

UNIVERSITÉ DU QUÉBEC

THÈSE PRÉSENTÉE À  
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DU DOCTORAT EN BIOPHYSIQUE ET BIOLOGIE CELLULAIRES

PAR  
CHRISTINE GAUDREAU

CONTRIBUTION DES OUTILS MOLÉCULAIRES À LA TAXINOMIE ET  
À LA PHYLOGÉNIE DES SIMULIES

DÉCEMBRE 2010

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

# UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN BIOPHYSIQUE ET BIOLOGIE CELLULAIRES (PH. D.)

Programme offert à l'Université du Québec à Trois-Rivières

CONTRIBUTION DES OUTILS MOLÉCULAIRES À LA TAXINOMIE  
ET À LA PHYLOGÉNIE DES SIMULIES

PAR

CHRISTINE GAUDREAU

---

Guy Charpentier, directeur de recherche                      Université du Québec à Trois-Rivières

---

Hélène Glémet, présidente du jury                              Université du Québec à Trois-Rivières

---

Bernard LaRue, codirecteur de recherche                      Université du Québec à Trois-Rivières

---

Jean-Charles Côté, évaluateur                                      Agriculture et agro-alimentaire Canada  
Université du Québec à Trois-Rivières

---

Michel Cusson, évaluateur externe                                      Université Laval

Thèse soutenue le 19 novembre 2010

## AVANT-PROPOS

Conformément aux articles D45 à D47 du règlement des études de cycles supérieurs de l'Université du Québec à Trois-Rivières, il est possible de présenter les résultats obtenus sous forme d'articles scientifiques. Le chapitre 1 de cette thèse est composé de l'introduction et de la méthodologie non décrite dans les articles. Les chapitres 2 à 6 constituent les cinq articles scientifiques qui englobent les résultats accumulés. Le chapitre 7 met ces résultats en perspective dans le cadre d'une discussion globale. Pour des raisons de simplicité, la bibliographie associée aux chapitres 1 et 7 a été reportée à la fin de ce dernier.

Le premier article scientifique « Seasonal and spatial distributions of blackfly larvae (Diptera : Simuliidae) in two lake outlet streams of the Mauricie Region of Quebec, and species survey in parts of Southern Quebec Territory» (chapitre 2) englobe l'ensemble des données recueillies lors de nos sorties sur le terrain. Les articles scientifiques suivants, «Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies (Diptera: Simuliidae)» (chapitre 3), «Phylogenetic suitability of COI, COII and 18S in land invertebrates: a black fly (Diptera : Simuliidae) case study» (chapitre 4), «Molecular comparison of Quebec and Newfoundland populations of the blackfly, *Simulium vittatum*, species complex» (chapitre 5) et «Recent origins and genetic fragmentation of *Simulium (Hebridosimulium) laciniatum* (Diptera), the endemic Fijian simuliid» (chapitre 6) présentent l'ensemble des données moléculaires obtenues.

Chaque article est conforme aux normes de la revue dans lequel il a été publié/soumis, exception faite qu'au sein de la thèse, nous avons volontairement agencé chaque figure avec sa légende afin de faciliter la compréhension.

## REMERCIEMENTS

La réalisation d'un tel travail n'aurait pu avoir lieu sans la participation de certaines personnes que je tiens à mentionner.

Tout d'abord, à mon directeur de recherche, Guy Charpentier, mon gourou de la mouche noire, sans qui je n'aurais pu identifier autant de larves aux sourires radieux. Lutin numéro 1 dans la chasse aux larves, tu as su égayer les sorties terrains, apporter la gaieté au laboratoire et le soutien nécessaire à la réussite d'un tel projet. Merci pour toutes les heures consacrées à me transmettre ces connaissances.

À mon co-directeur de recherche, le Dr. Bernard LaRue, sans les nombreux talents et l'expertise de qui je n'aurais pu faire avancer un tel projet aussi rapidement. Merci pour les judicieux conseils et tout le bagage moléculaire que tu as pu me donner durant ces années.

À Lise Meilleur, lutin numéro 2, amie et partenaire de terrain, que d'heures passées à tes côtés à se faire manger par les petits vampires volants pendant que les larves nous apparaissaient miettes par miettes. Mais où sont-elles cachées?

Je tiens aussi à exprimer toute ma gratitude à mes parents, pour m'avoir encouragée et supportée. Vous qui avez passé des heures à essayer de saisir le charabia incompréhensible dans lequel j'évolue depuis 4 ans pour m'aider à comprendre les trucs qui font que j'ai moins de cheveux sur la tête... Merci d'avoir allumé la lanterne du phare de temps à autre. Merci « pôpa » d'avoir passé de nombreuses heures à découvrir les joies de la bioinformatique ainsi que d'avoir corrigé mes textes. Merci « môman » pour la place dans le frigo qui n'a jamais été aussi peuplé de larves; merci pour toute l'énergie et la volonté qu'une mère peut transmettre à sa fille.

Aux autres membres du laboratoire, Mylène St-Onge (lutin numéro 3), Carole Charbonneau et à nos stagiaires, Valérie Charbonneau, Hubert O. Bagre, Aude Pochon et Pauline Marquer, merci pour l'environnement de travail harmonieux, les petits diners en groupe et les bons moments que nous avons eus.

Des remerciements sont également adressés aux techniciennes de laboratoire Micheline Boulianne et Diane Bégin pour leur générosité quant aux matériels empruntés ainsi que pour leurs précieux conseils. Un merci tout spécial à Nathalie Godbout, avec qui j'ai eu un réel plaisir durant ces années dans le laboratoire de Zoologie des invertébrés et qui a été à mes côtés pour me supporter dans mes projets les plus fous. Un gros merci à Catarina Leote Franco Pio pour le petit cours de Word 101. De plus, je remercie tout spécialement le Dr. Douglas A. Craig (University of Alberta) et le Dr. Murray Colbo (Memorial University) pour nous avoir fourni plusieurs spécimens de larves.

Enfin, merci à tous ces gens pour leurs encouragements constants et l'aide qu'ils ont pu m'apporter dans la réalisation d'un tel défi.

Mille Mercis  
Christine ☺

## TABLE DES MATIÈRES

<b>AVANT-PROPOS</b> .....	<b>iii</b>
<b>REMERCIEMENTS</b> .....	<b>iv</b>
<b>LISTE DES TABLEAUX</b> .....	<b>xii</b>
<b>LISTE DES FIGURES</b> .....	<b>xiii</b>
<b>LISTE DES SYMBOLES ET ABRÉVIATIONS</b> .....	<b>xv</b>
<b>Chapitre 1 Introduction et problématique</b> .....	<b>1</b>
1.1 Résumé .....	2
1.2 Introduction .....	3
1.2.1 Position taxinomique des similies .....	5
1.2.2 Inventaire des similies du Québec .....	6
1.2.3 Problème associé aux critères d'identification morphologique des similies.....	7
1.2.3.1 La couleur des larves et le patron des taches de l'apotome céphalique .....	9
1.2.3.2 Le nombre de rayons primaires sur les plumeaux céphaliques.....	10
1.2.3.3 Autres problèmes rencontrés lors de l'identification morphologique .....	11
1.2.4 Cytogénétique des similies .....	11
1.2.5 Vers une taxinomie moléculaire ? .....	14
1.2.6 Problématique et objectifs .....	16
1.3 Méthodologie.....	18
1.3.1 Constitution de la banque de spécimens et inventaire.....	18
1.3.2 L'identification morphologique.....	22
1.3.3 Méthodes d'identification moléculaire.....	23
1.3.3.1 Extraction d'ADN .....	23

1.3.3.2	Création des amorces.....	24
1.3.3.3	Amplification PCR.....	26
1.3.3.4	Séquençage.....	26
1.3.3.5	La construction du jeu de données .....	27
1.3.3.6	L’alignement des séquences.....	28
1.3.3.7	L’inférence phylogénétique.....	30
1.3.3.8	Modèle d’évolution des séquences.....	31
<b>Chapitre 2</b>	<b>Seasonal and Spatial Distributions of Black fly Larvae (Diptera: Simuliidae) in Two Lake Outlet Streams of the Mauricie Region of Québec, and Species Survey in Parts of Southern Quebec Territory.....</b>	<b>34</b>
2.1	Résumé.....	36
2.1.1	Contribution des auteurs de l’article.....	37
2.2	Abstract.....	37
2.3	Introduction.....	38
2.4	Field Site Description and Methods.....	40
2.5	Results and Discussion .....	42
2.5.1	Distribution of taxa in Québec.....	42
2.5.2	Temporal and Spatial Changes in Species Distribution .....	45
2.6	Remarks and Conclusion .....	51
2.6.1	Québec Simuliidae Survey .....	51
2.6.2	Limitations and Limits of Larval Identity .....	53
2.7	Acknowledgements.....	56
2.8	Litterature cited.....	56
2.9	Tables and figures.....	57



<b>Chapitre 3</b>	<b>Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies (Diptera: Simuliidae) .....</b>	<b>77</b>
3.1	Résumé.....	79
3.1.1	Contribution des auteurs de l'article.....	80
3.2	Abstract.....	80
3.3	Introduction.....	82
3.4	Materials and methods .....	84
3.4.1	Sample collection .....	84
3.4.2	DNA amplification .....	85
3.4.3	Purification and sequencing of PCR products.....	85
3.4.4	Sequence data analysis .....	85
3.5	Results.....	87
3.5.1	Taxonomic considerations.....	87
3.5.2	ITS1 comes in two basic sizes.....	88
3.5.3	Intraspecific and genomic ITS1 variability .....	90
3.5.4	A secondary structure for ITS1?.....	92
3.5.5	ITS2: a secondary structure .....	93
3.5.6	An ITS2-derived phylogeny .....	95
3.6	Discussion.....	97
3.7	Acknowledgements.....	100
3.8	References.....	100
3.9	Table and figures .....	108
<b>Chapitre 4</b>	<b>Phylogenetic suitability of COI, COII and 18S in land invertebrates: a black fly (Diptera : Simuliidae) case study .....</b>	<b>116</b>
4.1	Résumé.....	118
4.1.1	Contribution des auteurs de l'article.....	118
4.2	Abstract.....	119
4.3	Introduction.....	120
4.4	Methods .....	122
4.4.1	General collecting.....	122

4.4.2	DNA amplification and sequencing .....	122
4.4.3	Sequence data analysis .....	123
4.4.4	Simulating the evolutionary process .....	123
4.5	Results.....	124
4.5.1	COI and COII: general considerations .....	124
4.5.2	Genetic differentiation within tight groups .....	125
4.5.3	A tentative phylogeny from the 18S-5.8S rDNA cluster.....	126
4.6	Discussion.....	128
4.7	Acknowledgements.....	132
4.8	References.....	133
4.9	Table and figures .....	141
<b>Chapitre 5</b>	<b>Molecular comparison of Quebec and Newfoundland populations of the blackfly, <i>Simulium vittatum</i>, species complex .....</b>	<b>146</b>
5.1	Résumé.....	148
5.1.1	Contribution des auteurs de l'article.....	148
5.2	Abstract.....	148
5.3	Short Communication.....	149
5.4	Acknowledgements.....	154
5.5	References.....	155
5.6	Table and figures .....	158
<b>Chapitre 6</b>	<b>Recent origins and genetic fragmentation of <i>Simulium (Hebridosimulium) laciniatum</i> (Diptera), the endemic Fijian simuliid.....</b>	<b>160</b>
6.1	Résumé.....	162
6.1.1	Contribution des auteurs de l'article.....	163
6.2	Abstract.....	163
6.3	Introduction.....	164
6.4	Methods .....	166

6.4.1	Sample collection .....	166
6.4.2	DNA amplification .....	167
6.4.3	Purification and sequencing of PCR products .....	167
6.4.4	Analysis of sequence data.....	168
6.5	Results.....	168
6.5.1	COII .....	168
6.5.2	ITS1 .....	171
6.6	Discussion.....	174
6.7	Acknowledgements.....	180
6.8	References.....	180
6.9	Table and figures .....	186
<b>Chapitre 7 Résultats et discussion .....</b>		<b>192</b>
7.1	Résultats et discussion.....	193
7.1.1	Un critère d'interfécondité difficilement vérifiable.....	193
7.1.2	Problématique de la définition moléculaire de l'espèce .....	193
7.1.3	Étude moléculaire de populations .....	195
7.1.3.1	Cas de <i>S. laciniatum</i> .....	195
7.1.3.2	Cas de <i>S. vittatum</i> .....	196
7.1.3.3	Cas de <i>S. euryadminiculum</i> / <i>S. annulus</i> .....	198
7.1.3.4	Cas de <i>S. longistylatum</i> / <i>S. pictipes</i> .....	199
7.1.4.5	Cas des <i>Prosimulium</i> et <i>Helodon</i> .....	201
7.1.4	Utilité des divers marqueurs pour établir une phylogénie des simulies.....	202
7.1.4.1	L'unité de transcription ADNr .....	203
7.1.4.2	Espaceur 2a et ADNr 5.8S et 2S .....	203
7.1.4.3	ADNr 18S.....	204
7.1.4.4	ITS .....	206
7.1.4.5	L'ADN mitochondrial .....	208

7.1.4.6	Confirmation externe et interne de nos phylogénies.....	209
7.2	Problèmes liés aux banques de données.....	210
7.3	Conclusion et perspective.....	212
7.4	Bibliographie.....	214

## LISTE DES TABLEAUX

1.1	Liste des amorces.....	25
2.1	List of species, source of identification, months, and sites from which black fly larvae were collected through Québec .....	67
2.2	Physical and chemical parameters of Lac Souris and Lac du Barrage .....	69
2.3	Distribution and cytotypes for each species .....	70
2.4	Percentage comparison of each species found in February and in March/April of 2006 and 2008 for Lac Souris and Lac du Barrage.....	72
3.1	AT% and length for ITS1 and ITS2 in 15 black fly species from eastern Canada.....	108
4.1	List of species .....	141
4.2	Average base composition of the COI + COII genes .....	141
4.3	Amino acid interchanges in the COI and COII polypeptides.....	142
5.1	Divergence levels for the COI/COII aggregate in comparisons involving pairs of sequences (Kimura 2 parameters distance).....	158
6.1	Geographic and physical data for localities of <i>Simulium</i> ( <i>Hebridosimulium</i> ) <i>laciniatum</i> used in this study.....	186
6.2	Distribution of ITS1 alleles on Fijian islands.....	187

## LISTE DES FIGURES

1.1	Plumeaux céphaliques primaires et secondaires.....	8
1.2	Carcéristiques morphologiques de la tête des larves utilisées pour l'identification des mouches noires. ....	9
1.3	Distribution des taches céphaliques chez <i>S. vittatum</i> .....	10
1.4	Lieux d'échantillonnage .....	20
1.5	Indication des nouvelles aires de colonisation pour les simulies du Québec représentées sur les cartes provenant d'Adler <i>et al.</i> (2004) .....	21
1.6	Larves parasitées par un mermithide et des microsporidies .....	23
1.7	Électrophorégramme de deux séquences sœurs distinguées par une insertion simple.....	27
2.1	Study sites for temporal and spatial distribution .....	73
2.2	Seasonal variation in species distribution in 2006.....	74
2.3	Species abundance and distribution at each of the six stations .....	75
2.4	Hypostomal tooth patterns by which <i>Prosimulium</i> species were identified.....	76
3.1	Aligned ITS1 sequences from <i>Twinnia</i> (1), <i>Prosimulium</i> 1 consensus (2), <i>Prosimulium</i> 2 consensus (3), <i>S. quebecense</i> (4) and <i>S. mutata</i> (5) .....	109
3.2	Internal repeats and polymorphisms in ITS1 .....	110
3.3	The ITS1-5.8S rDNA region .....	111
3.4	Schematic ITS1 secondary structure in <i>Prosimuliini</i> .....	112
3.5	(A) The three stem-loop domains of ITS2 (as RNA), using <i>S. tuberosum</i> variant 1 as an example and (B) Structural consequences of base changes involving near-twin sequences .....	113
3.6	An ITS2 phylogeny generated through the GTR model, with the root arbitrarily set between <i>Prosimuliini</i> and <i>Simuliini</i> .....	115
4.1	Unrooted tree for COI/COII .....	143

4.2	Rooted tree for rDNA (18S and 5.8S region).....	144
4.3	Mutation map for aligned black fly rDNA sequences.....	145
5.1	Individual COI (Iceland and Nebraska: unavailable) and COII haplotypes.....	159
5.2	Sense strand sequence of the 103-105 bp internal transcribed spacer-1 (ITS1).....	159
6.1	Map of Fiji.....	188
6.2	Partial sequence of the COII polypeptide from <i>Simulium</i> ( <i>Hebridosimulium</i> ) <i>laciniatum</i> , with position numbering based on the <i>Cnephia dacotensis</i> polypeptide (inferred through DNA sequence, Genbank AF425841) .....	189
6.3	COII versus COI gene divergence (K2P distance) for various North American Simuliidae .....	190
6.4	Characterization of ITS1 .....	191
7.1	Opéron de l'ADNr.....	203
7.2	Mutations et phylogénie de la région englobant l'espaceur 2a et les ADNr 5.8S et 2S.....	204
7.3	Mutations et phylogénie de l'ADNr 18S.....	205
7.4	Arbre phylogénétique d'ITS1 .....	207
7.5	Arbre phylogénétique du mtADNr 16S.....	209
7.6	Origine des données du BOLD sur les simuliés .....	211

## LISTE DES SYMBOLES ET ABRÉVIATIONS

Cette section comporte une description bilingue des abréviations utilisées dans les portions francophones et anglophones du manuscrit. Afin de simplifier la consultation de cette liste, l'ordre alphabétique des abréviations fut préféré à l'ordre d'apparition dans le texte.

Selon le contexte, le nom abrégé d'une protéine peut aussi servir à désigner son gène.

12S ADNr	ADN ribosomal 12S des mitochondries
16S ADNr	ADN ribosomal 16S des mitochondries
A	Adénine
ADN (DNA)	Acide désoxyribonucléique (deoxyribonucleic acid)
ADNmt (mtDNA)	ADN mitochondrial (mitochondrial DNA)
ADNr (rDNA)	ADN ribosomal (ribosomal DNA)
AIC	Akaike Information Criterion
ARN (RNA)	Acide ribonucléique (ribonucleic acid)
ARNr (rRNA)	ARN ribosomal (ribosomal RNA)
BDWD	BioSystematic Database of World Diptera
BOLD	Barcode of life data systems
C	Cytosine
CAD	Gène de la carbamoylphosphate synthase
COI	Sous-unité I de la cytochrome oxydase
COII	Sous-unité II de la cytochrome oxydase
COIII	Sous-unité III de la cytochrome oxydase
Cytb	Cytochrome b
DDC	Dopa decarboxylase
EF $\alpha$ 1	Facteur d'élongation $\alpha$ 1
F81	Modèle felsenstein à six paramètres
G	Guanine
GTR ou (REV)	General Time Reversible
H3	Gène d'histone H3
H4	Gène d'histone H4



HKY85	Hasegawa, Kishino et Yano à six paramètres
I	Désoxyinosine
ITS	Espaceur interne transcrit (internal transcribed spacer)
ITS1	Espaceur transcrit 1 de l'ADNr nucléaire
ITS2	Espaceur transcrit 2 de l'ADNr nucléaire
JC	Jukes-Cantor
K2P	Kimura à deux paramètres
LRT	Likelihood Ratio Test
NADH	Nicotinamide adénine dinucléotide déshydrogénase (forme réduite)
NCBI	National Center for Biotechnology Information
ND1	Sous-unité 1 de la NADH déshydrogénase
ND2	Sous-unité 2 de la NADH déshydrogénase
ND4	Sous-unité 4 de la NADH déshydrogénase
ND5	Sous-unité 5 de la NADH déshydrogénase
Ng	nanogramme
NJ	Neighbor-Joining
OMS	Organisation mondiale de la santé
pb (bp)	Paire de bases (base pair)
PAUP*	Phylogenetic Analysis Using Parsimony (and Other Methods)
PCR	Réaction en chaîne de la polymérase (polymerase chain reaction)
PEPCK	Phosphoenolpyruvate carboxykinase
PGD	Phosphogluconate deshydrogénase
PhyML	Phylogenetic estimation using Maximum Likelihood
SDS	Dodécyl sulfate de sodium
T	Thymine
Tm	température de fusion ou (« melting temperature »)
TN	Tajima-Nei
TPI	Triose phosphate isomérase
UPGMA	Unweighted Pair Group Method with Arithmetic mean

# **CHAPITRE I**

## **Introduction et problématique**

## 1.1 Résumé

Lors de l'échantillonnage, vingt-deux des soixante douze espèces de simulies précédemment recensées au Québec ont été retrouvées dans cette étude. Une nouvelle espèce a été répertoriée pour le Québec, *Prosimulium magnum* et de nouvelles aires de distribution ont été trouvées pour quelques espèces. De plus, la colonisation d'un site artificiel par *Simulium aureum* dans un environnement péri-urbain a été observée. Un échantillon de 15 de ces espèces de mouches noires recouvrant cinq genres a permis de réaliser la première étude de systématique moléculaire avec les espaceurs intergéniques 1 (ITS1) et 2 (ITS2) de l'unité nucléaire de transcription de l'ADNr des Simuliidae. ITS1 fournit un outil performant en génétique des populations de simulies. Avec la structure secondaire d'ITS2 servant de guide commun pour l'alignement des séquences entre espèces, nous avons obtenu une phylogénie qui illustre de façon convaincante la divergence fondamentale entre Simuliini et Prosimuliini. Comme la résolution des relations de parenté parmi les mouches noires est particulièrement difficile, nous avons évalué trois des marqueurs d'ADN les plus courants, COI, COII et l'ADNr 18S, pour leurs capacités respectives à investiguer des niveaux distincts de profondeur évolutive à l'intérieur de la sous-famille des Nématocères. Ces données de séquences nous ont permis de résoudre certaines ambiguïtés taxinomiques, notamment pour les couples *S. pictipes*/*S. longistylatum*, et *S. annulus*/*S. euryadminiculum*, et de souligner des problèmes d'identification chez les genres *Prosimulium* et *Helodon*. De plus, des haplotypes spécifiques identifiés sur cinq positions des gènes mitochondriaux COI et COII différentient en partie les populations de *S. vittatum* selon leur provenance, en l'occurrence Québec-Ontario ou Terre-Neuve. Une étude des échantillons de *S. laciniatum* en provenance de cinq îles des Fidji démontre la nature monospécifique de l'espèce par une absence flagrante d'hétérogénéité de séquence du gène mitochondrial COII et suggère une date d'environ  $10^5$  années pour l'ancêtre immédiat de l'espèce actuelle, probablement un émigrant en provenance du Vanuatu.

## 1.2 Introduction

La classification des espèces est constituée de tous les organismes vivants divisés en groupes homogènes par catégories hiérarchiques. Nous ne pouvons que constater que les méthodes de classification ont changé et évolué au fil des ans. La méthode la plus ancienne, dite populaire, a permis de distinguer les genres et les espèces en se fondant sur des critères distinctifs comme les caractéristiques morphologiques, les mœurs, l'habitat, etc. Elle ne s'embarrasse guère des données scientifiques, mais conserve encore de nos jours une certaine importance. Par la suite, Carl Von Linné introduisit la méthode classique ou traditionnelle qui dominat jusqu'à la moitié du XX<sup>e</sup> siècle. Son plus important ouvrage, « Systema Naturae » publié en 1735, connut plusieurs éditions successives, mais ce n'est qu'en 1758 qu'il généralisa la nomenclature binominale (genre, espèce). Les espèces étaient classées selon la présence ou l'absence de facteurs biologiques, phénotypiques, anatomiques et physiologiques. Embrassant un point de vue fixiste selon lequel chaque espèce est immuable, elle est actuellement considérée comme obsolète, mais encore défendue par quelques auteurs. Par contre, elle reste importante dans la mesure où elle imprègne encore de nombreux écrits, ainsi que certains manuels scolaires. Vers 1859, Darwin publia « L'origine des espèces » et introduisit la théorie qui veut que toute évolution d'une espèce soit tributaire des modifications héréditaires subies par l'espèce dans le passé. Cette optique considère une espèce comme un point sur la flèche du temps et amène une notion de continuité. Par la suite, Hennig a introduit en 1950 les notions de classifications phylogénétiques. Depuis, la classification traditionnelle s'est vue de plus en plus remplacée par cette dernière, qui est uniquement basée sur les rapports de proximité évolutive entre espèces et la notion d'ascendance. Les termes concernés par les différentes classifications ne bénéficient pourtant pas d'une définition unanimement admise, chaque ouvrage, dictionnaire et auteur ayant la sienne. Comme l'écrivait Small en 1989 : « L'ironie est de constater que les spécialistes en classification biologique n'ont pas réussi à se doter d'une systématique claire et d'une nomenclature précise à l'intérieur de leur propre champ d'activité et de ses composants [...] » (cité dans Sytsma et Pires, 2001).

Depuis des siècles, on assiste à un remaniement constant de la classification des espèces. Certains utilisent la taxinomie classique, d'autres préfèrent la cytogénétique, et depuis quelques décennies, l'approche moléculaire semble la plus à la mode. Cependant, plusieurs questions semblent semer la controverse. Ces questions sont par ailleurs souvent passées sous silence et méritent impérativement d'être soulevées. Qu'en est-il des remaniements qui touchent les trois domaines (morphologie, cytogénétique, moléculaire)? Est-ce que les résultats de la classification seront les mêmes peu importe la méthode utilisée ou seront-ils totalement différents? Existe-t-il un consensus sur la façon de changer le statut d'une espèce, ou un chercheur (ou groupe de chercheurs) peut-il décider d'un tel changement? Avec toutes ces questions, nous en arrivons finalement à nous poser la question que peu osent aborder: qu'en est-il des critères d'admissibilité d'une espèce? Et voire même : qu'est-ce qu'une espèce?

L'espèce est un concept flou dont il existe une multitude de définitions. Selon le multi dictionnaire de la langue française, édition 2003, le terme espèce se définit comme suit : groupe d'êtres, d'éléments du même genre. Par ailleurs, le terme espèce vient de "species" (latin) qui signifie « catégorie » ou « apparence » (Campbell et Mathieu, 1995). De fait, nous apprenons à distinguer les animaux, les végétaux de par leur aspect physique ou morphologique; il s'agit de la définition morphologique de l'espèce qui, encore aujourd'hui, guide le plus souvent les classifications (Campbell et Mathieu, 1995). Ainsi, il est quelquefois difficile de discerner si un ensemble d'organismes comporte plusieurs espèces ou une seule ayant une grande variation phénotypique. De plus, deux populations impossibles à différencier d'après certains critères morphologiques pourraient en réalité être des espèces distinctes selon d'autres critères comme la cytogénétique (Campbell et Mathieu, 1995).

Il existe aussi la définition biologique de l'espèce. La plus admise est celle d'Ernst Mayr datant de 1942 : « une espèce est une population ou un ensemble de populations dont les individus peuvent effectivement ou potentiellement se reproduire entre eux et engendrer une descendance viable et féconde, dans des conditions naturelles » (Cité dans Campbell et Mathieu, 1995 et Harry, 2001). L'espèce au sens biologique s'appuie donc sur l'isolement reproducteur et non sur les critères de ressemblance morphologique.

Cependant, cette définition ne peut s'appliquer à toutes les situations, car les barrières de reproduction (ou le critère d'interfécondité) ne sont pas toujours absolues, et de nombreuses exceptions existent. Pensons à certains végétaux, protistes et mycètes, puis aux procaryotes qui ont une reproduction asexuée, des fossiles qui ne peuvent plus se reproduire sous nos yeux et des organismes dits parthénogénétiques. De plus, on ne peut appliquer la définition biologique de l'espèce si deux populations sont géographiquement séparées, car dans ce cas on ignore si les populations ont le potentiel de se croiser en nature à moins d'en faire la vérification expérimentale, même si morphologiquement elles se ressemblent au point qu'on puisse croire qu'il s'agit d'une seule espèce. Avec ces quelques définitions, on peut facilement constater que la notion d'espèce est très complexe. Au fil du temps, personne n'a réussi à énoncer des règles qui soient immuables et universellement applicables à toutes les espèces tout en tenant compte des facteurs géographiques, d'interfécondité, de spéciation, de mutation, d'évolution, etc. Darwin, dans une lettre adressée à Hooker, en 1856, soulignait déjà que l'emploi de ces critères était différent selon les auteurs : « J'étais en train de comparer plusieurs définitions de l'espèce [...] C'est vraiment amusant de voir quelles différentes idées prédominent dans les pensées des naturalistes, quand ils parlent d'espèce; pour certains la ressemblance est tout et la descendance de peu d'importance [...] - pour certains, la descendance est la clef- pour certains la stérilité est le test infaillible, avec d'autres encore, cela ne vaut rien. Je crois que tous cherchent à définir l'indéfinissable » (Harry, 2001)

### **1.2.1 Position taxinomique des simulies**

Les Simuliidae, communément appelées mouches noires ou simulies, constituent l'une des très nombreuses familles au sein de l'ordre des Diptères (sous-ordre des Nématocères, infra-ordre des Culicomorpha, superfamille des Chironomoïdes) (Crosskey, 1990). Selon le dernier relevé d'Adler et Crosskey (2009), autorités dans le domaine, cette famille comprendrait, à l'échelle mondiale, 2060 espèces actuelles distribuées entre 26 genres. Selon la cladistique traditionnelle (Currie et Adler, 2008) et les données moléculaires récentes (Moulton, 2000), il y a présentement consensus à l'effet que les simulies se subdivisent en trois sous-familles :

- les Parasimuliini, ne comprenant que quatre espèces répertoriées à l'ouest des Rocheuses et qui représenteraient, selon leurs traits primitifs (Currie et Adler, 2008) et les données moléculaires (Moulton, 2000), la branche la plus ancienne.
- les Prosimuliini, constitués d'espèces exclusivement néarctiques : six genres dont *Prosimulium* qui comprend le plus grand nombre d'espèces connues.
- Les Simuliini, incluant le genre *Simulium* qui englobe à lui seul 80% de toutes les espèces de simulies.

La famille des simulies comprend relativement peu d'espèces comparativement aux autres familles de diptères. Des types de gîtes similaires associés à un mode de développement spécialisé des stades immatures lui confèrent une grande uniformité morphologique, réduisant ainsi le nombre de caractères distinctifs permettant de séparer les espèces. C'est le cas notamment des espèces appartenant au genre *Prosimulium*, toutes semblables tant par leur morphologie que leur écologie (Peterson, 1970). L'identification à l'espèce demeure ardue et ne fait pas toujours référence précisément à l'espèce, mais à un complexe d'espèces soeurs. Ces complexes qui contiennent deux ou plusieurs espèces cytologiquement distinctes au niveau du caryotype, sont souvent morphologiquement indiscernables et de détermination particulièrement difficile. C'est le cas notamment du complexe *S. venustum/ S. verecundum* (Wolfe et Peterson, 1959). Ceci nous laisse facilement croire que les études morphologiques, bien que couramment utilisées seules, ne sont peut-être pas aussi précises qu'on le voudrait. Pour s'en convaincre, on n'a qu'à consulter les divers inventaires réalisés ces dernières années pour constater la variabilité du nombre de genres, une division taxinomique normalement immuable pour beaucoup d'autres groupes d'insectes. Cependant, chez les mouches noires, on n'a d'autre choix que de passer au départ par la morphologie avant de poursuivre au niveau cytogénétique ou moléculaire.

### 1.2.2 Inventaire des simulies du Québec

Il est d'un grand intérêt écologique d'effectuer de nouveaux inventaires pour les mouches noires du Québec, car l'entomofaune des simulies y a été très peu étudiée. Un nouvel inventaire nous permettrait non seulement de voir comment les espèces se

distribuent par rapport aux changements climatiques et environnementaux, mais aussi de mettre à jour les différents critères constituant les clés d'identifications morphologiques. Des 72 espèces retrouvées sur le territoire québécois, aucune n'est spécifique à celui-ci. Aujourd'hui, cet inventaire repose encore essentiellement sur d'anciens ouvrages datant d'avant 1980.

Les premières mentions proviennent de la liste des diptères du Québec de Winn et Beaulieu (1932) que Twinn a complétée en 1936, dans une monographie incluant la description des stades immatures des nouvelles espèces de l'est du Canada. Quelques auteurs ont décrit la morphologie des insectes dans chacune des phases de leur développement (Davies *et al.*, 1962; Grenier, 1953; Smart, 1944; Wood *et al.*, 1963), tandis que d'autres se concentrent surtout sur les stades immatures (Chance, 1970; Craig, 1975; Crosby, 1974; Ross et Craig, 1979). En effet, depuis les travaux effectués par Wolfe et Peterson (1959) à Baie-Comeau, par Mackay (1969) au Mont-Saint-Hilaire, par Back et Harper (1978, 1979) à Saint-Hippolyte-de-Kilkenny et par Laparé (1980) dans la région de la Baie James, aucune contribution importante n'a été faite à l'étude des simulies, contrairement à d'autres groupes d'insectes qui, d'une façon générale, sont mieux connus tant au point de vue faunistique qu'écologique. D'autres études se sont ajoutées depuis quelques années, mais seulement pour les provinces maritimes (Mc Creadie *et al.*, 1995; Minhas *et al.*, 2005) et l'extrême nord du Canada (Currie et Adler, 2000).

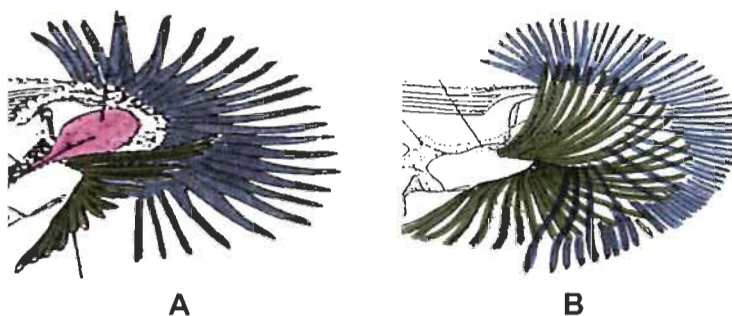
### **1.2.3 Problème associé aux critères d'identification morphologique des simulies**

Le présent travail ne s'est intéressé qu'aux larves de simulies, beaucoup plus faciles à récolter et identifier que les individus des autres stades de développement. Les critères larvaires utilisés ici sont basés sur quelques clés d'identification morphologique (Adler *et al.*, 2004; Adler et Kim, 1984; Currie, 1986; Wood *et al.*, 1963).

Tout d'abord, il est possible de faire une distinction grossière au niveau des genres à l'aide d'un seul critère : les plumeaux céphaliques. Nous sommes en présence du genre *Twinnia* quand aucun plumeau primaire n'est présent. Puis, en regardant la forme du



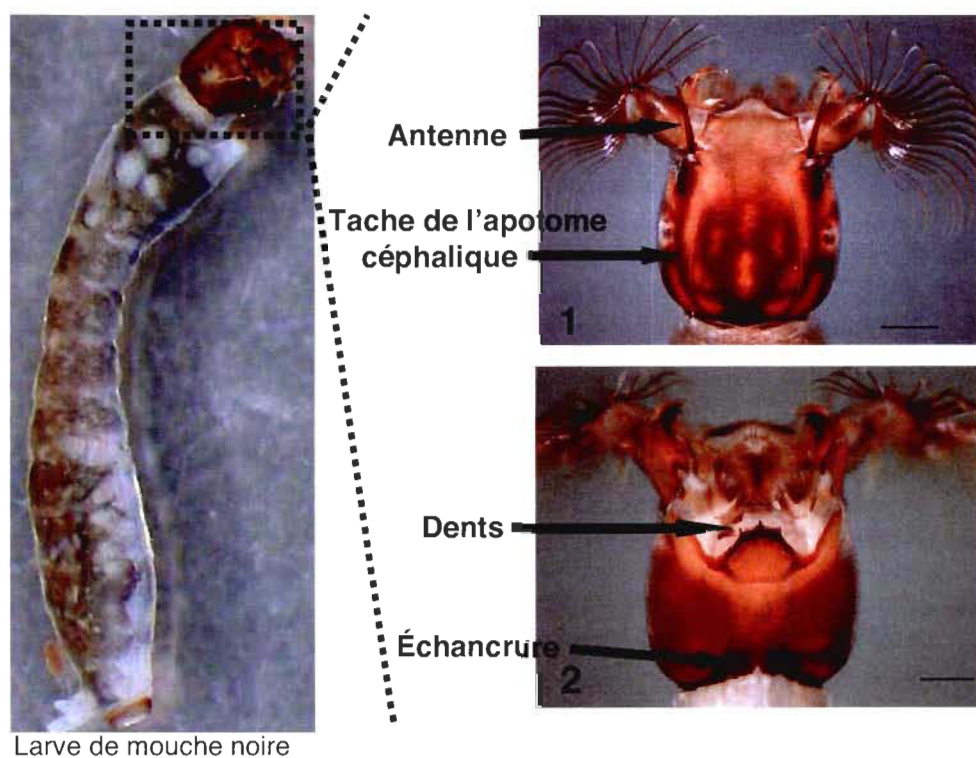
plumeau céphalique secondaire, nous pouvons différencier *Prosimulium* («triangular fan») des autres genres comme *Simulium*, *Stegopterna* et *Cnephia* («cupped fan») (Wood *et al.*, 1963) (figure 1.1).



**FIGURE 1.1** Plumeaux céphaliques primaires (en bleu) et secondaires (en vert). (A) « triangular fan » (B) « cupped fan ». (Figures tirées de Wood *et al.*, 1963).

Pour l'identification à l'espèce, les critères généralement employés sont la longueur des antennes, la distribution des taches de l'apotome céphalique et la forme de l'échancrure<sup>1</sup> (figure 1.2). Par exemple, pour les espèces appartenant au genre *Prosimulium*, lesquelles sont difficiles à séparer morphologiquement, nous avons essayé de les identifier à l'aide de la morphologie du submentum, les dents. Celle-ci sont visibles au stéréomicroscope, mais leur identification est facilitée si on les retire pour un montage histologique (Adler *et al.*, 2004). D'autres critères largement critiquables comme la couleur, le nombre de rayons au niveau des plumeaux céphaliques, le nombre de crochets du pseudopode et de la ventouse anale n'ont pas initialement été retenus, car plusieurs chercheurs (Alvan-Aguilar *et al.*, 2005; Lucas et Hunter, 1999; Zang, 2005) ont montré récemment que ceux-ci entraînaient des biais d'identification.

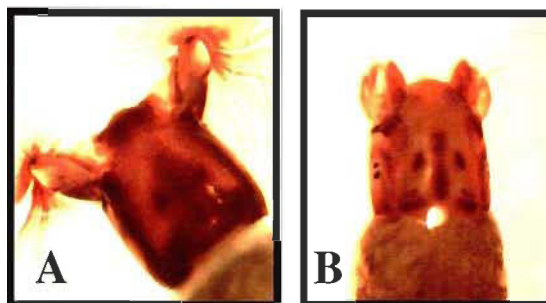
<sup>1</sup> Zone chitineuse mince et claire, située sur la face ventrale de la capsule céphalique. Cette zone chitineuse peut être réduite ou s'étendre plus ou moins loin vers l'avant et atteindre même la base du submentum.



**FIGURE 1.2** Caractéristiques morphologiques de la tête des larves utilisées pour l'identification des mouches noires. (tirée de Craig, 2004).

### 1.2.3.1 La couleur des larves et le patron des taches de l'apotome céphalique

Plusieurs facteurs font parfois varier la couleur des larves, incluant le sexe (Alvan-Aguilar *et al.*, 2005; Rothfels *et al.*, 1978) et le lieu de récolte (Alvan-Aguilar *et al.*, 2005). De plus, une étude menée par Adler et Kim (1984) sur les espèces sœurs du complexe *S. vittatum* montre que la différence entre les deux morphotypes connus faisant varier le patron foncé ou pâle des taches de l'apotome céphalique (figure 1.3) serait principalement due au type de substrat sur lequel la larve se trouve. Les larves sur les pierres sont plus foncées que celles qui vivent sur les herbages, quelque soit l'espèce, probablement à cause des différences dans leur alimentation.



**FIGURE 1.3** Distribution des taches céphaliques chez *S. vittatum*. (A) Forme foncée où les taches se rejoignent et (B) forme pâle où les taches sont séparées.

Fait encore plus intéressant, la couleur des larves varie selon le type de fixateur choisi. Bien que la pigmentation générale soit retenue, la couleur rouge peut se modifier dans le Carnoy pour atteindre des teintes grisâtres et brunâtres (Adler *et al.*, 2004). Dans l'éthanol, les larves en contact avec la végétation auront tendance à prendre une teinte verdâtre (Gaudreau, observation personnelle). Alors comment ce fait-il que certains auteurs utilisent la couleur comme critère? Donc, nous avons exclu ce critère lors de cette étude.

### 1.2.3.2 Le nombre de rayons primaires sur les plumeaux céphaliques

En ce qui concerne les plumeaux céphaliques, il est clairement démontré que le nombre de rayons primaires (figure 1.1) sur l'éventail du labre chez les larves de *S. rostratum*, *S. decorum* (Lucas et Hunter, 1999) et *S. noelleri* (Zhang, 2005) diminue lorsque la quantité de nourriture augmente. De plus, Zhang (2005) explique que dans un habitat combinant courant rapide et grande quantité de nourriture, les larves développent des petits plumeaux avec peu de rayons tandis que dans la combinaison faible courant/peu de nourriture, les larves développent de grands plumeaux comportant beaucoup de rayons (Palmer et Craig, 2000). Tout compte fait, l'utilisation du nombre de rayons dans la diagnose des espèces requiert de la prudence puisqu'il s'agit d'un caractère plastique qui varie en fonction de la charge du seston (Lucas et Hunter, 1999; Zhang, 2005) et de la vitesse du courant (Zhang, 2005). Les simules d'une même

espèce mais provenant de d'habitats variés peuvent différer dans la structure de leurs plumeaux (Palmer et Craig, 2000; Zhang et Malmqvist, 1996).

### **1.2.3.3 Autres problèmes rencontrés lors de l'identification morphologique**

Certains des autres problèmes taxinomiques rencontrés sont purement techniques, tandis que d'autres révèlent des ambiguïtés fondamentales quant à l'identification d'une espèce. Dans un premier temps, l'incohérence des critères morphologiques utilisés entre les clés anciennes et nouvelles rend difficile la correspondance entre les espèces reconnues hier et celles d'aujourd'hui. Les noms ayant à l'occasion changés plus d'une fois en cours de route au gré des auteurs, nous sommes souvent confrontés à des termes synonymes. Les clés morphologiques sont relativement anciennes, donc difficilement accessibles, sauf pour celle d'Adler *et al.* (2004), qui est beaucoup plus récente. Cette dernière est difficilement utilisable pour l'identification des larves, puisqu'on y retrouve pêle-mêle des critères géographiques nettement tranchés (longitude) associés à des critères morphologiques grossiers ou très pointus requérant des dissections fines (extraction et extension des filaments respiratoires) obligeant parfois des montages histologiques. Finalement, on peut remarquer à travers ces exemples et d'autres cas décrits plus loin que l'identification morphologique peut s'avérer très périlleuse si elle est prise à la légère. L'utilisation de plusieurs clés nous a donc permis non seulement de soulever ces incongruités, mais aussi d'accroître la crédibilité de nos identifications.

### **1.2.4 Cytogénétique des simulies**

Les chromosomes polytènes sont présents dans les cellules des glandes salivaires des larves de la plupart des diptères. Ils correspondent à un certain nombre de copies des chromatides (jusqu'à 1000) qui sont restées soudées entre elles. La coloration de Feulgen rend visible au microscope photonique des bandes fortement colorées alternant avec des bandes plus claires. Ce patron de bandes pâles et foncées permet au cytologiste de reconnaître les réarrangements chromosomiques (Adler et McCreadie, 1997). Chez les mouches noires, les travaux cytologiques initiés par Dunbar en 1953 ont été

poursuivis par de nombreux auteurs dans le but de différencier des espèces jumelles là où la morphologie n'y parvenait généralement pas.

Vu les implications médicales, les principaux travaux en cytogénétique furent réalisés sur *S. damnosum*, vecteur de l'onchocercose en Afrique (Rodhain et Perez, 1985). Au départ, la vaste distribution géographique de *S. damnosum* laissait soupçonner l'existence d'un complexe d'espèces morphologiquement indifférenciées, mais écologiquement différentes. L'étude des chromosomes géants des glandes séricigènes des larves permit à Dunbar, en 1966, d'entreprendre la cytotaxinomie du complexe *S. damnosum* (cité dans Quillévéré, 1979), où il identifia quatre types cytologiques différents pour les larves en provenance de l'Ouganda. En 1975, Vajime et Dunbar élèvent à huit le nombre de cytotypes ouest-africains au rang d'espèces (cité dans Quillévéré, 1979). Dès lors les résultats sont controversés : « Nous devons regretter que ces descriptions d'espèces aient été faites sans tenir compte ni d'aucun caractère morphologique ni du contexte écologique ou épidémiologique » (cité dans Quillévéré, 1979). Cette pratique semble courante, car Bedo (1975) publia une étude sur trois espèces du groupe *S. pictipes* et mentionna explicitement qu'aucun caractère morphologique n'avait été étudié. En 1976, lors d'une réunion internationale sur les complexes d'espèces vecteurs de maladie, on y décida de conserver les noms des espèces déjà publiés comme l'exigeait le Code International de Nomenclature Zoologique; cependant, aucun nouveau nom d'espèces ne pouvait être publié pour un cytotype sans que celui-ci ne fasse l'objet d'une étude morphologique ou bioécologique complémentaire (Quillévéré, 1979). En 1978, on reconnaissait dans le complexe *S. damnosum* au moins 25 formes taxinomiques que l'on ne peut distinguer avec certitude qu'en fonction de leurs caractéristiques chromosomiques; ce complexe en comportait au moins 40 en 1997 (Adler et McCreadie, 1997; Crosskey et Howard, 1997), 50 en 2004 (Crosskey et Howard, 2004) et 57 en 2007 (Post *et al.*, 2007). Par contre, ce ne sont pas tous les cytotypes qui permettent le développement du parasite et transmettent l'onchocercose (Day *et al.*, 2008). Un bulletin de l'OMS<sup>2</sup> paru en 1978 indique que la plus grande partie des travaux sur les mouches noires a été consacrée au

---

<sup>2</sup> Complexes d'espèces chez les Simuliidae. Bulletin de l'Organisation mondiale de la santé. 56 : 169-178.

complexe *S. damnosum*. D'ailleurs, on mentionne d'ores et déjà que la priorité devrait être donnée à des études morphologiques complètes, et ce pour chacun des stades biologiques de tous les membres du complexe tels qu'ils sont couramment définis à partir des investigations cytotoxonomiques sur les larves. De plus, on souligne aussi qu'« il faudrait continuer à examiner la possibilité de préparer et d'identifier les chromosomes polytènes des femelles adultes de ce complexe ». Quoiqu'il en soit, aujourd'hui, la cytotoxonomie s'applique presque uniquement aux larves des derniers stades (Post *et al.*, 2007; Rodriguez-Perez *et al.*, 2006; Thanwisai *et al.*, 2006; Xiong et Kocher, 1991). Les nymphes et les adultes ont aussi des chromosomes polytènes, mais de qualité généralement insuffisante pour la caryotypie (Morales-Hojas *et al.*, 2002). Cependant, quelques succès furent obtenus avec les chromosomes provenant des tubes de Malpighi (Bedo, 1976) de nymphes et d'adultes de mouches noires provenant d'Australie de même que chez certaines femelles du complexe *S. damnosum*. Cela permettrait l'identification des mouches femelles adultes, importantes épidémiologiquement (Krüger *et al.*, 2000). Par ailleurs, la cytogénétique, bien que capable de différencier deux espèces morphologiquement semblables, engendre encore plusieurs erreurs d'identification. C'est le cas notamment du sous-complexe Ketaketa de *S. damnosum* où il y a eu de mauvaises descriptions au sein des inversions chromosomiques (Krueger, 2006). De plus, bien que la définition d'espèce se base sur le critère de l'interfécondité/interstérilité, il faut savoir que certains cytotypes peuvent toutefois se croiser entre eux (Morales-Hojas et Krueger, 2009; Shields *et al.*, 2009).

Il est important de préciser ici que, dans le cadre de la présente étude, des analyses cytologiques n'ont pas été effectuées, malgré que plusieurs espèces du Québec soient considérées comme formant un complexe (tableau 2.3). Nous avons choisi de privilégier les analyses moléculaires, pour lesquelles on n'utilise pas le même fixateur et dont le taux de succès est plus élevé que celui de la caryotypie. (Spironello *et al.*, 2002). De plus, il était aussi impossible de suivre une formation afin de réaliser et comprendre la lecture des cartes chromosomiques. Cette expertise ne s'acquiert pas de façon magistrale, mais plutôt par le biais d'échange de connaissances entre chercheurs. Dans notre cas, les chercheurs approchés ont décliné nos offres.

### 1.2.5 Vers une taxinomie moléculaire ?

L'identification d'espèces par des critères moléculaires est une technique que plusieurs biologistes croient vouée à la réussite, tandis que d'autres la trouvent douteuse et pleine d'embûches. La justification de cette approche est que les génomes évoluent avec les espèces qui les portent et que les séquences d'ADN peuvent par conséquent identifier des spécimens ou distinguer des espèces très semblables. L'avènement du « barcoding » basé sur la séquence du gène COI suggère même que chaque espèce pourrait posséder sa propre clé d'identification moléculaire stockée dans une banque de données idéalement accessible à tous. Il ne faut cependant pas délaissier pour autant l'identification morphologique des taxinomistes au profit de l'information obtenue par l'ADN, car bon nombre de problèmes taxinomiques ne peuvent être résolus que par l'utilisation combinée de l'ADN, de la morphologie et de la cytogénétique.

Actuellement la reconnaissance des mouches noires par une identification morphologique requiert des clés complexes pour chacun des stades (oeuf, larve, nymphe et adulte) et nécessite très souvent la compétence de spécialistes capables d'utiliser à la fois la morphologie et la cytogénétique. Toutefois, même chez les spécialistes, le pourcentage d'erreur peut s'élever jusqu'à 9% (Carew *et al.*, 2003). Par contre, l'apprentissage des techniques moléculaires requiert beaucoup moins de temps que la formation de la part d'un taxinomiste. Ces techniques permettent aussi une identification fiable de tous les stades de développement (Tang *et al.*, 1996, 1998), chez les deux sexes, des espèces soeurs et des spécimens endommagés (Carew *et al.*, 2003). Pour ces raisons, plusieurs laboratoires de recherche ont commencé à accumuler des données en biologie moléculaire dans le but de simplifier l'identification au sein de certaines familles d'insectes afin de rendre disponible des outils d'identification rapides et fiables.

Chez les mouches noires, l'essor des techniques moléculaires est survenu dans les années 90, suite à quelques écrits précurseurs des années 70-80 (Adler *et al.*, 2004; Snyder et Linton, 1983; Sohn *et al.*, 1975; Teshima, 1972). Depuis l'article de Xiong et Kocher (1991), plusieurs études ont suivi, mais la majorité des travaux ne porte encore que sur certains groupes d'espèces comme le complexe *S. damnosum*. De plus, peu

d'études à grande échelle comportant plusieurs espèces et marqueurs, ont été réalisées; on parle plutôt d'échantillonnages basés sur un endroit à un temps donné. À l'exception de Bertone *et al.* (2008), Moulton (2003), Gaudreau *et al.* (2010a, b), LaRue *et al.* (2009) et Phayahasena *et al.* (2010), l'ensemble des études ne font usage que d'un ou deux marqueurs. Les régions les plus séquencées chez les insectes sont l'ADN mitochondrial et l'ADNr nucléaire (Caterino *et al.*, 2000). Pour les génomes mitochondriaux, COI (Day *et al.*, 2008; Finn et Adler, 2006; Pramual *et al.*, 2005; Rivera et Currie, 2009), COII (Pruess *et al.*, 2000; Tang *et al.*, 1998), les ADNr 16S (Krüger *et al.*, 2000; Tang *et al.*, 1996, 1998; Xiong et Kocher, 1991, 1993) et 12S (Moulton, 2003; Tang *et al.*, 1996, 1998) sont les plus fréquemment utilisés. D'autres marqueurs tels COIII, ND5, Cytb, ND2, ND4 (Rodriguez-Pérez *et al.*, 2006) et ND1 (Simon *et al.*, 1994) ont été utilisés à un degré moindre. Les autres représentants du génome mitochondrial ont plutôt été ignorés (Caterino *et al.*, 2000). L'ADNr est le marqueur nucléaire le plus utilisé pour les mouches noires (Brockhouse *et al.*, 1993; Krüger *et al.*, 2000; Krueger, 2006; Morales-Hojas *et al.*, 2002; Moulton, 2003; Rodriguez-Pérez *et al.*, 2006; Tang *et al.*, 1996, 1998; Thanwisai *et al.*, 2006). Parmi les autres marqueurs occasionnellement utilisés, on note CAD, TPI et PGD (Bertone *et al.*, 2008), PEPCK et DDC (Moulton, 2003), EFa1 (Caterino *et al.*, 2000; Moulton, 2003) et l'espaceur des gènes des histones H3 et H4 (Gaudreau, 2005).

Un questionnement se pose quant à l'utilité des marqueurs utilisés. Les différentes régions de l'ADN des organismes et les protéines qu'elles codent n'évoluent pas toutes à la même vitesse (Harry, 2001). En effet, certains gènes sont fortement contraints, car ils assurent des fonctions très importantes pour les organismes, et des mutations sur ces gènes limitent drastiquement la viabilité de leurs porteurs. À l'inverse, certaines régions évoluent plus rapidement, ce qui permet d'étudier des relations évolutives à faible échelle, par exemple au sein d'une population ou entre espèces proches (Day *et al.*, 2008; Krueger *et al.*, 2006; Schlötterer *et al.*, 1994). Pour étudier les relations entre tous les organismes vivants, des marqueurs tels que l'ARNr 12S et 16S (Harry, 2001) ou 18S et 28S sont généralement utilisés.



