clairement démontré que les patrons de connectivité chez les amphipodes marins arctiques étaient limités à une grande échelle géographique mais aussi à une échelle spatiale plus petite même chez des espèces pélagiques. Ces observations sont issues de l'analyse de la connectivité entre les trois océans canadiens (Chapitre II), à l'intérieur d'un genre strictement pélagique (Chapitre III) et à partir de l'étude de la structuration des populations de trois espèces d'amphipodes pélagiques (Chapitre IV et Chapitre V) à l'aide de marqueurs moléculaires évoluant à des vitesses différentes (Figure 6.1).





**Figure 6.1** Diagramme récapitulatif de la thèse. OTU= operational taxonomic unit, Mt DNA= ADN mitochondrial, RAD-seq= Restriction site associated DNA sequencing.

## 6.1 Connectivité à l'échelle tri-océanique (Chapitre II)

Le deuxième chapitre de ma thèse s'est focalisé sur la connectivité entre les trois océans canadiens chez les amphipodes marins. En particulier, nous nous sommes intéressées à l'influence des océans limitrophes de l'océan Arctique sur sa biodiversité. Comme le nombre d'espèce d'amphipodes est inconnu dans ces régions, nous avons utilisé les codes-barres ADN du cytochrome oxydase I (COI) provenant de l'échantillonnage et de la base de données BOLD pour définir le nombre d'espèces potentielles sous forme d'OTU (operationnal taxonomic unit) dans chaque océan. Nous avons d'abord délimité un seuil pour définir les OTUs, puis afin d'évaluer la connectivité entre chaque océan, nous avons recherché la présence d'OTUs partagées.

### 6.1.1 Échange tri-océanique

En comparant la composition en espèce des trois océans Arctique, Atlantique et Pacifique, j'ai mis en évidence une proximité entre les océans Atlantique et Arctique (26 OTUs en commun). Mon étude a aussi mis en évidence que le Pacifique semblait être plus isolé de l'Arctique car aucune espèce partagée entre les deux océans a été observé. Ces résultats suggèrent que la connectivité entre les trois océans n'est pas égale, mais semble se faire préférentiellement entre l'océan Atlantique et Arctique. Un résultat similaire avait aussi été observé chez les polychètes (Carr *et al.*, 2011; Budeava et Rogacheva, 2013). L'isolement que l'on observe entre l'océan Pacifique et Arctique bien qu'il y ait un apport continu d'eau du Pacifique en Arctique important pourrait refléter une adaptation de ces taxons aux conditions environnementales du Pacifique. Toutefois, les données de code-barres des amphipodes du Pacifique disponibles dans BOLD étaient plus limitées que celles des océans Arctique et Atlantique. Cependant, en normalisant les données la même tendance a été retrouvée (données non présentées); mais en augmentant l'effort de caractérisation de la biodiversité des amphipodes de cet océan ce résultat pourrait être confirmé.

La proximité entre les faunes marines arctiques et atlantiques que nous avons observée démontre que l'Atlantique joue un rôle important sur la biodiversité marine de l'Arctique. Mes résultats contrastent avec la plupart des études sur les échanges transarctiques (revues dans Vermeij, 1991) où une proximité entre les taxons du Pacifique et de l'Arctique semble être plus souvent observée. C'est le cas chez les algues brunes *Fucus distichus* L. (Coyer *et al.*, 2011), les oursins (Addison et Hart, 2004; 2005), les moules (Nikula *et al.*, 2003, Väinölä *et al.*, 2008) ou chez certains poissons (Dodson *et al.*, 2007; Pålsson *et al.*, 2009). Plusieurs hypothèses comme une différence initiale dans le nombre des espèces entre les biotes, le patrons de circulations des courants, ou les conditions des environnements receveurs (revues dans Vermeij, 1991) ont été avancées pour expliquer cette asymétrie en faveur du Pacifique dans la majorité des échanges transarctiques, mais il semble que l'opportunité écologique soit la principale hypothèse qui permettrait de l'expliquer (Cunningham et Collins, 1998). Notamment, l'océan Atlantique semble avoir expérimenté de plus grandes perturbations lors des périodes de glaciation survenues entre le Pliocène et le Pléistocène conduisant à un plus grand appauvrissement en espèces (Cunningham et Collins, 1998; Vermeij, 1991). Pour certains auteurs, le Nord-Est Atlantique apparaît aussi comme une zone ayant joué un rôle important dans la recolonisation du Nord-Ouest Atlantique mais aussi de l'Arctique (Cunningham et Collins, 1998; Ingólfsson, 2009). Notre étude s'est limitée aux trois océans canadiens. Il serait intéressant de reconduire la même étude en incluant aussi le Nord-est Atlantique afin de vérifier l'influence de ce dernier sur la biodiversité arctique et de vérifier si comme certains auteurs l'avancent, le Nord-est Atlantique serait à l'origine de la biodiversité du Nord-ouest Atlantique.

Durant les dernières glaciations, il est probable que le littoral ait été plus fortement perturbé que le milieu pélagique favorisant ainsi la recolonisation des milieux littoraux à partir de d'autres endroits (Cunningham et Collins, 1998). Il est donc

probable que les espèces littorales aient été plus affectées par la couverture de glace que les espèces pélagiques. En comparant les patrons d'échanges transarctiques en fonction du milieu occupé (i.e. littoral versus pélagique), il serait possible de mettre en évidence comment les échanges transarctiques et autres processus historiques ont influencé la diversité marine actuelle. De plus, il serait pertinent de faire aussi la distinction entre les espèces benthiques et pélagiques afin de caractériser une différence dans les échanges en fonction des courants.

#### 6.1.2 Seuil pour définir une espèce

Mon étude a confirmé l'utilité du code-barres ADN basé sur l'utilisation du gène COI dans la délimitation des espèces d'amphipodes. Notamment, grâce à cette technique, j'ai mis en évidence la présence de nombreuses espèces cryptiques au niveau des trois océans canadiens. Ces résultats confirment que la biodiversité marine a largement été sous-estimée, comme d'autres études l'ont suggéré chez les amphipodes pélagiques (Lörz *et al.*, 2012; Havermans *et al.*, 2013); les polychètes (Carr *et al.*, 2011), les copépodes (Goetze, 2005), les mollusques (Layton *et al.*, 2014), et de nombreuses autres espèces marines (Radulovici *et al.*, 2010; Raupach et Radulovici 2015). Néanmoins, cette technique a soulevé de nombreuses critiques notamment sur le seuil à utiliser pour déterminer les espèces potentielles (revues dans Collins et Cruickshank, 2013). Traditionnellement un seuil de 3% de divergence est utilisé pour discriminer les espèces (Hebert *et al.*, 2003a). Chez les crustacés un seuil plus élevé (16%) a été proposé par Lefébure *et al.*, (2006). Plusieurs méthodes ont par ailleurs été développées et améliorées pour mieux délimiter les espèces en se basant sur une seule donnée moléculaire (Pons *et al.*, 2006; Lefébure *et al.*, 2006; Puillandre *et al.*, 2012). En utilisant deux de ces méthodes, j'ai obtenu des résultats qui contrastaient avec ceux déjà connus chez les amphipodes. Premièrement, j'ai montré qu'un seuil de divergence de 7% au niveau du COI, plutôt qu'un seuil de 3%, réussit mieux à

discriminer la plupart des espèces d'amphipodes sauf les Gammaridae où le seuil de 16% reste le plus approprié. Deuxièmement, en explorant la distribution des distances inter et intra-spécifiques dans une dizaine de familles d'amphipodes, j'ai démontré que le seuil de 16% avancé par Lefébure (*et al.*, 2006) sous-estimait le nombre d'espèces d'amphipodes. Une autre étude (Trontelj *et al.*, 2009) avait aussi constaté que ce seuil n'était pas adapté pour délimiter les espèces de Niphargidae (famille de l'ordre Amphipoda). Il apparaît que les différences d'évolution au sein des familles pourraient faire varier la valeur du seuil de délimitation approprié pour discriminer les espèces (Collins et Cruickshank, 2013), ce qui veut dire qu'utiliser un seuil unique conduit à sous-estimer ou surestimer le nombre d'espèces. Il est donc important d'évaluer le seuil propre pour chaque espèce qui permettra de discriminer au mieux celles-ci plutôt que d'utiliser un seuil général.

## 6.2 Phylogénie (Chapitre III)

Le troisième chapitre de ma thèse s'est focalisé sur la connectivité au sein d'un genre d'amphipode strictement pélagique: *Themisto*. Les *Themisto* sont des espèces holozooplanktoniques qui passe leur vie entière dans la colonne d'eau leur conférant un potentiel de dispersion en théorie très grand. Pourtant la présence de différentes espèces de *Themisto* indiquent la présence de barrière qui a conduit à leur spéciation. Afin de caractériser les liens phylogénétiques entre les différentes espèces retrouvées dans tous les océans de l'hémisphère Nord, une phylogénie des espèces a été reconstruite en combinant des données mitochondriales et nucléaires. J'ai démontré par cette approche la présence d'une espèce cryptique au niveau de la Béringie. J'ai aussi proposé une phylogénie mises à jour du genre *Themisto* dont la dernière phylogénie moléculaire datait de 1986 (Schneppenheim et Wagman-Hass, 1986). Ensuite j'ai résolu le positionnement de *Themisto* au sein de la famille Hyperiididae (Hurt *et al.*, 2013).

### 6.2.1 Connectivité limitée chez les espèces pélagiques

En explorant les liens phylogénétiques des *Themisto*, j'ai mis en évidence l'existence d'une espèce cryptique, appelée provisoirement *Themisto sp.*, qui était confondue avec *T. libellula* ou *T. pacifica*. Ma phylogénie indique que *Themisto sp.* est plus proche des espèces de l'Arctique (*T. libellula*) et de l'Atlantique (*T. abyssorum*/*T. compressa*) que de celles du Pacifique (*T. japonica*/*T. pacifica*). Ce résultat confirme le fort isolement entre le Pacifique et l'Arctique que j'avais caractérisé au chapitre II, ainsi qu'avec ceux obtenus par Darling (*et al.*, 2007) chez les foraminifères. Même si le Pacifique et l'Arctique sont géographiquement proches, il semble que leurs échanges soient limités. Darling *et al.*, (2007) ainsi que d'autres études (Goetze, 2005; Sobel *et al.*, 2010) suggèrent que les préférences écologiques seraient principalement responsables de la spéciation en milieu marin. Chez les *Themisto*, les préférences écologiques de chacune des espèces ne sont pas connues. Néanmoins, chez certaines espèces *T. abyssorum* et *T. libellula* ou *T. gaudichaudii* et *T. japonica*, des préférences écologiques dans leur diètes et leur tolérance à la température ont été décrites (Auel *et al.*, 2002; Dalpadado *et al.*, 2008; Ikeda, 2009). Par exemple, même si les *Themisto* sont principalement carnivores, il semble que *T. abyssorum* soit plus détritivore (Auel *et al.*, 2002). Afin de vérifier si les préférences écologiques sont la principale cause de leur divergence, il serait intéressant de voir si chaque espèce de *Themisto* est associée préférentiellement à certains paramètres environnementaux.

Une autre hypothèse qui pourrait influencer la spéciation des *Themisto* est le comportement parasitaire. En plus d'avoir un mode de vie entièrement pélagique, une autre caractéristique des *Themisto* ainsi que de tous les autres représentants du sous-ordre Hyperiididae est leur comportement parasitaire envers les organismes zooplanctoniques gélatineux (Laval, 1980; Gasca *et al.*, 2004). Il semble qu'il existe

des associations préférentielles entre certains cnidaires et les Hyperiidæ (Laval, 1980) et que les Hyperiidæ ont développé des appendices spécifiques de l'hôte qu'ils parasitent (Browne *et al.*, 2007). Cependant, ces associations sont souvent basées sur des observations *in situ* et il est probable que toutes les associations ne soient pas connues. Il serait intéressant de voir si une coévolution entre les cnidaires et les Hyperiidæ existe, en comparant par exemple les arbres phylogénétiques de chacun. Actuellement, il existe encore des zones d'ombre dans l'arbre phylogénétique des cnidaires (Collins, 2009), mais il est probable qu'en comparant certaines branches, des associations pourraient être vérifiées. Par exemple, Legendre *et al.*, (2002) ont développé un test simple qui permet de vérifier si une association hôte-parasite a plus de chances de survenir par hasard en comparant les phylogénies des hôtes avec celles des parasites.

