

Université de Montréal

Étude du polymorphisme du gène majeur d'histocompatibilité de classe II β (MHII β)
chez l'omble de fontaine (*Salvelinus fontinalis*)

par

Sébastien Croisetière

Département de Microbiologie et Immunologie
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en microbiologie et immunologie

Octobre, 2009

© Sébastien Croisetière, 2009

Université de Montréal

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
Étude du polymorphisme du gène majeur d'histocompatibilité de classe II β (MHI β)
chez l'omble de fontaine (*Salvelinus fontinalis*)

présentée par :
Sébastien Croisetière

a été évaluée par un jury composé des personnes suivantes :

Président-rapporteur	Dr Marc Drolet
Directeur de recherche	Dr Pierre Belhumeur
Co-directeur de recherche	Dr Louis Bernatchez
Membre du jury	Dr Jacques Thibodeau
Examinatrice externe	Dre Céline Audet
Représentant du doyen de la FES	Dr Louis de Repentigny

RÉSUMÉ

Les molécules classiques du complexe majeur d'histocompatibilité de classe II (CMHII) sont des glycoprotéines de surface spécialisées dans la présentation de peptides, principalement dérivés de pathogènes extracellulaires, aux récepteurs des lymphocytes T CD4⁺ afin d'initier la réponse immunitaire adaptative. Elles sont encodées, avec celles du CMH de classe I, par les gènes les plus polymorphiques identifiés jusqu'à maintenant, avec plusieurs loci et une grande diversité allélique à chacun d'eux. De plus, le polymorphisme des gènes du CMHII n'est pas limité qu'aux séquences codantes. Il est également observé dans les promoteurs où on a démontré ses effets sur le niveau d'expression des gènes. La variation de la régulation d'un gène est considérée comme un facteur important et pour laquelle des modifications morphologiques, physiologiques et comportementales sont observées chez tous les organismes. Des séquences d'ADN répétées impliquées dans cette régulation ont été identifiées dans les régions non-codantes des génomes. D'un autre côté, la sélection par les pathogènes permettrait l'évolution et le maintien du polymorphisme des gènes du CMH chez les vertébrés. À ce sujet, plusieurs études ont montré l'implication de différents allèles du CMH dans la résistance ou la susceptibilité aux maladies. Cette étude avait pour objectifs de caractériser le polymorphisme du gène MHIIβ chez l'omble de fontaine (*Salvelinus fontinalis*) et de documenter ses effets au niveau de la survie conférée par des allèles et/ou génotypes particuliers lors d'une infection, ainsi que sur la variation du niveau d'expression du gène dans différentes conditions.

Dans une première partie, nous avons identifié un total de 6 allèles du gène MHIIβ, désignés *Safo-DAB*0101* à *Safo-DAB*0601*, qui montrent une grande similarité avec les séquences codantes provenant de poissons téléostéens et de l'humain. L'analyse des séquences du domaine β1 a permis de détecter l'effet d'une pression sélective positive pour maintenir le polymorphisme dans cette région de la molécule. Quatre de ces allèles ont été testés lors d'une expérience d'infection avec le pathogène *Aeromonas salmonicida* afin d'évaluer l'effet qu'ils pouvaient avoir sur la survie des poissons. Nous avons trouvé que l'allèle *DAB*0101* était significativement associé à la résistance à la furunculose. En plus d'avoir été identifié chez les individus homozygotes pour cet allèle, l'effet a également été

remarqué au niveau de la survie les poissons de génotype *DAB*0101/*0201*. À l'opposé, les facteurs de risque élevé obtenus pour les génotypes *DAB*0201/*0301* et *DAB*0301/*0401* suggèrent plutôt une association à la susceptibilité. Étant donné la faible fréquence à laquelle l'allèle *DAB*0101* a été retrouvé dans la population, le modèle de la sélection dépendante de la fréquence pourrait expliquer l'avantage conféré par ce dernier et souligne l'importance de ce mécanisme pour le maintien du polymorphisme du gène *MHIIβ* chez l'omble de fontaine.

Dans une seconde partie, nous avons rapporté la présence d'un minisatellite polymorphique formé d'un motif de 32 nucléotides dans le second intron du gène *MHIIβ*, et pour lequel un nombre exclusif de répétitions du motif a été associé à chaque allèle (69, 27, 20, 40, 19 et 25 répétitions pour les allèles *DAB*0101* à *DAB*0601* respectivement). L'expression relative de quatre allèles a été évaluée dans des poissons hétérozygotes aux températures de 6 °C et 18 °C. Les résultats indiquent que les allèles possédant un long minisatellite montrent une réduction de l'expression du gène d'un facteur 1,67 à 2,56 par rapport aux allèles qui en contiennent un court. De même, des allèles qui incluent des minisatellites de tailles similaires n'affichent pas de différence significative au niveau de l'abondance du transcrit aux deux températures. De plus, l'effet répressif associé aux longs minisatellites est amplifié à la température de 18 °C dans des poissons de trois génotypes différents. Nous avons finalement observé une augmentation significative par un facteur 2,08 de l'expression totale du gène *MHIIβ* à la température de 6 °C. Ces résultats appuient l'implication des séquences d'ADN répétées dans la régulation de l'activité transcriptionnelle d'un gène et suggèrent qu'un minisatellite sensible aux différences de températures pourrait être soumis aux forces sélectives et jouer un rôle important dans l'expression de gènes et l'évolution des organismes poïkilothermes.

Mots clés : Complexe majeur d'histocompatibilité; antigène d'histocompatibilité de classe II; polymorphisme; séquences d'ADN répétitives; régulation de l'expression génique; résistance aux maladies; infection bactérienne; sélection; évolution; salmonidés.

SUMMARY

Classical major histocompatibility complex class II (MHCII) molecules are cell-surface glycoproteins specialized in the presentation of peptides, mainly derived from extracellular pathogens, to the antigen receptors of CD4⁺ T cells in the adaptive immune system. They are encoded, with those of the MHC class I, by the most polymorphic genes known to date, with multiple loci and high allelic diversity at each one. Moreover, the polymorphism within MHCII genes is not restricted to coding sequences. It has also been observed in promoters where it was shown to affect the expression level of the genes. Variation in gene regulation is believed to be an important factor from which modification in morphology, physiology or behaviour can be observed in all organisms. Repeated DNA sequences with functional roles in this regulation have been identified within the non-coding parts of the genomes. On the other hand, pathogen-driven selection is also believed to be important in the evolution and maintenance of the polymorphism of the MHC genes in vertebrates. Studies have shown the implication of different MHC alleles in disease resistance or susceptibility. In this study, our aims were to characterize the polymorphism of the MHII β gene in brook charr (*Salvelinus fontinalis*), to document its effects on the survival conferred by specific alleles and/or genotypes following an infection and on the variation of the expression level of the gene in different environmental conditions.

In a first part, we identified a total of 6 MHII β alleles, designated *Safo-DAB*0101* to *Safo-DAB*0601*, showing a high similarity to coding sequences from teleost fish and human. Analysis of the β 1 domain sequences indicates the effect of a positive selection pressure to select polymorphic mutations in that region of the molecule. Four of these alleles were tested in a challenge experiment against the pathogen *Aeromonas salmonicida* to evaluate their effect on fish survival. We found that one allele, *DAB*0101*, was significantly associated with resistance to furunculosis. In addition to homozygotes for this allele, its resistance effect was also detected in the heterozygote individuals of the *DAB*0101/*0201* genotype. In contrast, other allelic combinations, namely heterozygous genotypes *DAB*0201/*0301* and *DAB*0301/*0401* were significantly associated with increased susceptibility.

Given that its frequency was relatively low in the population, the negative frequency dependant selection hypothesis could explain the advantage associated with the allele *DAB*0101* over the other alleles and highlight the importance of this mechanism to sustain variation at the MHC in brook charr.

In a second part, we reported the identification of a polymorphic minisatellite formed of a 32 nucleotides motif in the second intron of *MHII β* gene, and for which distinctive repeat numbers of the motif were associated to each alleles (69, 27, 20, 40, 19 and 25 repeats for the *DAB*0101* to *DAB*0601* alleles respectively). Relative expression levels of four alleles were determined in heterozygous fish at temperature of 18 °C and 6 °C. Results indicate that alleles carrying the longest minisatellite showed a 1.67 to 2.56-fold reduction in the transcript expression relatively to the shortest one. In contrast, no significant differences were seen in the expression levels between alleles with comparable minisatellite length at both temperatures. Furthermore, the repressive activity associated to the longest minisatellite was more effective at temperature of 18 °C in fish from three different genotypes. We finally observed a significant 2.08-fold up-regulation of the total *MHII β* transcript amount at 6 °C. The results support the implication of repeated DNA sequences in the regulation of the gene transcriptional activity and suggest that a temperature-sensitive minisatellite could potentially be submitted to selective forces and therefore play an important role in gene expression and evolution in ectothermic organisms.

Keywords : Major histocompatibility complex; Histocompatibility antigens class II; polymorphism; repetitive DNA sequences; gene expression regulation; disease resistance; bacterial infection; selection; evolution; salmonids.

TABLE DES MATIÈRES

RÉSUMÉ	iii
SUMMARY	v
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX.....	xii
LISTE DES FIGURES.....	xiii
LISTE DES ABRÉVIATIONS.....	xvi
REMERCIEMENTS.....	xix
CHAPITRE 1 : REVUE DE LA LITTÉRATURE	1
1.1. Le complexe majeur d'histocompatibilité.....	2
1.1.1. Les régions du CMH.....	2
1.1.1.1. La région du CMH de classe I	3
1.1.1.1.1. Les gènes de la région du CMH de classe I.....	3
1.1.1.1.2. La molécule présentatrice d'antigènes du CMH de classe I.....	5
1.1.1.2. La région du CMH de classe II.....	5
1.1.1.2.1. Les gènes de la région du CMH de classe II.....	5
1.1.1.2.2. La molécule présentatrice d'antigènes du CMH de classe II.....	9
1.1.1.3. La région du CMH de classe III.....	9
1.1.2. Le contrôle de l'expression des gènes du CMH de classe II.....	10
1.1.2.1. Le promoteur des gènes du CMH de classe II	10
1.1.2.2. Le transactivateur des gènes de classe II	12
1.1.2.2.1. Le CIITA : sa structure et son mode d'action.....	12
1.1.2.2.2. Les régions régulatrices du gène MHC2TA	13
1.1.2.2.3. La modulation de l'expression du CIITA	13
1.1.3. La présentation antigénique	16
1.1.3.1. Les cellules présentatrices d'antigènes	17
1.1.3.2. Les mécanismes de présentation antigénique	19
1.1.3.2.1. La présentation antigénique par le CMH de classe I	22
1.1.3.2.2. La présentation antigénique par le CMH de classe II	22
1.1.4. La structure de la molécule du CMH de classe II	23
1.1.4.1 L'interaction du peptide avec la molécule du CMH de classe II	25
1.1.5. Le CMH chez les vertébrés non-mammaliens	29

1.1.5.1. Le CMH chez les oiseaux (poulet).....	30
1.1.5.2. Le CMH chez les reptiles.....	30
1.1.5.3. Le CMH chez les amphibiens	31
1.1.5.4. Le CMH chez les poissons osseux (les téléostéens)	31
1.1.5.4.1. Les gènes majeurs d’histocompatibilité chez les salmonidés	32
1.1.5.5. Le CMH chez les poissons cartilagineux (les chondrichthyens)	34
1.1.6. Les pressions évolutives sur les gènes du CMH	35
1.1.6.1. La polygénie et le polymorphisme des gènes du CMH	35
1.1.6.2. Le polymorphisme des gènes du CMH <i>en équilibre</i>	41
1.1.6.2.1. Les mécanismes de sélection par les pathogènes.....	41
1.1.6.2.2 Les mécanismes de sélection par la reproduction.....	44
1.1.6.3. L’association entre les gènes du CMH et les maladies.....	44
1.2. Les séquences d’ADN répétitives	47
1.2.1. Les séquences d’ADN répétitives dispersées (transposons)	49
1.2.1.1. Les LINE (<i>long interspersed nuclear element</i>).....	51
1.2.1.2. Les SINE (<i>short interspersed nuclear element</i>).....	54
1.2.1.3. Les rétrotransposons à LTR.....	55
1.2.1.4. Les transposons à ADN	57
1.2.2. Les séquences d’ADN répétées en tandem	60
1.2.2.1. Les ADN satellites	62
1.2.2.2. Les ADN minisatellites et microsatellites.....	64
1.2.2.3. Les mécanismes impliqués dans l’instabilité des mini- et microsatellites.....	68
1.2.2.3.1 Les structures secondaires de l’ADN.....	68
1.2.2.3.2. Les mini- et microsatellites et la réplication de l’ADN	72
1.2.2.3.2.1. Les effets de la réplication de l’ADN sur les microsatellites	74
1.2.2.3.2.2. Les effets de la réplication de l’ADN sur les minisatellites	75
1.2.2.3.3. Le réarrangement des mini- et microsatellites lors de la recombinaison homologue	76
1.2.2.4. La régulation de l’activité des gènes par les séquences d’ADN répétitives ..	80
1.2.2.4.1. Les microsatellites géniques	82
1.2.2.4.2. Les minisatellites géniques	88
1.2.2.5. Les séquences d’ADN répétitives et l’évolution des espèces	92
1.3. Mise en contexte	97

CHAPITRE 2: IDENTIFICATION OF MHC CLASS IIβ RESISTANCE/ SUSCEPTIBILITY ALLELES TO <i>AEROMONAS SALMONICIDA</i> IN BROOK CHARR (<i>SALVELINUS FONTINALIS</i>).....	100
2.1. Abstract.....	103
2.2. Introduction.....	104
2.3. Materials and methods.....	107
2.3.1. Fish and challenge test.....	107
2.3.2. MH class II β genotyping of brook charr.....	107
2.3.3. RNA isolation from peripheral blood leukocytes (PBL) of brook charr.....	108
2.3.4. Reverse Transcriptase (RT)-PCR from PBL.....	109
2.3.5. Data analysis.....	109
2.4. Results.....	111
2.4.1. Sequences of brook charr MH class II β gene.....	111
2.4.2. Polymorphism of MH class II β	112
2.4.3. Single expressed locus.....	113
2.4.4. Analysis of MH class II β in resistance to furunculosis.....	114
2.5. Discussion.....	116
2.5.1. Identification of brook charr MH class II β alleles.....	116
2.5.2. Variability of brook charr MH class II β alleles.....	116
2.5.3. Polymorphism of MH class II β leader sequences.....	118
2.5.4. Minisatellite in MH class II β intron 2.....	119
2.5.5. Association of MH class II β alleles with resistance or susceptibility.....	120
2.5.6. Mechanism for pathogen-driven balancing selection.....	121
2.6. Acknowledgements.....	125
2.7. References.....	126
2.8. Tables and Figures.....	135
 CHAPITRE 3: TEMPERATURE AND LENGTH-DEPENDENT MODULATION OF THE MH CLASS IIβ GENE EXPRESSION IN BROOK CHARR (<i>SALVELINUS FONTINALIS</i>) BY A <i>CIS</i>-ACTING MINISATELLITE.....	 143
3.1. Abstract.....	146

3.2. Introduction.....	148
3.3. Materials and Methods.....	153
3.3.1. Fish.....	153
3.3.2. RNA isolation from peripheral blood leukocytes (PBL) and head kidneys.....	154
3.3.3. Polymerase chain reaction (PCR), cloning and sequencing.....	155
3.3.4. Reverse transcriptase (RT).....	156
3.3.5. Multiplex PCR.....	157
3.3.6. Quantitative PCR.....	157
3.3.7. Sequence analysis.....	159
3.4. Results.....	160
3.4.1. Nucleotides sequences and minisatellite from the second intron of MHII β gene.	160
3.4.2. MHII β gene expression level is inversely correlated to minisatellite length.....	161
3.4.3. MHII β gene expression level is up-regulated at low temperature.	162
3.4.4. The minisatellite from MHII β gene is a functional cis-acting polymorphism.....	164
3.5. Discussion.....	166
3.5.1. Instability of the minisatellite found in the second intron of MHII β gene.	166
3.5.2. Minisatellite modulation of the MHII β gene expression level.	167
3.5.3. Reduced immune gene expression at high temperature in brook charr?.....	169
3.5.4. Other elements contributing to the polymorphism of the MHII β gene in salmonids.	173
3.5.5. Identification of a new regulatory element submitted to natural selection?	176
3.6. Supplementary Material.....	178
3.7. Acknowledgements.....	179
3.8. Literature Cited.....	180
3.9. Tables and Figures.....	193
 CHAPITRE 4: DISCUSSION.....	 205
 4.1 Le gène MHII β chez l'omble de fontaine (<i>Salvelinus fontinalis</i>).....	 206
4.1.1. Identification des allèles du gène MHII β	206
4.1.1.1. Polymorphisme des allèles du gène MHII β	207
4.1.1.2. Polymorphisme de la séquence signal.....	208
4.1.2. Variabilité du second intron du gène MHII β	209

4.1.2.1. Instabilité du minisatellite chez l'omble de fontaine	210
4.1.2.2. L'origine du minisatellite chez les salmonidés	211
4.1.2.3. Un seul locus MHII β exprimé chez l'omble de fontaine.....	213
4.2. Régulation de l'activité transcriptionnelle du gène MHII β	213
4.2.1. Modulation de l'expression du gène MHII β par le minisatellite	214
4.2.1.1. Identification de l'effet du minisatellite dans différents tissus	214
4.2.1.2. Identification de l'effet du minisatellite dans plusieurs conditions expérimentales	215
4.2.2. Modulation de l'expression du gène MHII β par la température.....	216
4.2.2.1 Effet de la température sur l'activité du minisatellite	217
4.2.2.2. Modulation de l'expression de certains gènes du système immunitaire par la température.....	218
4.2.3. Analyse du motif du minisatellite et mécanisme fonctionnel	220
4.2.4. Implications fonctionnelles de la variation de l'expression du MHII β	221
4.3 Le gène MHII β et la résistance ou susceptibilité à une infection	224
4.3.1. Performances de survie associées aux allèles et aux génotypes.....	225
4.3.2. Conditions expérimentales de l'infection avec <i>Aeromonas salmonicida</i>	226
4.3.3. Analyse des positions polymorphiques de l'allèle <i>Safo-DAB*0101</i>	227
4.4 Pressions évolutives sur les gènes du CMH.....	229
4.4.1. Modèles de sélection par les pathogènes	229
4.4.2. Modèle du “ <i>genomic design</i> ” pour l'évolution du second intron.....	231
CHAPITRE 5 : CONCLUSION	235
CHAPITRE 6 : BIBLIOGRAPHIE	238
CHAPITRE 7 : ANNEXES	273
7.1. Figure supplémentaire.....	274
7.2. Curriculum vitae	275

LISTE DES TABLEAUX

CHAPITRE 1

Tableau 1.1. Nombre d'allèles pour les gènes du complexe HLA.....	37
Tableau 1.2. Variation en acides aminés entre différents allèles du complexe HLA.....	39
Tableau 1.3. Nombre de publications rapportant une association positive ou négative entre le CMH et des maladies.....	45
Tableau 1.4. Pathologies associées à l'amplification d'un microsatellite (SSR).....	85

CHAPITRE 2

Table 2.1 Mean diversity in the nucleotides (nt.) and amino acid (AA) sequences, mean proportion of observed synonymous (d_s) and nonsynonymous substitutions (d_N) and the ratio between d_N and d_s of brook charr MH class II β gene	135
Table 2.2 Hazard ratio in the single allele groups with at least one of the given class II β (DAB) alleles present	136
Table 2.3 Hazard ratio and mortality in groups with class II β (DAB) genotypes.....	137

CHAPITRE 3

Table 3.1 Primer sequences for multiplex and quantitative RT-PCR in brook charr.....	193
Table 3.2 Characteristics of the second intron in MHII β alleles from brook charr (<i>Safo-DAB*</i>) and other salmonid fishes	194

LISTE DES FIGURES

CHAPITRE 1

Figure 1.1. Localisation et organisation génétique du complexe majeur d'histocompatibilité chez l'humain (HLA).....	4
Figure 1.2. Structure des gènes du CMH retrouvés chez les mammifères.....	6
Figure 1.3. Structure des molécules du CMH de classe I et de classe II.....	7
Figure 1.4. Régulation de la transcription des gènes du CMH de classe II.....	11
Figure 1.5. Région régulatrice du gène CIITA.....	14
Figure 1.6. Activation et inhibition de l'expression du gène CIITA.....	15
Figure 1.7. Voies de présentation antigénique.....	20
Figure 1.8. Structure tridimensionnelle de la molécule du CMH de classe II, HLA-DR1 ...	24
Figure 1.9. Le peptide antigénique en conformation étendue.....	26
Figure 1.10. Modélisation tridimensionnelle de la niche peptidique de la molécule HLA-DR1.....	27
Figure 1.11. Structures tridimensionnelles montrant la variabilité de séquence des domaines $\alpha 1\beta 1$ des molécules du CMH de classe II.....	40
Figure 1.12. Structures tridimensionnelles montrant la superposition des domaines $\alpha 1\beta 1$ de molécules du CMH de classe II.....	42
Figure 1.13. Séquences d'ADN répétitives retrouvées dans les génomes eucaryotes.....	48
Figure 1.14. Caractéristiques des séquences d'ADN répétitives dispersées des génomes eucaryotes.....	50
Figure 1.15. Mécanisme d'intégration du LINE L1.....	53
Figure 1.16. Mécanisme d'intégration des rétrotransposons à LTR.....	56
Figure 1.17. Mécanisme d'intégration d'un transposon à ADN.....	59
Figure 1.18. Taille des motifs, longueur et abondance des ADN satellites chez les eucaryotes.....	61
Figure 1.19. Représentation schématique d'un monomère d'ADN satellite et des processus dans lesquels il peut être impliqué.....	63
Figure 1.20. Conformations non-B de l'ADN impliquées dans des réarrangements.....	69
Figure 1.21. Le modèle de réplication avec glissement pour les séquences d'ADN répétées en tandem.....	73
Figure 1.22. Modèles de recombinaison.....	79

Figure 1.23. Modèle de réparation d'un bris double-brin par glissement (<i>DSB repair slippage</i>) pour des séquences répétées en tandem.....	81
Figure 1.24. Fonctions régulatrices des microsatellites	83
Figure 1.25. Localisation des microsatellites responsables de pathologies humaines	84

CHAPITRE 2

Figure 2.1. Nucleotide sequences of the MH class II β cDNA and partial gDNA of six different alleles from brook charr	138
Figure 2.2. Alignment of brook charr MH class II β amino acid sequences with MH class II β sequences from other vertebrates	139
Figure 2.3. Neighbor-joining tree of the second exon of the MH class II β gene in brook charr and other salmonids	141
Figure 2.4. Kaplan-Meier plot showing the survival of different MH class II β genotypes in brook charr exposed to the challenge experiment with <i>Aeromonas salmonicida</i>	142

CHAPITRE 3

Figure 3.1. Properties of the second intron of the MHII β gene in salmonids	195
Figure 3.2. Example of a minisatellite sequence within in the second intron of brook charr MHII β gene, allele <i>Safo-DAB*0301</i>	196
Figure 3.3. MHII β expression level in brook charr head kidney as a function of minisatellite number of repeats	197
Figure 3.4. MHII β expression level in brook charr head kidney as a function of temperature.....	198
Figure 3.5. Expression level of MHII β alleles as a function of both minisatellite length and temperature.....	199
Supplementary Figure S3.1. Alignment of the second intron nucleotide sequences preceding the minisatellite found in the MHII β gene from brook charr and other salmonids	200

Supplementary Figure S3.2. Alignment of the second intron nucleotide sequences following the minisatellite found in the MHII β gene from brook charr and other salmonids	202
--	-----

CHAPITRE 4

Figure 4.1. Structure secondaire hypothétique pour le minisatellite de type I.....	222
--	-----

CHAPITRE 7

Figure 7.1. Niveaux d'expression de la β_2 -microglobuline en fonction de la température...	274
---	-----

LISTE DES ABRÉVIATIONS

A : Adénine

ADN : Acide désoxyribonucléique

ADNc : ADN complémentaire

AMPc : Adénosine monophosphate cyclique

ARN : Acide ribonucléique

ARNm : ARN messenger

ARNt : ARN de transfert

APC : "*Antigen-presenting cell*"

β_2m : β_2 -microglobuline

C : Cytosine

CIITA : "*Class II transactivator*"

CD : Cellule dendritique

cM : Centimorgans

CMH : Complexe majeur d'histocompatibilité

CMHII : CMH de classe II

G : Guanine

HA : Hémagglutinine

HLA : "*Human leukocyte antigen*"

HSP70 : "*Heat shock protein 70*"

INF- γ : Interféron gamma

kb : Kilobase

kDa : Kilodalton

LINE : "*long interspersed nuclear element*"

LMP : "*Low molecular mass polypeptide*"

LTR : *“long terminal repeats”*

MIIC : *“MHC class II compartment”*

Mb : Mégabase

MHII β : (gène) Majeur d’histocompatibilité de classe II bêta

ORF : *“open reading frame”*

Pb : paires de bases

PBR : *“peptide-binding region”*

RE : réticulum endoplasmique

SINE : *“short interspersed nuclear element”*

SSR : *“simple sequence repeats”*

T : Thymine

TAP : *“Transporters associated with antigen processing”*

TCR : *“T cell receptors”*

TEC : *“Thymic epithelial cells”*

TGF- β : *“transforming growth factor-beta”*

TIR : *“terminal inverted repeats”*

TNF- α : *“Tumor necrosis factor alpha”*

UTR : *“Untranslated region ”*

VNTR : *“variable number of tandem repeats”*

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

Robert Frost (1920)

REMERCIEMENTS

J'aimerais d'abord remercier mon directeur de recherche, le Dr Pierre Belhumeur, pour m'avoir permis de poursuivre ces recherches dans son laboratoire sous sa supervision. En plus d'avoir contribué significativement au développement de mon esprit scientifique par ses conseils et ses critiques, il m'a influencé par son approche de la recherche et ses qualités interpersonnelles. J'ai aussi apprécié la grande liberté et la confiance qu'il m'a témoignées au cours de ces années. Je tiens également à remercier mon co-directeur de recherche, le Dr Louis Bernatchez, pour l'aide précieuse qu'il m'a offerte au cours de la rédaction des articles présentés dans cette étude. Ses commentaires, ses encouragements et son intérêt dans ce projet de recherche m'ont permis de poursuivre ce travail et donné l'envie de me dépasser. Merci à vous deux, je vous serai toujours reconnaissant.

Je voudrais aussi remercier les collègues du laboratoire que j'ai côtoyés. Merci Martin et Julie pour les discussions, les suggestions et le support que vous m'avez apporté tout au long de la durée de ces travaux. Vous avez contribué à plusieurs niveaux à l'accomplissement de cette thèse. Merci Philippe pour ton excellente compagnie, travailler avec toi était un vrai plaisir. J'adresse également mes remerciements aux membres du département de Microbiologie et Immunologie pour leur collaboration.

Un merci spécial aux membres de ma famille pour votre appui toujours renouvelé et l'intérêt que vous avez démontré pendant ces études. Finalement, merci beaucoup Julie, ce travail n'aurait pu être mené à terme sans ta contribution et ton soutien inestimable. Rose, Victor et Angélie, vous m'inspirez et me poussez au dépassement à tous les jours, merci!

Merci aux membres du jury pour la lecture et la critique de ce manuscrit.

Chapitre 1

Revue de la littérature

1.1. Le complexe majeur d'histocompatibilité

Le complexe majeur d'histocompatibilité (CMH), originalement identifié pour ses effets sur le système immunitaire lors de transplantation tissulaire, est une région du génome où l'on retrouve une concentration importante de gènes dont les produits sont impliqués dans la reconnaissance du soi et du non-soi. Les CMH qui sont les mieux caractérisés sont ceux retrouvés chez l'homme et la souris. L'organisation génétique du CMH comprend trois régions. Les deux premières régions contiennent plusieurs gènes codant pour les récepteurs glycoprotéiques de classe I et de classe II qui sont structurellement et fonctionnellement différents. Leur fonction principale est de lier et présenter à la surface cellulaire les peptides antigéniques dérivés de pathogènes afin de permettre leur reconnaissance par les récepteurs des lymphocytes T appropriés (TCR, *T cell receptors*) (Klein et Sato 2000; Janeway et al. 2001). Des gènes du CMH ont également été identifiés chez d'autres vertébrés; depuis les poissons cartilagineux (ex. les requins) en passant par les poissons osseux (*bony fish*), les amphibiens, les reptiles, les oiseaux et différents mammifères (Trowsdale 1995). Étant donné l'importance cruciale du CMH dans le fonctionnement du système immunitaire, une forte pression sélective favorise l'émergence de pathogènes possédant de nouveaux moyens pour éviter leur reconnaissance et leur élimination. Ainsi, certains gènes du CMH ont évolué en tirant profit de différentes stratégies de défense (Trowsdale et Parham 2004). Dans le même ordre d'idée, il n'est pas surprenant de constater que le CMH joue un rôle primordial dans la résistance ou la susceptibilité à plusieurs infections de même que dans la plupart des maladies auto-immunes (Blackwell et al. 2009; Caillat-Zucman 2009).

1.1.1. Les régions du CMH

Découvertes à la surface des lymphocytes, les premières molécules produites par les gènes du CMH humain ont été désignées comme antigènes lymphocytaires. Cette appellation fut conservée pour désigner le CMH humain en tant que complexe HLA, ou *Human Leukocyte Antigen complex*. Les CMH les mieux caractérisés incluent celui retrouvé sur le bras court du chromosome 6 chez l'humain

et celui retrouvé sur le chromosome 17 chez la souris, nommé complexe H-2. Chez l'humain, la première séquence du complexe HLA obtenue en 1999 couvrait $3,6 \times 10^6$ paires de bases (pb) (MHC Sequencing Consortium 1999). On a identifié dans cette région 224 loci, desquels seulement 128 seraient exprimés et où 40 % de ces derniers auraient une fonction associée au système immunitaire. Depuis, le séquençage complet du chromosome 6 humain en 2003 a permis d'étendre jusqu'à $7,6 \times 10^6$ pb cette région, qu'on appelle le CMH étendu (*extended MHC*, Mungall et al. 2003). Le complexe HLA est divisé en trois régions : la région de classe I télomérique, la région de classe II centromérique et la région de classe III centrale aux deux précédentes. L'organisation générale des gènes du CMH humain est présentée en Figure 1.1. La prochaine section décrit le contenu et les molécules produites par ces régions du complexe HLA.

1.1.1.1. La région du CMH de classe I

1.1.1.1.1. Les gènes de la région du CMH de classe I

Il y a dans la région du CMH de classe I des gènes de type classique (de classe Ia) et d'autres de type non-classique (de classe Ib). Les gènes classiques codent pour les molécules présentatrices d'antigènes. On retrouve donc dans cette région les trois gènes qui forment en association avec la β_2 -microglobuline (β_2m) les molécules présentatrices d'antigènes du CMH de classe I. Les trois gènes classiques, appelés HLA-A, HLA-B et HLA-C, codent pour la chaîne alpha (α) des molécules hétérodimériques du CMH de classe I. Le gène codant pour la chaîne bêta (β) de la molécule, la β_2m , n'est pas retrouvé à l'intérieur du complexe HLA, mais plutôt sur le chromosome 15. Les gènes non-classiques diffèrent des gènes classiques à plusieurs niveaux : leur expression est irrégulière, plusieurs sont des pseudogènes, leur degré de polymorphisme est plus faible et la nature de leur fonction est différente (Klein et O'hUigin 1994; Janeway et al. 2001).