Lorsqu'on établit des phylogénies moléculaires, il faut utiliser une molécule dont l'évolution est en adéquation avec les niveaux de différenciation des taxons étudiés. Selon le gène, l'horloge moléculaire varie et les substitutions s'accumulent au cours du temps plus ou moins rapidement en fonction des pressions de sélection. Les gènes évoluant lentement permettront ainsi d'établir une phylogénie au niveau des familles; au contraire, on se servira de marqueurs plus variables, comme les gènes mitochondriaux pour étudier des genres proches, des espèces du même genre ou des populations conspécifiques (Harry, 2001; Hwang et Kim, 1999). De plus, l'utilisation de plusieurs gènes peut être utile pour valider une phylogénie (Harry, 2001). Par ailleurs, différents facteurs peuvent conduire à des conflits phylogénétiques entre données morphologiques et moléculaires sur les mêmes taxons étudiés. Ces conflits peuvent être dus à des vitesses d'évolution différentes entre les caractères morphologiques et les sites moléculaires, au nombre de caractères ou à la méthode de reconstruction phylogénétique. De plus, si des phylogénies contradictoires sont obtenues, n'oublions pas que tous les gènes ne retracent pas la même histoire évolutive, comme les gènes mitochondriaux exclusifs de la lignée maternelle (Tang *et al.*, 1996).

### **1.2.6 Problématique et objectifs**

Avant nos travaux, il n'existait pas d'inventaire récent des similies du Québec et aucune étude systématique de taxinomie moléculaire des espèces d'Amérique du Nord n'était disponible.

Dans le but d'établir pour les espèces du Québec un code barre simplifié ne faisant pas appel au séquençage, mes travaux de maîtrise ont porté sur la caractérisation de la taille de l'espaceur ITS1, de l'espaceur entre les gènes d'histones H3 et H4, ainsi que des gènes mitochondriaux de COII et des ARNr 12S et 16S. L'étude fut complétée, dans certains cas, par l'analyse des profils de restriction et des hétéroduplexes. Bien que s'en approchant, les résultats n'ont pas permis d'obtenir un code barre distinctif de chaque espèce, ce qui indiquait qu'il fallait aller jusqu'au niveau de la séquence nucléotidique.

Dans le cadre du doctorat, l'objectif global est d'utiliser les outils moléculaires pour résoudre certains problèmes taxinomiques et phylogénétiques chez les simulies. Au niveau des objectifs spécifiques, nous avons choisi de faire une série d'étude de cas à partir de notre banque de spécimens (chapitre II, V et VI) de façon à couvrir la progression suivante.

- Études de populations d'une même espèce où deux cas seront examinés :

Le cas de *Simulium laciniatum* (chapite VI) nous semblait intéressant, car il nous permettait d'étudier la diversité génétique dans le cadre de l'isolement insulaire. Ce cas nous a été soumis par le Dr Douglas A. Craig qui nous a fourni les spécimens et qui est co-auteur de l'article.

Le cas de *Simulium vittatum* (chapitre V) représente une progression par rapport à *S. laciniatum*. On y retrouve encore le thème de l'isolement insulaire, mais moins marqué puisque *S. vittatum* se retrouve non seulement à Terre-Neuve, mais aussi sur le vaste continent Nord Américain. Les spécimens larvaires provenant de Québec et d'Ottawa ont été fournis par nous et ceux de Terre-Neuve par le Dr Murray Colbo.

- Au niveau suivant, nous nous servirons des outils moléculaires afin de faire une distinction entre espèce proches. Trois cas seront analysés : *S. annulus*/*S. euryadminiculum*, *S. pictipes*/*S. longistylatum* et les genres *Prosimulium*/*Helodon*.

La littérature (Adler *et al.*, 2004) laissait déjà soupçonner que *S. annulus*/*S. euryadminiculum* ne constituaient qu'une seule espèce. Nous avons vérifié se point en utilisant divers marqueurs moléculaires (COI, COII, ITS1, ITS2 et 18S; voir chapitre IV).

Des différences subtiles semblaient capables de distinguer *S. pictipes* de *S. longistylatum*. À partir de tous les marqueurs précédents, nous avons pu

confirmer qu'il s'agit bien de deux espèces distinctes, mais ayant récemment divergé.

Les genres *Prosimulium* et *Helodon* présente peu de diversité morphologique. Nous avons comparé l'identification morphologique basée sur les dents hypostomiales au données moléculaires (COI, COII, ITS1) afin de voir la concordance entre les deux systèmes de classification.

- Comme dernier objectif nous nous sommes servi des outils moléculaires pour poser les bases d'une classification rationnelle basée sur la phylogénie.

### **1.3 Méthodologie**

Plusieurs procédures furent utilisées surtout en ce qui a trait aux expérimentations moléculaires. Nous n'en ferons pas état si elles sont déjà décrites dans les différents articles (chapitre II à VI).

#### **1.3.1 Constitution de la banque de spécimens et inventaire**

Pour constituer la banque d'espèces, un prélèvement sur le terrain à chacune des stations prospectées a été effectué. La banque avait pour but de mettre à jour et d'améliorer l'inventaire faunistique des simuliés du Québec et d'obtenir un large éventail de larves d'espèces distinctes pour la réalisation du volet moléculaire. Pour ce faire, nous avons sillonné les routes du Québec afin d'échantillonner les cours d'eau bordant celles-ci (figure 1.4), incluant une étude exhaustive de deux décharges de la région Mauricienne (Chapitre II). Les préférences écologiques des stades préimaginaux résident presque toujours dans les eaux douces courantes (insectes rhéophiles et oxyphiles), qu'elles soient permanentes ou temporaires; les gîtes larvaires sont souvent localisés au niveau des cascades, rapides et radiers ou sur les pentes des barrages. Les récoltes de larves ont été faites principalement sur des rameaux d'arbres ou d'arbustes exposés au courant et servant de support à la faune. Les larves recueillies ont été

directement fixées dans l'éthanol 95% et conservées à 4°C pour ensuite être identifiées à l'espèce (Fukatsu, 1999; Koch *et al.*, 1998; Post *et al.*, 1993; Wood *et al.*, 1963).

Le type de fixateur utilisé dépend du type d'analyse que l'on veut faire. Pour l'identification basée strictement sur la morphologie, les spécimens peuvent être placés dans de l'éthanol 80-95%<sup>3</sup>. Cependant, pour extraire un maximum d'informations morphologiques (extension des plumeaux céphaliques et des processus rectaux, puis conservation de certaines pigmentations), Adler *et al.* (2004) recommande d'utiliser du Carnoy<sup>4</sup>, tout comme pour les spécimens destinés aux analyses cytologiques. Pour une meilleure analyse cytologique on conseille un millilitre de Carnoy par larve. Même en prenant le meilleur fixateur, la qualité des chromosomes varie beaucoup. La température de l'eau plus chaude durant l'été et une faible qualité de nourriture produisent souvent des caryotypes peu exploitables (Adler *et al.*, 2004; Alvan-Aguilar *et al.*, 2005; Quillévéré, 1979). Les spécimens destinés aux analyses moléculaires doivent être sacrifiés dans l'éthanol 95%, car l'acide acétique du Carnoy dégrade l'ADN, rendant difficile son amplification. Aucun fixateur ne peut fournir de spécimens qui soient à la fois morphologiquement, cytologiquement et moléculairement adéquats. La meilleure façon à ce jour demeure de couper la larve en trois morceaux : la tête dans l'éthanol 80% ou le Carnoy pour l'étude morphologique, le thorax et les quatre premiers segments abdominaux dans l'éthanol absolu pour l'analyse moléculaire et la portion postérieure dans le Carnoy pour l'analyse chromosomique (Castillo, 2008).

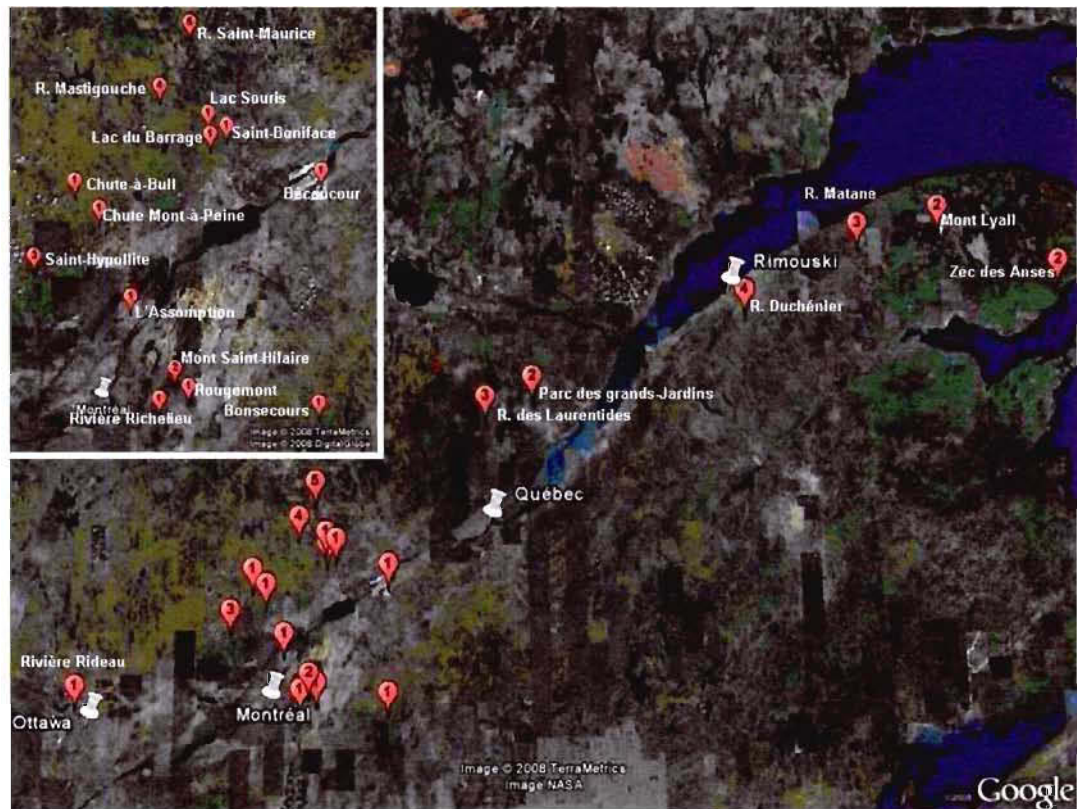
Des similies ont été récoltées entre 2003 et 2008 dans plusieurs cours d'eau du sud-est du Québec. Les larves récoltées des 24 aires d'échantillonnage (figure 1.4) proviennent d'une grande variété de cours d'eau depuis les ruisselets jusqu'aux grandes rivières. Nous avons répertorié 22 des 72 espèces reconnues au Québec à ce jour

---

<sup>3</sup> Pour l'obtention de spécimens parfaits, on recommande l'utilisation d'éthanol 80 à 95%. L'alcool à 70% s'est avéré peu satisfaisant, vu que l'eau prise avec le matériel produit une sur-dilution, ce qui entraîne la décomposition partielle du spécimen. De plus, les processus rectaux ne sont pas évaginés (Belquat et Dakki, 2004).

<sup>4</sup> Le Carnoy est un mélange : ¼ d'acide acétique glacial et ¾ d'éthanol (95-100%).

(Gaudreau et Charpentier, 2010) (chapitre 2). Elles appartiennent aux genres *Twinnia*, *Prosimulium*, *Helodon*, *Metacnephia*, *Stegopterna*<sup>5</sup>, *Cnephia*<sup>5</sup> et *Simulium*.

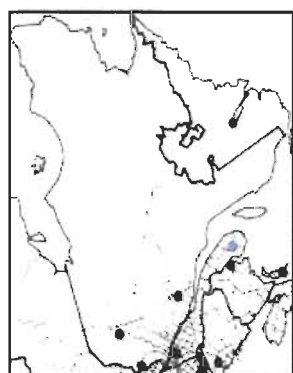


**FIGURE 1.4 :** Lieux d'échantillonnage (Gouttes rouges : lieu d'échantillonnage, le chiffre dans la goutte représente le nombre de site à chaque lieu; punaises blanches : ville).

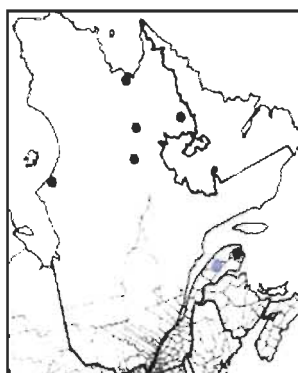
Notre échantillonnage, qui exclut les espèces des zones typiquement arctiques du Québec, n'est évidemment pas aussi exhaustif que la liste dressée par divers chercheurs au cours des dernières décennies. En plus de la découverte d'un membre du complexe *Prosimulium magnum*, qui n'avait pas encore été répertorié au Québec, plusieurs nouvelles aires de colonisation ont été relevées pour la majorité des autres espèces (Gaudreau et Charpentier, 2010) (figure 1.5). Ces nouvelles aires devront être ajoutées au recensement qui a paru sous forme de cartes dans Adler *et al.* (2004). Par ailleurs,

<sup>5</sup> Ici le genre *Cnephia* englobe l'espèce *Cnephia dacotensis*. *Cnephia mutata* a été renommé *Stegopterna mutata* par certains auteurs.

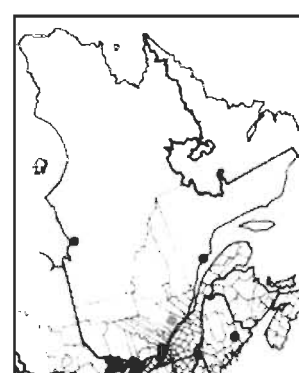
nous avons découvert qu'une espèce, *S. aureum*, pouvait aussi coloniser des sites artificiels, tels des jardins d'eau en milieu péri-urbain, ce qui n'avait jamais été mentionné auparavant dans la littérature.



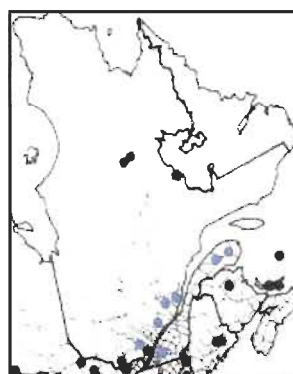
A



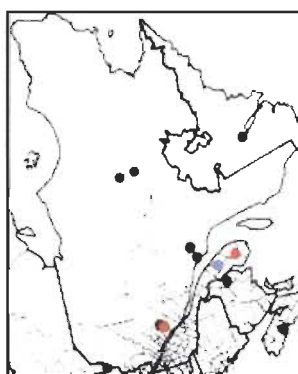
B



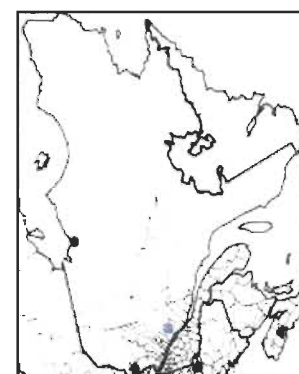
C



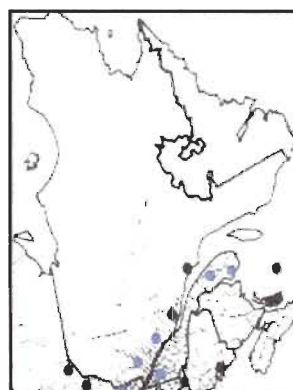
D



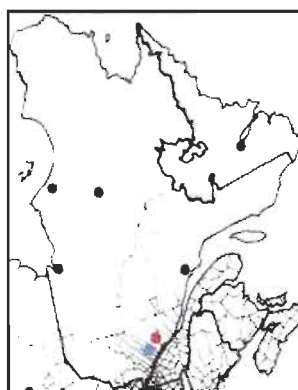
E



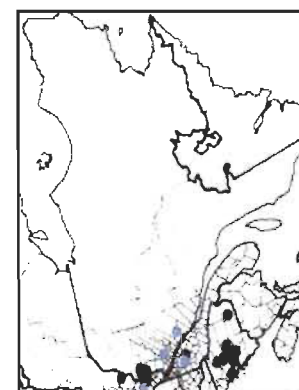
F



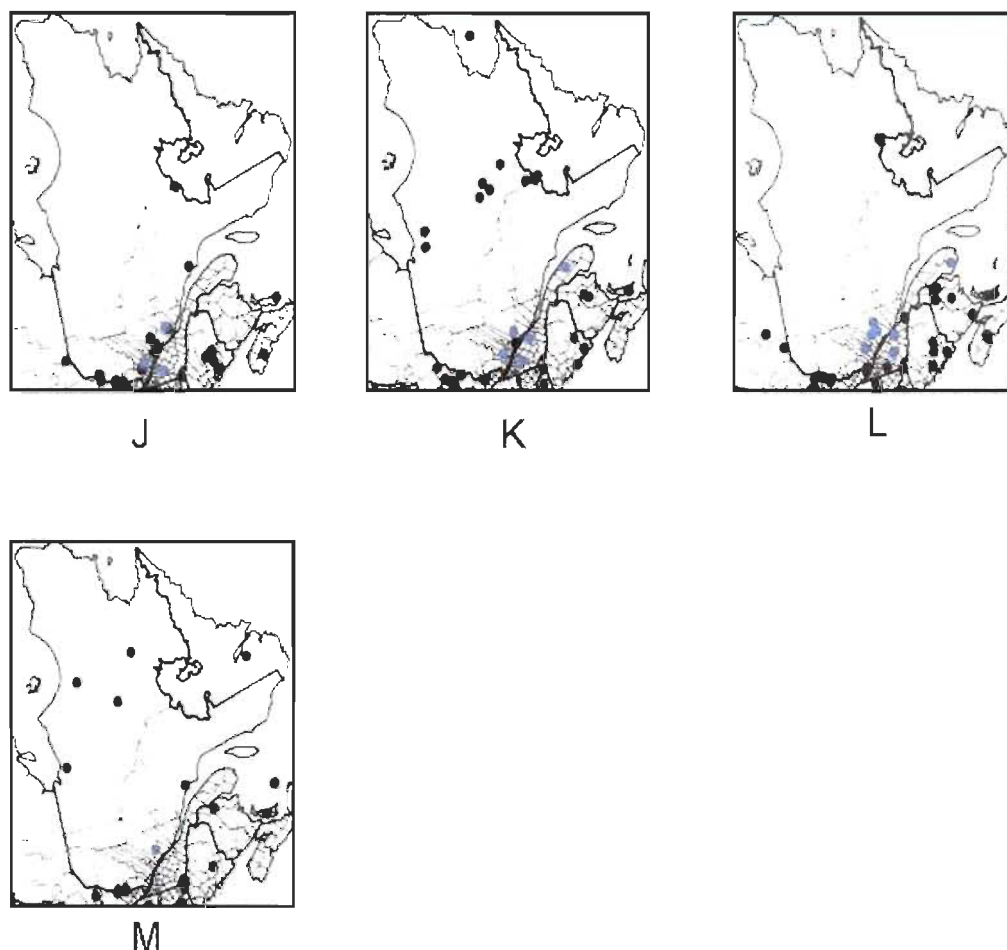
G



H



I



**FIGURE 1.5 :** Indication des nouvelles aires de colonisation (points de couleur bleu et rouge) pour les simuliés du Québec représentées sur les cartes provenant d'Adler *et al.* (2004). A à M dans l'ordre : *T. tibblesi*, *M. saileri*, *C. dacotensis*, *S. mutata*, *S. annulus* et *S. euryadminiculum* (rouge), *S. quebecense*, *S. vittatum*, *S. pictipes* et *S. longistylatum* (rouge), *S. jenningsi*, *S. decorum*, *S. tuberosum*, *S. venustum* et *S. sylvestre*.

### 1.3.2 L'identification morphologique

Avant de procéder à l'identification morphologique proprement dite, nous avons d'abord éliminé les larves parasitées, car l'ADN du parasite peut contaminer celui extrait de la simule (St-Onge *et al.*, 2008). On observe deux formes de parasites chez les larves, soient les mermithides (Sharp et Hunter, 2008) et les microsporidies (Adler *et al.*, 2004). Les mermithides sont visibles à travers la cuticule; ils se présentent enroulés sur

eux-mêmes dans la cavité abdominale à la base des glandes salivaires, tandis que les microsporidies sont des petits kystes blanchâtres (figure 1.6). Le niveau d'infestation varie d'une larve à l'autre, et les parasites envahissent souvent toute la cavité coelomique des larves, parfois jusque sous la capsule céphalique. La plupart des larves parasitées deviennent difformes et décolorées; de plus, leur taille excède quelquefois celle des larves saines à maturité. (Laparé, 1980; Maurand, 1973).



**FIGURE 1.6 :** Larves parasitées par un mermithide (gauche) et des microsporidies (droite) (Laparé, 1980).

L'identification morphologique à l'espèce des larves a été faite sous loupe stéréoscopique à l'aide de quatre clés d'identification morphologique (Adler *et al.*, 2004; Adler et Kim, 1984; Currie, 1986; Wood *et al.*, 1963). De plus, pour *Prosimulium*, les dents ont été disséquées et montées entre lame et lamelle dans une résine commerciale (Aquamount montant; BDH Chemicals Ltd, Poole England) afin d'être visualisées sous microscope à l'huile à immersion et utilisées dans le processus d'identification avec la clé d'Adler *et al.* (2004).

### 1.3.3 Méthodes d'identification moléculaire

Pour la réalisation de ce projet, différentes techniques ont été mises à profit pour typer 21 des espèces inventoriées.

#### 1.3.3.1 Extraction d'ADN

Chaque larve était réhydratée et broyée individuellement dans un tampon d'extraction. Le broyat était incubé une heure à 65°C avant que l'on ajoute du CH<sub>3</sub>COOK, qui par excès d'ions K<sup>+</sup> fera précipiter le détergent SDS contenu dans le



tampon. Le tout est transféré sur glace 30 minutes, puis centrifugé pour récupérer le surnageant contenant l'ADN. L'ADN est ensuite purifié avec la trousse « Wizard DNA Clean-up System » (Promega) dans lesquels il est mélangé à une suspension de résine de silice à haute force ionique. Le complexe résine-ADN est passé dans une mini-colonne munie d'un filtre et lavé par de l'isopropanol 80%, ce qui enlève toute trace de sel et d'impuretés. L'ADN est élué de la mini-colonne avec de l'eau stérile (rendement moyen de 200 ng d'ADN par larve).

### 1.3.3.2 Création des amorces

Au départ, plusieurs amorces provenant directement de la littérature furent utilisées pour amplifier les différents marqueurs choisis. Dans le cas de COI, ces amorces, bien que dites universelles, fonctionnaient de façon erratique. À partir de séquences fragmentaires obtenues de quelques simules, nous avons conçu deux amorces donnant pleine satisfaction dans la plupart des cas. Pour ITS1, les amorces originellement décrites par Tang *et al.* (1998) avaient le défaut d'amplifier en plus environ 600 pb correspondant à la fin de l'ARNr 18S. Nous les avons remplacées par deux amorces ciblant beaucoup plus étroitement ITS1 (chapitre 6). S'inspirant d'amorces utilisées par Thanwisai *et al.* (2006), nous avons créé similairement pour ITS2 de nouvelles amorces jugées ici plus efficaces (chapitre 3). Pour la liste complète des amorces utilisées, voir le tableau 1.1.

Les nouvelles amorces se devaient de respecter les critères usuels, à savoir un contenu en G et C d'environ 50% et des Tm quasi identiques pour la paire spécifique d'une cible donnée. Afin de s'assurer d'une hybridation parfaite à l'extrémité 3' de l'amorce pour une séquence codante, celle-ci devait préférentiellement se terminer sur un codon de méthionine (ATG) ou de tryptophane (TGG), ou sur une séquence  $G/C-G/C$  représentant les deux premières bases des codons d'un acide aminé conservé. Quant aux codons internes à l'amorce, nous avons parfois introduit sur leur 3<sup>e</sup> position une dégénérescence partielle ( $G/A$  ou  $C/T$ ) ou totale, la désoxyinosine (I), servant de base indifférente capable d'accepter comme vis-à-vis n'importe quel base partenaire.

**TABLEAU 1.1**  
Liste des amorces

Séquence, 5' vers 3' <sup>1</sup>	Tm <sup>2</sup>	Cible <sup>3</sup>	Hybridation <sup>4</sup>
CCGTAGGTGAACCTGCGGAAGG	72°	ITS1-F	18S <sup>5</sup> / 1954
TTAGCTGCGGTCTTCATCGACCC	72°	ITS1-R	5.8S / 30
TTGAACGCATATCGCACTTCTTGC	70°	ITS2-F	5.8S / 121
CCTTATTAATATGCTTAAATTCAGGG	68°	ITS2-R	28S <sup>5</sup> / 25
CTGGTTGATCCTGCCAGTAG	62°	18S-F	18S <sup>5</sup> / 23
GTGGTGCCCTTCCGTCAATTCC	70°	18S-R	18S <sup>5</sup> / 1214
GGATCGAAGGCGATTAGATAC	62°	18S-F	18S <sup>5</sup> / 1083
CTCCGCAGGTTACCTACG	64°	18S-R	18S <sup>5</sup> / 1936
CATCAATTGATACTGAAGTTATGA	60°	COII-F	COII <sup>6</sup> / 326
CAATTGGTATAAACTATGATTTGC	64°	COII-R	COII <sup>6</sup> / 604
ATAATTCNCATATTATTAGHCAAGARAG	72°	COI-F	COI <sup>6</sup> / 771
AGATTTTGGTATCATTCAATAGATGATT	72°	COI-R	COI <sup>6</sup> / 1451

<sup>1</sup> N: mélange des 4 bases. H: A/T/C. R: A/G.

<sup>2</sup> Calculé selon  $Tm = 2(\text{nombre de A} + \text{T}) + 4(\text{nombre de G} + \text{C})$ .

<sup>3</sup> -F (-R) : amorce d'orientation sens (antisens).

<sup>4</sup> Gène sur lequel l'amorce s'hybride / position 3' terminale de l'amorce sur le gène (orientation sens).

<sup>5</sup> Par rapport à la séquence de *S. vittatum* (U48383).

<sup>6</sup> Par rapport à la séquence de *C. dacotensis* (AF425841).

Les sites d'hybridation des amorces ont été choisis à partir de considérations évolutives, à savoir qu'une séquence conservée chez des espèces de moins en moins apparentées aux simulies avait toutes les chances de constituer une cible adéquate chez celles-ci. Les espèces utilisées provenant de GenBank et de la littérature sont *S. damnosum* (Diptera : Simuliidae), *Drosophila melanogaster* (Diptera), *Apis mellifera* (Hymemoptera) et *Caenorhabditis elegans* (Nematoda). La validité de la conception de nos amorces fut vérifiée par expérimentation sur *D. melanogaster*. Dans tous les cas, la taille obtenue pour l'amplicon correspondait à celle prévue et sa séquence à celle répertoriée dans GenBank (NCBI<sup>6</sup>). Le contrôle effectué chez la drosophile et nos simulies donne des résultats positifs et reproductibles, tant au niveau de la taille de l'amplicon que des données de séquençage.

<sup>6</sup> National Center for Biotechnology Information, Bethesda, MD, USA; <http://www.ncbi.nlm.nih.gov>

### 1.3.3.3 Amplification PCR

La réaction en chaîne de la polymérase (PCR) est la deuxième étape menant à l'identification moléculaire des similies. Chaque série de PCR comprend un contrôle négatif (eau stérile) pour s'assurer de l'absence de contamination par de l'ADN exogène lors des manipulations. Les conditions d'amplification particulières à chaque cible sont mentionnées dans les différents articles.

### 1.3.3.4 Séquençage

Avant séquençage, on purifie l'ADN amplifié par migration sur gel d'agarose ou de polyacrylamide, puis on extrait la bande d'ADN du gel à l'aide de la trousse QIAquick Gel extraction Kit (QIAGEN). Pour les gels d'agarose, le protocole fourni dans la trousse fut utilisé; pour les gels de polyacrylamide, il fallait suivre un protocole expérimental disponible sur Internet<sup>7</sup>. Avant l'envoi au séquençage, on effectue toujours deux contrôles. D'abord, sous peine d'être exclu, chaque amplicon doit avoir la taille attendue telle qu'estimée par rapport à un standard approprié. Ensuite, la quantité d'ADN est évaluée par comparaison visuelle avec une quantité connue d'ADN du bactériophage  $\lambda$  digéré avec *Hind*III et *Eco*R1. Tous les amplicons furent séquencés à l'Université Laval. Au retour du séquençage, l'analyse des résultats fut réalisée à l'aide d'outils bioinformatiques.

Afin de pouvoir tirer profit de nos données de séquençage et d'obtenir des résultats intéressants, incluant une phylogénie fiable, il a fallu passer par quelques étapes. La première est la construction du jeu de données qui inclut les séquences nucléotidiques et/ou protéiques<sup>8</sup> de nos différents amplicons. La deuxième englobe l'alignement des séquences par un programme d'alignement de type local ou global, suivi d'une édition manuelle qui corrige les imperfections des algorithmes.

---

<sup>7</sup> [http://stanxterm.aecom.yu.edu/wiki/data/Product\\_manuals\\_attach/qiaquick.pdf](http://stanxterm.aecom.yu.edu/wiki/data/Product_manuals_attach/qiaquick.pdf) (1 juin 2010)

<sup>8</sup> La séquence nucléotidique et protéique fut utilisée pour COI et COII. Pour l'ADNr, seule la séquence nucléotidique fut analysée; ITS1 et ITS2 ne sont pas des séquences codantes, mais plutôt des espaceurs internes transcrits destinés à l'élimination.



En plus de visualiser directement le résultat du séquençage afin de pouvoir repérer les erreurs encourues par le séquenceur, ce logiciel nous permet par ailleurs de transposer facilement la séquence dans sa forme complémentaire et d'enregistrer les séquences en format fasta (.fasta ou .fa). Le format fasta est le type de fichier nécessaire pour transiter entre la base de données du logiciel GENEIOUS (Drummond *et al.*, 2009) et les logiciels de bioinformatique.

**Étape 2.** Les séquences en format fasta sont éditées si l'électrophorégramme montre des imperfections, assemblées dans certains cas puis annotées. L'édition des séquences implique qu'on élimine le début (accumulation de déchets) et la fin (baisse de qualité) de la séquence. On identifie et élimine les pics parasites, on corrige les ambiguïtés (N : base non déterminée) et finalement, on repère et enlève les séquences des amorces. Toutes ces corrections peuvent se faire manuellement dans l'application bloc-notes ou dans le logiciel BioEdit. Cependant, dans certains cas, les séquences éditées doivent, en plus, être complétées (superposées dans le cas du séquençage du même gabarit dans deux directions) ou assemblées (mise bout-à-bout des fragments successifs séquencés). Toutes les données corrigées sont entrées dans la base de données GENEIOUS, puis chacune des séquences est annotée, étape qui consiste à identifier chacun des bouts de séquences. Par exemple, pour l'unité de transcription de l'ADNr, chacun des gènes (5.8S, 2S, 18S et 28S) ainsi que chacun des espaceurs (ITS1 et ITS2) fut repéré et annoté, ce qui facilite l'analyse des séquences. Une fois cette étape terminée, quelques-unes des séquences corrigées et éditées furent déposées dans GenBank.

#### 1.3.3.6 L'alignement des séquences

**Étape 3.** L'alignement consiste à disposer les unes sous les autres des portions de séquences homologues en minimisant leurs différences. Les séquences d'ADN se composent de caractères qui peuvent avoir cinq états différents : A, G, C, T ou - (délétion). Les sites qui conservent les mêmes états pour chacune des séquences constituent des portions conservées, tandis qu'un changement d'état fait référence à une substitution ou délétion.

Chacun de nos alignements a été fait à l'aide de divers logiciels dans le but d'optimiser l'alignement de nos différentes séquences. Nous avons décidé de prendre cette précaution, car les programmes sont souvent efficaces pour les régions de fortes similitudes, mais beaucoup moins pour les séquences divergentes. Dans le cas d'ITS2, on a même dû aller jusqu'à reconstruire la structure secondaire afin de garantir un alignement plus juste. Après l'alignement automatique fait par le logiciel, il est important d'affiner manuellement notre alignement et de retirer de l'analyse les régions où l'alignement est ambigu (Lopez *et al.*, 2002). Cette étape manuelle souvent négligée par certains engendre parfois un alignement erroné.

Les logiciels utilisés sont Geneious Align, Muscle et Clustal W, déjà implantés dans le logiciel GENEIOUS, et T-Coffee<sup>9</sup> disponibles sur Internet. La comparaison des différents alignements que ces logiciels génèrent nous permettait d'avoir une meilleure vision de l'alignement optimal. Pour les gènes mitochondriaux COI et COII, les alignements sont totalement identiques peu importe le logiciel, car aucune insertion/délétion ou répétition interne n'est observée. Pour les gènes 5.8S, 2S, et l'espaceur 2a, il fut facile de faire l'alignement, mais il en fut autrement pour le 18S et les deux espaceurs ITS1 et ITS2. Afin de comprendre pourquoi nous ne pouvons utiliser n'importe lequel des logiciels pour ces derniers, il faut savoir qu'il existe deux types d'alignement : globaux<sup>10</sup> et locaux<sup>11</sup>. Les alignements globaux sont utilisés quand les séquences mises en jeu sont similaires et de taille égale comme COI et COII, tandis que les alignements locaux sont utilisés quand deux séquences dissemblables sont soupçonnées de posséder des motifs homologues. Certains programmes comme T-coffee combinent les deux sortes d'alignements. Ce dernier compare toutes les séquences deux par deux, produisant ainsi un alignement global et une série d'alignements locaux. Il combine par la suite tous ces alignements au sein d'un seul alignement multiple. Ce type d'alignement est très utile puisqu'il permet d'identifier les régions qui sont conservées d'une espèce à une autre. Ce logiciel fut spécialement employé pour les marqueurs 18S, ITS1 et ITS2. Pour le 18S, l'alignement obtenu est assez performant de sorte que seules

---

<sup>9</sup> <http://www.ebi.ac.uk/t-coffee/>

<sup>10</sup> L'alignement global est basé sur l'algorithme de Needleman-Wunsch (J. Mol. Biol. 1970 48:443-5)

<sup>11</sup> L'alignement local est basé sur l'algorithme de Smith-Waterman (J. Mol. Evol. 1981 18:38-46).

quelques corrections manuelles durent être effectuées. C'était à première vue le chaos pour l'espaceur ITS1, mais l'alignement permettait de repérer huit blocs consensus. Par contre, les séquences entre ces blocs sont difficiles à aligner. Pour ITS2, malgré une multitude d'insertions/ délétions, la présence de longs blocs consensus facilite l'alignement. Nous avons également obtenu un alignement manuel en se guidant sur la structure secondaire déduite de MFOLD. La connaissance de cette structure garantit en principe un alignement optimal (Kjer, 2004) et non biaisé pour la suite des procédures, soient l'analyse du jeu de données et la construction d'arbres phylogénétiques.

### 1.3.3.7 L'inférence phylogénétique

**Étape 4.** La phylogénie moléculaire est l'utilisation d'un alignement de séquences pour obtenir des informations sur l'histoire évolutive des êtres vivants et notamment sur leurs liens de parenté, soit leur phylogénie. Les espèces y sont donc classées en fonction du caractère plus ou moins récent de leur dernier ancêtre commun. Il existe actuellement trois méthodes de reconstruction phylogénétique : cladistique (repose sur l'étude des états de caractères), phénétique (méthodes basées sur les distances) et probabiliste (méthodes basées sur les caractères) (Harry, 2001).

Les méthodes de distance (UPGMA, NJ) sont basées sur une matrice de dissimilarités, qui s'obtient en comparant les séquences deux-à-deux. Nous n'avons pas utilisé la méthode UPGMA qui présente le désavantage de supposer des taux d'évolution identiques sur toutes les lignées (toutes les feuilles sont à la même distance de la racine) et satisfont donc le critère d'horloge moléculaire. Quant à l'algorithme Neighbor-Joining (NJ), il ne suppose pas cette horloge, mais fonctionne mieux quand elle est respectée.

Les méthodes de caractères sont fondées sur des algorithmes statistiquement plus robustes que ceux de distances. Cette approche regroupe quelques méthodes dont celle de maximum de vraisemblance (Maximum Likelihood), laquelle nous avons utilisée. Dans celle-ci, les bases ou acides aminés de toutes les séquences sont considérées séparément à chaque site et le logarithme de la vraisemblance est calculé pour une topologie donnée en utilisant un modèle probabiliste. Ce logarithme est cumulé sur tous les sites et la somme est maximisée pour estimer la longueur de branches de l'arbre.

Cette procédure est répétée pour toutes les topologies possibles et la topologie ayant la plus haute vraisemblance est choisie. Cette méthode, considérée comme la plus fiable de toutes les méthodes phylogénétiques, conduit à l'arbre évolutif le plus probable. Par contre, elle demande la plus grande puissance de calcul et prend plus de temps. Comparée à la parcimonie, autre méthode, elle est beaucoup plus reproductible et moins sensible aux effets de l'attraction de longues branches. En plus, elle permet d'appliquer les différents modèles d'évolution et d'estimer la longueur des branches en fonction du changement évolutif.

### 1.3.3.8 Modèle d'évolution des séquences :

Le calcul de la fonction de vraisemblance dans sa forme la plus simple nécessite un modèle décrivant les probabilités de substitution d'un nucléotide par un autre. Pour être performantes et statistiquement robustes, les méthodes probabilistes nécessitent l'utilisation de modèles qui décrivent de la façon la plus réaliste possible les processus biologiques d'évolution des séquences. Ces modèles incorporent des paramètres prenant en compte les fréquences nucléotidiques, les taux de substitution, la présence d'une fraction de sites invariables et l'hétérogénéité des taux de substitution entre sites. Voici les principaux modèles proposés :

- Jukes-Cantor à un paramètre (JC) : le plus simple modèle d'évolution de séquences nucléotidiques. Les 4 bases ont les mêmes fréquences ( $\pi_A = \pi_T = \pi_G = \pi_C$ ) et les transitions ( $\alpha$ ), puis les transversions ( $\beta$ ) sont équiprobables.

- Kimura à deux paramètres (K2P): les quatre bases ont les mêmes fréquences ( $\pi_A = \pi_T = \pi_G = \pi_C$ ) et les transitions, puis les transversions ne sont pas équiprobables.

- Tajima-Nei (TN) = modèle Felsenstein à six paramètres (F81) : similaire à Jukes-Cantor, si ce n'est qu'il admet des fréquences différentes pour chaque nucléotide au lieu de considérer  $\frac{1}{4}$  pour chacun, mais considère les transitions, puis les transversions équiprobables.



- Hasegawa, Kishino et Yano à six paramètres (HKY 85) : les 4 bases ont des fréquences différentes ( $\pi_A \neq \pi_T \neq \pi_G \neq \pi_C$ ) et les transitions puis les transversions ne sont pas équiprobables.

- GTR (General Time Reversible) ou REV, englobe l'ensemble des modèles disponibles à ce jour (JC, K2P, TN, F81, HKY85). Il comporte dix paramètres; les quatre bases ont des fréquences différentes ( $\pi_A \neq \pi_T \neq \pi_G \neq \pi_C$ ) et il existe six types de substitutions.

Une critique souvent adressée à l'encontre des méthodes probabilistes en phylogénie moléculaire est leur dépendance vis-à-vis du modèle d'évolution sous-jacent. Ainsi, il est important d'employer des modèles qui s'ajustent le mieux possible aux données analysées. Pour ce faire, des tests statistiques probabilistes comme celui du rapport des vraisemblances (LRT pour « Likelihood Ratio Test ») et le critère d'information d'Akaike (AIC pour « Akaike Information Criterion ») ont été proposés pour comparer les différents modèles. Posada et Crandall (1998) ont développé le programme MODELTEST qui permet de choisir le modèle d'évolution qui s'ajuste le mieux aux données parmi toute une gamme disponible sur la base du LRT et du AIC. L'estimation simultanée des paramètres du modèle est menée par le logiciel PAUP\* (Swofford, 2002) sous 54 modèles d'évolution différents définis par Posada et Crandall (1998, 2001). Dans notre cas, le modèle sélectionné sur la base du critère LRT et AIC par le programme MODELTEST est le GTR.

Afin d'affiner encore plus, nous avons aussi demandé la valeur du « bootstrap » lors de la construction des arbres phylogénétiques. Cette méthode utilisée pour tester la validité d'un arbre (Harry, 2001; Pevsner, 2003) consiste à effectuer un tirage des sites au hasard avec remise. Donc, dans chaque réplique de bootstrap, certains sites peuvent être présents plusieurs fois, tandis que les autres peuvent être absents. Chaque réplique produit un nouvel alignement artificiel qui est utilisé pour construire un arbre artificiel. Pour chaque branche interne, on calcule le pourcentage des arbres artificiels contenant cette branche. Plus la valeur est élevée, plus la validité du nœud est grande; ce dernier est alors dit robuste (Harry, 2001). Même si une valeur est élevée, cela ne

signifie pas pour autant que l'arbre est juste, mais indique seulement qu'il est le plus probable selon le jeu de données et le type d'analyses utilisées. Cependant, on considère généralement que les branches définies par une valeur de bootstrap de plus de 95% sont fiables (Diallo, A. comm. pers.).

Finalement, la démarche décrite ci-dessous donne une idée de la complexité du travail requis pour produire une phylogénie.

- 1) Effectuer la construction du jeu de données et aligner les séquences.
- 2) Trouver le meilleur modèle d'évolution pour chacun des marqueurs à l'aide de MODELTEST/PAUP\*. Pour ce faire, il faut dans un premier temps exporter notre alignement, tel que visualisé dans GENEIOUS, en un fichier ayant une autre extension (.nex); on le nomme format nexus. Ce nouveau fichier est par la suite modifié dans l'application bloc-notes, où on insère les lignes de programmation nécessaires afin que le logiciel PAUP\* 4.0b10 puisse tester les 54 modèles d'évolution. Après plusieurs heures/jours selon la série de données, PAUP\* génère un fichier portant l'extension (.scores) qui peut être ouvert à l'aide des commandes dans une fenêtre DOS; ceci a pour but de modifier le fichier portant l'extension .scores en un fichier .txt pouvant être visualisé à l'aide de l'application bloc-notes.
- 3) Construction de l'arbre phylogénétique dans PhyML. Il existe plusieurs programmes de maximum de vraisemblance. Nous avons utilisé PhyML (Guindon et Gascuel, 2003), lequel permet de traiter un très grand nombre d'espèces en un temps raisonnable. Pour ce faire, il faut reprendre notre même fichier contenant l'alignement, tel que visualisé dans GENEIOUS, puis l'exporter avec l'extension (.phy). Par la suite, il suffit simplement de copier ce nouveau fichier dans le dossier .exe de phymI\_v2.4.4, puis de lancer le programme et d'entrer une ligne de programmation afin de pouvoir ouvrir notre fichier. On procède à l'ajustement de certains paramètres dont le modèle d'évolution (nous voulions le GTR) et le nombre de bootstraps, soit ici 500 ou 1000. Le programme génère par la suite plusieurs fichiers et il suffit de convertir encore une fois l'extension .txt en .tree et d'en importer le fichier dans GENEIOUS pour visualiser l'arbre.

## **CHAPITRE II**

**« Seasonal and Spatial Distributions of Black fly Larvae (Diptera: Simuliidae) in Two Lake Outlet Streams of the Mauricie Region of Québec, and Species Survey in Parts of Southern Québec Territory. »**

**Seasonal and Spatial Distributions of Black fly Larvae (Diptera: Simuliidae) in  
Two Lake Outlet Streams of the Mauricie Region of Québec, and Species Survey in  
Parts of Southern Québec Territory.**

**Christine Gaudreau<sup>1</sup> and Guy Charpentier<sup>1</sup>**

**<sup>1</sup>Département de Chimie-biologie, Université du Québec à Trois-Rivières, 3351  
Boulevard des Forges, Box 500, Trois-Rivières, Canada G9A 5H7.**

**Corresponding author:   Guy Charpentier**

Box 500, Université du Québec à Trois-Rivières

Trois-Rivières, Québec, G9A 5H7 Canada.

819-376-5053 (phone)

819-376-5084 (fax)

E-mail: [guy.charpentier@uqtr.ca](mailto:guy.charpentier@uqtr.ca)

**Key words :** Black fly, Simuliidae, distribution patterns, *Prosimulium*, *Cnephia*,  
*Twinnia*, *Simulium*, *Metacnephia*.

## 2.1 Résumé

Il est d'un grand intérêt écologique d'effectuer un nouvel inventaire des simules du sud du Québec, incluant un échantillonnage exhaustif de deux décharges de la région Mauricienne, car les dernières données publiées provenant du Québec datent d'avant 1980. Il pourrait être important d'effectuer un échantillonnage plus fréquent afin d'établir les aires de distribution des espèces selon les changements climatiques et environnementaux. Vingt-deux des soixante douze espèces précédemment recensées ont été retrouvées dans cette étude. Une nouvelle espèce a été répertoriée pour le Québec, *Prosimulium magnum*, et de nouvelles aires de distribution ont été trouvées pour quelques espèces. L'échantillonnage exhaustif de deux ruisseaux montre clairement une variation de la distribution temporelle et spatiale des espèces univoltines et polyvoltines. Les variations saisonnières des espèces sont synchronisées dans nos deux sites à l'étude, mais le nombre de générations est plus petit au Lac du Barrage qu'au Lac Souris. Il y a un effet environnemental sur le processus de colonisation des espèces, par exemple pour *Simulium annulus* et *S. jenningsi* en ce qui concerne la largeur du cours d'eau et le type de substrat, et pour *S. longistylatum* et *S. pictipes* qui préfèrent fortement des cours d'eau à fort débit. De plus, la colonisation d'un site artificiel par *S. aureum* dans un environnement péri-urbain n'a jamais été observée chez les simules avant cette étude, contrairement à ce qui est observé chez plusieurs espèces de moustiques.

### 2.1.1 Contribution des auteurs de l'article

La partie expérimentale a été effectuée entièrement par Christine Gaudreau. De plus, l'article a été rédigé par celle-ci et fut révisé par le Dr Guy Charpentier. Ce manuscrit sera publié sous peu dans le journal *Northeastern Naturalist*.

### 2.2 Abstract

There is great ecological interest in carrying out a new inventory of southern Québec simuliids, including an exhaustive sampling in two lake outlet streams of the Mauricie region, because data from Québec were last reported only prior to 1980. Frequent sampling is needed to determine how species distributions change in response to climatic and environmental changes. Twenty-two of the 65 Québec species were collected in the present study. One new species for Québec was collected (*Prosimulium magnum* complex) and new distribution areas were found for some species. Exhaustive sampling clearly showed time and space-related distribution variations of the univoltine and polyvoltine species. Seasonal appearance of species was synchronized at the two sites studied, but generation number was smaller at Lac du Barrage than at Lac Souris. The species-colonizing process seems to be environmentally determined stream width and substrate type influenced as *Simulium annulus* and *S. jenningsi*. *S. pictipes* and *S. innoxium*, were found in high current velocity falls. Artificial site colonization by *S. aureum* in a peri-urban environment had not been observed among simuliids prior to this study, in contrast to many mosquito species.

### 2.3 Introduction

Black flies are biting pests, some species of which attack human beings, mammals and birds, inducing high livestock-related economic losses and having a somewhat negative impact on tourism (Adler et al. 2004). In North America, black flies may kill wild and domestic mammals when biting or when large numbers are inhaled, thus blocking the respiratory tract (Adler et al. 2004). Black flies can transmit nematodes, protozoa and viruses, and thus some species are vectors of disease, including parasites such as *Leucocytozoon*, *Ornithofilaria* and *Trypanosoma* to wild and domestic birds (Bennett 1961; Hunter et al. 1997; Twinn 1933). Vesicular stomatitis New Jersey virus (VSNJV) can be transmitted by *Simulium vittatum* to cows, pigs and horses (Adler et al. 2004). VSNJV is involved in one of the most severe viral diseases of livestock (Mead et al. 2004). Although Québec is close to New Jersey, no pathogen or parasite transmission to mammals has yet been observed in this area. Only minor troubles due to bites or allergic reactions have been observed. In the forested areas of the Canadian Shield, *Simulium venustum* populations are so thick that outdoor activities are impossible during the day-time. Ultimately, however, investigations of black flies are more important for the economic repercussions than for medical importance.

Black flies have been extensively studied almost worldwide. But their distribution in Québec is not well known (Back and Harper 1978, 1979; Twinn 1936; Winn and Beaulieu 1932; Wolfe and Peterson 1959; Wood et al. 1963). Indeed, the most recent studies of Québec areas were published before 1980. Adler et al. (2004) only reported Québec records from the locations appearing in these references. Seventy-two

or sixty-five black fly species, depending on the authors and with some synonymies, have been identified from Québec (Adler et al. 2004, listed 65; Adler and Crosskey 2009, listed another one; Back and Harper 1978, 1979, listed four more; Wolfe and Peterson 1959, added two more). Species reports are important in that they are needed to assess the dispersal or appearance/ disappearance of various species among locations and through time. Species turnover may be accelerated due to human activities (international trade and tourism) or climate change. Effects of climate change on entomofauna in temperate zones are of growing interest (Ayres and Lombardero 2000), since insect movements are an early indicator of rising temperatures (Epstein 2000), which contribute to more rapid development of insects. Furthermore, milder winters, and longer frost-free periods will probably allow some temperate zone species whose distribution is presently cold-limited to become more widespread (Patterson et al. 1999), (e.g. *Twinnia tibblesi*, which is confined to northern and mountainous regions). Finally, regional warming could increase the rate of insect-transmitted viral diseases (Dohm et al. 2002) and the distribution of insect vectors (Kozár and Dávid 1986). Onchocerciasis (Adler et al. 2004; Blacklock 1926; Dalmat 1955), which is transmitted by black flies in Africa and in central and South America, is expected to show an increased incidence related to climate change (Rogers and Packer 1993). The same phenomenon could occur with other *Leucozytozoon* parasites, such as *L. simondi* Mathis and Leger 1910, which infects ducks, and *L. smithi* Laveran and Lucet 1905, which infects turkeys and geese, and with some viruses, such as VSV and arboviruses (Snowshoe Hare), for instance.

Between 2003 and 2008, we compiled a list of the species living in various southern Québec streams, for comparison with lists established mainly before 1980. It



could then be shown whether or not there has been any movement or any new distribution among some species. Next, to improve knowledge of Québec black fly species, a survey of Québec streams was carried out. Surprisingly, few reports exist about Québec simuliids. The major studies relate to the Baie-Comeau area (Wolfe and Peterson 1959), Mont Saint-Hilaire (Mackay 1969), two streams located at Saint-Hippolyte de Kilkenny in the Laurentians (Back and Harper 1978, 1979), and the Magdalen Islands (Adler et al. 2005). New studies have been published subsequently, but they all relate to the Maritime provinces (McCreadie et al. 1995; Minhas et al. 2005) or to the Northwest Territories and Nunavut (Currie and Adler 2000). Finally, in 2006, we carried out an exhaustive sampling program at two lake outlets located in the Mauricie area. In this paper we report on seasonal and spatial black fly larva succession at these two lake outlets.

#### **2.4 Field Site Description and Methods**

During the 2003-2008 survey in Québec in two climatic regions (Boreal and Great Lakes-Saint Lawrence) (Table 2.1), black fly larvae were collected several times in 24 streams chosen for their accessibility. At each sampling site, larvae were hand-collected from vegetation and rocks, and fixed in ethanol (95%) for subsequent use in molecular systematic studies (LaRue et al. 2009); they were then morphologically identified in the laboratory. All streams sampled were apparently permanent.

Two sampling sites were chosen for the seasonal species succession study: the Lac Souris outlet (Fig. 2.1A) located at Saint-Mathieu du Parc (45°35'N; 72°58'W) and

the Lac du Barrage outlet (Fig. 2.1B) (46°29'N; 72°58'W) located at Saint-Élie de Caxton. Both sites are in the Mauricie area where climate is temperate-cold, of the continental type (climatic zone 4A): average temperature is -7°C in February and 22°C in July (Environment Canada). The environmental parameters of the two streams are summarized in Table 2.2. These two sites were chosen for their accessibility and the presence of flow during all seasons, even during the coldest winter days. Sampling protocol was the same for both sites. Six sampling stations (numbered 1-6 extending from a lake outlet) were delineated by placing beacons (red bricks) every 15 m of stream length. Sampling was performed 0.5 m upstream and 0.5 m downstream of the beacon across the total width of the stream. The one exception occurred at the Lac du Barrage outlet, where the distance between stations 1 and 2 was 35 m. As in previous studies of this type (Corkum and Currie 1987; McCreadie and Adler 2006; McCreadie and Colbo 1991), it was assumed that species found in each sampling site were representative of local occurrences. About 50 to 100 larvae per collection day for each sampling station were hand collected from all types of substrates: vegetation, rocks and artificial substrates such as yellow/white polypropylene ropes attached to red bricks. The Lac Souris and Lac du Barrage sites were sampled bimonthly from February to September, except in April, June and September, when they were sampled only once. These streams were also observed in October, where winter larvae species were found but were too small and hence difficult to identify. No larvae were found in Lac du Barrage in June or at stations 1, 3 and 4 in April. The same phenomenon was seen at Lac Souris in April for stations 3, 5 and 6, where larvae might have drifted away due to the high water level. This phenomenon was observed too by Wolfe and Peterson (1959).