En estimant le taux de divergence entre les *Themisto*, j'ai trouvé la plus faible valeur de divergence entre une espèce de l'hémisphère nord (*T. compressa*) et une de l'hémisphère sud (*T. gaudichaudii*). de Vargas (*et al.*, 2000) ont montré que la distribution le long de l'Atlantique des génotypes du foraminifère *Orbulina* correspondait à des zones hydrographiques précises, suggérant que la distribution des espèces est surtout liée à l'adaptation écologique. Darling *et al.*,(2000) ont démontré que certains foraminifères arctiques et antarctiques étaient génétiquement proches, suggérant d'importants échanges entre les deux pôles. Plus récemment, Pawlowsky (*et al.*, 2007) ont démontré que très peu de différences génétiques existaient chez des foraminifères benthiques entre l'Arctique et l'Antarctique. Dernièrement, Havermans (*et al.*, 2013) ont caractérisé une faible divergence entre l'Arctique et l'Atlantique chez un amphipode *Eurythenes gryllus*. Dans mon étude, je n'ai pas eu accès aux spécimens de *Themisto* de l'hémisphère Sud. Il aurait été néanmoins intéressant de vérifier tout d'abord la présence d'espèces cryptiques de *Themisto* dans l'hémisphère Sud et ensuite d'estimer la divergence entre l'ensemble de ces espèces.

### 6.2.2 Limites de la phylogénie Hyperiidae

Ma phylogénie de la famille des Hyperiidae n'a pas permis d'éclaircir les relations entre certains genres comme *Hyperia* et *Hyperoche*. Récemment, Zeidler (2015) a proposé une révision taxonomique du genre *Hyperia* pour inclure les nouvelles espèces identifiées au niveau du Golfe du Mexique. Il serait intéressant de mettre à jour la phylogénie de la famille à la lumière de ces nouvelles descriptions. De plus, au sein de la famille Hyperiidae, plusieurs genres étaient représentés dans la présente étude par un seul spécimen, comme *Iulopis* ou *Hyperiella*. Augmenter l'effort d'échantillonnage en milieu pélagique permettrait de construire une meilleure phylogénie de cette famille. Notamment, il serait intéressant de voir les associations entre les liens phylogénétiques entre ces espèces et les conditions environnementales auxquelles elles sont associées. Beaucoup d'Hyperiidae sont associés à des masses d'eaux tropicales ou subtropicales alors que les *Themisto* en particulier semblent être plus restreints à des masses d'eau froide (revu par Lavenigos, 2014). Souvent les *Themisto* sont utilisés comme indicateurs de masses d'eau arctiques ou subarctiques (Pinchuck *et al.*, 2013; Lavenigos, 2014). Avant d'utiliser les espèces comme indicateur écologique, une meilleure caractérisation de la diversité spécifique serait nécessaire ainsi que leur écologie afin d'éviter une mauvaise interprétation de leur présence avec la présence d'une masse d'eau particulière.

### 6.3 Connectivité à l'échelle des populations (Chapitre IV et V)

Les quatrième et cinquième chapitres de ma thèse se sont intéressés à la connectivité entre les populations chez trois espèces pélagiques (*Themisto*) au niveau de l'archipel canadien et de la Béringie. La divergence entre les populations est la première étape de la spéciation. Celle-ci peut être due à une combinaison de facteurs historiques et environnementaux. En premier lieu j'ai caractérisé leur diversité génétique ainsi que la présence éventuelle de populations en combinant des données mitochondriales et du polymorphisme nucléotidiques (SNP). Dans un deuxième temps, j'ai évalué les



échanges entre les populations en estimant des indices de différenciations ( $\phi_{ST}$ ). J'ai montré la présence d'une forte diversité génétique chez *T. abyssorum* ainsi que la présence de groupes génétiques distincts chez chacune des trois espèces de *Themisto* au niveau de l'archipel canadien et de la Béringie. C'est la première fois qu'une étude de génétique des populations est entreprise chez les *Themisto*. C'est aussi la première étude à utiliser les Rad-tags en Arctique.

### 6.3.1 Refuge en profondeur versus zone contact

En se focalisant à une échelle locale la Béringie ou l'Archipel canadien, j'ai mis au jour des patrons phylogéographiques contrastés entre trois espèces d'amphipodes du genre *Themisto*. Tout d'abord, l'analyse mitochondriale de chacune des espèces a montré une très forte diversité chez *T. abyssorum* par rapport aux deux autres espèces de *Themisto*. Il est communément admis que la diversité génétique diminue vers les latitudes élevées (Hewitt, 2004; Hellberg *et al.*, 2002; 2007), notamment en raison des glaciations du Pléistocène lorsque l'Arctique étaient couvert de glace. Cet épisode a forcé la migration plus au Sud des espèces de l'Arctique à la recherche d'habitats plus favorables (Hewitt, 1996; 2004). Lors de la fonte de la glace, une recolonisation des espaces s'en est suivie par quelques groupes d'individus, conduisant à une plus faible diversité génétique en Arctique en raison des effets fondateurs (Hewitt, 1996). Cependant, plusieurs études (revues dans Hardy *et al.*, 2011) ont démontré la présence de zones libres de glace en Arctique pendant le Pléistocène, fournissant un refuge aux populations. Ces zones refuges sont caractérisées par la présence d'une forte diversité génétique et notamment par la présence d'allèles privés (Slatkin, 1995; Maggs *et al.*, 2008). J'ai avancé que la diversité observée chez *T. abyssorum* pourrait être expliquée par la présence d'un refuge en profondeur, ce qui est concordant avec son habitat actuel (Bowman, 1960; Vinonagrov *et al.*, 1996; Dalapadado, 2002). Il est probable que lors des dernières glaciations, les populations des espèces vivant en

profondeur ont subi moins de perturbations que celles vivants en surface, ce qui leur a permis de conserver une plus grande diversité génétique (Allcock et Strungell, 2012). Dans l'archipel canadien, une seule étude génétique réalisée sur les populations de polychètes *Harmothoe* a avancé l'hypothèse d'une zone de contact secondaire (Hardy *et al.*, 2011). Les zones de contact secondaires se caractérisent aussi par une forte diversité génétique car elles sont le résultat de la rencontre entre différentes populations qui ont été isolées pendant un temps. Les refuges se caractérisent par une absence de lien géographique ente les haplotypes tandis que les zones de contact secondaire se caractérisent par des groupes d'haplotypes séparés par plusieurs mutations (Maggs *et al.*, 2008). Mes résultats ne semblent pas en accord avec ceux trouvés par Hardy *et al.*, (2011) pour les polychètes et plusieurs explications peuvent être avancées. Tout d'abord, la bathymétrie de l'Archipel canadien n'est pas encore bien connue et il est probable que certaines zones profondes restent encore à être caractérisées (Archambault *et al.*, 2010). L'étude approfondie de ces zones permettrait de voir si des données géologiques ou de sédimentation particulières corroboreraient ou infirmeraient la présence de zone libre de glace en profondeur. Il est aussi probable que les différences observées entre mes résultats et ceux d'Hardy *et al.*, (2011) soient dues à mon échantillonnage. En effet, pour déterminer la présence d'une zone de contact secondaire, toute la zone de distribution de l'espèce doit être échantillonnée afin de voir si la plus forte diversité observée est issue d'un mélange d'haplotypes provenant de zones géographiques distinctes. En étudiant la diversité génétique des *T. abyssorum* provenant de la partie européenne de l'Arctique, je pourrais confirmer ou infirmer la présence d'une zone de refuge en profondeur ou d'une zone de contact secondaire au niveau de l'archipel canadien.

Contrairement à la forte diversité mitochondriale observée chez *T. abyssorum*, j'ai caractérisé une plus faible diversité chez *T. libellula* au même endroit. Cette différence dans les patrons de diversité entre des espèces sœurs retrouvées au niveau

de l'archipel canadien pourrait être due à un plus fort impact d'une perturbation chez *T. libellula*. Contrairement à *T. abyssorum*, *T. libellula* est une espèce qui est retrouvée à de plus faibles profondeurs (Bowman, 1960). Je pourrais supposer que cette plus faible diversité génétique soit due à une diminution passée de la population. Cette diminution pourrait être due aux dernières glaciations, cependant l'estimation temporelle (50 ans) ne correspond pas à cette période. Néanmoins, plusieurs facteurs peuvent aussi affecter mon étude. Tout d'abord, une horloge moléculaire inappropriée peut conduire à une mauvaise estimation du temps depuis l'expansion. Ensuite en se focalisant seulement sur l'archipel canadien, je n'ai pas l'image complète de la diversité de cette espèce, et il est probable qu'intégrer des individus provenant de la partie européenne de l'Arctique m'aiderait à obtenir une meilleure estimation des changements démographiques expérimentés par les *Themisto* de l'Arctique. En incluant des échantillons du détroit de Béring, je pensais pouvoir vérifier si le détroit de Béring a été une zone de refuge ou de contact secondaire pour *T. libellula*. Cependant, grâce aux identifications moléculaires, j'ai mis en évidence la présence à cet endroit d'une espèce cryptique que j'ai étudié par la suite. Toutefois, la présence de cette espèce cryptique ne me renseigne pas sur la présence d'une zone refuge au niveau du détroit de Béring pour *T. libellula*.

### 6.3.2 Les Rad-tags

Grâce à l'analyse des Rad-tags réalisée sur chacun des *Themisto* de l'Archipel canadien, deux sous-groupes génétiques ont été mis en évidence. Cette technique a aussi permis de mettre en évidence la présence de sous-groupes chez des invertébrés comme le homard américain (Benestan *et al.*, 2015). Mes résultats contrastent avec la seule étude connue à ce jour sur un invertébré pélagique, *Euphausia superba*, où aucune subdivision n'a été détectée après l'analyse de plusieurs milliers de SNPs, malgré une forte diversité du génome mitochondriale (Deagle *et al.*, 2015). La

présence des deux sous-groupes génétiques distincts chez *T. abyssorum* et *T. libellula* indique qu'à une échelle locale, les échanges sont limités chez ces espèces pélagiques. Chez *T. libellula*, il semble que les deux sous-groupes correspondent à la subdivision écologique connue au niveau de l'archipel canadien (Ardayna *et al.*, 2007). Néanmoins, l'absence de données écologiques associées aux génotypes ne me permet pas de vérifier cette hypothèse. Pour *T. abyssorum*, aucune hypothèse écologique claire n'a pu être avancée car les deux sous-groupes se retrouvaient tout au long de l'archipel canadien. Cependant, comme les données mitochondriales suggéraient la présence d'un refuge en profondeur, il serait envisageable que les sous-groupes identifiés par les rad-tags correspondent à une stratification verticale dans la colonne d'eau. Chez d'autres amphipodes, et notamment ceux associés aux profondeurs, il semble que la diversité génétique se stratifie avec la profondeur (Lörz *et al.*, 2012; Havermans *et al.*, 2013), mais encore peu d'études en tiennent compte. Récemment, Weiner *et al.*, (2012) ont caractérisé une répartition verticale de trois génotypes de foraminifères *Hastigerina pelagica*, chaque génotype étant associé à une profondeur particulière. Cette dernière étude a montré que sur la longueur d'une région, la diversité génétique ne changeait pas. Toutefois, lorsque la profondeur de chaque échantillon était prise en compte, cette structuration devenait visible. Il serait donc intéressant d'explorer cette possibilité chez *T. abyssorum*, en associant des données de profondeurs de chaque individu à leur génotype.

Avec un nombre d'individus relativement petit, j'ai observé des patrons phylogéographiques qui concordaient entre les données génétiques mitochondriales et nucléaires chez *Themisto sp.* Mon analyse basée sur les données du cytochrome oxydase I et plus de 4800 marqueurs neutres nucléaires a révélé la présence de deux populations distinctes une correspondant au Pacifique Nord et une spécifique à la Béringie. Un résultat similaire a aussi été trouvé chez le flétan du Pacifique avec la population du golfe de l'Alaska arborant une très grande diversité par rapport aux

populations des îles aléoutiennes et de la Béringie (Nielsen *et al.*, 2010). Contrairement au flétan du Pacifique, j'ai déterminé des niveaux de différenciation relativement forts entre le golfe de l'Alaska et la Béringie ( $\phi_{ST}=0.26$ ). Ce fort niveau de différenciation suggère des échanges génétiques très limités entre ces deux régions chez *Themisto sp.*, ce qui vient encore appuyer mes résultats obtenus au chapitre II et III. Cet isolement peut être dû à la présence d'une zone refuge au niveau du golfe de l'Alaska qui aurait permis la recolonisation de la Béringie lorsque l'environnement a été plus favorable aux *Themisto sp.* Néanmoins, pour s'en assurer, il faudrait augmenter le nombre d'individus dans cette analyse.