Les gènes codant pour la molécule du CMH de classe I sont généralement très conservés dans leur structure. Le gène type du CMH de classe I est composé de huit exons et de sept introns (Figure 1.2).

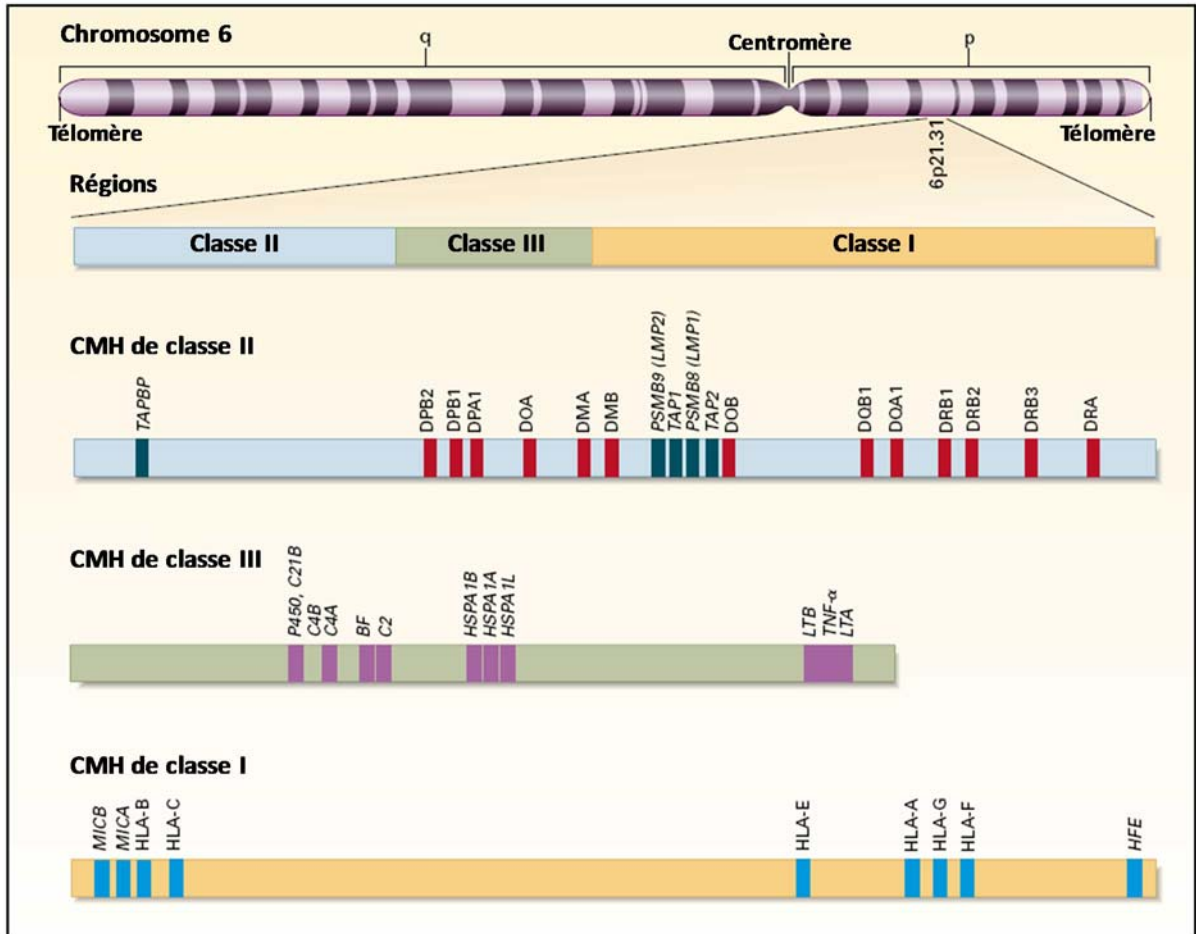


Figure 1.1. Localisation et organisation génétique du complexe majeur d'histocompatibilité chez l'humain (HLA). Le complexe HLA situé sur le bras court du chromosome 6 est divisé en trois régions: le CMH de classe II (bleu), le CMH de classe III (vert) et le CMH de classe I (jaune). Chaque région renferme plusieurs gènes (loci) et seulement quelques-uns sont indiqués (Traduit de Klein et Sato 2000).

La séquence 5'UTR et celle codant pour le peptide signal sont retrouvées dans l'exon 1. Les exons 2, 3 et 4 codent respectivement pour les domaines $\alpha 1$, $\alpha 2$ et $\alpha 3$. Le peptide de connexion ainsi que le domaine transmembranaire sont encodés par l'exon 5, alors que les exons 6 et 7 contiennent le domaine cytoplasmique. On retrouve dans le dernier exon l'extrémité C-terminale de la queue cytoplasmique ainsi que sa séquence 3'UTR. Il est à noter que le troisième intron est le plus long avec une longueur pouvant atteindre 2 kb et que tous les introns sont de phase 1; le codon est coupé entre le premier et le second nucléotide (Klein 1986).

1.1.1.1.2. La molécule présentatrice d'antigènes du CMH de classe I

La molécule du CMH de classe I est une glycoprotéine hétérodimérique retrouvée à la surface cellulaire. La chaîne α comporte cinq domaines distincts : deux domaines pour la liaison au peptide ($\alpha 1$ et $\alpha 2$), un domaine immunoglobuline-*like* ($\alpha 3$), la région transmembranaire et la queue cytoplasmique. La chaîne β de la molécule, la β_2m , consiste en un seul domaine associé de manière non-covalente au domaine $\alpha 3$ de la chaîne α (Figure 1.3). La fonction de cette molécule est de présenter des peptides antigéniques produits dans le cytoplasme aux TCR des lymphocytes T cytotoxiques CD8⁺. Les molécules du CMH de classe I sont exprimées à la surface de toutes les cellules somatiques nucléées (Klein et Sato 2000; Janeway et al. 2001).

1.1.1.2. La région du CMH de classe II

1.1.1.2.1. Les gènes de la région du CMH de classe II

La région du CMH de classe II contient les gènes classiques codant pour les chaînes α et les chaînes β de trois familles de molécules hétérodimériques du CMH de classe II appelées : HLA-DR, HLA-DP et HLA-DQ. Comme dans la région du CMH de classe I, on retrouve également des gènes non-classiques. Les gènes HLA-DM α et HLA-DM β codent pour une molécule dont la fonction est de catalyser la liaison du peptide aux molécules présentatrices d'antigènes du CMH de classe II.

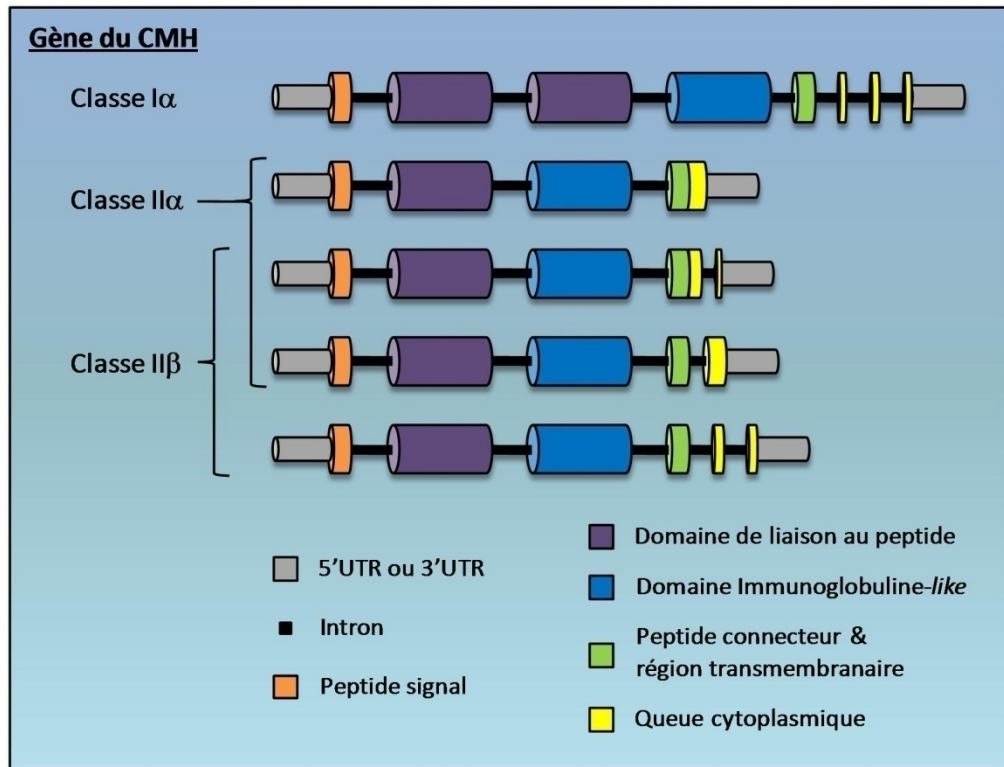


Figure 1.2. Structure des gènes du CMH retrouvés chez les mammifères. La légende indique quel exon (ou partie d'exon) code pour les domaines protéiques ou régions non-traduites. La figure n'est pas construite à l'échelle (Adapté de Dixon et al. 1995).

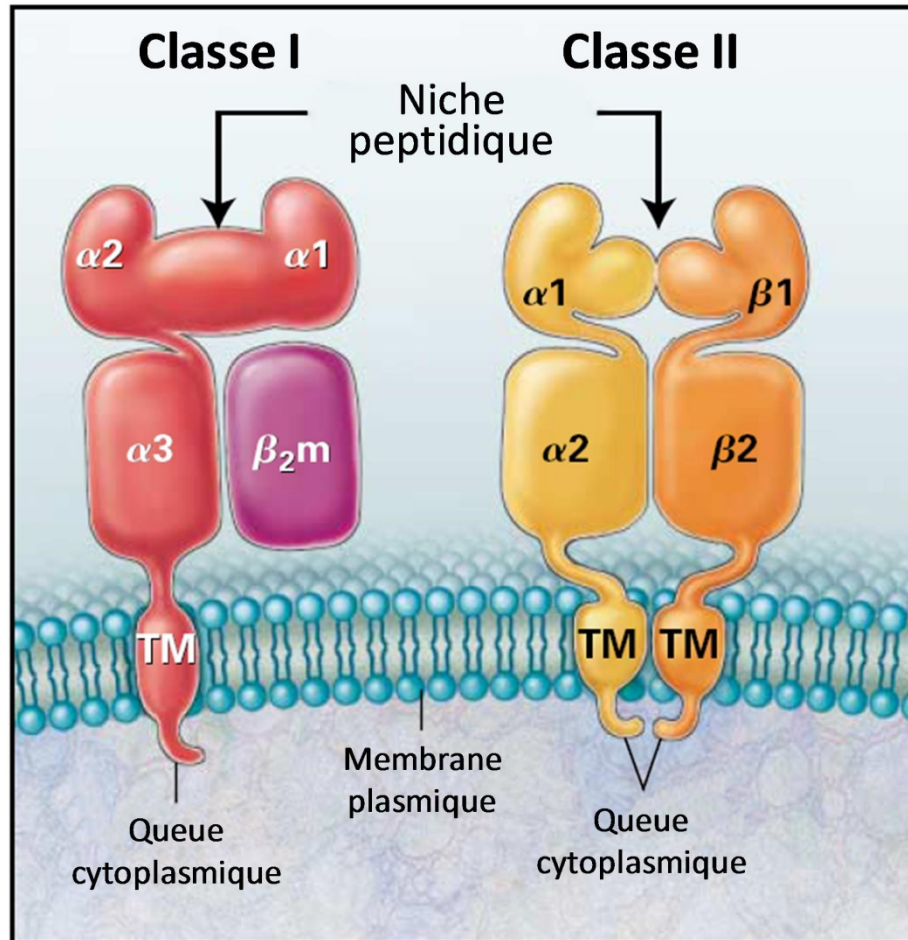


Figure 1.3. Structure des molécules du CMH de classe I et de classe II. La Bêta₂-microglobuline (β_2m) complète la molécule du CMH de classe I. La chaîne α du CMH de classe I a deux domaines de liaison au peptide (α_1 et α_2), un domaine immunoglobuline-like (α_3), une région transmembranaire (TM) et une queue cytoplasmique. Chaque chaîne du CMH de classe II a un domaine de liaison au peptide (α_1 ou β_1), un domaine immunoglobuline-like (α_2 ou β_2), une région transmembranaire et une queue cytoplasmique. La niche peptidique des molécules du CMH est formée par les deux domaines les plus éloignés de la membrane plasmique (Traduit de Klein et Sato 2000).

L'activité de cette dernière est modulée par un régulateur négatif, la molécule HLA-DO, qui est encodé par deux autres gènes non-classiques retrouvés dans cette région : HLA-DO α /HLA-DN α et HLA-DO β . De plus, on y retrouve plusieurs gènes codant pour des molécules impliquées dans l'apprêtement et la présentation antigénique par les molécules du CMH de classe I. Notons la présence des gènes pour les sous-unités du protéasome (LMP2 et LMP7), les sous-unités du transporteur de peptides TAP au réticulum endoplasmique (*transporters associated with antigen processing*, TAP1 et TAP2) et de la tapasine (TAPBP) (Janeway et al. 2001).

La structure des gènes du CMH de classe II est variable chez les mammifères. Ils peuvent être formés de quatre, cinq ou six exons séparés respectivement par trois, quatre ou cinq introns (Figure 1.2). Il y a une certaine similarité avec les gènes du CMH de classe I. La séquence 5'UTR ainsi que le peptide signal sont retrouvés sur l'exon 1. Dans le cas des gènes codant pour la chaîne β , on y retrouve en plus la séquence pour les quatre premiers acides aminés du domaine $\beta 1$. Dépendant du gène, le deuxième exon contient les séquences codantes pour les domaines $\alpha 1$ et $\beta 1$ alors que les domaines $\alpha 2$ et $\beta 2$ sont encodés par les séquences d'un troisième exon. La variabilité de structure des gènes du CMH de classe II se retrouve au niveau du peptide de connexion, du domaine transmembranaire, de la queue cytoplasmique et du domaine 3'UTR. Pour les gènes du CMH de classe II α , ces éléments sont encodés par un ou deux exons (quatrième et cinquième). Dans les situations où deux exons sont retrouvés, un intron sépare la séquence codante du reste de la molécule juste avant ou en plein cœur du domaine cytoplasmique. Ces structures sont également retrouvées pour les gènes du CMH de classe II β . Pour ces gènes, il y a en plus une structure à trois exons où le peptide de connexion et le domaine transmembranaire sont encodés par un quatrième exon, une partie du domaine cytoplasmique par un cinquième exon et la fin de ce domaine et la séquence 3'UTR par un sixième exon. Pour les gènes du CMH de classe II α , le premier intron est le plus long avec généralement plus de 1 kb, alors que pour les gènes du CMH de classe II β les premier et deuxième introns sont les plus longs et leur longueur

peut varier de 1 à 5 kb. Finalement, les introns des gènes du CMH de classe II sont également de phase 1 (Klein 1986).

1.1.1.2.2. La molécule présentatrice d'antigènes du CMH de classe II

Les molécules du CMH de classe II sont des glycoprotéines hétérodimériques, constituées d'une chaîne α et d'une chaîne β , retrouvées à la surface cellulaire. Chaque chaîne du CMH de classe II α et de classe II β est composée de quatre domaines : le domaine de liaison au peptide ($\alpha 1$ ou $\beta 1$), le domaine immunoglobuline-*like* ($\alpha 2$ ou $\beta 2$), la région transmembranaire et la queue cytoplasmique (Figure 1.3). La fonction des molécules hétérodimériques du CMH de classe II est de présenter des peptides antigéniques, provenant principalement d'antigènes exogènes, aux TCR des lymphocytes T auxiliaire CD4⁺. Leur reconnaissance provoque l'activation et la prolifération des lymphocytes T CD4⁺ spécifiques, initiant ainsi une réponse immunitaire dirigée contre l'antigène à partir duquel était dérivé le peptide antigénique. Les molécules du CMH de classe II ont un profil d'expression beaucoup plus limité que celles du CMH de classe I. Elles sont exprimées constitutivement à la surface des cellules spécialisées dans la capture et la présentation d'antigènes extracellulaires. Ces dernières sont appelées cellules présentatrices d'antigènes (APC, *antigen-presenting cells*) et incluent les cellules dendritiques (DC, *dendritic cell*), les lymphocytes B et les macrophages (M Φ). L'expression est également retrouvée à la surface et les cellules épithéliales thymiques (TEC, *thymic epithelial cells*) et des lymphocytes T activés humains. De plus, comme nous le verrons dans la section suivante, l'expression du CMH de classe II peut être modulée par différents stimuli, notamment l'INF- γ , dans plusieurs types cellulaires (Klein et Sato 2000; Janeway et al. 2001).

1.1.1.3. La région du CMH de classe III

En position centrale par rapport aux deux régions précédentes, on retrouve la région du CMH de classe III. On peut noter dans cette région la présence de certains gènes codants pour des protéines

sécrétées aux fonctions immunitaires tels que des composantes du complément (ex : C2, C4 et le facteur B) ou des cytokines (TNF- α et lymphotoxines) (Janeway et al. 2001).

1.1.2. Le contrôle de l'expression des gènes du CMH de classe II

Les molécules du CMH de classe II ont un rôle primordial à jouer dans l'initiation de la réponse immunitaire adaptative en présentant les peptides antigéniques pouvant induire l'activation des lymphocytes T auxiliaire CD4⁺. Cependant, l'expression des gènes codant pour ces molécules est restreinte à certains types cellulaires et peut être modulée quantitativement. Le contrôle de l'expression des gènes du CMH de classe II est principalement transcriptionnel et repose sur l'interaction de plusieurs protéines au niveau du promoteur des gènes. Le plus important de ces facteurs est le transactivateur des gènes de classe II (CIITA, *class II transactivator*) encodé par le gène MHC2TA. La vaste majorité des signaux qui activent ou répriment l'expression des gènes du CMH de classe II convergent vers un ou plusieurs des trois promoteurs alternatifs qui contrôlent l'expression du CIITA (Revue par Ting et Trowsdale 2002; LeibundGut-Landmann et al. 2004).

1.1.2.1. Le promoteur des gènes du CMH de classe II

L'expression des gènes du CMH de classe II est régulée principalement au niveau de la transcription. Un promoteur spécifique aux gènes du CMH de classe II est caractérisé par la présence de quatre motifs appelées les boîtes S (ou W), X, X2 et Y (Figure 1.4A). Il y a de fortes contraintes au niveau de l'ordre, de l'orientation et de l'espacement de ces boîtes qu'on appelle aussi le module S(W)-X-X2-Y. Un promoteur ayant ces caractéristiques est retrouvé sur les gènes codant pour les chaînes α et les chaînes β chez tous les vertébrés étudiés. On le retrouve également sur le gène du CMH de classe I, de la chaîne invariante (Ii), de même que sur les gènes non-classiques du CMH de classe II HLA-DM et HLA-DO. Les boîtes du promoteur des gènes du CMH de classe II lient plusieurs facteurs de transcription ubiquitaires. La boîte X permet la liaison du complexe hétérotrimérique RFX

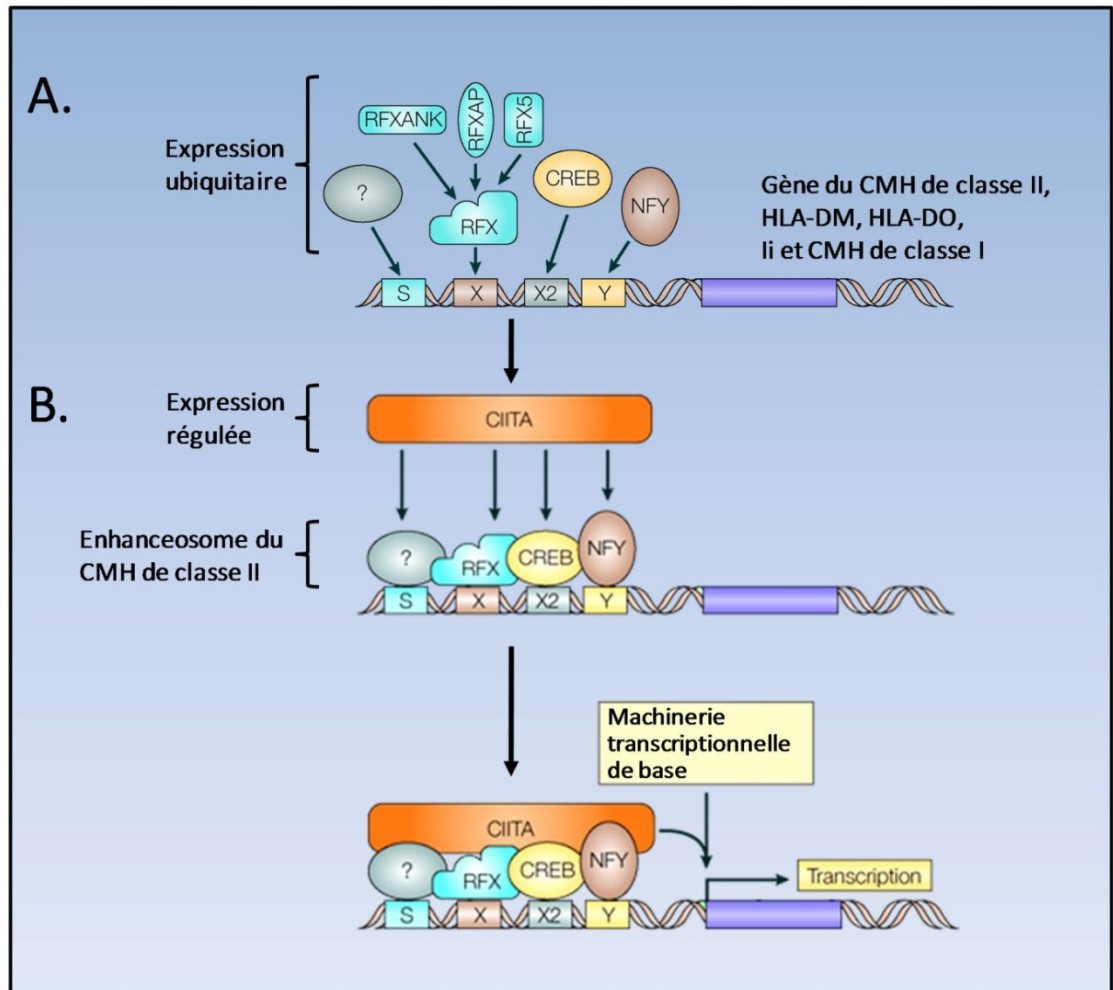


Figure 1.4. Régulation de la transcription des gènes du CMH de classe II. **A.** Liaison des facteurs de transcription au promoteur des gènes du CMH de classe II. Le promoteur des gènes du CMH de classe II ainsi que celui du gène HLA-DM, HLA-DO, de la chaîne invariante (Ii) et du CMH de classe I contiennent le module S(W)-X-X2-Y. Les facteurs de transcription ubiquitaires RFX, CREB, NF-Y et un autre non-identifié se lient aux différents motifs du promoteur pour former l'enhanceosome du CMH de classe II. **B.** Liaison du CIITA à l'enhanceosome du CMH de classe II. La liaison du cofacteur CIITA aux gènes du CMH de classe II s'effectue au niveau de l'enhanceosome par des interactions protéines-protéines, sans liaison directe avec l'ADN. Le CIITA coordonne ensuite le recrutement de la machinerie transcriptionnelle de base pour initier la transcription des gènes (Adapté de Reith et al. 2005).

(*regulatory factor X*) composé des protéines RFX5, RFXAP et RFXANK. La boîte X2 est reconnue par le facteur CREB (*cyclic-AMP-responsive-element-binding-protein*) alors que la boîte Y lie le complexe hétérotrimérique NF-Y (*nuclear transcription factor Y*) composé de NF-YA, NF-YB et NF-YC. Plusieurs facteurs ont montré une interaction avec la boîte S (ou W) *in vitro*, mais aucun n'a pu être démontré comme étant fonctionnel *in vivo*. Ce complexe multiprotéique, appelé l'enhanceosome du CMH de classe II constitue le plateau de liaison pour le CIITA (Ting et Trowsdale 2002; LeibundGut-Landmann et al. 2004).

1.1.2.2. Le transactivateur des gènes de classe II

1.1.2.2.1. Le CIITA : sa structure et son mode d'action

Le CIITA ne se lie pas directement avec l'ADN mais démontre plusieurs interactions protéines-protéines avec différents partenaires (Figure 1.4B). L'extrémité N-terminale de la protéine renferme une région riche en acides aminés acides formant un domaine d'activation de la transcription. Ce domaine permet l'interaction avec des composantes de la machinerie transcriptionnelle de base et autres coactivateurs, ainsi qu'avec des protéines impliquées dans le remodelage de la chromatine. Les deux-tiers de la partie C-terminale du CIITA renferme des motifs riches en leucines (LRR, *leucine-rich repeats*). Cette extrémité de la protéine est responsable de l'association avec elle-même, de sa localisation nucléaire et des interactions avec les partenaires de l'enhanceosome. De plus, l'activité transcriptionnelle du CIITA est modifiée par la phosphorylation et l'ubiquitylation de certains acides aminés (Tosi et al. 2002; Greer et al. 2003; 2004). Contrairement aux composantes de l'enhanceosome qui sont exprimées ubiquitairement, l'expression du CIITA est fortement régulée par quatre promoteurs distincts qui contrôlent l'expression du gène MHC2TA. Ultiment, c'est l'expression du CIITA qui dictera dans quelles cellules et à quels niveaux seront exprimés les gènes du CMH de classe II.

1.1.2.2.2. Les régions régulatrices du gène MHC2TA

La transcription du gène codant pour le CIITA est contrôlée par une large région régulatrice qui renferme quatre promoteurs différents (pI à pIV) (Muhlethaler-Mottet et al. 1997) (Figure 1.5). L'expression du gène dans un type cellulaire spécifique, de manière inductible par les cytokines et selon son stade de différenciation, repose sur l'activité différentielle au niveau des trois promoteurs pI, pIII et pIV, l'activité du promoteur pII étant moins bien définie. L'utilisation alternative de ces promoteurs permet l'incorporation d'un premier exon variable menant à la production de protéines de poids moléculaire différents : 121 kDa (pI, pIII et pIV), 124 kDa (pIII) et 132 kDa (pI). Le promoteur pI est actif dans les cellules présentatrices d'antigènes d'origine myéloïde, comme les cellules dendritiques conventionnelles et les macrophages. Le promoteur III est plutôt utilisé dans les cellules d'origine lymphoïde telles que les lymphocytes B, les lymphocytes T activés humains et les cellules dendritiques plasmacytoïdes. Le promoteur IV est responsable de l'expression constitutive dans les cellules épithéliales thymiques et inductible par l'INF- γ dans les cellules qui ne sont pas d'origine hématopoïétique (épithéliales, endothéliales, fibroblastes, astrocytes). C'est la présence des motifs GAS (*INF- γ activated site*), boîte E (*E-box*) et IRF-E (*INF-regulatory factor element*) dans ce promoteur qui permet l'induction de ce dernier par l'INF- γ (Muhlethaler-Mottet et al. 1997). Il a été démontré que l'activité transcriptionnelle de l'isoforme de 132 kDa était supérieure à celle de l'isoforme de 124 kDa pour l'induction d'un gène du CMH de classe II (Nickerson et al. 2001).

1.1.2.2.3. La modulation de l'expression du CIITA

Il a été clairement établi que le profil d'expression du CIITA corrèle avec celui des gènes du CMH de classe II. En plus de l'expression constitutive du CIITA dans les cellules ci-haut mentionnées ou induite par l'INF- γ dans la plupart des cellules, l'expression de ce dernier peut être réprimée par différents mécanismes (Figure 1.6). L'expression du CIITA induite par l'INF- γ peut être inhibée par le TGF- β dans des lignées cellulaires d'astrogliomes et de fibrosarcomes humains (Lee et al. 1997), par

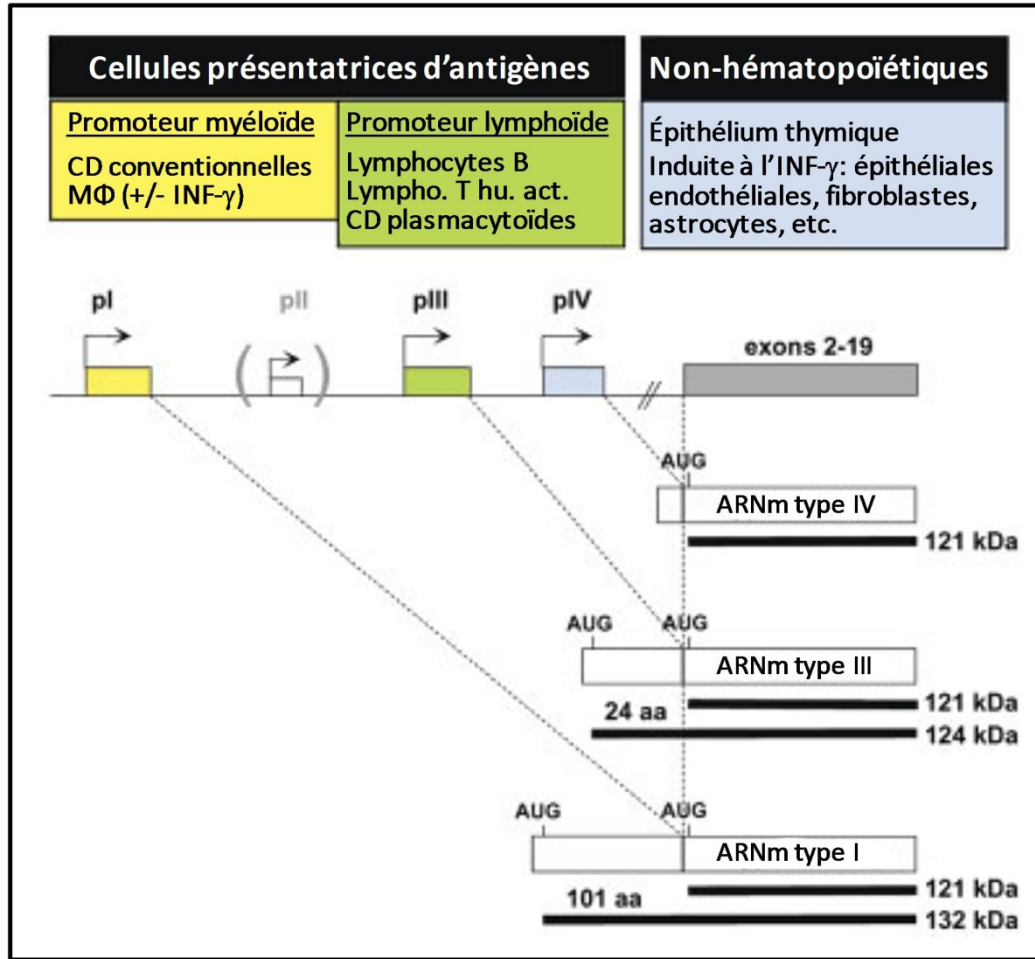


Figure 1.5. Région régulatrice du gène CIITA. Trois promoteurs indépendants contrôlent l'expression du CIITA (pI, pIII et pIV). pI est actif dans les cellules présentatrices d'antigènes d'origine myéloïdes et pIII dans celles d'origine lymphoïdes. pIV est essentiel pour l'induction à l'INF- γ dans les cellules qui ne sont pas d'origine hématopoïétiques et pour l'expression constitutive dans les cellules épithéliales thymiques. CD, cellule dendritique; M Φ , macrophages; hu., humain; act., activés. Trois types d'ARN messagers du CIITA (bandes blanches) sont initiés au pI, pIII et pIV pour encoder trois isoformes de la protéine (bandes noires) (Traduit de LeibundGut-Landmann et al. 2004).