All black fly larvae collected for the succession study or the species distribution survey were identified in the laboratory. Taxa were identified to species, where possible, using many morphological keys (Adler et al. 2004; Adler and Kim 1984; Currie 1986; Wood et al. 1963). Each *Prosimulium* larva's teeth were dissected, slide-mounted using a commercial setting resin (Aquamount mountant, BDH Inc., Ville St-Laurent, Québec, Canada) and examined beneath a compound microscope for identification. We were unable to differentiate some members of the *S. venustum/verecundum* complex on the basis of morphology only. Because we did not perform cytotyping, we maintained the appellation *S. venustum/verecundum* complex for these individuals. In the case of *S. euryadminiculum* put in synonymy with *S. annulus* by Adler et al. (2004), we agreed with the synonymy since our molecular data fit with that (Gaudreau et al. unpublished data).

## 2.5 Results and Discussion

### 2.5.1 Distribution of Taxa in Québec

The samples collected throughout Québec provided data on presence, abundances and distribution among instars of the various species at each of the studied areas. We found 22 species in the genera *Twinnia*, *Prosimulium*, *Helodon*, *Cnephia*, *Stegopterna*, *Metacnephia* and *Simulium* (Table 2.1). Comparison of our collection records with North American simuliid distribution maps (Adler et al. 2004) indicated that at least one new species record was added to the Québec list; *Prosimulium magnum* complex found in May at Lac Souris. Furthermore, our observations expanded the

distribution of several previously reported Québec species (Table 2.1). The World distribution for these same species is summarized in Table 2.3.

Some factors are widely recognized to influence larval distribution and persistence at a location, especially current velocity. Simuliid larvae as a group can develop in various types of streams or rivers, ranging from very small ones, where the flow is barely detectable (e.g., spring-fed streams that do not exceed 0.3 m width for *T. tibblesi* (Stone and Jamnback, 1955)), to much larger ones, that support cascades, torrents and waterfalls, where we found *S. innoxium* (= *S. pictipes*) (Chutes Monte-à-Peine) and *S. pictipes* (= *S. longistylatum*) (Chute Dunbar and Six Chutes) (Table 2.1). In contrast, *St. mutata* (= *Cnephia mutata*) and *S. vittatum*, are not limited to a specific water velocity range and are found in most; these species are widely distributed throughout Québec (Table 2.1).

*Twinnia tibblesi* is a univoltine species (Davies et al. 1962) usually found high above sea level. In earlier studies cited by Wood (1978), it was found in Québec at Old Chelsea, Gatineau (1971), Mont Saint-Hilaire (1966), La Vérendrye Provincial Park (1965) and Laurentide Provincial Park (1954). In the present study, *T. tibblesi* was also discovered for the first time, in May 2005, in a Gaspesia stream close to Mont Lyall (Table 2.1). According to Wood (1978), *T. tibblesi* species is associated with species such as *C. abdita* Peterson 1962, *S. rivuli* Twinn 1936, *H. decemarticulatus* in Algonquin Park, Ontario but supposed in the previous referenced studies to be found alone in Québec sites. The present study, demonstrates that *T. tibblesi* in Québec can be associated with some other species, such as *Prosimulium*, *St. mutata* and *S. annulus*.

*Metacnephia saileri* (Malmquist et al. 2004) is seldom mentioned in the literature (Laparé 1980); it was found in only three sites across Québec: Fort Chimo, Grande-Rivière-de-la-Baleine (Shewell 1957), and close to Caniapiscou River (Laparé 1980). The present study adds a new collecting site for *M. saileri* at the Duchénier Reserve. *Cnephia dacotensis* can be found in Québec from the Saint-Lawrence Valley to Baie-Comeau (Shewell 1958).

*Simulium aureum* was confined to small streams with abundant vegetation (Wotton 1976) and was absent in a stony bed. In the Lanaudière area, however, this species can be found during the whole summer in small, artificial water gardens (Preformed Waterfall kit, Algreen Products, Cambridge, Ontario) with a very low flow rate and sparse vegetation (*Ajuga reptans* L.), or no vegetation at all. Larvae cling directly to the plastic bottom of the artificial fall. *Simulium aureum* has been found on hygroscopic surfaces in cavernicolous areas (Grenier 1953): unsurprisingly then, it was collected in this low-water type of habitat. Whereas Ezenwa (1974) considered *S. aureum* to be univoltine, Malmqvist et al. (2004) reported it as being multivoltine. The present data also show it to be multivoltine.

*Simulium quebecense* is observed around the Great lakes, in Ontario and Québec, and southward along the Appalachian range (Adler et al. 2004). In Québec, in 1978, the only habitats in which it was observed were in the Outaouais area and in the streams of the Station de Biologie de l'Université de Montréal (SUBM) at Saint-Hippolyte de Kilkenny. The present data show them also in Lac Souris in the Mauricie region.

Widespread in southern Canada and the eastern United States, *S. jenningsi* was also found in the Mauricie, Montérégie and Lanaudière regions. *Simulium annulus* was found in the Lower Saint Lawrence area. *Simulium venustum/ verecundum* complex has the distribution pattern already described and is found as far as the Ungava Bay. Unsurprisingly, this species was collected in the Lanaudière, Laurentides and Mauricie areas.

### 2.5.2 Temporal and Spatial Changes in Species Distribution

As formerly mentioned by Back and Harper (1979), the length of time a species persists in a stream varies according to the years of the survey, but the species replacement patterns are approximately the same throughout the year: univoltine species are found as larvae from January to May, and polyvoltine species from May to October. This survey clearly shows this pattern. But more precise studies that use determinations of cytospecies and/ or molecular taxonomy might reveal a more complex pattern for species complexes, with a mix of univoltine and polyvoltine cytospecies not distinguished now. As regards univoltine species of *Prosimulium* genus (*P.spp*) (Snyder and Linton 1983), *C. dacotensis* (Gersabeck and Merritt 1979) and *St. mutata* (Wolfe and Peterson 1959), larval develop to their last instars during spring before mid-May, whereas polyvoltine species, with larvae developing later, are seen only after May. These polyvoltine species all belong to genus the *Simulium* (*S. venustum/ verecundum*, *S. tuberosum*, *S. decorum*, *S. jenningsi* and *S. silvestre*) (Adler et al. 2004; Back and Harper 1979); they appear at the time that larval populations of *Prosimulium* and *Cnephia* species decline. Two species of *Simulium*, however, behave differently:

*Simulium vittatum* is present year-round and *S. annulus* appears concomitantly with univoltine species. Finding *S. vittatum* all year round appears to be normal since there is no egg diapause, and larvae are produced continuously (Back and Harper 1979). On the other hand, surprisingly, *S. annulus*, considered as polyvoltine, was found on March 27th and April 27th, which is at the same time as univoltine species. As regards *S. silvestre*, a large peak is observed on May 18th at Lac Souris, whereas at Lac du Barrage, low numbers were found from May 18th to July 4th.

The times at which the first generation of species first appeared in the season were approximately synchronized between the two sites that were studied in detail. However, they did not remain synchronized for the following generation. Most multivoltine species exhibited one less generation at Lac du Barrage than was seen at Lac Souris. Except for *S. jenningsi* which has only one generation, two out of three generation arrival peaks are observed but the time of appearance can be shifted as already noted by Back and Harper (1979). *Prosimulium* sp., *C. dacotensis* (not distinguishable from *C. ornithophilia* as larvae without cytological study) and *St. mutata* show only one generation by year in the present study following the pattern observed in Ezenwa's (1974) study, indicating the univoltine status. *Simulium vittatum* is found all year round (Laparé 1980). Larvae of the overwintering generation are present at the same time as the univoltine species in winter, whereas the three following generations are synchronized with those of the polyvoltine species. At Lac du Barrage, two generations were observed: a first one between July 18th and August 15th and a second one, probably, after September 14th. Back and Harper (1979) dismiss the hypothesis of a third generation, but we clearly found this third generation at Lac Souris during the

summer. Three generations have also been reported in the literature. Four generations have been observed in Alaska, Saskatchewan, Illinois and South Carolina. In contrast, only one generation was found in Labrador and two were observed in Algonquin Park, Ontario (Wolfe and Peterson 1959). In the present study, *S. decorum* exhibited two generations at Lac du Barrage, and three at Lac Souris: these data are similar to those of Adler et al. (2004), who reported two to four generations a year. According to Wolfe and Peterson (1959), this species has two generations a year, the first one being the largest. *Simulium tuberosum* yields three generations at Lac Souris and two at Lac du Barrage. Four generations, however, were found in the Baie Comeau area, as well as in the Adirondacks (Wolfe and Peterson 1959). The third generation was much larger than the others. In the present study, this distribution pattern was not clearly observed. *Simulium tuberosum* larvae, which could be morphologically differentiated from those of *S. venustum* in this study, were found associated with them. *Simulium tuberosum* has the same generation pattern as *S. venustum*. It was observed that generations become less abundant when following one another. Usually, multivoltine species like *S. tuberosum* and *S. vittatum* have a small first generation and the population builds up gradually. This was not observed in the present study, except for *S. tuberosum* at Lac du Barrage. It has been sometimes observed that the *S. venustum* population may be larger than the *S. decorum* population. The present study gives the same results.

Species distribution along the outlet streams changes according to seasons. During winter and early spring, most species found belong to genus *Prosimulium*, whereas *S. venustum/verecundum* were found late in spring and during summer (Fig. 2.2). At Lac Souris, *S. venustum/verecundum* constituted 56% (1 to 98%) of black



fly larvae collected, whereas Wolfe and Peterson (1959) found that *S. venustum* constituted 75 to 80% of the population. At Lac Barrage, we found that 75% of larvae were *S. venustum/ verecundum* (52 to 96%) in accordance with these authors. It may be difficult or sometimes impossible to distinguish the species identity of first instar larvae of the *S. venustum/ verecundum* complex, and the same was true for species of *Prosimulium*. For the following accounts of species succession these complexes will be simply called *Prosimulium* spp. and *S. venustum/ verecundum* complex. As concerns the two outlet streams studied, however, we were able to identify 400 *Prosimulium* and *Helodon* larvae to the species level, and at least seven morphologically distinct species were found: *H. decemarticulatus*, *P. fontanum*, *P. fuscum*, *P. magnum* complex, *P. mixtum*, *P. multidentatum* and *H. vernalis*. *Prosimulium fuscum* represented 67.5% of the larvae collected at Lac Souris and 81.8% at Lac du Barrage. At Lac Souris, along with *P. fuscum*, 11.1% of *P. mixtum*, 14.9% of *P. fontanum*, 2.4% of unknown, 1.4 % of *H. vernalis*, 1.1% of *H. decemarticulatus*, 0.5% of *P. magnum* and 0.8% of *P. multidentatum* were found. At Lac du Barrage, there were 9.8% of *P. fontanum*, 4.8% of unknown, 1.7% of *P. mixtum*, 1.7% of *H. decemarticulatus*, 0.2% of *H. vernalis* and 0% of *P. magnum* complex and *P. multidentatum*.

At Lac Souris (Fig. 2.3A), more *S. venustum/ verecundum* were found at station 1 and 2, close to the lake outlet and more *Prosimulium* spp. at station 4 and 5, but in Lac du Barrage (Fig. 2.3B) these two complexes were evenly distributed along the stream. Among other species, there were more distributional differences between stations at Lac Souris than at Lac du Barrage, where no differences were observed except for *St. mutata*, *S. jenningsi* and *S. annulus*. Of the few larvae that could be collected, more

*C. dacotensis* were found at Lac Souris (16 larvae) than at Lac du Barrage (2 larvae). Similarly, more *St. mutata* larvae were found at Lac Souris (281 larvae) than at Lac du Barrage (29 larvae); this appears to be normal since these larvae preferentially adhere to substrates below small impoundment outflows (Adler et al. 2004), a condition occurring at Lac Souris and not at Lac du Barrage, which is much too wide. Large numbers of these two species are also found in the same region, in a stream at Saint-Boniface (140 *C. dacotensis*, 47 *St. mutata*) and also in the Laurentian area at SBUM of Saint-Hyppolite (53 *C. dacotensis*, 400 *St. mutata*). These two streams indeed, are a suitable habitat for *C. dacotensis* and *St. mutata* because they are less than 1 m wide. According to these data, *C. dacotensis* is found in the portion of the stream closest to the lake outlet at Lac Souris. The same pattern was observed for *St. mutata*, *S. decorum* and *S. vittatum*, which are more numerous in the stations closer to the lake outlet than in the others. *Simulium decorum* larvae, were equally abundant at the two sites studied, but *S. silvestre* larvae were obviously numerous at stations 3 and 4 of Lac Souris.

Back and Harper (1979) noted that *S. venustum* and *S. tuberosum* are mostly found at stations where *S. vittatum* are less common, but this was not the case in our study. These authors also said that *S. vittatum* is mainly observed in fast downstream areas at stations quite close to the lake. We observed this in Lac Souris where *S. vittatum* were only found in fast-flowing locations, whereas at Lac du Barrage, *S. vittatum* were evenly distributed along the stream.

Environment type also has an influence upon some species colonization. *Simulium annulus* and *S. jenningsi* larvae could not be found at Lac Souris (Fig. 2.3A).

Their absence is related to stream width (Adler et al. 2004). Indeed, *S. annulus*, are abundant in streams 2 - 10 m wide (Adler et al. 2004) and *S. jenningsi* in streams more than 6 m wide (Adler et al. 2004). In the present study, these two species were found at Lac du Barrage (Fig. 2.3B), mostly in stations 3 to 6 where the stream is wider. In this area, the substrate also plays a role. *Simulium annulus* larvae were found at stations 5 and 6, where the vegetation is the most abundant, whereas *S. jenningsi* were found in rockier areas. These data are similar to those of Adler et al. (2004) who reported that *S. annulus* larvae are found in wide shallow streams, quite often close to the lake outlet.

In order to confirm the consistency of the results of this wide distribution study in terms of time and space, larvae were collected twice more at the same stations in 2008, on February 7th and April 10th, for comparison with the 2006 data (Table 2.4). At Lac Souris in 2006, the reduced density of *Prosimulium* spp. seemed to leave space for more *St. mutata*, whereas the reverse effect was observed in 2008. The proportion of *St. mutata* decreased as the *Prosimulium* spp. proportion increased. As for Lac du Barrage, in February 2008, the *Prosimulium* spp. percentage was lower than in 2006, whereas the *St. mutata* and *S. vittatum* percentages are higher in 2008 than in 2006. These variations were not observed in larval collections at the end of March or at the beginning of April. All these differences between percentages for years 2008 and 2006 might be partially due to exceptional climatic conditions at the beginning of year 2008 (high amounts of snow, late spring). However, *Prosimulium* spp. was clearly always the predominant species at first, whereas the other species constitute a minority. The data from the 2008 collections and some others taken before 2006 (data not shown), confirm the accuracy of the species succession patterns at Lac Souris and Lac du Barrage.

## 2.6 Remarks and Conclusions

### 2.6.1 Québec Simuliidae Survey

Our study found only 22 of the 65 or 72 morphologically distinct black fly species that have been reported to occur in Québec. This is possibly the consequence of sampling, which was limited in time and space, and of the occurrence of cytotypes that were not recognized here because of our use of morphological and molecular criteria for distinguishing species. In order to develop a more comprehensive map of simuliid distribution and diversity, more than 24 sites should be visited, and they should not be restricted to southern Québec. Nevertheless, our black fly study provides the first update of the black fly species distribution patterns for southern Québec in over 20 years, adding at least one new species (*P. magnum* complex) and many new collecting sites. No comparative study of our sampling sites can be made, because with one exception, our collection points differ from those described in the published literature. However, we investigated the SBUM (Biology Station of the Montreal University) site, which had previously been studied by Back and Harper (1978, 1979). The species (*St. mutata*, *C. dacotensis*, *S. vittatum* and *Prosimulium* sp.) that we found in March 2003, were consistent with the records of those authors. This new collecting campaign added one or more new Québec locations to the distribution maps of Adler et al. (2004) for almost all of species collected here. The additional records include 1 more location for *T. tibblesi* (map 13), *C. dacotensis* (map 87), *M. saileri* (map 98) *S. quebecense* (map 150) and *S. decorum* (map 226), 2 more locations for *S. innoxium/ pictipes* (map 189 and 190), 3 more for *S. annulus* (map 114), 4 more for *S. jenningsi* (map 200), 5 more for *S. tuberosum* (map 237), 6 more for *S. vittatum* (map 172) in Québec and one other for

Newfoundland (Gaudreau et al. 2010), and finally 7 more for *S. venustum* (map 251)/*verecundum* (map 254) complex.

Environmental changes can modify some species distribution areas. This was indeed observed with *S. jenningsi* and *S. aureum* species. Simuliidae tend not to be found in polluted waters (Adler et al. 2004): therefore, no *S. jenningsi* had previously been found in the Saint Maurice River because of the floating logs. A prohibition on floating logs imposed in 1998 (Plante 2005) contributed to a decrease in environmental disturbance and to an increase in water quality, consequently resulting in recolonization of the Saint Maurice River by *S. jenningsi*. Quite interesting too, in the development of new habitat for early instar simuliids, a population was observed in small artificial aquatic habitats with a weak stream flow due to the water recycling in domestic gardens. Water is indeed pumped out towards a plastic-made waterfall with a mock gravel bottom. The falls portion of the artificial structure may be colonized by simuliids in summer, but populations cannot persist all year long, because these artificial aquatic habitats were operational only from the beginning of summer (May) to November. Given such forms of colonization, simuliid larvae might be expected to occur in similar habitats around or even inside the city of Montréal. These types of habitat should be investigated in the future.

No exhaustive survey of larval distribution throughout Québec has ever been carried out, thus making it difficult to undertake a comparative study of climate-related changes of species distribution. The goal of our sampling is to eventually be able to study effects of climate change on insects that are of great interest and raise many

questions (Ayres and Lombardero 2000). Progressive climate change is being recognized worldwide, and has an influence on insect distribution (Kozár and Dávid 1986). While new species invasions can occur, there can be a decrease in species number and their populations implanted for a long time.

### **2.6.2 Limitations and Limits of Larval identity**

One of the most significant challenges in relating current distribution and community composition to historical reports deals with reconciling the contemporary understanding of systematic relationships and species identities with historical reports. Most of the published reports of simuliid larval distribution in Québec tabulated species occurrences using morphological criteria. However, the most recent comprehensive classifications (e.g., Adler et al. 2004; Adler and Crosskey 2009) rely on the diagnostic power of cytotaxonomy and molecular methods to resolve and map the distribution of sibling members of species complexes. In this study, we have elected to use a conservative, morphologically-based approach to facilitate comparison with the available historical records. Research in progress using molecular methods (e.g. LaRue et al. 2009) will clarify some of the uncertainties of the current study. However, to facilitate comparison of historical species records with contemporary classification, Table 2.1 lists the synonymies between our usage of terms and the usage of Adler and Crosskey (2009).

Black flies are one of the most homogeneous of the insect families in term of morphology, thus making species identification quite difficult. Morphology-based larval identification is can be especially difficult and of limited value in distinguishing species

within a complex across the intricate geographic range. Some authors have described the morphology for all (Davies et al. 1962; Grenier 1953; Wood et al. 1963) or some (Chance 1970; Ross and Craig 1979) larval instars. If *Prosimulium* larvae are clearly differentiated from other genera through a morphological criterion: the secondary rays of the fans (Wood et al. 1963), identification to the species level is a little bit more complicated. Number of cephalic rays, head background, stain colours, tooth shape and length are often used but can be highly criticized by many authors (Alvan-Aguilar et al. 2005; Lucas and Hunter 1999; Zhang 2005). No single trait, like number and shape of teeth (Fig. 2.4) (Gaudreau et al.; unpublished data), number of cephalic fan rays (which may overlap in number among species (Snyder and Linton 1983), seston particle size ingested (Lucas and Hunter 1999; Zhang 2005) and stream velocity preference (Zhang 2005), can be used alone for obtaining an accurate identification. Moreover, larval pigmentation varies according to sex, collecting area (Alvan-Aguilar et al. 2005; Rothfels et al. 1978; Zettler et al. 1998) and fixative (Adler et al. 2004). *Simulium vittatum* larvae, for instance, which are found on rocks display dark head spots, whereas larvae with pale head spots are seen on vegetation (Adler and Kim 1984). We have recently shown that there is no genetic differentiation between these two morphotypes (Gaudreau et al. 2010).

Methods other than morphological study that are available for larval identification include cytogenetic and molecular techniques. Use of cytotaxonomy does permit some species to be recognized as members of a complex (Table 2.3) but can require a large number of specimens in order to yield interpretable material (Spironello et al. 2002) if larvae developed under particular environmental conditions: for instance,

only 1 in 10 larvae could display a proper chromosome spread on a slide (Spironello et al. 2002). Moreover, inappropriate larval storage at high temperature may lead to poor results, due to chromosome morphology alteration, and inexperienced staff could misidentify larvae; some chromosomal inversions have been misidentified in the subcomplex Ketaketa of *Simulium damnosum* Theobald 1903 (Krueger et al. 2006). Furthermore, cytotaxonomy can only be applied to late larval instars (Xiong and Kocher 1991) and does not allow adult female blackfly identification, which is quite important for epidemiological studies (Krüger et al. 2000). Recently, Morales-Hojas and Krueger (2009) point out a lack of clear genetic differentiation between the cytoforms of members in the *Simulium damnosum* complex.

Molecular methods can be much more appropriate than morphological and cytological studies. They are often very useful in clarifying taxonomic relationships and resolving the identity of morphologically similar species (Tang et al. 1998). Moreover, they can be used for both insect sexes and all life cycle stages (Tang et al. 1995, 1998). Finally, molecular markers are easy to work with, and their use requires only a small number of insects. Molecular tools could be of great value for broader investigations of insect species distribution and could assist in studying more precisely the impact of climate changes on North American simuliid populations. Our preliminary work on molecular genetic and phylogeny of black flies indicated some interesting problems in morphological taxonomy, especially in the genera *Prosimulium* and *Helodon* (LaRue et al. 2009). These studies also clarified distinction of species (LaRue et al. 2009), geographical origins and phylogeny (Gaudreau et al. 2008; 2010). Molecular markers can distinguish siblings within a complex and improve the resolution of studies of



simuliid seasonal and spatial distribution, and their environmental living conditions in connection with cytotyping and a revision of morphological taxonomic criteria.

## 2.7 Acknowledgements

We are grateful to the Université du Québec à Trois-Rivières, who provided C. Gaudreau's scholarship and funding for this research. We also thank Christian Back for the identification of some blackfly larvae and special thanks to Lise Meilleur, Pauline Marquer, Aude Pochon and Mylène St-Onge for their help on the field. We also want to thank Esther Lévesque and Benoit Tremblay for their help in identification of aquatic plants.

## 2.8 Literature Cited

- Adler, P.H. and Crosskey, R.W. 2009. World blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory. <http://www.entweb.clemson.edu/biomia/pdfs/blackflyinventory.pdf> (Accessed 28 November 2009)
- Adler, P.H., Giberson, D.J. and Purcell, L.A. 2005. Insular black flies (Diptera: Simuliidae) of North America: tests of colonization hypotheses. *Journal of Biogeography* 32: 211-220.

- Adler, P.H. and Kim, K.C. 1984. Ecological characterization of two sibling species, IIL-1 and IS-7, in the *Simulium vittatum* complex (Diptera: Simuliidae). Canadian Journal of Zoology 62: 1308-1315.
- Adler, P.H., Currie, D.C. and Wood, D.M. 2004. The blackflies (Simuliidae) of North America. ROM publication in sciences. New York, USA.
- Alvan-Aguilar, M.A., Hamada, N., Adler, P.H. and Luz, S.L.B. 2005. Cytotaxonomy of *Simulium cauchense* Floch & Abonnenc and *Simulium quadrifidum* Lutz (Diptera: Simuliidae) in Brazilian Amazonia. Memórias do Instituto Oswaldo Cruz, Rio de Janeiro. 100: 249-257.
- Ayres, M.P. and Lombardero, M.J. 2000. Assessing the consequences of global change for forest disturbance from herbivores and pathogens. The Science of the Total Environment 262: 263-286.
- Back, C. and Harper, P.P. 1978. Les mouches noires (Diptera : Simuliidae) de deux ruisseaux des Laurentides, Québec. Annales de la Société Entomologique du Québec 23: 55-66.
- Back, C. and Harper, P.P. 1979. Succession saisonnière, émergence, voltinisme et répartition de mouches noires des Laurentides (Diptera: Simuliidae). Canadian Journal of Zoology 57: 627-639.

- Bedo, D.G. 1975. Polytene chromosomes of three species of blackflies in the *Simulium pictipes* group (Diptera: Simuliidae). Canadian Journal of Zoology 53: 1147-1164.
- Bennett, G.F. 1961. On the specificity and transmission of some avian trypanosomes. Canadian Journal of Zoology 39: 17-33.
- Blacklock, D.B. 1926. The development of *Onchocerca volvulus* in *Simulium damnosum*. Annals of Tropical Medicine and Parasitology 20: 1-48.
- Chance, M.M. 1970. The functional morphology of the mouthparts of blackfly. Quaestiones Entomologicae 6: 245-284.
- Corkum, L.D. and Currie, D.C. 1987. Distributional patterns of immature Simuliidae (Diptera) in northwestern North America. Freshwater Biology 17: 201-221.
- Currie, D.C. 1986. An annotated list of and keys to the immature black flies of Alberta (Diptera: Simuliidae). Memoirs of the Entomological Society of Canada No 134, 90 pages.
- Currie, D.C. and Adler, P.H. 2000. Update on a survey of the black flies (Diptera: Simuliidae) from the Northwest territories and Nunavut project. Artic insect News 11: 6-9.

- Dalmat, H.T. 1955. The black flies (Diptera: Simuliidae) of Guatemala and their role as vectors of onchocerciasis. *Smithsonian miscellaneous Collections* 125: 1-425.
- Davies, D.M., Peterson, B.V. and Wood, D.M. 1962. The black flies (Diptera: Simuliidae) of Ontario. Part I. Adult identification and distribution with descriptions of six new species. *Proceedings of the Entomological Society of Ontario* 92: 69-154.
- Dohm, D.J., O'Guinn, M.L. and Turell, M.J. 2002. Effect of environmental temperature on the ability of *Culex pipiens* (Diptera: Culicidae) to transmit West Nile virus. *Journal of Medical Entomology* 39: 221-225
- Epstein, P.R. 2000. Is global warming harmful to health? *Scientific American* 283: 50-57.
- Ezenwa, A.O. 1974. Ecology of Simuliidae, Mermithidae and Microsporida in Newfoundland freshwaters. *Canadian Journal of Zoology* 52: 557-565.
- Gaudreau, C., LaRue, B., Charbonneau, V., Charpentier, G. and Craig, D.A. 2008. Recent origins and genetic fragmentation of *Simulium* (*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid. *Invertebrate Systematics* 22: 555-562.

- Gaudreau, C., LaRue, B. and Charpentier, G. 2010. Molecular comparison of Quebec and Newfoundland populations of the blackfly, *Simulium vittatum*, species complex. *Medical and Veterinary Entomology* 24: 214-217.
- Gersabek, E.F. and Merritt, R.W. 1979. The effect of physical factors on the colonization and relocation behavior of immature Black flies (Diptera: Simuliidae) *Environmental Entomology* 8: 34-39.
- Grenier, P. 1953. Simuliidae de France et d'Afrique du Nord. Editeur-libraire pour les sciences naturelles, Paris, 171 pages.
- Hunter, D.B., Rohner, C. and Currie, D.C. 1997. Mortality in fledgling great horned owls from black fly hematophaga and leucocytozoonosis. *Journal of Wildlife Diseases* 33: 486-491.
- Kozár, F. and Dávid, A.N. 1986. The unexpected northward migration of some species of insects in Central Europe and the climatic changes. *Anzeiger für Schädlingskunde* 59: 90-94.
- Krüger, A., Gelhaus, A. et Garms, R. 2000. Molecular identification and phylogeny of East African *Simulium damnosum s.l.* and their relationship with West African species of the complex (Diptera: Simuliidae). *Insect Molecular Biology* 9: 101-108.

- Krueger, A., Mustapha, M., Kalinga, A.K., Tambala, P.A.J., Post, R.J. and Maegga, B.T.A. 2006. Revision of the Ketaketa subcomplex of blackflies of the *Simulium damnosum* complex. *Medical and Veterinary Entomology* 20: 76-92.
- Laparé, R. 1980. Écologie des mouches noires (Diptera: Simuliidae) de LG2 Radisson (Territoire de la Baie James, Québec). Master's thesis in Environmental Sciences, Université du Québec à Trois-Rivières, 107 pages.
- LaRue, B., Gaudreau, C., Bagre, H.O. and Charpentier, G. 2009. Generalized structure and evolution of ITS1 and ITS2 in black flies (Diptera: Simuliidae). *Molecular Phylogenetics and Evolution* 53: 749-757.
- Lucas, P. and Hunter, F. 1999. Phenotypic plasticity in the labral fan of simuliid larvae (Diptera): effect of seston load on primary-ray number. *Canadian Journal of Zoology* 77: 1843-1849.
- Mackay, R.J. 1969. Aquatic insect communities of a small stream on Mont St-Hilaire, Québec. *Journal of the Fisheries Research Board of Canada* 26: 1157-1183.
- Malmqvist, B., Adler, P.H. and Strasevicius, D. 2004. Testing hypotheses on egg number and size in black flies (Diptera: Simuliidae). *Journal of Vector Ecology* 29: 248-256.

- McCreadie, J.W. and Adler, P.H. 2006. Ecoregions as predictors of lotic assemblages of blackflies (Diptera: Simuliidae). *Ecography* 29: 603-613.
- McCreadie, J.W. and Colbo, M.H. 1991. Spatial distribution patterns of larval cytotypes of the *Simulium venustum/ verecundum* complex (Diptera: Simuliidae) on the Avalon Peninsula, Newfoundland: factors associated with occurrence. *Canadian Journal of Zoology* 69: 2651-2659.
- McCreadie, J.W., Adler, P.H. and Colbo, M.H. 1995. Community structure of larval black flies (Diptera: Simuliidae) from the Avalon Peninsula, Newfoundland. *Annals of the Entomological Society of America* 88: 51-57.
- Mead, D.G., Howerth, E.W., Murphy, M.D., Gray, E.W., Noblet, R. and Stallknecht, D.E. 2004. Black fly involvement in the epidemic transmission of vesicular stomatitis New Jersey virus (*Rhabdoviridae: Vesiculovirus*). *Vector-Borne Zoonotic Diseases* 4: 351-359.
- Minhas, M.S., Brockhouse, C.L. and Adler, P.H. 2005. The black fly (Diptera: Simuliidae) fauna of Prince Edward Island, Canada. *Northeastern naturalist* 12: 67-76.
- Morales-Hojas, R. and Krueger, A. 2009. The species delimitation problem in the *Simulium damnosum* complex, blackfly vectors of onchocerciasis. *Medical and Veterinary Entomology* 23: 257-268.

- Moulton, J.K. and Adler, P.H. 1995. Revision of the *Simulium jenningsi* Species-Group (Diptera: Simuliidae). Transactions of the American Entomological Society 121: 1-57.
- Patterson, D.T., Westbrook, J.K., Joyce, R.J.V., Lingren, P.D. and Rogasik, J. 1999. *Weeds, Insects and Diseases*. Climatic Change. 43: 711-727.
- Plante, L. 2005. Le Nouvelliste: "Une rivière en Or". August 27 to September 1, 2005.  
Available online at  
[http://www.bvsm.ca/arret\\_du\\_flottage.htm#une%20riviere%20en%20or](http://www.bvsm.ca/arret_du_flottage.htm#une%20riviere%20en%20or). Accessed December 27, 2008.
- Rogers, D.J. and Packer, M.J. 1993. Vector-Borne diseases, models and global change. The Lancet 342: 1282-1284.
- Ross, D.H. and Craig, D.A. 1979. The seven larval instars of *Prosimulium mixtum* Syme and Davies and *P. fuscum* Syme and Davies (Diptera: Simuliidae). Canadian Journal of Zoology 57: 290-300.
- Rothfels, K.H. 1979. Cytotaxonomy of black flies (Simuliidae). Annual Review of Entomology 24: 507-539.



Rothfels, K.H. 1981. Cytotaxonomy: Principles and their application to some Northern Species-Complexes in *Simulium*. In "Blackflies. The future for biological methods in integrated control.", M. Laird ed., Academic Press, New York, pp.19-29.

Rothfels, K.H., Fereday, R. and Kaneps, A. 1978. A cytological description of sibling species of *Simulium venustum* and *S. verecundum* with standard maps for the subgenus *Simulium* Davies (Diptera). Canadian Journal of Zoology 56: 1110-1128.

Rubtsov, I.A. 1989. Blackflies (Simuliidae). Second Edition, Amerind Publishing Co. Pvt. Ltd., New Delhi. 1042 pages.

Shewell, G.E. 1957. Interim report on distributions of the blackflies obtained in the Northern Insect Survey. Contribution No. 3684, Entomology Division, Science Service, Department of Agriculture, Ottawa, Canada.

Shewell, G.E. 1958. Classification and distribution of arctic and subarctic Simuliidae. Proceeding 10<sup>th</sup> International Congress of Entomology, Montréal 1: 635-643.

Snyder, T.P. and Linton, M.C. 1983. Electrophoretic and morphological separation of *Prosimulium fuscum* and *P. mixtum* larvae (Diptera: Simuliidae). Canadian Entomologist 115: 81-87.

- Spironello, M., Hunter, F.F. and Craig, D.A. 2002. A cytological study of the Pacific black fly *Simulium cataractarum* (Diptera: Simuliidae). Canadian Journal of Zoology 80: 1810-1816.
- Stone, A. and Jamnback, H.A. 1955. The black flies of New York State (Diptera: Simuliidae). Bulletin N.Y. State Museum 349: 1-144.
- Tang, J., Pruess, K. and Unnasch, T.R. 1995. Genotyping North American black flies by means of mitochondrial ribosomal RNA sequences. Canadian Journal of Zoology 74: 39-46.
- Tang, J., Moulton, J.K., Pruess, K., Cupp, E.W. and Unnasch, T. 1998. Genetic variation in North American black flies in the subgenus *Psilopelmia* (*Simulium*: Diptera: Simuliidae). Canadian Journal of Zoology 76: 205-211.
- Twinn, C.R. 1933. The blackfly, *Simulium venustum* Say, and a protozoan disease of ducks Canadian Entomologist 65: 1-3.
- Twinn, C.R. 1936. The black flies of eastern Canada. Canadian Journal Research 14: 97-150.
- Winn, A.F. and Beaulieu, G. 1932. Part II. Diptera. Revised and supplemented. In Petch, C.E. and J.B. Maltais. A preliminary list of the insects of the Province of Quebec. Report of the Quebec Society for the Protection of Plants 24: 1-100.

- Wolfe, L.S. and Peterson, D.G. 1959. Black flies (Diptera: Simuliidae) of the forest of Quebec. *Canadian Journal of Zoology* 37: 137-159.
- Wood, D.M. 1978. Taxonomy of the Nearctic species of *Twinnia* and *Gymnopais* (Diptera: Simuliidae) and a discussion of the ancestry of Simuliidae. *The Canadian Entomologist* 110: 1297-1338.
- Wood, D.M., Peterson, B.V., Davies, D.M. and Gyorkos, H. 1963. The blackflies (Diptera: Simuliidae) of Ontario. Part II. Larval identification, with description and illustration. *Proceedings of the Entomological Society of Ontario* 93: 99-129.
- Wotton, R.S. 1976. The distribution of blackfly larvae (Diptera: Simuliidae) in upper Teesdale streams northern England. *Hydrobiologia* 51: 259-263.
- Xiong, B. and Kocher, T.D. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). *Genome* 34: 306-311.
- Zettler, F.A., Adler, P.H. and Mc Creadie, J.W. 1998. Factors influencing larval color in the *Simulium vittatum* complex (Diptera: Simuliidae). *Invertebrate Biology* 117: 245-252.
- Zhang, Y. 2005. Balancing food availability and hydrodynamic constraint: phenotypic plasticity and growth in *Simulium noelleri* blackfly larvae. *Oecologia* 147: 39-46.

## 2.9 Tables and Figures

**Table 2.1: List of species, source of identification, months, and sites from which black fly larvae were collected throughout Québec.**

Species	Identified by the key of	January	February	March	April	May	June	July	August	September
<i>Helodon</i> sp.	Adler et al. 2004	11	10, 24	10,14,16, 22	10,22	1,2,3,4,5,6,7, 8,10,11	8	9		
<i>Prosimulium</i> sp.	Adler et al. 2004	11	10, 24	10,14,16, 22	10,22	1,2,3,4,5,6,7, 8,10,11	8	9		
<i>Helodon</i> ( <i>Parahelodon</i> ) <i>decemarticulatus</i> , Twinn 1936	Adler et al. 2004		10	10,22	10,22	10				
<i>Helodon</i> ( <i>Parahelodon</i> ) <i>vernalis</i> , Shewell 1952	Adler et al. 2004		10	10,22	10,22	10				
<i>Prosimulium</i> <i>fontanum</i> , Syme & Davies 1958	Adler et al. 2004		10	10,22	10,22	10				
<i>Prosimulium fuscum</i> , Syme & Davies 1958	Adler et al. 2004		10	10,22	10,22	10				
<i>Prosimulium mixtum</i> , Syme & Davies 1958	Adler et al. 2004		10	10,22	10,22	10				
<i>Prosimulium magnum</i> complex, Dyar & Shannon 1927	Adler et al. 2004					10				
<i>Prosimulium</i> <i>multidentatum</i> , Twinn 1936	Adler et al. 2004		10	10,22	10,22	10				
<i>Twinnia tibblesi</i> , Stone & Jamnback 1955	Adler et al. 2004					1				
<i>Cnephia dacotensis</i> , Dyar & Shannon 1927	Wood et al. 1963		10,11,22	10,16	10	10,11	9	9		
<i>Metacnephia saileri</i> , Stone 1952	Adler et al. 2004; Currie, 1986					4				
<i>Simulium</i> ( <i>Boreosimulium</i> ) <i>annulus</i> , Lundstöm 1911	Adler et al. 2004			22	22	1, 4				
<i>Simulium</i> ( <i>Eusimulium</i> ) <i>aureum</i> , Fries 1924	Currie, 1986								14	
<i>Simulium</i> ( <i>Hemicnetha</i> ) <i>innoxium</i> , Comstock & Comstock 1895 (Identified as <i>S.</i> <i>pictipes</i> in Wood et al. 1963)	Wood et al. 1963									15
<i>Simulium</i> ( <i>Hemicnetha</i> ) <i>pictipes</i> , Hagen 1880 (Identified as <i>S. longistylatum</i> in Wood et al. 1963)	Currie, 1986; Wood et al. 1963						8, 9			
<i>Simulium</i> ( <i>Nervermannia</i> ) <i>quebecense</i> , Twinn 1936	Adler et al. 2004; Wood et al. 1963					10				
<i>Simulium</i> ( <i>Nervermannia</i> ) <i>silvestre</i> , Rubtsov 1956	Back, C. Comm. Pers.					10, 22		22		

<i>Simulium (Psilozia) vittatum</i> , Zetterstedt 1838	Wood et al. 1963	10, 22	10,16,22	10, 22	2,5,10	9,10,19,21	10,22	10,22, 24	10,12,22
<i>Simulium (Simulium) jenningsi</i> , Malloch 1914	Wood et al. 1963				22	21	22	22	12,15,22
<i>Simulium (Simulium) decorum</i> , Walker 1848	Wood et al. 1963				10,22		10,22, 23	10,24	10
<i>Simulium (Simulium) tuberosum</i> , Lundström 1911	Wood et al. 1963				5,10,22	8,9,18	10,22, 23	10,22,24	10,22
<i>Simulium (Simulium) venustum/ verecundum complex</i> , Say 1823	Wood et al. 1963				2,10,11, 22	8,9,10, 17,18, 19,20	10,22, 23	10,22,24	10,22
<i>Stegopterna mutata</i> , Malloch 1914 (originally named <i>Cnephia mutata</i> )	Wood et al. 1963	10,11,22	12,13,16, 22	10	1,4,5,6,7,10, 11				

Site	# station/site	Coordinates
1. Mont Lyall (Gaspésie)	2	48°47'N 66°05'W
2. Zec des Anses (Gaspésie)	2	48°20'N 64°43'W
3. Réserve faunique de Matane (Gaspésie)	3	48°44'N 66°55'W
4. Réserve Duchénier (Bas-Saint-Laurent)	4	48°08'N 68°38'W
5. Réserve faunique de Rimouski (Bas-Saint-Laurent)	3	48°03'N 68°15'W
6. Réserve faunique des Laurentides (Road 175; km 83, 188, 199) (Québec)	3	
7. Parc des Grands-Jardins (Charlevoix)	2	47°42'N 70°42'W
8. Réserve faunique Mastigouche (Mauricie)	4	46°40'N 73°14'W
9. Réserve faunique du Saint-Maurice (Mauricie)	5	46°56'N 73°05'W
10. Lac Souris (Mauricie)	1	46°34'N 72°58'W
11. Saint-Boniface (Mauricie)	1	46°31'N 72°51'W
12. Rivière Richelieu à Chambly (Montérégie)	1	45°25'N 73°13'W
13. Rougemont (Montérégie)	1	45°26'N 73°03'W
14. L'Assomption (Lanaudière)	1	45°50'N 73°25'W
15. Chute Mont-à-Peine (Lanaudière)	1	46°12'N 73°34'W
16. Saint-Hyppolite (Lac Croche) (Laurentide)	3	45°59'N 74°01'W
17. St-Gertrude (Centre du Québec)	1	46°18'N 72°16'W
18. Bécancour- Chemin Bouvreuil (Centre du Québec)	1	45°53'N 72°29'W
19. Rivière Yamaska Nord (Canton de l'est)	1	45°25'N 72°17'W
20. Bonsecours (Cantons de l'est)	1	45°25'N 72°17'W
21. Rivière Rideau (Ottawa)	1	45°22'N 75°41'W
22. Lac du Barrage (Mauricie)	1	46°29'N 72°57'W
23. Chute-à-Bull (Lanaudière)	1	46°17'N 73°46'W
24. Mont St-Hilaire (Montérégie)	2	45°32'N 73°09'W

**Table 2.2: Physical and chemical parameters of Lac Souris and Lac du Barrage.**

	Lac Souris	Lac du Barrage
Width (m)	6.9 (relatively constant)	8 (4 to 11.9)
Depth (cm)	19 (10 to 29)	15.2 (9 to 20)
Current velocity (m/s)	0.7	0.51
pH	6.2 to 6.9	5.8 to 7.1
Conductivity ( $\mu\text{S}/\text{cm}$ )	0.02	0.02
Streambed characteristics	mainly rocky	site 1 to 4 rocky; 5 to 6 gravelled with abundant vegetation*

\* (*Myosotis laxa* Lehm., *Galium palustre* L., *Sparganium chlorocarpum* Rydb,

*Hypericum ellipticum* Hook, *Callitriche heterophylla* Pursh, *Potamogeton pectinatus* L.,

*Najas flexilis* (Willd) Rostk and Schmidt and *Potamogeton gramineus* L.).

Table 2.3: Distribution and cytotypes for each species.

Species	Distribution**	Cytotypes**
<i>H. decemarticulatus</i>	Alaska, Canada and northern USA <sup>1</sup>	Two sibling species by Rothfels (1979) but contested by Adler et al. 2004.
<i>H. vernalis</i>	Eastern North America <sup>1</sup>	IIS-2 <sup>7</sup>
<i>P. fontanum</i>	Northeastern America <sup>1</sup>	16 Rothfels 1956 ( <i>P.</i> cytoform) <sup>6</sup> IIL-1 <sup>7</sup>
<i>P. fuscum</i>	Canada (from eastern Coast to Saskatchewan) and eastern USA <sup>1</sup>	1 Rothfels 1956 ( <i>hirtipes</i> cytoform) <sup>6</sup> IIL-1 <sup>7</sup>
<i>P. mixtum</i>	Eastern Canada and eastern USA <sup>1,6</sup>	2 Rothfels 1956 ( <i>hirtipes</i> cytoform) <sup>6</sup> IIL-1 <sup>7</sup>
<i>P. magnum</i> complex	Eastern USA <sup>1,6</sup>	IIS-1 <sup>7</sup>
<i>P. multidentatum</i>	Southeastern Canada and northeastern USA <sup>1,6</sup>	IIS-1 <sup>7</sup>
<i>T. tibblesi</i>	Northeastern Canada to Pennsylvania <sup>1,6</sup>	Fixed inversion: 1L-1, IIS-3; Rothfels and Freeman 1966, Rothfels 1979 <sup>6</sup>
<i>C. dacotensis</i>	USA and Canada <sup>1</sup>	
<i>M. saileri</i>	Holarctic <sup>1</sup>	IIS-2 Procnier 1982 (cytoform) <sup>6</sup> Tuk Procnier 1980 (cytoform; thesis) <sup>6</sup> IIL-3 + 4, IIL-3,5 Procnier 1982 (cytoform) <sup>6</sup>
<i>S. annulus</i>	Holarctic <sup>1</sup>	
<i>S. aureum</i>	Holarctic <sup>2</sup>	F Dunbar 1959 (cytoform) <sup>6</sup> A ( <i>Simulium bracteatum</i> ); Dunbar 1958 ( <i>aureum</i> cytoform) <sup>6</sup> C ( <i>Simulium exulatum</i> ); Dunbar 1959 ( <i>aureum</i> cytoform) <sup>6</sup> B ( <i>Simulium pilosum</i> ); Dunbar 1958 ( <i>aureum</i> cytoform) <sup>6</sup> D ( <i>Simulium violator</i> ); Dunbar 1959 ( <i>aureum</i> cytoform) <sup>6</sup>
<i>S. innoxium</i> (synonyme: <i>S. pictipes</i> )	USA and Canada <sup>1,6</sup>	B ( <i>Simulium innoxium</i> ); Bedo 1975 ( <i>pictipes</i> cytoform) <sup>6,8</sup>
<i>S. pictipes</i> (synonyme: <i>S. longistylatum</i> )	USA and Canada <sup>1,6</sup>	XoYo, Y <sub>1</sub> (IS-1) <sup>8</sup>
<i>S. quebecense</i>	Northeastern North America southward along the Appalachian <sup>1</sup>	Possibly more than one species (Adler et al. 2004)
<i>S. silvestre</i>	Holarctic <sup>1,6</sup>	IIL-19 for eastern population of North America (Adler et al. 2004)
<i>S. vittatum</i>	Transcontinental from arctic to 34th parallel <sup>1</sup>	IIL-1 ( <i>Simulium tribulatum</i> ); Pasternak 1964, Rothfels and Featherston 1981 <sup>6</sup> IS-7 ( <i>Simulium vittatum</i> ); Pasternak 1964, Rothfels and Featherston 1981 <sup>6</sup> XoYo Rothfels and Featherston 1981, Duncan et al. 2004, Minhas et al. 2005 <sup>9</sup> and *
<i>S. jenningsi</i>	Eastern USA and southeastern Canada <sup>1</sup>	Suggested as a complex <sup>10</sup>
<i>S. decorum</i>	Nearctic <sup>1</sup>	IIL-10 <sup>11</sup>
<i>S. tuberosum</i>	Holarctic <sup>3</sup>	CDE ( <i>Simulium appalachiense</i> ); Landau 1962 ( <i>tuberosum</i> cytoform) <sup>6</sup> CDE-2 ( <i>Simulium appalachiense</i> ); Mason 1982 ( <i>tuberosum</i> cytoform) <sup>6</sup> CDEM ( <i>Simulium appalachiense</i> ); Mason 1984 ( <i>tuberosum</i> cytoform) <sup>6</sup> FGH ( <i>Simulium conundrum</i> ); Landau 1962 ( <i>tuberosum</i> cytoform) <sup>6</sup> AB Landau 1962 (cytoform) <sup>6</sup> A Mason 1982 (cytoform) <sup>6</sup> Y <sub>2</sub> Mason 1982 (cytoform) <sup>6</sup> FG ( <i>Simulium vandalicum</i> ); Landau 1962 ( <i>tuberosum</i> cytoform) <sup>6</sup>

<i>S. venustum/</i> <i>verecundum</i> complex	Holarctic <sup>2</sup>	CC2 (Ca14; CTC4) ( <i>Simulium hematophilum</i> ); Rothfels et al 1978 ( <i>venustum</i> cytoform) <sup>6</sup>
		CC(gB)B ( <i>Simulium irritatum</i> ); Rothfels 1979 ( <i>venustum</i> cytoform) <sup>6</sup>
		CC3 ( <i>Simulium irritatum</i> ); Rothfels 1981 ( <i>venustum</i> cytoform) <sup>6</sup>
		A/C ( <i>Simulium molestum</i> ); Rothfels, Feraday and Kaneps 1978 ( <i>venustum</i> cytoform) <sup>6</sup>
		AC (gB)(AC) ( <i>Simulium piscicidium</i> ); Rothfels, Feraday and Kaneps 1978 ( <i>venustum</i> cytoform) <sup>6</sup>
		CH(gB) ( <i>Simulium tormentor</i> ); Rothfels 1979 ( <i>venustum</i> cytoform) <sup>6</sup>
		H/C ( <i>Simulium tormentor</i> ); Rothfels 1981 ( <i>venustum</i> cytoform) <sup>6</sup>
		EFG/C (CEFG, CEF) ( <i>Simulium truncatum</i> ); Rothfels et al. 1978 ( <i>venustum</i> cytoform) <sup>6</sup>
		EFG ( <i>Simulium venustum</i> ); Rothfels, Feraday and Kaneps 1978 (cytoform) <sup>6</sup>
		ACD ( <i>Simulium rostratum</i> ); Rothfels, Feraday and Kaneps 1978 ( <i>verecundum</i> cytoform) <sup>6</sup>
		A/C ( <i>verecundum</i> cytoform) <sup>11</sup>
CC ( <i>verecundum</i> cytoform) <sup>11</sup>		
<i>S. mutata</i>	Nearctic <sup>1,4,5</sup>	2n ( <i>Stegopterna diplomutata</i> ); Basrur 1957 (thesis) ( <i>mutata</i> cytoform) <sup>6</sup>
		diploïde ( <i>Stegopterna diplomutata</i> ); Basrur and Rothfels 1959 ( <i>mutata</i> cytoform) <sup>6</sup>
		3n Basrur 1957 (thesis) (cytoform) <sup>6</sup>
		triploïde Basrur and Rothfels 1959 (cytoform) <sup>6</sup>

\* Found in Iceland and in Northwest Atlantic islands, maybe possibly in Magdalen Islands.

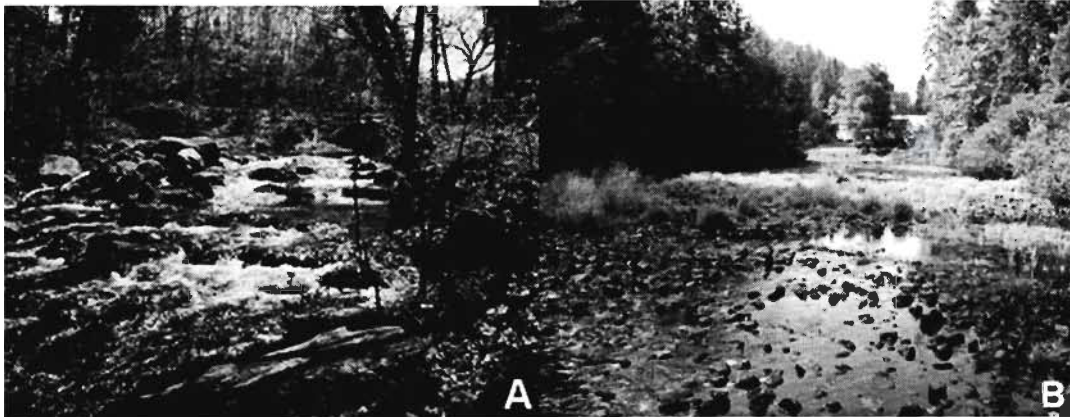
\*\* The references for distribution and cytotypes are the following; 1. Adler et al. 2004, 2. Grenier 1953, 3. Rubtsov 1989, 4. Back and Harper 1978, 5. Back and Harper 1979, 6. Adler and Crosskey 2009, 7. Rothfels 1979, 8. Bedo 1975, 9. Minhas et al. 2005, 10. Moulton and Adler 1995 and 11. Rothfels 1981.