Les sous-groupes génétiquement distinguables chez les *Themisto* de l'Archipel canadien ne sont pas complètement isolés. En effet, leurs coefficients de différenciation ( $\phi_{ST}$ ) sont faibles, autour de 0,02. Mes résultats correspondent tout à fait aux coefficients de différenciation chez les espèces marines pélagiques où ces faibles valeurs sont communes (Knusten *et al.*, 2011). Des valeurs de  $F_{ST}$  similaires ont été trouvées entre l'épinoche d'eau douce et océanique ( $\phi_{ST}=0.02$ , Hohenhole *et al.*, 2010) et entre les populations de lamproie du Pacifique ( $\phi_{CT}=0.02$ , Hess *et al.*, 2013). Néanmoins, mes valeurs sont plus élevées que celles rapportées chez d'autres crustacés marins, comme le homard américain ( $\phi_{CT}=0,0011$ , Benestan *et al.*, 2015), le mysid *Euphausia superba* (panmixie, Deagle *et al.*, 2015) ou le copépode *Centropages typicus* ( $\phi_{CT}=0,03$ ; Blanco-Bercial et Bucklin, 2016). Mon estimation du  $\phi_{ST}$  entre les populations de *Themisto sp.* au niveau de la Béringie est la plus forte ( $\phi_{ST}=0.28$ ) jamais reportée avec les rad-tags chez des organismes marins. Même si le faible nombre d'individu utilisé va influencer cet estimé, l'ensemble de ces résultats suggèrent que la connectivité entre les populations d'organismes marins, notamment des espèces holozooplanktoniques comme les copépodes, les mysides ou les amphipodes, va être influencée par de nombreux facteurs.

Dans ce projet de recherche doctoral, j'ai documenté la structuration des populations des *Themisto*, mais sur une seule campagne d'échantillonnage. Il serait intéressant aussi d'explorer la variation interannuelle et intra-annuelle sur la structure génétique des populations de *Themisto* en réalisant des échantillonnages à deux saisons différentes sur plusieurs années. Par exemple des évènements particuliers comme des tempêtes qui ont décimé une partie des populations de *Themisto* en Norvège (Eine et Daase, 2002) ont probablement influencé la diversité génétique de leur population. Chez les *Themisto*, il semble que les cycles de reproduction varient en fonction de la latitudes des individus (Dalpadado, 2002, Dalpadado *et al.*, 2008). Cependant il faudrait étudier les cycles de reproduction de chacune des espèces plus en détail pour mieux connaître la biologie de ces espèces et notamment leur temps de génération.

L'analyse sur plus de 800 marqueurs communs aux deux *Themisto* a démontré que *Themisto sp.* est génétiquement distinct de *T. libellula* comme le suggérait notre phylogénie. Il serait intéressant d'identifier les gènes qui sont impliqués dans la divergence entre ces deux espèces. Il serait aussi pertinent, afin de mieux comprendre les mécanismes de spéciation en milieu marin, de vérifier si les mêmes gènes sont impliqués dans la divergence entre les différentes espèces de *Themisto*. Il serait probable que des gènes impliqués dans la régulation de la température ou dans les pressions soient soumis à sélection dans les différentes espèces de *Themisto* puisque chacune semble être associée à des gammes de températures particulières. Il apparaît néanmoins que peu de gènes sont nécessaires à la spéciation et que la plupart d'entre eux n'ont que très peu d'effet sur l'isolement reproducteur (Wu et Ting, 2004). Les gènes impliqués dans l'isolement reproducteur ont été surtout étudiés chez la drosophile et beaucoup moins chez les espèces marines (Noor et Feder, 2006). Cependant, les gènes impliqués dans la différenciation entre les populations sont à la base de la spéciation (Nosil et Feder, 2012). En effet, l'adaptation à un environnement particulier va se refléter sur le génome d'un organisme. Par exemple,

Rocap *et al.*, (2003) ont étudié l'organisation du génome de deux écotypes de cyanobactérie marine *Prochlorococcus*, l'un associé à un environnement peu lumineux en profondeur et l'autre plus lumineux en surface. En comparant leurs génomes, d'importantes différences ont été identifiées, en particulier la présence de gènes uniques à chaque écotpe impliqués dans l'adaptation et la survie à leur environnement. Des gènes impliqués dans la tolérance à la salinité et la température ont montré des traces de sélection entre différentes populations de hareng *Clupea harengus* (Limborg *et al.*, 2012). Plus récemment, chez l'ormeau rouge *Haliotis rufescens*, De Wit et Palumbi (2013) ont montré que, pour des populations de différentes régions de la Californie, des gènes impliqués dans la tolérance aux stress environnementaux étaient soumis à différentes pressions sélectives le long des côtes californiennes. Chez le genre *Themisto*, j'ai mis en évidence plusieurs SNPs soumis à sélection, mais nous n'avons pas identifié les gènes concernés. Il serait intéressant d'identifier dans quelle région ou dans quel gène ces SNP sont trouvés entre les populations afin de mieux comprendre l'architecture de la divergence en milieu marin.

### 6.3.3 Limites des SNPs et de la génomique

Pour terminer, mon étude sur la connectivité locale est basée sur un faible nombre d'individus, entre 5 et 40 par station. Il est à noter que lorsque la différenciation est faible ( $F_{ST} < 0.01$ ), il serait préférable d'avoir plus de 100 individus pour pouvoir détecter avec confiance des structures génétiques subtiles (Benestan *et al.*, 2016). Cependant, aucune étude génétique n'avait encore été réalisée chez ces espèces, mon étude a donc permis d'apporter les premières informations au sujet de l'organisation de la diversité génétique en Arctique.

De plus la technique de Rad-sequencing que j'ai utilisée pour découvrir les SNPs comporte certains désavantages. Les régions dupliquées représentent un biais qui est

encore assez mal quantifié dans la plupart des études qui utilisent des Rad-tags. Dernièrement quelques études ce sont intéressées à les quantifier. Deagle *et al.*, (2015) ont remarqué que les marqueurs qui bénéficiaient d'une couverture suffisante étaient dûs à la présence de régions dupliquées qui sont relativement importantes dans le génome d'*Euphausia superba* (10Gb). Les méthodes utilisées pour appeler les SNPs peuvent être aussi source de biais. Pan *et al.*, (2015) ont comparé trois logiciels couramment utilisés dans la sélection des Rad-tag (Stacks, UNEAK et GATK). Il apparait que chacun comporte ses forces et ses faiblesses notamment face aux problèmes de régions dupliquées. C'est seulement en combinant plusieurs approches que les « vrais » SNPs pourront être mieux caractérisés. Dernièrement, Rodríguez-Ezpeleta *et al.*, (2016) ont démontré que le choix des paramètres pour définir et construire le catalogue de SNP dans Stacks influençait le nombre de SNPs retenus mais aussi la structure des populations. Dans mon étude, je n'ai pas comparé l'influence des différentes méthodes d'identification des SNPs. Afin de vérifier nos résultats, il serait intéressant d'utiliser ces différentes approches.

#### 6.4 Perspectives

Mon étude a permis de démontrer que la connectivité des amphipodes marins en Arctique peut être limitée alors que l'on s'attendait à ce qu'elle soit moins restreinte. Tout d'abord, les échanges entre les populations des océans Arctique et Pacifique semblaient très restreints. Toutefois, comme je l'ai démontré clairement dans les chapitres II et III, les échanges étaient beaucoup plus forts entre les populations des océans Arctique et Atlantique. Ensuite, à une échelle spatiale plus petite, que ce soit au niveau de l'archipel canadien ou de la Béringie, l'absence de panmixie observée chez des espèces holozooplanktoniques prouve qu'en milieu marin, même pour des espèces qui passent leur vie entière dans la colonne d'eau, les échanges entre populations peuvent être limités, comme je l'ai montré dans les chapitres IV et V.



Mon étude a clairement démontré l'importance des outils génétiques dans la caractérisation de la biodiversité, dans la découverte d'espèces cryptiques et dans l'évaluation de la connectivité des populations en milieu marin. Toutefois, même si j'ai rencontré des limitations au niveau de l'échantillonnage, j'ai soulevé d'autres hypothèses qui permettraient de mieux comprendre la biodiversité marine arctique ainsi que son organisation.

#### 6.4.1 Comparaison de la connectivité chez d'autres espèces pélagique en Arctique

Les *Themisto* sont des acteurs importants de la chaîne trophique arctique, mais il existe d'autres espèces holozooplanctoniques qui sont toutes aussi importantes comme les copépodes dont la connaissance sur la connectivité est encore largement incomplète. Goetze (2005) avait déjà démontré que chez certaines espèces de copépodes du genre *Eucalanus*, la connectivité pouvait être limitée à l'échelle globale à cause de leur adaptation à des environnements spécifiques. Dernièrement, Blanco-Bercial et Bucklin (2016) ont démontré la présence d'une structuration des populations dans l'Atlantique Nord chez le copépode *Centropages typicus*. Cependant à ce jour encore aucune étude n'a été faite sur la connectivité des copépodes en Arctique. Les copépodes dominent la faune arctique en termes d'abondance (Bucklin *et al.*, 2010a), et sont la principale proie des *Themisto*. Comparer les patrons de diversité génétiques des copépodes avec ceux de leur prédateur permettrait de voir comment les événements passés, comme la vicariance, les ont affectés (Zink, 2002).

#### 6.4.2 Comparaison de la connectivité entre les deux pôles

Ce projet de doctorat s'est principalement concentré sur l'Arctique, mais afin de mieux comprendre la connectivité en milieu polaire, il serait intéressant d'évaluer la connectivité de l'océan Antarctique. L'océan Antarctique est entouré par les océans

Pacifique, Atlantique et Indien. Il serait intéressant de voir comment les échanges autour de cet océan se font, notamment en regardant l'influence des océans Pacifique et Atlantique mais aussi de l'océan Indien. En particulier, en fonction des courants, il est probable que les échanges soient orientés. L'océan Antarctique a aussi expérimenté les glaciations du Pléistocène. La présence de zones libres de glace dans cet océan du sud offrant des refuges aux espèces marines antarctiques a aussi été documenté (Allock et Strungell, 2012). Ces refuges auraient permis de recoloniser les mers adjacentes prises dans les glaces durant les périodes interglaciaires. Il serait intéressant de vérifier les liens entre les océans adjacents et l'océan antarctique en comparant la composition en OTUs de chaque océan.

En étudiant la connectivité de l'océan Antarctique, une comparaison pourrait être faite avec l'océan Arctique. En effet les deux océans partagent plusieurs espèces. Un inventaire récent effectué par Census of Marine Life a estimé qu'au moins 235 espèces avaient une distribution bipolaire (Havermans *et al.*, 2013). Quelques études ont étudié la différenciation génétique entre les pôles (Havermans *et al.*, 2013; Norris et de Vargas, 2000; Darling *et al.*, 2000; 2007; Pawlowski *et al.*, 2007). Toutes ces études ont démontré peu de différenciation génétique entre l'Arctique et l'Antarctique suggérant une forte connectivité entre eux. La présence d'un courant Antarctique qui circule à travers le monde pourrait expliquer les forts échanges bipolaires. Il serait intéressant de vérifier si toutes les espèces bipolaires qui présentent peu de différenciation génétique sont associées à ce courant ou si la faible différenciation observée est la conséquence d'une adaptation à un environnement similaire. Comparer les patrons de connectivité entre les océans polaires permettrait de mieux comprendre la dynamique des échanges entre deux océans menacés par les changements climatiques.

## BIBLIOGRAPHIE

- Addison, J. A., et Hart, M. W. (2005). Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins (*Strongylochnitrotus droebachiensis*). *Evolution*, 59(3), 532–543.
- Agnalt, A. L., Pavlov, V., Jørstad, K. E., Farestveit, E., et Sundet, J. (2011). The snow crab, *Chionoecetes opilio* (Decapoda, Majoidea, Oregoniidae) in the Barents Sea. In *In the Wrong Place-Alien Marine Crustaceans: Distribution, Biology and Impacts*. Springer Netherlands.
- Allcock, A. L., et Strugnell, J. M. (2012). Southern Ocean diversity: new paradigms from molecular ecology. *Trends in Ecology and Evolution*, 27(9), 520–528.
- Anderson, P. J., et Piatt, J. F. (1999). Community reorganization in the Gulf of Alaska following ocean climate regime shift. *Marine Ecology. Progress series*, 189, 117–123.
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., et Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92.
- Angel, M. V. (1993). Biodiversity of the pelagic ocean. *Conservation Biology*, 7(4), 760–772.
- Anisimova, M., et Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55(4), 539–552.
- Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., et al. (2012). The magnitude of global marine species diversity. *Current Biology*, 22(23), 2189–2202.