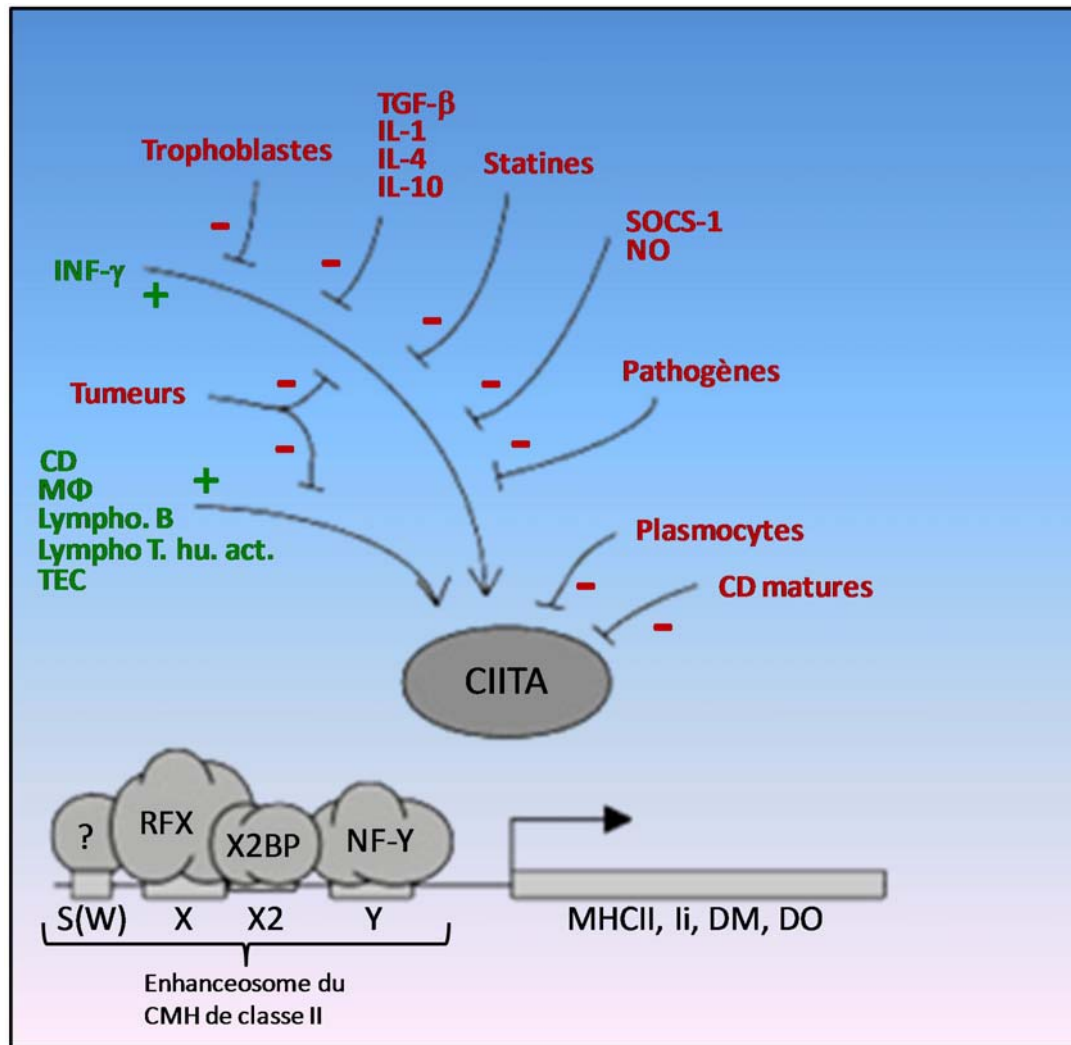


Figure 1.6. Activation et inhibition de l'expression du gène CIITA. Les effets positifs ou négatifs sur la transcription sont indiqués en rouge et vert respectivement. Les CD (cellules dendritiques), MΦ (macrophages), Lympho. B (lymphocytes B), Lympho T hu. (humain) Act. (activé) et TEC expriment le CIITA constitutivement. L'induction de l'expression par l'INF-γ peut être inhibée par plusieurs facteurs. L'expression du CIITA est réprimée lors de la différenciation terminale des lymphocytes B en plasmocytes et des CD immatures en CD matures. L'expression constitutive du CIITA ou induite par l'INF-γ est fréquemment inhibée par des tumeurs de différentes origines (Traduit de LeibundGut-Landmann et al. 2004).

l'IL-1 dans des astrocytes humains (Rohn et al. 1999), par l'IL-4 et l'IL-10 sur des cellules de la microglie chez la souris (O'Keefe et al. 1999), par les statines dans une lignée primaire de cellules endothéliales humaines (Kwak et al. 2000), par la protéine SOCS-1 qui est un régulateur négatif de la voie d'activation par l'INF- γ (O'Keefe et al. 2001) et par l'oxyde nitrique (NO, *nitric oxide*) produit lors de l'activation des macrophages (Kielar et al. 2000). Plusieurs pathogènes ont également développé des mécanismes pour inhiber l'expression du CIITA, principalement en interférant avec la voie d'induction par l'INF- γ (Pai et al. 2003; Abendroth et al. 2000). Des mécanismes de modification de la chromatine sont aussi impliqués dans la répression du CIITA. La région régulatrice du gène est retrouvée hyperméthylée au niveau des dinucléotides CpG dans plusieurs types de tumeurs (van den Elsen et al. 2003) et dans les cellules trophoblastiques fœtales (van den Elsen et al. 2000), ce qui les rend également insensibles à l'induction par l'INF- γ . De plus, il a été démontré que l'inhibition de la synthèse des gènes de classe II lors de la maturation des cellules dendritiques est causée par la déacétylation des histones sur un large domaine couvrant la région régulatrice du CIITA (Landmann et al. 2001). Finalement, un mécanisme impliquant le facteur PRDI-BF1/Blimp-1 (*human positive regulatory domain 1 binding factor-1/B-lymphocyte-induced maturation protein-1*) au niveau du promoteur pIII du CIITA a été proposé pour expliquer son inhibition lors de la différenciation des lymphocytes B en cellules plasmiques (Piskurich et al. 2000).

1.1.3. La présentation antigénique

La présentation antigénique par les molécules du CMH est un élément crucial dans l'initiation de la réponse immunitaire permettant la reconnaissance de peptides antigéniques dérivés de pathogènes viraux, bactériens ou parasitaires. Cette conversion de protéines endogènes ou exogènes en peptides reconnus par les récepteurs des lymphocytes T appropriés nécessite une série d'évènements qui mènent à la formation d'un complexe immun formé d'un peptide antigénique et d'une molécule du CMH de classe I ou de classe II. La molécule du CMH de classe I présente à la surface de toutes les cellules

nucléées pourra, par exemple, présenter un peptide dérivé d'une protéine virale aux TCR des lymphocytes T CD8⁺ cytotoxiques, ce qui mènera à la destruction de la cellule présentatrice infectée. De plus, des études récentes indiquent que les lymphocytes T CD8⁺ cytotoxiques sont également impliqués dans la réponse immunitaire face à plusieurs bactéries intracellulaires par les mécanismes de présentation croisée (*cross presentation*) et d'amorçage croisé (*cross priming*) (Kaufmann et Schaible 2005). La présentation d'antigènes exogènes, d'origine bactérienne ou de tumeurs, par la molécule du CMH de classe II aux TCR des lymphocytes T CD4⁺ auxiliaires aura pour conséquence leur prolifération et différenciation afin d'induire une immunité à médiation cellulaire ou humorale. La présentation antigénique par les molécules du CMH de classe II est essentielle à plusieurs autres niveaux. Elle permet également le processus de sélection positive et négative qui forme la spécificité du répertoire des TCR de la population de lymphocytes T CD4⁺ lors de leur développement dans le thymus. Elle est finalement impliquée dans l'homéostasie de la population de lymphocytes T CD4⁺ matures en périphérie, la maintenance de la tolérance au soi de même que dans le bris de tolérance dans les maladies auto-immunes. Dans la prochaine section nous aborderons plus en détails les mécanismes de présentation antigénique pour les deux classes de molécules du CMH. Dans un premier temps, nous allons nous intéresser aux caractéristiques des cellules présentatrices d'antigènes (Klein et Sato 2000; LeibundGut-Landmann et al. 2004).

1.1.3.1. Les cellules présentatrices d'antigènes

L'apprêtement et la présentation des peptides antigéniques aux lymphocytes T CD4⁺ auxiliaires par différents types cellulaires exprimant le CMH de classe II se font avec une efficacité variable (Schneider et Sercarz 1997). Ceci a mené à l'appellation "professionnel" pour un certain groupe de cellules présentatrices d'antigènes (APC) qui inclue les macrophages (MΦ), les cellules dendritiques (CD) et les lymphocytes B. L'autre groupe de cellules dites "non-professionnelles" exprime généralement un niveau plus faible ou variable des molécules du CMH de classe II, des

molécules co-stimulatrices et des molécules accessoires. Ceci mène à des capacités hétérogènes d'apprêtement et de présentation antigénique qui conduisent ensuite à une présentation plus restreinte de peptides (Muntasell et al. 2004).

Le macrophage a une capacité extraordinaire pour l'endocytose, ce qui lui permet d'internaliser jusqu'à 50 % de sa surface cellulaire en un seul évènement de phagocytose (Steinman et al. 1983). Il peut internaliser tout types d'antigènes, soluble ou cellulaire, de manière non-spécifique ou par des récepteurs spécifiques. Le MΦ exprime les molécules du CMH de classe I, du CMH de classe II ainsi que les molécules co-stimulatrices B7 (CD80 et CD86) dont il augmente les niveaux d'expression lorsqu'il est activé par des cytokines inflammatoires ou des produits bactériens. Comparativement aux cellules dendritiques et aux lymphocytes B, les niveaux d'expression de ces molécules demeurent cependant plus bas, ce qui expliquerait pourquoi les MΦ sont généralement considérés comme des APC moins efficaces (Mellman et al. 1998). Ils sont néanmoins présents en grand nombre aux sites d'infection et d'inflammation chronique pour y présenter les peptides antigéniques.

Contrairement aux MΦ et aux lymphocytes B, la fonction primaire des cellules dendritiques semble être la présentation antigénique. Tout comme les MΦ, les CD complètent leur différenciation en quittant la circulation sanguine pour s'établir en périphérie. Les CD sont aussi très efficaces au niveau de l'endocytose et peuvent présenter les peptides antigéniques sur les molécules du CMH de classe I ou de classe II, qu'elles expriment d'ailleurs à un niveau très élevé lorsqu'elles atteignent leur maturité. Elles ont la propriété de pouvoir transporter un antigène capturé en périphérie vers les organes lymphoïdes secondaires, ou de l'internaliser directement de la lymphe (Itano et al. 2003). Dans les organes lymphoïdes, on les retrouve plus nombreuses dans des régions où les lymphocytes T naïfs sont activés; dans une position qui permet un flot continu de lymphocytes T à la recherche du complexe peptide-CMH apparenté (Itano et Jenkins 2003). Elles sont significativement plus efficaces que les MΦ pour stimuler les lymphocytes T (Mellman et al. 1998) et pour présenter des peptides

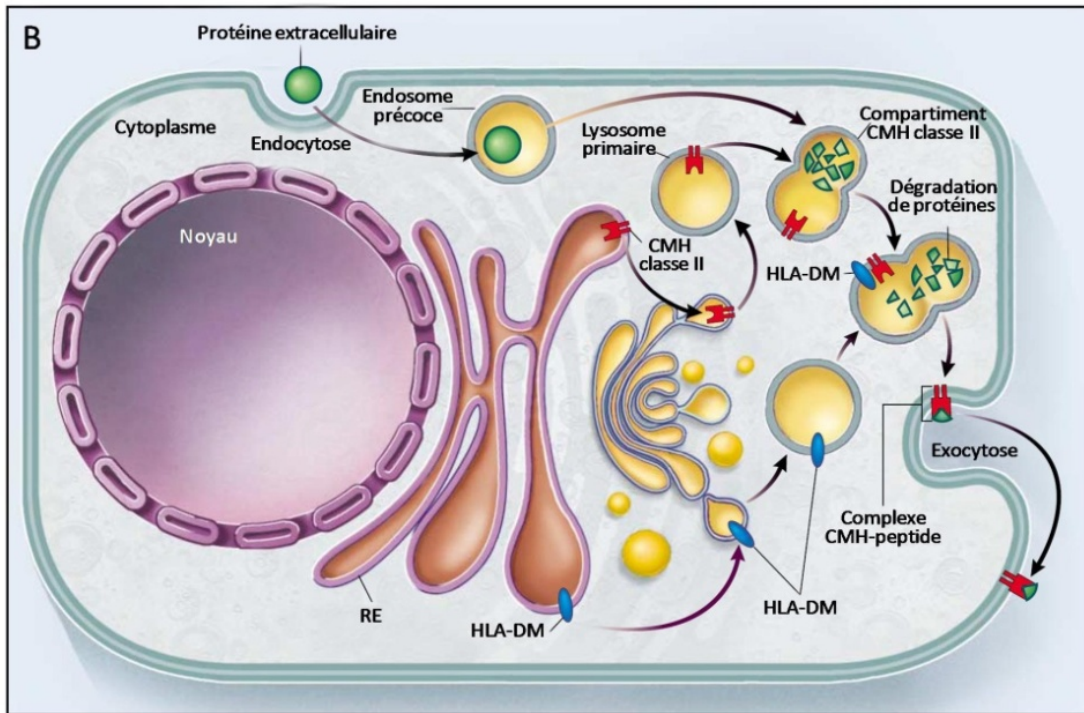
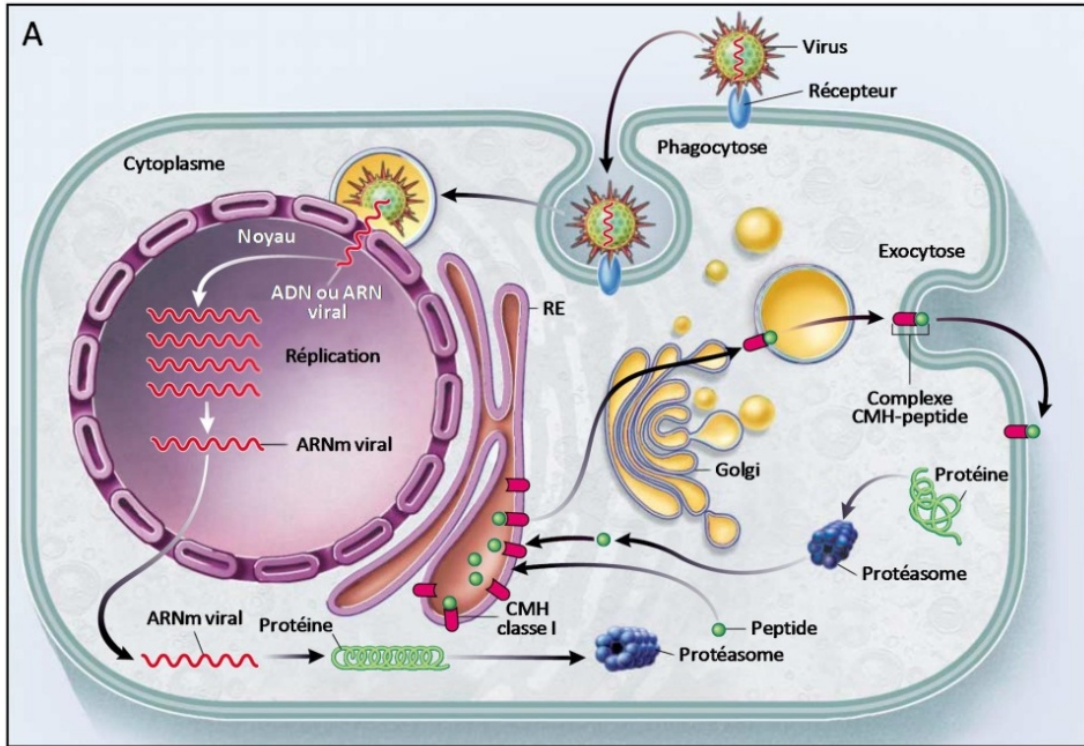
exogènes sur les molécules du CMH de classe I par le mécanisme de présentation croisée (*cross presentation*) (Delamarre et al. 2003).

Les lymphocytes B ne sont pas très efficaces pour internaliser des antigènes, à l'exception de ceux reconnus par les récepteurs immunoglobulines spécifiques à la surface. De cette manière, leur activité se concentre sur la présentation d'antigènes capables de stimuler les lymphocytes T CD4⁺ auxiliaires qui vont en retour les activer pour la sécrétion d'anticorps. Les lymphocytes B expriment constitutivement un niveau élevé de la molécule du CMH de classe II ainsi que les molécules HLA-DM et HLA-DO qui contribuent au chargement du peptide sur la molécule du CMH de classe II. Malgré leur faible capacité pour l'endocytose, la liaison d'antigènes aux immunoglobulines induit l'internalisation des récepteurs et accélère la livraison aux compartiments du CMH de classe II (MIIC, *MHC class II compartment*) où ils sont dégradés (West et al. 1994).

1.1.3.2. Les mécanismes de présentation antigénique

Le caractère distinctif des mécanismes de présentation antigénique du CMH de classe I et du CMH de classe II a depuis longtemps été lié à la source des peptides. Les ligands du CMH de classe I sont généralement dérivés de protéines cytosoliques et ceux du CMH de classe II proviennent d'une source exogénique. On sait maintenant que l'inverse est également possible. La présentation par la molécule du CMH de classe I d'antigènes provenant de bactéries présentes dans le phagosome donne lieu à deux voies différentes : la présentation croisée (*cross presentation*) (Pfeifer et al. 1993; Canaday et al. 1999) et l'amorçage croisé (*cross priming*) (Yrlid et Wick 2000; Schaible et al. 2003). La présentation d'antigènes intracellulaires par la molécule du CMH de classe II a également été rapportée (Rudensky et al. 1991; Dongre et al. 2001). La Figure 1.7 illustre les deux voies classiques pour la présentation antigénique (Klein et Sato 2000).

Figure 1.7. Voies de présentation antigénique. A. Voie de présentation par les molécules du CMH de classe I. Les protéines cytoplasmiques, du soi ou du non-soi, sont dégradées en peptides par le protéasome. Ceux-ci sont ensuite chargés sur la molécule du CMH de classe I au niveau du réticulum endoplasmique (RE) pour former un complexe peptide-CMH classe I. Ce dernier passe par le Golgi pour y être glycosylé avant d'être dirigé à la surface pour présentation aux lymphocytes T CD8⁺ cytotoxiques. **B.** Voie de présentation par les molécules du CMH de classe II. Une protéine exogène est endocytée (ou phagocytée) et dirigée par la voie endosomale vers le compartiment du CMH de classe II pour y être dégradée en peptides. La molécule du CMH de classe II, glycosylée au niveau du Golgi, se retrouve également dans ce compartiment où le chargement du peptide est catalysé par la molécule HLA-DM. Le complexe peptide-CMH classe II est ensuite dirigé à la surface cellulaire pour présentation aux lymphocytes T CD4⁺ auxiliaires (Traduit de Klein et Sato 2000).



1.1.3.2.1. La présentation antigénique par le CMH de classe I

Dans la cellule, il y a un renouvellement constant des protéines au cours duquel elles sont fragmentées en courts fragments peptidiques par un complexe protéolytique présent dans le cytoplasme et qui est appelé le protéasome (Glickman et Ciechanover 2002). La présentation antigénique par le CMH de classe I débute justement à ce niveau (Figure 1.7A) (Pamer et Cresswell 1998). L'élimination des protéines du soi défectueuses et leur recyclage constitue la fonction première de ce système. Les protéines retrouvées dans le cytoplasme peuvent également être d'origine virale, bactérienne ou fongique suite à une infection. Dans le cas d'une infection virale par exemple, l'ADN ou l'ARN du virus sera répliqué dans le noyau pour ensuite être traduit en protéines dans le cytoplasme. Les fragments peptidiques générés par le protéasome sont ensuite transportés dans la lumière du réticulum endoplasmique (RE) par le transporteur de peptides TAP formé des protéines TAP1 et TAP2 (Androlewicz et al. 1993). Dans le RE, les peptides sont chargés sur la molécule du CMH de classe I nouvellement synthétisée avec l'aide des chaperonnes moléculaires calnexine et calreticuline ainsi que les protéines ERp57 et tapasine. Le complexe peptide-CMH de classe I est finalement transporté au travers du Golgi vers la surface cellulaire pour être présenté aux TCR des lymphocytes T DC8⁺ cytotoxiques (Groothuis et al. 2005).

1.1.3.2.2. La présentation antigénique par le CMH de classe II

Les antigènes extracellulaires présentés par la voie du CMH de classe II sont d'abord internalisés par la cellule présentatrice d'antigènes par les mécanismes de phagocytose ou d'endocytose (Figure 1.7B). Ils sont acheminés dans la voie endosomale où ils sont dégradés par une série de protéases de la famille des cathepsines, ce qui permet de générer différents peptides antigéniques (Villadangos et al. 1999). De leur côté, les chaînes α et β du CMH de classe II sont produites au niveau du RE où elles sont assemblées avec la chaîne invariante (Ii) dans un complexe nanomérique $(\alpha\beta Ii)_3$ orienté vers la lumière du RE. Le complexe, associé aux membranes de

vésicules, est ensuite acheminé au travers du Golgi vers une région du cytoplasme pour rencontrer son cargo qui est contenu dans les endosomes tardifs. Les deux types de vésicules fusionnent pour former ce qui est appelé le compartiment du CMH de classe II (*MHC class II compartment*) (Stern et al. 2006). Les peptides antigéniques sont alors chargés sur la molécule du CMH de classe II avec l'aide de la chaperonne HLA-DM, avant d'être transportés à la membrane plasmique pour présentation aux TCR des lymphocytes T CD4⁺ auxiliaires.

1.1.4. La structure de la molécule du CMH de classe II

Plus d'une vingtaine de structures tridimensionnelles de la molécule du CMH de classe II humaine ont été obtenues par cristallographie depuis la toute première publiée par Brown et collaborateurs (1993) pour l'allèle *HLA-DRI* (Stern et al. 1994; Dessen et al. 1997; Smith et al. 1998). En comparant ces structures tridimensionnelles à celles connues pour les molécules du CMH de classe I, on se rend compte que malgré les différences au niveau de leur fonction et composition en sous unités, l'aspect général des deux types de molécules est conservé (Batalia et Collins 1997). La molécule du CMH de classe II est formée par deux chaînes, α et β , traversant la membrane plasmique et qui sont assemblées sans liaison covalente. Les domaines en position N-terminale de chaque chaîne, $\alpha 1$ et $\beta 1$, sont repliés de manière à ce que chacun forme la moitié de la niche peptidique située à l'extrémité de la molécule la plus distale de la membrane cellulaire (Figure 1.8A). Ces domaines forment une structure qui consiste en deux hélices- α segmentées qui reposent sur une base de huit feuillets- β antiparallèles. Le repliement de ces domaines crée entre les deux hélices- α un long sillon, la niche peptidique, qui constitue le site de liaison du peptide à la molécule du CMH de classe II (Figure 1.8B). Les deux autres domaines, $\alpha 2$ et $\beta 2$, sont situés en position proximale de la membrane et adoptent une structure en sandwich- β qui est homologue à celles de la superfamille des immunoglobulines. La molécule est finalement ancrée à la membrane plasmique par la région suivante

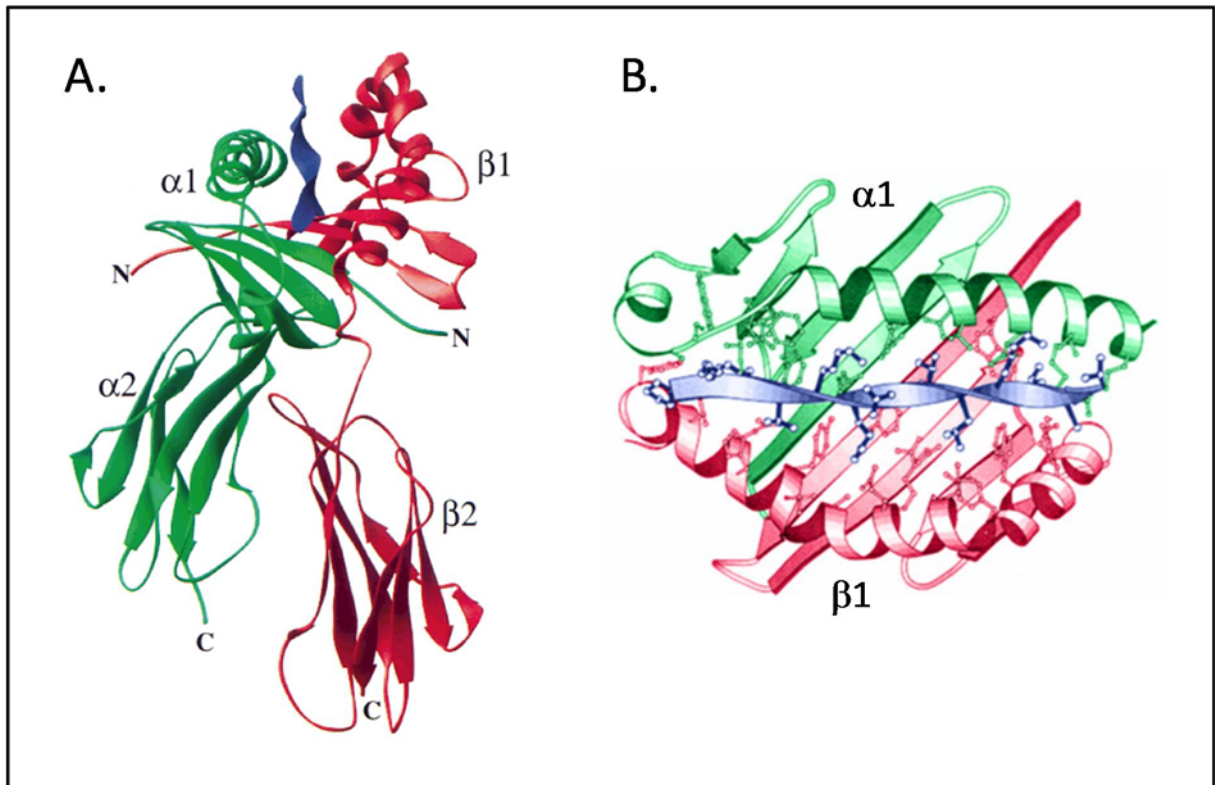


Figure 1.8. Structure tridimensionnelle de la molécule du CMH de classe II, HLA-DR1. **A.** Les chaînes α (vert) et β (rouge) s'associent au peptide (bleu) pour former le complexe ternaire. Les domaines à l'extrémité N-terminale de chaque chaîne ($\alpha 1$ et $\beta 1$) forment la niche peptidique. Les domaines $\alpha 2$ et $\beta 2$ se replient en domaines immunoglobulines. **B.** Vue en plongée de la niche peptidique, les domaines $\alpha 1$ (vert), $\beta 1$ (rouge) et le peptide (bleu) sont représentés. Les chaînes latérales des acides aminés du peptide ainsi que celles de la molécule du CMH de classe II en contact avec le peptide sont illustrées (Adapté de Batalia et Collins 1997, Stern et al. 1994).

constituée de 25 acides aminés, pour laisser ensuite une courte queue cytoplasmique formée de 8 à 15 acides aminés en position C-terminale des chaînes.