**Table 2.4: Percentage comparison of each species found in February and in March/April of 2006 and 2008 for Lac Souris and Lac du Barrage.**

species	Lac Souris				Lac du Barrage			
	2 February 2006	7 February 2008	27 March 2006	10 April 2008	2 February 2006	7 February 2008	27 March 2006	10 April 2008
<i>Prosimulium</i> sp	69.3	96.4	71.2	83.3	95.3	82.5	95.4	96.7
<i>C. mutata</i>	23.6	0	23.8	10.0	2.8	5.3	3.3	3.3
<i>C. dacotensis</i>	1.2	0	1.0	0	0.7	0	0	0
<i>S. euryadmiculum</i>	0	0	0	0	0	0	1.0	0
<i>S. vittatum</i>	5.9	3.6	4.0	6.7	1.2	12.3	0.3	0
<b>Total number of larvae</b>	<b>254</b>	<b>50</b>	<b>359</b>	<b>50</b>	<b>253</b>	<b>50</b>	<b>302</b>	<b>50</b>

**Figure 2.1 : Study sites for temporal and spatial distribution A) Lac Souris and B) Lac du Barrage.**



**Figure 2.2 : Seasonal variation in species distributions in 2006 A) Lac Souris and B) Lac du Barrage.**  
 (X-axis: date of collection, Y-axis: number of larvae per species, Z-axis: species name)

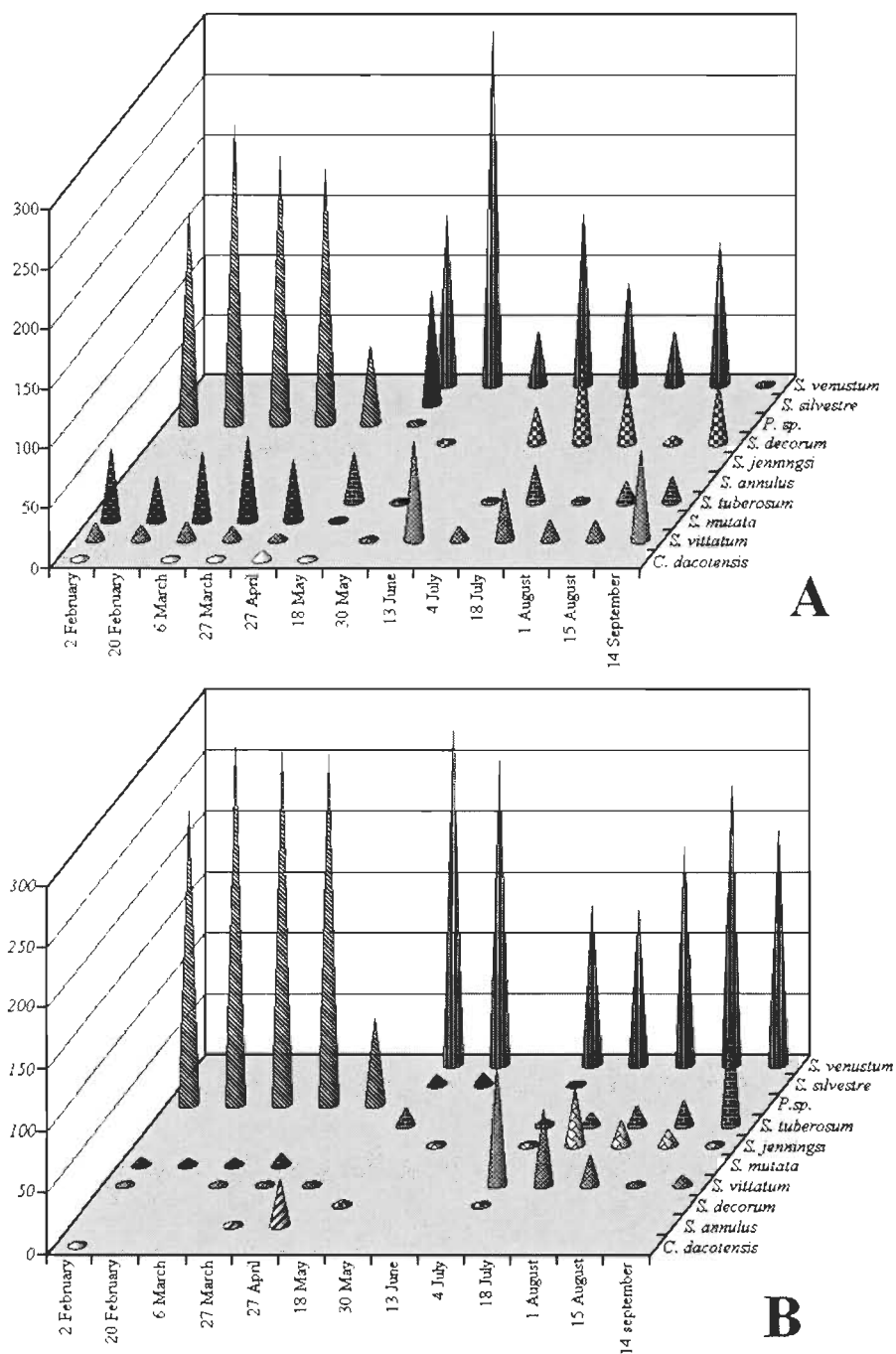


Figure 2.3 : Species abundance and distribution at each of the six stations of A) Lac Souris and B) Lac du Barrage. (X-axis: station number, Y-axis: number of larvae per species, Z-axis: species name)

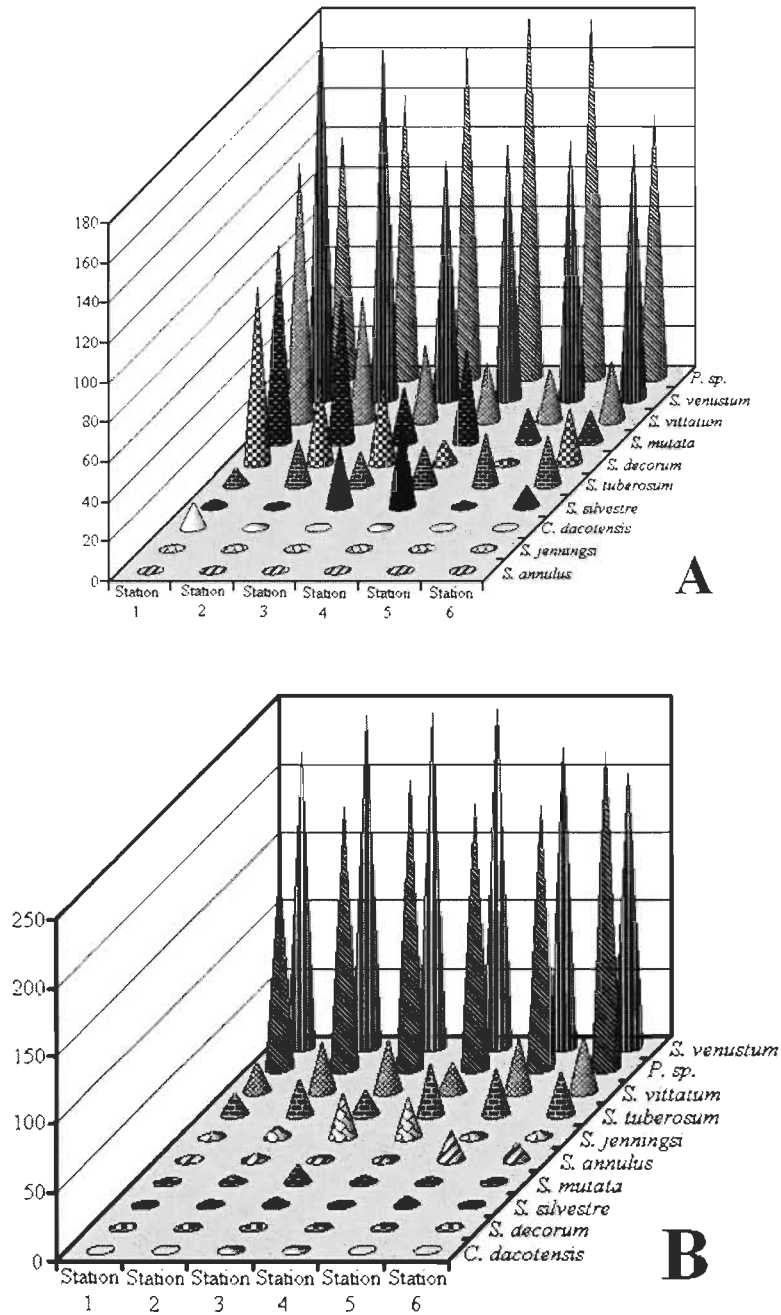
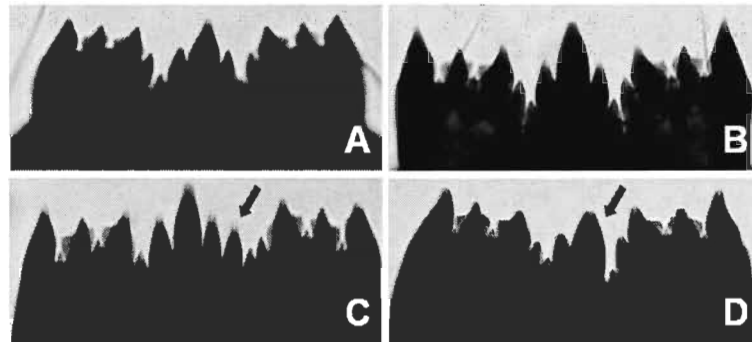


Figure 2.4: Hypostomal tooth patterns by which *Prosimulium* species were identified. A) *P. fuscum*; B) *P. fontanum*; C and D) Unknown. Arrow in figure C indicates location of an extra tooth; Arrow in figure D indicates locations of a missing tooth.



## **CHAPITRE III**

**« Generalized structure and evolution of ITS1 and ITS2 rDNA  
in black flies (Diptera: Simuliidae) »**

**Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies  
(Diptera : Simuliidae)**

Bernard LaRue <sup>1,2</sup>, Christine Gaudreau <sup>1</sup>, Hubert O. Bagre <sup>1</sup>, Guy Charpentier <sup>1</sup>

<sup>1</sup>Département de Chimie-biologie, Université du Québec à Trois-Rivières, 3351

Boulevard des Forges, Box 500, Trois-Rivières, Canada G9A5H7

<sup>2</sup> Corresponding author. Phone: (819) 376-5052. Fax: (819) 376-5084.

Bernard LaRue: [bernard.larue@uqtr.ca](mailto:bernard.larue@uqtr.ca)

Christine Gaudreau: [christine.gaudreau@uqtr.ca](mailto:christine.gaudreau@uqtr.ca)

Hubert O. Bagre: [obagreca@yahoo.fr](mailto:obagreca@yahoo.fr)

Guy Charpentier: [guy.charpentier@uqtr.ca](mailto:guy.charpentier@uqtr.ca)

**Key words: ITS1, ITS2, Simuliidae, Secondary structure, Evolution.**

### 3.1 RÉSUMÉ

Un échantillon de 15 espèces de mouches noires recouvrant 5 genres a permis de réaliser la première étude systématique de l'espaceur intergénique 1 (ITS1) de l'unité nucléaire de transcription de l'ADNr des Simuliidae. Alors que ITS1 se caractérise chez la tribu des Prosimuliini par une séquence conservée, sans répétitions internes et fortement structurée d'environ 490 nucléotides (nt), les Simuliini possèdent une version soit courte, d'au minimum 95 nt, soit de taille moyenne. Toutes ces versions ont un noyau commun de 39 nt constitué de 8 courts blocs consensus dispersés au milieu de séquences hautement variables. Cette variabilité s'avère d'une ampleur telle qu'elle interdit en général l'emploi d'ITS1 dans les études phylogéniques tout en fournissant potentiellement un outil performant en génétique des populations à cause du polymorphisme souvent observé. Une étude menée en parallèle sur ITS2, l'autre espaceur intergénique transcrit de l'ADNr, démontre que ce dernier a une longueur peu variable centrée autour de 270 nt et se replie en trois domaines ancrés par quatre régions conservées représentant environ 40% de sa séquence. Avec la structure secondaire servant de guide commun pour l'alignement des séquences entre espèces, nous avons obtenu via le modèle GTR une phylogénie qui illustre de façon convaincante la divergence fondamentale entre Simuliini et Prosimuliini. Cependant, le faible niveau de support observé pour plusieurs nœuds intermédiaires de cet arbre pourrait indiquer que les genres actuels de Simuliini seraient issus de divergences plus ou moins simultanées survenues dans le passé.



### 3.1.1 CONTRIBUTION DES AUTEURS DE L'ARTICLE

L'article a été rédigé par Bernard LaRue et révisé par Christine Gaudreau et Guy Charpentier. La partie expérimentale fut réalisée par Bernard LaRue, Christine Gaudreau et Hubert O. Bagre. Quant aux données provenant des logiciels associés à la bioinformatique, elles ont été obtenues par Christine Gaudreau. Ce manuscrit a été publié en 2009 dans le journal *Molecular Phylogenetics and Evolution*, volume 53, pages 749-757.

### 3.2 ABSTRACT

A sample of 15 Nearctic black fly species spread over five genera is used to perform the first systematic study of the internal transcribed spacer 1 (ITS1) from the nuclear rDNA transcription unit of Simuliidae. ITS1 from the Prosimuliini tribe is a conserved, repeat-free and highly structured sequence of about 490 nucleotides (nt), while Simuliini exhibit a medium-sized or short version, the latter minimally 95 nt long. All size versions possess a common 39 nt core made from eight short blocks interspersed among highly variable sequences. Conversely, that variability which generally excludes ITS1 from phylogenetic applications translates for many species into polymorphisms suggesting the general feasibility of ITS1-based population studies. We show in a parallel investigation that ITS2, the other rDNA transcribed spacer, is length-constrained around 270 nt and possesses a three-domain fold anchored by four conserved regions representing about 40% of the whole sequence. An alignment guided by this secondary structure leads to a phylogeny, derived through the GTR model, which

convincingly displays the basal divergence between Simuliini and Prosimuliini. However, the poorer support of some intermediate nodes could indicate rapid divergence events within *Simulium*.

### 3.3 INTRODUCTION

With approximately 2000 species spanning 26 genera (Crosskey and Howard, 2004), black flies (Diptera: Simuliidae) constitute a taxonomic challenge due to the general lack of distinctive characters. Molecular data (Bertone et al., 2008) indicate Thaumaleidae as the closest relatives to Simuliidae and suggest a divergence time of  $130 \pm 55$  million years (My). The earliest authentic fossils dating to late Jurassic (Kalugina, 1991) attest to the antiquity of the Simuliidae family and there are indications that some extant genera were already well emplaced during the following Cretaceous (Currie and Grimaldi, 2000; Jell and Duncan, 1986). Cladistics (Adler et al., 2004) and molecular analyses (Moulton, 2000) provide evidence for the basal position of the *Parasimulium* genus. All the remaining species, now designated as Simuliinae, are subdivided into two tribes (Grenier and Rageau, 1960; Currie, 1988), the holarctic Prosimuliini and the worldwide-distributed Simuliini, a scheme strongly supported by Moulton (2000).

The intergenic spacer ITS2 from the nuclear rDNA cluster, located on a single chromosome pair for most black fly species (Rothfels, 1979), constitutes a recent and valuable addition to the molecular toolbox used in phylogenetic studies of many groups of organisms. The concerted evolution of this multigene family usually ensures a high redundancy between most of the copies (Arnheim, 1983), although each part of the transcription unit changes at its own rate. ITS2 has a rather constant length and exhibits long sections that are essentially preserved within a given class of organisms. At the same time, much of the remaining sequence seemingly varies at a near-neutral rate

(Schlötterer et al., 1994) and comprises multiple indels, hence sequence alignment problems and possibly a lack of resolving power above the genus level. However, comparisons between yeast and vertebrates (Joseph et al., 1999) brought evidence for a common secondary structure core, which applies even to very distant groups like green algae and flowering plants (Mai and Coleman, 1997). In yeast, this fold occurs independently from the rest of the rDNA primary transcript, is an absolute functional requirement and contains key elements for pre-rRNA processing (Peculis and Greer, 1998; Van Nues et al., 1995). Once elucidated, the secondary structure guides the alignment step and allows the recovery of an otherwise hidden phylogenetic information. The usefulness of this approach was demonstrated in dipterans by Young and Coleman (2004) as they disproved an earlier *D. melanogaster* secondary structure proposal from Schlötterer et al. (1994), in addition to providing a phylogeny for drosophilids and true fruitflies in general. For black flies, about 40 *Simulium* species from Thailand were clearly resolved into several clades in spite of an alignment omitting secondary structure considerations (Thanwisai et al., 2006).

The ITS1 spacer located between the 18S and 5.8S genes is generally known to vary extensively in both sequence and length. In spite of some phylogenetic applications in other dipterans (Perrin et al., 2006; Toma et al., 2002), most studies of ITS1 in Simuliidae have concentrated on the onchocerciasis carriers from the *Simulium damnosum* complex. The observed length variation, from 240 to 400 nt, which enabled a partial distinction between members of the complex (Brockhouse et al., 1993; Krueger et al., 2006; Tang et al., 1996) was attributed to a variable number of repeats from two distinct sequence segments (Krüger et al., 2000; Morales-Hojas et al., 2002). At the

opposite, the sequence from *Simulium vittatum* encompasses only 103 nt (Miller et al., 1997) and exhibits no obvious repeats.

The present study of 15 Nearctic black fly species compares ITS1 structure between five genera, namely *Prosimulium*, *Twinnia*, *Stegopterna*, *Cnephia* and *Simulium*, to see whether it may lead to a unified model in spite of length and sequence differences. Following reports indicating intraspecific polymorphisms in *Simulium kilibanum* (Krüger et al., 2000), *Simulium ochraceum* s.l. (Rodriguez-Pérez et al., 2006) and *Simulium laciniatum* (Gaudreau et al., 2008), we also evaluate more broadly the extent of such occurrences prior to future applications in population genetics. Finally, we validate for the first time a secondary structure of ITS2 from Simuliidae and demonstrate its potential in helping to resolve deep phylogenetic relationships.

### 3.4 MATERIALS AND METHODS

#### 3.4.1. Sample collection

Details of black fly larvae collection, performed over a 3-year period in various streams from Québec and Eastern Ontario, will be published elsewhere as part of a field study. Specimens were transferred to 95% ethanol upon collection, carried back to the lab and stored at 4°C. Their identification under the binocular microscope used the keys of Wood et al. (1963) and Adler et al. (2004). *S. vittatum* specimens from Newfoundland and *Simulium (Hebridosimulium) laciniatum* larvae from Fijis were gifts from M. Colbo (Memorial University) and D. A. Craig (University of Alberta), respectively.

### *3.4.2. DNA amplification*

ITS1 amplification of DNA extracted and purified following St-Onge et al. (2008) was performed as described by Gaudreau et al. (2008). ITS2 primers were 5'-TTGAACGCATATCGCACTTCTTGC-3' (forward; -54 relative to ITS2 start) and 5'-CCTTATTAATATGCTTAAATTCAGGG-3' (reverse; +25 into the 28S gene), with reagents and thermocycling conditions set identical to those used for ITS1.

### *3.4.3. Purification and sequencing of PCR products*

As described by St-Onge et al. (2008), amplification products were cleaned by a short run on a 1% agarose gel or underwent a polyacrylamide-urea gel fractionation. Following extraction from the gel, they were directly sequenced at the automated sequencing facility of Laval University using standard dye-termination methods.

### *3.4.4. Sequence data analysis.*

ITS1 amplification (a less common occurrence for ITS2) often resulted into a mixture of two sequences, rarely even more. Sequences of pure components could be retrieved from sequencing electrophoregrams under any of the following circumstances: (i) two identical variants save for a single indel shifting them relative to each other, with a further sequence check through backward sequencing; (ii) amounts differing much between two components then readable as major versus minor from the peak profiles; (iii) prior polyacrylamide gel fractionation of heteroduplexes, with unique and distinct sequences often obtained by sequencing from one and then the other direction. Regarding this last occurrence, hybrid structures accumulating during the final PCR

plateau often had their apparent size increased by 20-50% relative to homoduplexes from which they were clearly resolved.

The beginning and the end of ITS1 were assigned through alignment of *Drosophila melanogaster* 18S and 5.8S rDNA sequences (Tautz et al., 1988) with their black fly counterparts. They may occasionally be shifted  $\pm 1-2$  nt relative to those used by some authors for dipterans in general. ITS2 borders were set according to Thanwisai et al. (2006) and sequence alignment was performed using ClustalW on Geneious version 3.8.5 software (Drummond et al., 2007) or a t-coffee online version available at <http://www.ebi.ac.uk/t-coffee/>. The few positions obviously misplaced relative to the consensus blocks (see Section 3) were realigned manually. Sequences were also analyzed for secondary structure through mfold (Zuker, 2003), after which interspecific comparisons provided the common folding pattern which defined the 2D-constrained alignment. The 2D-guided alignment that was used for ITS2 is available from the authors upon request.

PAUP\* 4.10b (Swofford, 2003) was used with Modeltest 3.4 (Posada and Crandall, 1998) to find the best model for nucleotide substitution. The best choice was GTR following the Aikake information criterion and likelihood ratio test. The maximum likelihood tree for each type of ITS2 alignment was inferred with PhyML 2.4.4 (Guindon and Gascuel, 2003). All PhyML parameters were set to default except for the choice of GTR and bootstrap replicates (500).

Most of the ITS1 and ITS2 sequences were deposited in Genbank as parts of the rDNA transcription unit under accession numbers EU779824, FJ436345-FJ436353, FJ437564-FJ437567 and FJ595476-FJ595477. While doing so, the single-nucleotide

polymorphisms (SNPs) noticed within several species were indicated as dual-base positions (save for *Prosimulium* 1, see Section 3), thus providing at a few locations a composite picture rather than unique sequences. Besides those already listed as part of the rDNA cluster, ITS1 indel variants from *Simulium jenningsi* and *Simulium tuberosum* were deposited separately as FJ807778 and FJ807775.

### 3.5 RESULTS

#### 3.5.1. Taxonomic considerations

Except for the unique *Twinnia tibblesi* specimen, ITS1 sequences originated from at least two larvae per species. With morphology as the prime criterion, the species attribution of several specimens was also corroborated by comparing their COI/COII (C. Gaudreau, unpublished) or ITS2/ITS1 sequences to independent Genbank entries, whenever available. Matches were: *Simulium vittatum* (U48383: ITS1 and ITS2, 100% similar), *Cnephia dacotensis* (AF425841: COI, 99.8%), *Simulium aureum* (AF083862, listed as *S. pilosum* = *S. aureum* “B”: COII, 100%), *S. jenningsi* (DQ284509: COII, 98.8%), *S. tuberosum* (DQ284515: COII, 98%) and *Simulium decorum* (DQ284522: COII, 99.6%).

The 23 larvae from the *Prosimulium* genus, notorious for its taxonomic difficulties, were identified chiefly through the hypostomal teeth pattern, apparently a discriminating feature among all the characters reported in Currie (1986) and Adler et al. (2004). Our sample then consisted of six morphological species, mostly *P. fuscum* and *P. mixtum*. However, this classification appeared incongruent with another one derived from COI and COII gene sequences since specimens from the same morphological unit



spread randomly between two major and quite divergent branches of the molecular tree (C. Gaudreau, unpublished). Each of these two groups also coincided with specific ITS1 and ITS2 structural features. Given the morphological pitfalls, we finally settled onto the molecular designations of *Prosimulium* 1 and 2. The average COII gene divergence (252 nt partial sequence) between the two “species” stood at 9.5%, with *P. dicentum* (AF083858, listed as *Helodon dicentum*; 13.4% divergent) showing up as the closest Genbank match. *Prosimulium* 1 and 2 sequences from Fig. 3.1 represent the ITS1 consensus obtained after neglecting the individual mutations found in some individuals specimens at a few polymorphic sites (see below).

Adler et al. (2004) have recently proposed not to consider any more *Simulium longistylatum* as a valid species and to move it to *Simulium pictipes*, while the former *S. pictipes* should shift to the already described *Simulium innoxium*. Given the absence of adequate morphological keys, we chose instead to follow Wood et al. (1963) by keeping the *S. longistylatum* and *S. pictipes* former designations.

### 3.5.2. ITS1 comes in two basic sizes

ITS1 had an average A + T content of 74.9%, with extremes of 65.3% and 83%, and its length varied about fivefold from 95 to 492 nt (Table 3.1). This base composition fitted into the range previously reported for other nematocerans with split 5.8S and 2S rRNA genes (Miller et al., 1997), but was much enriched in A + T content relative to the Culicidae having an undivided 5.8S rDNA gene. Broadly speaking, ITS1 size reflected the basal divergence between the two Simuliinae tribes, since a long version close to 500 nt distinguished the Prosimuliini (*Prosimulium* and *Twinnia*) from the middle-sized or small one seen in Simuliini (*Cnephia*, *Stegopterna* and *Simulium*).

Only eight mutations distinguished the *Prosimulium* 1 and 2 sequences, including a 4-nt insertion which typified the latter (Fig. 3.1). Neither the *Twinnia* nor the *Prosimulium* sequence showed internal repeats of a significant length. The excess size relative to Simuliini could have reflected the presence of a transposable unit, which was dismissed given the absence of direct terminal repeats and a Genbank blast that revealed no known homolog. Most likely, a long and internally originated ITS1 version characterized the *Twinnia* and *Prosimulium* common ancestor. When considering consecutive deletions as single mutations, ITS1 divergence was 1.63% between *Prosimulium* 1 and 2. In comparison, their partial COI and COII gene sequences differed by 7.7% over a 915 nt length, thus an age of about 3.3 My for their common ancestor as based on the currently admitted divergence rate of 2.3% My<sup>-1</sup> for these genes in arthropods (Brower, 1994). Standardization against COI/COII yielded for ITS1 a molecular clock rate of about 0.5% My<sup>-1</sup>, providing a date around 30 My for the *Prosimulium*-*Twinnia* divergence since their ITS1 sequences differed by 15-16 %.

*Stegopterna mutata* (292 nt) and *Simulium quebecense* (329 nt), the lone *Simulium* representative within this size class, exhibited a repeat-free middle-sized version. The 5' half of their sequences appeared moderately homologous to the Prosimuliini counterparts, but, starting at positions 220-240, most of the 3' half was deleted until a tight alignment with *Prosimulium* and *Twinnia* resumed around position 470 (Fig. 3.1). All the other Simuliini sequences represented the short (95-123 nt) ITS1 version, either directly or after subtracting the internal repeats which sometimes inflated their basic length. Such repeats occurred in *S. decorum*, *S. tuberosum* (Fig. 3.2) and *C. dacotensis* (155-159 nt, not shown). The major repeats appeared quite homologous between *S. tuberosum* and *S. decorum*, including quite similar locations for their start

and end points. Removing the extra copies of the duplicated elements resulted in sequences trimmed down to 113, 97 and 120 nt for *S. tuberosum*, *S. decorum* and *C. dacotensis*, respectively.

The manual search for a common sequence signature revealed a core of eight short and highly conserved blocks that appeared under a fixed order and altogether added up to only 39 nt (Fig. 3.3). Most of the interspecific ITS1 length variation originated from a 12-409 nt expansion segment intercalating the third and fourth blocks. Omitting the latter, sequences between conserved blocks had short and quite fixed lengths, but diverged considerably between species, which suggested them to act merely as anonymous stuffers ensuring the appropriate spacing of core elements. Given the extreme variability for the remainder of the sequence, chance alone could hardly explain the preservation of the 39 nt core, of a still undefined function. Sequences from other nematocerans, namely *Ablabesmya ramphae* (Chironomidae, U48384), *Culicoides variipennis* (Ceratopogonidae, U48380), *Dixella cornuta* (Dixidae, U48381), *Aedes punctor* (Culicidae, U48378) and *Eucorethra underwoodi* (Chaoboridae, U07981), all lacked this specific core array.

### 3.5.3. Intraspecific and genomic ITS1 variability

Only *S. annulus*, *S. pictipes*, *S. mutata* and *Twinnia* did not exhibit ITS1 sequence variation between and/or within individuals, a finding quite at odds with the intraspecific near-homogeneity reported for drosophilids (Young and Coleman, 2004). On a species-based average and considering each indel of any length as single mutations, these polymorphisms involved 1.8% of nucleotide positions. Although the minute sampling size for most species likely led to some undervaluation, this figure

nevertheless amounted at most to one tenth the level of interspecific variation found within *Simulium*.

Sequence variants from the short ITS1 often combined base substitutions with indels or microsatellites, as illustrated by Fig. 3.2B. This variability occurred between individuals, between clusters of presumably heterozygous specimens (as surmised from the 1:1 abundance ratio of two variants) and at times within clusters as the presence in a given specimen of more than two variants or a ratio far from unity indicated internal heterogeneity. This applied for instance to four *S. aureum* variants found under equimolar ratios in two out of three specimens. Sequences 1 and 2 from *S. tuberosum* appeared once under the 1:1 proportion hinting at the heterozygous state, but a second specimen had a 4:1 ratio for sequences 3 and 4. Similarly, two indel variants from *C. dacotensis* (not illustrated) showed up in distinct individuals as 1:1 and 4:1 mixtures, respectively. Last, *S. longistylatum* variants were distinguished by an indel of four consecutive A residues and, further downstream, by an expansion segment filled by an  $A_MCA_N$  microsatellite. In an extreme case, data from one specimen (sequences 5-6, Fig. 3.2B) suggested a heterozygous state for the  $A_4$  deletion combined with a 2:1:1:1 mixture of  $A_{11}CA_{8,9,10,11}$  variants, for a possible total of eight distinct sequences within the same genome.

In 18 sequences out of 19 from *Prosimulium* 1, one to five positions showed two mixed bases or, exceptionally, a single-base indel. With an average of 2.1 mixed-identity sites per sequence, thirteen polymorphic sites were recorded overall, most of them clustered within a 160 nt length span (Fig. 3.1). The pool of sequencing electrophoregrams was most parsimoniously interpreted as fourteen unique variants coexisting in the population and four of them, including the consensus sequence,

represented about 80% of the total on a composition basis (Fig. 3.2B). Among the remaining ten (not shown), most were found as a minor component in a single specimen.

#### 3.5.4. A secondary structure for ITS1?

The average A to T ratio of 1.9 for the short ITS1 version, already quite A + T rich, impeded folding opportunities and 2D modeling generally revealed  $\leq 35\%$  of paired bases, quite the level seen in random sequences of identical base composition. The low free energy of the folds, usually  $-3.5$  to  $-4.5$  kcal mol<sup>-1</sup>, almost implied that the short ITS1 was functionally constrained for excluding secondary structure elements. The medium-length versions from *S. quebecense* and *S. mutata* generated a diversity of thermodynamically equivalent solutions quite unrelated to each other or to domain I from *Prosimuliini* (see below). For both, the incorporation of 50-55% of the sequence in paired stems indicated some potential for acquiring a yet to be defined fold.

The long ITS1 exhibited two large domains having a well balanced A to T ratio and conserved between *Twinnia* and *Prosimulium* (Fig. 3.4). The 3'-located domain II consisted of a long stem, including several small bulges, which anchored a terminal bifurcation. In addition to a solution shared by most *mfold* predictions, isomorphy considerations strongly supported this model. Most of the interspecific differences occurred in common bulges or loops (10 instances) and as concerted (5) or hemi-compensatory (3) base substitutions, the so-called CBCs or hemi-CBCs, which preserved the stem pairing. The terminal bifurcation and its adjacent sequences, which concentrated many such changes, were further scrutinized with respect to mutation pattern. As shown by the inset to Fig. 3.4, nearly all the mutations either preserved or

even reinforced the secondary structure. This was most striking for several minor *Prosimulium* 1 variants, since their identities had often been deduced from mixed-sequence profiles rather than formally proven. Further upstream, domain I appeared as a long stem-loop with a *Prosimulium*-specific lateral arm due to a 21-nt insertion relative to *Twinnia*. This insertion was accommodated without disturbing the remainder of the structure since it bulged from a region shaped as an internal loop in *Twinnia*. In spite of fewer available CBCs, the criteria used for validating domain II structure applied to domain I as well. Finally, the leftover sequences from both termini and from the short linker connecting domains I and II mimicked the short ITS1 in showing an A content over 50% and a very poor folding potential. This “A bubble” exhibited a stem-loop next to the CATCA consensus block (a 2D feature fitting all black fly sequences) and a possible pairing between the conserved 5' TTATTT and 3' AATAA termini.

#### 3.5.5. ITS2: a secondary structure

ITS2 spanned a narrow size range, typically  $273 \pm 13$  nt except for *S. aureum* with only 246 nt (Table 3.1). The A + T content averaged 76.3% and, contrary to ITS1, the A to T ratio always stayed very close to unity. Except for *Prosimulium* 1 (7 larvae) and *S. annulus* (2), sequences came from a single specimen per species and yet revealed for about half of them some genomic variability, usually limited to one or two polymorphic sites. Indels occurred exclusively in *S. jenningsi* (single nt deleted) and *S. tuberosum* (three scattered deletions, one nt each).

Comparisons of 2D models revealed a common fold of three stem-loop domains (Fig. 3.5A). Support for stem I and stem II structures came from several CBC and hemi-

CBC locations, including the demonstrative pattern seen for near-twin sequences such as intraspecific variants. Among fourteen mutations of the latter type, two resulted into CBCs (case 1, Fig. 3.5B), six in hemi-CBCs (case 2), two in a pairing switch increasing a helix size (case 3) and three occurred in unpaired regions, while a single one created a helical mismatch. The tips of the stem-loops I and II held species-distinctive sequences while the roots consisted of three conserved regions that provided a constant framework for 2D folding. Incidentally, these regions and the next two consensus blocks did also appear in the data from Thanwisai et al. (2006) although their structural implications were missed.

The more irregular domain III amounted to 50-58% of the total length.

Sequences at the low half of stem III varied widely except for pairs of closely related species, which left only short and few conserved helical segments to use for proper sequence alignment. Stem III frequently started three residues beyond the UAUGAA tract ending the third conserved block, but a 3' shift by four more residues was noticed in the shorter *S. aureum* sequence which also showed the deletion of a nearby 5-6 bp helix. The 3' end of the stem most often fell one or two nt ahead of the conserved 3' terminus, of which the first third was deleted in the unusual *S. laciniatum* sequence. The top of stem III consisted on its 5' side of a conserved 32 nt run having a CUU terminus close to or, more often, initiating a terminal loop. Its 3' counterpart was poorly conserved due to bulges intercalating short helices that often shifted positions between species. The 3' side of the whole stem III comprised on the average 14% more residues than its 5' partner. The *S. tuberosum* sequence provided a good example of this with two asymmetric bulges made from oligoU runs and a third one essentially from an oligoA stretch (Fig. 3.5A).

The 3' terminus comprised 18 near-invariant residues unable to fold back on themselves for more than three consecutive base-pairs. This absence of the small fourth domain shared by most groups of organisms could not result from a wrongly deduced helix III as modeling the 3' terminus together with nearby 5' sequences gave partial folds that varied between species in addition to forbidding the shortened helix III to re-form a common structure. We instead noticed for all nematoceran sequences a possible 6-7 bp stem due to the association of the 5' and 3' ends, as earlier proposed for green algae and flowering plants (Mai and Coleman, 1997). The boxed conformation of Fig. 3.5A appeared for Simuliidae the most plausible between two alternate structures since it could uniquely fit all sequences while also generating an uninterrupted helix.

#### *3.5.6. An ITS2-derived phylogeny*

Besides the high A + T content bias, the presence of strongly conserved regions restricting the number of informative positions could potentially impede phylogenetic inferences drawn from ITS2. The location of many such positions in the low half of the highly variable stem III did not entirely rule out alignment schemes differing slightly from the actual one, contrarily to the reliable guidance provided by the 5' consensus sequence from the upper part. Consequently, two slightly distinct alignments were used, with the relevant phylogenies to be later compared. The first was wholly constrained by the secondary structure, except for the three terminal loops being computer-aligned, and the other, of mixed type, aligned the low half of stem III in a purely sequence-dependent fashion. Finally, no outgroup sequence could be used due to the lack of a clear sequence homology between Simuliidae and other nematocerans



Using the 2D-assisted alignment, a phylogenetic tree was generated through the GTR model (Fig. 3.6), with identical results from HKY save for bootstrap values that differed slightly for the poorly supported branches. A mixed alignment produced the same tree topology with both models, except that *S. annulus* appeared instead as basal to *Simulium* or clustered with *S. laciniatum* (data not shown). The latter, a Fijian *Hebridosimulium* representative, was originally intended as an internal outgroup to check the cohesion of the *Simulium* clade. Removing the problematic *S. annulus* sequence increased bootstrap values to 60% or more for branches that previously showed lower support. All trees enabled a clear and expected distinction between Prosimuliini and Simuliini and indicated a *C. dacotensis/S. mutata* group diverging early from *Simulium*, which clearly stood out as a monophyletic clade. The data also recovered *S. jenningsi*, *S. venustum* and *S. decorum* in a single clade appearing as the nearest neighbor to *S. tuberosum*. *S. pictipes* and *S. longistylatum* were very closely related, with *S. vittatum* a more distant affiliate. A weakly supported group consisting of *S. aureum* and *S. quebecense* was also noticed. As far as we know, none of these groupings had been formulated earlier as phylogenetic relationships. We observed disproportionate branch lengths for *Simulium*, possibly because of the larger diversity of its representatives relative to other genera depicted as orphan species. This could also indicate that mutation saturation was seeping in at intermediate and basal tree levels.

For comparison purposes, Fig. 3.6 included the GTR tree inferred from an alignment based on sequence-only. A weaker support for the basal Prosimuliini divergence and for *Simulium* as a distinct clade indicated that neglecting secondary structure considerations reduced the limit depth of phylogenetic resolution. A few other differences were noticed, as the early divergence of *S. mutata* from the other Simuliini,

still a plausible scheme according to Adler et al. (2004) and Moulton (2003). Finally, *S. laciniatum* was oddly located as basal to the *S. tuberosum* branch. This anomaly likely originated from upstream sequences which, contrarily to secondary structure indications, were wrongly shifted 3' by the alignment program into the void left by the earlier mentioned *S. laciniatum* deletion.

### 3.6. DISCUSSION

The structural preservation of the long ITS1 highlights strong functional constraints likely pertaining to elaborate pre-rRNA processing mechanisms. Its low mutation rate and the knowledge of the secondary structure open the way for deciphering the phylogeny of Prosimuliini, still in doubt with respect to the branching of *Helodon*, *Urosimulium* and *Gymnopais*. Given the remnants of the long version still seen in some Simuliini, the parallel with the fundamental divergence within Simuliinae suggests a long version as the Simuliidae archetype, a view which data from the basal *Parasimulium* could eventually corroborate. The virtual disappearance of the 3' half of the expansion segment and the poor quality of the alignment for the 5' half likely identify the intermediate versions of *S. mutata* and *S. quebecense* as a molecular relic that finally evolved under a relaxed selection into the minimal structure now seen in a majority of modern Simuliini. Sampling other species could hopefully uncover other instances of the intermediate version and reveal details about the history of the size decrease, possibly due to a series of paraphyletic events.

Although the short ITS1 can provide few phylogenetic clues, its high level of interspecific variation indicates a good potential for molecular taxonomy. This is

exemplified by *S. longistylatum* and *S. pictipes*, which are related through a common ancestor only 2.2 My old according to a 5.1% divergence between their COI/COII sequences (Gaudreau et al., unpublished). ITS1 sequences from these species show the usual consensus blocks and a singular landmark of about 20 consecutive A residues filling the expansion segment. Once these elements are removed, leftover sequences aligned at a high cost of six indels still differ by 25%, a figure five times higher than for COI/COII and well above the level of ITS1 intraspecific variation. In other words, the ITS1 marker provides a sharp taxonomic signature, which view could be put to the ultimate test with siblings, or morphological twins distinguished only through their karyotypes. Duplications also appear quite sensitive with respect to evolutionary depth since the sequence from *S. decorum* shows for instance a partially re-duplicated structure while the moderately related *S. venustum* and *S. jenningsi* exhibit no duplication at all. A still subtler distinction was seen with length variations between members of the *S. damnosum* s.l. complex (Krüger et al., 2000). These examples emphasize the major role of indels in shaping the short ITS1 and it may be speculated that the homogenizing mechanisms acting selectively to preserve the structure of 18S, 28S and to a lesser degree ITS2 rDNA have the opposite effect on ITS1 due to relaxed constraints. In addition to taxonomic applications, ITS1 shows a significant potential in population studies due to a wide distribution of intraspecific variation. Complications could at times arise from rDNA cluster heterogeneity as seen for the multiple microsatellites of *S. longistylatum*. Along the same line of thought, the multiplicity of *Prosimulium* 1 variants suggests that some species undergo a fast rDNA cluster turnover.

As for other taxonomic groups, this investigation shows the resolving power of ITS2 to extend well beyond the genus level once sequence alignment incorporates the

now available 2D model. While still far from a fully documented phylogeny, our sampling of species spread over few, but quite divergent genera allows for tracking back remote events like the basal split within Simuliinae and outlines several *Simulium* clades. Whether a similar resolution can be maintained upon adding more species and genera to the tree remains to be seen. Due to a distinct coverage of species, our results cannot be directly compared to those from Thanwisai et al. (2006), limited to *Simulium*, and even less to the large number of genera studied by Moulton (2003). The weak support of intermediate tree levels reiterates the inability of a single molecular tool to fully decipher the entire evolutionary history of Simuliidae, which possibly involves a rapid divergence of genera at the end of the Mesozoic era. This task would require a combination of probes attuned to various time scales, from COI and COII which efficiently discriminate between closely related species and monospecific populations (Carew et al., 2007; Finn and Adler, 2006; Joy and Conn, 2001; Joy et al., 2007; Nolan et al., 2007; Pruess et al., 2000) to 28S rDNA which resolves basal relationships. The critical in-between period remains quite problematic since the nuclear genes studied so far, such as EF-1 $\alpha$ , DDC, PEPCK (Moulton, 2000) and CAD (Bertone et al., 2008), often evolve too fast and provide too few informative positions to enable a satisfactory resolution of intermediate nodes. The histone gene cluster is now evaluated as a candidate to fill the gap. The five genes cumulate about 700 codons and, contrary to ITS2, provide a straightforward sequence alignment. Moreover, they evolve mostly by synonymous third position mutations at the slower pace seen in other multigene families such as 18S and 28S rDNA, which would supply the right clock rate to resolve events from the critical period.

### 3.7 ACKNOWLEDGEMENTS

We acknowledge the research and student support provided to C. Gaudreau and H.O. Bagre by the Université du Québec à Trois-Rivières. We also thank M. Colbo and D.A. Craig for gifts of *S. vittatum* and *S. laciniatum* specimens, respectively.

### 3.8 REFERENCES

- Adler, P.H., Currie, D.C., Wood, D.M., 2004. The blackflies of North America. Cornell University Press, Ithaca, NY.
- Arnheim, N., 1983. Concerted evolution of multigene families. In: Evolution of genes and proteins (M. Nei and R.K. Koehn eds.). Sinauer Associates, Boston, MA, pp. 38-61.
- Bertone, M.A., Courtney, G.W., Wiegman, B.M., 2008. Phylogenetics and temporal diversification of the earliest true flies (Insecta; Diptera) based on multiple nuclear genes. Syst. Entomol. 33, 668-687.
- Brockhouse, C.L., Vajime, C.G., Marin, R., Tanguay, R.M., 1993. Molecular identification of onchocerciasis vector sibling species in black flies (Diptera: Simuliidae). Biochem. Biophys. Res. Commun. 194, 628-634.

Brower, A.B.Z., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91, 6491-6495.

Carew, M.E., Pettigrove, V., Cox, R.L., Hoffmann, A.A., 2007. DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae). J. N. Amer. Benthol. Soc. 26, 586-599.

Crosskey, R.W., Howard, T.M., 2004. A revised taxonomic and geographical inventory of world blackflies (Diptera: Simuliidae). The Natural History Museum, London. <http://www.nhm.ac.uk/research-curation/projects/blackflies>.

Currie, D.C., 1986. An annotated list of and keys to the immature black flies of Alberta (Diptera: Simuliidae). Mem. Entomol. Soc. Can. 134, 1-90.

Currie, D.C., 1988. Phylogeny of Primitive Simuliidae (Insecta:Diptera: Culicimorpha). PhD dissertation, University of Alberta, Edmonton, Canada.

Currie, D.C., Grimaldi, D., 2000. A new black fly (Diptera: Simuliidae) genus from mid Cretaceous (Turonian) amber of New Jersey. In D. Grimaldi (ed.), Studies on fossils in amber, with particular reference to the Cretaceous of New Jersey. Backhuis Publishers, Leiden, The Netherlands, pp. 473-485.

Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., Wilson, A., 2007. Geneious V3.6, available from <http://www.geneious.com/>

Finn, D.S., Adler, P.H., 2006. Population genetic structure of a rare high-elevation black fly, *Metacnephia coloradensis*, occupying Colorado lake outlet streams. *Freshw. Biol.* 51, 2240-2251.

Gaudreau, C., LaRue, B., Charbonneau, V., Charpentier, G., Craig, D.A., 2008. Recent origins and genetic fragmentation of *Simulium* (*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid. *Invertebr. Syst.* 22, 555-562.

Grenier, P., Rageau, J., 1960. Remarques sur la classification des Simuliidae. *Bull. Soc. Pathol. Exot.* 53, 727-742.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696-704.

Gutell, R.R., Weiser, B., Woese, C.R., Noller, H.F., 1985. Comparative anatomy of 16-S-like ribosomal RNA. *Progr. Nucl. Acids Res. Mol. Biol.* 32, 155-215.

Jell, P.A., Duncan, P.M., 1986. Invertebrates, mainly insects, from the freshwater, Lower Cretaceous, Koonwarra Fossil Bed (Korrumburra Group), South Gippsland, Victoria. *Mem. Assoc. Austral. Paleontol.* 3, 111-205.

Joseph, N., Krauskopf, E., Vera, M.I., Michot, B., 1999. Ribosomal internal spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.* 27, 4533-4540.

Joy, D.A., Conn, J.E., 2001. Molecular and morphological phylogenetic analysis of an insular radiation in Pacific black flies (*Simulium*). *Syst. Biol.* 50, 18-38.

Joy, D.A., Craig, D.A., Conn, J.E., 2007. Genetic variation tracks ecological segregation in Pacific island black flies. *Heredity.* 99, 452-459.

Kalugina, N.S., 1991. New Mesozoic Simuliidae and Leptoconopidae and the origins of blood-sucking in lower dipterans. *Paleontol. Zh.* 1, 69-80.

Krüger, A., Gelhaus, A., Garms, R., 2000. Molecular identification and phylogeny of East African *Simulium damnosum* s.l. and their relationship with West African species of the complex (Diptera: Simuliidae). *Insect Mol. Biol.* 9, 101-108.

Krueger, A., Kalinga, A.K., Kibweja, A.M., Mwaikonyole, A., Maegga, B.T.A., 2006. Cytogenetic and PCR-based identification of *S. damnosum* 'Nkusi J' as the anthropophilic blackfly in the Uluguru onchocerciasis focus in Tanzania. *Trop. Med. Int. Health* 11, 1066-1074.



Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.* 44, 258-271.

Miller, B.R., Crabtree, M.B., Savage, H.M., 1997. Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S sequences (Diptera: Nematocera). *Insect Mol. Biol.* 8, 105-114.

Morales-Hojas, R., Post, R.J., Wilson, M.D., Cheke, R.A., 2002. Completion of the sequence of the nuclear ribosomal DNA subunit of *Simulium sanctipauli*, with descriptions of the 18S, 28S genes and the IGS. *Med. Vet. Entomol.* 16, 386-394.

Moulton, J.K., 2000. Molecular sequence data resolves basal divergences within Simuliidae (Diptera). *Syst. Entomol.* 25, 95-113.

Moulton, J.K., 2003. Can the current molecular arsenal adequately track rapid divergence events within Simuliidae (Diptera)? *Mol. Phylogenet. Evol.* 27, 45-57.

Nolan, D.V., Carpenter, S., Barber, J., Mellor, P.S., Dallas, J.F., Mordue (Luntz), A.J., Piertney, S.B., 2007. Rapid diagnostic PCR assays for members of the *Culicoides obsoletus* and *Culicoides pulicaris* species complexes, implicated vectors of bluetongue virus in Europe. *Vet. Microbiol.* 124, 82-94.

Peculis, B.A., Greer, C.L., 1998. The structure of the ITS2 proximal stem is required for pre-rRNA processing in yeast. *RNA* 4, 1610-1622.

Perrin, A., Cêtre-Sossah, C., Mathieu, B., Baldet, T., Delecolle, J.-C., Albina, E., 2006. Phylogenetic analysis of *Culicoides* species from France based on ITS<sub>1</sub>-rDNA sequences. *Med. Vet. Entomol.* 20, 219-228.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformat.* 14, 817-818.

Pruess, K.P., Adams, B.J., Parsons, T.J., Zhu, X., Powers, T.O., 2000. Utility of the mitochondrial cytochrome oxidase II gene for resolving relationships among black flies (Diptera: Simuliidae). *Mol. Phylogenet. Evol.* 16, 286-295.

Rodriguez-Pérez, M.A., Núñez-Gonzalez, C.A., Lizarazo-Ortega, C., Sanchez-Varela, A., Wooten, M.C., Unnasch, T.R., 2006. Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH dehydrogenase subunit 4 mitochondrial genes of the onchocerciasis vector *Simulium ochraceum*. *J. Med. Entomol.* 43, 1-6.

Rothfels, K.H., 1979. Cytotaxonomy of black flies (Simuliidae). *Ann. Rev. Entomol.* 24, 507-539.

Schlötterer, C., Hauser, M.T., von Haeseler, A., Tautz, D., 1994. Comparative evolutionary analysis of rDNA regions in drosophila. *Mol. Biol. Evol.* 11, 513-522.

St-Onge, M., LaRue, B., Charpentier, G., 2008. A molecular revision of the taxonomic status of mermithid parasites of black flies from Quebec (Canada). *J. Invertebr. Pathol.* 98, 299-306.

Swofford, D.L., 2003. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods) version 4. Sinauer Associates, Boston, MA.

Tang, J., Toè, L., Back, C., Unnasch, R., 1996. Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex. *Mol. Biol. Evol.* 13, 244-252.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., Dover, G.A., 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5, 366-376.

Thanwisai, A., Kuvangkadilok, C., Baimai, V., 2006. Molecular phylogeny of black flies (Diptera: Simuliidae) from Thailand, using ITS2 rDNA. *Genetica* 128, 177-204.

Toma, T., Miyagi, I., Crabtree, M.B., Miller, B.R., 2002. Investigation of the *Aedes (Stegomyia) flavopictus* complex (Diptera: Culicidae) in Japan by sequence

analysis of the internal transcribed spacers of ribosomal DNA. *J. Med. Entomol.* 39, 461-468.

Van Nues, R.W., Rientjes, J.M.J., Morré, S.A., Mollee, E., Planta, R.J., 1995. Evolutionary conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. *J. Mol. Biol.* 250, 24-36.

Wood, D.M., Peterson, B.V., Davies, D.M., Gyorkos, H., 1963. The black flies (Diptera: Simuliidae) of Ontario. Part II. Larval identification, with description and illustration. *Proc. Entomol. Soc. Ont.* 93, 99-129.

Young, I., Coleman, A.W., 2004. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Mol. Phylogenet. Evol.* 30, 236-242.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 13, 3406-3415.

## 3.9 TABLE AND FIGURES

Species <sup>A</sup>	ITS1		ITS2	
	AT% <sup>B</sup>	Length, nt	AT% <sup>B</sup>	Length, nt
<i>Cnephia dacotensis</i> (2)	80.5	155-159	80.5	263
<i>Prosimulium</i> 1 (19)	71.0	487-490	75.5	274
<i>Prosimulium</i> 2 (4)	70.9	491-492	75.0	272
<i>Simulium annulus</i> (2)	73.6	123	74.0	271
<i>Simulium aureum</i> (3)	73.5	95-100	73.0	246
<i>Simulium decorum</i> (2)	71.4	182	77.6	273
<i>Simulium jenningsi</i> (1)	76.2	98-100	75.5	265-266
<i>Simulium longistylatum</i> (5)	83.0	100-110	80.8	276
<i>Simulium pictipes</i> (4)	81.7	115	84.3	278
<i>Simulium quebecense</i> (2)	65.3	329	70.0	272
<i>Simulium tuberosum</i> (2)	79.8	206-215	70.4	260-263
<i>Simulium venustum</i> (2)	76.2	102-105	78.1	265
<i>Simulium vittatum</i> (24)	77.6	103-105	75.5	259
<i>Stegopterna mutata</i> (2)	71.9	292	76.5	260
<i>Twinnia tibblesi</i> (1)	70.7	481	78.0	286
Average	74.9	255.5	76.3	269.8
SD	5.0	144.6	3.8	10.4

<sup>a</sup> *S. mutata* formerly known as *Cnephia mutata*. *S. annulus*: in synonymy with *S. euryadminiculum* (Adler et al., 2004)

<sup>b</sup> Value for longest variant if more than one found.

**Table 3.1**

AT% and length for ITS1 and ITS2 in 15 black fly species from eastern Canada.