- Archambault, P., Snelgrove, P. V., Fisher, J. A., Gagnon, J. M., Garbary, D. J., Harvey, M., Kenchington, E. L., Lesage, V., Levesque, M., Lovejoy, C., Mackas, D. L., McKindsey, C. W., Nelson, J. R., Pepin, P., Piché, L., and Pulin, M. (2010). From sea to sea: Canada's three oceans of biodiversity. *PLoS One*, 5(8), e12182.
- Ardyna, M., Gosselin, M., Michel, C., Poulin, M., et Tremblay, J. É. (2011). Environmental forcing of phytoplankton community structure and function in the Canadian High Arctic: contrasting oligotrophic and eutrophic regions. *Marine Ecology Progress Series*, 442, 37–57.
- Arrigo, K. R., van Dijken, G., et Pabi, S. (2008). Impact of a shrinking Arctic ice cover on marine primary production. *Geophysical Research Letters*, 35(19).
- Ashton, G. V., Stevens, M. I., Hart, M. C., Green, D. H., Burrows, M. T., Cook, E. J., et Willis, K. J. (2008). Mitochondrial DNA reveals multiple northern hemisphere introductions of *Caprella mutica* (Crustacea, Amphipoda). *Molecular Ecology*, 17(5), 1293–1303.
- Audzijonyte, A., et Vaecinoelae, R. (2006). Phylogeographic analyses of a circumarctic coastal and a boreal lacustrine mysid crustacean, and evidence of fast postglacial mtDNA rates. *Molecular Ecology*, 15(11), 3287–3301.
- Auel, H., Harjes, M., Da Rocha, R., Stübing, D., et Hagen, W. (2002). Lipid biomarkers indicate different ecological niches and trophic relationships of the Arctic hyperiid amphipods *Themisto abyssorum* and *T. libellula*. *Polar Biology*, 25(5), 374–383.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., et Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489–522.
- Ballard, J. W. O., et Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729–744.

- Beaumont, M. A., et Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1377), 1619–1626.
- Beheregaray, L. B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17(17), 3754–3774.
- Beheregaray, L. B., et Caccone, A. (2007). Cryptic biodiversity in a changing world. *Journal of Biology*, 6(4), 1.
- Benestan *et al.* (2016), Erratum- Benestan *et al.* 2015. *Molecular Ecology*, 25, 1626–1629.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., et Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13), 3299–3315.
- Benzie, J. A. (1999). Genetic structure of coral reef organisms: ghosts of dispersal past. *American Zoologist*, 39(1), 131–145.
- Berge, J., Cottier, F., Last, K. S., Varpe, Ø., Leu, E., Søreide, J., Eiane, K., Falk-Petersen, S., Willis, K., Nygård, N., Vogedes, D., Griffiths, C., Johnsen G., Lorentzen, D. et Brierley, A. S. (2009). Diel vertical migration of Arctic zooplankton during the polar night. *Biology Letters*, 5(1), 69–72.
- Berge, J., Johnsen, G., Nilsen, F., Gulliksen, B., et Slagstad, D. (2005). Ocean temperature oscillations enable reappearance of blue mussels *Mytilus edulis* in Svalbard after a 1000 year absence. *Marine Ecology Progress Series*, 303, 167–175.
- Bernatchez, L., et Wilson, C. C. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, 7(4), 431–452.

- Berry, O., Tocher, M. D., et Sarre, S. D. (2004). Can assignment tests measure dispersal?. *Molecular Ecology*, 13(3), 551–561.
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K. K., et Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22(3), 148–155.
- Blanco-Bercial, L., et Bucklin, A. (2016). New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*. *Molecular Ecology*, 25 (7), 1566–1580.
- Blanco-Bercial, L., Álvarez-Marqués, F., et Bucklin, A. (2011). Comparative phylogeography and connectivity of sibling species of the marine copepod *Clausocalanus* (Calanoida). *Journal of Experimental Marine Biology and Ecology*, 404(1), 108–115.
- Bluhm, B. A., Gebruk, A. V., Gradinger, R., Hopcroft, R. R., Huettmann, F., Kosobokova, K. N., Sirenko, B. I. et Weslawski, J. M. (2011). Arctic marine biodiversity: an update of species richness and examples of biodiversity change. *Oceanography*, 24(3), 232–248.
- Bluhm, B. A., Kosobokova, K. N., et Carmack, E. C. (2015). A tale of two basins: An integrated physical and biological perspective of the deep Arctic Ocean. *Progress in Oceanography*, 139, 89–121.
- Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, 21–45.
- Bohonak, A. J., et Roderick, G. K. (2001). Dispersal of invertebrates among temporary ponds: are genetic estimates accurate?. *Israel Journal of Zoology*, 47(4), 367–386.
- Bolnick, D. I., et Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, 459–487.

- Bouchet, P. (2006). The magnitude of marine biodiversity. In: Duarte CM, editor. *The Exploration of Marine Biodiversity: Scientific and Technological Challenges*. Fundación BBVA, Madrid.
- Bowman, T. E. (1960). *The pelagic amphipod genus ParaThemisto (Hyperiidea: Hyperiidae) in the north Pacific and adjacent Arctic Ocean*. Smithsonian Institution Press.
- Bowman, T. E., et Gruner, H. E. (1973). *The families and genera of Hyperiidea (Crustacea: Amphipoda)*. Smithsonian Institution Press.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. et Buckler E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635.
- Bradbury, I. R., Laurel, B., Snelgrove, P. V., Bentzen, P. et Campana, S. E. (2008). Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1644), 1803–1809.
- Brierley, A. S., et Kingsford, M. J. (2009). Impacts of climate change on marine organisms and ecosystems. *Current Biology*, 19(14), 602–614.
- Broquet, T., et Petit, E. J. (2009). Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40, 193–216.
- Browne, W. E., Haddock, S. H., et Martindale, M. Q. (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*, 47(6), 815–830.
- Brumfield, R. T., Beerli, P., Nickerson, D. A. et Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, 18(5), 249–256.

- Bucklin, A., Hopcroft, R. R., Kosobokova, K. N., Nigro, L. M., Ortman, B. D., Jennings, R. M. et Sweetman, C. J. (2010)a. DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(1), 40–48.
- Bucklin, A., Nishida, S., Schnack-Schiel, S., Wiebe, P. H., Lindsay, D., Machida, R. J. et Copley, N. J. (2010)b. A census of zooplankton of the global ocean. *Life in the World's Oceans: Diversity, Distribution, and Abundance*. edited by: McIntyre, A.
- Bucklin, A., Smolenack, S. B., Bentley, A. M. et Wiebe, P. H. (1997). Gene flow patterns of the euphausiid, *Meganyciophanes norvegica*, in the NW Atlantic based on mtDNA sequences for cytochrome b and cytochrome oxidase I. *Journal of Plankton Research*, 19(11), 1763–1781.
- Bucklin, A., Steinke, D. et Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science*, 3, 471–508.
- Bucklin, A., Wilson, R. R. et Smith, K. L. (1987). Genetic differentiation of seamount and basin populations of the deep-sea amphipod *Eurythenes gryllus*. *Deep Sea Research Part A. Oceanographic Research Papers*, 34(11), 1795–1810.
- Budaeva, N. E. et Rogacheva, A. V. (2013). Colonization of the Arctic Ocean by two cosmopolitan genera of marine invertebrates. *Invertebrate Zoology*, 10(1), 127.
- Byrne, M. (2012). Global change ecotoxicology: identification of early life history bottlenecks in marine invertebrates, variable species responses and variable experimental approaches. *Marine Environmental Research*, 76, 3–15.
- Campbell, V., Legendre, P. et Lapointe, F. J. (2009). Assessing congruence among ultrametric distance matrices. *Journal of classification*, 26(1), 103–117.



- Carmack, E., et Wassmann, P. (2006). Food webs and physical–biological coupling on pan-Arctic shelves: unifying concepts and comprehensive perspectives. *Progress in Oceanography*, 71(2), 446–477.
- Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A. et Hebert, P. D. (2011). A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. *PLoS One*, 6(7), e22232.
- Carstens, B. C., Pelletier, T. A., Reid, N. M. et Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, 22(17), 4369–4383.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552.
- Chan, F. T., Bailey, S. A., Wiley, C. J. et MacIsaac, H. J. (2013). Relative risk assessment for ballast-mediated invasions at Canadian Arctic ports. *Biological Invasions*, 15(2), 295–308.
- Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, 2(4), e64.
- Cheng, F. P., Wang, M. X., Sun, S., Li, C. L. et Zhang, Y. S. (2013). DNA barcoding of Antarctic marine zooplankton for species identification and recognition. *Advances in Polar Science*, 24(2), 119–127.
- Chu, K. H., Tsang, L. M., Ma, K. Y., Chan, T. Y. et Ng, P. K. L. (2009). Decapod phylogeny: what can protein-coding genes tell us. *Decapod Crustacean Phylogenetics*, 89–99.
- Clark, D. L. (1988). Early history of the Arctic Ocean. *Paleoceanography*, 3(5), 539–550.
- Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D. et Lewis, L. C. (2009). Comparative performance of single

- nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, 100(5), 556–564.
- Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G. et Gray, M.R. (1998). Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology*, 46, 419–437.
- Collins, A. G. (2009). Recent insights into cnidarian phylogeny. *Smithsonian Contributions to Marine Sciences*, 38, 139–149.
- Collins, R. A., et Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, 13(6), 969–975.
- Corbett, J. J., Lack, D. A., Winebrake, J. J., Harder, S., Silberman, J. A., et Gold, M. (2010). Arctic shipping emissions inventories and future scenarios. *Atmospheric Chemistry and Physics*, 10(19), 9689–9704.
- Costa, F. O., DeWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., et Hebert, P. D. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(2), 272–295.
- Costa, F. O., Henzler, C. M., Lunt, D. H., Whiteley, N. M., et Rock, J. (2009). Probing marine Gammarus (Amphipoda) taxonomy with DNA barcodes. *Systematics and Biodiversity*, 7(04), 365–379.
- Cowen, R. K., et Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Marine Science*, 1, 443–466.
- Cowen, R. K., Gawarkiewicz, G. G., Pineda, J., Thorrold, S. R. et Werner, F. E. (2007). Population connectivity in marine systems: an overview. *Oceanography*, 20(3), 14–21.
- Cowen, R. K., Paris, C. B. et Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, 311(5760), 522–527.

- Coyer, J. A., Hoarau, G., Van Schaik, J., Luijckx, P. et Olsen, J. L. (2011). Trans-Pacific and trans-Arctic pathways of the intertidal macroalga *Fucus distichus* L. reveal multiple glacial refugia and colonizations from the North Pacific to the North Atlantic. *Journal of Biogeography*, 38(4), 756–771.
- Coyne, J. A., et Orr, H. A. (2004). *Speciation*. Sunderland, MA, Sinauer Associates.
- Cronin, T. M., et Cronin, M. A. (2015). Biological response to climate change in the Arctic Ocean: the view from the past. *Arktos*, 1(1), 1–18.
- Crow, K. D., Munehara, H. et Bernardi, G. (2010). Sympatric speciation in a genus of marine reef fishes. *Molecular Ecology*, 19(10), 2089–2105.
- Cunningham, C. W. et Collins, T. M. (1998). Beyond area relationships: extinction and recolonization in molecular marine biogeography. In *Molecular approaches to ecology and evolution*. Birkhäuser Basel.
- Matzen da Silva, J. M., Creer, S., Dos Santos, A., Costa, A. C., Cunha, M. R., Costa, F. O. et Carvalho, G. R. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS One*, 6(5), e19449.
- Dalpadado, P. (2002). Inter-specific variations in distribution, abundance and possible life-cycle patterns of *Themisto* spp. (Amphipoda) in the Barents Sea. *Polar Biology*, 25(9), 656–666.
- Dalpadado, P., Borkner, N., Bogstad, B., et Mehl, S. (2001). Distribution of *Themisto* spp. (Amphipoda) in the Barents Sea and predator-prey interactions. *ICES Journal of Marine Science: Journal du Conseil*, 58(4), 876–895.
- Dalpadado, P., Yamaguchi, A., Ellertsen, B. et Johannessen, S. (2008). Trophic interactions of macro-zooplankton (krill and amphipods) in the Marginal Ice Zone of the Barents Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 55(20), 2266–2274.

- Darling, K. F., Kucera, M., et Wade, C. M. (2007). Global Molecular Phylogeography Reveals Persistent Arctic Circumpolar Isolation in a Marine Planktonic Protist. *Proceedings of the National Academy of Sciences of the United States of America*, 5002–5007.
- Darling, K. F., Wade, C. M., Stewart, I. A., Kroon, D., Dingle, R. et Brown, A. J. (2000). Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature*, 405(6782), 43–47.
- Darnis, G., Robert, D., Pomerleau, C., Link, H., Archambault, P., Nelson, R. J., Geoffroy, M., Tremblay, J-E., Lovejoy, C., Feguson, S. H., Hunt, B. P. V. et Fortier, L. (2012). Current state and trends in Canadian Arctic marine ecosystems: II. Heterotrophic food web, pelagic-benthic coupling, and biodiversity. *Climatic Change*, 115(1), 179–205.
- Davey, J. W., et Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in functional genomics*, 9(5-6), 416–423.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., et Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510.
- De Bie, T., Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, D., Denys, L., Vanhecke, L., Van der Grucht, K., Van Wichelen, J., Vyverman, W. et Declerck, S. A. J. (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters*, 15(7), 740–747.
- de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W., et Pawlowski, J. (1999). Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences*, 96(6), 2864–2868.