1.1.4.1 L'interaction du peptide avec la molécule du CMH de classe II

Contrairement à la molécule du CMH de classe I qui présente des peptides qui ont en général 9 acides aminés de longueur, celle des peptides présentés par la molécule du CMH de classe II peut varier considérablement, entre 11 et 17 acides aminés (Rammensee et al. 1995). L'explication de cette différence repose sur la position des deux hélices- α ainsi que sur la présence d'acides aminés avec des plus petites chaînes latérales qui laissent ouvertes les deux extrémités de la niche peptidique (Brown et al. 1993). L'obtention de la structure tridimensionnelle de l'allèle HLA-DR1 de la molécule du CMH de classe II complexé au peptide antigénique de l'hémagglutinine (HA) du virus de l'influenza a permis de montrer la nature des interactions entre le peptide et la niche peptidique (Stern et al. 1994). Le peptide est retrouvé dans une conformation linéaire étendue pour les positions P1 à P9. Ceci lui permet de diriger certaines chaînes latérales vers le fond de la niche peptidique pour stabiliser l'interaction avec la molécule du CMH de classe II alors que d'autres sont dirigées vers la surface du complexe vers le récepteur des lymphocytes T (Figure 1.9A, B). Des interactions à deux niveaux stabilisent le peptide dans la niche peptidique. Des ponts hydrogènes sont formés entre les atomes de la chaîne principale du peptide et d'autres retrouvés à des positions conservées sur les hélices- α et les feuillettes- β formant la niche peptidique afin de stabiliser l'interaction indépendamment de la séquence du peptide (Figure 1.9C). Les acides aminés aux positions P1, P4, P6 et P9 du peptide sont appelés résidus d'ancrage (*anchor residues*) parce que l'interaction de leurs chaînes latérales avec des poches (*pockets*) distinctives au fond de la niche peptidique stabilise davantage le complexe. La plupart des positions polymorphiques sont orientées vers l'intérieur de la niche peptidique pour interagir avec le peptide. La Figure 1.10A illustre différentes poches retrouvées au fond de la niche peptidique de la molécule HLA-DR1. Les chaînes latérales du peptide dirigées vers le fond de la niche peptidique

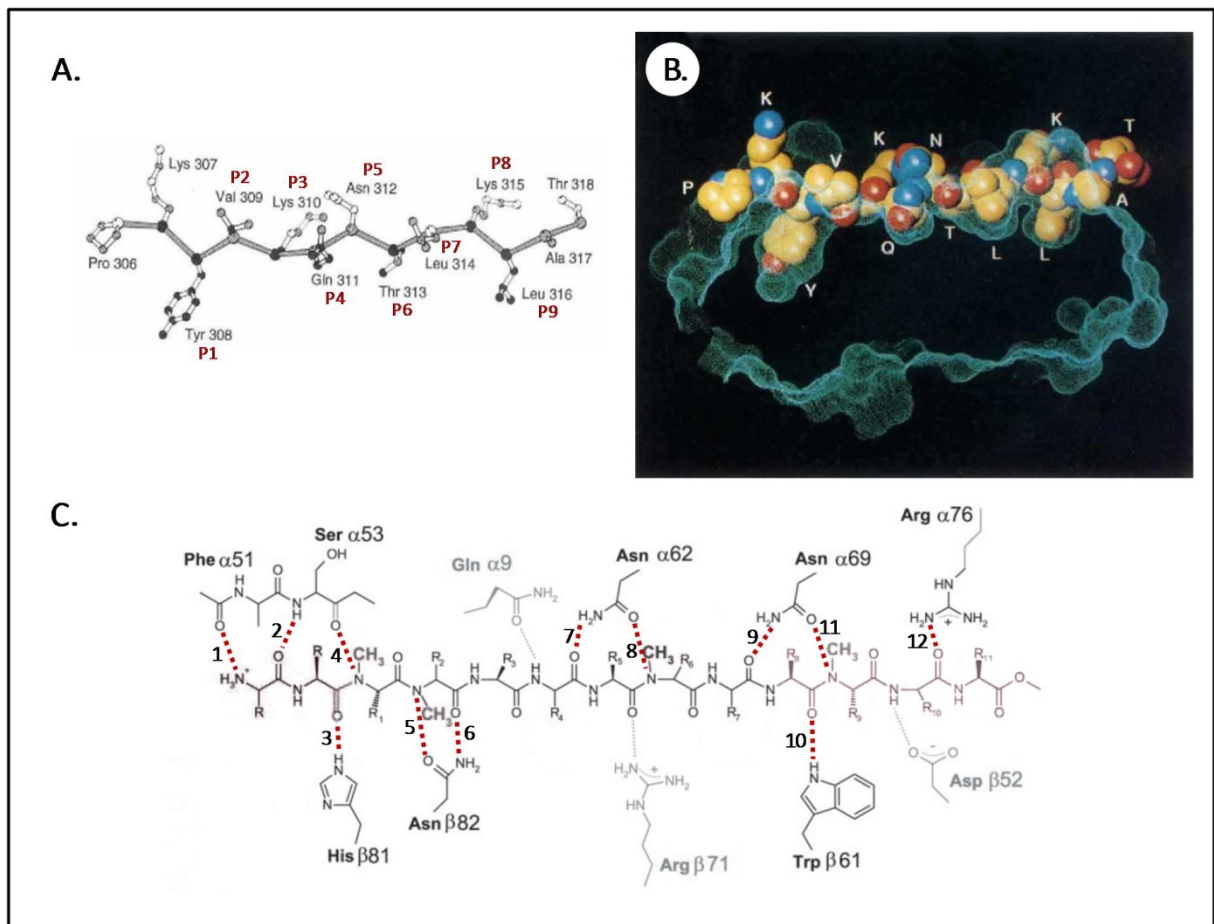
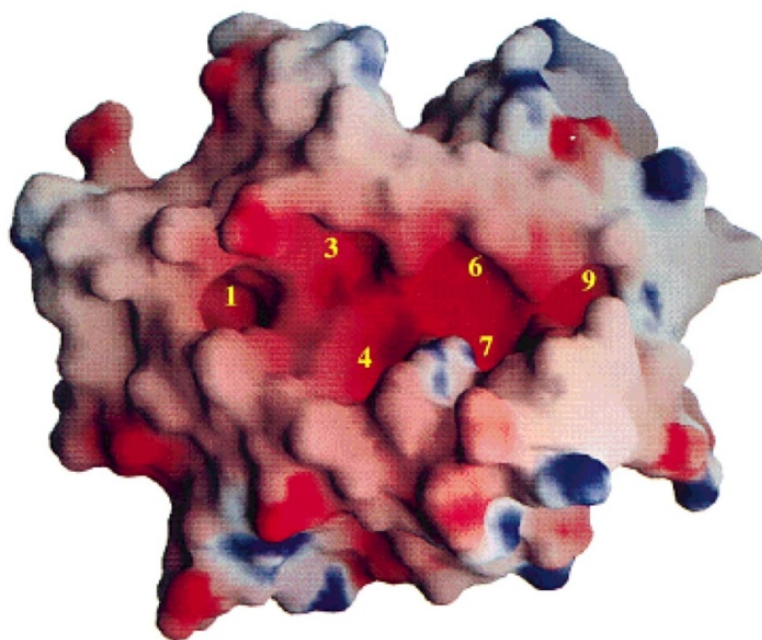


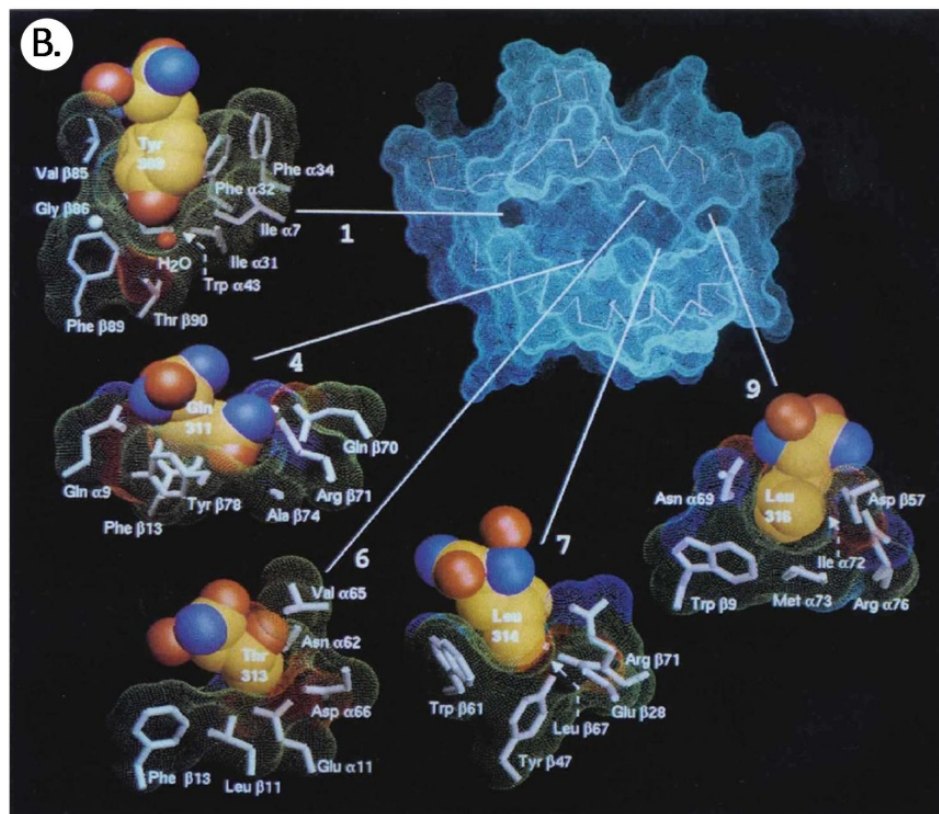
Figure 1.9. Le peptide antigénique en conformation étendue. A. Vue de côté du peptide HA (306-318) de l'hémagglutinine du virus de l'influenza. Les positions P1 à P9 du peptide sont indiquées en rouge. La coloration des atomes du peptide illustre l'interaction avec HLA-DR1. Les atomes en noir sont camouflés dans le complexe peptide HA/HLA-DR1; les atomes en gris sont en contact avec HLA-DR1 et accessibles au solvant; les atomes en blanc ne sont pas en contact avec HLA-DR1. **B.** Vue de côté de la surface moléculaire de la niche peptidique de HLA-DR1 (en bleu) en contact avec le peptide HA. Les chaînes latérales sont identifiées par le code à une lettre. **C.** Représentation schématique du réseau de ponts hydrogènes entre la molécule HLA-DR1 et la chaîne principale du peptide. Les ponts hydrogènes conservés sont indiqués par un pointillé rouge et numérotés de 1 à 12 à partir de l'extrémité N-terminale. Les acides aminés des chaînes α et β de la molécule HLA-DR1 sont illustrés (Adapté de Stern et al. 1994, Stratikos et al. 2004).

Figure 1.10. Modélisation tridimensionnelle de la niche peptidique de la molécule HLA-DR1. A. Les poches de liaison au peptide. Les poches de liaison au peptide (1, 3, 4, 6, 7, et 9) de la niche peptidique sont numérotées en jaune. Le potentiel électrostatique positif (bleu) et négatif (rouge) est illustré. **B.** Interaction des poches de liaison au peptide avec les chaînes latérales du peptide HA (306-318) de l'hémagglutinine du virus de l'influenza. Les détails du positionnement des chaînes latérales du peptide sont illustrés pour les poches: P1 qui reçoit la Tyr 308, P4 (Gln 311), P6 (Thr 313), P7 (Leu 314) et P9 (Leu 316). Les atomes des chaînes latérales sont représentés par le modèle CPK; atome de carbone (jaune), d'azote (bleu) et d'oxygène (rouge). Les chaînes latérales de la molécule HLA-DR1 sont indiquées par des bâtonnets blancs (Batalia et al. 1997, Stern et al. 1994).

A.



B.



sont accueillies par ces poches formées des chaînes latérales d'acides aminés situées à des positions polymorphiques (Figure 1.10B). Les propriétés de ces poches (taille, potentiel électrostatique et hydrophobicité), dépendant de la nature des acides aminés qu'on y retrouve, déterminent la spécificité du peptide pour différentes molécules du CMH de classe II.

1.1.5. Le CMH chez les vertébrés non-mammaliens

Nos connaissances actuelles sur le CMH proviennent en très grande partie d'études réalisées chez l'homme et la souris depuis les soixante-dix dernières années (Gorer 1937). Depuis ce temps, des gènes du CMH de classe I ou de classe II ont été identifiés chez tous les vertébrés à mâchoire (gnathostomes) étudiés, les poissons cartilagineux (ex : les requins) étant le plus ancien groupe (Kasahara et al. 1992). La prochaine section décrit les propriétés associées au CMH pour différents groupes d'animaux non-mammaliens à partir des oiseaux, en passant par les reptiles et les amphibiens jusqu'aux poissons osseux et poissons cartilagineux. Les études montrent que l'organisation des gènes du CMH diffère parmi ces groupes. On peut cependant distinguer deux caractéristiques qui semblent propres aux groupes des non-mammaliens par rapport aux mammifères. Premièrement, chez les oiseaux, les poissons osseux et possiblement les amphibiens et les poissons cartilagineux, on retrouve les gènes codants pour les sous-unités du protéasome (LMP) ou ceux pour le transporteur TAP en association avec les gènes classiques du CMH de classe I, et non avec ceux du CMH de classe II. Ceci indique que chez ces animaux, ces gènes ont vraisemblablement évolué de façon concertée étant donné les fonctions qui leur sont associées. Deuxièmement, la région du CMH de classe III n'est pas présente dans tous ces groupes d'animaux; elle est retrouvée chez les amphibiens mais est absente chez les poissons osseux et un seul gène est toujours présent chez les oiseaux (Flajnik et Kasahara 2001).

1.1.5.1. Le CMH chez les oiseaux (poulet)

Le CMH du poulet, auquel on réfère aussi en tant que locus *B*, est une région du génome de 92 kb qui contient 19 gènes (Kaufman et al. 1999). Des gènes orthologues à ceux des mammifères y sont retrouvés pour chacune des trois régions, mais plusieurs sont absents. Deux gènes du CMH de classe I, dont un seul semble exprimé à un niveau permettant la présence de la molécule à la surface cellulaire, sont situés de part et d'autre des gènes TAP1 et TAP2 codant pour le transporteur TAP. On ne retrouve cependant pas les gènes des sous-unités du protéasome (LMP2 et LMP7) qui ont probablement été supprimés du génome (Kandil et al. 1996). Les gènes codant pour la chaîne β du CMH de classe II et les chaînes α et β (DM α et DM β) de la molécule accessoire DM sont présents, alors que ceux codant pour la chaîne α du CMH de classe II et les deux chaînes de l'autre molécule accessoire, DO (DN/DZ α et DO β), sont absents. En fait, un seul gène non-polymorphique codant pour la chaîne α du CMH de classe II est situé à 5 centimorgans (cM) du locus *B*, ce qui laisse la chaîne β responsable du polymorphisme de cette molécule chez le poulet. Seul un gène d'histone et le gène C4 constitue la région du CMH de classe III, les autres étant retrouvés dans des régions différentes du génome. Cette région compacte du génome a amené Kaufmann et collaborateurs (1995) à qualifier le poulet comme ayant un "CMH minimal et essentiel", dans lequel les gènes non essentiels à la présentation antigénique ont été éliminés. Finalement, on retrouve chez d'autres oiseaux tels que la caille et le moineau un CMH dont le contenu et l'organisation des gènes sont similaires au poulet, cette région a cependant doublé en taille et en nombre de gènes suite à des duplications géniques (Freeman-Gallant et al. 2002; Shiina et al, 2004).

1.1.5.2. Le CMH chez les reptiles

L'information disponible sur le CMH des reptiles est limitée, mis à part la présence de gènes du CMH de classe I chez un serpent (*Nerodia sipedon*) et un lézard (*Amieva amieva*), et de classe II chez un alligator (*Alligator mississippiensis*) et un lézard (*Sphenodon punctatus*) (Grossberger et

Parham, 1992; Edwards et al. 1995; Miller et al. 2005). Chez le lézard *Amieva amieva*, cinq cDNA distincts ont été isolés d'un même individu, ce qui indique l'expression d'au moins trois loci du CMH de classe I.

1.1.5.3. Le CMH chez les amphibiens

Les CMH de deux espèces d'amphibiens ont été étudiés, chez la grenouille (*Xenopus laevis*) et l'axolotl (*Ambystoma mexicanum*), et chaque espèce montre des propriétés uniques. Chez la grenouille, on retrouve les régions du CMH de classe I, de classe II et de classe III; la région de classe I se trouvant entre les deux autres comme chez les oiseaux. Comme chez le poulet, il n'y a également qu'un seul locus du CMH de classe I exprimé à un niveau détectable (Shum et al. 1993), auquel on retrouve fortement associé les gènes TAP1, TAP2 et LMP7 (Nonaka et al. 1997). On retrouve dans la région du CMH de classe II trois gènes codant pour la chaîne β (Sato et al. 1993), alors que seulement deux des trois gènes codant pour la chaîne α semblent exprimés (Liu et al. 2002). Comme chez les mammifères, il y a dans la région du CMH de classe III les gènes C4, HSP70 et Bf en étroite association (Flajnik et al. 1999). Même si le génome de la grenouille est tétraploïde, son CMH est diploïde (Sato et al. 1993) et toutes les copies de gènes qui ont été dupliquées ont été rendues silencieuses, à l'exception de chez *Xenopus ruwenzoriensis* (Sammut et al. 2002). Chez l'axolotl, contrairement au CMH de la grenouille, on retrouve plusieurs loci exprimés du CMH de classe I auxquels est associé un seul locus du CMH de classe II (Sammut et al. 1999).

1.1.5.4. Le CMH chez les poissons osseux (les téléostéens)

Le groupe des poissons osseux inclut plus de 20 000 espèces dont plusieurs ont été étudiées au niveau du CMH dans des études d'association de gènes (*linkage studies*): le poisson zèbre (*zebrafish*, *Danio rerio*; Bingulac-Popovic et al. 1997), la carpe (*Cyprinus carpio*; Wiegertjes et al. 1997), le saumon atlantique (*Salmo salar*; Grimholt et al. 2002), la truite arc-en-ciel (*Oncorhynchus*

mykiss; Hansen et al. 1999), le médaka (*Oryzias latipes*; Naruse et al. 2000) et le *pufferfish* Japonais (*Takifugu rubripes*; Sambrook et al. 2002). Le séquençage de cette dernière espèce est complété (Aparicio et al. 2002) alors que celui du poisson zèbre est toujours en cours. De ces études d'association, il en ressort que l'organisation des gènes du CMH des téléostéens se distingue clairement de celle des mammifères à plusieurs niveaux. Premièrement, les gènes codants pour la chaîne α du CMH de classe I ne sont pas liés à ceux pour les chaînes α et β du CMH de classe II chez les poissons mentionnés ci-haut. Comme les gènes du CMH de classe I et de classe II sont situés sur des chromosomes différents, on devrait plutôt y faire référence comme étant des gènes majeurs d'histocompatibilité puisque l'utilisation du terme "complexe" n'a ici aucun sens (Dixon et Stet 2001). Deuxièmement, les gènes du CMH de classe II sont encodés à plus d'un locus chez le poisson zèbre, la carpe et le médaka. Troisièmement, on retrouve chez le poisson zèbre et le médaka deux gènes codant pour la chaîne α du CMH de classe I qui sont, comme chez la grenouille, liés aux gènes LMP2, LMP7 et TAP2 (Takami et al. 1997; Naruse et al. 2000). Finalement, il ne semble pas y avoir de région du CMH de classe III renfermant certains gènes codant pour des protéines du complément.

1.1.5.4.1. Les gènes majeurs d'histocompatibilité chez les salmonidés

Les gènes classiques du CMH de classe I et de classe II ont été identifiés chez plusieurs espèces de salmonidés. Les séquences de plusieurs molécules du CMH de classe I ont été obtenues chez le saumon atlantique (Grimholt et al. 1993), la truite arc-en-ciel (Hansen et al. 1996) et plusieurs espèces de saumon du Pacifique du genre *Oncorhynchus* (Miller et Withler 1997). Un seul locus du CMH de classe I est exprimé chez le saumon atlantique (Grimholt et al. 2002) alors que la molécule TAP est encodée par des loci polymorphiques, plus particulièrement les deux gènes identifiés pour TAP2 (Grimholt 1997). Comme observé chez le rat, la présence d'une molécule TAP polymorphique, dont les gènes codants sont ici étroitement liés à celui du CMH de classe I, permet probablement de produire différents peptides qui peuvent être présentés par un seul type de molécules du CMH de classe

I (Powis et al. 1996). Les séquences pour la chaîne α du CMH de classe II (Hordvik et al. 1993; Glamann 1995) et celles pour la chaîne β ont été également obtenues chez le saumon atlantique et la truite arc-en-ciel (Grimholt et al. 2000). Des séquences partielles des domaines $\beta 1$ et $\beta 2$ ont aussi été obtenues pour sept espèces de saumon du Pacifique (*Oncorhynchus*) (Miller et Withler 1996). On a confirmé chez le saumon atlantique que ces deux gènes codant pour la molécule du CMH de classe II ségrégaient sous un même haplotype et qu'un seul locus pour chacune des chaînes était exprimé (Stet et al. 2002). De plus, l'isolement des promoteurs de ces deux gènes montre qu'ils sont fonctionnels et que les boîtes X, X2 et Y sont conservées, suggérant ainsi que l'expression des gènes est régulée d'une manière similaire à celle retrouvée chez l'humain (Syed et al. 2003).

Chez les salmonidés, l'expression des gènes du CMH de classe II est observée dans les tissus lymphoïdes primaires (rein et thymus) aussi bien que dans les tissus lymphoïdes secondaires (rate, branchies, intestin) et les lymphocytes du sang périphérique (Juul-Madsen et al. 1992; Koppang et al. 1998; Nath et al. 2006). Il a été rapporté chez l'humain que la densité des molécules du CMH de classe II à la surface cellulaire a un effet sur la réponse immunitaire et dépend également du niveau d'expression des gènes (Maffei et al. 1989; Glimcher et Kara 1992). De plus, la régulation du gène s'effectue aux niveaux transcriptionnel et post-transcriptionnel (Rosa et Fellous 1988; Cullel-Young et al. 2001). Des études chez les salmonidés ont montré que durant une infection, le lipopolysaccharide (LPS) stimulait l'expression du CMH de classe II et que l'effet pouvait être accentué par la présence de TNF- α (Koppang et al. 1998; Knight et al. 1998).

D'autres études ont démontré que l'expression des molécules du CMH était régulée par la température. Chez la carpe exposée à une température de 6 °C, on a noté une réduction sévère de l'expression de la β_2 -microglobuline qui se traduit par une absence de la molécule du CMH de classe I à la surface des lymphocytes en périphérie (Rodrigues et al. 1998). Cet effet de la température n'a cependant pu être démontré dans une étude chez des salmonidés, où aucune baisse de la β_2m au niveau de l'ARN et de l'expression à la surface cellulaire n'a été remarquée après une exposition de 10 jours à

2 °C (Kales et al. 2006). Ces résultats indiquent que parmi les téléostéens, différents moyens sont utilisés pour réguler l'activité du système immunitaire. À cet effet, une étude sur l'expression du CMH de classe II réalisée par Nath et al. (2006) démontre une baisse de l'expression des gènes codant pour les deux chaînes α et β qui se traduit au niveau protéique. Deux études montrent finalement la régulation par la température de gènes du système immunitaire inné et adaptatif après immunisation (Raida et Buchmann 2007; 2008).

1.1.5.5. Le CMH chez les poissons cartilagineux (les chondrichthyens)

Le groupe des poissons cartilagineux est le plus ancien des vertébrés à mâchoire (*jawed vertebrates*) à avoir divergé d'un ancêtre commun avec l'homme. Malgré sa position ancestrale, la plupart des gènes typiques du CMH ont été identifiés; celui du CMH de classe I (Bartl et al. 1997) ainsi que les gènes codants pour TAP1, TAP2, LMP2, LMP7, HSP70 et Bf (Flajnik et al. 1999). L'étude des gènes du CMH de classe I montre une grande diversité entre les allèles et que des événements de recombinaison entre les allèles et les gènes en seraient grandement responsables. Les gènes codant pour les chaînes α et β du CMH de classe II ont également été identifiés (Kasahara et al. 1992; Bartl et Weissman 1994), et il a été montré que c'est la chaîne α qui est polymorphique avec la majorité des variations au niveau du domaine de liaison au peptide (Kasahara et al. 1993). Contrairement à ce que l'on retrouve chez les poissons osseux, le CMH des poissons cartilagineux montre une forte liaison des gènes de classe I, de classe II et certains gènes de classe III (Ohta et al. 2000, Terado et al. 2003). Ceci suggère que l'ancêtre commun de tous les vertébrés à mâchoire devait posséder un CMH ainsi organisé.

1.1.6. Les pressions évolutives sur les gènes du CMH

Face au grand nombre de pathogènes de différentes natures qui développent sans cesse de nouveaux moyens pour éviter d'être reconnus et éliminés lors d'infection, les gènes du système immunitaire doivent continuellement évoluer afin de s'adapter à ces moyens d'évasion toujours plus sophistiqués. Les gènes du CMH ont un avantage non-négligeable dans cette lutte en étant exprimés de manière codominante; chaque copie du gène à un locus particulier sera exprimée dans un individu hétérozygote. Dans la prochaine section, nous verrons comment la polygénie et le polymorphisme ont été exploités par les gènes du CMH de classe I et de classe II pour présenter une plus grande variété de peptides et ainsi permettre au système immunitaire de se défendre contre ces pathogènes en évolution constante. Nous aborderons également les mécanismes responsables du maintien de ce polymorphisme, ceux opérés par les pathogènes eux-mêmes et les autres, par le choix du partenaire pour la reproduction. Finalement, des allèles particuliers des gènes du CMH ont été associés à la susceptibilité ou résistance face aux pathogènes. Nous énoncerons brièvement ces résultats pour divers vertébrés avant de nous attarder plus particulièrement à ceux obtenus chez les salmonidés.

1.1.6.1. La polygénie et le polymorphisme des gènes du CMH

On retrouve dans le CMH de l'humain et de la souris trois gènes codant pour les molécules classiques du CMH de classe I; HLA-A, HLA-B, HLA-C et H2-K, H2-D, H2-L respectivement. Pour les molécules classiques CMH de classe II on compte trois paires de gènes chez l'humain, HLA-DR, HLA-DP, HLA-DQ et deux paires chez la souris, H2-A et H2-E. Cette polygénie pour chacune des classes de molécules permet la présentation d'une plus grande variété de peptides à la surface cellulaire par des molécules qui ont des spécificités de liaison particulière. L'analyse des gènes retrouvés dans la région du CMH de classe I chez l'humain a permis d'établir que des événements de duplications successives et de diversification d'une certaine unité de base ont engendré le CMH tel que nous le connaissons aujourd'hui (Shiina et al. 1999). La duplication d'un gène permet de conserver sa

fonction initiale pour une première copie, alors que la seconde a la liberté de muter pour adopter possiblement une nouvelle fonction. Ce processus de “naissance et de mort” est également retrouvé dans plusieurs autres familles de gènes impliqués dans la défense immunitaire : les récepteurs Fc, les cytokines et les KIR (*killer-cell Ig-like receptor*) (Trowsdale and Parham 2004).

Le polymorphisme des gènes du CMH a été étudié plus en détails chez l’humain où plusieurs variants (allèles) ont été identifiés à chacun des loci. La notation de ces différents allèles est donnée par un nombre précédé d’un astérisque. Par exemple, HLA-DRB1*0401 désigne l’allèle 0401 du gène 1 de la chaîne β de la molécule du CMH de classe II. Les gènes du CMH sont en fait les plus polymorphiques connus à ce jour. Chez l’humain, il y a jusqu’à maintenant (mai 2009) 2385 allèles qui ont été identifiés pour les gènes classiques du CMH de classe I et 1000 pour ceux du CMH de classe II (Tableau 1.1) (Robinson et al. 2003). Le nombre d’allèles, montrant une variation à au moins un acide aminé, est variable pour chacun des gènes; le plus grand nombre étant retrouvé pour les gènes HLA-B et HLA-DRB1 avec 1178 et 618 allèles respectivement. À l’opposé, seulement 3 allèles ont été identifiés pour le gène HLA-DRA. De plus, la variabilité des gènes codant pour la chaîne β de chacune des trois familles du CMH de classe II est plus grande. Le polymorphisme des molécules du CMH a aussi été observé pour les gènes orthologues du CMH chez d’autres espèces.

Un tel niveau de variation du CMH a été associé en partie à l’optimisation de la protection immunitaire contre les pathogènes. Ceci se reflète dans la localisation de la variabilité qui se situe plus précisément au niveau d’acides aminés de la région de liaison au peptide (PBR, *peptide-binding region*) formée des domaines $\alpha 1\alpha 2$ et $\alpha 1\beta 1$ des molécules du CMH de classe I et de classe II (Parham 1988). De plus, on remarque que le taux de substitutions non-synonymes (provoquant un changement d’acide aminé) est plus grand que le taux de substitutions synonymes (ne provoquant pas de changement d’acide aminé) pour les mutations se produisant dans les codons d’acides aminés impliqués dans la liaison du peptide (Hughes et Nei 1988; 1989). Une étude de 2003 établit précisément la divergence en acides aminés entre les différents allèles d’un même gène du complexe

Tableau 1.1. Nombre d'allèles pour les gènes du complexe HLA.

HLA	Gène	Nb. allèles
Classe I		
	HLA-A	767
	HLA-B	1178
	HLA-C	440
Classe II		
	HLA-DPA1	27
	HLA-DPB1	133
	HLA-DQA1	34
	HLA-DQB1	96
	HLA-DRA	3
	HLA-DRB1	618
	HLA-DRB3	50
	HLA-DRB4	13
	HLA-DRB5	18

Nombre d'allèles identifiés (mai 2009) pour les gènes classiques du CMH de classe I et de classe II chez l'humain.

IMGT/HLA Database, <http://www.ebi.ac.uk/imgt/hla/stats.html>

(Robinson et al. 2003).

HLA, en comparant la protéine entière ou seulement la région pour la liaison au peptide (Tableau 1.2) (Reche et Reinherz 2003). Il faut interpréter les données en gardant à l'esprit que l'analyse a été faite avec les allèles connus à cette date. La variation en acides aminés diffère d'un gène à l'autre. Ainsi, les allèles du locus HLA-DRA ne diffèrent que par un seul acide aminé alors qu'il y a une variation d'environ 20 acides aminés en moyenne pour HLA-A. Ces résultats d'analyse confirment également que la majeure partie de la variation se situe dans la région de liaison au peptide. Ainsi, il y a pour les allèles HLA-DRB une variation moyenne de 11,04 acides aminés représentant 4,15 % de la chaîne entière, alors que la seule région de liaison au peptide varie en moyenne de 10,40 acides aminés (11,96 %).

Les différentes molécules du CMH présentent une variabilité distincte, plus particulièrement au niveau des acides aminés impliqués dans la liaison au peptide, mais également dans le contact avec le récepteur des lymphocytes T. Cette différence dans la localisation de la variabilité peut être appréciée lorsqu'on observe la structure tridimensionnelle des molécules (Reche et Reinherz 2003). La Figure 1.11 illustre la variabilité à la surface des molécules HLA-DP, HLA-DQ et HLA-DR. L'analyse de ces structures permet de faire ressortir trois caractéristiques. Premièrement, les positions variables (en bleu, Figure 1.11A-C) se retrouvent à différents endroits sur les molécules. Par exemple, les trois positions les plus variables pour les molécules du CMH de classe II sont $\beta 9$, $\beta 57$ et $\beta 86$ pour HLA-DP, $\alpha 44$, $\alpha 50$ et $\beta 57$ pour HLA-DQ, et $\beta 11$, $\beta 13$ et $\beta 37$ pour HLA-DR. Deuxièmement, puisque la majorité de la variabilité se retrouve à des positions où l'acide aminé est en contact avec le peptide, les positions variables se retrouvent au niveau des feuillets- β qui forment la base ($\beta 9$ de HLA-DP; $\beta 11$, $\beta 13$ et $\beta 37$ de HLA-DR), et des hélices- α qui constituent les parois latérales de la niche peptidique ($\beta 57$ de HLA-DP et HLA-DQ). Troisièmement, la variabilité est aussi présente à la surface des molécules dans des régions où des acides aminés sont en contact avec le TCR (en bleu, Figure 1.11D-F). Cette variabilité est faible pour HLA-DP, intermédiaire pour HLA-DR ($\beta 70-71$) et plus élevée pour HLA-DQ ($\alpha 50$ et $\beta 70-71$). Finalement, une autre étude montre que les structures tridimensionnelles

Tableau 1.2. Variation en acides aminés entre différents allèles du complexe HLA.

Gène	Pleine longueur			PBR ($\alpha 1\alpha 2$ ou $\alpha 1\beta 1$)		
	Variation (AA)	Variation (%)	Longueur (AA)	Variation (AA)	Variation (%)	Longueur (AA)
HLA-A	19,79 ± 9,56	5,42	365	16,20 ± 6,69	9,05	179
HLA-B	18,15 ± 8,01	5,00	363	15,89 ± 5,66	8,88	179
HLA-C	13,33 ± 6,74	3,64	366	10,72 ± 3,71	5,99	179
HLA-DPA	2,90 ± 2,33	1,11	260	1,64 ± 1,01	2,22	74
HLA-DPB	9,01 ± 5,56	3,4	258	7,63 ± 3,54	8,7	87
HLA-DQA	16,72 ± 9,87	6,56	255	12,67 ± 7,49	16,90	75
HLA-DQB	13,75 ± 7,53	5,27	261	12,53 ± 6,04	14,40	87
HLA-DRA	1,0 ± 0,0	0,4	254	0	0	74
HLA-DRB	11,04 ± 4,68	4,15	266	10,40 ± 4,22	11,96	87

Les valeurs correspondent à la moyenne ± écart type. Les variations en acides aminés entre les allèles dépendent de l'identité des allèles considérés, calculées par la méthode *all-against-all pairwise comparisons* (Adapté de Reche et Reinherz, 2003).