In parentheses: number of specimens with known ITS1 sequences. Except for

*Prosimulium* 1 (7 specimens) and *S. annulus* (2), ITS2 sequences were obtained from a single specimen taken out of the previous ITS1 pool.

```

                20;                40;                60;                80;                100;
1. CTA--TTATTTT-CTCAA-----ACAAACAAAGAGTT-AAGAAAAAAAAAAAAACAAAATATAAAGTCGTAACAA--AATATCT---TGCTTACATGAT
2. CTA--TTATTTT-CTCAA-----ACAAAAAAGATTTTAAAGAAACAA-----AAA---AAAGTCGGAATAT---AATATCT---TACTTACATGGT
3. CTA--TTATTTT-CTCAA-----ACAAAAAAGATTTTAAAGAAACAA-----TACAAAA---AAAGTCGGAATAA---AATATCT---TACTTACATGGT
4. CTATATTATTTT-CACAAAATCAAGAAAACAAAGAGTT-AAGAACAAAA-----AAGTCGGATGAAGTGTAAATCAAGCTATTGTCATGGT
5. CTA--TTATTTTCACTCGAA-----ACATAAAAAGA-TT-AAGAAAAA-----AAGTCGGAACAA--AATATCTTGTCTGCTTACATGGT

                120;                140;                160;                180;                200;
1. CGAAGTACTAATACTTGTCTT---ATGCTCTGCTACAT-----GTAGGCA---ATCTTATGGAATTTCT--TATAGAAATACA
2. CGAAGTACGTATACTTGTCTT---ATGCTCTGCTACATAATTGGAATCAATAGATCTCTGTAGAGACTAATCTTGTGGACTTTCT--TATGAAACGAT
3. CGAAGTACGTATACTTGTCTT---ATGCTCTGCTACATAATTGGAATCAATGGATCTCTGTAGAGACTAATCTTGTGGACTTTCT--TATGAAACGAA
4. CAAA-TATGGACATTTGTCTATACATGCTCTGCTACATA-----GTTTATGRAGA-----TT---RATTTCT-TCATAACCGGCT
5. CGAA-TATATATATATAT-ATATACAT-CTCTGTTTCAT-----TC---AGACA---ATTTTATGGAATCCCTCACACGGTAGGAA

                220;                240;                260;                280;                300;
1. ATAATATTGTTATGGTGTAGTC--TAGATGATGGTAACTT CGACATAAATTGTAAGTGTAGCTCTATATAAAAACGCAAAATCTAGGCAAGTTT AGTTG
2. ACAATATAGTTATTCGTGTAGTACAAAGATGATAGGTAAC--GACATAAATTGTAAGTACAGCTCTATTAA--ACGCAAAATCTAGGTAAGTTT AGTTA
3. ACAATATAGTTATTCGTGTAGTACAAAGATGATAGGTAAC--GACATAAATTGTAAGTAAAGCTCTATTAA--ACGCAAAATCTAGGTAAGTTT AGTTA
4. CCTTACTG---ACG-GTACTCGAATAGGCT--GT--CT---ATAAT-GTT---GC--TATTG-----GTGAGTTT
5. ATAATATTGTTATTCG-GT-----GTGACT-----TTT

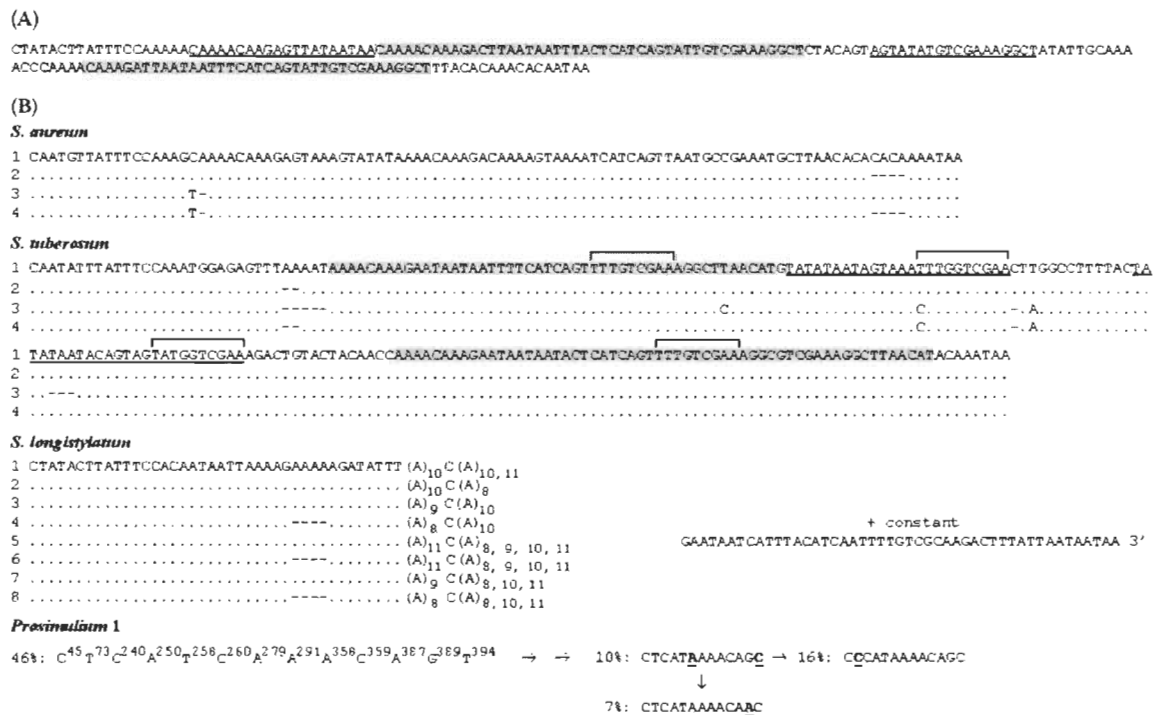
                320;                340;                360;                380;                400;
1. CATTGATCGAATTACAAAACAAGGATCTACATGACACTAGGTGACATTTACATGACTAACAATCAATGCGGACATTTCTAAAGACATTTCTTAATAAA
2. CATTGATCGAAGTACAAGACAAGGATCTACATGACACTAAATGACATTTAAACAACATAAAAATTGTGCGGACATTTCTAAATACGTT--TACTATG
3. CATTGATCGAAGTACAAGACAAGGATCTACATGACACTAAATTAACATTTAAACAAGACTAACAATTTGTGCGGACATTTCTAAATACGTT--TACTATG
4. GRGTGAT-----GA-----ATG-----GATTT-----GACAATCAT-----GT-----
5. GAGTGAT-----GAT-----ATG-----TGTGTTTT-----GACAAAAT-----GT-----

                420;                440;                460;                480;                500;
1. AATGACTTTAATATTGTGCGGTCAATCGGTGTGTAACCTTTGTTTCGATGCGTGACCAAACTCCTAAACAAAGAAATAGAAAAAATCATCAAT-TAT
2. AATGAATTTAATATTGTGCGGTCAATGGTGTGTAGCCTTTGTTTCGATGCGTTCCAAAACTCCTAAACAAAGATTAAGATTATACTCATCAAT-TAT
3. AATGAATTTAATATTGTGCGGTCAATGGTGTGTAGCCTTTGTTTCGATGCGTTCCAAAACTCCTAAACAAAGATTAAGATTATACTCATCAAT-TAT
4. -----TAGTTTGTG-TTTCAG-----GAAGCG-----CAAA-----GAAACAAAGATTAAGTAAATCCATCAGTATTT
5. -----ATAGTACAAGCTCTATT-----AAAGCG-----CATT-----TATAAAAAGATTAAGTAAATTTTCATCAGTAT-T

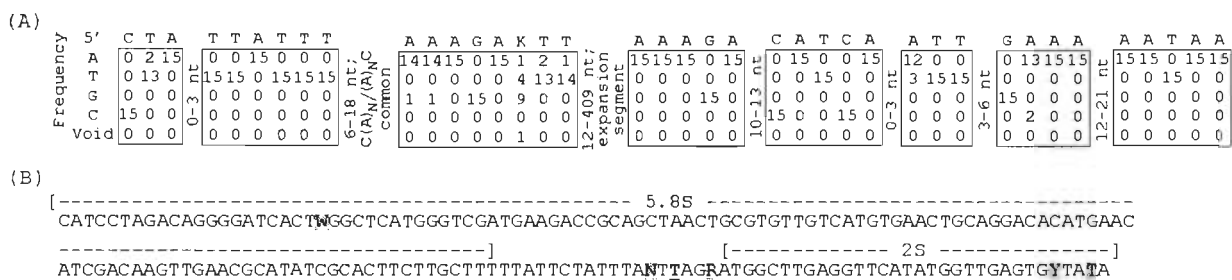
                520;                533;
1. AACGAAAGTTAT-AATGTATAA---CAAATAA
2. AACGAAAGTTAT-AA--TATAA---CA-AATAA
3. AACGAAAGTTAT-AA--TATAA---CA-AATAA
4. GTCGAAAGGCATCTACACACAAAACCAAATAA
5. AACGAAAGTTCT-TACACACA-----AAAATAA

```

**Figure 3.1.** Aligned ITS1 sequences from *Twinnia* (1), *Prosimulium* 1 consensus (2), *Prosimulium* 2 consensus (3), *S. quebecense* (4) and *S. mutata* (5). Overhead dot: polymorphic in *Prosimulium* 1. Light shading: differs systematically between *Prosimulium* 1 and 2. Dash, deletion; R, A/G.

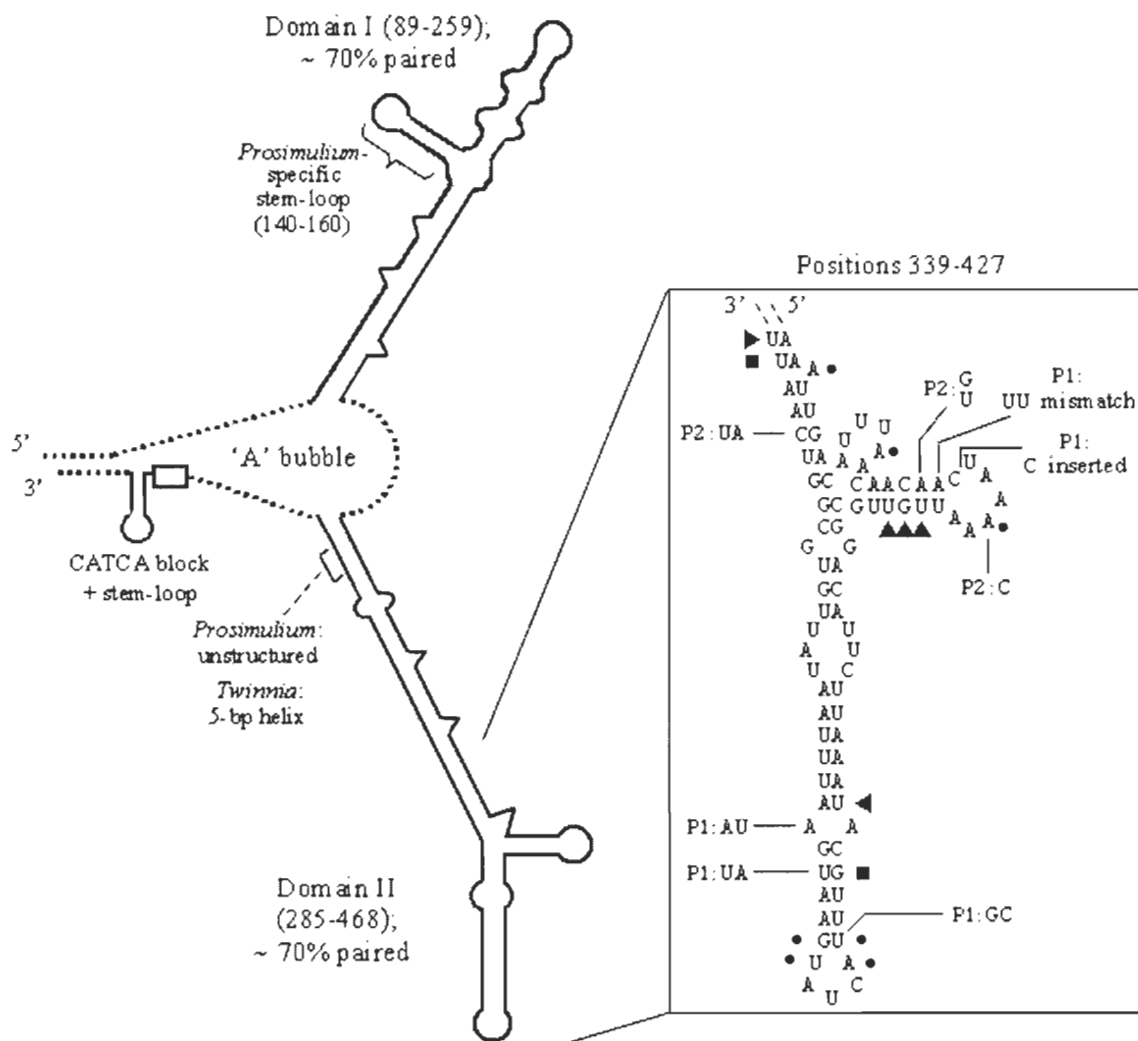


**Figure 3.2.** Internal repeats and polymorphisms in ITS1. (A) In *S. decorum*, the 5' copy of a large duplication (shaded) intercalates imperfect repeats (underlined) of its own 5' and 3' ends. (B) Instances of intraspecific sequence variants. A dot indicates identity with the top sequence and a dash, a gap. Among three specimens, *S. aureum* showed a total of four variants. *S. tuberosum* (2 specimens) had four; the 206-215 nt sequence exhibited two copies of a large duplication (shaded) flanking a smaller one (underlined) having an homologous segment (bracket) to a section of the major duplication. Thirteen *S. longistylatum* (5 specimens) variants of the central  $A_M C A_N$  microsatellite were recorded, in addition to an upstream  $A_4$  indel. The consensus for thirteen *Prosimulium 1* (19 specimens) polymorphic positions, numbered as exponents according to Fig. 1, is given at the bottom left; three other variants are also illustrated. Percent abundance relative to population composition is indicated and underlining shows base changes.

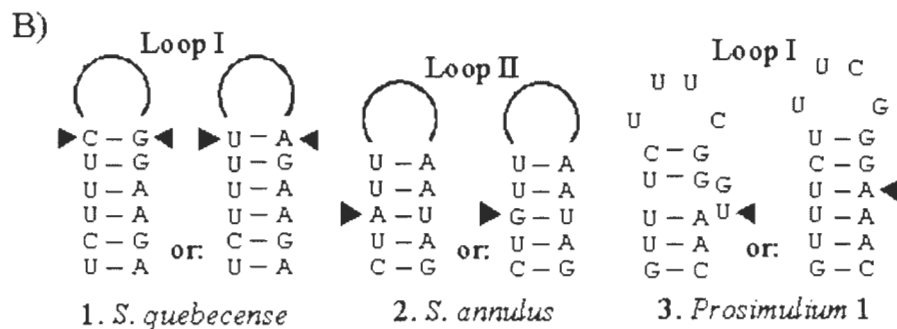
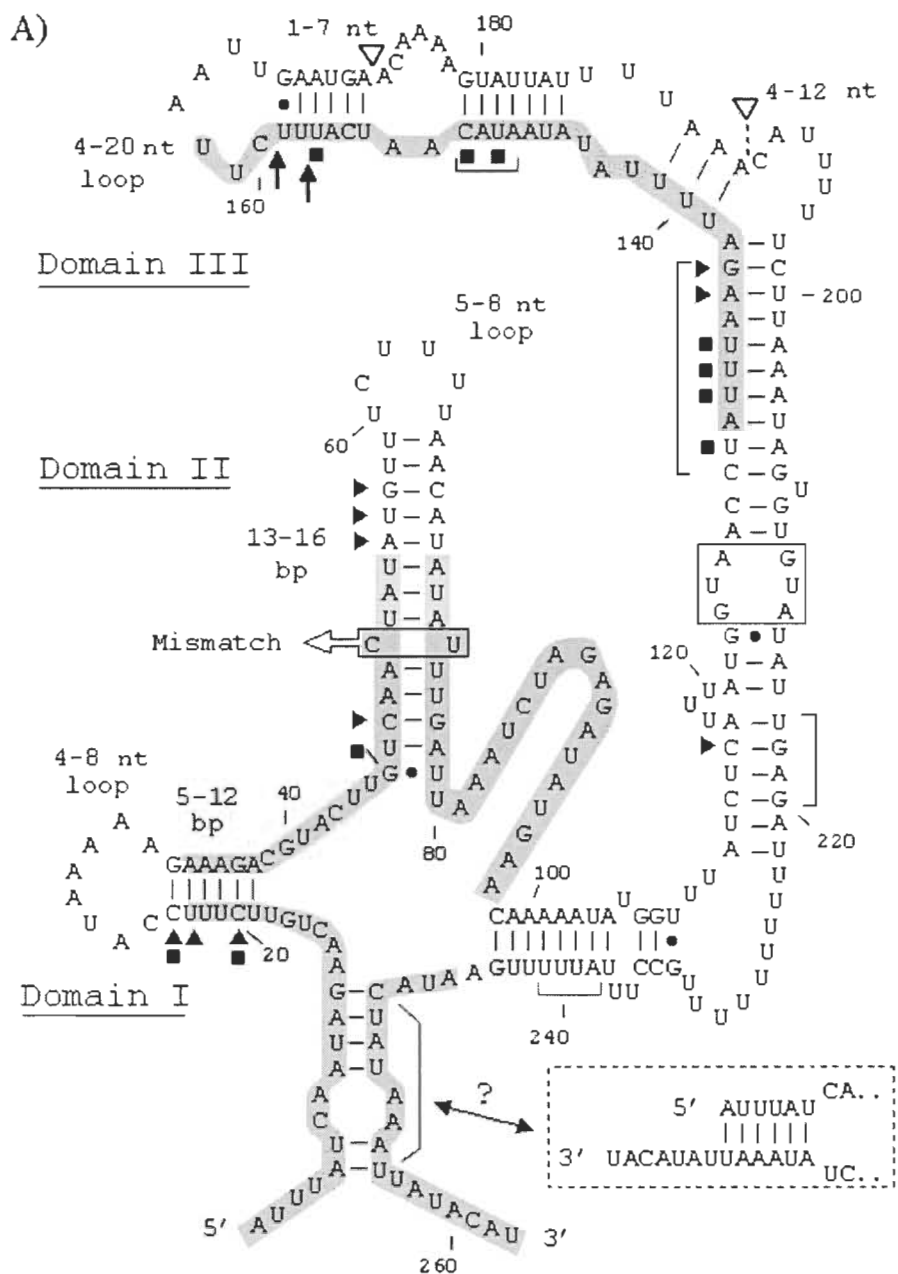


**Figure 3.3.** The ITS1-5.8S rDNA region. (A) ITS1: frequency of each base given below the consensus sequence, sizes of intercalary sequences indicated between the conserved blocks. K = T/G. (B) The split 5.8S + 2S genes. **W**, **Y**: T in Prosimuliini, A and C, respectively, in Simuliini. **R**: A/G for *Simulium*, G elsewhere. **T**: replaced by G in *S. laciniatum*. **N** varies within *Simulium*, but stands for G otherwise. **T**: followed by an inserted AT in *S. tuberosum*.





**Figure 3.4.** Schematic ITS1 secondary structure in *Prosimuliini*, with positions numbered as in Fig. 1. Spikes, bubbles: one-sided bulges, internal or terminal loops. See text for other explanations. (Inset) Mutation pattern for the terminal bifurcation of domain II (as RNA), with the consensus *Prosimulium* 1 sequence shown. ►, ■: true and hemi-CBCs, respectively, between *Prosimulium* 1 and *Twinnia*. •: other differences between them, including for *Twinnia* a deleted G•U pair ending the bottom stem. P1, P2: *Prosimulium* 1 variants, mutations distinguishing *Prosimulium* 2 from *Prosimulium* 1.



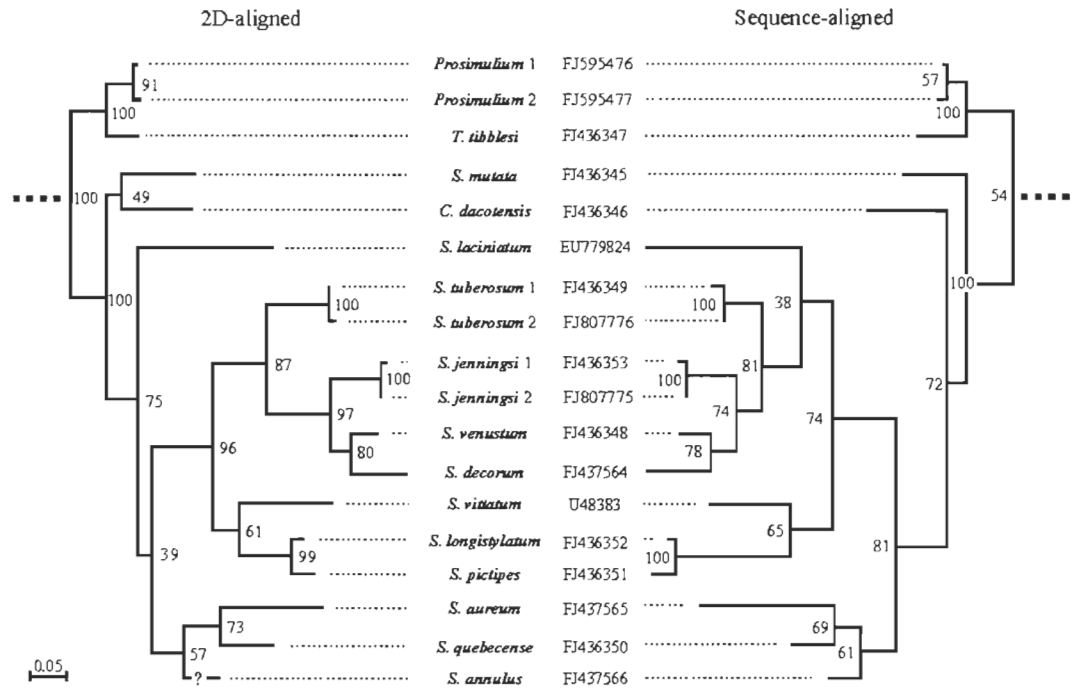
**Figure 3.5.** (A) The three stem-loop domains of ITS2 (as RNA), using *S. tuberosum* variant 1 as an example. Five highly conserved regions (121 nt) comprising on the average 44% of the whole sequence are shaded. On a 0 to 2 scale, the Conservation Value ( $C$ ) of Gutell et al. (1985) was calculated within these regions according to:

$$C = \sum P_i \log_2(4P_i) + P_{\Delta} \log_2(P_{\Delta}),$$

where  $P_i$  = frequency of base  $i$  and  $P_{\Delta}$  that of a deletion at that position (unusual *S. laciniatum* 3' terminus omitted from the calculation. See text for details). Each region has an average  $C$  over 1.8, save the second one from the 5' end ( $C = 1.7$ ), and usually contains no more than one residue with  $C < 1.0$ . When larger than shown here, stem I may have a mismatch near its top (*S. pictipes* and *T. tibblesi*). Stem II exhibits an invariant C-U mismatch in sixth position and may include one or two (*S. quebecense*) more mismatches in the upper half of stems over 14 bp. Filled triangles and squares indicate CBC and hemi-CBC ( $G \bullet U \leftrightarrow A-U$  or  $G \bullet U \leftrightarrow G-C$ ) locations, respectively. Multiple independent cases are found at a few of these locations. Other CBC sites in the top part of stem I, missing from *S. tuberosum*, are not indicated. For domain III, CBCs are attributed only at internal positions within conserved helices (brackets) used for aligning sequences and a box stands for the conserved position of an internal loop (4-13 nt and variable sequence). Together with size indications, open triangles in stem III indicate usual 3' one-sided bulges sometimes facing a small 5' bulge. Arrows pinpoint the normal starts of domain III terminal loop, 4-20 nt long, which possibly includes for the larger versions from *S. jenningsi*, *S. venustum* and *S. decorum* a short secondary

stem-loop. Two alternate pairing schemes between the 5' and 3' ends are suggested here.

(B) Structural consequences of base changes involving near-twin sequences (see text).



**Figure 3.6.** An ITS2 phylogeny generated through the GTR model, with the root arbitrarily set between Prosimuliini and Simuliini. The two variants from *S. jenningsi* and *S. tuberosum* are considered as distinct ‘species’ for treeing purposes. Bootstrap values (%) over 500 replicates are indicated for each node. At left, tree from 2D-constrained alignment; at right, from sequence-only alignment. Question mark: uncertain branching of *S. annulus* (see text for explanation).

## **CHAPITRE IV**

« Phylogenetic suitability of COI, COII and 18S in land invertebrates: a black fly (Diptera: Simuliidae) case study. »

**Phylogenetic suitability of COI, COII and 18S in land invertebrates: a black fly  
(Diptera: Simuliidae) case study.**

Christine Gaudreau<sup>A</sup>, Bernard LaRue<sup>A,B</sup>, and Guy Charpentier<sup>A</sup>

<sup>A</sup>Département de Chimie-biologie, Université du Québec à Trois-Rivières, 3351 Boulevard des Forges, Box 500, Trois-Rivières, Québec G9A 5H7, Canada.

<sup>B</sup>Corresponding author. Email : [bernard.larue@uqtr.ca](mailto:bernard.larue@uqtr.ca)

**Running title**

Phylogeny of black flies through COI, COII and 18S

**Additional Keywords :** Black fly, COI, COII, 18S rDNA, Simuliidae.

## 4.1 Résumé

Comme la résolution de parenté parmi les mouches noires est particulièrement difficile, nous avons évalué trois des marqueurs d'ADN les plus courants, COI, COII et l'ADNr 18S, pour leurs capacités respectives à investiguer des niveaux distincts de profondeur évolutive à l'intérieur de la sous-famille des Nématocères. En utilisant les données de 15 espèces de larves de simulies recueillies au Québec (Canada), les gènes COI et COII pourraient reculer à environ 16 millions d'années avant que la saturation en mutations n'embrouille la phylogénie. Toutefois, ces deux marqueurs d'ADNmt sont convenables pour distinguer les espèces relativement proches ou même des populations de la même espèce. Ces analyses moléculaires mettent aussi en doute la valeur de certains traits morphologiques utilisés pour l'identification des spécimens de *Prosimulium*, au point où la classification taxinomique devra être réévaluée. À l'opposé, le 18S résout mieux les divergences basales, mais s'avère si conservateur de séquence que son utilisation combinée à celle de COI et COII laisse en suspens l'intervalle phylogénétique s'étendant du genre jusqu'au sous-genre.

### 4.1.1 Contribution des auteurs de l'article

L'article a été rédigé par Christine Gaudreau et Bernard LaRue. La partie expérimentale et bioinformatique fut entièrement réalisée par Christine Gaudreau. Guy Charpentier a supervisé l'ensemble des travaux et aidé à l'identification morphologique des larves de simulies.

Ce manuscrit sera soumis sous peu.

## 4.2 Abstract

Since resolving relationships among black flies appears especially difficult, we evaluated three DNA markers, COI, COII and 18S rDNA, for their respective abilities to investigate distinct levels of evolutionary depth within this nematoceran subfamily. Using sequence data from fifteen species collected in Québec (Canada) as larval specimens, the COI and COII genes could reach at most ~16 million years back before mutation saturation blurred the phylogeny. However, both mtDNA markers were well suited for distinguishing between closely related species or even monospecific populations. Molecular evidence also cast doubts on the value of some morphological traits for classifying *Prosimulium* members, to the point of calling for a taxonomic reappraisal of the whole genus. At the opposite, 18S performed best in elucidating remote divergence events, such that these two categories of probes left an unresolved phylogenetic gap extending from the genus to the subgenus level.



### 4.3 Introduction

Due to morphological and ecological similarities, black flies, a nematoceran sub-family with 2072 known species (Adler and Crosskey 2009), seem sometimes like a taxonomist's nightmare. Also, an environmentally induced plasticity and a prevalence of biotypes and complexes of species bring additional complications (Currie and Adler 2008). This classification problem has relevance to both basic research and health issues, which involve mostly onchocerciasis. In this case, efficient selective control depends on a reliable identification of vector species from the *Simulium damnosum* complex and on an understanding of their respective roles in pathogen transmission (Morales-Hojas *et al.* 2002). Based on fixed-inversion differences in polytene salivary gland chromosomes of larvae, many nominal black fly species now appear as complexes of siblings. However, cytological techniques do not allow the identification of biting adult females (Krüger *et al.* 2000), usually require samples from the late larval instars (Rothfels 1979) and have other severe limitations such as the amount of work required and a high failure rate (Spironello *et al.* 2002). On the other hand, molecular methods hold the promise of rational classification schemes based on phylogeny-related information. They can be applied to both sexes at all development stages, rely on the PCR and usually aim at easily amplified DNA sequences.

The nuclear ribosomal DNA cluster (rDNA) and the mitochondrial cytochrome oxidase I (COI) and II (COII) genes are prominent among DNA markers given the availability of universal PCR primers (Folmer *et al.* 1994; Shepard *et al.* 2006). Due to variations in evolutionary conservatism, each one probes distinct levels of phylogenetic depth. The rDNA transcription unit, a mosaic of conserved and variable regions, exhibits little sequence divergence between its hundreds to thousands of individual copies, hence it is used as a model for the study of molecular drive (Dover 1986; Strachan *et al.* 1985) and concerted evolution (Arnheim 1983) in multigene families. The highly conserved 18S, 5.8S and 28S genes are used to construct higher

level phylogenies (Miller *et al.* 1997; Moulton 2000; Pace *et al.* 1986), while the intergenic spacers ITS1 and ITS2, discarded during rRNA maturation, evolve faster and frequently undergo insertions or deletions complicating sequence alignment (Schlötterer *et al.* 1994). ITS2 is used at low to intermediate levels of phylogenetic reconstructions (Coleman 2003; LaRue *et al.* 2009; Thanwisai *et al.* 2006) and ITS1 mostly for population studies, including black flies (Gaudreau *et al.* 2008; Rodríguez-Pérez *et al.* 2006), and species authentication (Brockhouse *et al.* 1993; Krüger *et al.* 2000).

Variation in the COI and COII genes is constrained by their coding function. The corresponding polypeptides are highly conserved, thus providing for an unambiguous DNA sequence alignment, but the genes themselves evolve at a quick pace, mainly through silent mutations of the third codon position (Gaudreau *et al.* 2009a). The proposal of COI as a standard for the molecular bar-coding of species (Hebert *et al.* 2003) has generated plenty of data for many animal groups and, more specifically, a widespread use in insect phylogenetics (Caterino *et al.* 2000). Current applications of COI and COII in black flies cover resolution of species (Day *et al.* 2008), phylogeographic variation (Finn and Adler 2006; Gaudreau *et al.* 2008, 2009a) and evolutionary history of subgenera (Joy and Conn 2001; Joy *et al.* 2007). The loss of resolution observed at higher taxonomic levels suggests the inability of COII to decipher relationships above the subgeneric level (Pruess *et al.* 2000), a likely conclusion for COI as well.

Based on a sample of fifteen Simuliidae species from Québec, the temporal horizon beyond which the phylogenetic usefulness of the COI and COII genes vanishes was estimated. Their ability to resolve pairs of black fly morphospecies with an ambiguous status was also probed and the validity of some traits currently used to classify members from the *Prosimulium* genus questioned by pitting them against molecular data. At a basal level, a phylogeny

constructed from the highly conserved 18S and 5.8S rDNA was finally compared to a previous one built through ITS2 (LaRue *et al.* 2009).

#### 4.4 Methods

##### 4.4.1 General collecting

Actual specimens of black fly larvae from Quebec were part of a larger sample from a field study, with collection and storage prior to DNA extraction performed as described earlier (Gaudreau *et al.* 2009b). Larvae were identified under a low-magnification microscope according to the morphological keys of Wood *et al.* (1963) and Adler *et al.* (2004).

##### 4.4.2 DNA amplification and sequencing

DNA was extracted and purified according to St-Onge *et al.* (2008). PCR primers and thermocycling conditions for COII (Gaudreau *et al.* 2008) and COI (Gaudreau *et al.* 2009a) amplification were described earlier. In rare instances of initial PCR failure, the COI region was recovered by using other pairs of primers described by Simon *et al.* (1994). The 18S gene was amplified in two separate sections overlapping on a 130 nucleotide (nt) length, with CTGGTTGATCCTGCCAGTAG (forward) and GTGGTGCCGTTCCGTCAATTCC (reverse) as 5' half primers, and GGATCGAAGGCGATTAGATAC (forward) and CTTCCGCAGGTTACCTACG (reverse) for the 3' half. Either half was amplified in 50 µL of a 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub> buffer containing 1 unit of Taq DNA polymerase (Roche), 200 µM of each dNTP and 4 µM of each primer. Thermocycling consisted of an initial 60 s denaturation step at 94°C followed by 38 amplification cycles (30 s, 94°C + 50 s, 51°C + 120 s, 72°C) and a final 180 s extension step at 72°C. Sequencing was performed in both directions so as to give an average redundancy of 85% for the entire 18S gene. The sequence of the split 5.8S gene (including 5.8S rDNA proper, spacer 2a and 2S rDNA) was recovered from the PCR products obtained through the combined use of the leftward ITS1

(Gaudreau *et al.* 2008) and rightward ITS2 primers (LaRue *et al.* 2009). As already described (Gaudreau *et al.* 2008), all PCR products underwent agarose gel electrophoresis followed by extraction and purification prior to automated dideoxy sequencing.

#### 4.4.3 Sequence data analysis

Alignment of DNA sequences from rDNA (18S and 5.8S), COI and COII was performed with the Clustal W program implanted in Geneious software version 4.7.6. (Drummond *et al.* 2009) and further refined by manual adjustments as required. For each gene, we performed maximum-likelihood (ML) analysis using PhyML, version 2.4.4 (Guindon and Gascuel 2003). Modeltest 3.4 (Posada and Crandall 1998) indicated general time-reversible (GTR) to be the appropriate substitution model for ML based on a hierarchical likelihood ratio test or Akaike's information criterion. The credibility of nodes in the tree was assessed by 500 bootstraps.

#### 4.4.4 Simulating the evolutionary process

COI and COII evolution was mimicked using a program built in Excel software from Microsoft Office 2003. The underlying model assumed mutation equilibrium in a strict sense, meaning that each nucleotide interchange occurred at equal rates in both directions. For any given transition to transversion ratio (Ts/Tv), a parameter that could be adjusted at will, an explicit solution is obtained for the probability value of each possible mutation. In practice, Ts/Tv had little impact on the final result except when set at extreme (and most unlikely) values. Starting from an arbitrary sequence having the same composition as the third codon position from COI/COII, individual mutations were randomly introduced according to the above probabilities and the process was repeated until reaching a specified average number of hits per position. Using the program, the divergence between daughter sequences with the same

progenitor could be followed and fictitious trees created by re-introducing a daughter sequence as the input for the next mutation cycle.

## 4.5 Results

Table 4.1 gives the list of the fifteen black fly species (five genera) from Québec analyzed here, together with GenBank accession numbers for their COI, COII and rDNA sequences. As indicated, morphological identifications were also validated by two molecular criteria: (i) externally, through matches with independent GenBank entries, if available, (ii) through the cohesion of molecular data within each morphological unit, as COI, COII and/or ITS1 (highly sensitive to interspecific differences; see LaRue *et al.* 2009) showed minimal sequence heterogeneity. COI or COII sequences varied in numbers from one to twenty-eight depending on the species under study. The rDNA sequence, with a presumed (through comparisons with *D. melanogaster*) 41 nt length missing at the 5' end of the 18S gene due to the forward primer location, was obtained for each species from a single specimen having already known COI and COII sequences.

### 4.5.1 COI and COII: general considerations

The PCR products from the COI and COII genes gave readable sequences through 651 nt (217 codons) and 252 nt (84 codons), respectively, without any gap being required for interspecific alignment. The two genes evolve at comparable rates (Gaudreau *et al.* 2008) and were thus treated as a single unit for phylogenetic purposes. The base composition was quite well balanced at the 1<sup>st</sup> codon position, but the excess of T at the 2<sup>nd</sup> position reflected the high hydrophobic amino acids content in the corresponding polypeptides (Table 4.2). As usual for arthropod mtDNA genes, the 3<sup>rd</sup> position, with only 15% of such sites remaining invariant, exhibited a strong A-T bias. Nucleotide interchanges recorded there resulted overwhelmingly in synonymous mutations. Among the 62 variable sites (~20% of the total) from 1<sup>st</sup> and 2<sup>nd</sup>

positions, nearly half implicated T↔C transitions simply generating 1<sup>st</sup> position silent switches between leucine codons. Much of the remainder produced presumably neutral interchanges between hydrophobic amino acids or within the serine/threonine/alanine trio (Table 4.3).

Base composition imbalances and high substitution rates can lead to mutation saturation and impair phylogenetic reconstructions (Bos and Posada 2005). A theoretical saturation limit was computed by assuming a randomized 3<sup>rd</sup> position within a fixed base composition (taken from Table 4.2) due to mutation equilibrium. The proportions of synonymous codon families having the XXR, XXY and XXN structures (R = A/G, Y = T/C and N = any base), respectively, were also weighted into the calculation according to their specific uses in the COI and COII genes. Random sequences generated through these rules were then expected to differ on the average by 46%, as opposed to 34% for aligned 3<sup>rd</sup> positions from actual COI/COII sequences. This 75% saturation level suggested the unreliability of phylogenetic reconstructions except for terminal branches, which was consistent with the poor to negligible bootstrap support found for most nodes of the tree (Fig. 4.1). Upon removing the three closely related groups described next, divergence levels were distributed narrowly (s.d./mean = 0.14), another indication of nearby saturation.

#### 4.5.2 Genetic differentiation within tight groups

COI/COII sequence homology and bootstrap support criteria pointed to three cases of recently diverged species or distinct monospecific populations. In the first instance, specimens of *S. pictipes* (Chutes Mont-à-Peine; 46°12'N, 73°34'W) and *S. longistylatum* (Chutes Dunbar; 46°57'N, 73°07'W) collected about 90 km apart in distinct streams were first identified through the eye spot area pattern and the postgenal cleft shape that enable the two species to be distinguished (Wood *et al.* 1963). Molecular data fully agreed with these attributions as the COI tree (inset to Fig. 4.1) exhibited two well resolved branches coinciding with morphotypes and

diverging from each other by 4.9% on average. Similar results were recorded for COII, with 25 and 4 specimens from *S. longistylatum* and *S. pictipes*, respectively (data not shown). A second case involved *S. annulus* (Rivière Duchénier; 48°08'N, 72°58'W) and *S. euryadminiculum* (Lac du Barrage; 46°29'N, 72°58'W), the latter recently put in synonymy with the former by Adler *et al.* (2004). *S. annulus* was identified through the key of Adler *et al.* (2004) and *S. euryadminiculum* according to Wood *et al.* (1963), which keys indicated significant differences in the pattern of various head spots. A split between two molecular lines that perfectly fitted morphology (inset to Fig. 4.1) was again obtained, but divergence decreased to 1.4% only. As discussed later, this low figure indicated phylogeographic groups from a unique species.

Based on Snyder and Linton (1983), there was no single character of larval morphology that could resolve the various species from the *Prosimulium* genus. Since the keys from Adler *et al.* (2004) et Wood *et al.* (1963) both describe it in full detail, the hypostomal teeth pattern was used here to identify 23 specimens all sampled during wintertime at the outlet of Lac Souris (46°35'N, 72°58'W). Our temporary classification into six nominal species was checked against COI data, which generated a clash between morphotypes and molecular phylogeny as the various 'species' were more or less intermingled across the whole tree which comprised only two major branches (inset to Fig. 4.1). Due to an average sequence divergence of 7.8% between these lineages, we recognized (at least) two species to be temporarily defined on molecular grounds. For lack of formal names, they were designated as *P.sp.1* (*Prosimulium* 1) and *P.sp.2* (*Prosimulium* 2).

#### 4.5.3 A tentative phylogeny from the 18S-5.8S rDNA cluster

Upon adding the 41 bases presumably missing from the 5' end, 18S rDNA varied in length from 1973 (*P.sp.1*, *P.sp.2*, *T. tibblesi*) to 1955 nt (*S. tuberosum*) while the remainder of the *Simulium* sequences stood uniformly at 1961 nt. Average base composition was 28.5% A,

28.5% T, 24.5% G and 18.5% C, with minimal interspecific variation. There was no evidence for dual-base positions in any species, showing virtual identity between all gene copies. The 5.8S region (including 5.8S, spacer 2a and 2S rDNA) provided 174 nt sequences that could be perfectly aligned, except for an AT dinucleotide inserted next to position 139 of the spacer 2a sequence from *S. tuberosum*. Due to its weak phylogenetic signal (Miller *et al.* 1997), the short and highly conserved 5.8S region was fused with 18S to generate the tree shown in Fig. 4.2.

Using two nematoceran outgroups, this tree convincingly resolved the Prosimuliini (*Prosimulium* and *Twinnia*) and Simuliini (other Simuliidae) tribes, a basal divergence earlier recognized on both morphological (Currie 1988) and molecular grounds (Moulton 2000). *Simulium* appeared as a tight clade although statistical evidence was too weak to pinpoint its exact rooting with respect to *Cnephia* and *Stegopterna*. Also, several bootstrap values within *Simulium* appeared not significant, which likely reflected a lack of sequence variation as exemplified by a maximum of only twenty-two differences (average: 11.4) between the two most distant species from this group. Consistent with this low resolving power, much of the 18S sequence was stringently conserved. Variability was largely concentrated inside two hot spots centered about 760 and 1480 nt after the start of the gene (regions I and II in Fig. 4.3a), including a *Simulium*-specific 7-nt deletion in region I (Fig. 4.3b), and mutations elsewhere appeared mostly as sporadic occurrences. At the opposite, many positions from the hot spots mutated repeatedly. When using the 18S phylogeny to reconstruct the mutation network, average densities of 1.4 and 0.8 mutations/site were recorded for the hypervariable cores (Fig. 4.3b) of regions I and II, respectively. As indicated, several diagnostic positions from the cores (and a few others scattered elsewhere) were detected, in the sense that base identity typified either Prosimuliini or *Simulium*. Similar correlations also occurred at positions 22 (Prosimuliini vs Simuliini: T vs A) and 172 (T to C substitution) of the 5.8S region.



#### 4.6 Discussion

To evaluate the time required for mutation saturation to blur the phylogenetic signal from the COI/COII genes, we used Brower's (1994) estimate of a divergence rate, averaged over all three codon positions, of 2.3% per million years (My). Our analysis was restricted to the 3<sup>rd</sup> codon position which cumulated ~ 90% of all nucleotide interchanges, thus an upgrade of the previous figure to ~6% My<sup>-1</sup> (meaning ~3% My<sup>-1</sup> per lineage). A computer simulation (see experimental procedures) that introduced nucleotide replacements at this rate while following a mutation equilibrium model was next performed on random sequences having the same nucleotide proportions as COI/COII. The 75% mutation saturation level found earlier to ruin phylogenetic inference was then obtained after a simulation equivalent to ~16 My, barely one eighth of the attested antiquity of black flies (Bertone *et al.* 2008; Kalugina 1991). In another test, the simulation generated fictitious trees extending from -30 My to present, with a single lateral branch growing out of an intermediary node every 2 My. The trees reconstructed from the terminal sequences faithfully reflected the simulated phylogeny for the first 8-10 My, but became more and more unreliable while reaching further back, as the bootstrap support fell below a critical 50% level at about -13 My and kept on diminishing. At the same time, branching points were increasingly inaccurate and apparent lengths for older branches much undervalued when compared to the simulation input. To summarize, a temporal horizon located 13 to 16 My ago restricts COI/COII uses to low taxonomic levels and population studies. With respect to real sequences, even adding the less mutable and presumably more informative 1<sup>st</sup> and 2<sup>nd</sup> codon positions would bring little improvement due to the huge noise from 3<sup>rd</sup> positions. Using the COI/COII polypeptide sequences would be of no help either since their conservatism restricts applications to much higher taxonomic levels.

Disagreements about black fly taxonomy remain commonplace and our results could resolve ambiguities with respect to *S. pictipes*/*S. longistylatum*, *S. annulus*/*S. euryadminiculum* and *Prosimulium*. In the first instance, a reexamination of museum specimens has recently led to the proposal of switching names from *S. longistylatum* to *S. pictipes* and equating the former *S. pictipes* with the already known *S. innoxium* (Adler *et al.* 2004), but morphological keys were not updated accordingly. Since Crosskey and Howard's (2004) inventory of blackflies does not yet acknowledge the change, we chose instead to follow Wood *et al.* (1963) with the former names and to use the accompanying full descriptions. *S. longistylatum* and *S. pictipes* are both characterized by unusually large larvae inhabiting swift current areas at the head of waterfalls. Although superficially quite similar, they are morphologically distinguishable, which was fully supported by molecular data. For argument's sake, the distance between the two groups could still reflect intra- rather than interspecific variation, but this interpretation was rejected on several grounds. First, COI/COII sequences diverge by about 5%, in the vicinity or higher than the usually recognized threshold values for species level in invertebrates (Carew *et al.* 2007; Hebert *et al.* 2003). Also, each group exhibits a reduced genetic diversity while being clearly differentiated from the other, a testimony to a comparatively early separation of ancestral lines. An age of ~2.1 My ago is inferred for their last common ancestor based on the COI/COII molecular clock rate. Finally, 18S sequences differ, albeit slightly (as expected for this highly conserved marker), and highly significant dissimilarities in ITS1 sequences (LaRue *et al.* 2009) provide independent evidence from the nuclear genome about the existence of two species.

The ornithophilic *S. (Boreosimulium) annulus* species group wore several labels since the original description as *Melusina annulus* Lundström, 1911, including *S. euryadminiculum* Davies, 1949, still used in the key of Wood *et al.* (1963). Adler *et al.* (1999) later recognized the morphological and chromosomal identities between *S. annulus* (Scandinavia) and *S. euryadminiculum* (North America), finally put in synonymy under the former's name (Adler *et*

*al.* 2004). Our specimens were initially categorized as one or the other 'species' based on minor morphological details reported in former keys. In the end, molecular evidence agreed with the proposed synonymy and, since the two collection sites were ~400 km distant, suggested phylogeographic variation as the best explanation for the observed differences. Indeed the COI/COII differentiation between *S. annulus* and *S. euryadminiculum* stayed well within the normal intraspecific range for invertebrates (Carew *et al.* 2007; Gaudreau *et al.* 2009a). Yet, this clue could lead to an erroneous conclusion due possibly to mtDNA introgression between true, but closely related species. However, nuclear genome data provided independent support for a monospecific unit as ITS1, ITS2 (LaRue *et al.* 2009) and 18S sequences were identical between the two groups.

*Prosimulium* species were globally identified as members of the *P. hirtipes* Fries complex prior to 1956, after which they were distinguished by morphology, ecology and, later on, according to cytoforms that were equated with true taxonomic units. Borders between species appear fuzzy at times, as for *P. mixtum* and *P. fuscum* that are related enough to hybridize with each other (Rothfels and Freeman 1977) in spite of a morphological distinctness attributed by Snyder and Linton (1983) to slight differences in ecological requirements. Characters such as the number of cephalic fan rays and the hypostomal tooth pattern are often described as discriminating criteria, but some overlap is occasionally observed with respect to fans (although current keys mention < 33 versus > 33 rays for *P. fuscum* versus *P. mixtum*), even after eliminating the common bias of accidentally broken rays. One may express doubts about a clear-cut demarcation, since changes in water velocity and seston availability can induce phenotypic plasticity of ray numbers in other species (Lucas and Hunter 1999; Palmer and Craig 2000). The tooth pattern discriminates quite efficiently between genera (Currie 1986), but much less at the species level. For instance, Snyder and Linton (1983) report that *P. fuscum* populations at three different sites exhibited a significant variability in tooth length and mention

that tooth arrangement is probably unfit for separating *P. mixtum* from *P. fuscum*. They even go to the point of suggesting that neither a single trait nor any given combination of morphological characters can identify with certainty *Prosimulium* larvae at the species level. The COI/COII tree lends support to this view as our classification into six morphospecies according to tooth pattern, checked separately by two of the authors, clearly disagrees with the molecular phylogeny. Worse, the *Helodon* and *Prosimulium* genera appear mixed with each other. One could still rely on morphology while hypothesizing a high introgression rate of the maternally transmitted mtDNA, hence a gene pool shared between all six species. However, the observed COI/COII distance (7.8% divergence for *P.sp.1* versus *P.sp.2*) would imply at the same time that the mtDNA flow transgresses normal interspecific barriers, which negates reproductive isolation. This self-denying scenario also contradicts the nuclear genome data since specific ITS1 features (LaRue *et al.* 2009) split the 23 *P.sp.1* and *P.sp.2* specimens exactly along the same two major lines as COI/COII, leaving only the option of recognizing *P.sp.1* and *P.sp.2* as distinct species. Moreover, the radiation of the *P.sp.1* COI sub-tree into two terminal branches diverging by 4% on average suggests that the *P.sp.1* line itself is undergoing a further speciation event. Since chromosomal maps of our specimens were not presently determined, neither *P.sp. 1* nor *P.sp. 2* can be equated now with known nominal species due to the inadequacies of morphology and the absence of GenBank sequence matches. Finally, the reliability of differences in chromosomal maps for defining true taxonomic units is far from absolute. For instance, the diversity of cytoforms in the *S. damnosum* complex vastly exceeds the number of cytospecies due to allopatric variation (Morales-Hojas and Krueger 2009), which possibly implies the same about *Prosimulium*. In conclusion, numerous inconsistencies and interrogations (in general, not only with respect to our results) indicate that time is ripe for a taxonomic reappraisal of the whole *Prosimulium* genus. Such a task would require a unique pool of specimens spread over an adequate range of nominal species and fully characterized with respect to morphology, cytology and ecology. Only then shall it be possible, by bringing in the relevant molecular phylogeny, to

separate traits that simply reflect intraspecific polymorphisms from those that truly establish species distinctness. In this context, only a systematic combination of methods can lead to a better understanding of the biology, genetics and evolution of the genus.

The conservatism of 18S rDNA allows the elucidation of phylogenies down to ordinal relationships in insects (Kjer 2004), but ITS2 provides better bootstrap support for resolving low to intermediate levels. For instance, an ITS2 tree based on an identical sampling of species clearly shows *Simulium* coming out as the latest diverged branch within Simuliini and is better resolved with respect to terminal clades (LaRue *et al.* 2009). Those include a *S. jenningsi/venustum/decorum* group, with *S. tuberosum* a more distant relative, neighboring another one which consists of *S. vittatum/pictipes/longistylatum*, while these relationships are blurred or distorted when using rDNA. On the other hand, *S. quebecense*, *S. aureum* and *S. annulus (euryadminiculum)* appear in each case as early diverged *Simulium* lineages. Both trees indicate a slowed down mutation rate for Prosimuliini. In fact, *P.sp.1* and *P.sp.2* rDNA sequences are totally identical and exhibit only ten substitutions relative to *Twinnia* although data from ITS1 suggest that the two genera diverged ~30 My ago (LaRue *et al.* 2009). As a general conclusion, 18S goes far back to several hundred My, while ITS2 and COI/COII perform best at the sub-family and subgeneric levels, respectively, such that combining all three markers could hopefully lead to a fully resolved phylogeny of Simuliidae.

#### 4.7 Acknowledgements

We are grateful to the Université du Québec à Trois-Rivières who provided C. Gaudreau's scholarship and funding during her research.

#### 4.8 References

Adler, P.H., Malmqvist, B. and Zhang, Y. (1999). Black flies (Diptera: Simuliidae) of northern Sweden: taxonomy, chromosomes, and bionomics. *Entomologica Scandinavica* **29**, 361-382. doi: 10.1163/187631298X00014

Adler, P.H. and Crosskey, R.W. (2009). World blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory. Available online at <http://entweb.clemson.edu/biomia/pdfs/blackflyinventory.pdf>. [Accessed on 22 January 2010]

Adler, P.H., Currie, D.C. and Wood, D.M. (2004). 'The Blackflies of North America.' (Cornell University Press: Ithaca, NY, USA.)

Arnheim, N. (1983). Concerted evolution of multigene families. In *Evolution of genes and proteins* (M. Nei and R.K. Koehn eds.). Sinauer Associates, Boston, MA, pp. 38-61.

Bertone, M.A., Courtney, G.W. and Wiegman, B.M. (2008). Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. *Systematic Entomology* **33**, 668-687. doi: 10.1111/j.1365-3113.2008.00437.x

Bos, D.H. and Posada, D. (2005). Using models of nucleotides evolution to build phylogenetic trees. *Developmental and Comparative Immunology* **29**, 211-227. doi: 10.1016/j.dci.2004.07.007

Brockhouse, C.L., Vajime, C.G., Marin, R. and Tanguay, R.M. (1993). Molecular identification of onchocerciasis vector sibling species in black flies (Diptera: Simuliidae). *Biochemical and Biophysical Research Communications* **194**, 628-634. doi: 10.1006/bbrc.1993.1867

Brower, A.B.Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceeding of the National Academy of Sciences of the United States of America* **91**, 6491-6495.

Carew, M.E., Pettigrove, V., Cox, R.L., Hoffmann, A.A. (2007). DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae). *Journal of the North American Benthological Society* **26**, 586-599. doi: 10.1899/06-120.1

Caterino, M.S., Cho, S. and Sperling, F.A. (2000). The current state of insect molecular systematics: a thriving tower of babel. *Annual Review of Entomology* **45**, 1-54. doi:10.1146/annurev.ento.45.1.1

Coleman, A.W. (2003). ITS2 is a doubled-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* **19**, 370-375. doi:10.1016/S0168-9525(03)00118-5

Crosskey, R.W. and Howard, T.M. (2004). A revised taxonomic and geographical inventory of world blackflies (Diptera: Simuliidae). The Natural History Museum, London. Available online at <http://www.nhm.ac.uk/research-curation/research/projects/blackflies/> [Accessed on 22 January 2010]

Currie, D.C. (1986). An annotated list of and keys to the immature black flies of Alberta (Diptera: Simuliidae). *Memoirs of the Entomological Society of Canada* **134**, 1-90.

Currie, D.C. (1988). Phylogeny of primitive Simuliidae (Insecta: Diptera: Culicimorpha). PhD dissertation, University of Alberta, Edmonton, Canada.

- Currie, D.C. and Adler, P.H. (2008). Global diversity of black flies (Diptera: Simuliidae) in freshwater. *Hydrobiologia* **595**, 469-475. doi: 10.1007/s10750-007-9114-1
- Day, J.C., Goodall, T.I. and Post, R.J. (2008). Confirmation of the species status of the blackfly *Simulium galaratum* in Britain using molecular taxonomy. *Medical and Veterinary Entomology* **22**, 55-61. doi: 10.1111/j.1365-2915.2008.00719.x
- Dover, G.A. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics* **2**, 159-165. doi: 10.1016/0168-9525(86)90211-8
- Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T. and Wilson, A. (2009). Geneious v4.7, Available from <http://www.geneious.com/>
- Finn, D.S. and Adler, P.H. (2006). Population genetic structure of a rare high-elevation black fly, *Metacnephia coloradensis*, occupying Colorado outlet streams. *Freshwater Biology* **51**, 2240-2251. doi: 10.1111/j.1365-2427.2006.01647.x
- Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294-299.
- Gaudreau, C., LaRue, B., Charbonneau, V., Charpentier, G. and Craig, D.A. (2008). Recent origins and genetic fragmentation of *Simulium* (*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid. *Invertebrate Systematics* **22**, 555-562. doi: 10.1071/IS08024



Gaudreau, C., LaRue, B., and Charpentier, G. (2009a). Molecular comparison of Quebec and Newfoundland populations of the *Simulium vittatum* species complex (Diptera: Simuliidae). *Medical and Veterinary Entomology* (In Press) doi: 10.1111/j.1365-2915.2009.00844.x

Gaudreau, C. and Charpentier, G. (2009b). Seasonal and spatial distributions of blackfly larvae in two lake outlet streams of the Mauricie Region of Québec, and species survey in parts of southern Québec Territory. *Northeastern Naturalist* (Under revision)

Guindon, S. and Gascuel, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704. doi: 10.1080/10635150390235520

Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings Biological sciences / The Royal Society of London* **270**, 313-321. doi: 10.1098/rspb.2002.2218

Joy, D.A. and Conn, J.A. (2001). Molecular and morphological phylogenetic analysis of an insular radiation in Pacific black flies (*Simulium*). *Systematic biology* **50**, 18-38. doi: 10.1080/106351501750107431

Joy, D.A., Craig, D.A. and Conn, J.E. (2007). Genetic variation tracks ecological segregation in Pacific island black flies. *Heredity* **99**, 452-459. doi:10.1038/sj.hdy.6801023

Kalugina, N.S. (1991). New Mesozoic Simuliidae and Leptoconopidae and the origins of blood-sucking in lower dipterans. *Paleontologicheskii Zhurnal* **1**, 69-80.