- De Wit, P., et Palumbi, S. R. (2013). Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, 22(11), 2884–2897.
- Deagle, B. E., Faux, C., Kawaguchi, S., Meyer, B., et Jarman, S. N. (2015). Antarctic krill population genomics: apparent panmixia, but genome complexity and large population size muddy the water. *Molecular Ecology*, 24(19), 4943–4959.
- Degnan, J. H., et Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, 24(6), 332–340.
- Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., et Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36 (Web Server issue), 465–469.
- Dionne, K., Vergilino, R., Dufresne, F., Charles, F. et Nozais, C. (2011). No evidence for temporal variation in a cryptic species community of freshwater amphipods of the *Hyaella azteca* species complex. *Diversity*, 3, 390–404.
- Dodson, J. J., Tremblay, S., Colombani, F., Carscadden, J. E., et Lecomte, F. (2007). Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*). *Molecular Ecology*, 16(23), 5030–5043.
- Drummond AJ, Suchard MA, Xie D et Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molecular Biology and Evolution*, 29, 1969–1973.
- Dunbar, M. J. (1946). On *Themisto libellula* in Baffin Island coastal waters. *Journal of the Fisheries Board of Canada*, 6(6), 419–434.
- Dunbar, M. J. (1957). The determinants of production in northern seas: A study of the biology of *Themisto libellula* Mandt. *Canadian Journal of Zoology*, 35(6), 797–819.

- Earl, D. A. et vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- Eiane, K., et Daase, M. (2002). Observations of mass mortality of *Themisto libellula* (Amphipoda, Hyperidae). *Polar Biology*, 25(5), 396–398.
- England, J. H., et Furze, M. F. (2008). New evidence from the western Canadian Arctic Archipelago for the resubmergence of Bering Strait. *Quaternary Research*, 70(1), 60–67.
- England, J. H., Furze, M. F., et Doupé, J. P. (2009). Revision of the NW Laurentide Ice Sheet: implications for paleoclimate, the northeast extremity of Beringia, and Arctic Ocean sedimentation. *Quaternary Science Reviews*, 28(17), 1573–1596.
- Ersts, P.J. (2016). Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Récupéré de [http://biodiversityinformatics.amnh.org/open\\_source/gdmg](http://biodiversityinformatics.amnh.org/open_source/gdmg). Accès le 5 avril 20156.
- Evanno, G., Regnaut, S., et Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–2620.
- Excoffier, L., Laval, G. et Schneider, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47–50.
- Fedorov, V. B., et Stenseth, N. C. (2002). Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1505), 2071–2077.

- Fitzpatrick, B. M., Fordyce, J. A., et Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, 22(11), 2342–2347.
- Foll, M., et Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977–993.
- Folmer, O., M. Black, Hoeh, W., Lutz, R., et Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.
- Folt, C. L., et Burns, C. W. (1999). Biological drivers of zooplankton patchiness. *Trends in Ecology and Evolution*, 14(8), 300–305.
- France, S. C., et Kocher, T. D. (1996). Geographic and bathymetric patterns of mitochondrial 16S rRNA sequence divergence among deep-sea amphipods, *Eurythenes gryllus*. *Marine Biology*, 126(4), 633–643.
- Fu, Y. X. et Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133 (3): 693–709.
- Gagnaire, P. A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., et Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary applications*, 8(8), 769–786.
- Gasca, R., et Haddock, S. H. (2004). Associations between gelatinous zooplankton and hyperiid amphipods (Crustacea: Peracarida) in the Gulf of California. *Hydrobiologia*, 530/531, 529–535.
- Giribet, G., Edgecombe, G. D., et Wheeler, W. C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. *Nature*, 413(6852), 157–161.

- Gladenkov, A. Y., Oleinik, A. E., Marinovich, L., et Barinov, K. B. (2002). A refined age for the earliest opening of Bering Strait. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 183(3), 321–328.
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS One* 9(2): e90346.
- Glover, K. A., Quintela, M., Wennevik, V., Besnier, F., Sørvik, A. G., et Skaala, Ø. (2012). Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. *PLoS One*, 7(8), e43129.
- Goetze, E. (2003). Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1531), 2321–2331.
- Goetze, E. (2005). Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution*, 59(11), 2378–2398.
- Goetze, E. (2010). Species discovery in marine planktonic invertebrates through global molecular screening. *Molecular Ecology*, 19(5), 952–967.
- Goetze, E., et Bradford-Grieve, J. (2005). Genetic and morphological description of *Eucalanus spinifer* T. Scott, 1894 (Calanoida: Eucalanidae), a circumglobal sister species of the copepod *E. hyalinus* ss (Claus, 1866). *Progress in Oceanography*, 65(1), 55–87.
- Goldstein, P. Z., et DeSalle, R. (2011). Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays*, 33(2), 135-147.
- Goodall-Copestake, W. P., Perez-Espona, S., Clark, M. S., Murphy, E. J., Seear, P. J., et Tarling, G. A. (2010). Swarms of diversity at the gene *cox1* in Antarctic krill. *Heredity*, 104(5), 513–518.



- Grant, W. S. (2015). Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity*, esv020.
- Grant, W. S., Liu, M., Gao, T., et Yanagimoto, T. (2012). Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Molecular Phylogenetics and Evolution*, 65(1), 203–212.
- Grebmeier, J. M., Overland, J. E., Moore, S. E., Farley, E. V., Carmack, E. C., Cooper, L. W., Frey, K. E., Helle, J. H., McLaughlin, F. A., et McNutt, S. L. (2006). A major ecosystem shift in the northern Bering Sea. *Science*, 311(5766), 1461–1464.
- Grosberg, R., et Cunningham, C. W. (2001). Genetic structure in the sea. *Marine community ecology*. Sinauer, Sunderland, MA.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., et Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321.
- Hanski, I., et Gaggiotti, O. E. (2004). *Ecology, genetics, and evolution of metapopulations*. Academic Press.
- Hardy, S. M., Carr, C. M., Hardman, M., Steinke, D., Corstorphine, E., et Mah, C. (2011). Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. *Marine Biodiversity*, 41(1), 195–210.
- Harris, S. A. (2005). Thermal history of the Arctic Ocean environs adjacent to North America during the last 3.5 Ma and a possible mechanism for the cause of the cold events (major glaciations and permafrost events). *Progress in Physical Geography*, 29(2), 218–237.

- Havermans C, Sonet G, d'Udekem d'Acoz C, Nagy ZT, Martin P, et al. (2013) Genetic and Morphological Divergences in the Cosmopolitan Deep-Sea Amphipod *Eurythenes gryllus* Reveal a Diverse Abyss and a Bipolar Species. *PLoS One*, 8(9): e74218.
- Heads, M. (2005). Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society*, 84(4), 675–723.
- Hebert, P. D., Ratnasingham, S., et de Waard, J. R. (2003)a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(S1), S96–S99.
- Hebert, P. D., Cywinska, A., et Ball, S. L. (2003)b. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1512), 313–321.
- Hebert, P. D. N., Stoeckle, M. Y., Zemplak, T. S., et Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2(10), e312.
- Hedgecock, D., Barber, P. H., et Edmands, S. (2007). Genetic approaches to measuring connectivity. *Oceanography*, 20(3), 70-79.
- Heled, J., et Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27(3), 570–580.
- Hellberg, M. E. (2009). Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution and Systematics*, 40, 291–310.
- Hellberg, M. E., Balch, D. P., et Roy, K. (2001). Climate-driven range expansion and morphological evolution in a marine gastropod. *Science*, 292(5522), 1707–1710.
- Hellberg, M. E., Burton, R. S., Neigel, J. E., et Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, 70(Supplement 1), 273–290.

- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., et Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, 11(s1), 123–136.
- 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.
- Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., et Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22(11), 2898–2916.
- Hess, J. E., Matala, A. P., et Narum, S. R. (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, 11(s1), 137–149.
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological journal of the Linnean Society*, 58(3), 247–276.
- Hewitt, G. M. (2004). The structure of biodiversity—insights from molecular phylogeography. *Frontiers in Zoology*, 1(4), 1–16.
- Hey, J., et Machado, C. A. (2003). The study of structured populations—new hope for a difficult and divided science. *Nature Reviews Genetics*, 4(7), 535–543.
- Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F., et Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, 54(1), 291–301.

- Hickerson, M. J., Meyer, C. P., et Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, 55(5), 729–739.
- Hoelzer, G. A. (1997). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution*, 51(2), 622–626.
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., et Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2), e1000862.
- Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I., et Xu, C. F. (2004). Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *European Journal of Human Genetics*, 12(5), 395–399.
- Hurt, C., Haddock, S. H., et Browne, W. E. (2013). Molecular phylogenetic evidence for the reorganization of the Hyperiid amphipods, a diverse group of pelagic crustaceans. *Molecular Phylogenetics and Evolution*, 67(1), 28–37.
- Huyse, T., Poulin, R., et Theron, A. (2005). Speciation in parasites: a population genetics approach. *Trends in Parasitology*, 21(10), 469–475.
- Ikeda, T. (2013). Metabolism and chemical composition of marine pelagic amphipods: synthesis toward a global bathymetric model. *Journal of oceanography*, 69(3), 339–355.
- Ilut, D. C., Nydam, M. L., et Hare, M. P. (2014). Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. *BioMed Research International*, 2014, 1–9.
- Ingölfsson, A. (1992). The origin of the rocky shore fauna of Iceland and the Canadian Maritimes. *Journal of Biogeography*, 705–712.
- Johnson, M. S., et Black, R. (1982). Chaotic genetic patchiness in an intertidal limpet, *Siphonaria sp.* *Marine Biology*, 70(2), 157–164.

- Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Jombart, T., et Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071.
- Jombart, T., Devillard, S., et Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94.
- Josefson, A. B., Mokievsky, V., (2013). Marine invertebrates. In *Arctic Council 2013. Arctic Biodiversity Assessment*.
- Kanno, Y., Vokoun, J. C., et Letcher, B. H. (2011). Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology*, 20(18), 3711–3729.
- Katoh, K., et Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.
- Kekkonen, M., et Hebert, P. D. (2014). DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. *Molecular Ecology Resources*, 14(4), 706–715.
- Kinlan, B. P., et Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology*, 84(8), 2007–2020.
- Knowles, L. L., et Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), 887–895.
- Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology and Systematics*, 189–216.

- Knowlton, N., et Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1412), 2257–2263.
- Knowlton, N., Weigt, L. A., Solorzano, L. A., Mills, D.K. et Bermingham, E. (1993). Mitochondrial DNA, and Reproductive Compatibility Across the Isthmus of Panama. *Science*, 260(5114), 1629–1632.
- Knutsen, H., Jorde, P. E., Albert, O. T., Hoelzel, A. R., et Stenseth, N. C. (2007). Population genetic structure in the North Atlantic Greenland halibut (*Reinhardtius hippoglossoides*): influenced by oceanic current systems?. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(6), 857–866.
- Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., et Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes ‘biologically meaningful’? A case study of coastal Atlantic cod. *Molecular Ecology*, 20(4), 768–783.
- Kool, J., Moilanen, A., et Treml, E. (2013). Population connectivity: recent advances and new perspectives. *Landscape Ecology*, 28(2), 165–185.
- Kosobokova, K. N., et Hopcroft, R. R. (2010). Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(1), 96–110.
- Koszteyn, J., Timofeev, S., Węśławski, J. M., et Malinga, B. (1995). Size structure of *Themisto abyssorum* Boeck and *Themisto libellula* (Mandt) populations in European Arctic seas. *Polar Biology*, 15(2), 85–92.
- Kraft, A., Berge, J., Varpe, Ø., et Falk-Petersen, S. (2013)b. Feeding in Arctic darkness: mid-winter diet of the pelagic amphipods *Themisto abyssorum* and *T. libellula*. *Marine Biology*, 160(1), 241–248.
- Kraft, A., Nöthig, E. M., Bauerfeind, E., Wildish, D. J., Pohle, G. W., Bathmann, U. V., Beszczynska-Möller, A. et Klages, M. (2013)c. First evidence of reproductive

- success in a southern invader indicates possible community shifts among Arctic zooplankton. *Marine Ecology Progress series*, 493, 291–296.
- Kuzmin, S. A., Akhtarin, S. M., et Menis, D. T. (1998). The first finding of snow crab *Chionoectes opilio* (Decapoda: Majidae) in the Barrents sea. *Zoologicheskii Zhurnal*, 77, 489–491.
- Laval, P. (1980). Hyperiid amphipods as crustacean parasitoids associated with gelatinous zooplankton. *Oceanography and Marine Biology: Annual Review*, 18, 11–56.
- Lavaniegos, B. E. (2014). Pelagic amphipod assemblage associated with subarctic water off the West Coast of the Baja California peninsula. *Journal of Marine Systems*, 132, 1–12.
- Layton, K. K., Martel, A. L., et Hebert, P. D. (2014). Patterns of DNA barcode variation in Canadian marine molluscs. *PLoS One*, 9(4), e95003.
- Lecointre, G., Philippe, H., V&acirc;n L&ecaron;, H. L., et Le Guyader, H. (1993). Species sampling has a major impact on phylogenetic inference. *Molecular Phylogenetics and Evolution*, 2(3), 205–224.
- Lee, C. E., et Frost, B. W. (2002). Morphological stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia*, 480(1-3), 111–128.
- Lef&ecirc;bure, T., Douady, C. J., Gouy, M., et Gibert, J. (2006). Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, 40(2), 435–447.
- Legendre, P. et F.-J. Lapointe. 2004. Assessing congruence among distance matrices: single malt Scotch whiskies revisited. *Australian and New Zealand Journal of Statistics*, 46, 615–629.
- Legendre, P., Desdevises, Y., et Bazin, E. (2002). A statistical test for host–parasite coevolution. *Systematic Biology*, 51(2), 217–234.