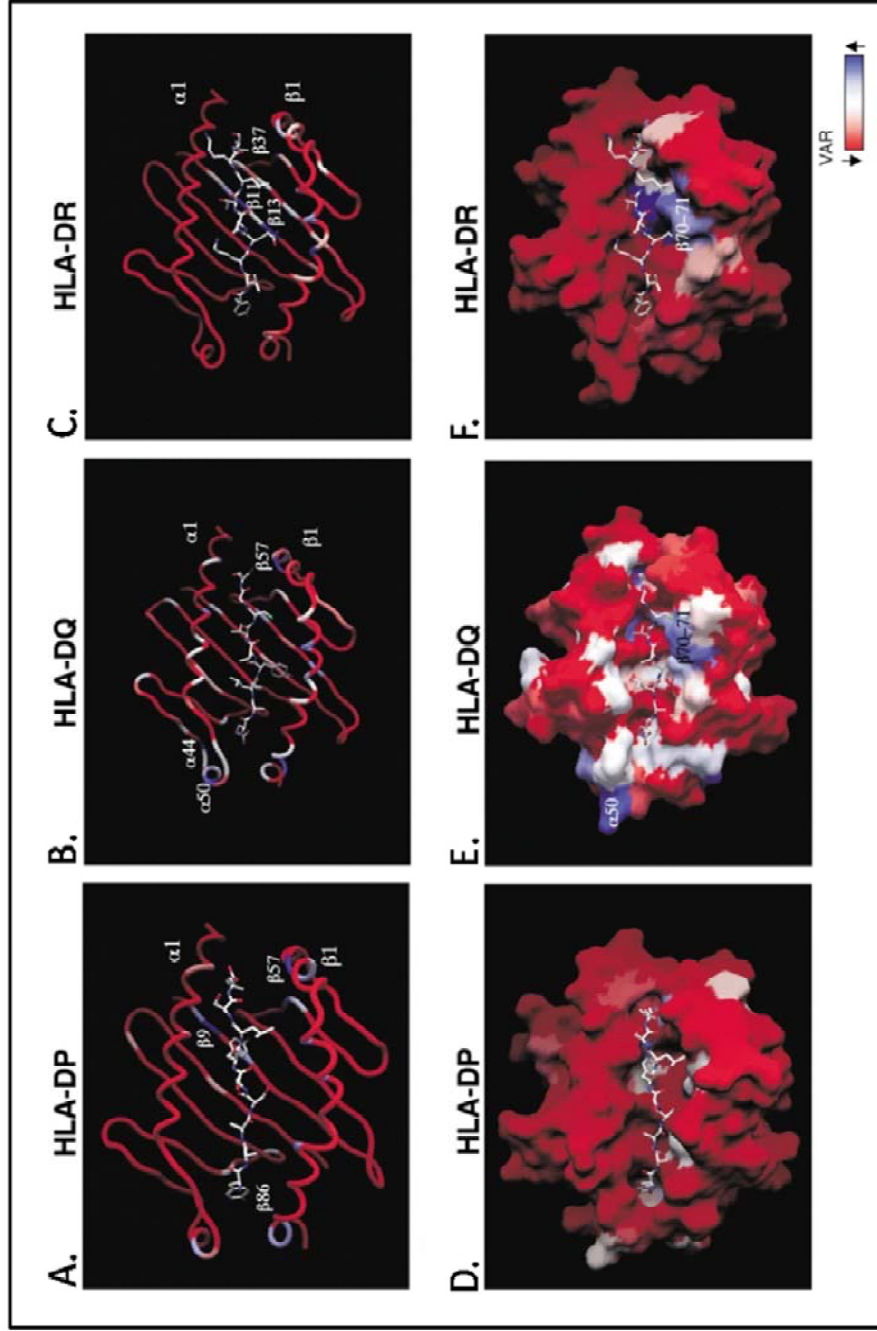


Figure 1.11. Structures tridimensionnelles montrant la variabilité de séquence des domaines $\alpha 1\beta 1$ des molécules du CMH de classe II. La variabilité de séquence des domaines $\alpha 1\beta 1$ des molécules HLA-DP, HLA-DQ et HLA-DR est marquée sur la représentation en tube (A, B et C) et à la surface de la molécule (D, E et F, respectivement). Les acides aminés consensus aux positions montrant la plus grande variabilité sont indiqués sur la représentation en tube. Les acides aminés consensus aux positions polymorphiques qui sont en contact avec le TCR sont indiqués à la surface des molécules. La variabilité est illustrée par le bleu (variable) et le blanc (invariable) et le rouge (variabilité intermédiaire). Des peptides de neuf acides aminés sont présents dans les représentations en tube (Adapté de Reche et Reinherz, 2003).

obtenues pour les gènes des différentes familles sont superposables, à l'intérieur d'une même ou entre les espèces (Figure 1.12) (Jones et al. 2006).

1.1.6.2. Le polymorphisme des gènes du CMH *en équilibre*

Trois destins sont possibles pour une mutation survenant dans un gène. La plupart du temps, elle sera rapidement éliminée de la population si elle ne procure pas d'avantages aux individus, on dira alors qu'elle s'est éteinte. Dans des rares cas, elle sera distribuée dans la population pour remplacer la version précédente du gène, de celle-ci on dira qu'elle s'est fixée. Un troisième type de mutation sera retrouvé en équilibre à une certaine fréquence dans une population et persistera sur une longue période de temps avant de s'éteindre ou de se fixer. La valeur de sélection est un facteur important qui influence la destinée de la mutation. Une mutation avantageuse soumise à la sélection positive pourra être fixée (sélection directionnelle) ou rester en équilibre à une certaine fréquence (sélection en équilibre, *balancing selection*). Ce dernier type de sélection est responsable du maintien du polymorphisme au niveau du CMH (Klein et al. 1998). La prochaine section exposera les différents mécanismes qui ont été proposés pour remplir cette tâche, ceux faisant intervenir les pathogènes et les autres qui reposent sur le choix du partenaire pour la reproduction (revue par Meyer et Thomson 2001; Bernatchez et Landry 2003).

1.1.6.2.1. Les mécanismes de sélection par les pathogènes

La sélection par les pathogènes opère lorsque qu'un allèle particulier d'un gène du CMH procure à l'individu une meilleure protection lors d'une infection. Trois mécanismes non-exclusifs ont été proposés pour expliquer le maintien du polymorphisme des gènes du CMH. Doherty et Zinkernagel (1975) ont proposé le modèle de l'avantage hétérozygote (*overdominant selection*) dans lequel les individus hétérozygotes sont en meilleure forme que les homozygotes lors d'une infection parce que les deux allèles qu'ils portent à un locus permettent de présenter une plus grande variété de

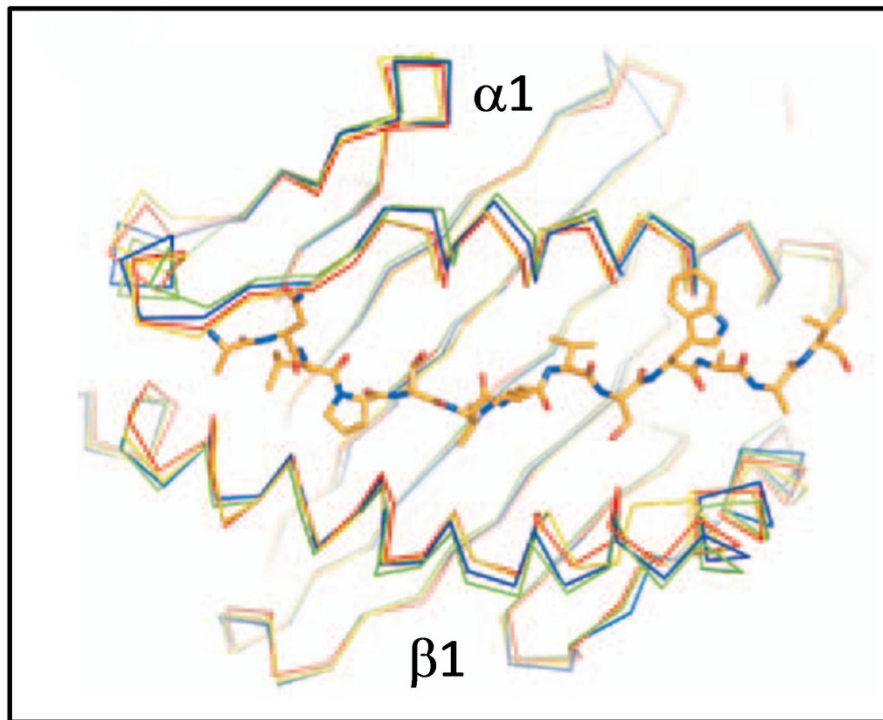


Figure 1.12. Structures tridimensionnelles montrant la superposition des domaines $\alpha 1\beta 1$ de molécules du CMH de classe II. Superposition de niches peptidiques pour certaines molécules du CMH de classe II de l'humain (HLA-DQ2, vert; HLA-DQ6.2, orange; HLA-DQ8, bleu) et de la souris (H2-IA^b, rouge; H2-IA^{g7}, jaune). Le peptide hypocrétine du complexe hypocrétine-HLA-DQ6.2 est illustré (Jones et al. 2006).

peptides antigéniques. Ainsi, ils peuvent également résister à une plus grande variété de pathogènes que l'homozygote (voir aussi Hughes et Nei 1988; 1989). Plusieurs études appuient ce modèle de l'avantage hétérozygote chez les humains et les animaux (Thursz et al. 1997; Carrington et al. 1999; Arkush et al. 2002; Penn et al. 2002).

Le second modèle proposé par Clarke et Kirby (1966) est celui de la sélection dépendante de la fréquence (*negative frequency dependent selection*). Il repose sur la coévolution des pathogènes avec les allèles du CMH; les individus possédant un allèle du CMH retrouvé à une faible fréquence (rare ou nouveau) dans la population sont avantagés puisque peu de pathogènes y ont été exposés et ont pu s'y adapter. Ainsi, la fréquence de ces allèles va augmenter jusqu'à être commune dans la population. Dans cette situation, les pathogènes qui vont s'adapter en exploitant une faille seront sélectionnés, et la forme relative des individus possédant cet allèle va décliner. Un processus cyclique est ainsi créé, où les allèles fluctuent en fréquence dans le temps selon l'adaptation des pathogènes (voir aussi Takahata et Nei 1990; Slade and McCallum 1992). En support à ce modèle, l'étude de Paterson et al. (1998) a associé des allèles du CMH et la résistance au parasitisme par des nématodes intestinaux. Une autre étude identifie des allèles de résistance à la malaria qui sont retrouvés à plus haute fréquence dans une population d'Afrique de l'ouest que dans d'autres groupes raciaux, possiblement suite à l'exposition au pathogène (Hill et al. 1991).

Le troisième modèle de la sélection fluctuante dans l'espace et le temps (*fluctuating selection*) a été proposé par Hedrick et Thompson en 1987 (voir aussi Hedrick 2002). Si la fréquence d'un pathogène change constamment, des molécules distinctes du CMH seront choisies à différents endroits ou moments. Le polymorphisme du CMH dans une population sera donc le reflet de la variation au niveau de la présence ou de la densité des pathogènes. Ce modèle se distingue du modèle de sélection dépendante de la fréquence; la bonne forme des individus varie en fonction de la fréquence des pathogènes plutôt qu'en fonction de la fréquence des allèles. Ce modèle a été validé par plusieurs

études qui appuient l'idée d'une variation dans la pression sélective dans le temps et l'espace (Landry et Bernatchez 2001; Miller et al. 2001; Charbonnel et Pemberton 2005; Dionne et al. 2007).

1.1.6.2.2 Les mécanismes de sélection par la reproduction

Des mécanismes de sélection par la reproduction contribuent également à former et maintenir le polymorphisme des gènes du CMH (Penn et Potts 1999). Un premier modèle qu'on appelle le croisement non-assorti (*disassortative mating*) repose sur le choix d'un partenaire sexuel qui possède des allèles différents au CMH, ce qui favorise la génération d'individus hétérozygotes. Des études ont observé ce mécanisme de sélection chez des souris en laboratoire, en milieu sauvage ou semi-naturel (Yamazaki et al. 1976; Egid et Brown 1989; Potts et al. 1991), chez le saumon atlantique (Landry et al. 2001) de même que chez l'humain (Ober et al. 1997; Génin et al. 2000). Deux autres modèles s'appuient sur les conséquences négatives provoquées par le croisement de proches parents ou l'incompatibilité de certains allèles du CMH : le croisement non-apparenté (*inbreeding avoidance hypothesis*; Potts et Wakeland 1990) ou de l'incompatibilité materno-fœtale (Ober et al. 1998). Finalement, des études indiquent que les gènes du CMH influencent l'odeur d'un individu et que la souris et l'homme y sont sensibles (Yamazaki et al. 1983; Wedekind et Fÿry 1997). Ainsi, l'effet des gènes du CMH sur l'odeur pourrait être direct par la liaison de molécules volatiles ou indirect en contrôlant la faune bactérienne qui génère ces odeurs (Jordan et Bruford 1998), ce qui pourrait influencer le choix du partenaire pour générer la descendance.

1.1.6.3. L'association entre les gènes du CMH et les maladies

Les gènes du complexe HLA ont été associés à plus d'une centaine de différentes maladies incluant la plupart, sinon toutes les maladies auto-immunes chez l'homme (Shiina et al. 2009). On peut mentionner entre autre les maladies communes comme le diabète, l'arthrite rhumatoïde, le psoriasis, l'asthme et plusieurs autres. Le tableau 1.3 dresse un portrait partiel des maladies infectieuses

Tableau 1.3. Nombre de publications rapportant une association positive ou négative entre le CMH et des maladies.

Maladie	Type d'association ^a	No. d'études rapportant une association avec:				
		CMH de classe I			CMH de classe II	
		A	B	C	DR	DQ
HIV/SIDA						
Infection	+	4	4	9	7	6
	-	3	4	6	4	5
Progression	+	5	2	25	12	6
	-	1	1	4	-	-
Transmission	+	2	1	3	5	2
	-	1	-	1	1	1
Hépatite B						
	+	6	1	6	19	10
	-	2	-	2	2	2
Hépatite C						
	+	7	9	9	23	16
	-	2	2	2	2	2
Lèpre						
	+	7	4	9	25	11
	-	2	2	3	1	1
Tuberculose						
	+	4	1	6	23	14
	-	4	2	4	2	2
Malaria						
	+	-	-	5	7	4
	-	-	-	-	-	-
Leishmaniose						
	+	3	2	3	5	3
	-	1	-	1	2	1
Schistosomiase						
	+	5	-	8	4	4
	-	1	-	2	-	-

^a +, positive; -, négative

(Adapté de Blackwell et al. 2009)

pour lesquelles un grand nombre de publications rapportent des associations positives ou négatives pour les différentes familles de gènes classiques du CMH (adapté de Blackwell et al. 2009). On remarque que dans la majorité des familles de gènes, des allèles de résistance et d'autres de susceptibilité ont été identifiés pour des infections de différents types : virales (HIV/SIDA, hépatites B et C), bactériennes (lèpre et tuberculose) et parasitaire (malaria, leishmaniose, schistosomiase). Par exemple, le gène HLA-B*27 qui reconnaît un épitope conservé d'une protéine de la capsid du HIV est associé à une survie accrue des individus, alors que HLA-B*35 est associé à une progression rapide de la maladie (Bailey et al. 2006, Tomiyama et al. 1997).

Le modèle animal de son côté procure les meilleures évidences d'association entre des allèles du CMH et la résistance ou susceptibilité aux maladies puisqu'il a l'avantage de permettre l'utilisation de pathogènes spécifiques pour des infections en milieu contrôlé. C'est chez le poulet qu'on retrouve la plus forte association entre le CMH et une maladie, la résistance de l'haplotype B21 aux tumeurs causées par le virus MDV (*Marek's disease virus*) (Kaufman 2000). Ces études ont permis entre autres, de développer par croisement sélectif des lignées de poulet offrant une meilleure résistance aux maladies et donc un meilleur rendement pour les éleveurs (Bacon et al. 2000). Les études d'association entre le CMH et les maladies sont certainement simplifiées par le fait qu'ils ont un "CMH minimal et essentiel" (Kaufman et al. 1995). Dans une population naturelle de moutons, une autre étude a montré que certains allèles du CMH étaient associés à une faible probabilité de survie et à un niveau élevé de parasitisme ou *vice versa* (Paterson et al. 1998).

Chez les poissons osseux, plusieurs études chez le saumon atlantique associent également certains allèles du CMH à la résistance ou susceptibilité aux maladies. Langefors et collaborateurs (2001) ont été les premiers à établir une corrélation entre les probabilités de survie et trois allèles différents du CMH de classe II β après une exposition à *Aeromonas salmonicida*. Leurs observations ont été confirmées dans une expérience où l'effet des gènes non-associés à ceux d'histocompatibilité majeure (MH) était contrôlé (Lohm et al. 2002). Toujours chez le saumon atlantique, deux autres

études ont identifié des associations significatives entre des allèles du CMH de classe I et de classe II et la résistance à des infections d'origine bactérienne ou virale (Grimholt et al. 2003; Kjøglum et al. 2008). Il est intéressant de constater que différents allèles furent associés à la résistance pour chaque type de pathogène. De plus, certaines combinaisons d'allèles du CMH de classe I et de classe II ont montré un effet suggérant une communication entre les deux voies (*cross-talk*). Une étude avait déjà démontré l'effet de telles combinaisons d'allèles qui étaient associées à la résistance ou la susceptibilité au virus ISAV (*infectious salmon anaemia virus*) (Kjøglum et al. 2006). Finalement, ce qui peut être interprété comme une manifestation de la sélection naturelle par les pathogènes dans un milieu sauvage a été observé au locus du CMH de classe II (de Eyto et al. 2007; Dionne et al. 2009).

1.2. Les séquences d'ADN répétitives

Le séquençage complet et l'analyse de plusieurs génomes eucaryotes nous procurent aujourd'hui une quantité phénoménale d'informations sur les séquences d'ADN, ouvrant ainsi la voie à une nouvelle ère en génétique : la génomique. Curieusement, une dizaine d'années avant le développement des méthodes de séquençage de l'ADN vers la fin des années 1970 (Maxam et Gilbert 1977; Sanger et al. 1977), on avait déjà posé l'hypothèse que des séquences d'ADN répétitives étaient présentes par centaines de milliers de copies dans le génome d'organismes supérieurs (Britten et Kohne 1968). Ces séquences, représentant 10 % de l'ADN du génome de la souris et faisant partie de la fraction d'ADN satellite, s'avéraient être des répétitions d'une courte séquence de nucléotides répétée en tandem (Waring et Britten 1966). Aujourd'hui, on sait que ces séquences d'ADN répétitives sont très abondantes dans les génomes eucaryotes, constituant près de 40 % et 50 % du génome de la souris et de l'humain respectivement (Mouse Genome Sequencing Consortium 2002; Levy et al. 2007). On peut classer les séquences d'ADN répétitives en deux grandes catégories : celles qui sont dites dispersées et celles qui sont répétées en tandem. Chacune est ensuite divisée en superfamilles et en familles dont les membres principaux sont indiqués à la Figure 1.13. Les séquences d'ADN répétitives

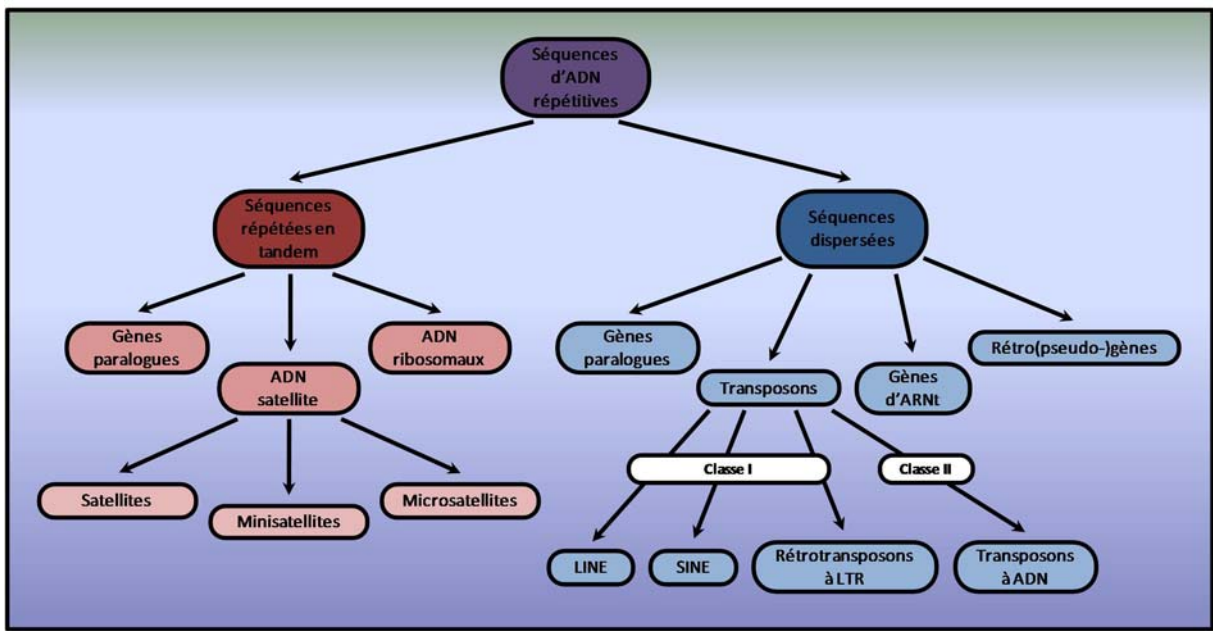


Figure 1.13. Séquences d'ADN répétitives retrouvées dans les génomes eucaryotes. Les deux catégories principales d'éléments répétés (dispersés et répétés en tandem) sont indiquées avec chacune de leurs subdivisions (Adapté de Richard et al. 2008).

dispersées incluent la superfamille des gènes d'ARNt, des gènes paralogues, des rétro(pseudo-)gènes et des transposons, eux-mêmes subdivisés en familles de LINE (*long interspersed nuclear element*), de SINE (*short interspersed nuclear element*), de rétrotransposons à LTR (*long terminal repeats*) et de transposons à ADN. Les séquences d'ADN répétitives répétées en tandem englobent la superfamille des gènes paralogues, des ADN ribosomiaux et des ADN satellites, qui sont à leur tour divisés en familles d'ADN satellites, minisatellites et microsatellites (Richard et al. 2008). Des études ont démontré que des éléments appartenant à chacune de ces deux catégories étaient impliqués dans la régulation de l'expression génique (Tomilin 2008). Dans les sections qui suivent, nous allons d'abord donner les caractéristiques des membres des familles de transposons et d'ADN satellites avant d'exposer les effets biologiques associés à leur présence. Nous distinguerons également les mécanismes moléculaires responsables de la création et de la propagation des séquences d'ADN répétitives pour chacune des deux catégories. Ceux impliqués dans l'instabilité des minisatellites et des microsatellites seront décrits plus en détails. Finalement, nous verrons comment les microsatellites et les minisatellites constituent des éléments importants pour l'évolution des espèces, permettant leur adaptation rapide aux conditions environnementales.

1.2.1. Les séquences d'ADN répétitives dispersées (transposons)

Le système de classification des transposons distingue deux classes en fonction de la nature de l'intermédiaire de transposition : l'ARN (classe I, rétrotransposon) ou l'ADN (classe II). La classe I inclut donc les LINE, les SINE et les rétrotransposons à LTR, alors que les transposons à ADN sont retrouvés dans la classe II (Figure 1.14). Certains rétrotransposons sont autonomes, c'est à dire qu'ils produisent les protéines requises pour leur rétrotransposition, alors que d'autres ne le sont pas. Le mécanisme de transposition des éléments de la classe I se fait sous le modèle du "copier-coller", alors qu'on parle plutôt du modèle "couper-coller" pour ceux de la classe II (Wicker et al. 2007). L'abondance relative de ces éléments est fortement variable entre les génomes eucaryotes. Par exemple,

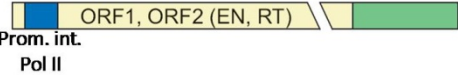

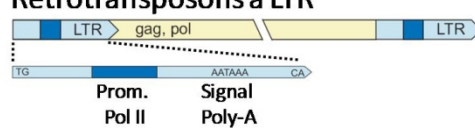
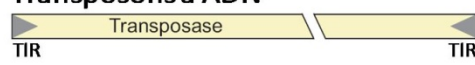
Familles de séquences d'ADN dispersées	Taille	Nb. copies	Fraction du génome
Classe I			
LINE 	6-8 kb	850 000	21%
SINE 	80-500 pb	1 500 000	13%
Rétrotransposons à LTR 	4-12 kb	450 000	8%
Classe II			
Transposons à ADN 	1,3-2,4 kb	300 000	3%

Figure 1.14. Caractéristiques des séquences d'ADN répétitives dispersées des génomes eucaryotes. Une représentation schématique est donnée pour chacune des familles d'éléments. Les régions bleu foncé indiquent des régions promotrices et les régions vertes les régions 3'UTR. Les boîtes A et B sont indiquées pour le promoteur des SINE. Le nombre (Nb.) de copies et la fraction du génome sont les données pour l'humain. LINE, long interspersed nuclear element; SINE, short interspersed nuclear element; LTR, long terminal repeat; TIR, terminal inverted repeats; Prom., promoteur; int., interne; Pol II et III, ARN polymérase II et III; EN, endonucléase; RT, transcriptase inverse (Adapté de IHGSC 2001; Jurka et al. 2007).

environ 90 % des transposons sont de classe I chez l'homme et la souris, alors qu'on en retrouve moins de 20 % chez l'amibe (*Entamoeba invadens*) et le nématode *Caenorhabditis elegans* (Feschotte et Pritham 2007). Des recombinaisons homologues entre transposons peuvent induire des réarrangements chromosomiques tels que des délétions, inversions et duplications de segments d'ADN, aussi bien que la mutation d'un gène lorsqu'ils y sont transposés (Belancio et al. 2008). Pour cette raison, les transposons affecteraient la structure des génomes et pourraient constituer des éléments importants dans l'évolution des espèces (Jurka et al. 2007). Les sections suivantes traiteront des caractéristiques et des effets biologiques associés aux membres des quatre familles principales de transposons.

1.2.1.1. Les LINE (*long interspersed nuclear element*)

Les LINE sont des rétrotransposons autonomes ne possédant pas de séquences LTR à leurs extrémités. Ces éléments, dont le membre le mieux caractérisé est le LINE-1 (L1), occupent environ 20% du génome humain. La transcription de L1 par l'ARN polymérase II, à partir d'un promoteur interne situé dans la région 5'-UTR, génère un ARN messager bicistronique de 6kb (Skowronski et al. 1988). Le produit du premier cadre de lecture ouvert (*open reading frame 1*, ORF1) possède une activité chaperonne pour acides nucléiques et forme un trimère qui se lie l'ARNm de L1 pour générer un complexe ribonucléoprotéique considéré comme un intermédiaire de transposition. Le second ORF encode une protéine ayant des activités d'endonucléase (EN) et de transcriptase inverse (*reverse transcriptase*, RT), toutes deux nécessaires à la transposition. La séquence de L1 se termine par une région 3'UTR contenant un signal de polyadénylation (Belancio et al. 2008). Il est intéressant de noter que la région 5'UTR de l'élément L1 permet également la transcription dans l'orientation anti-sens par rapport à L1, générant ainsi des ARNm complémentaires à l'ADN génomique (Nigumann et al. 2002). Plus étonnant encore, plusieurs LINE contiennent des inversions et la transcription normale à partir de ces promoteurs est aussi dans la direction opposée, toujours par rapport à L1 (Ostertag et Kazazian 2001).

Même si la plupart des éléments L1 sont inactifs, quelques-uns demeurent capables de rétrotransposition. La Figure 1.15 décrit les étapes d'intégration de l'élément LINE L1. Premièrement, l'activité endonucléase (EN) de la protéine encodée par l'ORF2 du rétrotransposon crée un bris d'ADN simple-brin entre le T et le A du brin négatif, à la séquence consensus 5'-TTTTAA-3' de l'ADN génomique. Ensuite, il y a appariement de la queue poly(A) de l'ARNm de L1 avec la séquence poly(T) au site du bris, séquence qui sert également d'amorce pour la synthèse de l'ADNc par l'activité transcriptase inverse (RT) de la protéine encodée par l'ORF2. La dernière partie du mécanisme permet la dégradation de l'ARNm de L1 et la synthèse du second brin suivi de ligations aux extrémités de l'insert LINE L1 (Jurka 1997; Belancio et al. 2008).

La présence des rétrotransposons LINE dans les génomes eucaryotes affecte plusieurs activités cellulaires (Buzdin 2004). Ces éléments sont des sites chauds (*hot spots*) pour la recombinaison, causant ainsi plusieurs maladies. Par exemple, la recombinaison homologue entre des éléments L1 présents dans le gène de la sous-unité bêta de la phosphorylase kinase (PHKB) cause la glycogénose (Burwinkel et Kilimann 1998). Lors de la méiose, des croisements inégaux entre des séquences L1 sur des chromosomes homologues peuvent créer des nouvelles familles de gènes, comme montré par la duplication des gènes de la gamma-globine (Kazazian 2000). L'insertion de séquences LINE dans des gènes entraîne également plusieurs autres problèmes. Ainsi, l'insertion dans le gène du facteur VIII cause l'hémophilie A (Van de Water et al. 1998), ou celle dans le gène de la dystrophine occasionne la dystrophie musculaire de Duchenne (Narita et al. 1993). En plus de provoquer de l'instabilité génomique, les éléments LINE régulent l'expression des gènes. Ils peuvent avoir un effet "*enhancer*" comme démontré pour l'apolipoprotéine A humaine, ou de répresseur transcriptionnel tel que rapporté pour le gène de l'insuline I chez le rat (Brosius 1999).

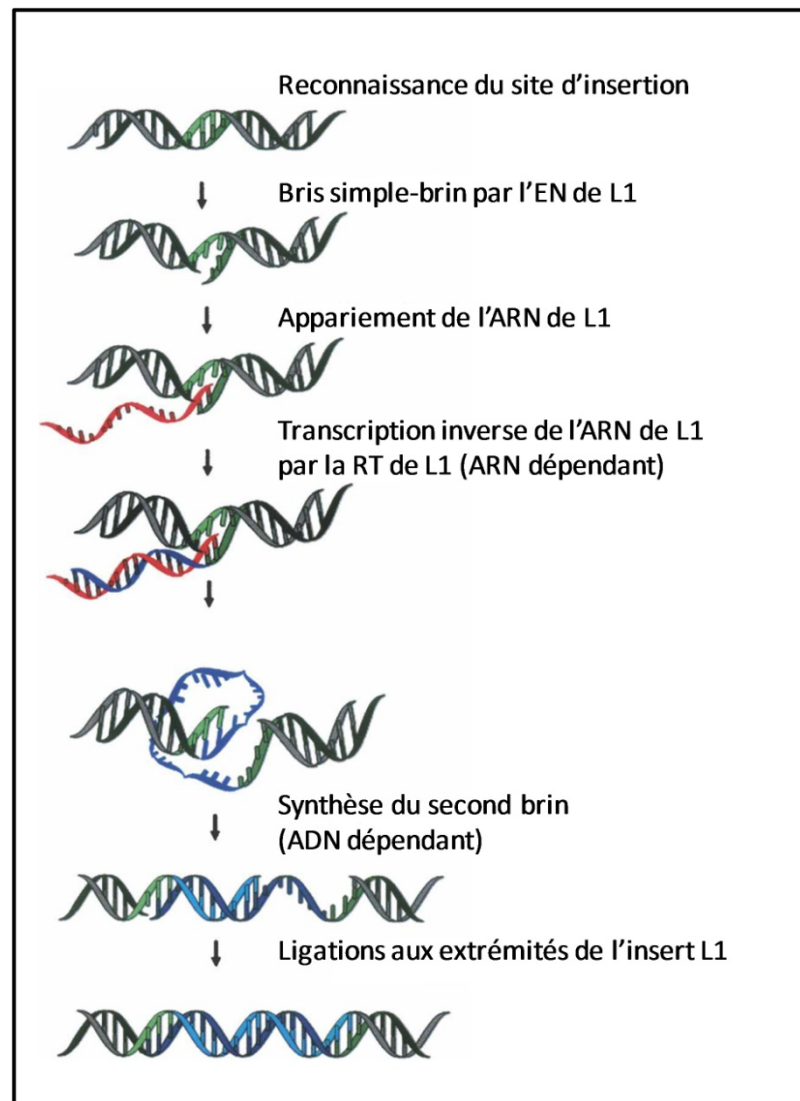


Figure 1.15. Mécanisme d'intégration du LINE L1. L'endonucléase (EN) encodée par l'ORF2 de L1 reconnaît le site consensus d'insertion (vert) et induit un bris d'ADN simple-brin sur l'ADN génomique. L'extrémité 3' résultant s'apparie à la queue poly-A de l'ARNm de L1 (rouge) et sert d'amorce pour la synthèse du premier brin d'ADNc par la transcriptase inverse (RT) de L1. Les étapes suivantes du processus d'intégration ne sont pas bien définies. Durant l'intégration de L1, un bris d'ADN simple-brin est fait sur l'autre brin et la synthèse du second brin (bleu clair) est faite. Finalement, les extrémités de L1 sont réparées pour compléter son intégration (Traduit de Belancio et al. 2008).