Kjer, K.M. (2004). Aligned 18S and insect phylogeny. *Systematic Biology* **53**, 506-514. doi: 10.1080/1063515049445922

Krüger, A., Gelhaus, A. and Garms, R. (2000). Molecular identification and phylogeny of East African *Simulium damnosum* s.l. and their relationship with West African species of the complex (Diptera: Simuliidae). *Insect Molecular Biology* **9**, 101-108. doi: 10.1046/j.1365-2583.2000.00163.x

LaRue, B., Gaudreau, C., Bagre, H.O. and Charpentier, G. (2009). Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies (Diptera: Simuliidae). *Molecular Phylogenetics and Evolution* **53**, 749-757. doi: 10.1016/j.ympev.2009.07.032

Lucas, P. and Hunter, F. (1999). Phenotypic plasticity in the labral fan of simuliid larvae (Diptera): effect of seston load on primary-ray number. *Canadian Journal of Zoology* **77**, 1843-1849. doi: 10.1139/cjz-77-12-1843

Miller, B.R., Crabtree, M.B. and Savage, H.M. (1997). Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S sequences (Diptera: Nematocera). *Insect Molecular Biology* **8**, 105-114. doi: 10.1111/j.1365-2583.1997.tb00078.x

Morales-Hojas, R., Post, R.J., Cheke, R.A. and Wilson, M.D. (2002). Assessment of rDNA IGS as a molecular marker in the *Simulium damnosum* complex. *Medical and Veterinary Entomology* **16**, 395-403. doi: 10.1046/j.1365-2915.2002.00391.x

Morales-Hojas, R. and Krueger, A. (2009). The species delimitation problem in the *Simulium damnosum* complex, blackfly vectors of onchocerciasis. *Medical and Veterinary Entomology* **23**, 257-268.

Moulton, J.K. (2000). Molecular sequence data resolves basal divergences within Simuliidae (Diptera). *Systematic Entomology* **25**, 95-113. doi: 10.1046/j.1365-3113.2000.00097.x

Pace, N.R., Olsen, G.J. and Woese, C.R. (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**, 325-326. doi: 10.1016/0092-8674(86)90315-6

Palmer, R.W. and Craig, D.A. (2000). An ecological classification of primary labral fans of filter-feeding black fly (Diptera: Simuliidae) larvae. *Canadian Journal of Zoology* **78**, 199-218. doi: 10.1139/cjz-78-2-199

Posada, D. and Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818. Available: [http://bioag.byu.edu/zoology/crandall\\_lab/modeltest.html](http://bioag.byu.edu/zoology/crandall_lab/modeltest.html)

Pruess, K.P., Adams, B.J., Parsons, T.J., Zhu, X. and Powers, T.O. (2000). Utility of the mitochondrial cytochrome oxidase II gene for resolving relationships among black flies (Diptera: Simuliidae). *Molecular Phylogenetics and Evolution* **16**, 286-295. doi:10.1006/mpev.2000.0807

Rodríguez-Pérez, M.A., Núñez-González, C.A., Lizarazo-Ortega, C., Sánchez-Varela, A., Wooten, M.C., and Unnasch, T.R. (2006). Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH dehydrogenase subunit 4 mitochondrial genes of the

onchocerciasis vector *Simulium ochraceum*. *Journal of Medical Entomology* **43**, 701-706. doi: 10.1603/0022-2585(2006)043[0001:ASFAD]2.0.CO;2

Rothfels, K.H. (1979). Cytotaxonomy of black flies (Simuliidae). *Annual Review of Entomology* **24**, 507-539. doi:10.1146/annurev.en.24.010179.002451

Rothfels, K.H. and Freeman, D.M. (1977). The salivary gland chromosomes of seven species of *Prosimulium* (Diptera: Simuliidae) in the *mixtum* (IILL-1) group. *Canadian Journal of Zoology* **55**, 482-507. doi: 10.1139/z77-065

Schlötterer, C., Hauser, M.T., von Haeseler, A. and Tautz, D. (1994). Comparative evolutionary analysis of rDNA regions in drosophila. *Molecular Biology and Evolution* **11**, 513-522.

Shepard, J.J., Andreadis, T.G. and Vossbrinck, C.R. (2006). Molecular phylogeny and evolutionary relationships among mosquitoes (Diptera: Culicidae) from the Northeastern United States based on small subunit ribosomal DNA (18S rDNA) sequences. *Journal of Medical Entomology* **43**, 443-454. doi: 10.1603/0022-2585(2006)43[443:MPAERA]2.0.CO;2

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651-701.

Snyder, T.P. and Linton, M.C. (1983). Electrophoretic and morphological separation of *Prosimulium fuscum* and *P. mixtum* larvae (Diptera: Simuliidae). *The Canadian Entomologist* **115**, 81-87. doi: 10.4039/Ent11581-1

Spironello, M., Hunter, F.F. and Craig, D.A. (2002). A cytological study of the Pacific black fly *Simulium cataractarum* (Diptera: Simuliidae). *Canadian Journal of Zoology* **80**, 1810-1816. doi: 10.1139/Z02-167

Strachan, T., Webb, D.A and Dover, G.A. (1985). Transition stages of molecular drive in multi-copy DNA families in *Drosophila*. *EMBO Journal* **4**, 1701-1708.

St-Onge, M., LaRue, B. and Charpentier, G. (2008). A molecular revision of the taxonomic status of mermithid parasites of black flies from Quebec (Canada). *Journal of Invertebrate Pathology* **98**, 299-306. doi: 10.1016/j.jip.2008.04.001

Thanwisai, A., Kuvangkadilok, C., Baimai, V. (2006). Molecular phylogeny of black flies (Diptera: Simuliidae) from Thailand, using ITS2 rDNA. *Genetica* **128**, 177-204. doi: 10.1007/s10709-005-5702-z

Wood, D.M., Peterson, B.V., Davies, D.M. et Gyorkos, H. (1963). The Black flies (Diptera: Simuliidae) of Ontario. Part II. Larval identification, with description and illustration. *Proceedings of the Entomological Society of Ontario* **93**, 99-129.

## 4.9. Table and figures

Table 4.1. List of species

Species (no specimens)*, author	COI / COII / rDNA (Genbank no.)	Molecular validation **
<i>Cnephia dacotensis</i> (2), Dyar and Shannon 1927	GQ280358 / GQ265806 / FJ436346	Int (ITS1). Ext (COI / AF25841)
<i>Prosimulium</i> 1 (19)	GQ355808 / GQ355810 / FJ595476	Int (COI, COII, ITS1)
<i>Prosimulium</i> 2 (4)	GQ355809 / GQ355811 / FJ595477	Int (COI, COII, ITS1)
<i>Simulium arevulus</i> (4), Lundström 1911	GQ280360 / GQ265808 / FJ437566	Int (COI)
<i>Simulium alpeum</i> (3), Fries 1824	GQ280361 / GQ280348 / FJ437565	Int (ITS1). Ext (COII / AF083862)
<i>Simulium decorum</i> (2), Walker 1848	GQ280362 / GQ280349 / FJ437564	Int (COI, COII, ITS1). Ext (COII / DQ284522)
<i>Simulium elryacmiraculum</i> (6), Davies 1949	GQ280363 / GQ380350 / FJ437566	Int (COI, COII)
<i>Simulium jervingsi</i> (1), Malloch 1914	GQ280364 / GQ280351 / FJ436353	Ext (COII / DQ284509)
<i>Simulium longistylatum</i> (25), Shewell 1959	GQ280366 / GQ280352 / FJ436352	Int (COI, COII, ITS1)
<i>Simulium pictipes</i> (4), Hagen 1880	GQ280367 / GQ280353 / FJ436351	Int (COI, COII, ITS1)
<i>Simulium quebecense</i> (2), Twinn 1936	GQ280368 / GQ280354 / FJ436350	Int (ITS1)
<i>Simulium tuberosum</i> (2), Lundström 1911	GQ280369 / GQ280355 / FJ436349	Int (ITS1, ITS2). Ext (COII / DQ284515)
<i>Simulium verutum</i> (3), Say 1823	GQ280370 / GQ280356 / FJ436348	Int (COI, COII, ITS1)
<i>Simulium vitatum</i> (28), Zetterstedt 1838	FJ231202 / FJ231200 / FJ231203	Int (COI, COII, ITS1). Ext (rDNA / U48383)
<i>Stegopterna mutata</i> (2), Malloch 1914 ***	GQ280359 / GQ265807 / FJ436345	Int (ITS1)
<i>Wixenia tiddlesi</i> (1), Stone and Jamnback 1955	GQ280371 / GQ280357 / FJ436347	NA

\* Specimens with known sequences of any type (COI, COII and/or ITS1)

\*\* Int: morphological identification supported by high sequence homology between all specimens (type of sequence). Ext: independent Genbank match (type of sequence / accession no)

\*\*\* Previously known as *Cnephia mutata*; see Adler et al. (2004)

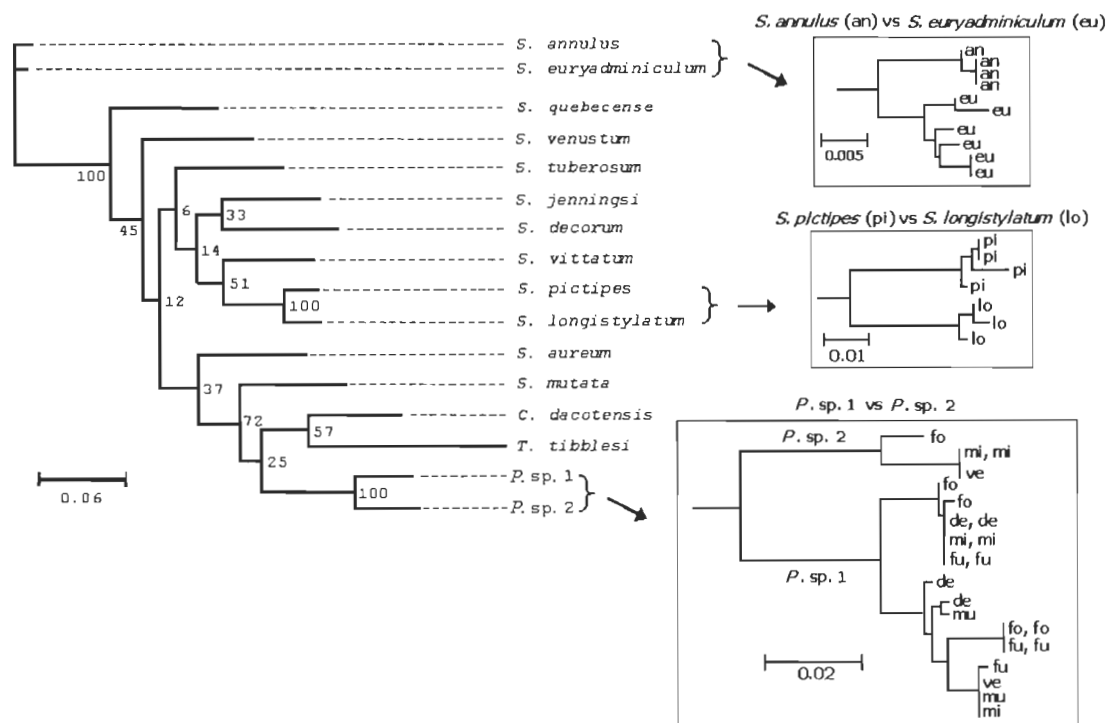
Table 4.2. Average base composition of the COI + COII genes.

	A%	T%	G%	C%	% variable
Position 1 <sup>st</sup>	27.0	27.1	29.3	16.6	18.6
Position 2 <sup>nd</sup>	19.5	41.3	15.5	23.7	2.0
Position 3 <sup>rd</sup>	42.3	41.0	4.2	12.5	85.0
All	29.6	36.5	16.3	17.6	35.2

**Table 4.3.** Amino acid interchanges in the COI and COII polypeptides. Dot: identity with the consensus sequence.

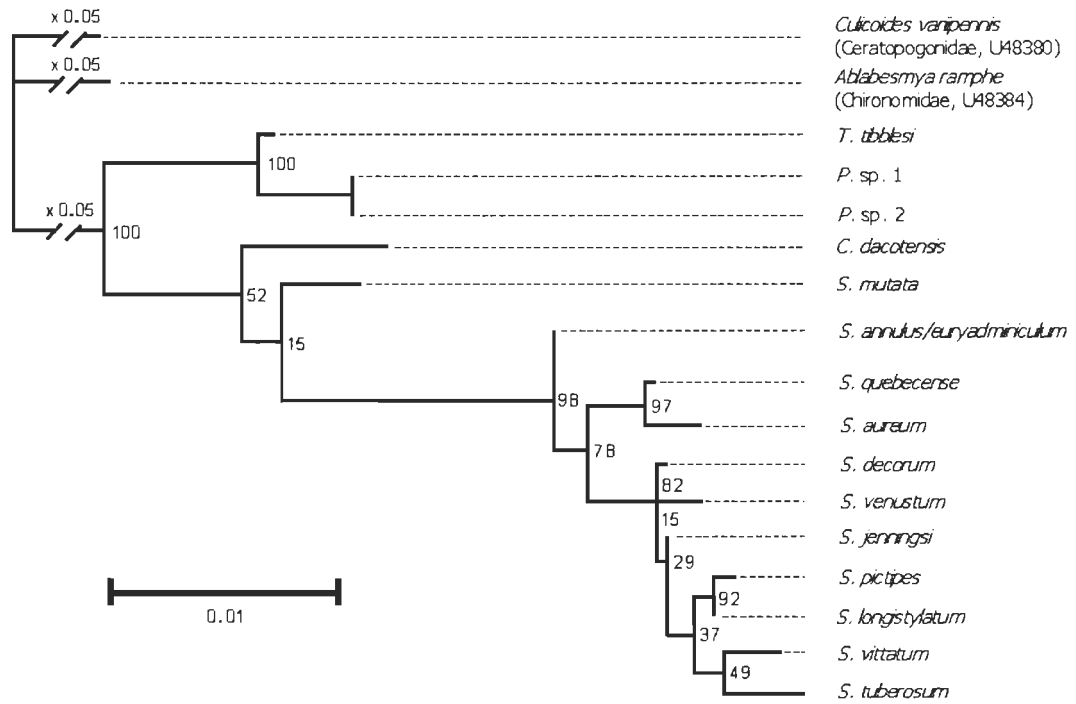
	COI																COII																			
No codon	6	6	7	7	2	3	4	4	4	4	5	7	8	8	0	0	0	1	1	1	6	8	3	6	9	3	5	7	8	3	8	3	7	8		
Consensus	T	Y	M	A	F	T	A	M	V	I	L	A	V	V	M	S	S	V	Y	I	V	T	T	V	M	I	I	V	T	L	I	V	G	L		
<i>T. tibblasi</i>	.	F	.	S	.	S	T	.	I	.	.	T	I	I	I	T	.	A	F	.	I	.	I	.	L	.	V	I	.	.	.	I	A	.		
<i>Prosimulium</i> 1	.	.	.	.	.	.	.	.	I	.	I	T	I	.	.	A	T	.	.	.	I	.	.	.	L	.	.	.	S	.	.	.	A	M	.	
<i>Prosimulium</i> 2	.	.	.	.	.	.	.	.	I	.	I	T	I	.	.	A	T	.	.	.	I	.	.	.	L	.	.	.	S	.	.	.	A	M	.	
<i>S. mutata</i>	A	.	.	.	L	.	.	.	I	.	.	T	.	.	.	.	T	.	.	.	I	.	.	I	L	.	.	.	.	.	.	.	A	.	.	
<i>C. dacotensis</i>	.	.	.	.	.	S	.	L	I	.	.	T	I	.	.	T	T	I	.	.	.	.	.	.	L	.	.	.	.	.	.	.	A	.	.	
<i>S. vittatum</i>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	L	.	.	.	.	.	.	.	.	.	.	I	V	.	.	.	
<i>S. longistylatum</i>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	.	L	I	.	M	.	.	.	.	.	.	.	I	V	.	.	.	
<i>S. pictipes</i>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	I	V	.	.	.	
<i>S. tuberosum</i>	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	S	.	.	.	.	.	.	.	.	.	V	.	.	.	.
<i>S. venustum</i>	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.
<i>S. decorum</i>	.	.	I	.	.	.	.	.	.	.	.	.	.	I	.	.	A	.	.	.	.	.	M	.	.	.	.	.	.	.	.	V	.	.	.	.
<i>S. jenningsi</i>	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>S. aureum</i>	.	.	I	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	
<i>S. annulus</i>	.	.	I	.	.	.	.	.	I	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	V	.	.	.	.	.	V	.	.	.	
<i>S. euryadmiculum</i>	.	.	I	.	.	.	.	.	I	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	V	.	.	.	.	.	V	.	.	.	
<i>S. quebecense</i>	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.

**Fig 4.1.** Unrooted tree for COI/COII. Bootstrap values (% over 500 replicates) are indicated. Insets: COI sub-trees for the *S. annulus*/*S. euryadmiculum*, *S. pictipes*/*S. longistylatum* and *P.sp.1*/*P.sp.2* groups. *Prosimulium* nominal species according to tooth pattern: fo (mi, fu, mu), *Prosimulium fontanum* (*mixtum*, *fuscum*, *multidentatum*); de (ve), *Helodon decemarticulatus* (*vernalis*).





**Fig. 4.2.** Rooted tree for rDNA (18S and 5.8S region). Bootstrap values (% over 500 replicates) are indicated. Branch lengths are set proportional to mutational distance, with the exception of outgroup sequences (20-fold size reduction).





## **CHAPITRE V**

« Molecular comparison of Quebec and Newfoundland populations of the blackfly, *Simulium vittatum*, species complex »

**Molecular comparison of Quebec and Newfoundland populations  
of the blackfly, *Simulium vittatum*, species complex**

Christine Gaudreau<sup>1</sup>, Bernard LaRue<sup>1</sup>, and Guy Charpentier<sup>1</sup>

<sup>1</sup>Département de Chimie-biologie, Université du Québec à Trois-Rivières, 3351  
Boulevard des Forges, Box 500, Trois-Rivières, Canada G9A 5H7

Correspondence:

Bernard LaRue

Box 500, Université du Québec à Trois-Rivières  
Trois-Rivières, Québec, Canada G9A 5H7

(819) 376-5052 (phone)

(819) 376-5084 (fax)

E-mail: [bernard.larue@uqtr.ca](mailto:bernard.larue@uqtr.ca)

**Key words** : *Simulium vittatum*, Black fly, COII, ITS1, mtDNA markers.

## 5.1 RÉSUMÉ

Des haplotypes spécifiques identifiés sur cinq positions des gènes mitochondriaux COI et COII différentient en partie les populations de *Simulium vittatum* selon leur provenance, en l'occurrence Québec-Ontario ou Terre-Neuve. Cette signature géographique s'ajoute à une quarantaine d'autres sites polymorphes globalement répertoriés, de sorte que la divergence entre séquences ne peut mener seule à une distinction nettement tranchée entre les deux populations. Jointe aux observations précédentes, l'apparition sporadique d'haplotypes intermédiaires entre les consensus des deux territoires respectifs suggère que la prédominance à Terre-Neuve d'un haplotype spécifique s'explique par une séquence fondatrice particulière issue du vaste réservoir de diversité typique des populations du bloc continental Nord Américain. L'espaceur interne transcrit (ITS1) de l'unité nucléaire de transcription de l'ADNr ne révèle pas mieux de démarcation géographique nette entre les deux populations que le critère basé sur COI /COII.

### 5.1.1 CONTRIBUTION DES AUTEURS DE L'ARTICLE

L'article a été rédigé par Christine Gaudreau et Bernard LaRue et révisé par Guy Charpentier. La partie expérimentale fut effectuée par Christine Gaudreau.

Ce manuscrit a été publié en 2010 dans *Medical and Veterinary Entomology*, volume 24, pages 214-217.

## 5.2 ABSTRACT

Specific haplotypes at five positions in the COI and COII mitochondrial genes allowed a partial differentiation of *Simulium vittatum* Zetterstedt (Diptera: Simuliidae) populations from Quebec-Ontario and Newfoundland, respectively. This geographical signature was superimposed on about 40 other polymorphic sites such that sequence

divergence alone did not enable a clear-cut distinction between the two populations. Together with the sporadic occurrence of haplotypes intermediate to the Newfoundland and Quebec-Ontario consensus, this suggested that one peculiar sequence among many found in populations from the North American landmass predominates in Newfoundland as a result of a founder effect. The internal transcribed spacer (ITS1) sequence from the nuclear rDNA transcription unit was no more able to resolve populations along geographical lines than the COI/COII criteria.

### 5.3 SHORT COMMUNICATION

*Simulium vittatum* (Diptera: Simuliidae) is a complex of multivoltine blackfly species widely distributed across North America and Atlantic islands such as Newfoundland and Prince Edward Island (Adler & Kim, 1984). The immature stages occur in a wide range of habitats, from small streams to medium-sized rivers. The adult female, a pest of domestic fowl and many wild and farm-raised mammals, causes significant economic losses that result from severe scabbing, irritation and ear dermatitis (Adler *et al.*, 2004). Feeding on humans remains uncommon (Wolfe & Peterson, 1959; Peterson, 1977), but *S. vittatum* s.l. can be a severe nuisance because it forms persistent swarms and enters the ears, eyes, nose and mouth. Although *S. vittatum* is a vector for the vesicular stomatitis virus and *Onchocerca* sp. (Adler *et al.*, 2004) under laboratory conditions, this role has yet to be proven in the wild.

The *S. vittatum* species complex consists of at least two cytological forms, IS-7 and IIII-I (Pasternak, 1964; Rothfels & Featherston, 1981), and the formal recognition of IS-7 as *Simulium vittatum* sensu stricto and IIII-I as *Simulium tribulatum* has been recently proposed (Adler *et al.*, 2004). This taxonomic revision left out the XoYo cytotype (undifferentiated sex chromosomes) which typifies the populations of Iceland and the northwest Atlantic islands (Rothfels & Featherston, 1981; Duncan *et al.*, 2004; Minhas *et al.*, 2005). On the North American continent, the distribution of the IS-7 and IIII-I morphological twins correlates generally with ecology and geography. The IIII-I type, which prefers warmer and less well oxygenated streams (Duncan *et al.*, 2004), extends from the Atlantic coast to the Prairies, whereas the IS-7 type is usually found further north, reaching as far as Alaska in the west. IIII-I and IS-7 coexist over a vast

area covering the south of Quebec and Ontario and the northeastern USA. In allopatry, larval populations of either one span early spring to early autumn. Under sympatric conditions, seasonal population peaks are somewhat shifted according to the preferred temperature range of each cytotype (Adler & Kim, 1984), which favours reproductive isolation.

The molecular data record is still ambiguous with respect to the existence of two distinct entities. Allozymes (Zhu *et al.*, 1998) and directed heteroduplex analysis of 12S and 16S mt rRNA genes (Tang *et al.*, 1996) failed to distinguish between the two siblings, whereas random amplification polymorphic DNA (RAPD) procedures enabled a partial separation and also suggested western and eastern origins for IS-7 and IIIIL-1, respectively (Duncan *et al.*, 2004). In an attempt to differentiate between the *S. vittatum* populations from Quebec and Ontario (IS-7/IIIIL-1) and Newfoundland (XoYo), we studied the sequence diversity of the COI and COII (cytochrome oxidase) mitochondrial genes and the intergenic spacer ITS1 located between the 18S and 5.8S nuclear rRNA genes. These targets were attractive as their variability often allows for distinguishing between siblings and characterizing population genetic structure (Day *et al.*, 2008; Gaudreau *et al.*, 2008).

Blackfly larvae were collected from five streams: two in Quebec (Rivière Richelieu, 45°25'N; 73°13'W and Lac Souris, 46°34'N; 72°58'W), one in Ontario (Rivière Rideau, 45°22'N; 75°41'W) and two in Newfoundland at the easternmost tip of North America (Broad Cove River, 47°35'N; 53°05'W and Bristol Cove Brook, 46°39'N; 53°05'W). Larvae fixed in 95% ethanol upon collection were first identified as *S. vittatum* (28 specimens overall) under a low-magnification microscope using the keys of Wood *et al.* (1963) and Adler *et al.* (2004) and secondly as of dark or light form according to head spot pattern (Adler & Kim, 1984). Head spots are confluent on the dark form but appear as separated on the light form.

DNA extraction and purification, amplification of COII or ITS1 and the recovery and sequencing of polymerase chain reaction (PCR) products were described in Gaudreau *et al.* (2008). The PCR for COI used the same conditions, except for an annealing temperature set at 52°C for the custom-designed primers SCOIL-1 (forward: ATAATTCNCATATTATTAGHCAAGARAG) and SCOIR-2 (reverse:

AGATTTTGGTATCATTCAATAGATGAATT). Alignment of DNA sequences was performed with Geneious version 3.8.5. (Drummond *et al.*, 2007). COI and COII consensus sequences for Newfoundland and Quebec-Ontario populations, respectively, have been deposited in Genbank under accession numbers FJ231199-FJ231202 and the CA<sub>6</sub>C ITS1 variant (see below) as FJ231203. The general time-reversible (GTR) phylogenetic tree for COI/COII was built with PhyML (Guindon & Gascuel, 2003).

PCR products gave readable sequences of 252 and 651 nucleotides (nt) for COII and COI, respectively. Given that both are mtDNA genes with quite similar evolution rates (Gaudreau *et al.*, 2008), COI and COII were considered as a 0.9- kb aggregate rather than separately. Forty-nine variable positions and 27 distinct haplotypes were noted within a sample including the 28 specimens from this study as well as the *S. vittatum* Iceland (AF083865, XoYo cytotype) and Nebraska (M76433, IIIL-1) COII sequences (Pruess *et al.*, 2000), which were not outstandingly distinct from the rest (Fig. 5.1). In agreement with the strong conservatism of the corresponding polypeptides (Pruess *et al.*, 2000; C. Gaudreau, B. LaRue, G. Charpentier, 2009, unpublished data), mutations were all synonymous, except for one (COI position 628) producing an A to T amino acid change. Quite remarkably, more than half of them appeared only once. COI and COII sequence diversity was outstanding in the Quebec-Ontario data, with 16 distinct variants spread between only 17 specimens. Although these findings must be interpreted with caution, given the small sampling sizes, the average sequence divergence between individuals at each Quebec-Ontario location appeared to be the same as for all the data clumped together (Table 5.1), indicating little or no local clustering of haplotypes. The Newfoundland population also showed less internal heterogeneity, especially as much of the average divergence value was contributed by a single specimen. By contrast with the Quebec-Ontario population, all other sequences were very closely related (up to two mutations) to the Newfoundland consensus sequence, which was indeed found as such in three specimens. Overall, the average and maximum divergence levels remained well within the usual limits of intraspecific COI sequence comparisons involving invertebrates (Carew *et al.*, 2007). An attempt to relate all Quebec-Ontario haplotypes in a neighbor-joining tree (data not shown) indicated that many branches involved between two and five mutations. If we assume quite reasonably



that COI/COII evolved through point mutations, this implies that many variants were either missed by the sampling in this study or simply became extinct.

Five positions enabled a distinction to be drawn along geographical lines because a consensus rule set separately for each area identified haplotypes ATAGG (from left to right in Fig. 5.1) and GCGTA as clearly predominant in Quebec-Ontario and Newfoundland, respectively. The distinction remained partial, with a Quebec-Ontario haplotype slipping into the Newfoundland sample and vice versa. A few sequences differing from either ATAGG or GCGTA at a single position were classified with the closest distinctive haplotype for calculation purposes and three others (\* in Fig 5.1) that did not adhere to the 'four out of five' rule were left out. The statistical significance of these observations, assessed by a Monte Carlo simulation ( $10^4$  individual runs) performed in Microsoft Excel software, showed the chance occurrence of such an asymmetric distribution between Newfoundland and Quebec-Ontario to be  $<0.1\%$ . A few specimens showed haplotypes that were intermediate between the respective Newfoundland and Quebec-Ontario consensus, suggesting that the Newfoundland haplotype originated through migration and gene drift from one peculiar sequence found among a pool of pre-existing North American variants. Because of the lack of suitable habitats, any ensuing founder effect occurring under an insular context must obviously be no older than the end of the last glacial episode, which took place about 12,000 years ago. As the ice sheet retreated, Quebec and Ontario would also have been repopulated, but the gene flow within the North American landmass was more likely to maintain more genetic diversity and leave less opportunity for the development of local isolates, an interpretation to be eventually validated by molecular data from all over North America. Finally, a cursory examination of Fig. 5.1 revealed no differentiation of haplotypes with respect to head spot pattern. The light and dark morphs were found in both Quebec-Ontario and Newfoundland, coexisted at several locations and sometimes corresponded to identical or very similar sequences. This was not unexpected as Zettler *et al.* (1998) and Adler & Kim (1984) have already demonstrated that larval head morphology is influenced by environmental parameters such as ultraviolet exposure, diet and substrate background color, but is not linked to cytological form or sex.

Further differentiation of DNA specimens through ITS1, located between the 18S and 5.8S genes and usually the most variable part of the nuclear rDNA transcription unit, was attempted. As Fig. 5.2 shows, sequencing revealed a length of either 103 bp or 105 bp as a result of an oligoadenylate tract alternating between four ( $A_4$  variant) and six ( $A_6$ ) A residues in length. Two additional polymorphisms occurred elsewhere, leading to A/C and C/T dichotomies, respectively. This amounted theoretically to eight distinct sequences, with possible mixes within a same individual adding a further complication as a result of a heterozygous state or an internal heterogeneity of the rDNA cluster. A third of the specimens provided a pure sequence and, in another third, the presence of a single mixed-identity position led to an unequivocal interpretation about the two current sequences. For the remaining specimens, including exceptional ones containing up to four distinct variants, denaturing polyacrylamide gel electrophoresis allowed distinct heteroduplexes accumulated during the final amplification plateau to be distinguished from one another and independently sequenced (Gaudreau *et al.*, 2008). Among the eight possible variants,  $CA_4C$  (polymorphic site identity listed 5' to 3' with respect to Fig. 5.2) predominated clearly in Quebec-Ontario and, in Newfoundland, coexisted at comparable frequency levels with  $CA_6C$ . Sequences  $AA_4C$  and  $AA_4T$  were never detected and the four remaining variants all occurred as minor components (global frequency of 0.14) mixed under a submolar ratio with either  $CA_4C$  or  $CA_6C$ , an obvious clue to rDNA cluster heterogeneity. Given this heterogeneity and the lack of a clear geographical fragmentation, ITS1 was finally deemed to be insufficiently discriminating to be used here as a population marker, quite unlike the situation frequently observed with other dipterans (Polanco *et al.*, 1998; Rodríguez-Pérez *et al.*, 2006; Gaudreau *et al.*, 2008).

Because of ethanol fixation, cytotype could not be determined. However, under the assumption of a XoYo cytotype for the Newfoundland specimens (Rothfels & Featherston, 1981), COI/COII data alone could not distinguish them from the Quebec-Ontario sample, presumably IS-7/IIIL-1, because the average divergence level between the two populations (0.0083, see Table 5.1) was similar to that within Quebec and Ontario. Thus, the XoYo cytotype and the GCGTA COI/COII haplotype would be

incidental to recent genetic isolation and would not implicate enough characters for distinct species status to be granted at present.

#### **5.4 ACKNOWLEDGEMENTS**

We are grateful to the Université du Québec à Trois-Rivières, which provided C. Gaudreau with a scholarship and funding during her research. Newfoundland specimens were a gift from M. Colbo (Memorial University).

## 5.5 REFERENCES

- Adler, P.H. & Kim, K.C. (1984) Ecological characterization of two sibling species, IILL-1 and IS-7, in the *Simulium vittatum* complex (Diptera: Simuliidae). *Canadian Journal of Zoology*, **62**, 1308-1315.
- Adler, P.H., Currie, D.C. & Wood, D.M. (2004) The blackflies (Simuliidae) of North America. ROM publication in sciences. New York, USA.
- Carew, M.E., Pettigrove, V., Cox, R.L. & Hoffman, A.A. (2007) DNA identification of urban Tanytarsini chironomids (Diptera:Chironomidae). *Journal of the North American Benthological Society*, **26**, 586-589.
- Day, J.C., Goodall, T.I. & Post, R.J. (2008) Confirmation of the species status of the blackfly *Simulium galeratum* in Britain using molecular taxonomy. *Medical and Veterinary Entomology*, **22**, 55-61.
- Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T. & Wilson, A. (2007) Geneious v.3.8.5, Available from <http://www.geneious.com/>.
- Duncan, G.A., Adler, P.H., Pruess, K.P. & Powers, T.O. (2004) Molecular differentiation of two sibling species of the black fly *Simulium vittatum* (Diptera: Simuliidae) based on random amplified polymorphic DNA. *Genome*, **47**, 373-379.
- Gaudreau, C., LaRue, B., Charbonneau, V., Charpentier, G. & Craig, D.A. (2008) Recent origins and genetic fragmentation of *Simulium* (*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid. *Invertebrate Systematics*, **22**, 555-562.
- Guindon, S. & Gascuel, O. (2003) A simple, fast and accurate algorithm to estimate 618 large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696-704.

- Minhas, M.S., Brockhouse, C.L. & Adler, P.H. (2005) The black fly (Diptera:Simuliidae) fauna of Prince Edward island, Canada. *Northeastern Naturalist*, **12**, 67-76.
- Pasternak, J.J. (1964) Chromosome polymorphism in the blackfly *Simulium vittatum* (Zett.). *Canadian Journal of Zoology*, **42**, 135-158.
- Peterson, B.V. (1977) The black flies of Iceland (Diptera:Simuliidae) *Canadian Entomologist*, **109**, 449-472.
- Polanco, C., González, A.I., de la Fuente, Á. & Dover, G. (1998) Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. *Genetics*, **149**, 243-256.
- Pruess, K.P., Adams, B.J., Parsons, T.J., Zhu, X. & Powers, T.O. (2000) Utility of the mitochondrial cytochrome oxidase II gene for resolving relationship among black flies (Diptera:Simuliidae). *Molecular phylogenetics and evolution*, **16**, 286-295.
- Rodríguez-Pérez, M.A., Núñez-González, C.A., Lizarazo-Ortega, C., Sánchez-Varela, A., Wooten, M.C. & Unnasch, T.R. (2006) Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH dehydrogenase subunit 4 mitochondrial genes of the onchocerciasis vector *Simulium ochraceum*. *Journal of Medical Entomology*, **43**, 1-6.
- Rothfels, K. & Featherston, D. (1981) The population structure of *Simulium vittatum* (Zett.): the IILL-1 and IS-7 sibling species. *Canadian Journal of Zoology*, **59**, 1857-1883.

Tang, J., Pruess, K. & Unnasch, T.R. (1996) Genotyping North American black flies by means of mitochondrial ribosomal RNA sequences. *Canadian Journal of Zoology*, **74**, 39-46.

Wolfe, L.S. & Peterson, D.G. (1959) Black flies (Diptera:Simuliidae) of the forests of Quebec. *Canadian Journal of Zoology*, **37**, 137-159.

Wood, D.M., Peterson, B.V., Davies, D.M. & Gyorkos, H. (1963) The blackflies (Diptera : Simuliidae) of Ontario. Part II. Larval identification, with description and illustration. *Proceedings of the Entomological Society of Ontario*, **93**, 99-129.

Zettler, J.A., Adler, P.H. & McCreadie J.W. (1998) Factors influencing larval color in the *Simulium vittatum* complex (Diptera:Simuliidae). *Invertebrate Biology*, **117**, 245-252.

Zhu, X. Pruess, K.P. & Powers, T.O. (1998) Mitochondrial DNA polymorphism in a black fly, *Simulium vittatum* (Diptera:Simuliidae). *Canadian Journal of Zoology*, **76**, 440-447.

## 5.6 TABLE AND FIGURES

Table 5.1. Divergence levels for the COI/COII aggregate in comparisons involving pairs of sequences (Kimura 2 parameters distance).

Type of comparison (specimens)	Average	Maximum
All (28)*	0.0082	0.023
Within Newfoundland (11)	0.0041	0.012
Within Québec-Ontario (17)	0.0084	0.022
Québec, location 4 (9)	0.0096	0.022
Ontario, location 5 (5)	0.0103	0.017

\* Omits Iceland and Nebraska COII sequences.

Location 3 omitted (only three specimens).





## CHAPITRE VI

« Recent origins and genetic fragmentation of *Simulium*  
(*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian  
simuliid »

**Recent origins and genetic fragmentation of *Simulium*  
(*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid**

Christine Gaudreau<sup>A</sup>, Bernard LaRue<sup>A,C</sup>, Valérie Charbonneau<sup>A</sup>, Guy Charpentier<sup>A</sup> and  
Douglas A. Craig<sup>B</sup>

<sup>A</sup> Département de Chimie-biologie, Université du Québec à Trois-Rivières, 3351  
Boulevard des Forges, Box 500, Trois-Rivières, Canada G9A 5H7

<sup>B</sup> Department of Biological Sciences, University of Alberta, Edmonton, Canada T6G  
2E9.

<sup>C</sup> Correspondence: Bernard LaRue  
Box 500, Université du Québec à Trois-Rivières  
Trois-Rivières, Québec, G9A 5H7 Canada.  
E-mail: [bernard.larue@uqtr.ca](mailto:bernard.larue@uqtr.ca)

**Keywords : Biogeography, black fly, COII, ITS1, Simuliidae.**

## 6.1 RÉSUMÉ

*Simulium laciniatum* Edwards, 1924, est non seulement endémique aux îles Fidji, mais aussi la seule espèce de mouche noire répertoriée en ces lieux, soit une situation très inusitée puisque l'archipel voisin du Vanuatu héberge quatorze espèces également endémiques du même sous-genre *Hebridosimulium* Grenier et Rageau, 1961. L'uniformité morphologique indique en apparence le caractère monospécifique de *S. (H.) laciniatum* sur l'ensemble de l'archipel Fidjien, le cas a été examiné plus en profondeur en adoptant une approche moléculaire. Des échantillons en provenance de cinq îles des Fidji démontrent une absence flagrante d'hétérogénéité de séquence du gène mitochondrial COII et suggèrent une date d'environ  $10^5$  années pour l'ancêtre immédiat de l'espèce actuelle, probablement un émigrant en provenance du Vanuatu. Le maintien de l'homogénéité de COII serait aussi facilité par les capacités de dispersion des femelles en quête d'un repas de sang, situation probablement amplifiée dans le passé alors que l'abaissement du niveau marin concomitant avec les pics des ères glaciaires a fortement diminué les distances entre îles. Par ailleurs, l'espaceur ITS1 situé entre les gènes des ADNr 18S et 5.8S possède six haplotypes dont les deux principaux se distribuent de façon clairement asymétrique selon un axe nord-sud. Cette dernière observation suggère que les distances entre îles lors de l'époque récente sont demeurées suffisantes pour permettre une différenciation génétique entre populations locales.

### 6.1.1 CONTRIBUTION DES AUTEURS DE L'ARTICLE

La rédaction du manuscrit fut effectuée par Christine Gaudreau, Bernard LaRue et Douglas A. Craig. Guy Charpentier a quant à lui effectué la révision. Une fraction substantielle de la partie expérimentale fut effectuée par Valérie Charbonneau dans le cadre du Stage du Baccalauréat en Biologie médicale, sous la supervision de Christine Gaudreau et Bernard LaRue.

Ce manuscrit a été publié en 2008 dans le journal *Invertebrate Systematics*, volume 22, pages 555-562.

### 6.2 ABSTRACT

*Simulium laciniatum* Edwards, 1924 is endemic to Fiji and is also the only known black fly species in the Fiji archipelago, a most unusual situation since neighboring Vanuatu harbors fourteen precinctive species from the same subgenus *Hebridosimulium* Grenier & Rageau, 1961. Morphology indicates that *S. (H.) laciniatum* is monospecific throughout the Fiji archipelago, so the species was further scrutinised from a molecular standpoint. Samples taken from five islands of Fiji show a major lack of heterogeneity in the mitochondrial COII gene and a probable date around  $10^5$  years has been inferred for the *S. (H.) laciniatum* ancestor, a likely migrant from Vanuatu. COII homogeneity may also reflect the dispersal abilities of blood-seeking females of this species, abetted possibly by shortened distances between islands during ice ages concurrent with depressed sea levels. However, the ITS1 spacer between the 18S and

5.8S rDNA genes exhibits six haplotypes, with the two predominant ones clearly arrayed in a north-south distribution. This suggests sufficient distance between suites of islands in recent times to allow precinctive population development.

### 6.3 INTRODUCTION

In the central and south-western Pacific, insular isolation and colonization by oceanic dispersal frequently lead to precinctive evolution and prevalent endemism. Species of Simuliidae from this general area, a group moderately well examined over the last two decades (Takaoka 1983, 1996, 2003; Craig 1987, 1997; Craig *et al.* 2001, 2006a, 2006b), also fit this scheme and the regional faunas have been shown to be markedly more speciose than previously thought. For instance, in Polynesia, the increase in known species from a subgenus of *Simulium* Latreille, 1802, *Inseliellum* Rubtsov, 1974, was about an order of magnitude in a study by Craig (1997). Further, mitochondrial DNA evidence indicated speciation to be rapid, with suggestions that *Simulium* (*I.*) *oviceps* Edwards, 1933 and its close relative *S. (I.) dussertorum* Craig, 1997 arose ~1.8-2.0 and 0.87 million years ago (Mya) respectively (Joy *et al.* 2007).

For Vanuatu, an archipelago that was never part of a continental landmass, Craig *et al.* (2006b) showed a marked increase, from one to fourteen, in the number of recognisable simuliid morphospecies, all from the *Hebridosimulium* Grenier & Rageau, 1961 subgenus of *Simulium*. It was noted that collections were peripheral on the islands and that more species could be expected. The originally described species, named *S. (H.) jolyi* by Roubaud (1906), was from Malekula, New Hebrides (now Vanuatu). *Simulium*

*(H.) jolyi* and *S. (H.) laciniatum* Edwards, 1924, later collected in Viti Levu, Fiji, were raised by Grenier and Rageau (1961) to genus level as *Hebridosimulium*, which was finally relegated to subgeneric status by Crosskey (1967) along with other such segregates of *Simulium* from the western Pacific. The phylogenetic position of *Hebridosimulium* is not clear, a point well made by Takaoka (1996) in an examination of the distribution of world Simuliidae. He considers the subgenera *Gomphostilbia* Enderlein, 1921, *Morops* Enderlein, 1930, *Wallacellum* Takaoka, 1983, *Inseliellum* and *Hebridosimulium* to all have a common ancestor, noting that the first two subgenera have distributions that centre on Papua New Guinea, with the last three having exclusively insular distributions. *Hebridosimulium* inhabits only Vanuatu and Fiji, with the possible exception of a single enigmatic *S. (H.) anaticum* specimen collected in Tahiti (Craig, 1987). The larvae of *Hebridosimulium* do not show the morphological modifications seen in those from *Inseliellum*, of which many species have adapted to cascade dwelling. Indeed, larvae of *Hebridosimulium* have a rather generalized structure.

The Fiji archipelago, located on the Melanesian Arc system where the edges of the Indo-Australian and Pacific Plates meet, has been considered by Evenhuis and Bickel (2005) to harbour one of the most unusual arthropod fauna of the south-western Pacific. Given the species radiation seen in Vanuatu, the occurrence in Fiji of a single black fly species, the endemic *S. (H.) laciniatum*, then appears as most intriguing. Since many simuliids are comprised of species that are cryptic or members of complexes (Crosskey 1990; Adler *et al.* 2004), is indeed *S. (H.) laciniatum* a single species? Adults of *Hebridosimulium* species have a highly conserved morphology and are difficult to separate, a problem that led to early taxonomic difficulties (Dumbleton 1973; Crosskey 1974; Craig 1987). Most of the useful morphological characters are embodied in the

larval stage. For *S. (H.) laciniatum*, Craig *et al.* (2006b) could find little to distinguish any population anywhere, Fiji-wide. Slight colour differences in the larvae were of no significance because colour was sex-linked in some populations and, similarly, small differences in hairiness of adults were not consistent. Numbers of hypostomal setae in the larvae were enhanced in the Kadavu population, but no other character distinguished that from others.

That morphological uniformity suggested the presence of *S. (H.) laciniatum* in Fiji to be a recent phenomenon involving too little elapsed time for evolution of discrete populations or species. In the present investigation, we probed this apparent lack of heterogeneity from a molecular standpoint by using the cytochrome oxidase subunit II (COII) mitochondrial gene and the intergenic spacer ITS1, located between the 18S and 5.8S nuclear rDNA genes.

## 6.4 METHODS

### 6.4.1 Sample collection

The *S. (H.) laciniatum* larval specimens came from a larger collection, sampled November-December 2004 as described by Craig *et al.* (2006b). Voucher specimens from this collection were deposited at the Bernice P. Bishop Museum (Honolulu, Hawaii). Larvae of *S. (H.) laciniatum* were fixed in 98% ethyl alcohol soon after collecting. The material of the present investigation comprised 27 specimens from eleven locations on five islands (Table 6.1). The majority of larvae were from Viti Levu and Vanua Levu, the two largest Fijian islands (Fig. 6.1). The smaller islands of Taveuni

and Ovalau provided four specimens each and the southernmost island of Kadavu, only two.

#### 6.4.2 DNA amplification

DNA that had been extracted and purified as described by St-Onge *et al.* (2008) was recovered in 50  $\mu$ L of sterile water and stored at  $-20^{\circ}\text{C}$  until further needed. COII amplification, including the use of C2-J-3407 and C2-N-3686 primers, was performed following Tang *et al.* (1998) with a fifty-fold final dilution of the DNA in the reaction mixture. For ITS1, the PCR was set in a 50  $\mu$ L volume containing 200  $\mu$ M of each deoxyribonucleotide triphosphate (dNTP), 0.2  $\mu$ M of each primer, 2 units of *Taq* DNA polymerase (Roche, [www.roche.com](http://www.roche.com)) and 1  $\mu$ L of DNA, all in a buffer consisting of 10-mM Tris-HCl, pH 8.3, 50 mM KCl and 1.5 mM  $\text{MgCl}_2$ . The thermocycling profile started with a 180 s denaturation step at  $95^{\circ}\text{C}$ , followed with 38 amplification cycles (30 s at  $95^{\circ}\text{C}$  + 50 s at  $52^{\circ}\text{C}$  + 60 s at  $72^{\circ}\text{C}$ ) and ended by a final extension of 120 s at  $72^{\circ}\text{C}$ . ITS-1s (forward: CCGTAGGTGAACCTGCGGAAGG), hybridizing at -8 before the 3' end of the 18S rDNA gene, and ITS-2s (reverse: TAGCTGCGGTCTTCATCGACCC), annealing at +30 downstream relative to the 5.8S rDNA start, were our custom-designed ITS1 primers.

#### 6.4.3 Purification and sequencing of PCR products

The COII or ITS1 PCR product was freed from contaminants by electrophoresis on a 1% agarose gel, after which the band of interest was cut and the DNA extracted using the QIAquick PCR purification kit (QIAGEN, [www.qiagen.com](http://www.qiagen.com)). ITS1 products



were also characterized by a 90- min electrophoresis at  $14 \text{ Vcm}^{-1}$  on a 5% acrylamide, 2.7 M urea gel. Depending on the input DNA being used, it revealed from one to four bands after staining with ethidium bromide. Individual bands from a few specimens were cut out and the gel slices soaked at  $50^{\circ}\text{C}$  for 60 min in 250  $\mu\text{L}$  of 0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1% sodium dodecyl sulfate (SDS), 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 ('elution buffer' of Sambrook *et al.* 1989). The liquid phase was recovered and DNA cleansing then proceeded as described above for agarose-extracted material. The purified DNA was sequenced at the automated sequencing facility of Laval University, Québec using standard dye-termination methods.

#### *6.4.4 Analysis of sequence data*

Alignment of COII DNA sequences was performed using the global alignment program Geneious version 3.8.5 (Drummond *et al.* 2007). ITS1 sequences were aligned manually since, in addition to their short length, all but one showed total identity except for a small microsatellite region. Sequences were deposited in GenBank under accession numbers EU779823 and EU779824.

## **6.5 RESULTS**

### *6.5.1 COII*

The amplification product gave a readable sequence for nucleotides 1981 through 2232 (84 codons), as based on the numbering system of the *Cnephia dacotensis*

Dyar & Shannon, 1927 COII GenBank sequence (AF425841). As expected for insect mtDNA, nucleotide frequencies were strongly AT-biased at the third codon position, with T = 45.2%, A = 45.2%, G = 2.4% and C = 7.2%. The population as a whole revealed two unusual features relative to the sequence diversity normally seen for the related COI marker. The sequence showed up 100% identical in 20 out of 27 specimens and, more importantly, each of the seven remaining sequences differed from the previous one by a distinct single-base mutation, which likely identified the most common variant as their direct progenitor (Fig. 6.2). Five of the seven mutations were synonymous and the two others resulted into V to I and A to T amino acid changes respectively. Although three mutations were recorded for the Vanua Levu sample (one-third of all specimens), each of the remaining four islands yielded a single one. Given the diversity of sampling locations, it was judged that an increase in the number of specimens or collection sites was unlikely to alter this global picture of a minimal COII gene diversity. The observed lack of clustering indicated that populations could not be differentiated through specific COII haplotypes, a conclusion already reached by Cywinska (unpubl. obs.; quoted in Craig *et al.* 2006b) using instead the COI marker.

The average COII divergence in comparisons between pairs of DNA sequences stood at 0.21%. This level of intraspecific variability falls within the inferior range of values usually reported in invertebrates for COI, one of the favorite genes for the bar-coding of species (Hebert *et al.* 2003). To more accurately assess the meaning of the COII figure for *S. (H.) laciniatum*, a parallel study (sequences not shown) was undertaken on the same gene segment from *Simulium longistylatum* Shewell, 1959 (now proposed to be recognised as *S. pictipes* Hagen, 1880, see Adler *et al.* 2004) and *Simulium vittatum* Zetterstedt, 1838, two North American simuliids. Using a comparable

number of specimens (24 and 28 respectively), the intraspecific divergence levels averaged 0.36% and 0.86%. Only synonymous base substitutions were observed, a strong indication (but not a stringent proof) that the COII gene evolves mostly in a neutral fashion. Conversely, the probability of a total sequence identity between any two individuals picked at random stood at 0.54, 0.27 and 0.09 for *S. (H.) laciniatum*, *S. longistylatum* and *S. vittatum*, respectively, again showing less diversity for the Fijian simuliid. In addition to mutations recorded only once, *S. longistylatum* and *S. vittatum* also exhibited a diversity of haplotypes found repeatedly and often differing from each other by several mutations, likely the result of a longer evolutionary history than for *S. (H.) laciniatum*.

A hypothetical tree for the *S. (H.) laciniatum* sequences would take the simplest possible shape, with a few individual branches, each only one mutation in length, emerging from the most common sequence taken as their direct ancestor. It was also supposed that each of the actual sequences originated from a distinct maternal line of descent, still a reasonable approximation given the number and dispersal of collection sites relative to total sample size. This allowed a global treatment of the data envisioned as a sequence space of 27 initially identical copies constituting a mutation target. The mutation rate for COII was evaluated by comparing COII and COI divergence for various pairs of North American black flies spanning a large spectrum of relatedness. While doing this, interspecific differences were found to occur mostly as synonymous changes, strongly suggesting that the two genes evolve neutrally. As shown by Fig. 6.3, COI and COII divergences were linearly related even when levels close to mutation saturation were reached. Using the average rate of  $1.15 \times 10^{-8}$  nucleotide substitutions  $\text{site}^{-1} \text{year}^{-1}$  currently suggested for mtDNA in arthropods (Brower 1994), the slope seen

in Fig. 6.3 yielded for COII a slightly higher rate of  $1.23 \times 10^{-8}$ . Assuming this last value to apply to *S. (H.) laciniatum* as well,  $8.4 \times 10^4$  years would be required on the average for the seven recorded mutations to occur over the available COII sequence space. The previous definition of the mutation target allowed the use of the Poisson distribution to simulate the data, which led to lower and upper age estimates of  $4.1$  and  $17.2 \times 10^4$  years (95% confidence level) for the most recent common ancestor of the actual *S. (H.) laciniatum* population. However, these remarkably low values could be unexpectedly biased by unwarranted assumptions introduced into the model. To get an independent confirmation, a comparison was drawn with the *S. (I.) oviceps* river ecotype dated back to a common ancestor arisen 0.89 Mya (mean estimate; Joy *et al.* 2007) as based on COI sequence data. The phylogenetic tree from the *S. oviceps* clade showed several ramifications and was rooted seven to eight times deeper than the one from *S. (H.) laciniatum*. Although this comparison provided only a semiquantitative estimate, it strongly suggested for the most recent common ancestor of *S. (H.) laciniatum* a maximum age of  $\sim 2 \times 10^5$  years. An identical conclusion was reached when using instead *S. (I.) dussertorum* (mean age: 0.87 Mya) for the outside comparison.