- Lessios, H. A., Kessing, B. D., et Robertson, D. R. (1998). Massive gene flow across the world's most potent marine biogeographic barrier. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1396), 583–588.
- Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, 46(3), 282–297.
- Librado, P. and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452.
- Limborg, M. T., Helyar, S. J., de Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., Carvalho, G. R., FTP CONSORTIUM et Bekkevold, D. (2012). Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, 21(15), 3686–3703.
- Liu, L. (2008). BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, 24(21), 2542–2543.
- Lörz, A. N., et Held, C. (2004). A preliminary molecular and morphological phylogeny of the Antarctic Epimeriidae and Iphimediidae (Crustacea, Amphipoda). *Molecular Phylogenetics and Evolution*, 31(1), 4–15.
- Lörz, A. N., Linse, K., Smith, P. J., et Steinke, D. (2012). First molecular evidence for underestimated biodiversity of Rhachotropis (Crustacea, Amphipoda), with description of a new species. *PLoS One*, 7(3), e32365.
- Lowe, W. H., et Allendorf, F. W. (2010). What can genetics tell us about population connectivity?. *Molecular Ecology*, 19(15), 3038–3051.
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, S. D. et Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, 9(1), e1003215.



- Machida, R., Hashiguchi, Y., Nishida, M. et Nishida, S. (2009). Zooplankton diversity analysis through single-gene sequencing of a community sample. *BMC Genomics*, 10(1): 438.
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic biology*, 46(3), 523–536.
- Maddison, W. P., et Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, 55(1), 21–30.
- Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., Olsen, J., Perez, K. E., Stam, W., Väinölä, R, Viard, F. et Wares, J. (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, 89(sp11), S108–S122.
- Manel, S., Gaggiotti, O. E., et Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20(3), 136–142.
- Marion, A., Harvey, M., Chabot, D., et Brêthes, J. C. (2008). Feeding ecology and predation impact of the recently established amphipod, *Themisto libellula*, in the St. Lawrence marine system, Canada. *Marine Ecology Progress Series*, 373, 53–70.
- Marko, P. B. (2002). Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution*, 19(11), 2005–2021.
- Marko, P. B. 2004. ‘What’s larvae got to do with it?’ Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology*, 13, 597–611.
- Marko, P. B., et Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology and Evolution*, 26(9), 448–456.

- Marta-Almeida, M., Dubert, J., Peliz, Á., et Queiroga, H. (2006). Influence of vertical migration pattern on retention of crab larvae in a seasonal upwelling system. *Marine Ecology Progress Series*, 307, 1–19.
- Martin, J. W., et Davis, G. E. (2001). *An updated classification of the recent Crustacea*. Science Series 39, Natural History Museum of Los Angeles County.
- Matthiessen, J., Knies, J., Vogt, C., et Stein, R. (2009). Pliocene palaeoceanography of the Arctic Ocean and subarctic seas. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1886), 21–48.
- McLaughlin, P. A., Camp, D. K., Angel, M. V., Bousfield, E. L., Brunel, P., Brusca, R. C., et Cadien, D. (2005). *Common and scientific names of aquatic invertebrates from the United States and Canada: crustaceans*. American Fisheries Society.
- Meier, R., Zhang, G., et Ali, F. (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology*, 57(5), 809–813.
- Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24 (13), 3223–3231.
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31–46.
- Milne I, Wright F, Rowe G, Marshal DF, Husmeier D et McGuire G (2004), TOPALi: Software for Automatic Identification of Recombinant Sequences within DNA Multiple Alignments. *Bioinformatics*, 20 (11), 1806–1807.
- Mitarai, S., Siegel, D. A., Watson, J. R., Dong, C., et McWilliams, J. C. (2009). Quantifying connectivity in the coastal ocean with application to the Southern California Bight. *Journal of Geophysical Research: Oceans*, 114(C10).
- Mix, A. C., Bard, E., et Schneider, R. (2001). Environmental processes of the ice age: land, oceans, glaciers. *Quaternary Science Reviews*, 20(4), 627–657.

- Molnar, J. L., Gamboa, R. L., Revenga, C., et Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, 6(9), 485–492.
- Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 49 (4), 718–726.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., et Worm, B. (2011). How many species are there on Earth and in the ocean?. *PLoS Biology*, 9(8), e1001127.
- Morin, P. A., Luikart, G., et Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19(4), 208–216.
- Morin, P. A., Martien, K. K., et Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9(1), 66–73.
- Moritz, C., et Faith, D. P. (1998). Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, 7(4), 419–429.
- Mueter, F. J. et M. A. Litzow (2008). Sea ice retreat alters the biogeography of the Bering Sea continental shelf. *Ecological Applications*, 18(2), 309–320.
- Mulligan, T. J., Chapman, R. W., et Brown, B. L. (1992). Mitochondrial DNA analysis of walleye pollock, *Theragra chalcogramma*, from the eastern Bering Sea and Shelikof Strait, Gulf of Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 49(2), 319–326.
- Neigel, J. E. (2002). Is  $F_{ST}$  obsolete?. *Conservation Genetics*, 3(2), 167–173.
- Nelson, R. J., Carmack, E. C., McLaughlin, F. A., and Cooper, G. A. (2009). Penetration of Pacific zooplankton into the western Arctic Ocean tracked with molecular population genetics. *Marine Ecology Progress Series*, 381, 129–138.
- Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology and Evolution*, 16(7), 358–364.

- Nielsen, J. L., Graziano, S. L., et Seitz, A. C. (2010). Fine-scale population genetic structure in Alaskan Pacific halibut (*Hippoglossus stenolepis*). *Conservation Genetics*, 11(3), 999–1012.
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218.
- Nikula, R., Strelkov, P., et Väinölä, R. (2007). Diversity and trans-arctic invasion history of mitochondrial lineages in the north Atlantic *Macoma balthica* complex (Bivalvia: Tellinidae). *Evolution*, 61(4), 928–941.
- Noor, M. A., et Feder, J. L. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics*, 7(11), 851–861.
- Norris, R. D. (2000). Pelagic species diversity, biogeography, and evolution. *Paleobiology*, 26(sp4), 236–258.
- Norris, R. D. (2000). Pelagic species diversity, biogeography, and evolution. *Paleobiology*, 26(sp4), 236–258.
- Norris, R. D., et de Vargas, C. (2000). Evolution all at sea. *Nature*, 405(6782), 23–24.
- Norris, R. D., et Hull, P. M. (2012). The temporal dimension of marine speciation. *Evolutionary Ecology*, 26(2), 393–415.
- Nosil, P., et Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1587), 332–342.
- Noyon, M., Gasparini, S., et Mayzaud, P. (2009). Feeding of *Themisto libellula* (Amphipoda Crustacea) on natural copepods assemblages in an Arctic fjord (Kongsfjorden, Svalbard). *Polar Biology*, 32(11), 1559–1570.
- O'reilly, P. T., Canino, M. F., Bailey, K. M., et Bentzen, P. (2004). Inverse relationship between  $F_{ST}$  and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, 13(7), 1799–1814.

- Pálsson, S., Kállman, T., Paulsen, J., et Árnason, E. (2009). An assessment of mitochondrial variation in Arctic gadoids. *Polar biology*, 32(3), 471–479.
- Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology and Evolution*, 7(4), 114–118.
- Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, 547–572.
- Palumbi, S. R. (1996). What can molecular genetics contribute to marine biogeography? An urchin's tale. *Journal of Experimental Marine Biology and Ecology*, 203(1), 75–92.
- Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, 13(sp1), 146–158.
- Palumbi, S. R. (2004). Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management. *Annual Review of Environment and Resources*, 29, 31–68.
- Palumbi, S. R., et Kessing, B. D. (1991). Population biology of the trans-Arctic exchange: mtDNA sequence similarity between Pacific and Atlantic sea urchins. *Evolution*, 1790–1805.
- Pan, J., Wang, B., Pei, Z. Y., Zhao, W., Gao, J., Mao, J. F., et Wang, X. R. (2015). Optimization of the genotyping-by-sequencing strategy for population genomic analysis in conifers. *Molecular Ecology Resources*, 15(4), 711–722.
- Paradis, E., Claude, J., et Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290.
- Paterson, A. M., et Banks, J. (2001). Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. *International journal for parasitology*, 31(9), 1012–1022.

- Pawlowski, J., Fahrni, J., Lecroq, B., Longet, D., Cornelius, N., Excoffier, L., Cedhagens, T., et Gooday, A. J. (2007). Bipolar gene flow in deep-sea benthic foraminifera. *Molecular Ecology*, *16*(19), 4089–4096.
- Pearse, D. E., et Crandall, K. A. (2004). Beyond  $F_{ST}$ : analysis of population genetic data for conservation. *Conservation Genetics*, *5*(5), 585–602.
- Peijnenburg, K. T. C. A., Fauvelot, C., Breeuwer, J. A. J., et Menken, S. B. J. (2006). Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology*, *15*(11), 3319–3338.
- Peijnenburg, K. T. C. A., Van Haastrecht, E. K., et Fauvelot, C. (2005). Present-day genetic composition suggests contrasting demographic histories of two dominant chaetognaths of the North-East Atlantic, *Sagitta elegans* and *S. setosa*. *Marine Biology*, *147*(6), 1279–1289.
- Peijnenburg, K. T., et Goetze, E. (2013). High evolutionary potential of marine zooplankton. *Ecology and Evolution*, *3*(8), 2765–2781.
- Pelc, R. A., Warner, R. R., et Gaines, S. D. (2009). Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, *36*(10), 1881–1890.
- Percy, J. A. (1993). Reproduction and growth of the Arctic hyperiid amphipod *Themisto libellula* Mandt. *Polar Biology*, *13*(2), 131–139.
- Pfenninger, M., et Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, *7*(1), 221.
- Pilar Cabezas, M., Cabezas, P., Machordom, A., et Guerra-García, J. M. (2013). Hidden diversity and cryptic speciation refute cosmopolitan distribution in *Caprella penantis* (Crustacea: Amphipoda: Caprellidae). *Journal of Zoological Systematics and Evolutionary Research*, *51*(2), 85–99.

- Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., et Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344(6187), 1246752.
- Pinchuk, A. I., Coyle, K. O., Farley, E. V., et Renner, H. M. (2013). Emergence of the Arctic *Themisto libellula* (Amphipoda: Hyperiididae) on the southeastern Bering Sea shelf as a result of the recent cooling, and its potential impact on the pelagic food web. *ICES Journal of Marine Science: Journal du Conseil*, fst031.
- Pinet, P. R. (2009) *Invitation to Oceanography*. 5th edn. Jones and Bartlett Publishers, Burlington, MA, USA.
- Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M., Jennings, A. E., Miller, G. h., O'Regan, M., Savelle, J., Serreze, M., St. John, K., White, J. W. C., et Wolff, E. (2010). History of sea ice in the Arctic. *Quaternary Science Reviews*, 29(15), 1757–1778.
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., et Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic biology*, 55(4), 595–609.
- Pritchard, J. K., Stephens, M., et Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Prokopowicz, A. J., Rueckert, S., Leander, B. S., Michaud, J., et Fortier, L. (2010). Parasitic infection of the hyperiid amphipod *Themisto libellula* in the Canadian Beaufort Sea (Arctic Ocean), with a description of *Ganymedes Themistos* sp. n. (Apicomplexa, Eugregarinorida). *Polar Biology*, 33(10), 1339–1350.
- Provan, J., et Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology et Evolution*, 23(10), 564–571.