1.2.1.2. Les SINE (*short interspersed nuclear element*)

Les SINE sont des rétrotransposons non-autonomes, sans séquences LTR et d'origine distincte; ils ne dérivent pas d'une délétion sur un LINE, mais proviennent de rétrotranspositions accidentelles de différents ARNm produits par la polymérase III. Ces séquences d'ADN, dont la taille varie de 80 à 500 paires de bases, sont les éléments répétitifs les plus abondants et comptent pour environ 13 % du génome humain (Wicker et al. 2007). Le plus connus des SINE est l'élément *Alu* qui est présent à au moins 10^6 copies chez l'homme (Batzer et Deininger 2002), soit environ une copie à tous les 3 kb. Le corps des éléments *Alu*, qui font approximativement 300 pb de longueur, est composé de deux monomères d'environ 130 pb séparés par une région riche en bases adénine. L'origine des SINE (ARNt, ARN 7SL et ARN 5S) est révélée par l'extrémité 5' qui contient un promoteur interne pour l'ARN polymérase III (Pol III), alors qu'une queue poly(A) est retrouvée à l'extrémité 3'. Puisqu'aucun signal de terminaison pour la Pol III n'est retrouvé sur l'élément *Alu*, la transcription s'étend dans la région adjacente jusqu'à la rencontre d'une séquence formée d'au moins quatre bases thymine consécutives. Les éléments *Alu* ont gardé leur activité et se transposent par un mécanisme similaire à celui des LINE, nécessitant cependant une transcriptase inverse (RT) "exogène" puisqu'ils ne contiennent généralement pas de séquences codantes. La rétrotransposition de ces éléments est médiée par les LINE qui permettent la production de l'endonucléase et de la RT par l'ORF2 (Dewannieux et al. 2003).

Les effets provoqués par l'insertion d'éléments SINE sont généralement du même type que ceux des LINE, par exemple, des régions chromosomiques sont supprimées ou déplacées par la recombinaison d'éléments *Alu* (Brosius 1999). L'insertion d'éléments *Alu* dans certains gènes est également la cause de plusieurs maladies comme l'hémophilie et le développement de tumeurs, et peut jouer un rôle d'"*enhancer*" ou de répresseur de l'activité transcriptionnelle en plus d'être impliquée dans l'épissage alternatif (Buzdin 2004). Enfin, on a démontré une activité d'"*enhancer*" pour les SINE de type AmnSINE1, desquels 25 % sont localisés près de gènes impliqués dans le

développement du cerveau. Ceci pourrait suggérer l'implication de ce type de SINE dans le développement du système nerveux central des mammifères (Sasaki et al. 2008).

1.2.1.3. Les rétrotransposons à LTR

Les rétrotransposons à LTR sont des éléments autonomes formant environ 8 % du génome humain et dont la taille varie généralement de 4 à 12 kb. Ils se distinguent des autres membres de la classe I par la présence, comme leur nom l'indique, de séquences LTR flanquant la séquence codante interne. Ces LTR sont des séquences répétées dans la même orientation (directe) et dont la taille peut varier d'une centaine de paires de base à plus de 5 kb. Un exemple bien connu est l'élément Ty1 qui est responsable de la plupart des réarrangements chromosomiques chez la levure (Umezu et al. 2002). Le rétrotransposon à LTR est transcrit par l'ARN polymérase II (Pol II) à partir d'un promoteur situé dans le LTR à l'extrémité 5'. On retrouve généralement deux ORF dans la séquence de ces éléments. Le gène *GAG* produit une protéine structurale formant des particules semblables à un virus (*virus-like particle*, VLP) et à l'intérieur de laquelle la transcription inverse a lieu. Le gène *POL* encode une protéase (*aspartic proteinase*, AP), une transcriptase inverse (RT), une RNase H (RH) et une intégrase (IN). Un signal de polyadénylation est finalement retrouvé près du LTR situé en à l'extrémité 3' du rétrotransposon (Wicker et al. 2007).

Leur mécanisme de transposition des rétrotransposons à LTR se distingue de ceux des LINE et des SINE (Figure 1.16). Ils partagent en fait le même mécanisme de réplication de certains rétrovirus, à l'exception qu'ils ne peuvent former de particules infectieuses puisqu'ils ne possèdent pas le gène *ENV* codant pour la protéine de l'enveloppe. Une première étape du mécanisme implique la transcription du rétrotransposon par la Pol II cellulaire produisant l'ARNm qui est ensuite traduit dans le cytoplasme afin de générer les protéines qui s'assembleront en VLP. Habituellement, deux molécules d'ARN du rétrotransposon sont assemblées dans une particule, où ont ensuite lieu les réactions de transcription inverse, impliquant aussi la RNase H, pour générer une copie ADN double-

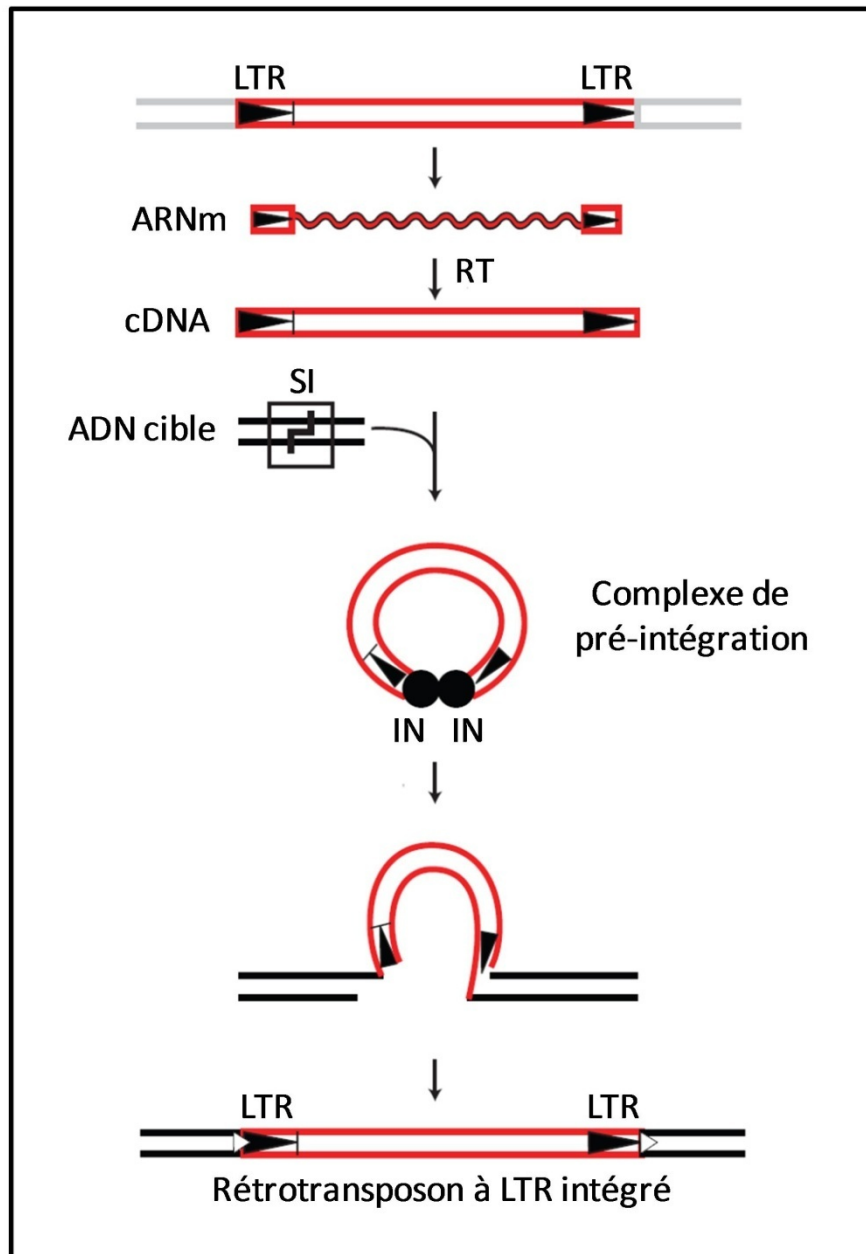


Figure 1.16. Mécanisme d'intégration des rétrotransposons à LTR. Le rétrotransposon est coloré en rouge, les séquences LTR (long terminal repeat) sont représentées par des boîtes avec une flèche et le chromosome donneur par des lignes grises. Le transcrite du rétrotransposon (ligne ondulée rouge) est traduit dans le cytoplasme et transformé en cDNA par sa transcriptase inverse (RT). L'intégrase (IN) représentée par les points noirs, permet l'accès au site d'intégration (SI) sur le chromosome receveur (lignes noires) (Traduit de Beauregard et al. 2008).

brin du rétrotransposon. Cette dernière est par la suite transférée au noyau pour être insérée dans un chromosome par un évènement de recombinaison impliquant l'intégrase (Wilhelm et Wilhelm 2001; Havecker et al. 2004).

Comme mentionné précédemment, un des effets des rétrotransposons à LTR se fait sentir par des réarrangements de structure de l'ADN. Il est intéressant de constater que la recombinaison homologue entre les deux LTR d'un rétrotransposon élimine l'élément du génome (*popping out*) en laissant en cicatrice un seul LTR. Les rétrotransposons à LTR ont également des implications dans la transcription de plusieurs gènes. Un mécanisme important repose sur la présence dans ces éléments de séquences pouvant lier des récepteurs hormonaux et des régulateurs transcriptionnels avec une forte affinité. Notons par exemple des sites de liaison pour le récepteur de l'acide rétinoïque (RAR), des glucocorticoïdes (GR) et de la progestérone (PR), et aussi pour les facteurs de transcription NF- κ B, Myb, Oct-1, Sp1, Sp3. Ces éléments peuvent être impliqués dans des fonctions normales du génome, par exemple l'activité "*enhancer*" du LTR de l'élément ERV-9 qui détermine l'expression du gène de la β -globine dans les érythrocytes (Buzdin 2004).

1.2.1.4. Les transposons à ADN

Contrairement à la première classe des rétrotransposons, la seconde classe de transposons à ADN se transpose en déplaçant le contenu en ADN d'un site chromosomique à un autre sans passer par un intermédiaire ARN. Cette famille d'éléments, dont la taille varie de 1,3 à 2,4 kb, constitue environ 3 % du génome humain. L'information disponible sur les transposons est limitée par rapport aux rétrotransposons, ce qui s'expliquerait par leur découverte relativement récente dans les génomes mammifères. De plus, il y aurait eu une activité intense des transposons à ADN au cours de l'évolution des mammifères et des premiers primates, alors que l'activité de transposition aurait cessé au cours des dernières phases évolutives qui ont mené à l'homme (Pace et Feschotte 2007). On distingue trois sous-classes de transposons à ADN : (i) ceux qui sont excisés sous forme d'ADN double-brin et réinsérés

ailleurs dans le génome par un mécanisme “couper-coller” (ex. Tc1/Mariner, hAT, Merlin, élément P); (2) les *Helitrons* qui utilisent probablement le mécanisme du cercle roulant (*rolling-circle*) pour leur répliation; et (3) les *Mavericks* qui encoderaient leur propre polymérase pour se réplier (Feschotte et Pritham 2007). La vaste majorité des transposons se retrouve dans la première sous-classe et sont caractérisés par la présence d’un gène codant pour une transposase et des séquences terminales répétées-inversées (*terminal inverted repeats*, TIR) contenant un ou deux sites de liaison pour la transposase (Miskey et al. 2005). Il est intéressant de noter que la protéine RAG1 ainsi que ses séquences signal (SSR) pour les recombinaisons V(D)L des immunoglobulines seraient dérivées d’une transposase de la sous-famille de transposon *Transib* (Kapitonov et Jurka 2005).

Les éléments de la première sous-classe se transposent par un mécanisme “couper-coller” (Figure 1.17), et peuvent augmenter en nombre dans un génome en passant d’une région de chromosome déjà répliquée à une autre qui ne l’est pas encore. Le mécanisme implique la transposase qui initie l’excision du transposon par une coupure double-brin de l’ADN générant des bouts cohésifs pour le réintégrer ensuite à un autre site qui est fréquemment un dinucléotide TA. Le remplissage de l’extrémité simple-brin du transposon au site d’intégration et la réparation du bris double-brin au site donneur est effectué par la machinerie cellulaire de réparation de l’ADN. Comme pour les rétrotransposons à LTR, une légère cicatrice est laissée au site donneur après l’excision.

Les transposons à ADN peuvent affecter l’évolution de leur hôte à différents niveaux (Feschotte et Pritham 2007). Premièrement, l’insertion dans un gène peut altérer sa fonction. Ils sont justement de puissants outils utilisés dans des études d’analyse mutationnelles au cours desquelles on étudie la fonction d’un gène d’intérêt en l’inactivant à l’aide d’un transposon. Une autre approche, qui tire profit de la préférence des transposons à s’insérer près ou dans un gène, permet d’obtenir des phénotypes mutants en introduisant des mutations au hasard dans le génome pour ensuite identifier le gène muté (Spradling et al. 1995). Contrairement aux rétrotransposons, les transposons à ADN ont la propriété de pouvoir s’exciser spontanément. Par exemple, l’activité des éléments *Ac* et *Spm* du maïs

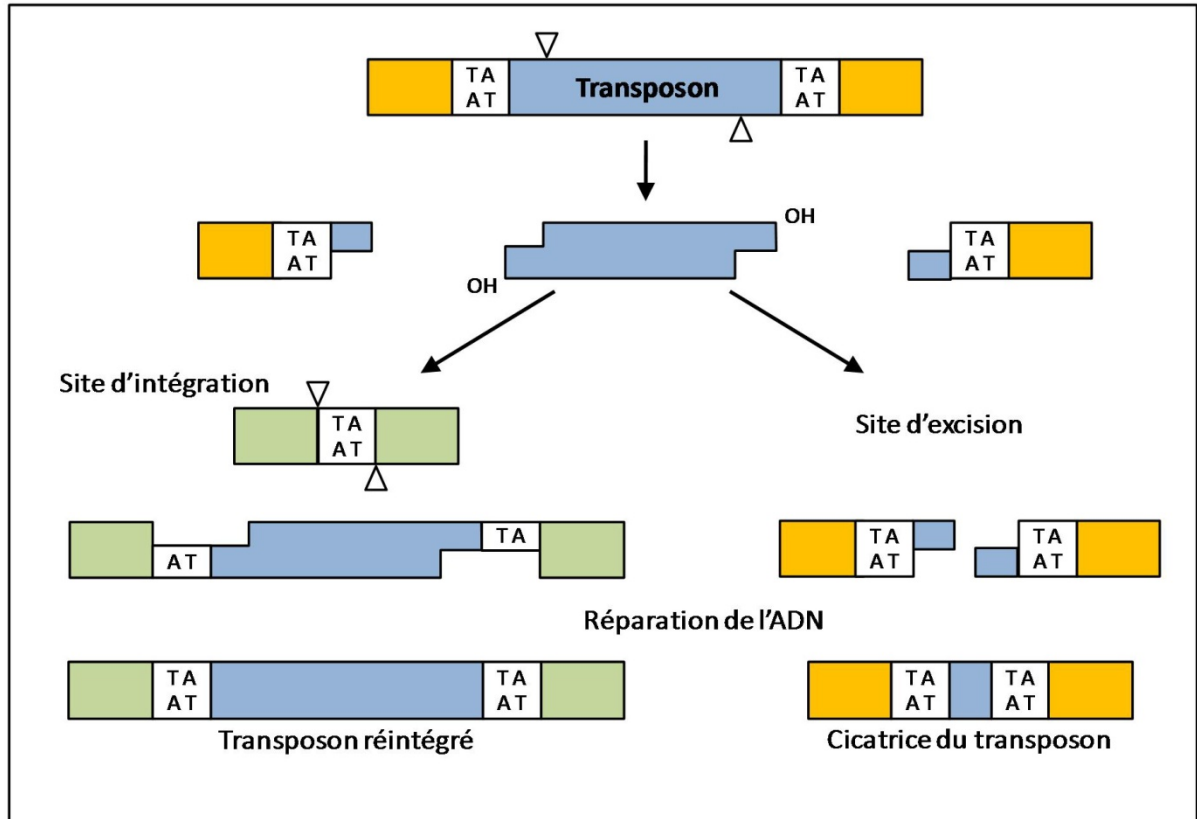


Figure 1.17. Mécanisme d'intégration d'un transposon à ADN. La transposase initie l'excision du transposon en faisant une coupure double-brin générant des bouts cohésifs et le réintègre à un dinucléotide cible TA. Le manque simple-brin au site d'intégration et la cassure double-brin au site d'excision sont réparés par la machinerie cellulaire. Après la réparation, le dinucléotide cible TA est dupliqué au site d'intégration et une cicatrice est laissée au site d'excision (Reproduit de Miskey et al. 2005).

explique les différents phénotypes observés dans la coloration mauve ou la morphologie du grain (Wessler 1988). Puisque le mécanisme de transposition de ces éléments implique des bris d'ADN double-brins, ils sont prédisposés à être responsables d'évènements de réarrangements chromosomiques. À cet effet, on avait remarqué la fréquence et l'amplitude élevée des réarrangements provoqués par la recombinaison d'éléments *FB* chez la drosophile (Collins et Rubin 1984). On a récemment identifié dans le génome du riz la présence de 3000 fragments d'un transposon (Pack-MULE) contenant les séquences de plus d'un millier de gènes, dont environ le cinquième provenait d'exons (Jiang et al. 2004). Cette étude démontre la capacité des transposons à ADN à dupliquer, transporter des séquences régulatrices et créer de nouveaux gènes. Des résultats semblent également indiquer que les transposons et certains rétrotransposons constitueraient des cibles mobiles pour la formation locale d'hétérochromatine (Slotkin et Martienssen 2007).

1.2.2. Les séquences d'ADN répétées en tandem

La classification des ADN satellites les sépare en trois classes : les satellites, les minisatellites et les microsatellites. Ces éléments répétés se distinguent par trois caractéristiques : la taille de la séquence répétée (le motif), la longueur totale de la région générée par la répétition du motif et le nombre d'éléments retrouvés par génome. La figure 1.18 illustre les différences entre ces trois types d'ADN répété en tandem. Cette catégorie d'éléments répétés ne compte que pour 5 % des 50 % des séquences d'ADN répétées présentes dans le génome humain, dont 3,5 % sont des ADN satellites (Levy et al. 2007). Deux orientations sont possibles pour des séquences répétées consécutives : l'orientation directe (tête-à-queue) dans laquelle les séquences ont la même orientation, et l'orientation inversée (tête-à-tête) où c'est le contraire. La vaste majorité des séquences répétées en tandem retrouvées dans les génomes eucaryotes est dans l'orientation directe, puisque l'orientation inversée mène à de larges réarrangements chromosomiques qui sont néfastes pour les organismes (Lobachev et al. 2000). Dans les sections suivantes, nous allons d'abord exposer les caractéristiques des trois classes

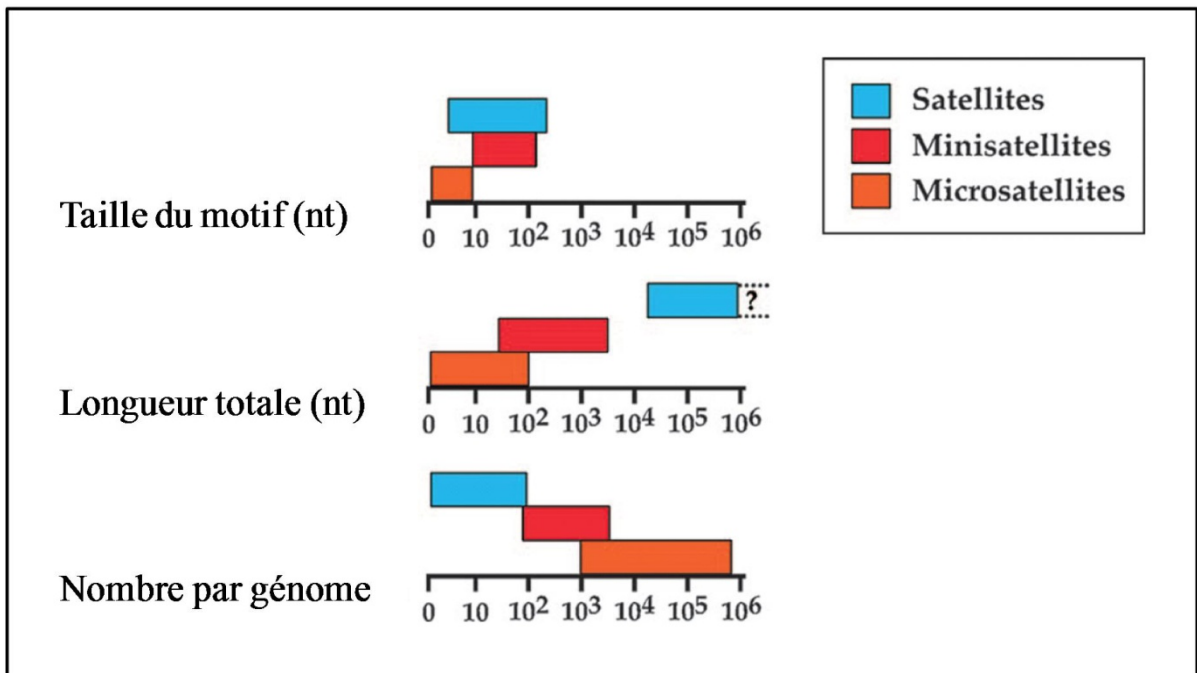


Figure 1.18. Taille des motifs, longueur et abondance des ADN satellites chez les eucaryotes. Pour chaque catégorie (satellites, minisatellites et microsatellites) la taille des motifs, la longueur totale du segment répétitif et la fréquence par génome eucaryote sont indiquées en échelle logarithmique. Les ADN satellites peuvent s'étendre sur des mégabases mais la longueur maximale est inconnue par manque d'information (point d'interrogation) (Traduit de Richard et al. 2008).

d'ADN satellites. Par la suite, les dernières sections s'attarderont aux minisatellites et microsatellites : aux mécanismes impliqués dans leur instabilité, à leurs effets biologiques ainsi qu'à leurs implications dans l'évolution des espèces.

1.2.2.1. Les ADN satellites

Historiquement, l'ADN satellite a été identifié comme une fraction qui sédimente en une bande distincte de la bande principale lors d'une centrifugation sur un gradient de chlorure de césium. Les ADN satellites sont des séquences répétées directes généralement étalées sur une longueur de plusieurs mégabases, et localisées dans les régions d'hétérochromatines péri-centromériques et télomériques (Charlesworth et al. 1994). L'unité de base répétée (le motif) est souvent riche en bases adénine et thymine (A+T) et varie en longueur de 5 nucléotides à plus de 1 kb. Les tailles de 150 à 180 pb et de 300 à 360 pb fréquemment retrouvées dans plusieurs satellites seraient le reflet des longueurs nécessaires pour l'enroulement de l'ADN autour d'un ou deux nucléosomes (Henikoff et al. 2001). L' α -satellite (171 pb) représente le principal élément structurel des régions centromériques et péri-centromériques chez l'humain (Schueler et al. 2001). Dans une espèce, on retrouve plusieurs ADN satellites différents qui sont soumis à l'influence de la conversion génique (*gene conversion*) et aux croisements non-équivalents (*unequal crossing over*) des chromosomes. Le nombre de copies d'un ADN satellite peut changer lors de croisements non-équivalents, alors que des réarrangements de séquences peuvent générer de nouveaux motifs (Ugarković et Plohl 2002). L'ADN satellite des régions télomériques est constitué d'une répétition de la séquence (TTAGGG)_n, dont la longueur totale atteint de 10 à 15 kb chez l'humain à la naissance. Elle est le site de formation d'un complexe nucléoprotéique dont la fonction est de protéger l'extrémité des chromosomes (Palm et de Lange 2008).

Des effets structuraux sur la chromatine sont imposés par la nature des ADN satellites et d'autres fonctions sont associées à leur transcription (Figure 1.19) (Plohl et al. 2008). L'étude de la

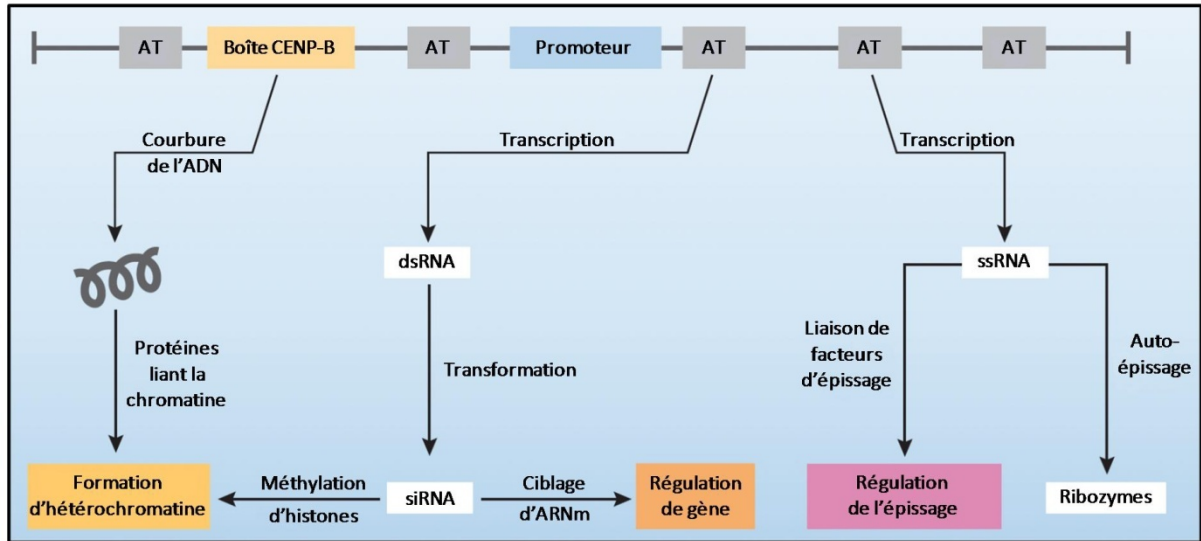


Figure 1.19. Représentation schématique d'un monomère d'ADN satellite et des processus dans lesquels il peut être impliqué. Les éléments fonctionnels communs comme les séquences riches en AT, la boîte CENP-B et un promoteur sont indiqués. Une structure hélicoïdale induite par une courbure de l'ADN, avec la liaison de protéines liant la chromatine tel CENP-B, peut contribuer à la formation d'hétérochromatine. La transcription des ADN satellites mène à la formation d'ARN double-brin (dsRNA) ou d'ARN simple-brin (ssRNA). Les dsRNA sont transformés en petits ARN d'interférence (siRNA) qui participent à la formation d'hétérochromatine et au contrôle de l'expression de gènes possédant des séquences complémentaires. Les transcrits simple-brin avec une structure secondaire *hammerhead-like* possèdent une activité de clivage auto-catalytique et agissent comme ribozymes. Des ssRNA recrutent des facteurs et régulent l'épissage (Traduit de Ugarkovic 2005).

variabilité des ADN satellites montre qu'elle n'est pas uniforme sur toute la longueur du motif, ce qui suggère des contraintes fonctionnelles sur les séquences qui pourraient être associées à des sites spécifiques de liaison à des protéines. À cet effet, la protéine CENP-B (*centromere protein-B*) lie un motif de 17 pb (CENP-B box) sur l' α -satellite (Masumoto et al. 1989). D'autres motifs de liaison ont également été identifiés sur des satellites différents. De plus, le regroupement à intervalles réguliers de séquences riches en dinucléotides AT provoque généralement la courbure de l'axe de l'hélice d'ADN, ce qui serait important pour la formation de l'hétérochromatine (Fitzgerald et al. 1994). Certaines études ont démontré que l'ADN satellite serait temporairement transcrit à des stades de développement ou exprimé différemment dans certains types cellulaires, tissus ou organes. Ce patron d'expression suggère qu'ils ont une fonction régulatrice, qui reste cependant à définir. Des promoteurs et des éléments régulateurs des ARN polymérase II et III ont été identifiés dans certains ADN satellites. Il a été suggéré que la séquestration de facteurs d'épissage par un transcrite du *satellite III* pourrait réguler l'épissage des ARNm durant un stress (Valgardsdottir et al. 2005). Une autre étude a montré qu'un transcrite d'ADN satellite pouvait adopter une structure secondaire semblable aux ribozymes *hammerhead* et effectuer des clivages en *cis* ou en *trans* (Ferbeyre et al. 1998). Finalement, des études ont rapporté que des transcrits d'ADN satellites étaient la source de petits ARN de 20 à 25 pb (*small interfering RNA*, siRNA) impliqués dans l'assemblage de l'hétérochromatine médié par l'interférence ARN (*RNA interference*, RNAi) (Volpe et al. 2002) et possiblement dans la régulation post-transcriptionnelle des gènes. C'est d'ailleurs la présence de séquences semblables à celle de l' α -satellite dans les régions 5'-UTR et 3'-UTR de plusieurs ARNm qui suggère une régulation de l'expression de ces gènes par des dérivés siRNA de l' α -satellite (Li et Kirby 2003).

1.2.2.2. Les ADN minisatellites et microsatellites

Les minisatellites et les microsatellites se distinguent principalement de l'ADN satellite par des motifs plus courts et une localisation dans des régions d'euchromatine. Les minisatellites (aussi

appelés VNTR, *variable number of tandem repeats*) ont un motif répété en tandem, généralement riche en bases guanine et cytosine (G+C), dont la taille varie de 10 à une centaine de paires de bases. Les microsatellites (aussi appelés SSR, *simple sequence repeats*) consistent en une répétition d'un motif de 1 à 9 pb. Pour le moment, il ne semble pas y avoir de consensus sur la nature des deux types d'éléments; certains auteurs ne considèrent pas les répétitions d'un même nucléotide en tant que microsatellite, alors que d'autres positionnent la séparation des deux types d'éléments entre 6 et 10 pb pour le motif (Debrauwere et al. 1997).

Au cours des dernières années, plusieurs études ont regardé la distribution des microsatellites dans les génomes eucaryotes. Cependant, comme les chercheurs ne s'entendent pas sur une définition commune et que des algorithmes de recherche utilisant des paramètres différents ont été employés, des résultats très variables sur le nombre de microsatellites ont été obtenus (Richard et al. 2008). L'analyse des génomes eucaryotes a tout de même permis d'obtenir d'autres types d'informations. Chez toutes les espèces étudiées, les trinuécléotides et les hexanucléotides sont les plus abondants dans les exons, alors que des patrons de distribution spécifiques aux espèces sont retrouvés dans les régions intergéniques et les introns. Chez *Saccharomyces cerevisiae*, les microsatellites ne sont généralement pas retrouvés dans les gènes, à l'exception de trinuécléotides et d'hexanucléotides qui ne changent pas le cadre de lecture du gène (Tóth et al. 2000). L'analyse des acides aminés encodés par ces éléments montre une surreprésentation de certains : glutamine, acide glutamique, asparagine, acide aspartique et sérine. De plus, les répétitions d'acides aminés polaires ou chargés sont significativement associés aux facteurs de transcription et aux protéines kinases, alors que les répétitions de sérines sont associées aux transporteurs membranaires (Mar Albà et al. 1999).