### 6.5.2 ITS1

Electrophoresis of PCR products from ITS1 on acrylamide-urea gels generated a total of five distinct profiles, from A showing a single band to B-E with three to four (Fig. 6.4a). The five profiles shared a common band with an apparent size of  $\sim 195$  bp and each one was perfectly reproducible when amplifying again the same input DNA material, excluding possible artifacts arising from erratic *Taq* DNA polymerase slippage

upon encountering repeated sequences during the PCR. Each band from unique specimens having the B (its upper band being a barely resolved doublet) and D profiles, respectively, was individually extracted and amplified again. Instead of the single band just extracted, the products from the second amplification showed upon electrophoresis the same complex profile as before, a seemingly awkward result clarified by sequencing from both directions. For each specimen, the two slow-moving bands then appeared as reciprocal heteroduplexes of two slightly divergent sequences, while the 195 bp band was an unresolved mixture of the corresponding homoduplexes. The electrophoretic data were explained in retrospect by an early PCR plateau giving way to subsequent melting and annealing cycles, hence the final mixture of four various structures. Except for hybrid molecules generated here during the PCR itself, this result mimicked exactly a classical heteroduplex analysis during which two homologous sequences are separately amplified and then mixed before undergoing a last thermal cycle.

For the remaining samples, we simplified the procedure by directly sequencing the single band always seen on agarose gels. It then appeared that each A-type profile consisted of a unique sequence while B-E indicated mixtures of two sequences. ITS1 length varied between extremes of 110 and 117 bp and all sequences save one differed only by a microsatellite region containing five to eight AG repeats followed by seven or eight consecutive A residues (Fig. 6.4*b*). For mixed sequences, the individual components could still be retrieved from the sequencing electrophoregram after hitting the microsatellite region by reading manually the longest variant as the 'echo' of the shortest one. When this occurred, the ratio between variants always stayed close to unity as judged from the peak intensities of the echoing one relative to its shorter counterpart. Following others (Polanco *et al.* 1998; Rodríguez-Pérez *et al.* 2006), these findings

allowed the treatment of variants as alleles by assuming that most copies of the transcription unit were internally homogenised within each of the two rDNA clusters. Complex acrylamide gel profiles thus defined the ‘heterozygous’ state, and the single band of profile A indicated the specimen to be ‘homozygous’. Alleles were designated as  $(AG)_M A_N$  according to the numbers of AG and A repeats. Only one allele (‘special’ in Fig. 6.4b) did not fit this standard pattern, showing both an abnormal microsatellite region and an AC insertion located elsewhere. It was found in a single heterozygous specimen having the COII consensus sequence, such that the presence of an unrelated intruder in the *S. (H.) laciniatum* sample could be excluded.

Five of the eight theoretical  $(AG)_{5-8}A_{7-8}$  alleles were detected among the 27 specimens pool (Table 6.2) and their frequency pattern was clearly dominated by the  $(AG)_6A_8$  and  $(AG)_7A_8$  versions found under both the homozygous and heterozygous conditions. A casual inspection of data revealed no obvious correlation between the altitude of the collection site and the presence/absence of a given allele type. However, a clear geographical division stood out as allele  $(AG)_6A_8$ , found at every collection site of the three southernmost islands -Viti Levu, Ovalau and Kadavu- and having a global frequency of 0.5 for this area, did not show up at all in our sampling for the northern group of Vanua Levu and Taveuni. Given the dispersal of collection sites, a local sampling bias appeared an unlikely explanation for such an uneven distribution. Following a computer simulation to sample thousands of fictitious populations generated by randomly picking one of the actual 54 alleles at a time, the empirical probability of such a chance event was estimated at less than one in  $10^4$ . In addition, the Poisson distribution indicated that the northern frequency of  $(AG)_6A_8$ , if some occurred there, could be no greater than 0.11 (95% probability). In spite of the small figures involved,

the four rarer alleles (including 'special'), all from heterozygous individuals, also seemed to cluster preferentially on the northern group of islands. On a purely random basis, a distribution showing an asymmetry such as found or greater should occur with a probability of ~2% only.

## 6.6 DISCUSSION

Two DNA markers, COII and ITS1, provide distinct perspectives on *S. (H.) laciniatum*. As shown by Pruess et al. (2000) and by our data, the COII coding unit seemingly evolves in a neutral fashion for other Simuliidae and should accordingly perform as a molecular clock attuned to divergence time. The lack of COII heterogeneity seen in *S. (H.) laciniatum* could then hardly be attributed to selection, hence allowing an age inference for its last common ancestor. On the other hand, the nucleolar organizer found on two homologous chromosomes from the same pair in *S. (H.) laciniatum* (Craig et al. 2006b) contains many rDNA tandem copies evolving in a concerted fashion. Apart from the eight short consensus blocks shown in Fig. 6.4b, the ITS1 sequence from Simuliidae generally obeys few selective constraints and appears as simple stuffing material inserted between the conserved 18S and 5.8S genes (B. LaRue, unpubl, data). Up to 25% of the *S. (H.) laciniatum* sequence is made from homopolymeric tracts of four or more A residues in addition to a near-perfect 14-15 bp repeat, indicating an evolutionary scheme predominantly based on indels. This sequence context likely set the stage for recombination slippages occurring as sudden, unpredictable events, which led to the creation and subsequent establishment of the (AG)<sub>M</sub>A<sub>N</sub> microsatellite variants.

Such ITS1 isolates are also known for continent-based dipterans, including *Drosophila melanogaster* Meigen, 1830 (Polanco *et al.* 1998) and *Simulium ochraceum* Walker, 1861 (Rodríguez-Pérez *et al.* 2006), which shows that an insular setting does not constitute the prerequisite for this to happen. Two conflicting hypotheses, both implicating an interruption of gene flow, can explain the apparent absence of the (AG)<sub>6</sub>A<sub>8</sub> allele in the northern group of islands. This variant either appeared south first and did not spread north, or was lost through drift within the original colonising northern population and was not reintroduced later.

Notwithstanding the lack of a *Hebridosimulium* phylogeny, the precinctive radiation into fourteen species seen for Vanuatu strongly suggests movement of a *S. (H.) laciniatum* ancestor from there to Fiji, now located 1200 km south-east to Vanuatu. From the COII-inferred age of  $\sim 8.4 \times 10^4$  years only, one could argue for a synonymy between *S. (H.) laciniatum* and some species from Vanuatu. Indeed, morphological similarities apparently equated *S. (H.) jolyi* with *S. (H.) laciniatum* (Crosskey 1974) at a time when only these two *Hebridosimulium* species, collected as adult female specimens, were known. This was later dismissed following a redescription of the subgenus, including larvae which provided most of the discriminating characters (Craig *et al.* 2006b). Geological reconstructions show that a distance of  $\sim 100$  km only separated Viti Levu (Fiji) from Malekula (central province of proto-Vanuatu)  $\sim 7.5$  Mya (Dickinson 2001; Begg and Gray 2002), but this relationship is long gone owing to the expansion of the North Fiji Basin initiated at the same time. Elevated land was then already available on Viti Levu and became so  $\sim 5.8$  Mya for Vanua Levu and 2.8 Mya for Kadavu, with Taveuni and Ovalau in between. The hinterland of the larger and older



Viti Levu is complex and provides many aquatic habitats, including mature running water systems such as large rivers that are formed over long periods of time (D.A. Craig, pers. obs.). Similarly for Taveuni, the amount of erosion into the underlying basalt is concomitant with continuous running water since the formation of the island. This relatively stable environment differs much from the younger Tahiti and other Polynesian islands where massive erosion and regular episodes of volcanism continuously remodel the landscape, which incidentally resulted for larvae of some black fly species, in specialised adaptations such as cascade dwelling. Yet, these geological hot spots had enough time to develop a lot of endemics, often over short distances, while the older Fiji harbours a single simuliid species. In summary, suitable habitats in Fiji were available to colonisation for the last few millions of years and the presence of the singular *S. (H.) laciniatum* remains an oddity.

The simplest explanation of the young age inferred for the last common ancestor of *S. (H.) laciniatum* would involve an accidental and recent introduction relying on long-range dispersal mechanisms such as wind or birds. The question then revolves around explaining the lack of subsequent development within Fiji of COII endemic haplotypes. Central to the MacArthur and Wilson (1963, 1967) model of island biogeography is influence of distance between land masses – the greater the distance the less gene flow, with adaptive radiation then occurring in more distant islands. The effect of distance has been investigated for simuliids of Polynesia (Craig *et al.* 1995; Craig 2003; Spironello and Brooks 2003) and North American islands (Adler *et al.* 2005), leading to the general conclusions that gene flow between islands occurs rather freely for distances up to ~20 km and becomes quite restricted above ~100 km. However, there is

exception to this since a 15 km distance between Hiva Oa and Mohotani in the Marquesas Islands appears sufficient to have allowed precinctive development of *S. (I.) englundii* on Mohotani from *S. (I.) gallinum* on Hiva Oa (Craig 2004). In short, ~40 km is of sufficient distance to restrict gene flow, but not colonisation, and a distance above ~200 km frequently leads to endemism. The vicious biting behaviour of *S. (H.) laciniatum* females, which sometimes fly kilometres from breeding habitat to obtain a blood meal for egg development (Crosskey 1974; D.A. Craig, pers. obs.), could have enhanced gene flow. Having already mated before feeding, the female is essentially independent in the following search for an egg-laying habitat. This behaviour is quite common in biting simuliids (Wenk 1981), but non-biting or autogenous female simuliids rarely fly far from the breeding site. Similar behaviour is seen in *S. (I.) buissoni* Roubaud, 1906 and *S. (I.) gallinum* Edwards, 1932, both serious biters in the Marquesas, and widely distributed. The same applies to *S. (I.) cheesmanae* Edwards, 1927, the only well substantiated biting simuliid on Tahiti, where it has been collected remarkably widely on the island. The other non-biting species on Tahiti are rarely encountered more than a few metres from running water. Elsewhere in the world, female simuliids documented as travelling hundreds of kilometres (Crosskey 1990) are always blood feeders. Since searching for a blood meal must raise the probability of wind dispersal, often associated with weather fronts, a lack of COII haplotypes as seen now for *S. (H.) laciniatum* would be expected.

For flying insects in general, the gene exchange may have been periodical during the last 0.5 million years owing to at least five ice ages, which entailed a maximum sea-level drop of 130 m (Voris 2000). Since much of Fiji is surrounded mostly by shallow

coral lagoon, Vanua Levu, Viti Levu Ovalau and Taveuni became almost connected to each other at this low point, with the exception of some 10 km of the Bligh Water (Fig. 6.1). Kadavu, on the other hand, remained as well separated from Viti Levu as it is now. Although low relief land does not generally provide the high velocity running water preferred by larvae of *S. (H.) laciniatum* (Craig *et al.* 2006b), dispersing adults would not have to fly over great distances of sea. Such broad land connections as seen in Fiji did not occur in Vanuatu, since the steep submarine topography of the islands mostly precluded that (Woodward and Naibitakele 1995), thus restricting dispersal and explaining endemism.

Given that the COII-inferred age refers to the last common female ancestor, a severe genetic bottleneck occurring around that time could as well explain the lack of heterogeneity in the actual population. As discussed previously, a temporary and radical decrease of habitat availability can likely be ruled out. On the other hand, one cannot entirely dismiss the possibility of a symbiont-induced sweep. As demonstrated for the *Wolbachia*-infected *Drosophila recens* Wheeler, 1960 (Shoemaker *et al.* 2004), the mitochondrial genome of the first infected female then spreads to the entire population and becomes the founder for all subsequent mtDNA haplotypes. Several similar observations have even led to the radical proposal that mtDNA should be dismissed for bar-coding purposes at the species level (Hurst and Jiggins 2005) and be replaced by nuclear markers since those, contributed equally from each parent, are not subjected to such sweeps. Although fitting nicely to the actual differences between COII and ITS1, this scenario remains very hypothetical owing to the lack of supporting data. It should also be reminded that ITS1, as part of a co-evolving unit, has peculiarities of its own with respect to a tempo of sudden changes.

In summary, the bulk of evidence suggests that *S. laciniatum* colonised Fiji from Vanuatu quite recently, at most  $2 \times 10^5$  years ago. Although explaining the low COI and COII heterogeneity, this short period of time still allowed for the segregation of distinct ITS1 haplotypes between northern and southern island groups, an event most likely following the climax of the last ice age when most of the archipelago was nearly connected through land bridges.

## 6.7 ACKNOWLEDGEMENTS

We acknowledge the research and student support provided by the Université du Québec à Trois-Rivières. Financial support for D.A. Craig was from a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

## 6.8 REFERENCES

- Adler, P.H., Currie, D.C., and Wood, D.M. (2004). 'The blackflies of North America'. (Cornell University Press, Ithaca, NY, USA)
- Adler, P.H., Giberson, D.J., and Purcell, L.A. (2005). Insular black flies (sic) (Diptera: Simuliidae) of North America: test of colonization hypotheses. *Journal of Biogeography* **32**, 211-220. doi: 10.1111/j.1365-2699.2004.1156.x
- Begg, G., and Gray, D.R. (2002). Arc dynamics and tectonic history of Fiji based on stress and kinematic analysis of dikes and faults of the Tavua Volcano, Viti Levu Island, Fiji. *Tectonics* **21**, 1023. 10.1029/2000TC001259. doi:10.1029/2000TC001259
- Brower, A.V.Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the USA* **91**, 6491-6495. doi: 10.1073/pnas.91.14.6491
- Craig, D.A. (1987). A taxonomic account of the black flies (Diptera: Simuliidae) of the Society islands – Tahiti, Moorea and Raiatea. *Questiones Entomologicae* **23**, 372-492.

- Craig, D.A. (1997). A taxonomic revision of the Pacific black fly subgenus *Inseliellum* (Diptera: Simuliidae). *Canadian Journal of Zoology* **75**, 855-904. doi: 10.1139/z97-108
- Craig, D.A. (2003). Geomorphology, development of running water habitats and evolution of black flies on Polynesian islands. *BioScience* **53**, 1079-1093. doi: 10.1641/0006-3568(2003)053[1079:GDORWH]2.0.CO;2
- Craig, D.A. (2004). Three new species for *Inseliellum* (Diptera: Simuliidae) from Polynesia. *Zootaxa* **450**, 1-18.
- Craig, D.A., Fossati, O., and Séchan, Y. (1995). Black Flies (Diptera: Simuliidae) of the Marquesas Islands, French Polynesia: redescrptions and new species. *Canadian Journal of Zoology* **73**, 775-800. doi:10.1139/z95-091
- Craig, D.A., Currie, D.C., and Joy, D.A. (2001). Geographical history of the central-western Pacific black fly subgenus *Inseliellum* (Diptera: Simuliidae: *Simulium*) based on a reconstructed phylogeny of the species, hot-spot archipelagoes, and hydrological considerations. *Journal of Biogeography* **28**, 1101-1128. doi: 10.1046/j.1365-2699.200100619.x
- Craig, D.A., Englund, R., and Takaoka, H. (2006a). Solomon Island Simuliidae: new records and species. *Zootaxa* **1328**, 1-26.
- Craig, D.A., Currie, D.C., Hunter, F.F., and Spironello, M. (2006b). A taxonomic revision of the southwestern Pacific subgenus *Hebridosimulium* (Diptera: Simuliidae: *Simulium*). *Zootaxa* **1380**, 1-90.
- Crosskey, R.W. (1967). The classification of *Simulium* Latreille (Diptera: Simuliidae) from Australia, New Guinea and the western Pacific. *Journal of Natural History* **1**, 23-51. doi:10.1080/00222936700770611

- Crosskey, R.W. (1974). The identity and taxonomic status of the man-biting black-fly of Fiji (Simuliidae: Diptera). *Journal of Entomology* **43**, 149-158 [B]
- Crosskey, R.W. (1990). 'The natural history of blackflies' (J. Wiley and Sons: New York, USA.)
- Dickinson, W.R. (2001). Petrology and geological provenance of sand tempers in prehistoric potsherds from Fiji and Vanuatu, South Pacific. *Geoarchaeology* **16**, 275-322. doi: 10.1002/1520-6548(200103)16:3<275::AID-GEA1005>3.0.CO;2-E
- Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kears, M., Moir, R., Stones-Havas, S., Thierer, T., and Wilson, A. (2007). *Geneious* v3.0. Available from <http://www.geneious.com>.
- Dumbleton, L.J. (1973). *Simulium laciniatum* Edwards - a valid species (Simuliidae: Diptera). *Journal of the Royal Society of New Zealand* **3**, 453-456.
- Evenhuis, N.L., and Bickel, D.J. (2005). The NSF-Fiji terrestrial arthropod survey: overview. Fiji Arthropods I. Edited by Neil L. Evenhuis & Daniel J. Bickel. *Bishop Museum Occasional Papers* **82**, 3-25.
- Grenier, P., and Rageau, J. (1961). Un nouveau genre de Simulies: *Hebridosimulium* N.G. (Diptera, Simuliidae) des Nouvelles-Hébrides. Description des stades préimaginaux et du mâle, redescription de la femelle de *H. jolyi* (Roubaud), 1906. *Bulletin de la Société de Pathologie exotique (1990)* **54**, 95-102.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. (2003). Biological identification through DNA barcodes. *Proceedings of the Royal Society of London* **270**, 313-321. doi: 10.1098/rsbp.2002.2218

- Hurst, G.D.D., and Jiggins, F.M. (2005). Problems with mitochondrial DNA as a marker in population phylogeographic and phylogenetic studies: the effect of inherited symbionts. *Proceedings of the Royal Society of London. Series B. Biological sciences* **272**, 1525-1534. doi: 10.1098/rsbp.2005.3056
- Joy, D.A., Craig, D.A., and Conn, J.E. (2007). Genetic variation tracks ecological segregation in populations of Pacific island black flies. *Heredity* **99**, 452-459. doi: 10.1038/sj.hdy.6801023
- MacArthur, R.H., and Wilson, E.O. (1963). An equilibrium theory of insular zoogeography. *Evolution* **17**, 373–387. doi: 10.2307/2407089
- MacArthur, R.H., and Wilson, E.O. (1967). ‘The Theory of Island Biogeography’ (Princeton University Press: Princeton, NJ, USA.)
- Polanco, C., González, A.I., de la Fuente, Á., and Dover, G. (1998). Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. *Genetics* **149**, 243-256.
- Pruess, K.P., Adams, B.J., Parsons, T.J., Zhu, X., and Powers, T.O. (2000). Utility of the mitochondrial cytochrome oxidase II gene for resolving relationships among black flies (Diptera:Simuliidae). *Molecular Phylogenetics and Evolution* **16**, 286-295. doi: 10.1006/mpev.2000.0807
- Rodríguez-Pérez, M.A., Núñez-González, C.A., Lizarazo-Ortega, C., Sánchez-Varela, A., Wooten, M.C., and Unnasch, T.R. (2006). Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH dehydrogenase subunit 4 mitochondrial genes of the onchocerciasis vector *Simulium ochraceum*. *Journal*



*of Medical Entomology* **43**, 1-6. doi : 10.1603/0022-2585(2006)043[0001 :ASFIAD]2.0.CO;2

Roubaud, M.E. (1906). Sur deux types intéressants de Simuliides de l'Afrique équatoriale et des Nouvelle-Hébrides. *Bulletin du Museum d'Histoire Naturelle* **12**, 140-143.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). 'Molecular cloning. A laboratory manual', 2<sup>nd</sup> edn. (Cold Spring Harbor Laboratory Press, New York, USA.)

Shoemaker, D.D., Dyer, K.A., Ahrens, M., McAbee, K., and Jaenike, J. (2004). Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics* **168**, 2049-2058. doi: 10.1534/genetics.104.030890

Spironello, M., and Brooks, D.R. (2003). Dispersal and diversification: macroevolutionary implications of the MacArthur–Wilson model, illustrated by *Simulium (Inseliellum) Rubstov* (Diptera: Simuliidae). *Journal of Biogeography* **30**, 1563-1573. doi: 10.1046/j1365-2699.2003.00945.x

St-Onge, M., LaRue, B., and Charpentier, G. (2008). A molecular revision of the taxonomic status of mermithid parasites of black flies from Quebec (Canada). *Journal of Invertebrate Pathology* **98**, 299-306. doi: 10.1016/j.jip.2008.04.001

Takaoka, H. (1983). 'The blackflies (Diptera: Simuliidae) of the Philippines' (Japan Society for the Promotion of Science, Tokyo, Japan.)

Takaoka, H. (1996). The geographical distribution of the genus *Simulium* Latreille in the Oriental and Australasian regions. *Japanese Journal of Tropical Medicine and Hygiene* **24**, 113-124.

- Takaoka, H. (2003). 'The Black Flies (Diptera: Simuliidae) of Sulawesi, Maluku and Irian Jaya' (Kyushu University Press, Fukuoka, Japan.)
- Tang, J., Moulton, J.K., Pruess, K., Cupp, E.W., and Unnasch, R. (1998). Genetic variation in North American black flies in the subgenus *Psilopelmia* (*Simulium*: Diptera: Simuliidae). *Canadian Journal of Zoology* **76**, 205-211. doi: 10.1139/cjz.76-2-205
- Voris, H.K. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time duration. *Journal of Biogeography* **27**, 1153-1167. doi: 10.1046/j.1365.2699.2000.00489.x
- Wenk, P. (1981). 'Bionomics of adult blackflies. Blackflies: the future for biological methods in integrated control' (Ed. M. Laird) pp. 259-279. (Academic Press, London, UK.)
- Woodward, P., and Naibitakele, N. (1995). Bathymetric map of Vanuatu – EEZ. 1:2,500,000. *SOPAC, Bathymetric Series Map 24*.

## 6.9 TABLE AND FIGURES

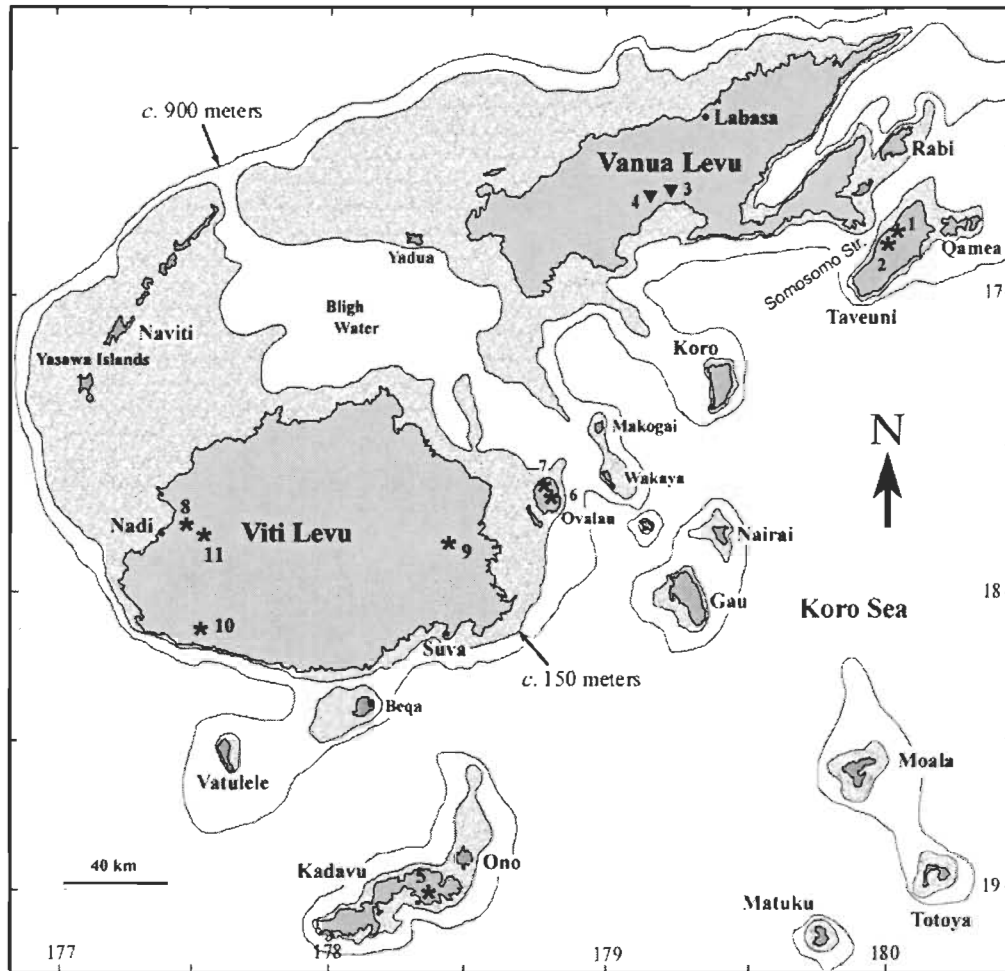
**Table 6.1. Geographic and physical data for localities of *Simulium* (*Hebridosimulium*) *laciniatum* used in this study**

No. of specimens	Island	(Location No.) and local name	Latitude / Longitude	Alt. (mabs)	Habitat/ substrate <sup>A</sup>
2	Taveuni	(1) Waitavala Waterslides	S16.79720° W179.99064°	60	V
2	Taveuni	(2) Upper Wairiki River	S16.81939° W179.99456°	350	RV
4	Vanua Levu	(3) Ravini Stream Cascade	S16.65305° E179.22107°	120	RV
5	Vanua Levu	(4) Nakasa River	S16.67770° E179.19881°	90	RV
2	Kadavu	(5) Wailea Stream cascade	S19.03674° E178.37830°	10	R
2	Ovalau	(6) Waitovu Stream cascade	S17.66545° E178.82551°	60	R
2	Ovalau	(7) Rukuruku Stream	S17.65023° E178.76643°	76	R
2	Viti Levu	(8) Vereni Creek cascade, Abaca	S17.66866° E177.54200°	510	RV
2	Viti Levu	(9) Uru's Waterfall	S17.78154° E178.43657°	120	RV
2	Viti Levu	(10) Savunamatelaya Cascade	S18.17973° E177.73834°	100	RV
2	Viti Levu	(11) Savuione Falls, Abaca	S17.67429° E177.55430°	600	R

<sup>A</sup> R = rock, V = vegetation. Predominant habitat listed first.

**Table 6.2. Distribution of ITS1 alleles on Fijian islands**

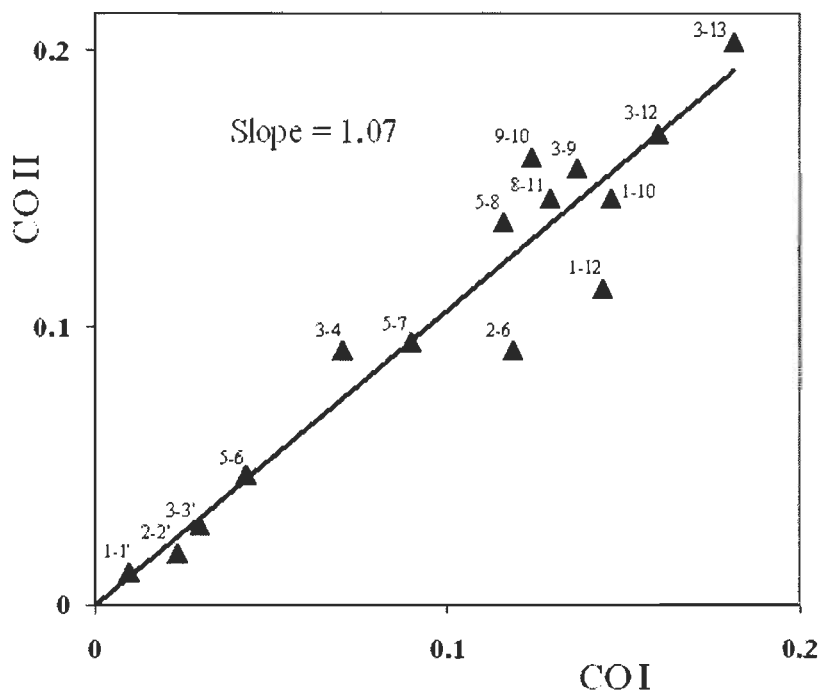
	Island	(AG) <sub>5</sub> A <sub>7</sub>	(AG) <sub>5</sub> A <sub>8</sub>	(AG) <sub>6</sub> A <sub>8</sub>	(AG) <sub>7</sub> A <sub>8</sub>	(AG) <sub>8</sub> A <sub>8</sub>	Special
Southern group	Viti Levu	-	-	9	7	-	-
	Kadavu	-	-	3	1	-	-
	Ovalau	-	-	2	5	1	-
Northern group	Vanua	1	-	-	14	3	-
	Levu						
	Taveuni	-	1	-	5	1	1
None detected, -.							



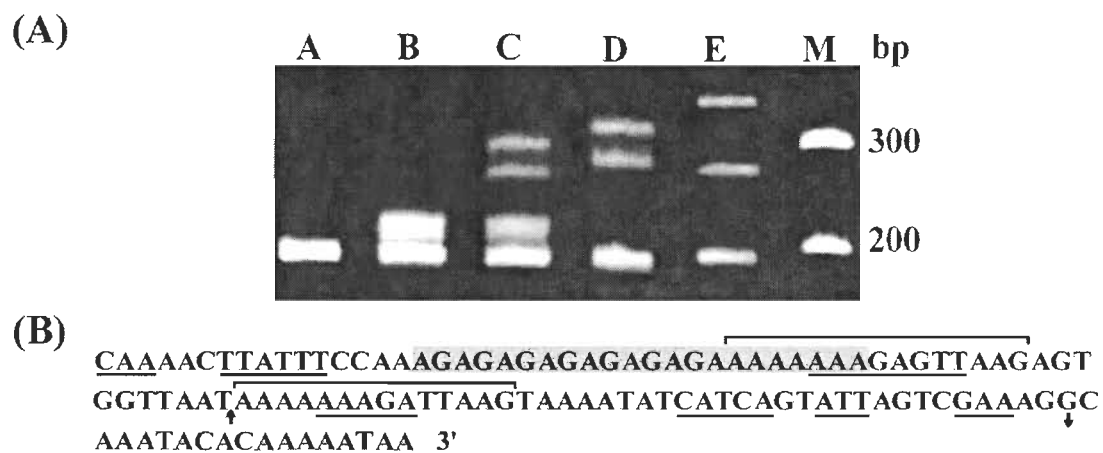
**Fig 6.1.** Map of Fiji. Smaller islands omitted. Full description of collection sites (indicated here by numbers) given in Table 1. Two specimens were sampled for DNA analysis at each location, except when triangles indicate more. The ~150 m bathymetric contour indicates extent of land at maximum sea-level depression during ice ages over the last 500,000 years. Based on 1 : 1 000 000 scale, geological map of Fiji, first edition, 1965, Geological Survey, Suva.



**Fig. 6.2.** Partial sequence of the COII polypeptide from *Simulium* (*Hebridosimulium*) *laciniatum*, with position numbering based on the *Cnephia dacotensis* polypeptide (inferred through DNA sequence, Genbank AF425841). 20 specimens out of 27 have exactly the same DNA sequence coding for the polypeptide segment shown on top. Point mutations are indicated by a shaded box when an amino acid change occurs and by an open box if synonymous.



**Fig. 6.3.** COII versus COI gene divergence (K2P distance) for various North American Simuliidae. Following is a list of pairs of species or conspecific geographical variants that are compared. 1 and 1': variants of *Simulium euryadminiculum* Davies, 1949. 2 and 2': variants from *S. vittatum*. 3 and 3': variants of a *Prosimulium* Roubaud, 1906 species. 4: another *Prosimulium* sp. 5: *Simulium longistylatum*. 6: *Simulium pictipes* Hagen, 1880; recently proposed by Adler *et al.* (2004) to be moved to *Simulium innoxium* Comstock and Comstock, 1895. 7: *Simulium decorum* Walker, 1848. 8: *Simulium jenningsi* Malloch, 1914. 9: *Stegopterna* (formerly *Cnephia*) *mutata* Malloch, 1914. 10: *Cnephia dacotensis* Dyar & Shannon, 1927. 11: *Simulium venustum* Say, 1823. 12: member of the *Simulium aureum* Rubtsov, 1956 species group. 13: *Twinnia tibblesi* Stone and Jamnback 1955. Full COI (663 bp) and COII (252 bp) sequences to be published elsewhere.



**Fig. 6.4.** Characterization of ITS1. (a) From A to E: the five types of acrylamide-urea gel patterns for PCR products. M, size marker. (b) Sense strand sequence of the 110-117 bp spacer, with the microsatellite region (variant (AG)<sub>7</sub>A<sub>8</sub> shown) shaded. Bordering 18S and 5.8S rDNA ends are omitted. Simuliidae consensus blocks are underlined and brackets show the two copies of a 14-15 bp internal repeat. A single 'special' sequence differs from the above in having a (AG)<sub>5</sub>A<sub>9</sub> microsatellite, a G to A change (position shown by down arrow) and an AC insertion (up arrow).



## **CHAPITRE VII**

### **Résultats et discussion**

## 7.1 Résultats et discussion

### 7.1.1 Un critère d'interfécondité difficilement vérifiable

Chez les mouches noires (simulies), l'attribution à l'espèce peut s'effectuer à partir de critères morphologiques, cytogénétiques et même moléculaires. Ceci ne confère pas à une espèce donnée une identité fixe pour autant, car les phénomènes de spéciation semblent fréquents chez les mouches noires. De plus, il est pratiquement impossible de vérifier chez ce groupe la capacité de reproduction. L'élevage complet des Simuliidae au laboratoire n'a pas souvent été réalisé jusqu'à présent. Même s'il est assez facile d'en faire l'élevage des œufs aux adultes, en partant d'une ponte recueillie d'un cours d'eau (Gaudreau, observation personnelle), peu d'accouplements en captivité sont observés. En fait, il existe quelques cas dont ceux de *Cnephia lascivum* signalé par Twinn (1936) (cité dans Grenier, 1953), de *S. decorum* où l'on a atteint dix générations avec accouplements (Simmons et Edman, 1981) et de *S. vittatum*, cytotype IS-7, maintenu en élevage continu (Gray, 2010). L'accouplement en captivité étant très difficile, il ne nous est donc pas possible de vérifier si des individus de groupes phylogéographiques distincts, comme les populations de *S. vittatum* du Québec, de Terre-Neuve et d'Islande sont interféconds. De plus, Davies a suggéré que la parthénogénèse puisse être le mode courant de reproduction chez *C. mutatum*<sup>12</sup>, une espèce triploïde avec une proportion de femelles extrêmement élevée (cité dans Grenier, 1953).

### 7.1.2 Problématique de la définition moléculaire de l'espèce

On peut considérer que la taxinomie moléculaire a gagné ses lettres de noblesse avec la proposition d'Hebert *et al.* (2003) d'utiliser des codes barres à ADN. Ces auteurs ont avancé que des marqueurs comme COI étaient suffisamment variables entre espèces pour servir de critère très fiable pour l'identification d'espèces animales et végétales. Ils ont entre autres démontré que la séquence de COI permettait d'identifier correctement et à l'aveugle 200 espèces apparentées de lépidoptères. La notion de code barres ne signifie pas que tous les individus de la même espèce possèdent une séquence unique, mais que

---

<sup>12</sup> Synonyme de *S. mutata*.

ces séquences se rangent à l'intérieur d'un groupe distinct de tous les autres. En effet, on s'attend à ce que l'interruption du flux génique lors d'une spéciation produise des discontinuités génétiques importantes qu'on pourra alors utiliser comme définition opérationnelle de l'espèce.

La question est alors de savoir quel est le marqueur le plus approprié à cet objectif et quel seuil de divergence tracera la frontière entre deux espèces. Depuis Hebert *et al.* (2003), qui estimaient le niveau maximum de variation intraspécifique à environ 3% pour COI, plusieurs travaux dont ceux de Carew *et al.* (2003a) et Day *et al.* (2008) estiment qu'une divergence de 5% ou plus marque de façon à peu près certaine la présence de deux espèces. En effet, le taux de divergence de COI est estimé à 2.3% par million d'années (Brower, 1994) ce qui fait que le seuil de 5% sera facilement atteint à l'intérieur de la durée moyenne de vie d'une espèce. Il peut sembler incongru d'utiliser un critère quantitatif pour délimiter les espèces, mais on doit rappeler que la taxinomie basée sur la morphologie utilise la même démarche de façon qualitative. Même si ce seuil de 5% semble fermement établi, une valeur inférieure pourrait cacher une spéciation en cours ou très récente. On pourrait citer par exemple les invasions à *Wolbachia* qui génèrent une spéciation quasi instantanée via la barrière reproductrice qu'elles dressent (Shoemaker *et al.*, 2004). Cela pourrait aussi survenir à la suite d'un effet fondateur suivi d'une diversification rapide, comme on le voit fréquemment pour les espèces endémiques dans les îles du Pacifique (Joy *et al.*, 2007). Dans le cas d'une spéciation cryptique non détectée par COI, un autre marqueur, comme ITS1 qui suit un mode d'évolution saltatoire, pourrait apporter une réponse distincte (Krüger *et al.*, 2000)

Vu que l'horloge moléculaire varie d'un gène à l'autre, le choix correct du marqueur peut s'avérer crucial par rapport à la profondeur phylogénétique qu'on cherche à sonder. Alors que la période couverte par COI et COII ne dépasse pas 10 millions d'années environ, ITS2 (chapitre 3) permet pratiquement de remonter à l'origine des mouches noires, vieille d'environ 130 millions d'années (Bertone *et al.*, 2008). Enfin, d'autres marqueurs nucléaires tels l'ADNr 18S, 28S, et les séquences codantes CAD, TPI, PGD (Bertone *et al.*, 2008) sont capables de retracer l'origine des familles et même

celle des ordres. Vu leur conservatisme, ces marqueurs seront par contre mal adaptés à la phylogénie des groupes récemment divergés (chapitre 6).

Dans le cadre de la problématique précédente, nous suivrons une progression similaire dans l'exposé et la discussion des principaux résultats obtenus. Nous partirons de l'analyse de populations d'une espèce ou en voie d'une possible spéciation en deux espèces pour ensuite examiner les cas litigieux d'espèces proches ou difficiles à distinguer. Nous tenterons finalement de poser les bases d'une phylogénie globale des espèces de simules du Québec.

### **7.1.3 Étude moléculaire de populations**

#### **7.1.3.1 Cas de *S. laciniatum***

*S. laciniatum* est le cas flagrant d'espèce endémique apparue à la suite d'un effet fondateur survenu dans le cadre d'un isolement géographique important. Vu que c'est la seule simule des Fidji et qu'elle est présente sur toutes les îles de l'archipel, on peut légitimement se poser la question de son statut monospécifique. On peut répondre positivement, car la morphologie est homogène et la caractérisation de COII révèle un haplotype unique largement prédominant qui n'est accompagné que de rares variants individuels distribués au hasard dans l'archipel. Une modélisation évolutive a permis de calculer pour l'origine de l'espèce un âge probable légèrement inférieur à 100 000 ans.

Cependant ITS1 a permis d'établir une différenciation génétique partielle entre populations des diverses îles (tableau 6.2), événement possiblement relié à une augmentation de l'isolement géographique dû à la remontée des niveaux marins lors de la fin de la dernière ère glaciaire. Ces différences entre les deux marqueurs peuvent s'expliquer par leurs modes d'évolution distincts. COII évolue par mutation ponctuelle suivant un rythme d'horloge moléculaire, tandis que la famille multigénique d'ADNr dont fait partie ITS1 suit un mode saltatoire basé sur une évolution de groupe. Un variant en copie unique apparu dans un génome pourrait se répandre ensuite à l'intérieur de celui-ci, puis dans la population via la reproduction sexuée. Ce cas n'est pas unique et

l'exemple des simules du Québec (chapitre 3) montre qu'ITS1 pourrait se comporter fréquemment de cette façon.

### 7.1.3.2 Cas de *S. vittatum*

*S. vittatum* comporte trois cytotypes en Amérique du Nord; IS-7 (*S. vittatum*), IIL-1 (*S. tribulatum*) et XoYo qu'on retrouve seul dans les provinces maritimes (Duncan *et al.*, 2004; Minhas *et al.*, 2005; Pasternak, 1964; Rothfels et Featherston, 1981). En plus de ces cytotypes, on répertorie aussi deux morphotypes, soient les formes pâles et foncées (figure 1.3). Les deux sont très différents l'un de l'autre, laissant même croire à un novice qu'il pourrait s'agir de deux espèces distinctes. De plus, seul l'article d'Adler et Kim (1984) relate et illustre ces morphotypes, même la plus récente clé d'Adler *et al.* (2004) ne les mentionne pas. D'ailleurs, nos données moléculaires ne permettent pas de différencier les deux morphotypes, lesquels partagent souvent sur le même site, des haplotypes COI et COII voisins, sinon identiques.

En comparant la cytogénétique aux données moléculaires, Tang *et al.* (1996) remarquent qu'il y a des contradictions : « This result was in contrast to morphological and ecological studies which suggested that *S. vittatum* s.l. consists of two or more sibling species (Adler et Kim, 1984; Rothfels et Featherston, 1981) ». Ses analyses, basées sur la technique des hétéroduplexes appliquée sur le 12S et 16S n'ont révélé que des variations minimales entre deux espèces soeurs de *S. vittatum* (IIL-1 et IS-7) récoltées en quelques endroits (Colorado, Islande, Montana, Nebraska, Terre Neuve, New Hampshire et Texas). De même Zhu *et al.* (1998) mentionnent : « Allozymes often fail to reveal differences in population structure in black flies, and found no differences in allozyme allele frequency between the two siblings of *S. vittatum*. Both cytospecies are morphologically variable but characters separating the siblings have not been found ». Enfin, Pruess *et al.* (2000) n'ont trouvé que très peu de différences au sein de la séquence nucléotidique de COII, soient trois substitutions silencieuses entre les larves provenant de l'Islande et du Nebraska.

Nos données moléculaires appuient indirectement celles de ces auteurs. Même si nous n'avons pas fait la cytogénétique, la distribution connue des cytotypes (Adler *et al.*, 2004) permet d'affirmer que nous sommes en présence d'au moins deux cytotypes : XoYo pour les spécimens de Terre-Neuve et IIIIL-1 ou IS-7 pour ceux du Québec (Mauricie, Montérégie) et d'Ontario (Ottawa). Un haplotype constitué de cinq positions spécifiques sur COI et COII différencie en partie les populations de *S. vittatum* selon leur provenance, en l'occurrence Québec-Ontario ou Terre-Neuve. Cette signature géographique s'ajoute à une quarantaine d'autres sites polymorphes globalement répertoriés, de sorte que la divergence entre séquences ne peut mener seule à une distinction nettement tranchée entre les deux populations (Gaudreau *et al.*, 2010a; chapitre 5).

Que conclure de l'opposition apparente entre la cytogénétique et l'ensemble des données moléculaires? Comme on le voit chez *S. damnosum*, rien ne garantit de prime abord que la présence de deux cytotypes signale forcément l'existence d'espèces distinctes (Morales-Hojas et Krueger, 2009). La différence entre espèces sœurs de similies se réduit en général à une inversion chromosomique unique, ce qui entraîne pour l'éventuel hybride un léger désavantage sélectif lié aux produits aberrants de méiose résultant d'une recombinaison à l'intérieur de la région inversée. Parmi tous les facteurs sélectifs en action, il s'agit possiblement d'un élément mineur, et ce d'autant plus que l'inversion est de petite taille. Chez le genre *Prosimulium*, les cytoespèces *P. mixtum* et *P. fuscum* sont par exemple suffisamment apparentées pour s'hybrider en nature (Rothfels et Freeman, 1977) malgré des distinctions morphologiques dues à des facteurs écologiques (Snyder et Linton, 1983). Pour les cytotypes IIIIL-1 et IS-7 de *S. vittatum*, qui coexistent sur un vaste territoire en Amérique du Nord (Adler *et al.*, 2004), on ne dispose pas de données confirmant ou infirmant une possibilité d'hybridation. L'absence de différenciation moléculaire du complexe *S. vittatum*, notée pour des régions aussi éloignées que l'Islande et le Nebraska, suggérerait alors un flux génique substantiel entre "espèces", bien qu'on observe un certain décalage temporel dans le développement larvaire de IIIIL-1 et IS-7 (Adler et Kim, 1984). De même, le lien entre le caryotype XoYo et l'haplotype COI/COII relativement spécifique de Terre-Neuve pourrait n'être que fortuit si on admet un effet fondateur issu d'un haplotype

voisin de certains retrouvés à l'état minoritaire au Québec et en Ontario (chapitre 5). Enfin, les taux moyen et maximum de divergence que nous observons pour COI et COII demeurent dans la fourchette intraspécifique normale chez les invertébrés (Hebert *et al.*, 2003).

### 7.1.3.3 Cas de *S. euryadminiculum* / *S. annulus*

Le cas de *S. euryadminiculum* / *S. annulus* est quelque peu différent de celui de *S. vittatum*, car on n'y retrouve pas de morphotypes ou de cytotypes distincts. Dans ce cas-ci, l'attribution du "nom" de l'espèce est ambiguë et dépend de la clé morphologique utilisée au laboratoire. On serait en présence de *S. euryadminiculum* selon la clé de Wood *et al.* (1963) et de *S. annulus* selon Adler *et al.* (2004). Par ailleurs, les deux clés morphologiques ne correspondant pas en tout point et il est donc difficile d'établir une équivalence entre elles.

L'appellation de cette espèce a varié au cours du temps et selon l'auteur. Dans « A Catalog of Diptera of America North of Mexico » (Stone *et al.*, 1965) et dans la clé de Wood *et al.* (1963), on faisait seulement état de *S. euryadminiculum*, de même que la collection nationale des insectes du Canada<sup>13</sup> en 1991. Ce n'est qu'au cours des dernières années que *S. annulus* fit son apparition. Bien que la pratique la plus courante aujourd'hui soit d'utiliser *S. annulus* pour désigner ces spécimens, il reste qu'à partir des données d'inventaires, on n'a pas encore osé trancher nettement la question. *S. annulus* ne figurait pas encore dans le BDWD<sup>14</sup> en 2009, mais portait plutôt le nom de *S. euryadminiculum*, alors que l'inventaire d'Adler et Crosskey (2009) faisait état du contraire. Les deux inventaires publiés en 2009 n'étaient donc pas en accord et démontrent encore une fois qu'aucune règle ne régit l'appellation des espèces de mouches noires. Dans les différents articles qui constituent cette thèse, et à la demande des reviseurs pour les articles publiés, nous avons choisi pour des raisons de clarté

<sup>13</sup> <http://www.nadsdiptera.org/Catalogs/CNCtypes/Part1.pdf>

<sup>14</sup> La BDWD (BioSystematic Database of World Diptera) est une base de données où l'on retrouve des informations au sujet des diptères.

d'utiliser toujours le nom *S. annulus*, pour désigner ces deux espèces qui en réalité n'en constitue qu'une seule.

Lorsqu'on regarde les données moléculaires, les deux espèces correspondent en fait à une seule, bien que les arbres phylogénétiques (COI, COII) permettent de distinguer les deux populations échantillonnées. Par contre, dans ce cas-ci, le pourcentage de divergence entre celles-ci se situe à environ 1.4%, ce qui cadre à l'intérieur du niveau normal de variations intraspécifiques mentionnées par Hebert *et al.* (2003). D'ailleurs, l'identité des séquences d'ITS1, ITS2 et 18S entre les deux groupes nous permet de conclure au statut monospécifique de la population. Nous sommes donc en accord avec ce que proposent Adler *et al.* (2004), les deux noms étant synonymes pour la même espèce.

#### 7.1.3.4 Cas de *S. longistylatum* / *S. pictipes*

*S. longistylatum* et *S. pictipes* sont de grosses larves très foncées et habitent principalement les chutes. Tout comme les deux cas précédents, ces espèces ont vu leur nom changer au cours des dernières années. Au début des recherches, soit avant 2004, nous n'avions accès qu'à deux clés d'identification, celles de Wood *et al.* (1963) et de Currie (1986). Nous avons toutefois utilisé principalement celle de Wood *et al.* (1963), élaborée originellement pour l'Ontario et qui, par extension, devrait aussi s'appliquer au Québec. Sur cette base, nous avons donc séparé les larves des deux espèces morphologiquement très similaires, mais néanmoins distinguables grâce à quelques critères. Ceux-ci n'étaient pas tous directement cités dans la clé d'identification écrite de Wood *et al.* (1963), mais repérables sur les représentations imagées annexées à celle-ci. Seul le premier critère faisant référence aux antennes provient directement de la clé d'identification écrite :

→*S. pictipes*: antennes pâles. *S. longistylatum*: antennes foncées comportant trois taches pâles (Wood *et al.*, 1963).

→Échancrure plus ronde chez *S. pictipes*.

→La région des ébauches oculaires plus pâle chez *S. pictipes*.



Avec l'arrivée de la clé d'Adler *et al.* (2004), nous avons voulu comparer nos identifications précédentes à celles obtenues en utilisant cette nouvelle clé, mais *S. longistylatum* y était absente. Adler *et al.* (2004) l'avaient renommée *S. pictipes*. Selon eux, ces deux espèces appartiendraient en fait au groupe *S. pictipes* qui comprend les espèces *S. claricentrum*, *S. innoxium* et *S. pictipes* (*S. longistylatum* Shewell 1959A, nouveau synonyme). Adler *et al.* (2004) affirment donc que *S. longistylatum* est en fait *S. pictipes*, relégué au rang de *S. innoxium*.

En fouillant dans la littérature, nous avons pu constater que le nom de ces espèces n'avait pas été statique. *S. pictipes* et *S. longistylatum* figurent dans « A Catalog of the Diptera of America North of Mexico » (Stone *et al.*, 1965) comme deux espèces différentes. *S. innoxium* et *S. aldrichiana* seraient des sous-espèces de *S. pictipes*. Cependant, en 1975, après avoir examiné les chromosomes des glandes salivaires chez des individus du groupe *S. pictipes* par des méthodes de coloration classique et de fluorescence à la quinacrine, Bedo (1975) reconnaissait cytologiquement trois espèces dont l'une était inconnue jusque-là. Des trois espèces jumelles, *S. longistylatum* est la seule à pouvoir être identifiée avec certitude selon la morphologie. Il désignait alors les deux autres *S. pictipes* A et B. Finalement, Crosskey et Howard (2004) dans « A revised taxonomic and geographical inventory of world blackflies » laissent *S. longistylatum* comme espèce à part entière et non comme synonyme à *S. pictipes*. La BDWD<sup>14</sup> reconnaît aussi en 2009 *S. longistylatum* comme une espèce à part entière, mais *S. pictipes* est renommé *S. innoxium*. Encore une fois, les correspondances entre les espèces des anciennes clés, de la clé d'Adler *et al.* (2004) et des inventaires sont assez contradictoires.

Par contre, la distribution des espèces se ressemble entre Bedo (1975) et Adler *et al.* (2004), le premier affirmant : « *S. pictipes* is distributed from southern United States to southern Ontario and Quebec, while *S. longistylatum* occurs to the north with a zone of overlap in southern Ontario ». Selon Adler *et al.* (2004), la distribution de *S. innoxium* (anciennement *S. pictipes*) se situe autour des grands lacs et, plus au sud, de l'Alabama au Missouri. L'espèce est commune le long des Appalaches, mais aucun point de

collection (figure 1.5) n'est mentionné pour le Québec (Adler *et al.*, 2004). Par contre, *S. pictipes* (anciennement *S. longistylatum*) se retrouve à quelques endroits (figure 1.5) au Canada, mais non aux USA (Adler *et al.*, 2004).

Relativement au statut de ces quasis jumeaux morphologiques séparés par les trois critères mentionnés ci-dessus, nos analyses moléculaires montrent bel et bien la présence de deux espèces distinctes, quoique très voisines. Les arbres phylogénétiques issus de COI, COII, ITS1, ITS2 et 18S les différencient bien. De plus, le pourcentage de divergence pour COI se situe à environ 4.9%, au niveau du seuil critique pour reconnaître deux espèces. Cependant, nous avons décidé de conserver pour l'instant l'ancienne nomenclature pour diverses raisons, la première étant que les critères morphologiques furent établis à partir de celle de Wood *et al.* (1963), puis parce que nous ne pouvons pour l'instant établir de correspondances avec la clé incomplète d'Adler *et al.* (2004). Évidemment, il faudrait éventuellement communiquer avec Adler, puis aller examiner les larves conservées dans la collection du Royal Museum of Toronto pour fin de comparaison.

#### 7.1.4.5 Cas des *Prosimulium* et *Helodon*

Même si nous pouvons d'emblée distinguer *Prosimulium* de tous les autres genres grâce aux plumeaux céphaliques secondaires (figure 1.1), l'attribution subséquente à l'espèce peut s'avérer difficile, voire impossible. Comme discuté ci-dessus, certains critères ne peuvent pas être appliqués pour cette étape finale (Alvan-Aguilar *et al.*, 2005; Lucas et Hunter, 1999; Zhang, 2005). Un critère unique, tel le nombre de rayons céphaliques, l'échancrure et les dents, ne doit surtout pas être utilisé. Le nombre de rayons céphaliques, qui varie trop selon divers paramètres environnementaux (section 1.2.3.2), n'a pas été retenu. Par ailleurs, Snyder et Linton (1983) ont aussi soulevé le fait que ce critère ne peut être utilisé seul dû aux chevauchements retrouvés entre espèces. De même, il est très difficile d'effectuer l'identification à l'espèce à partir de l'échancrure, car son aspect plus ou moins arrondi et assez uniforme est soumis à la subjectivité du taxinomiste. Finalement, restent les dents qui s'avèrent taxonomiquement intéressantes, mais un critère tout aussi discutable que les autres. Encore une fois, deux

clés existent (Adler *et al.*, 2004; Wood *et al.*, 1963). Cependant, celle d'Adler *et al.* (2004) fut privilégiée, car elle est visuellement beaucoup plus complète. Même en procédant en double aveugle, et ce à deux reprises, il ne fut pas facile d'en venir à une identification basée sur les dents. Par ailleurs, il faut noter que 7 à 8% des spécimens examinés (total de 800) de *Prosimulium* et du genre voisin *Helodon* ont dû être écartés de l'analyse moléculaire, puisqu'ils exhibaient un patron inhabituel, avec dents sous ou surnuméraires et de forme aberrante, rendant impossible leur classification selon les critères d'Adler *et al.* (2004). Bien que ce critère semble discriminant aux yeux de Currie (1986) et d'Adler *et al.* (2004), notre classification<sup>15</sup> sur la base de critères morphologiques comprenant six espèces (*P. fuscum*, *P. mixtum*, *P. fontanum*, *P. multidentatum*, *H. vernalis* et *H. decemarticulatus*) apparaît non congruente avec les données moléculaires (figure 4.1).