- Puillandre, N., Lambert, A., Brouillet, S., et Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864–1877.
- Radulovici, A. (2012). *A tale of two biodiversity levels inferred from DNA barcoding of selected North Atlantic crustaceans*. PhD thesis. University of Quebec in Rimouski.
- Radulovici, A. E., Archambault, P., et Dufresne, F. (2010). DNA barcodes for marine biodiversity: moving fast forward?. *Diversity*, 2(4), 450–472.
- Radulovici, A. E., Sainte-Marie, B., et Dufresne, F. (2009). DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: a regional-scale approach. *Molecular Ecology Resources*, 9(s1), 181–187.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., et Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, 3(1), 431.
- Rambaut A, Suchard MA, Xie D et Drummond AJ (2014) Tracer v1.6, Récupéré de <http://beast.bio.ed.ac.uk/Tracer>.
- Raupach, M. J., et Radulovici, A. E. (2015). Looking back on a decade of barcoding crustaceans. *ZooKeys*, (539), 53–81.
- Raupach, M. J., Barco, A., Steinke, D., Beermann, J., Laakmann, S., Mohrbeck, I., Neumann, H., Kihara, T. C., Pointner, K., Radulovici, A., Segelken-Voigt, A., Wesse, C. et Kneblsberger, T. (2015). The application of DNA barcodes for the identification of marine crustaceans from the North Sea and adjacent regions. *PLoS One*, 10(9), e0139421.
- Raupach, M. J., Malyutina, M., Brandt, A., et Wägele, J. W. (2007). Molecular data reveal a highly diverse species flock within the munnopsoid deep-sea isopod *Betamorpha fusiformis* (Barnard, 1920)(Crustacea: Isopoda: Asellota) in the



- Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(16), 1820–1830.
- Raupach, M. J., Thatje, S., Dambach, J., Rehm, P., Misof, B., et Leese, F. (2010). Genetic homogeneity and circum-Antarctic distribution of two benthic shrimp species of the Southern Ocean, *Chorismus antarcticus* and *Nematocarcinus lanceopes*. *Marine Biology*, 157(8), 1783–1797.
- Rees, D. J., Dufresne, F., Glémet, H., et Belzile, C. (2007). Amphipod genome sizes: first estimates for Arctic species reveal genomic giants. *Genome*, 50(2), 151–158.
- Reid, P. C., Johns, D. G., Edwards, M., Starr, M., Poulin, M, et Snoeijs, P. (2007). A biological consequence of reducing Arctic ice cover: arrival of the Pacific diatom *Neodenticula seminae* in the North Atlantic for the first time in 800 000 years. *Global Change Biology*, 13(9), 1910–1921.
- Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., et Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, 22(11), 2953–2970.
- Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, 424(6952), 1042–1047.
- Rocha, L. A., Craig, M. T., et Bowen, B. W. (2007). Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, 26(3), 501–512.
- Rode, K. D., Peacock, E., Taylor, M., Stirling, I., Born, E. W., Laidre, K. L., et Wiig, Ø. (2012). A tale of two polar bear populations: ice habitat, harvest, and body condition. *Population Ecology*, 54(1), 3–18.
- Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., et Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and

- hierarchical SNP selection. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12518.
- Rooney, A. P., Piontkivska, H., et Nei, M. (2002). Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Molecular Biology and Evolution*, 19(1), 68–75.
- Saenz-Agueldo, P., Jones, G. P., Thorrold, S. R., et Planes, S. (2009). Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, 18(8), 1765–1776.
- Sanford, E., et Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3, 509–535.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Von Horn, D. J., et Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541.
- Schneppenheim, R., et Weigmann-Haass, R. (1986). Morphological and electrophoretic studies of the genus *Themisto* (Amphipoda: Hyperiidea) from the South and North Atlantic. *Polar Biology*, 6(4), 215–225.
- Selkoe, K. A., et Toonen, R. J. (2011). Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, 436, 291–305.
- Selkoe, K. A., Gaines, S. D., Caselle, J. E., et Warner, R. R. (2006). Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, 87(12), 3082–3094.

- Selkoe, K. A., Watson, J. R., White, C., Horin, T. B., Iacchei, M., Mitarai, S., Siegel, D. A., Gaines, S. D., et Toonen, R. J. (2010). Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, *19*(17), 3708–3726.
- Serejo, C. S., et Sittrop, D. J. (2009). Hyalidae. *Zootaxa*, *2260*, 440–452.
- Shanks, A. L., Grantham, B. A., et Carr, M. H. (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, *13*(sp1), 159–169.
- Shanks, A. L. (2009). Pelagic larval duration and dispersal distance revisited. *The Biological Bulletin*, *216*(3), 373–385.
- Sher, A. (1999). Traffic lights at the Beringian crossroads. *Nature*, *397*(6715), 103–104.
- Simmons, M. P., Ochoterena, H., et Freudenstein, J. V. (2002). Amino acid vs. nucleotide characters: challenging preconceived notions. *Molecular Phylogenetics and Evolution*, *24*(1), 78–90.
- Sirenko, B. I. (Ed.). (2001). *List of species of free-living invertebrates of Eurasian Arctic seas and adjacent deep waters*. Russian Academy of Science, Zoological Institute.
- Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462.
- Sobel, J. M., Chen, G. F., Watt, L. R., et Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*(2), 295–315.
- Song, H., Buhay, J. E., Whiting, M. F., et Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *PNAS*, *105*(36).

- Spies, I. (2012). Landscape genetics reveals population subdivision in Bering Sea and Aleutian Islands Pacific cod. *Transactions of the American Fisheries Society*, 141(6), 1557–1573.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123 (3), 585–595.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., et Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), 2731–2739.
- Thiel, M. et Haye, P. A. (2006) The ecology of rafting in the marine environment III: biogeographical and evolutionary consequences. *Oceanography and Marine Biology*, 44,323–429.
- Thomas, J. A., Welch, J. J., Lanfear, R., et Bromham, L. (2010). A generation time effect on the rate of molecular evolution in invertebrates. *Molecular Biology and Evolution*, 27(5), 1173–1180.
- Thomas, L., et Bell, J. J. (2013). Testing the consistency of connectivity patterns for a widely dispersing marine species. *Heredity*, 111(4), 345–354.
- Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., et Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, 466(7310), 1098–1101.
- Toews, D. P., et Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930.
- Toussaint, R. K., Scheel, D., Sage, G. K., et Talbot, S. L. (2012). Nuclear and mitochondrial markers reveal evidence for genetically segregated cryptic speciation in giant Pacific octopuses from Prince William Sound, Alaska. *Conservation Genetics*, 13(6), 1483–1497.

- Trebitz, A. S., Hoffman, J. C., Grant, G. W., Billehus, T. M., et Pilgrim, E. M. (2015). Potential for DNA-based identification of Great Lakes fauna: match and mismatch between taxa inventories and DNA barcode libraries. *Scientific reports*, 5, 12162.
- Trontelj, P., Douady, C. J., Fiser, C., Gibert, J., Goricki, Š., Lefébure, T., Sket, B., et Zaksek, V. (2009). A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts?. *Freshwater Biology*, 54(4), 727–744.
- Turakulov, R., et Easteal, S. (2003). Number of SNPS loci needed to detect population structure. *Human Heredity*, 55(1), 37–45.
- Unal, E., et Bucklin, A. (2010). Basin-scale population genetic structure of the planktonic copepod *Calanus finmarchicus* in the North Atlantic Ocean. *Progress in Oceanography*, 87(1), 175–185.
- Underwood, J. N., Smith, L. D., Van Oppen, M. J. H., et Gilmour, J. P. (2007). Multiple scales of genetic connectivity in a brooding coral on isolated reefs following catastrophic bleaching. *Molecular Ecology*, 16(4), 771–784.
- Uthicke, S., et Benzie, J. A. H. (2003). Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology*, 12(10), 2635–2648.
- Väinölä, R. (2003). Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. *Marine Biology*, 143(5), 935–946.
- Väinölä, R., Witt, J. D. S., Grabowski, M., Bradbury, J. H., Jazdzewski, K., et Sket, B. (2008). Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia*, 595(1), 241–255.
- van Hannen, E. J., Mooij, W., van Agterveld, M. P., Gons, H. J., and Laanbroek, H. J. (1999). Detritus-Dependent Development of the Microbial Community in an

- Experimental System: Qualitative Analysis by Denaturing Gradient Gel Electrophoresis. *Applied and Environmental Microbiology*, 65(6), 2478–2484.
- Van Oppen, M. J. H., Draisma, S. G. A., Olsen, J. L., et Stam, W. T. (1995). Multiple trans-Arctic passages in the red alga *Phycodrys rubens*: evidence from nuclear rDNA ITS sequences. *Marine Biology*, 123(1), 179–188.
- Vermeij, G. J. (1991). Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology*, 17(03), 281–307.
- Vignal, A., Milan, D., SanCristobal, M., et Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), 275–306.
- Vihma, T. (2014). Effects of Arctic sea ice decline on weather and climate: a review. *Surveys in Geophysics*, 35(5), 1175–1214.
- Vinogradov M. E., Volkov, A. F., et Semenova, T. N. (1996). *Hyperiid amphipods (Amphipoda, Hyperiidea) of the world oceans*. Lebanon, USA: Science Publishers.
- Vinogradov, M.E. et Semenova, T.N. (2009). *Illustrated Keys to Free-Living Invertebrates of Eurasian Arctic Seas and Adjacent Deep Waters*, Vol. 1. Rotifera, Pycnogonida, Cirripedia, Leptostraca, Mysidacea, Hyperiidea, Caprellidea, Euphausiacea, Natantia, Anomura, and Brachyura. Alaska Sea Grant. University of Alaska Fairbanks. Vinogradov Vassilenko, S.V., and V.V. Petryashov (eds.).
- Vogler, A. P., et M. T. Monaghan. Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*, 45.1 (2007), 1–10.
- Waples, R. S., et Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419–1439.
- Wassmann, P. (2011). Arctic marine ecosystems in an era of rapid climate change. *Progress in Oceanography*, 90(1), 1–17.

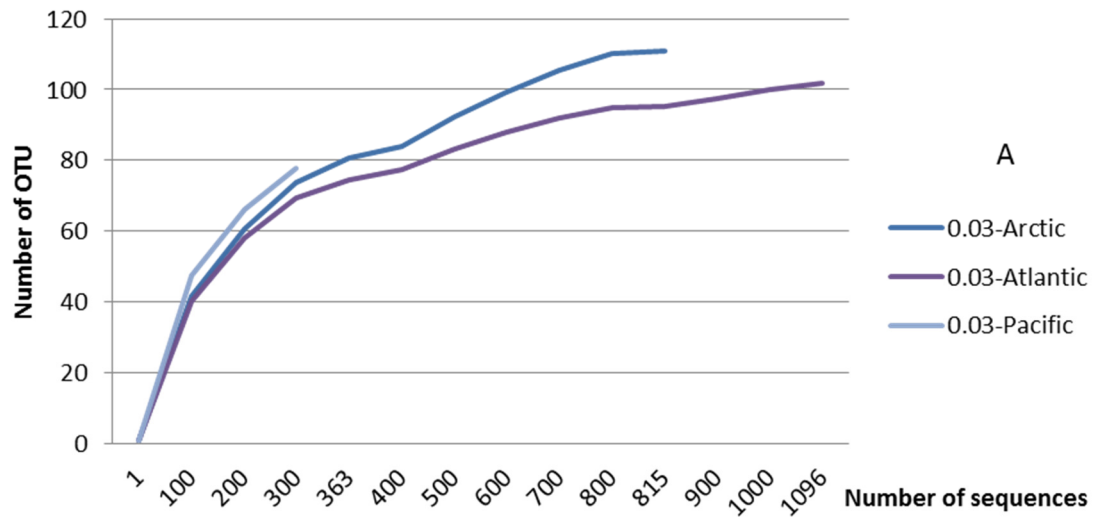
- Wassmann, P., Duarte, C. M., Agusti, S., et Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology*, 17(2), 1235–1249.
- Wassmann, P., Kosobokova, K. N., Slagstad, D., Drinkwater, K. F., Hopcroft, R. R., Moore, S. E., Ellingsen, I., Nelson, R. J., Carmack, E., Popova, E. et Berge, J. (2015). The contiguous domains of Arctic Ocean advection: trails of life and death. *Progress in Oceanography*, 139, 42–65.
- Waters, J. M., et Roy, M. S. (2004). Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow?. *Molecular Ecology*, 13(9), 2797–2806.
- Watts, J., et Tarling, G. A. (2012). Population dynamics and production of *Themisto gaudichaudii* (Amphipoda, Hyperiidae) at South Georgia, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, 59, 117–129.
- Webb, C. O., Slik, J. F., et Triono, T. (2010). Biodiversity inventory and informatics in Southeast Asia. *Biodiversity and Conservation*, 19(4), 955–972.
- Weersing, K. A. (2007). *Population genetics, larval dispersal, and demographic connectivity in marine systems*. Master Thesis. University of Hawaii.
- Weersing, K., et Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, 393, 1–12.
- Weiner, A., Aurahs, R., Kurasawa, A., Kitazato, H., et Kucera, M. (2012). Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Molecular Ecology*, 21(16), 4063–4073.
- Wencki, K., Morza, Z. E., Pan, I. O., et Warszaway, P. (2000). Interannual variability in the occurrence of *Themisto* (Amphipoda) in the north Norwegian Sea. *Polish Polar Research*, 21, 143–152.