Chez l'humain, la densité des microsatellites est relativement uniforme sur tous les chromosomes, variant de 10 à 40 SSR/Mb pour les différentes tailles (Subramanian et al. 2003a; IHGSC 2001). Les trinuécléotides AGC et AGG et CCG sont les éléments les plus abondants dans les séquences codantes (Subramanian et al. 2003b). Une étude sur les séquences codantes des primates a

montré que certains trinuécléotides sont enclins à la prolifération, par exemple les répétitions de CGG, CCG, CAG et GAA codants pour des segments protéiques polyalanine, polyglycine, polyproline, polyglutamine et polylysine (Borstnik et Pumpernik 2002). Comme nous le verrons dans une section suivante, l'expansion anormale de certains trinuécléotides retrouvés dans les gènes est la cause de plusieurs maladies et désordres neurologiques (Orr et Zoghbi 2007).

Les minisatellites ont été beaucoup moins étudiés que les microsatellites. Deux études ont rapporté l'analyse complète du génome de *S. cerevisiae* et confirmé la présence des minisatellites (Verstrepen et al. 2005; Richard et Dujon 2006). Celles-ci ont identifié respectivement 44 et 49 gènes contenant des minisatellites, où les motifs sont généralement uniques et dont la taille est un multiple de trois. Onze VNTR furent également identifiés dans des régions intergéniques (Louis et al. 1994). Parmi les gènes contenant des minisatellites, 50 à 60 % de ceux-ci encodent des protéines de la paroi cellulaire ou d'autres ayant une implication dans son propre métabolisme. Les minisatellites qu'on y retrouve encodent principalement des acides aminés sérines et thréonines, générant ainsi des sites de O-mannosylation, une modification importante pour le maintien des protéines à la paroi cellulaire (Ecker et al. 2003). L'étude du génome de plusieurs levures montre également qu'un minisatellite différent peut-être retrouvé dans le gène orthologue chez d'autres espèces (Richard et Dujon 2006). Par exemple, on retrouve dans le gène *PRY2* six répétitions d'un motif de 18 pb chez *S. cerevisiae*, neuf répétitions d'un motif différent de 15 pb chez *Saccharomyces paradoxus*, un pseudo-minisatellite chez *Debaryomyces hansenii* et aucun minisatellite chez *Candida glabrata*. Ceci suggère que l'évolution des minisatellites est plus rapide que celle des gènes qui les contiennent ou que ces éléments ont la capacité de s'intégrer à un site spécifique comme le font les transposons. Finalement, *C. glabrata* possède de très longs minisatellites, souvent retrouvés dans des protéines impliquées dans l'adhésion cellule-cellule, faisant de ces VNTR des éléments susceptibles d'influencer la pathogénicité (Thierry et al. 2008).

À partir de l'analyse des séquences des chromosomes 21 et 22, on estime à environ 6000 le nombre de minisatellites retrouvés dans le génome humain, desquels une majorité (75 %) semble montrer du polymorphisme (Denoeud et al. 2003). La distribution des VNTR dans le génome humain montre un biais pour les régions subtélomériques (Amarger et al. 1998). Une catégorie particulière de minisatellites appelés "hypermutable" montre un taux de mutation moyen plus grand que 0,5 % dans les cellules germinales. L'analyse des séquences adjacentes à ces minisatellites a révélé qu'ils étaient localisés près d'un point chaud (*hot spot*) de recombinaison méiotique situé en amont (Bois 2003). Parce qu'ils contiennent des variations internes, ces éléments ont constitué une source importante d'informations sur les mécanismes menant à l'instabilité des minisatellites. Comme pour les microsatellites, des minisatellites sont retrouvés dans les gènes et nous verrons dans une section à venir que la variation de taille des minisatellites est également associée à des maladies et des variations du comportement humain.

Depuis la découverte du premier minisatellite par Wyman et White (1980), qui ont identifié un locus ayant une variabilité de longueur pour des individus avec différents degrés de parenté, le polymorphisme extrême des minisatellites a été utilisé à plusieurs niveaux comme marqueur génétique pour : l'identification d'individus, la localisation de gènes et des études populationnelles (Bois et Jeffreys 1999). Les microsatellites, dont le premier fut caractérisé par Weller et collaborateurs (1984) comme une répétition variable de la séquence (GGAT)_n dans le gène de la myoglobine humaine, ont également été utilisés et ensuite adoptés pour ce genre d'études puisqu'ils sont plus abondants que les minisatellites (Richard et al. 2008). Par génération, les séquences d'ADN des eucaryotes supérieurs ont un taux de mutation approximatif de 10^{-10} à 10^{-11} par nucléotide (Drake et al. 1998), alors que le taux de mutation des microsatellites est supérieur par plusieurs ordres de grandeur; évalué entre 10^{-3} et 10^{-4} par locus (Weber et Wong 1993). La section qui suit exposera les mécanismes moléculaires qui ont été proposés pour expliquer l'instabilité des micro- et minisatellites.

1.2.2.3. Les mécanismes impliqués dans l'instabilité des mini- et microsatellites

À partir d'études sur l'instabilité des minisatellites humains au cours de la méiose, il a d'abord été proposé que l'expansion et la contraction des minisatellites se faisait par recombinaison homologue, alors que la variation de taille des microsatellites était le résultat de glissement entre les brins d'ADN au cours de sa réplication en phase S. Des études subséquentes ont montré que ces mécanismes n'étaient pas exclusifs à chaque type d'élément répétitif (Richard et al. 2008). L'étude des microsatellites a permis d'établir que l'instabilité observée au niveau des séquences répétées était médiée par les processus de réplication, de réparation et de recombinaison de l'ADN. Celle-ci est grandement influencée par les propriétés biophysiques des SSR qui permettent l'adoption de structures non-B de l'ADN (Wells et al. 2005). Les mécanismes de réparation de l'ADN ne seront pas abordés dans la section suivante puisque peu d'informations sont disponibles sur l'influence qu'ils ont dans l'instabilité des microsatellites.

1.2.2.3.1 Les structures secondaires de l'ADN

Un grand nombre de séquences répétées simples peuvent exister sous deux types de conformations. La plupart du temps elles adoptent la structure classique de l'ADN de type B qui repose sur les interactions Watson-Crick (WC) A•T et G•C pour former une double-hélice d'ADN caractérisée par un sillon majeur et un autre mineur. Dans certaines situations, ces séquences permettent également l'adoption de structures secondaires non-B de l'ADN, desquelles une dizaine ont jusqu'à maintenant été identifiées. Celles-ci incluent les structures glissantes/épingle à cheveux (*hairpin*), cruciformes, ADN-Z, ADN triplexe et tetraplexe (Figure 1. 20). L'appariement des bases des structures cruciformes et ADN-Z sont de type WC, alors qu'on peut retrouver des appariements non-WC dans les structures glissantes/épingle à cheveux, ADN triplexe et tetraplexe. Ces structures transitoires sont formées au niveau d'une séquence d'ADN spécifique par l'accumulation de surenroulement négatif généré en partie par la transcription, la liaison de protéines et d'autres facteurs.


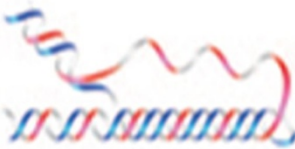
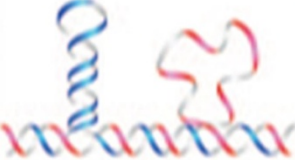
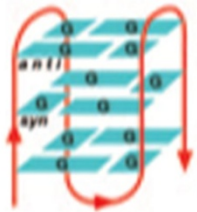

Nom	Conformation	Propriété de la séquence	Séquence
Structure cruciforme		Répétition inversée	$\begin{array}{c} \text{TCGGTACCGA} \\ \text{AGCCATGGCT} \end{array}$
ADN triplexe		$(R \cdot Y)_n$ Répétition miroir	$\begin{array}{c} \text{AAGAGG} \text{GGAGAA} \\ \text{TTCTCC} \text{CCTCTT} \end{array}$
Structure glissante/épingle à cheveux		Répétition directe	$\begin{array}{c} \text{TCGGTTCGGT} \\ \text{AGCCAAGCCA} \end{array}$
ADN tetraplexe		Oligo (G) _n	AG ₃ (T ₂ AG ₃) ₃ simple-brin
ADN-Z		$(YR \cdot YR)_n$	$\begin{array}{c} \text{CGCGTGC GTGTG} \\ \text{GCGCACGCACAC} \end{array}$

Figure 1.20. Conformations non-B de l'ADN impliquées dans des réarrangements. Le nom et une représentation schématique de diverses conformations non-B de l'ADN sont présentés. Les propriétés requises pour le motif ou au niveau du type de répétition, ainsi qu'un exemple sont indiqués (Traduit de Bacolla et Wells 2004).

Elles seraient responsables d'au moins 46 maladies génétiques chez l'humain impliquant des réarrangements génomiques et d'autres types de mutations (Bacolla et Wells 2004).

La répétition directe d'une séquence peut mener à la formation d'une structure glissante/épingle à cheveux. La réplication et la transcription de l'ADN provoquent la séparation des brins du duplexe, ce qui donne à l'ADN simple-brin l'opportunité de se replier et de former de nouveaux appariements avec des bases de son propre brin, d'où la structure en épingle à cheveux. Par exemple, la séquence répétée (CTG)_n associée à la dystrophie myotonique de type 1 (DM1) forme une structure glissante/épingle à cheveux qui est dépendante de la longueur, de la pureté et de l'orientation de la séquence répétée (Kang et al. 1995; Tam et al. 2003). Une autre étude, sur le tétranucléotide répété (CCTG)_n associé à la dystrophie myotonique de type 2 (DM2), a montré que le brin formé de la séquence complémentaire (CAGG)_n avait la propriété de former une structure épingle à cheveux plus stable que sa séquence complémentaire (Dere et al. 2004). Finalement, les ADN polymérase eucaryotes alpha, delta et epsilon peuvent s'arrêter aux structures en épingle à cheveux, ce qui peut alors provoquer leur glissement et éventuellement de l'instabilité génomique (Kamath-Loeb et al. 2001).

Les structures cruciformes sont produites lorsque les séquences répétées sont dans une orientation inversée, formant ainsi un palindrome. Dans ce cas, les ADN simple-brin générés lors de la réplication ou transcription se replient au centre de symétrie pour reconstituer un ADN double-brin intramoléculaire coiffé d'une boucle simple-brin. Il en résulte une structure où deux structures épingles à cheveux sont situées à la même position sur chaque brin. Ces séquences sont répliquées à un taux plus lent que des séquences d'ADN normales dans *E. coli* (Lindsey et Leach 1989). L'analyse de plusieurs séquences palindromique riches en A+T sur les chromosomes humains 11 et 22 a démontré qu'une cassure d'ADN double-brin survient au centre de ces séquences et qu'elle serait possiblement responsables de la translocation fréquemment observée entre ces chromosomes (Kurahashi et Emanuel 2001). L'étude *in-vitro* de ces séquences a d'ailleurs montré qu'elles avaient la

capacité d'adopter des structures cruciformes qui ont pu être visualisées en microscopie atomique (Kurahashi et al. 2004).

La structure ADN triplexe (triple-hélice, H-DNA) est adoptée par une séquence continue de plusieurs purines ou pyrimidines et favorisée par la présence d'un centre de symétrie. Dans cette structure, le brin de purines est apparié au brin de pyrimidines dans une conformation de type B, et fait en plus des interactions non-WC avec un troisième brin qui se localise dans le sillon majeur de l'ADN. Une étude a montré que la répétition d'un dinucléotide $(GA)_n$ avait la capacité d'adopter une structure ADN triplexe qui bloquait la synthèse de l'ADN *in-vitro* (Baran et al. 1991). De plus, il a été rapporté que de longues répétitions du trinucleotide $(GAA)_n$ retrouvé dans l'intron 1 du gène frataxine causaient l'ataxie de Friedrich en réduisant la transcription du gène par l'adoption de cette structure (Grabczyk et Usdin 2000).

Une structure à quatre brins appelée ADN tetraplexe (G-quadruplex, G-quartet) est composée de tétrades de guanine empilées les unes sur les autres et favorisée par la présence de quatre répétitions de trois guanine ou plus dans la séquence. Il a été suggéré que la répétition du trinucleotide $(CGG)_n$ dans le gène *FMRI* qui est associé au syndrome du X fragile, pouvait adopter une telle structure (Fojtík et Vorlícková 2001), bloquer la réplication de l'ADN (Usdin et Woodford 1995) et inhiber la synthèse de la protéine encodée par le gène dans lequel la séquence répétée est retrouvée (Darnell et al. 2001). Des minisatellites riches en guanine ont également été identifiés dans les régions promotrices des gènes de l'insuline humaine (Catasti et al. 1996) et de la β -globine du poulet (Howell et al. 1996). La nature de ces minisatellites permet l'adoption d'une structure ADN tetraplexe menant respectivement au blocage de la transcription ou de la synthèse de l'ADN.

Parce que les bases purines peuvent adopter une conformation en *syn*, les séquences qui sont constituées d'une alternance de purines et pyrimidines (ex. CG ou CA) peuvent être converties en une structure ADN-Z (Rich et Zhang 2003). Des études ont montré que des séquences ayant le potentiel d'adopter une structure ADN-Z étaient associées à des sites de recombinaison chez les eucaryotes

(Wahls et al. 1990) et à des évènements de glissement de la polymérase chez *E. coli* (Freund et al. 1989). Ces structures sont fréquemment retrouvées près des sites d'initiation de la transcription (Schroth et al. 1992). Cette structure retrouvée en amont d'un gène permettrait de garder l'ADN ouvert pour la liaison de facteurs de transcription et l'initiation de la transcription par l'ARN polymérase, suggérant ainsi le déclenchement de la transcription par la formation d'ADN-Z (Liu et al. 2001).

1.2.2.3.2. Les mini- et microsatellites et la réplication de l'ADN

Il est généralement accepté que les procédés métaboliques qui incluent la synthèse d'ADN, tels que la réplication, la réparation et la recombinaison, sont impliqués dans la variation de la taille des mini- et microsatellites. Cette instabilité est dépendante de facteurs comme la localisation chromosomique, la longueur et la pureté de la séquence répétée. La réplication de l'ADN est cependant considérée comme la première étape qui peut faire intervenir les autres procédés par la suite (Wells et al. 2005). Le mécanisme du glissement de la polymérase au cours de la réplication de l'ADN (*replication slippage model*, Figure 1.21) a d'abord été proposé pour expliquer l'instabilité des microsatellites et des expériences récentes semblent montrer qu'il s'appliquerait également aux minisatellites. Dans ce modèle, un glissement de l'ADN polymérase sur le brin retardé peut survenir lorsque celle-ci rencontre une structure secondaire qui la force à s'arrêter. Une dissociation partielle du brin nouvellement synthétisé se produit alors et peut permettre la formation d'une structure secondaire sur ce brin ou sur le brin matrice dans une région d'ADN simple-brin. La présence de cette structure provoque un ré-appariement hors cadre du brin naissant qui se traduit par un glissement de la polymérase lorsqu'elle reprend la synthèse. Il en résulte une expansion de la séquence répétée lorsque la structure secondaire se forme sur le brin naissant, alors qu'une contraction est observée lorsque cette dernière est formée sur le brin matrice. L'analyse des mutations observées sur les microsatellites montre que celles-ci varient presque toujours par l'addition ou la suppression d'un seul motif répété, suggérant un

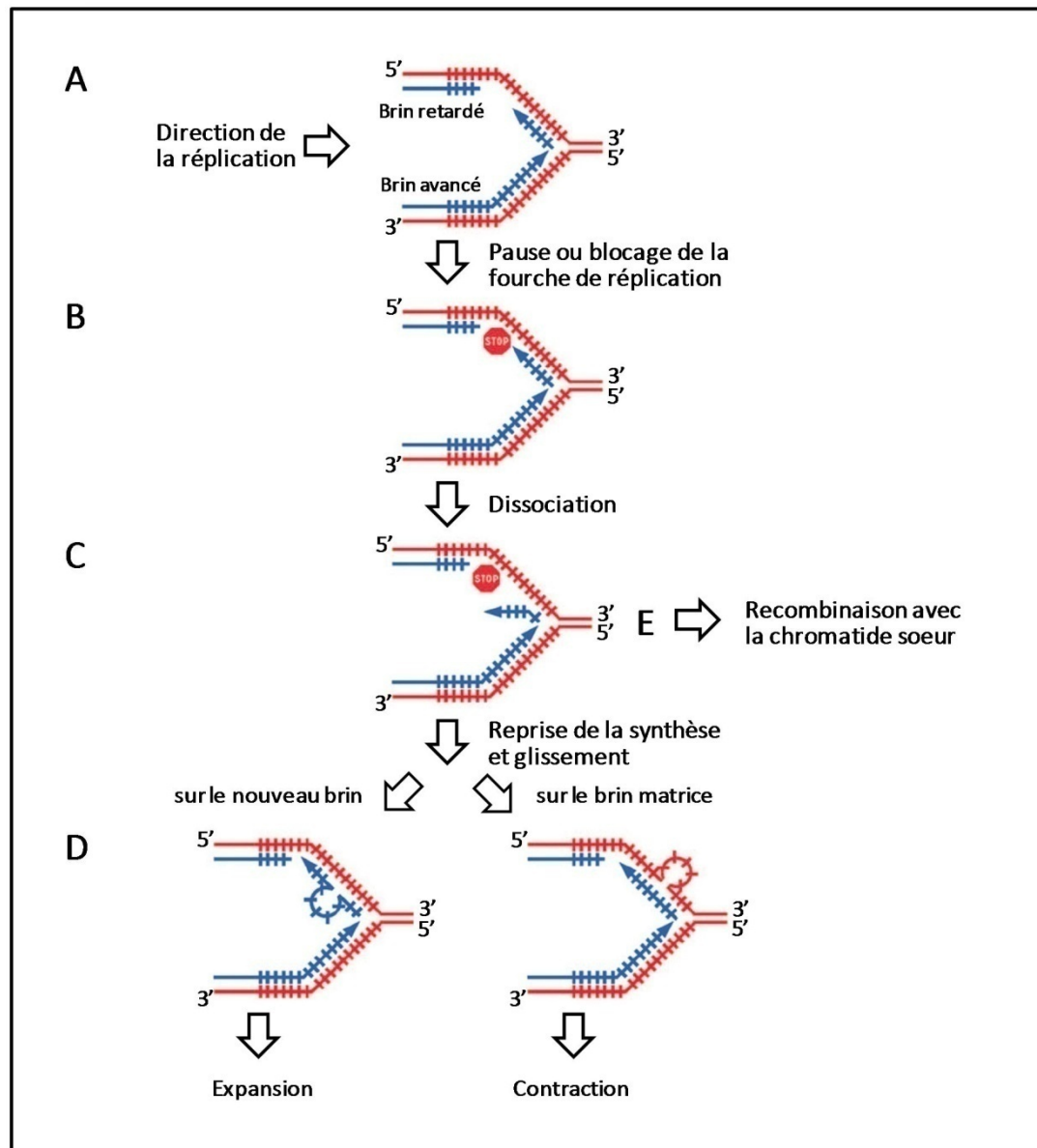


Figure 1.21. Le modèle de réplication avec glissement pour les séquences d'ADN répétées en tandem. Les brins matrices sont illustrés en rouge et les nouveaux brins synthétisés en bleu. Durant la réplication d'une séquence contenant une répétition (A), la machinerie de réplication peut faire une pause sur le brin retardé lorsqu'elle rencontre une structure secondaire ou une lésion (B). (C) Une dissociation partielle du brin retardé peut mener à un glissement de la réplication lorsqu'elle reprend (D), causant une expansion ou une contraction du segment répété, en fonction du brin sur lequel le glissement a lieu. (E) Alternativement, la dissociation partielle du brin retardé peut passer outre la lésion par recombinaison homologue avec la chromatide sœur (Traduit de Richard et al. 2008).

modèle par étapes (*stepwise mutation model*) pour l'évolution des SSR (Zhu et al. 2000). Alternativement, la dissociation partielle du brin nouvellement synthétisé peut mener à une recombinaison homologue avec la chromatide sœur comme nous le verrons dans une prochaine section.

1.2.2.3.2.1. Les effets de la réplication de l'ADN sur les microsatellites

Comme nous l'avons vu dans une section précédente, plusieurs structures secondaires de l'ADN ont la capacité d'arrêter sa réplication, la conséquence pouvant être observée dans la variation de taille des microsatellites. L'étude de différents microsatellites a permis d'identifier des facteurs qui influencent l'activité mutationnelle. Chez *S. cerevisiae*, il a été montré que l'instabilité du microsatellite (GT)_n dépendait de la longueur de la région répétée, avec un taux de mutation variant d'un facteur 500 entre (GT)₁₅ et (GT)₁₀₅, la plus longue séquence étant la plus instable (Wierdl et al. 1997). L'expansion des trinuécléotides est également locus spécifique, ce qui a été observé lors de l'analyse de familles affectées par la maladie de Huntington (répétitions de (CAG)_n). L'étude a noté que du polymorphisme était retrouvé au locus associé à la dystrophie myotonique de type 1 (répétitions de (CTG)_n), alors que le locus associé à l'ataxie spinocérébelleuse de type 1 (répétitions de (CAG)_n) et d'autres microsatellites du génome étaient stables (Goellner et al. 1997). De plus, la distance entre la séquence répétée et l'origine de réplication jouent un rôle important dans l'instabilité, celles retrouvées à proximité montrant plus d'instabilités que celles qui en sont éloignées (Dere et al. 2004). Cette même étude a démontré que l'orientation du microsatellite associé à la dystrophie myotonique de type 2 (DM2) avait également une influence sur son instabilité, la répétition du tétranuécléotide (CAGG)_n la favorisant par rapport à son complément (CCTG)_n. À l'opposé, pour le microsatellite associé à la dystrophie myotonique de type 1 (DM1), le brin formé de la répétition de la séquence (CTG)_n forme une structure plus stable que la séquence (CAG)_n (Kang et al. 1995). Finalement, la stabilité des

trinuécléotides est augmentée par la présence d'interruptions dans la séquence répétée. Il a été observé que l'interruption par des trinuécléotides CAT et AGG des répétitions de trinuécléotides (CAG)_n et (CGG)_n associées respectivement à l'ataxie spinocérébelleuse de type 1 (SCA1) et au syndrome du X fragile (FRAXA) diminuait la formation de structures glissantes (Pearson et al. 1998). On retrouve justement des interruptions AGG au locus FRAXA dans la population normale, alors que les patients atteints du syndrome du X fragile montrent une répétition ininterrompue de (CGG)_n (Dombrowski et al. 2002).

1.2.2.3.2.2. Les effets de la réplication de l'ADN sur les minisatellites

En comparaison avec l'abondante littérature qui est retrouvée sur l'instabilité des microsatellites, seulement quelques études ont été faites pour étudier celle des minisatellites. On avait déjà remarqué qu'en plus du taux élevé de réarrangement du minisatellite MS32 dans les cellules germinales (0,8 %), on observait également de l'instabilité dans les cellules sanguines (< 0,06 %). L'analyse des rares mutants retrouvés dans le sang a révélé l'ajout ou la suppression simple de motifs sur toute la longueur du minisatellite (Jeffreys et Neumann 1997). Puisqu'aucune évidence de recombinaisons interalléliques ne fut trouvée, un mécanisme de glissement de la polymérase ou de recombinaison avec la chromatide sœur fut proposé pour expliquer les mutations. Plusieurs études ont utilisé différentes souches de levures mutantes pour la réplication de l'ADN dans l'étude des minisatellites. La déstabilisation des microsatellites dans des levures mutantes pour *RAD27*, qui est impliqué dans la transformation des fragments d'Okasaki sur le brin retardé, avait déjà été démontrée. Une étude a reproduit cet effet et montré qu'un minisatellite de 20 pb était aussi déstabilisé dans une souche Δ *RAD27*, et également dans une souche mutante thermosensible *pol-3t*, qui affiche un taux de synthèse réduit du brin retardé (Kokoska et al. 1998). À partir de ces résultats les auteurs ont suggéré le mécanisme de glissement de la polymérase pour expliquer la variation de taille des minisatellites. Une autre étude a proposé qu'en plus de ce mécanisme, un autre impliquant un bris d'ADN double-

brin et une réparation par recombinaison intra-allélique était possiblement impliqué (Lopes et al. 2002). Finalement, l'étude de minisatellites distincts a permis d'établir que même s'il y a une corrélation positive entre la longueur du minisatellite et l'instabilité, c'est la structure interne du minisatellite qui est le facteur déterminant (Maleki et al. 2002). Ceci est démontré par l'observation du minisatellite MS205-2.7 qui est plus stable que le CEB1-1.8, même si le premier est plus long et montre un plus grand nombre de répétitions que le second.

1.2.2.3.3. Le réarrangement des mini- et microsatellites lors de la recombinaison homologue

Les premières études sur les minisatellites humains ont établi que plusieurs d'entre eux sont très variables (Jeffreys et al. 1985). D'autres ont par la suite montré que cette variabilité est le résultat d'évènements ayant lieu dans les cellules germinales (Buard et Vergnaud 1994). Il avait déjà été proposé que les évènements mutationnels menant au gain ou à la perte de plus de 200 motifs étaient possiblement le résultat de recombinaisons méiotiques (Jeffreys et al. 1988). L'analyse des mutations observées dans les cellules germinales a révélé des caractéristiques typiques du processus mutationnel (Buard et Vergnaud 1994; Jeffreys et al. 1994). Premièrement, une polarité est observée dans les réarrangements, c'est-à-dire que les motifs sont gagnés ou perdus préférentiellement à une extrémité du minisatellite. Ensuite, la présence de réarrangements complexes suggère que les nouveaux allèles mutants sont produits à partir de l'information génétique retrouvée sur la chromatide sœur (évènement intra-allélique) et sur le chromosome homologue (évènement interallélique). Troisièmement, l'échange des marqueurs sur les séquences flanquantes est rare, indiquant que les évènements de recombinaison n'impliquent pas le croisement des brins (*crossover*) lors de la résolution. De plus, un site chaud (*hot spot*) de recombinaison a été identifié en amont du minisatellite humain MS32 et démontré comme étant responsable de son instabilité méiotique (Jeffreys et al. 1998). Ceci indique que l'hypervariabilité est retrouvée à proximité de ce type d'élément reconnu pour causer des bris d'ADN double-brin qui sont des instigateurs puissants de recombinaisons (Bishop et Schiestl 2000).

Cela explique par le fait même la polarité mutationnelle observée au niveau des minisatellites puisque les bris d'ADN double-brin sont retrouvés à une extrémité du minisatellite. Par ailleurs, un taux de mutation plus élevé des minisatellites a été observé dans les cellules germinales de la population exposée aux radiations après l'accident de Tchernobyl, suggérant que l'irradiation a induit des réarrangements au niveau des minisatellites, possiblement à la suite de bris d'ADN double-brin qui ont stimulé la recombinaison homologue (Dubrova et al. 1996).

S. cerevisiae s'est révélé un outil puissant pour approfondir l'étude de la recombinaison dans l'instabilité des minisatellites humains. Le minisatellite MS1 a d'abord été introduit dans le génome d'une levure haploïde. Après avoir observé une variation fréquente de la taille du minisatellite dans cette souche, les auteurs ont conclu que le réarrangement entre des chromosomes homologues n'était pas essentiel pour générer la variabilité des minisatellites (Cederberg et al. 1993). L'instabilité des minisatellites MS32 et CEB1 dans la levure a confirmé que ces derniers étaient plus instables durant la méiose que durant la mitose, et montré la perte ou le gain de motifs par des événements de recombinaisons intra- et interalléliques (Appelgren et al. 1997; Debrauwère et al. 1999). De plus, il a été rapporté que l'instabilité méiotique du minisatellite CEB1 repose sur la présence de l'endonucléase Spo11 qui initie la recombinaison en générant un bris d'ADN double-brin et sur l'activité de Rad50p qui est impliquée dans sa réparation. Une étude a présenté des évidences que de longues séquences palindromiques (> 50 pb) peuvent générer des bris d'ADN double-brin à haute fréquence durant la méiose chez la levure. La formation de ces cassures dépend de la composition en nucléotides, de la position et de la longueur des séquences répétées inversées (Nasar et al. 2000). Il a finalement été démontré que la protéine Rad1, qui est impliquée dans l'enlèvement des extrémités non homologues lors de la recombinaison, augmentait la fréquence des expansions de minisatellites (Jauert et al. 2002).

Plusieurs auteurs ont également utilisé la levure *S. cerevisiae* pour comprendre le rôle des bris d'ADN double-brin et de la recombinaison dans l'instabilité des microsatellites. Durant la croissance mitotique, il a été rapporté que de longues séquences répétées de (CTG)₁₃₀₋₂₅₀ constituaient des sites

fragiles et induisaient des bris d'ADN double-brin longueur-dépendants, au niveau du microsatellite ou tout près, qui stimulaient ensuite la recombinaison (Freudenreich et al. 1998). Cette étude a également démontré que la recombinaison était diminuée dans des souches mutantes pour les protéines Rad50, Rad52 et Rad1, toutes impliquées pour la réparation du bris d'ADN double-brin. On a montré que l'instabilité d'un microsatellite composé du trinuécléotide $(CAG)_n$ était plus grande durant la méiose que durant la mitose, lorsque ce dernier était inséré près d'un point chaud de recombinaison (*hot spot*) (Jankowski et al. 2000). À l'opposé, l'utilisation de courtes séquences répétées de $(CAG)_n$ ou l'interruption du microsatellite par un autre motif ne permet pas de reproduire cette instabilité (Schweitzer et al. 2001). Tout comme pour les minisatellites, la plupart des réarrangements méiotiques dépendent de la présence de l'endonucléase Spo11 (Jankowski et Nag 2002). Des expériences menées avec de longues séquences répétées du trinuécléotide $(CAG)_{98-255}$ n'ont cependant pas permis d'associer ces microsatellites à des points chauds (*hot spot*) de recombinaison méiotique dans la levure (Richard et al. 2003). Les recherches sur l'instabilité des microsatellites ne semblent donc pas indiquer de différences majeures par rapport aux minisatellites.