En fait, deux marqueurs (COI et COII) pour chacun des 23 individus représentant nos six espèces ont été amplifiés et séquencés. Les deux phylogénies obtenues (figure 4.1 et autres données non illustrées) sont parfaitement congruentes, même au niveau du branchement de chaque spécimen, et indiquent deux regroupements majeurs identiques à ceux observés pour ITS1 (chapitre 3). De plus, nous constatons deux choses. D'abord, des individus des genres *Helodon* et *Prosimulium* se trouvent entremêlés sur une même branche, ce qui est en contradiction avec l'analyse morphologique. Ensuite, aucune espèce morphologique ne forme de regroupement moléculaire unique. On peut donc affirmer qu'un critère unique, tel les dents, ne peut être utilisé pour identifier un individu à l'espèce.

#### 7.1.4 Utilité des divers marqueurs pour établir une phylogénie des simulies

Nous examinerons ci-dessous les divers marqueurs moléculaires utilisés quant aux niveaux (ordre, famille, genre, espèce) de l'arbre phylogénétique qu'ils explorent chez les simulies. Nous verrons si ces marqueurs permettent de retrouver les branchements de la phylogénie cladistique classique, de distinguer des espèces voisines, ou des sous-

---

<sup>15</sup> Nous avons analysé, pour chaque espèce, au minimum trois spécimens.

populations d'une espèce. Nous parlerons par la suite des applications possibles de ces outils moléculaires à l'étude des similies.

#### 7.1.4.1 L'unité de transcription ADNr

L'opéron de l'ADNr (figure 7.1) est présent chez les diptères en plusieurs centaines de copies, presque toutes regroupées sous forme d'unités de transcription répétées en tandem (Hwang et Kim, 1999; Miller *et al.*, 1997; Schlötterer *et al.*, 1994), ce qui en facilite la détection par PCR (Caterino *et al.*, 2000; Young et Coleman, 2004). Chaque unité comprend quatre régions conservées après maturation du pré-ARN (18S, 5.8S, 2S et 28S) et trois régions appelées « Internal Transcribed Spacer » (2a, ITS1 et ITS2) (Miller *et al.*, 1997; Thanwisai *et al.*, 2006) et destinées à l'élimination lors de la maturation en ARNr. L'unité de transcription permet l'étude de différents niveaux d'évolution, des réseaux basaux de la vie aux relations entre espèces voisines et aux études de populations, puisqu'elle comporte à la fois des éléments variables et d'autres conservés (Caterino *et al.*, 2000). La variabilité la plus forte provient des espaceurs intergéniques qui évoluent plus vite. Les substitutions et autres mutations surviennent fréquemment dans ces régions qui subissent peu de pression de sélection, car elles sont destinées à l'élimination (Hwang et Kim, 1999; Schlötterer *et al.*, 1994). Les régions conservées sont fonctionnelles et spécifient des composantes essentielles de la machinerie de traduction.

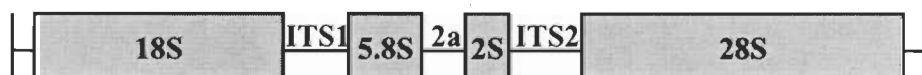
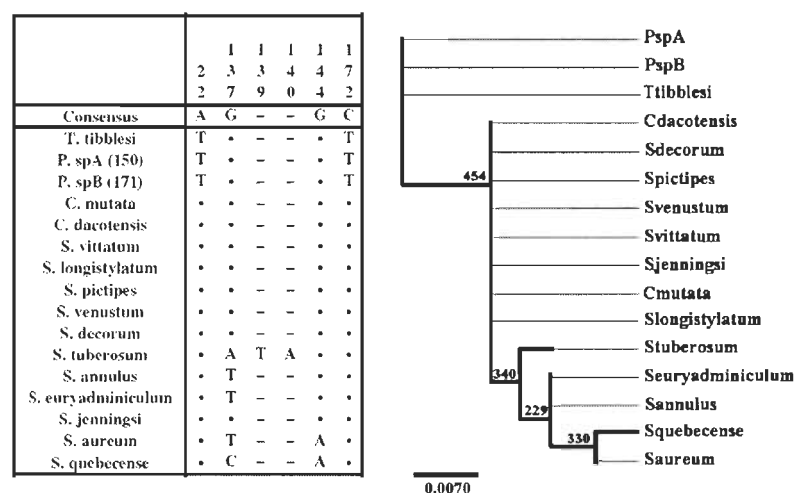


FIGURE 7.1 : Opéron de l'ADNr.

#### 7.1.4.2 Espaceur 2a et ADNr 5.8S et 2S

La séquence concaténée de l'ADNr 5.8S, l'espaceur 2a et l'ADNr 2S est de 176 bases chez les mouches noires. Chez les espèces étudiées ici, on observe seulement six positions mutationnelles au total, se réduisant à quatre lorsqu'on exclut l'insertion d'une séquence TA au niveau de l'espaceur 2a chez *S. tuberosum* (figure 7.2). Bien que l'arbre

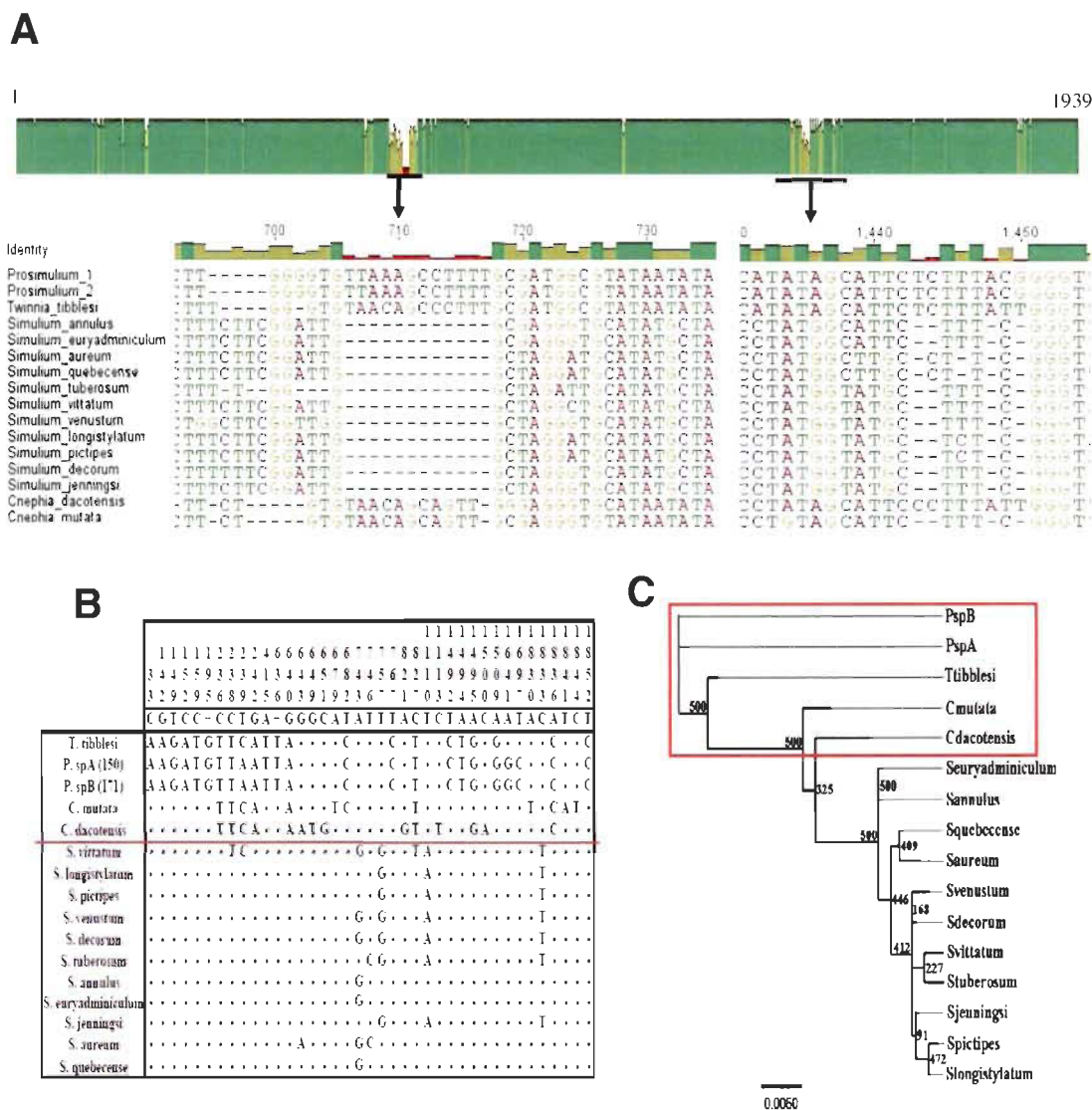
issu de ce marqueur soit capable de donner des groupements ciblant les branches profondes (divergences anciennes) d'un arbre phylogénétique (Miller *et al.*, 1997), comme la division entre Prosimuliini (*Prosimulium* et *Twinnia*) et Simuliini (*Cnephia*, *Stegopterna* (*C. mutata*) et *Simulium*), la taille limitée de ce bout de séquence ne peut être considérée comme fiable pour l'obtention d'une phylogénie résolue au niveau des espèces (figure 7.2). Il est trop court pour contenir assez d'informations phylogénétiques (Miller *et al.*, 1997; Hwang et Kim, 1999).



**FIGURE 7.2 :** Mutations et phylogénie de la région englobant l'espaceur 2a et les ADNr 5.8S et 2S. À gauche, mutations listées par position (depuis le début du 5.8S) et par espèce. À droite, l'échelle de l'arbre phylogénétique correspondant est donnée en unités de divergence relative. Chiffres indiquant le bootstrap sur 500 répliques.

#### 7.1.4.3 ADNr 18S

Le 18S (figure 7.1) semble plus informatif concernant des espèces moins phylogénétiquement reliées, car plus conservé, suite aux contraintes de structures secondaire et tertiaire, et beaucoup plus long (environ 1800 à 2000 bases) (Miller *et al.*, 1997). Sa longueur varie quelque peu, allant de 1973 bases, pour *Prosimulium* et *Twinnia*, à 1955 bases pour *S. tuberosum* (Gaudreau *et al.*, 2010b). La majorité des espèces du genre *Simulium* se situent toutes très près de 1961 nucléotides. L'ensemble du gène semble bien conservé sauf pour deux régions hypervariables (figure. 7.3a; figure 4.3). La première comprend, par rapport à *Simulium*, une insertion de douze bases chez



**FIGURE 7.3 :** Mutations et phylogénie de l'ADNr 18S. A) Haut : représentation schématique des positions variables le long de la séquence (vert au rouge : de 100 à 0% d'identité); Bas : zoom sur les deux régions hypervariables B) Identification des positions signatures en dehors des régions hypervariables C) Arbre phylogénétique du 18S; les espèces ayant l'insertion de 12 bases sont encadrées (bootstrap sur 500 répliques).

*Prosimulim*, *Twinnia*, *Cnephia* et *Stegopterna*. L'arbre phylogénétique montre bel et bien cette distinction. Les *Simulium* apparaissant comme un clade distinct (figure 7.3c). Cette distinction est aussi renforcée par la présence dans les régions hypervariables (figure 4.3) et ailleurs (figure 7.3b) d'un certain nombre de « positions signatures » qui typifient en bloc les Prosimuliini, l'ensemble Prosimuliini-*Cnephia* ou le genre *Simulium*.

#### 7.1.4.4 ITS

L'espaceurs intergéniques ITS1 (figure 7.1) a été employé pour résoudre des problèmes phylogénétiques aux niveaux inférieurs du genre, de l'espèce et des populations. Il contient les régions les plus variables de séquence et de longueur, même entre espèces étroitement apparentées (Morales-Hojas *et al.*, 2002). Ici, sa taille fluctue entre 95-120 pb (la plupart des Simuliini) et 492 pb (Prosimuliini) (LaRue *et al.*, 2009). De plus, huit courts blocs consensus, 39 pb au total, définissent une structure minimale commune.

À la lumière de plusieurs travaux, certaines questions demeurent en suspens quant à l'utilité de ce marqueur. Nous savons qu'il s'agit d'une famille multigénique dont les membres peuvent varier d'une copie à l'autre chez un même individu, et à plus forte raison, entre individus de la même espèce. Ce point devient problématique par rapport à la définition d'un marqueur d'identification idéal. En effet, un bon marqueur devrait soit refléter des différences absolues et fixes entre espèces, soit au pire délimiter une frontière quantitative nette entre les variations intra et interspécifiques. Même si nos travaux ont révélé une variation intraspécifique chez la plupart des espèces, celle-ci demeure inférieure d'au moins un ordre de grandeur comparativement aux différences observées entre espèces même très voisines. Cette caractéristique fait d'ITS1 un marqueur presque idéal pour l'obtention d'une signature d'espèce et justifie son emploi occasionnel en génétique des populations (Gaudreau *et al.*, 2008; Larue *et al.*, 2009; Rodríguez-Pérez *et al.*, 2006) et pour différencier entre membres d'un complexe (Brockhouse *et al.*, 1993; Krüger *et al.*, 2000). En contrepartie, on pourrait cependant conclure à son incapacité de résoudre les niveaux phylogénétiques les plus élevés et à l'échec assuré des tentatives en ce sens. Juste à l'alignement des séquences, on remarque

en effet que seuls les segments du début et à la fin sont conservés, le milieu ne constituant souvent que du simple « remplissage ». Malgré tout, et à notre grande surprise, l'arbre issu d'ITS1 (figure 7.4) a généré quelques groupements véridiques confirmés par la morphologie et par les phylogénies basées sur d'autres marqueurs.

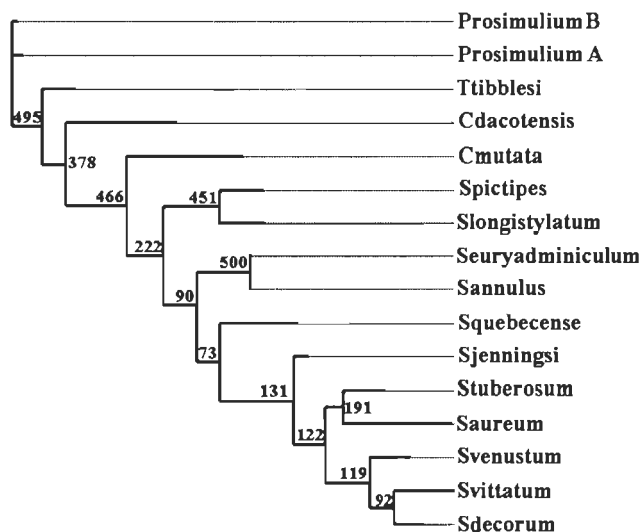


FIGURE 7.4 : Arbre phylogénétique d'ITS1 (bootstrap sur 500 répliques)

ITS2 (figure 7.1) a une longueur peu variable d'environ 270 pb et contient cinq longs blocs fortement conservés. Nos résultats concordent avec ceux d'autres études dont celle de Thanwisai *et al.* (2006) où la taille de l'ITS2 d'espèces de mouches noires provenant de Thaïlande se situe entre 247 et 308 pb. ITS2 est beaucoup plus conservé qu'ITS1, ce qui permet, grâce aux cinq blocs conservés, d'établir la structure secondaire (Chapitre 3). Cette structure secondaire facilite l'alignement des séquences en plus de garantir l'obtention d'une phylogénie globale des simulies qui soit plus juste, alors qu'un alignement basé sur la séquence seule nous aurait fait perdre une partie de l'information phylogénétique (Young et Coleman, 2004). L'arbre obtenu (figure 3.6), confirme la séparation fondamentale entre Prosimuliini et Simuliini en plus de définir un certain nombre de clades. ITS2 pourrait constituer un outil précieux pour révéler des parentés évolutives insoupçonnées, en plus de retracer l'histoire globale des simulies, dont l'origine remonte au moins au crétacé inférieur (Bertone *et al.*, 2008).



#### 7.1.4.5 L'ADN mitochondrial

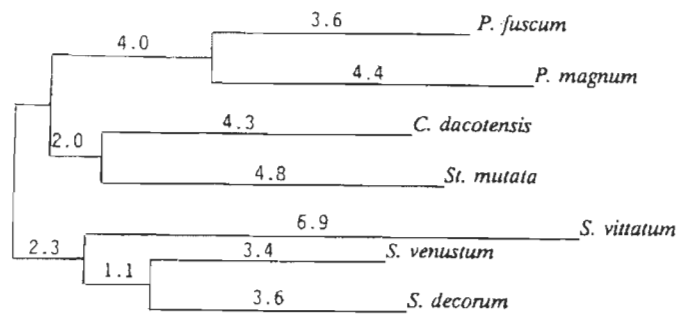
L'étude de l'ADN mitochondrial (ADNmt) jouit encore plus de la faveur des biologistes moléculaires depuis l'avènement du « barcoding of life » associé à COI. Ce dernier est le plus long des trois gènes codant pour des sous-unités de la cytochrome oxydase (Moralais et Severson, 2002) et semble plus conservé (Hwang et Kim, 1999) que COII et COIII. Nous avons aussi étudié COII vu que des séquences complètes étaient déjà disponibles chez certaines Simuliidae, mais aussi chez des Chironomidae, ainsi que chez *Anopheles quadrimaculata* et *Culex quinquefasciatus* (Pruess *et al.*, 2000). Les gènes COI et COII s'alignent sans ambiguïté tout le long de leur séquence (Beckenbach et Borkent, 2003) et sont présents en de nombreux exemplaires par cellule (Dusfour *et al.*, 2004; Le Guyader, 2006; Simon *et al.*, 1994), ce qui facilite le PCR.

Les séquences de COI et COII fourniraient de l'information qui pourrait nous aider à comprendre l'évolution et la phylogénie des simulies. En outre, elles révéleraient assez de détails pour aider dans la résolution des espèces du genre *Simulium* (Rodriguez-Perez *et al.*, 2006) et des complexes d'espèces (Xiong et Kocher, 1991). La séquence en acides aminés est fortement conservée et ne convient qu'à l'étude des niveaux taxinomiques élevés (famille, ordre) (Hebert *et al.*, 2003; Hwang et Kim, 1999). À l'intérieur de groupes plus restreints, on utilisera plutôt la séquence du gène, dont la variation rapide s'explique essentiellement par des mutations synonymes en troisième position de codon. Les séquences comparatives des gènes COI et COII apparaissent comme des outils performants en génétique des populations et permettent de distinguer des espèces apparentées comme le démontrent ici les cas du genre *Prosimulium* et de la paire *S. longistylatum*/*S. pictipes*. Le taux d'évolution très rapide de ces séquences empêche cependant leur utilisation phylogénétique au-delà d'une limite d'environ 15 millions d'années. Une phylogénie combinant COI/ COII, ITS2 et le 18S en un seul corpus de données permettrait cependant, on l'espère, d'afficher et de résoudre en simultanément et de façon adéquate tous les niveaux de l'histoire évolutive des mouches noires. Malgré la fréquence d'utilisation de COI et COII, quelques problèmes potentiels ont été relatés, dont le risque de confusion, vu leur transmission maternelle exclusive, entre polymorphisme intraspécifique et variation interspécifique, ainsi que la création

fréquente de pseudogènes nucléaires issus de la migration vers le noyau de gènes mitochondriaux. Dans ce dernier cas, cette migration conduirait à des conclusions erronées quant à l'identification à l'espèce (Pineau *et al.*, 2005).

#### 7.1.4.6 Confirmation externe et interne de nos phylogénies

Les arbres obtenus à partir d'ITS2 et du 18S ressemblent à ceux déjà cités dans la littérature, même s'ils proviennent d'autres marqueurs et utilisent des paramètres bioinformatiques (modèle d'évolution) distincts. Ceci a pour conséquence d'accroître le niveau de confiance accordé à nos résultats. Si on prend l'arbre de Xiong et Kocher (1991) on peut remarquer que leurs groupements indiquent une tendance identique même si les spécimens proviennent du New Hampshire, ce qui montre le peu d'importance de l'origine géographique des spécimens. Par contre, il faut mentionner qu'ils ont étudié les séquences (347 pb) de l'ARNr mitochondrial 16S. Les inférences phylogénétiques ont confirmé les regroupements tribaux et génériques traditionnels, c'est-à-dire que *P. fuscum* est près de *P. magnum* tandis que *S. decorum*, *S. venustum* et *S. vittatum* forment une deuxième branche et que *C. dacotensis* et *C. mutata*, appartenant à la tribu Cnephiini, se retrouvent sur une troisième (figure 7.5).



**FIGURE 7.5 :** Arbre phylogénétique du mtADNr 16S (adapté de Xiong et Kocher (1991))

Au niveau des confirmations internes nous observons que :

→Les phylogénies profondes obtenues via ITS2 et 18S sont essentiellement congruentes.

→La variabilité intraspécifique d'ITS1 est faible et confirme nos identifications morphologiques puisque tous les individus attribués à la même espèce donnent des séquences très voisines.

→Un arbre de COI construit à partir de l'ensemble des données, soient 75 séquences issues de 16 espèces, regroupe tous les individus d'une même espèce sur un nœud unique (figure 4.1 et autres résultats non illustrés), démontrant l'homogénéité du clade correspondant.

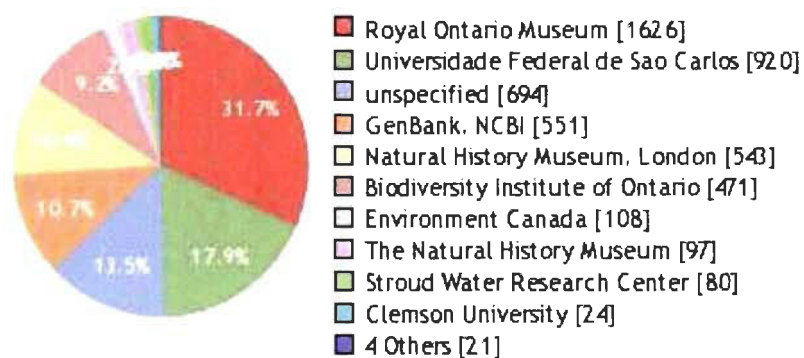
## 7.2 Problèmes liés aux banques de données

Encore aujourd'hui, bien des données morphologiques et cytogénétiques n'ont pas été revues depuis plusieurs années. Quant aux données moléculaires, elles sont plus récentes, mais j'avoue que pour un néophyte qui ne connaît rien en bioinformatique, y obtenir accès demande temps et patience. La majorité des séquences se trouvent dans GenBank, mais beaucoup s'entrecourent avec celles qui sont aussi disponibles sur d'autres sites comme le BOLD<sup>16</sup> (Barcode of life data systems). Cependant, il est impératif de souligner que la majorité des individus séquencés représentant une espèce donnée proviennent en général d'échantillonnages qui sont de type ponctuel tant pour le site que pour le moment. Il n'existe que de rares études comparatives de gènes pour une même espèce, mais provenant d'endroits différents, alors qu'il serait possible de l'effectuer (chapitres 5 et 6).

Cependant, il faut user de prudence lorsque nous voulons effectuer des comparaisons. Certains auteurs comme Castillo (2008) et Dawnay *et al.* (2007), font une mise en garde, car beaucoup de séquences erronées peuvent se retrouver dans les bases de données. À mon avis, les erreurs les plus fréquentes sont surtout liées à l'entrée de données dans la banque. Elles peuvent aussi résulter d'une mauvaise identification morphologique du spécimen au départ (Crosskey et Howard, 2004), problème destiné à

s'aggraver dû à l'extinction prévisible des taxinomistes. Également, les séquences amplifiées par les différents auteurs ne ciblent pas toujours la même région du gène. La solution est pourtant simple : se servir des séquences qui sont disponibles dans la banque de données pour tenter de valider le tout. Nous avons souvent utilisé cette approche afin de faire concorder nos séquences à celles des autres. Par exemple, les séquences (GQ280358) (tableau 4.1), FJ524395 (Rivera et Currie, 2009) et AF425841 (Mitchell *et al.*, 2002) proviennent toutes du gène COI de *C. dacotensis*, mais les deux premières ne peuvent être comparées puisqu'elles n'ont aucun recouvrement.

Parmi les bases de données utilisées, nous avons choisi GenBank du NCBI et avons aussi consulté le BOLD<sup>16</sup>. GenBank est la banque la plus consultée, tandis que celle du BOLD est beaucoup moins connue. Bien qu'elle soit beaucoup moins exhaustive que celle du NCBI, elle s'adresse surtout aux données issues de COI, mais on peut y retrouver aussi ITS, matK et rbcL. Au fil des dernières années, le contenu du BOLD<sup>16</sup> s'est beaucoup amélioré. Il y figure désormais quelques images représentant certaines espèces de mouches noires et on peut maintenant télécharger quelques séquences. Le seul problème est qu'il y a dans cette banque 4719 entrées de séquences de mouches noires, dont seulement 83 séquences de COI pouvant être visualisées. De ces 83 séquences, presque la moitié est issue de *Simulium reptans*. Globalement, le tiers des données proviendrait du Royal Ontario Museum (figure 7.6<sup>16</sup>). Où sont les autres séquences ? Le NCBI quant à lui nous fournit 2631 séquences.



**FIGURE 7.6 :** Origine des données du BOLD sur les similies.

<sup>16</sup> <http://www.boldsystems.org/views/taxbrowser.php?taxid=707> (3 juin 2010)

### 7.3 Conclusion et perspective

Encore beaucoup trop d'études ne se basent que sur un seul niveau de classification, morphologique, cytologique ou moléculaire. Je recommande fortement une approche intégrée qui prendrait en considération les trois niveaux pour ainsi travailler sur chacun des individus représentant une espèce. À ce jour, la façon de conserver les spécimens fait que toutes les analyses, morphologique (éthanol et Carnoy), cytologique (Carnoy) et moléculaire (éthanol), sont difficilement réalisables sur un seul et même individu. Seules trois études (Collins *et al.*, 1987; Conflitti *et al.*, 2010; Pramual *et al.*, 2005) rapportent l'isolement de l'ADN à partir de spécimens fixés dans le Carnoy. Comment fait-on pour être certain à 100% de la correspondance entre les trois niveaux sachant très bien que la plupart du temps, les chercheurs n'utilisent pas une seule larve pour effectuer toutes les analyses requises? Cette approche est problématique, car plus d'une espèce sœur est souvent présente sur un site d'échantillonnage (Rivera et Currie, 2009).

Ce problème constitue l'un des enjeux principaux pour l'avenir. Nous appuyons donc ce que dit et recommande Castillo (2008) : « Accordingly, matching a chromosomally verified sibling species with a corresponding COI sequence can be difficult or even impossible unless both analyses (molecular and cytological) are performed on the same individual. In other words, it is often impossible to cytologically confirm the identity of specimens used for DNA analysis. Thus, I recommend that in the future a single individual must be sectioned into three parts soon after collection and while still alive: the abdomen, where the salivary glands are located, must be preserved in Carnoy's fixative for cytological analysis; the thorax must be preserved in ethanol for molecular analysis; and the head, where most taxonomic characters needed for morphological identification are found, should be preserved in ethanol and retained as a voucher ». Rivera et Currie (2009) proposent aussi cette solution. Or, une publication récente (Conflitti *et al.*, 2010), mentionne qu'il serait possible d'obtenir, à l'intérieur de deux mois, des données moléculaires sur des spécimens fixés dans le Carnoy. Cependant, rien de tel ne sera possible si aucun esprit de coopération ne s'installe entre les laboratoires spécialisés.

Dans cette perspective, un éventuel programme de recherche devrait cibler prioritairement deux cas actuellement problématiques.

Le cas de *S. vittatum* semble intéressant, car il existe deux morphotypes et plusieurs cytotypes en Amérique du Nord. Tout d'abord, la question des morphotypes pourrait être réglée par des expériences en laboratoire, clarifiant le rôle de l'environnement sur l'expression phénotypique du patron des taches de l'apotome céphalique. En effet, en utilisant les oeufs issus d'une seule ponte, il suffit de scinder celle-ci en deux groupes et d'élever chacun sur des substrats différents. Quant à la différenciation moléculaire entre cytotypes, il faut rappeler qu'aucune des études actuelles n'a pu démontrer une distinction nette. De plus, il n'est pas clair que celle-ci ne puisse être attribuée à des variations biogéographiques. En appliquant une approche intégrée, il faudrait constituer un corpus de spécimens de diverses provenances en Amérique du Nord afin d'établir s'il existe une corrélation, et de quel type, entre la biogéographie, la cytogénétique et les données moléculaires.

Finalement, il existe de sérieux doutes quant à la taxinomie des genres *Prosimulium* et *Helodon*. En effet, la clé actuelle permettant d'identifier à l'espèce des spécimens de ces genres est en réalité un amalgame de divers critères géographiques, morphologiques et cytogénétiques entre lesquels la corrélation est souvent loin d'être nette. Ces problèmes ne pourront être résolus qu'à partir d'une approche intégrée où chacun des spécimens sera pleinement caractérisé selon les critères précédents et les marqueurs moléculaires appropriés, dont on devra évaluer en cours de route.

## 7.4 Bibliographie

Adler, P.H., Crosskey, R.W. (2009) "World blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory", <http://www.entweb.clemson.edu/biomia/pdfs/blackflyinventory.pdf> (3 juin 2010)

Adler, P.H., Currie, D.C., Wood, D.M. (2004) "The blackflies of North America". Cornell University Press, Ithaca, NY, USA, 941 pages.

Adler, P.H., Kim, K.C. (1984) "Ecological characterization of two sibling, IIL-1 and IS-7, in the *Simulium vittatum* complex (Diptera: Simuliidae)", Canadian Journal of Zoology 62: 1308-1315.

Adler, P.H., McCreadie, J.W. (1997) "The hidden ecology of black flies: sibling species and ecological scale", American Entomologist Fall: 153-161.

Alvan-Aguilar, M.A., Hamada, N., Adler, P.H., Luz, S.L.B. (2005) "Cytotaxonomy of *Simulium cauchense* Floch & Abonnenc and *Simulium quadrifidum* Lutz (Diptera: Simuliidae) in Brazilian Amazonia", Memórias do Instituto Oswaldo Cruz, Rio de Janeiro 100: 249-257.

Back, C., Harper, P.P. (1978) "Les mouches noires (Diptera : Simuliidae) de deux ruisseaux des Laurentides, Québec", Annales de la Société Entomologique du Québec 23: 55-66.

Back, C., Harper, P.P. (1979) "Succession saisonnière, émergence, voltinisme et répartition de mouches noires des Laurentides (Diptera: Simuliidae)", Canadian Journal of Zoology 57: 627-639.

Beckenbach, A.T., Borkent, A. (2003) "Molecular analysis of the biting midges (Diptera: Ceratopogonidae), based on mitochondrial cytochrome oxidase subunit 2", Molecular Phylogenetics and Evolution 27: 21-35.

Bedo, D.G. (1975) "Polytene chromosomes of three species of blackflies in the *Simulium pictipes* group (Diptera: Simuliidae)", Canadian Journal of Zoology 53: 1147-1164.

Bedo, D.G. (1976) "Polytene chromosomes in pupal and adult blackflies (Diptera: Simuliidae)", Chromosoma 57: 387-396.

Bertone, M.A., Courtney, G.W., Wiegman, B.M. (2008) "Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes", Systematic Entomology 33: 668-687.

Brockhouse, C.L., Vajime, C.G., Marin, R., Tanguay, R.M. (1993) "Molecular identification of onchocerciasis vector sibling species in black flies (Diptera: Simuliidae)", *Biochemical and Biophysical Research Communications* 194: 628-634.

Brower, A.V.Z. (1994) "Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution", *Proceedings of the National Academy of Sciences* 91:6491-6495.

Campbell, N.A., Mathieu, R. (1995) "Biologie", ERPI, édition du nouveau pédagogique inc., Canada, 1990 pages.

Carew, M.E., Pettigrove, V., Hoffmann, A.A. (2003) "Identifying chironomids (Diptera: Chironomidae) for biological monitoring with PCR-RFLP", *Bulletin of Entomological Research* 93: 483-490.

Carew, M.E., Pettigrove, V., Cox, R.L., Hoffmann, A.A. (2003a). "DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae)", *Journal of the North American Benthological Society* 26:586-599.

Castillo, J.M.R. (2008) "Utility of the cytochrome oxidase I (COI) for species identification and phylogeographic analysis in black flies (Diptera: Simuliidae)", *Masters in Sciences, Université de Toronto*.

Caterino, M.S., Cho, S., Sperling, F.A. (2000) "The current state of insect molecular systematics: a thriving tower of Babel", *Annual Review of Entomology* 45: 1-54.

Chance, M.M. (1970) "The functional morphology of the mouthparts of blackfly", *Quaestiones Entomologicae* 6: 245-284.

Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffex, P.C., Besansky, N.J., Finnerty, V. (1987). "A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex", *American Journal of Tropical Medicine and Hygiene* 37: 37-41.

Conflitti, I.M., Kratochvil, M.J., Spironello, M., Shields, G.F., Currie, D.C. (2010). "Good species behaving badly: Non-monophyly of black fly sibling species in the *Simulium articum* complex (Diptera: Simuliidae)", *Molecular Phylogenetics and Evolution* 57: 245-257.

Craig, D.A. (1975) "The larvae of Tahitian Simuliidae", *Journal of Medical Entomology* 12: 463-476.

Craig, D.A. (2004) "Three new species of *Inseliellum* (Diptera: Simuliidae) from Polynesia", *Zootaxa* 450: 1-18.

Crossby, T.K. (1974) "Life history stages and taxonomy of *Autrosimulium tillyardianum*", *New Zealand Journal of Zoology* 1: 2-28.



Crosskey, R.W. (1990) "The natural history of blackflies", Edition John Wiley & Sons Inc, New York, U.S.A., 711 pages.

Crosskey, R.W., Howard, T.M. (1997) "A new taxonomic and geographical inventory of world blackflies (Diptera: Simuliidae)", The Natural History Museum, London.

Crosskey, R.W., Howard, T.M. (2004) "A revised taxonomic and geographical inventory of world blackflies (Diptera: Simuliidae)", <http://www.nhm.ac.uk/research-curation/research/projects/blackflies/> (3 juin 2010)

Currie, D.C. (1986) "An annotated list of and keys to the immature black flies of Alberta (Diptera: Simuliidae)", *Memoirs of the Entomological Society of Canada* No 134, 90 pages.

Currie, D.C., Adler, P.H. (2000) "Update on a survey of the black flies (Diptera: Simuliidae) from the Northwest territories and Nunavut project", *Arctic Insect News* 11: 6-9.

Currie, D.C., Adler, P.H. (2008) "Global diversity of black flies (Diptera: Simuliidae) in freshwater", *Hydrobiologia* 595: 469-475.

Davies, D.M., Peterson, B.V., Wood, D.M. (1962) "The black flies (Diptera: Simuliidae) of Ontario. Part I. Adult identification and distribution with descriptions of six new species", *Proceedings of the Entomological Society of Ontario* 92: 69-154.

Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R., Thorpe, R.S. (2007) "Validation of the barcoding gene COI for use in forensic genetic species identification", *Forensic Science International* 173: 1-6.

Day, J.C., Goodall, T.I., Post, R.J. (2008) "Confirmation of the species status of the blackfly *Simulium galeratum* in Britain using molecular taxonomy", *Medical and Veterinary Entomology* 22: 55-61.

Duncan, G.A., Adler, P.H., Pruess, K.P., Powers, T.O. (2004) "Molecular differentiation of two sibling species of the black fly *Simulium vittatum* (Diptera: Simuliidae) based on random amplified polymorphic DNA", *Genome* 47: 373-379.

Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., Wilson, A. (2009) "Geneious v4.7", <http://www.geneious.com/>

Dusfour, I., Linton, Y.-M., Cohuer, A., Harbach, R.E., Baimai, V., Trung, H.D., Seng, C.M., Matusop, A., Manguin, S. (2004) "Molecular evidence of speciation between island and continental population of *Anopheles (Cellia) sundaicus* (Diptera : Culicidae), a principal malaria vector taxon in southeast Asia", *Journal of Medical Entomology* 41: 287-295.

- Finn, D.S., Adler, P.H. (2006) "Population genetic structure of a rare high-elevation black fly, *Metacnephia coloradensis*, occupying Colorado outlet streams", *Freshwater Biology* 51: 2240-2251.
- Fukatsu, T. (1999) "Acetone preservation: a practical technique for molecular analysis", *Molecular Ecology* 8: 1935-1945.
- Gaudreau, C. (2005) "Identification moléculaire des simulies (Diptera: Simuliidae) du Québec", Mémoire en sciences de l'environnement, Université du Québec à Trois-Rivières.
- Gaudreau, C., LaRue, B., Charbonneau, V., Charpentier, G. et Craig, A.D. (2008) "Recent origins and genetics fragmentation of *Simulium* (*Hebridosimulium*) *laciniatum* (Diptera), the endemic Fijian simuliid", *Invertebrate Systematics* 22 :555-562.
- Gaudreau, C., Charpentier, G. (2010) "Seasonal and spatial distributions of blackfly larvae in two lake outlet streams of the Mauricie Region of Quebec, and species survey in parts of southern Quebec Territory", *Northeastern Naturalist* (en revision).
- Gaudreau, C., LaRue, B., Charpentier, G. (2010a) "Molecular comparaison of Quebec and Newfoundland populations of the blackfly, *Simulium vittatum*, species complex", *Medical and Veterinary Entomology* 24: 214-217.
- Gaudreau, C., LaRue, B., Charpentier, G. (2010b) "Phylogenetic suitability of COI, COII and 18S in land invertebrates : a black fly (Diptera) case study", manuscrit en préparation.
- Gray, E. (2010) "The University of Georgia black fly colony", *Wing Beats* 22-25.
- Grenier, P. (1953) "Simuliidae de France et d'Afrique du Nord", Editeur-libraire pour les sciences naturelles, Paris, 171 pages.
- Guindon, S., Gascuel, O. (2003) "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood", *Systematics Biology* 52: 696-704.
- Harry, M. (2001) "Génétique moléculaire et évolutive", Maloine, Paris, France, 325 pages.
- Hebert, P.D.N., Cywinska, A., Ball, S.L, deWaard, J.R. (2003) "Biological identifications through DNA barcodes", *Proceeding of the Royal Society of London B* 270: 313-321.
- Hwang, U., Kim, W. (1999) "General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics", *The Korean Journal of Parasitology* 37: 215-228.

- Joy, D.A., Craig, D.A., Conn, J.E. (2007) "Genetic variation tracks ecological segregation in Pacific island black flies", *Heredity* 99:452-459.
- Kjer, K.M. (2004) "Aligned 18S and insect phylogeny", *Systematics Biology* 53: 506-514.
- Koch, D.A., Duncan, G.A., Parsons, T.J., Pruess, K.P., Powers, T.O. (1998) "Effects of preservation methods, parasites, and gut contents of black flies (Diptera: Simuliidae) on polymerase chain reaction products", *Journal of Medical Entomology* 35: 314-318.
- Krueger, A. (2006) "Guide to blackflies of the *Simulium damnosum* complex in eastern and southern Africa", *Medical and Veterinary Entomology* 20: 60-75.
- Krüger, A., Gelhaus, A., Garms, R. (2000) "Molecular identification and phylogeny of East African *Simulium damnosum* s.l. and their relationship with West African species of the complex (Diptera: Simuliidae)", *Insect Molecular Biology* 9: 101-108.
- Krueger, A., Mustapha, M., Kalinga, A.K., Tambala, P.A.J., Post, R.J., Maegga, B.T.A. (2006) "Revision of the Ketaketa subcomplex of blackflies of the *Simulium damnosum* complex", *Medical and Veterinary Entomology* 20: 76-92.
- Laparé, R. (1980) "Écologie des mouches noires (Diptera : Simuliidae) de LG2 Radisson (Territoire de la Baie James, Québec)", *Mémoire en sciences de l'environnement, Université du Québec à Trois-Rivières*.
- LaRue, B., Gaudreau, C., Bagre, H.O., Charpentier, G. (2009) "Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies (Diptera : Simuliidae)", *Molecular Phylogenetic and Evolution* 53: 749-757.
- Le Guyader, H. (2006) "Un code-barres à ADN pour identifier les espèces?", *Pour la science* 340: 16-19.
- Lopez, P., Casane, D., Philippe, H. (2002) "Phylogénie et évolution moléculaires", *Médecine/ Sciences* 18: 1146-1154.
- Lucas, P., Hunter, F. (1999) "Phenotypic plasticity in the labral fan of simuliid larvae (Diptera): effect of seston load on primary-ray number", *Canadian Journal of Zoology* 77: 1843-1849.
- Mackay, R.J. (1969) "Aquatic insect communities of a small stream on Mont St-Hilaire, Québec", *Journal of the Fisheries Research Board of Canada* 26: 1157-1183.
- Maurand, J. (1973) "Recherches biologiques sur les microsporidies des larves de simulies", *Thèse de doctorat, Université des Sciences et Technologies du Languedoc*.

- McCreadie, J.W., Adler, P.H., Colbo, M.H. (1995) "Community structure of larval black flies (Diptera : Simuliidae) from the Avalon Peninsula, Newfoundland", *Annals of the Entomological Society of America* 88: 51-57.
- Miller, B.R., Crabtree, M.B., Savage, H.M. (1997) "Phylogenetic relationship of the Culicomorpha inferred from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera)", *Insect Molecular Biology* 6: 105-114.
- Minhas, M.S., Brockhouse, C.L., Adler, P.H. (2005) "The black fly (Diptera: Simuliidae) fauna of Prince Edward Island, Canada", *Northeastern Naturalist* 12: 67-76.
- Mitchell, A., Sperling, F.A.H., Hickey, D.A. (2002) "Higher level phylogeny of mosquitoes (Diptera: Culicidae): mtDNA data support a derived placement for Toxorhynchites", *Insect Systematics and Evolution* 33: 163-174.
- Moralais, I., Severson, D.W. (2002) "Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome C oxidase subunit I (COI) from *Aedes aegypti*", *DNA Sequence* 13: 123-127.
- Morales-Hojas, R., Post, R.J., Cheke, R.A., Wilson, M.D. (2002) "Assessment of rDNA IGS as a molecular marker in the *Simulium damnosum* complex", *Medical and Veterinary Entomology* 16: 395-403.
- Morales-Hojas, R., Krueger, A. (2009) "The species delimitation problem in the *Simulium damnosum* complex, blackfly vectors of onchocerciasis", *Medical and Veterinary Entomology* 23: 257-268.
- Moulton, J.K. (2000) "Molecular sequence data resolves basal divergences within Simuliidae (Diptera)", *Systematic Entomology* 25:95-113.
- Moulton, J.K. (2003) "Can the current molecular arsenal adequately track rapid divergence events within Simuliidae (Diptera)?", *Molecular Phylogenetics and Evolution* 27: 45-57.
- Palmer, R.W., Craig, D.A. (2000) "An ecological classification of primary labral fans of filter-feeding black fly (Diptera : Simuliidae) larvae", *Canadian Journal of Zoology* 78: 199-218.
- Pasternak, J. (1964) "Chromosome polymorphism in the blackfly *Simulium vittatum* (Zett)", *Canadian Journal of Zoology* 42: 135-158.
- Pevsner, J. (2003) "Bioinformatics and functional genomics", Wiley-Liss, USA, 752 pages.
- Peterson, B.V. (1970) "The Prosimulium of Canada and Alaska (Diptera: Simuliidae)", *Memoirs of the Entomological Society of Canada* 69: 216 pages.

Phayuhasena, S., Colgan, D.J., Kuvangkadilok, C., Pramual, P., Baimai, V. (2010). "Phylogenetic relationships among the black fly species (Diptera:Simuliidae) of Thailand based on multiple gene sequences", *Genetica* 138 : 633-648.

Pineau, P., Henry, M., Suspène, R., Marchio, A., Dettai, A., Debruyne, R., Petit, T., Lécu, A., Moisson, P., Dejean, A., Wain-Hobson, S., Vartanian, J-P. (2005) "A universal primer set for PCR amplification of nuclear histone H4 genes from all animal species", *Molecular Biology and Evolution* 22: 582-588.

Posada, D., Crandall, K.A. (1998) "MODELTEST: testing the model of DNA substitution", *Bioinformatics* 14: 817-818.

Posada, D., Crandall, K.A. (2001) "Selecting the best-fit model of nucleotides substitution", *Systematics Biology* 50: 580-601.

Post, R.J., Flook, P.K., Millest, A.L. (1993) "Methods for the preservation of insects for DNA studies", *Biochemical Systematics and Ecology* 21: 85-92.

Post, R.J., Mustapha, M., Krueger, A. (2007) "Taxonomy and inventory of the cytospecies and cytotypes of the *Simulium damnosum* complex (Diptera:Simuliidae) in relation to onchocerciasis", *Tropical Medicine and International Health* 12: 1342-1353.

Pramual, P., Kuvangkadilok, C., Baimai, V., Walton, C. (2005) "Phylogeography of the black fly *Simulium tani* (Diptera : Simuliidae) from Thailand as inferred from mtDNA sequences", *Molecular Ecology* 14: 3989-4001.

Pruess, K.P., Adams, B.J., Parsons, T.J., Zhu, X., Powers, T.O. (2000) "Utility of the mitochondrial cytochrome oxidase II gene for resolving relationships among black flies (Diptera:Simuliidae)", *Molecular Phylogenetics and Evolution* 16: 286-295.

Quillévére, D. (1979) "Contribution à l'étude des caractéristiques taxonomiques, bioécologiques et vectrices des membres du complexe *Simulium damnosum* présents en Côte d'Ivoire", *Travaux et documents de l'ORSTOM* 109. 304 pages.

Rivera, J., Currie, D.C. (2009) "Identification of Nearctic black flies using DNA barcodes (Diptera : Simuliidae)", *Molecular Ecology Resources* 9: 224-236.

Rodhain, F., Perez, C. (1985) "Les simulies : systématique, biologie, importance médicale", Dans *Précis d'entomologie médicale et vétérinaire*. Maloine s.a. éditeur, Paris, Pages 177-199.

Rodríguez-Pérez, M.A., Núñez-González, C.A., Lizarazo-Ortega, C., Sánchez-Varela, A., Wooten, M.C., Unnasch, T.R. (2006) "Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH dehydrogenase subunit 4 mitochondrial genes of the Onchocerciasis vector *Simulium ochraceum*", *Journal of Medical Entomology* 43: 701-706.

Ross, D.H., Craig, D.A. (1979) "The seven larval instars of *Prosimulium mixtum* Syme and Davies and *P. fuscum* Syme and Davies (Diptera: Simuliidae)", Canadian Journal of Zoology 57: 290-300.

Rothfels, K.H., Freeman. (1977) "The salivary gland chromosomes of seven species of *Prosimulium* (Diptera, Simuliidae) in the mixtum (IIIL-1) group", Canadian Journal of Zoology 55:482-507.

Rothfels, K.H., Fereday, R., Kaneps, A. (1978) "A cytological description of sibling species of *Simulium venustum* and *S. verecundum* with standard maps for the subgenus *Simulium* Davies (Diptera)", Canadian Journal of Zoology 56: 1110-1128.

Rothfels, K., Featherston, D. (1981) "The population structure of *Simulium vittatum* (Zett.): the IIIL-1 and IS-7 sibling species", Canadian Journal of Zoology 59: 1857-1883.

Schlötterer, C., Hauser, M.T., Haeseler, V., Tautz, D. (1994) "Comparative evolutionary analysis of rDNA ITS regions in drosophila", Molecular Biology and Evolution 11: 513-522.

Sharp, A.L., Hunter, F. (2008) "Chiasmata meiosis in male black fly (Diptera: Simuliidae) larvae associated with mermithid infections (Nematoda: Mermithidae)", Canadian Journal of Zoology 86: 1198-1202.

Shields, G.F., Christiaens, B.A., Van Leuven, M.L., Hartman, A.L. (2009) "Reproductive status and continuity of taxa of the *Simulium arcticum* complex (Diptera: Simuliidae) at the Clearwater river, Montana (2007, 2008, and 2009)", Western North American Naturalist 69: 511-520.

Shoemaker, D.D., Dyer, K.A., Ahrens, M., McAbee, K., Jaenike, J. (2004) "Decreased diversity but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection", Genetics 168:2049-2058.

Simon, C., Frati, F., Beckenback, A., Crespi, B., Liu, H., Flook, P. (1994) "Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers", Annals of the Entomological Society of America 87: 651-701.

Simmons, K.R., Edman, J.D. (1981) "Sustained colonization of the black fly *Simulium decorum* Walker (Diptera: Simuliidae)", Canadian Journal of Zoology 59: 1-7.

Smart, J.S. (1944) "The British Simuliidae with keys to the species in the adult, pupal and larval stages", Freshwater Biology Association of the British Empire Scientific Publication 9: 1-57.

Snyder, T.P., Linton, M.C. (1983) "Electrophoretic and morphological separation of *Prosimulium fuscum* and *P. mixtum* larvae (Diptera: Simuliidae)", Canadian Entomologist 115: 81-87.

Sohn, U.I., Rothfels, K.H., Straus, N.A. (1975) "DNA : DNA hybridization studies in black flies", *Journal of Molecular Evolution* 5: 75-85.

Spironello, M., Hunter, F.F., Craig, D.A. (2002) "A cytological study of the Pacific black fly *Simulium cataractarum* (Diptera: Simuliidae)", *Canadian Journal of Zoology* 80: 1810-1816.

St-Onge, M., LaRue, B., Charpentier, G. (2008) "A molecular revision of the taxonomic status of mermithid parasites of black flies from Québec (Canada)", *Journal of Invertebrate Pathology* 98: 299-306.

Stone, A. (1965) "Family Simuliidae", pp. 181-189. In Stone, A., Sabrosky, C.W., Wirth, W.W., Foote, R.H., Coulson, J.R. (eds). "A Catalog of the Diptera of America north of Mexico. United States Department of Agriculture Handbook No 276, Agricultural Research Service, Washington, D.C. 1696 pages."

Swofford, D.L. (2002) "PAUP\*. Phylogenetic Analysis Using Parsimony (\*and other methods)", Sinauer associates, Sunderland, Massachusetts.

Sytsma, K.J., Pires, J.C. (2001) "Plant systematics : a half-century of progress (1950-2000) and futur challenges-Plant systematics in the next 50 years—Re-mapping the new frontier", *Taxon*. 50: 713-732.

Tang, J., Kenneth, P., Unnasch, T.R. (1996) "Genotyping North American black flies by means of mitochondrial ribosomal RNA sequences", *Canadian Journal of Zoology* 74: 39-46.

Tang, J., Moulton, J.K., Pruess, K., Cupp, E.W., Unnasch, T.R. (1998) "Genetic variation in North American black flies in the subgenus *Psilopelmia* (Simulium: Diptera: Simuliidae)", *Canadian Journal of Zoology* 76: 205-211.

Teshima, I. [E.] (1972) "DNA : DNA hybridization in blackflies (Diptera : Simuliidae)", *Canadian Journal of Zoology* 50: 931-940.

Thanwisai, A., Kuvangkadilok, C., Baimai, V. (2006) "Molecular phylogeny of black flies (Diptera : Simuliidae) from Thailand, using ITS2 rDNA", *Genetica* 128: 177-204.

Twinn, C.R. (1936) "The black flies of eastern Canada", *Canadian Journal Research* 14: 97-150.

Winn, A.F., Beaulieu, G. (1932) "Part II. Diptera. Revised and supplemented. In Petch, C.E. and J.B. Maltais. A preliminary list of the insects of the Province of Quebec", *Report of the Quebec Society for the Protection of Plants* 24: 1-100.

Wolfe, L.S., Peterson, D.G. (1959) "Black flies (Diptera: Simuliidae) of the forest of Quebec", *Canadian Journal of Zoology* 37: 137-159.

- Wood, D.M., Peterson, B.V., Davies, D.M., Gyorkos, H. (1963) "The Blackflies (Diptera : Simuliidae) of Ontario. Part II. Larval identification, with description and illustration", *Proceedings of the Entomological Society of Ontario* 93: 99-129.
- Xiong, B., Kocher, T.D. (1991) "Comparaison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera : Simuliidae)", *Genome* 34: 306-311.
- Xiong, B., Kocher, T.D. (1993) "Phylogeny of sibling species of *Simulium venustum* and *S. verecundum* (Diptera: Simuliidae) based on sequences of the mitochondrial 16S rRNA gene", *Molecular Phylogenetics and Evolution* 2: 293-303.
- Young, I., Coleman, A.W. (2004) "The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example", *Molecular Phylogenetics and Evolution* 30: 236-242.
- Zhang, Y. (2005) "Balancing food availability and hydrodynamic constraint: phenotypic plasticity and growth in *Simulium noelleri* blackfly larvae", *Oecologia* 147: 39-46.
- Zhang, Y., Malmqvist, B. (1996) "Relationships between labral fan morphology, body size and habitat in North Swedish blackfly larvae (Diptera : Simuliidae)", *Biological Journal of the Linnean Society* 59: 261-280.
- Zhu, X., Pruess, K.P., Powers, T.O. (1998) "Mitochondrial DNA polymorphism in a black fly, *Simulium vittatum* (Diptera: Simuliidae)", *Canadian Journal of Zoology* 76: 440-447.