- Węśławski, J., Opanowski, A., Legeżyńska, J., Maciejewska, B., Włodarska-Kowalczyk, M., et Kędra, M. (2010). Hidden diversity in Arctic crustaceans. How many roles can a species play?. *Polish Polar Research*, 31(3), 205–216.
- White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., et Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20092214.
- White, T. A., Perkins, S. E., Heckel, G., et Searle, J. B. (2013). Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, 22(11), 2971–2985.
- Whitlock, M. C., et McCauley, D. E. (1999). Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm + 1)$ . *Heredity*, 82(2), 117–125.
- Willing, E. M., Dreyer, C., et Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by  $F_{ST}$  do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, 7(8), e42649.
- Wing, B. L. (1976). *Ecology of Parathemisto Libellula and P. Pacifica (Amphipoda, Hyperiidea) in Alaskan Coastal Waters*. PhD thesis. Northwest Fisheries Center Auke Bay Fisheries Laboratory. National Marine Fisheries Service, NOAA.
- Winston, J. E. (2012). Dispersal in marine organisms without a pelagic larval phase. *Integrative and Comparative Biology*, ics040.
- Witt, J. D., Threlhoff, D. L., et Hebert, P. D. (2006). DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*, 15(10), 3073–3082.
- Woodson, C. B., et McManus, M. A. (2007). Foraging behavior can influence dispersal of marine organisms. *Limnology and Oceanography*, 52(6), 2701–2709.
- Wu, C. I., et Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics*, 5(2), 114–122.

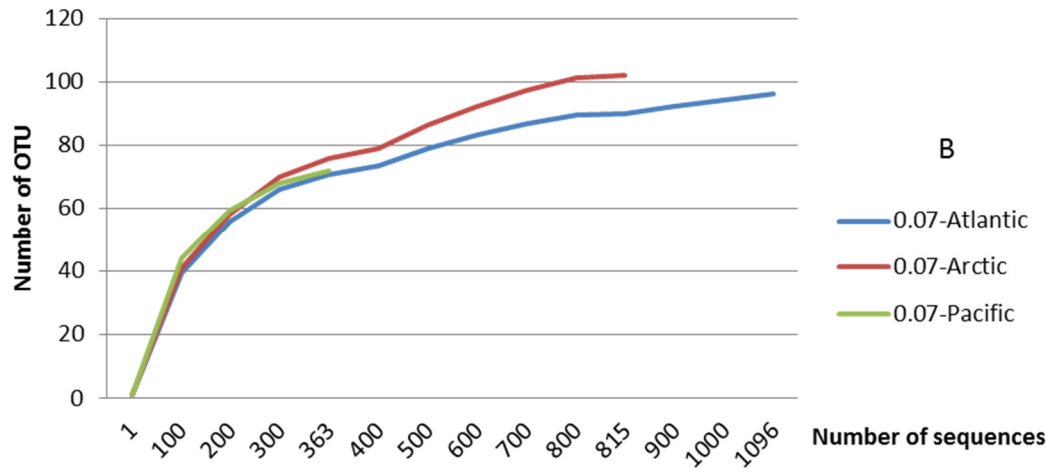


- Xia X, (2009) Assessing substitution saturation with DAMBE. In: Lemey P, Salemi, M., Vandamme, A.,M. (eds) *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, 2nd edn. Cambridge University Press, Cambridge.
- Xia, X., et Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92(4), 371–373.
- Xia, X., Xie, Z., Salemi, M., Chen, L., et Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1), 1–7.
- Yamada, Y., Ikeda, T., et Tsuda, A. (2004). Comparative life-history study on sympatric hyperiid amphipods (*Themisto pacifica* and *T. japonica*) in the Oyashio region, western North Pacific. *Marine Biology*, 145(3), 515–527.
- Yang, Z., et Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13(5), 303–314.
- Zeidler, W. (2004). *A review of the families and genera of the hyperiidean amphipod superfamily Phronimoidea Bowman et Gruner, 1973 (Crustacea: Amphipoda: Hyperiidea)*. Magnolia Press.
- Zeidler, W. (2015). A review of the hyperiidean amphipod genus *Hyperoche* Bovallius, 1887 (Crustacea: Amphipoda: Hyperiidea: Hyperiididae), with the description of a new genus to accommodate *H. shihi* Gasca, 2005. *Zootaxa*, 3905(2), 151–192.
- Zhang, A. B., He, L. J., Crozier, R. H., Muster, C., et Zhu, C. D. (2010). Estimating sample sizes for DNA barcoding. *Molecular Phylogenetics and Evolution*, 54(3), 1035–1039.
- Zhang, D. X., et Hewitt, G. M. (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, 11(6), 247–251.
- Zink, R. M. (2002). Methods in comparative phylogeography, and their application to studying evolution in the North American aridlands. *Integrative and Comparative Biology*, 42(5), 953–959.

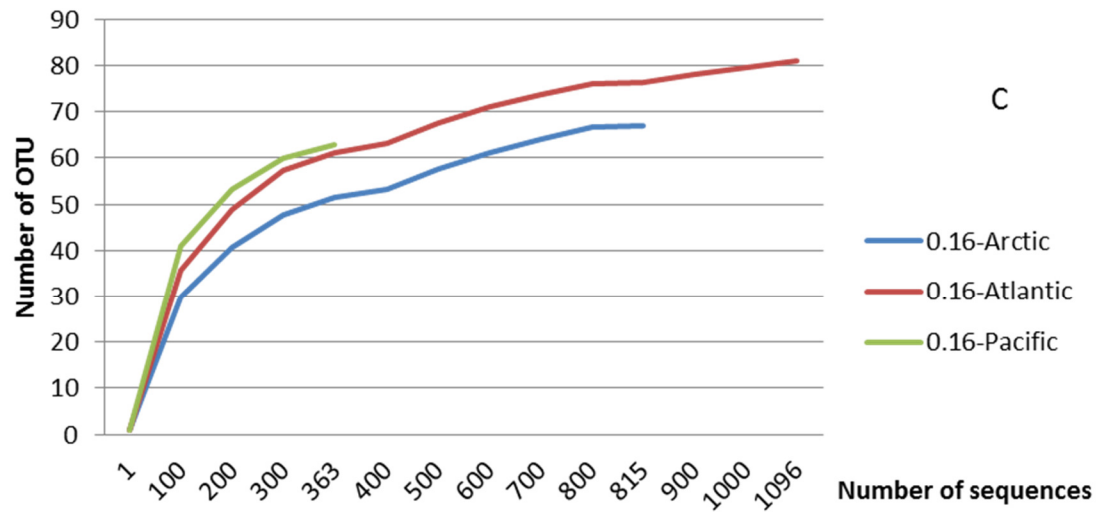
## APPENDICE A



**Figure A.1** Rarefaction curve showing the number of OTUs found in each ocean according to the number of sequences sampled. (A) at 3% threshold.

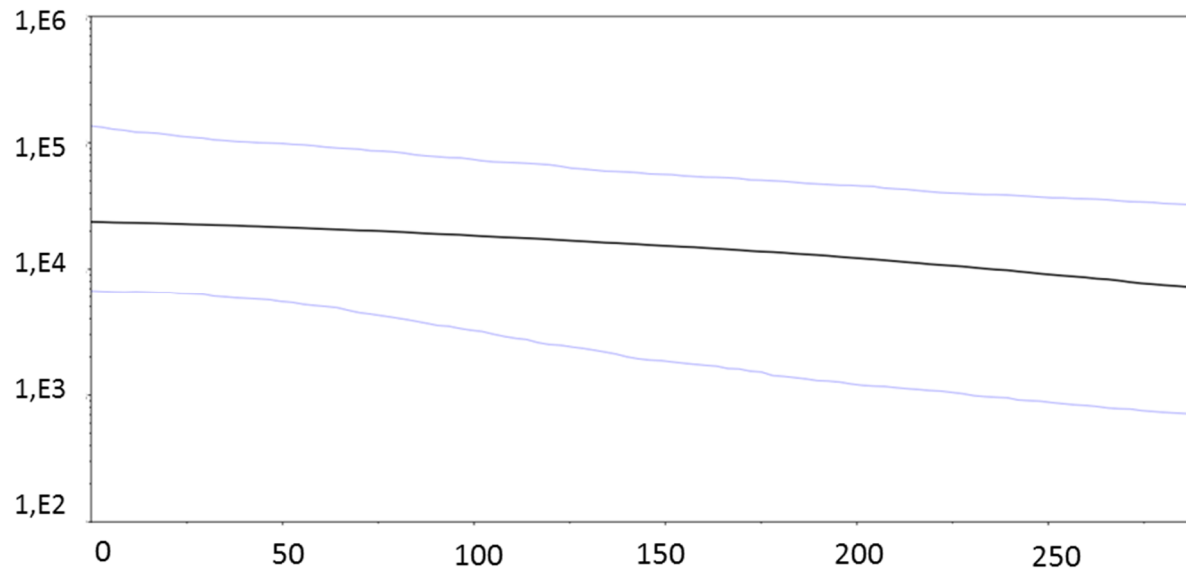


**Figure A.2** Rarefaction curve showing the number of OTUs found in each ocean according to the the number of sequence sampled (B) at 7% threshold.

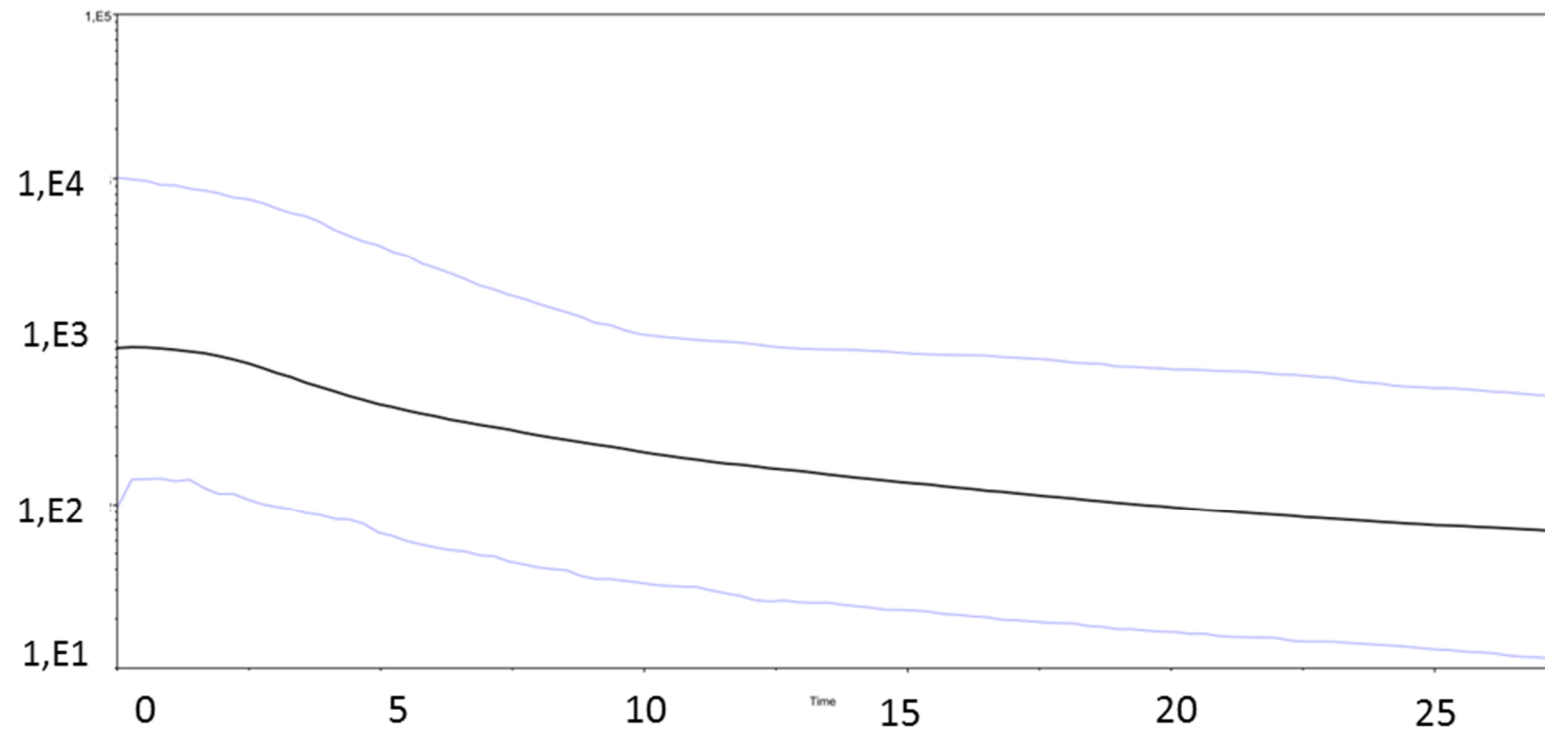


**Figure A.3** Rarefaction curve showing the number of OTUs found in each ocean according to the number of sequences sampled (C) at 16% threshold.

## APPENDICE B



**Figure B.1** Bayesian Skyline Plots of effective population size through time obtained using a 27%/My rate for *T. abyssorum*. Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year.



**Figure B.2** Bayesian Skyline Plots of effective population size through time obtained using a 27%/My rate for *T. libellula*. Blue lines represent interval confidence and black line the mean estimate. The y axis represents the estimated effective population size. The x-axis represents the time in year.