Le modèle de réparation des bris d'ADN double-brin qui fut d'abord proposé chez la levure implique la résolution des deux structures à quatre brins appelées "jonctions de Holliday". Le résultat de cette conversion génique est associé au croisement des chromatides (*crossover*) dans 50 % des cas (Figure 1.22A). Cependant, les séquences d'ADN répétées en tandem sont généralement associées à des conversions géniques qui n'impliquent pas le croisement des chromatides. Ainsi, un autre modèle qui n'implique pas ces structures a été proposé pour expliquer ces recombinaisons : le modèle d'appariement des brins dépendant de la synthèse (*synthesis-dependent strand annealing*, SDSA, Figure 1.22B). Dans ce modèle, l'étape initiale implique également l'invasion d'une séquence homologue par l'extrémité 3' au site de cassure double-brin pour initier la synthèse d'ADN. À la différence du premier modèle, celui-ci propose que les brins nouvellement synthétisés se dissocient de

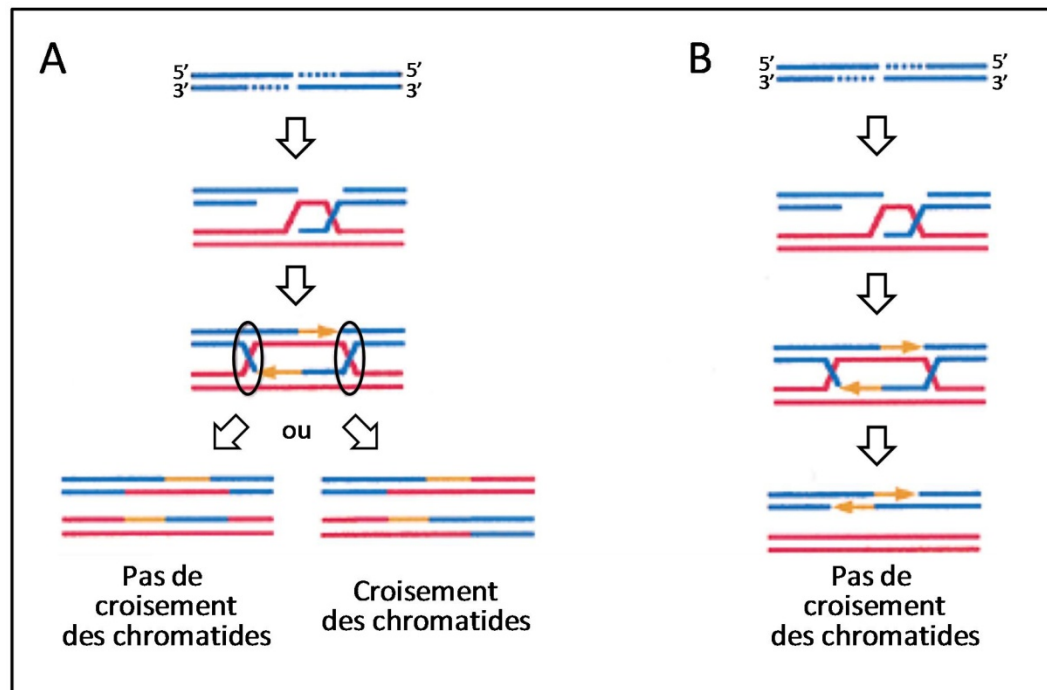


Figure 1.22. Modèles de recombinaison. La molécule d'ADN brisée (donneuse) est illustrée en bleu et la molécule matrice (receveuse) en rouge, l'ADN nouvellement synthétisé est en orange. Dans les deux modèles, la conversion génique est initiée par un bris double-brin, et l'invasion de l'ADN matrice homologue est initiée par l'extrémité 3'. (A) Modèle impliquant la résolution de deux structures à quatre brins appelées "jonctions de Holliday" (cercle noir) résultant en une conversion génique associée à un croisement des chromatides dans 50 % des cas. (B) Modèle SDSA (*synthesis-dependent strand annealing*) dont la résolution est accomplie par une dissociation des deux brins nouvellement synthétisés de l'ADN matrice et leur appariement ne produisant pas de croisement des chromatides (Traduit de Richard et Pâques 2000).

leur brin matrice et s'apparient ensuite ensemble sans croisement des chromatides (Richard et Pâques, 2000).

Un modèle unifié, dérivé du modèle SSDA, a été proposé pour expliquer les réarrangements méiotiques et mitotiques des microsatellites pendant la conversion génique (Richard et al. 2003). Celui-ci permet d'accommoder quatre caractéristiques importantes de la conversion génique associées aux séquences d'ADN répétées en tandem. Premièrement, la formation d'une cassure d'ADN double-brin dans une des séquences répétées. Ensuite, l'importance des structures secondaires de l'ADN pour générer l'instabilité de l'élément. Troisièmement, l'absence d'évidences pour des échanges par croisement de chromatides. Finalement, l'absence de variations au niveau de la séquence matrice utilisée pour réparer la cassure. Le modèle s'explique par les étapes suivantes (Figure 1.23). Suite à la formation d'une cassure double-brin dans une séquence répétée, la conversion génique est initiée par l'invasion d'une séquence homologue, par un ou les deux brins d'ADN au niveau de la cassure, suivie de la synthèse d'ADN. De manière indépendante sur chacun des brins, cette synthèse peut être directe ou associée au glissement de la polymérase suite à la formation de structures secondaires. Les événements de glissement vont provoquer une expansion de la séquence répétée s'ils surviennent sur le brin nouvellement synthétisé, ou une contraction s'ils ont lieu sur le brin matrice. Alternativement, dans une situation où la synthèse aura été directe sur les deux brins, ceux-ci pourront se dissocier et se ré-apparier dans le bon cadre, ou hors-cadre menant à une expansion ou une contraction. Plusieurs étapes de ce modèle permettent donc de générer de la variabilité sur les séquences d'ADN répétées.

1.2.2.4. La régulation de l'activité des gènes par les séquences d'ADN répétitives

Dans les sections antérieures, nous avons vu que les séquences d'ADN répétées en tandem, principalement les microsatellites, jouent un rôle dans plusieurs phénomènes biologiques. La distribution de certains dans les régions centromériques ou télomériques, ainsi que l'adoption de structures secondaires de l'ADN par d'autres suggèrent leur implication dans la structure des chromosomes

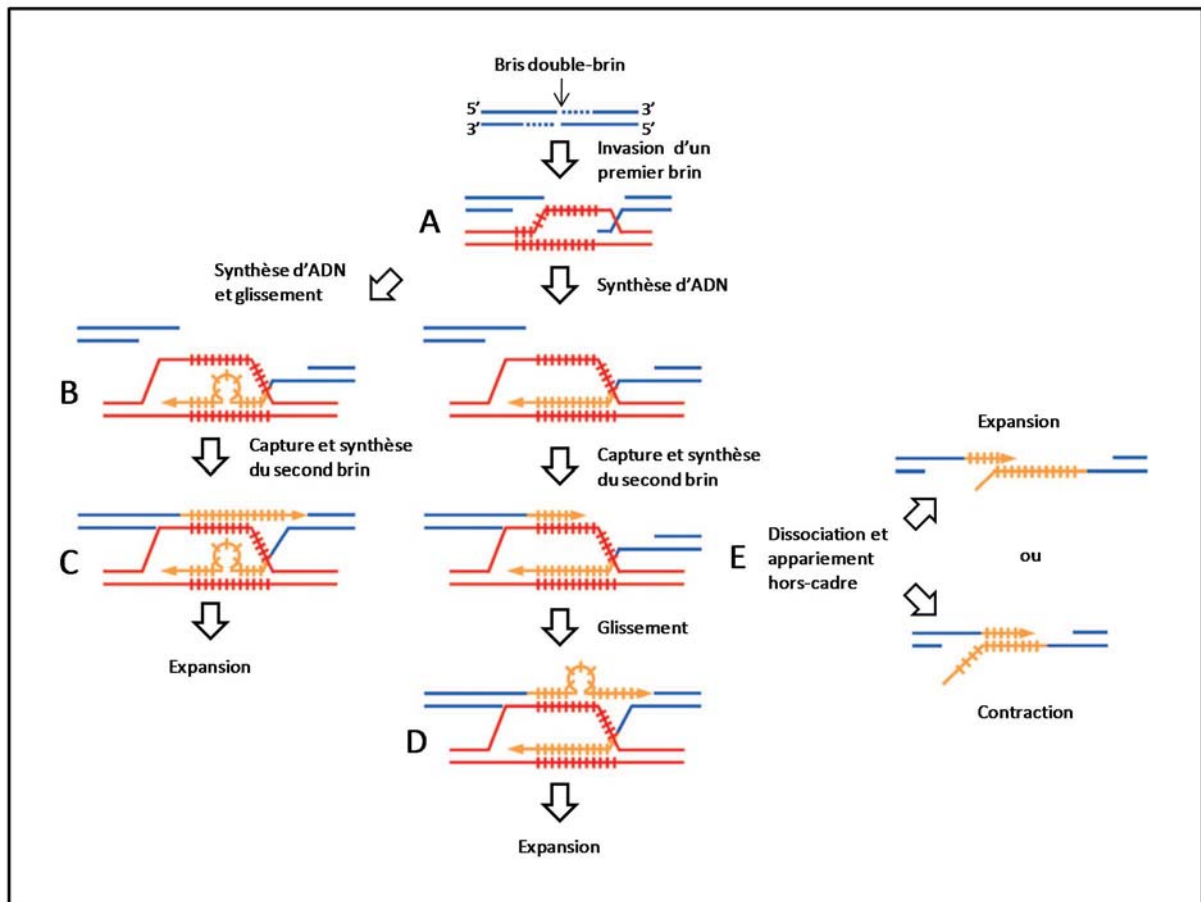


Figure 1.23. Modèle de réparation d'un bris double-brin par glissement (*DSB repair slippage*) pour des séquences répétées en tandem. La molécule d'ADN brisée (donneuse) est illustrée en bleu et la molécule matrice (receveuse) en rouge, l'ADN nouvellement synthétisée est en orange. (A) À la suite d'un bris double-brin, la conversion génique est initiée par l'invasion de l'ADN matrice homologue par l'extrémité 3'. (B) La synthèse de l'ADN peut être directe ou associée au glissement de la polymérase. Après la capture du second brin, la synthèse de l'ADN sur ce second brin peut être directe (C) ou associée au glissement de la polymérase (D). Les événements de glissement vont mener à l'expansion du segment contenant de l'ADN répétitif (C et D) ou à la contraction si le glissement survient sur le brin matrice. Alternativement, après la capture des deux extrémités suivie de la synthèse d'ADN, les deux brins nouvellement synthétisés peuvent se dissocier et s'apparier ensemble dans le bon cadre ou hors-cadre, menant à l'expansion ou la contraction du segment contenant l'ADN répétitif(E) (Traduit de Richard et al. 2008).

et l'organisation de la chromatine. Comme illustré à la section précédente, plusieurs évidences démontrent l'effet des microsatellites et des minisatellites sur les activités métaboliques telles que la réplication, la réparation et la recombinaison de l'ADN. La localisation des ADN répétés en tandem dans un gène peut être dans la séquence codante, dont l'expansion génère une protéine possédant une répétition d'un ou plusieurs acides aminés, ou dans une région non-codante (5'-UTR, 3'-UTR, intron). Des études ont fait la revue de différentes maladies neurologiques ou de comportements associés à l'amplification de séquences d'ADN répétées à l'intérieur d'un gène (Li et al. 2004; Orr et Zoghbi 2007; Usdin 2008; Fondon et al. 2008). Dans cette section, nous allons exposer les évidences qui soutiennent que les SSR et VNTR sont impliqués dans l'activité des gènes par divers mécanismes moléculaires.

1.2.2.4.1. Les microsatellites géniques

Plusieurs études montrent que les microsatellites ne sont pas distribués de manière aléatoire à l'intérieur des séquences codantes, des régions 5'-UTR et 3'-UTR et des introns. La variation de taille des SSR retrouvées dans ces régions peut provoquer une modulation de l'expression du gène, son inactivation, un glissement du cadre de lecture (*frameshift*), ou une modification de la fonction ou de l'activité de la protéine menant éventuellement à un changement phénotypique (Figure 1.24). Chez l'humain, la variation de taille de ces séquences peut causer des cancers et plusieurs maladies neurologiques, dont divers types d'ataxie spinocérébelleuses (SCA) et de dystrophies myotoniques (DM). La Figure 1.25 illustre la variété des séquences et la localisation de ces microsatellites dans les gènes associés à diverses maladies humaines, alors que le Tableau 1.4 donne les détails sur le gène affecté et la variation de taille de la SSR menant au développement des symptômes. La variation de taille des SSR est beaucoup plus importante dans les séquences non-codantes que dans les exons. En plus des effets sur la transcription déjà mentionnés, les microsatellites modifient l'activité génique de différentes manières, qu'ils soient situés dans une séquence codante ou non-codante.

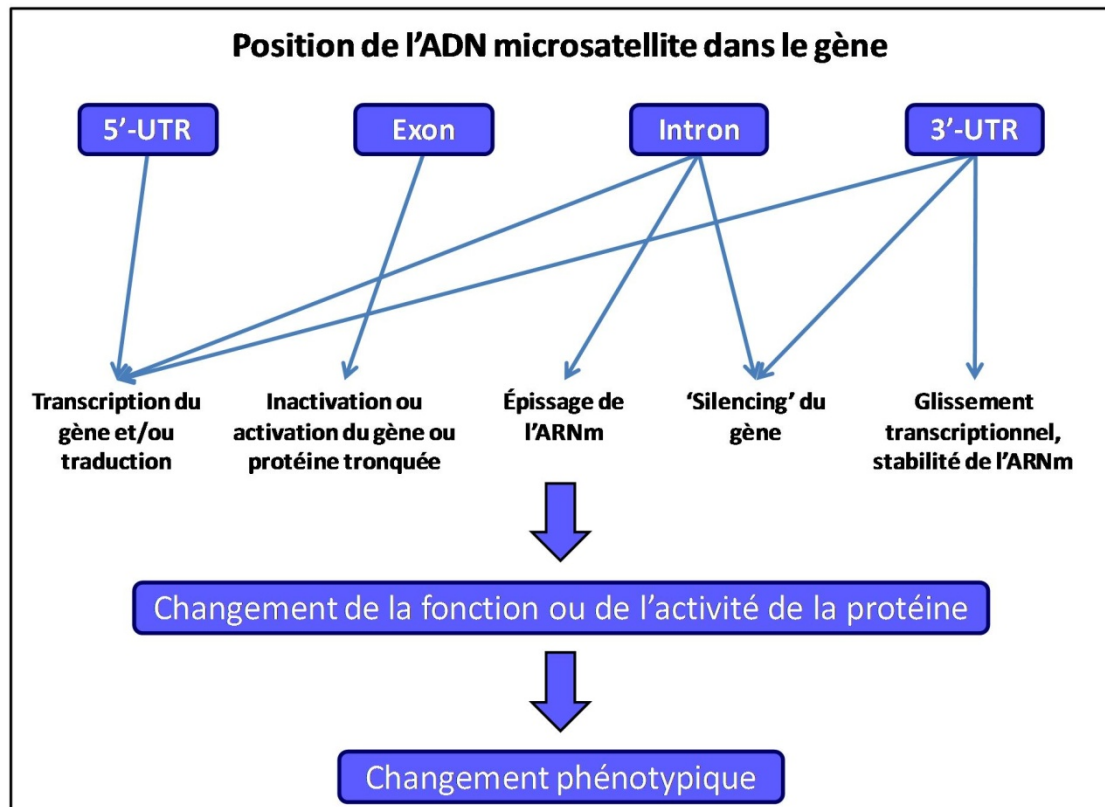


Figure 1.24. Fonctions régulatrices des microsatellites. Les microsatellites sont retrouvés dans les régions 5'-UTR, les exons, les introns et les régions 3'-UTR. Les fonctions associées aux microsatellites retrouvés à ces endroits sur un gène sont indiquées (Traduit de Li et al. 2004).

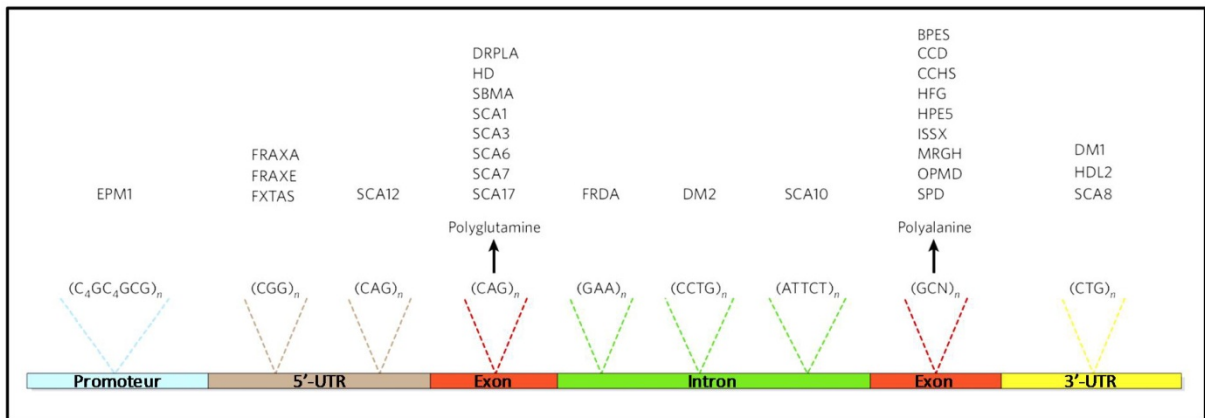


Figure 1.25. Localisation des microsatellites responsables de pathologies humaines. La séquence et la position dans un gène type des microsatellites qui causent des pathologies humaines sont indiquées, et les pathologies sont données. BPES, blepharophimosis, ptosis and epicanthus inversus; CCD, cleidocranial dysplasia; CCHS, congenital central hypoventilation syndrome; DM, myotonic dystrophy; DRPLA, dentatorubral–pallidolusian atrophy; EPM1, progressive myoclonic epilepsy 1; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation associated with *FRAXE* site; *FRDA*, *Friedreich's ataxia*; *FXTAS*, *fragile X* tremor and ataxia syndrome; HD, Huntington's disease; HDL2, Huntington's-disease-like 2; HFG, hand–foot–genital syndrome; HPE5, holoprosencephaly 5; ISSX, X-linked infantile spasm syndrome; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; SPD, synpolydactyly (Traduit de Mirkin 2007).

Tableau 1.4. Pathologies associées à l'amplification d'un microsatellite (SSR).

Pathologie	Gène affecté	Séquence du SSR	Position du SSR	Nb. normal de SSR	Nb. symptomatique de SSR
SSR codant					
DRPLA	ATN1	CAG	ORF	7-25	49-88
HD	HTT	CAG	ORF	6-34	36-180
SBMA	AR	CAG	ORF	11-24	40-62
SCA1	ATXN1	CAG	ORF	6-39	39-83
SCA2	ATXN2	CAG	ORF	15-29	34-59
SCA3	ATXN3	CAG	ORF	13-36	55-84
SCA6	CACNA1A	CAG	ORF	4-16	21-30
SCA7	ATXN7	CAG	ORF	4-35	34->300
SCA17	TBP	CAG	ORF	25-44	45-66
SSR non-codant					
DM1	DMPK	CTG	3'UTR	5-37	>50->2000
DM2	CNBP	CCTG	Intron 1	<27	75-11000
FXS	FMR1	CGG	5'UTR	6-52	~55->2000
FRA XE MR	AFF2/FMR3	CCG	5'end	6-25	>200
FRA12A MR	DIP2B	CGG	5'UTR	6-23	?
FRDA	FXN	GAA	Intron 1	7-22	>66->900
SCA10	ATXN10	ATTCT	Intron 9	10-29	280-4500
SSR codant et non-codant					
SCA8	ATXN8/ ATXN8OS	CAG et CTG	ORF et TNC	6-37	~107-250

(DM1) myotonic dystrophy type 1; (DM2) myotonic dystrophy type 2; (DRPLA) dentatorubral-pallidoluysian atrophy; (FXS) Fragile X syndrome; (MR) mental retardation; (FRDA) Friedreich ataxia; (HD) Huntington disease; (SBMA) spinobulbar muscular atrophy; (SCA) spinocerebellar ataxia; (TNC) transcrit non-codant (Adapté de Usdin, 2008).

Dans toutes les maladies pour lesquelles la séquence d'ADN répétée est située dans un exon, l'unité répétée est un trinuécléotide de séquence $(CAG)_n$ ou $(GCN)_n$, dont l'amplification mène respectivement à une protéine ayant une région polyglutamine (polyQ) ou polyalanine (polyA). La maladie de Huntington et plusieurs types d'ataxies spinocérébelleuses sont associées à l'amplification du trinuécléotide $(CAG)_n$. Le mécanisme de pathogénèse de ces maladies neurodégénératives serait associé directement ou indirectement à la formation d'agrégats causés par un changement conformationnel de la protéine (Orr et Zoghbi 2007). Il a d'ailleurs été démontré que les segments polyQ étaient toxiques pour les neurones et les cellules périphériques (Galvão et al. 2001). De plus, les protéines ayant un segment polyQ interagissent avec des facteurs de transcription et des cofacteurs riches en glutamines tels Sp1 et CBP (*CREB-binding protein*), ce dernier ayant été retrouvé dans les agrégats (Nucifora et al. 2001). La séquestration de ces protéines pourrait expliquer l'effet généralisé des protéines polyQ sur la transcription. Les segments polyA, pouvant être produits lors d'un glissement du cadre de lecture sur une séquence $(CAG)_n$, ont été impliqués dans plusieurs désordres congénitaux, mais le mécanisme menant au développement des pathologies est présentement inconnu.

Pour d'autres maladies, le microsatellite est localisé dans des régions non-codantes, montrant que l'association entre son amplification et le développement d'une pathologie repose sur d'autres facteurs que les propriétés intrinsèques de la protéine. Des effets sur l'expression des gènes par des SSR situés en région 5'-UTR ont été démontrés. Par exemple, la séquence formée par l'expansion du trinuécléotide $(CGG)_n$ associée au syndrome du X fragile, est retrouvée hyperméthylée et le gène FMR1 qui la contient est inactivé lorsque l'allèle possède plus de 230 répétitions (Jin et Warren 2000). À l'opposé, le locus n'est pas méthylé et l'ARNm est abondant avec un nombre plus faible de répétitions (Kenneson et al. 2001). Une variété de protéines liant les séquences CGG et ayant le potentiel d'affecter la transcription ont été décrites (ex. CGGBP (*CGG-binding protein*), Pur et Egr) (Usdin 2008). De plus, une étude a rapporté que l'expansion du trinuécléotide $(CTG)_n$ dans la région 5'-UTR

d'un gène rapporteur diminue la traduction de l'ARNm par la formation d'une structure épingle à cheveux qui inhiberait l'étape de balayage par le complexe de pré-initiation (Raca et al. 2000).

Les séquences d'ADN répétitives dans une région 3'-UTR peuvent mener à un glissement de la transcription. Il a été observé chez la levure que la transcription d'un gène rapporteur contenant une séquence (CAG)_n répétée en 3'-UTR produit des ARNm plus long de plusieurs kb de ce qu'il est attendu (Fabre et al. 2002). L'adoption d'une structure secondaire induirait une pause et un glissement arrière de l'ARN polymérase suivis d'une resynthèse d'une certaine séquence. Celle-ci générerait ensuite un plus long ARNm après plusieurs cycles de pause, glissement et synthèse. Une autre étude a rapportée que la répétition (CUG)_n retrouvée en structure épingle à cheveux sur l'ARNm activait la protéine kinase dépendante de l'ARN double-brin (PKR) pour interférer avec les fonctions cellulaires normales (Tian et al. 2000). On retrouve une répétition (CAG)_n dans la région 3'-UTR du gène DPMK associé à la dystrophie myotonique de type 1 et des effets ont été associés à son amplification anormale. Premièrement, l'ARNm de DPMK est retrouvé dans des foci nucléaires *in vivo* et le transcrit n'est pas exporté correctement au cytoplasme (Davis et al. 1997). Ensuite, l'expansion de cette séquence d'ADN répétée réduit l'expression du gène en aval DMAHP, probablement par un remodelage local de la chromatine (Thornton et al. 1997). Finalement, l'ARNm lie les protéines MBLN (*muclblind-like protein*) et CUG-BP (*CUG-binding protein*), dérégulant l'épissage de plusieurs transcrits (Philips et al. 1998; Miller et al. 2000).

D'autres séquences d'ADN répétées en tandem sont retrouvées dans des introns et plusieurs ont montré des effets sur l'activité transcriptionnelle. Un tétranucléotide (TCAT)_n situé dans le premier intron du gène de la tyrosine hydroxylase joue un rôle de régulateur sur la transcription (Meloni et al. 1998). Cet élément est conservé dans le gène de plusieurs primates, suggérant qu'une contrainte évolutive agit sur cette séquence (Meyer et al. 1995). L'expansion de SSR introniques peut également mener à certaines maladies, comme l'ataxie de Friedrich et le cancer du sein, suite à l'expansion respective d'un (GAA)_n qui inhibe l'expression du gène FRDA (Sakamoto et al. 2001) ou

d'un (CA)_n et stimule la transcription du gène du récepteur de l'EGF (*epidermal growth factor*) (Tidow et al. 2003). Un microsatellite complexe (TG)_m(T)_n situé au site d'épissage accepteur de l'intron8/exon9 du gène CFTR (*cystic fibrosis transmembrane conductance regulator*) contrôle l'épissage de l'ARNm et provoque la fibrose kystique lorsqu'il est muté (Pagani et al. 2000). Comme pour DM1, l'ARNm du gène ZNF9 (*zinc finger protein 9*) contenant un tétranucléotide amplifié (CCTG)_n dans l'intron 1 est retrouvé dans des foci nucléaires, indiquant la rétention des transcrits (Liquori et al. 2001).

1.2.2.4.2. Les minisatellites géniques

Même si le nombre d'études est plus limité, les minisatellites géniques semblent également modifier ou moduler la fonction de la protéine lorsqu'ils sont codants, et avoir un effet régulateur sur l'activité transcriptionnelle des gènes dans lesquels ils se trouvent lorsqu'ils sont non-codants. Pour certains VNTR, des mécanismes semblables à ceux démontrés pour les microsatellites ont été proposés et l'association est faite avec une maladie. Dans plusieurs cas cependant, on observe seulement une association entre une taille particulière d'un minisatellite et la variation du niveau d'expression du gène. La démonstration directe du lien entre la variation de l'expression du gène, la fonction de la protéine affectée et les effets phénotypiques observés est rarement présente. Dans cette section, l'implication des minisatellites codants et non-codants sur l'activité des gènes sera abordée.

Plusieurs minisatellites ont été identifiés dans les régions codantes des gènes. Le minisatellite retrouvé dans l'exon 3 du gène du récepteur humain de la dopamine (DRD4) est un bon exemple de la difficulté à établir un lien entre la présence d'un minisatellite et l'effet phénotypique. La taille de ce minisatellite varie entre 2 et 10 répétitions d'un motif de 48 pb qui est situé dans une région codant pour une boucle cytoplasmique du récepteur. Plusieurs études ont associé l'allèle contenant sept répétitions à des désordres du comportement, tels que la schizophrénie, l'hyperactivité/déficit d'attention et l'abus de drogues. En plus de dix ans, les résultats de ces études d'association ont été

alternativement réfutés et supportés par des études subséquentes, faisant en sorte que l'effet de la variation de taille du minisatellite à ce locus demeure incertain. L'allèle possédant sept répétitions du motif a cependant montré une affinité légèrement différente pour la liaison du ligand au récepteur (Van Tol et al. 1992), une modulation différente de l'AMPc intracellulaire par rapport aux allèles contenant deux et quatre répétitions (Asghari et al. 1995). Des évidences suggèrent même qu'il augmenterait en fréquence dans la population par sélection positive (Ding et al. 2002). La découverte d'un minisatellite de 120 pb dans la région 5'-UTR, pour lequel une corrélation inverse a été démontrée entre la longueur du VNTR et l'activité transcriptionnelle, suggère un autre mécanisme par lequel les minisatellites pourraient possiblement affecter l'activité du gène du récepteur de la dopamine (D'Souza et al. 2004). L'apolipoprotéine A (Apo(a)) montre un polymorphisme très grand avec plus de 30 allèles et une taille allant de 300 à 800 kDa. Cette variabilité dépend du nombre de séquences codantes répétées de 342 pb qui peut varier entre 12 et 51. Une corrélation inverse est généralement observée entre la taille des allèles Apo(a) et le niveau de lipoprotéine A (Lp(a)) dans le sérum. Un niveau élevé de Lp(a) a été identifié comme facteur de risque de maladies coronariennes et cérébrovasculaires (Pati et Pati 2000). Un autre minisatellite codant est retrouvé dans le gène de la protéine prion (PRNP) dont la structure normale comprend un segment de quatre répétitions d'un minisatellite de 24 nucléotides codant pour un octapeptide (PHGGGWGQ) dans la région N-terminale (Goldfarb et al. 1991). Cette région non-structurée contient cinq à six sites de liaison à forte affinité pour le cuivre et il a été suggéré que la liaison du cation divalent aux octapeptides facilitait le repliement et stabilisait la conformation de la protéine (Brown et al. 1997; Cereghetti et al. 2003). Des mutations impliquant l'insertion de quatre à huit octapeptides supplémentaires ont été décrites dans des familles héritant de manière autosomale dominante d'une maladie neuropathologique, tel que la maladie de Creutzfeldt-Jakob (Cochran et al. 1996; Laplanche et al. 1999). Les études se limitent encore une fois à de simples associations sans l'identification du mécanisme moléculaire. Des minisatellites sont retrouvés dans des familles de gènes, par exemple des minisatellites codants pour une séquence riche en sérines et thréonines de 17, 16 et 59 acides aminés ont été respectivement identifiés dans le gène des mucines *MUC3A*, *MUC4*, et

MUC17, desquels les molécules produites sont ensuite retrouvées à la surface de plusieurs cellules épithéliales (Desseyn et al. 2008).

Le minisatellite non-codant qui est probablement le mieux caractérisé est celui retrouvé en amont du promoteur du gène de l'insuline (INS) et qui constitue un locus de susceptibilité au diabète de type 1 (Bennett et al. 1995). Trois classes de minisatellites ont été identifiées sur la base du nombre de répétitions d'un motif de 14 à 15 pb riche en G+C (ACAGGGGT(G/C)(T/C)GGGG)_n : classe I (n = 20 à 63), classe II (n = 64 à 139) et classe III (n = 140 à 210). La classe I est généralement associée à la susceptibilité au diabète de type 1, avec le plus haut risque pour les homozygotes, alors que la classe III est associée à une protection, avec quelques exceptions (Bennett et al. 1997). Le transcrit de l'insuline a été détecté dans le thymus et les cellules β-pancréatiques, et l'expression de l'ARNm de l'insuline corrèle avec la variation de taille du minisatellite. La classe III protectrice est associée à une transcription du gène plus élevée que la classe I dans le thymus, qui elle montre un niveau plus élevé d'ARNm dans le pancréas (Vafiadis et al. 1996; Pugliese et al. 1997). La concentration plus élevée dans le thymus favoriserait l'élimination des lymphocytes T autoréactifs spécifiques à cette protéine, ce qui expliquerait l'effet dominant protecteur des allèles possédant un minisatellite de classe III. L'augmentation du risque de développer une auto-immunité associée aux allèles avec un VNTR de classe I serait le résultat combiné d'une faible expression au thymus et d'une forte expression dans le pancréas. Cet exemple montre bien que les minisatellites ont des effets spécifiques en fonction du type cellulaire. Une structure ADN tétraplexe a également été montrée pour ce minisatellite et l'adoption de cette dernière est directement corrélée avec l'activité transcriptionnelle (Catasti et al. 1996). De plus, on a identifié des sites de liaison pour le facteur de transcription Pur-1 qui stimulerait l'activité transcriptionnelle du gène (Kennedy et al. 1995).

Un minisatellite au motif de 12 pb (CCCCGCCCCGCG)_n retrouvé dans le promoteur du gène de la cystatine B (CSTB) est associé à un type d'épilepsie (*progressive myoclonus epilepsy*, EPM1) lorsque l'amplification atteint plus de 30 répétitions (Lafrenière et al. 1997; Lalioti et al. 1997). Les

patients ayant une expansion du VNTR montrent une expression réduite de l'ARNm du gène CSTB dans les cellules lymphoblastoïdes, et conséquemment une réduction de l'activité de cet inhibiteur de protéases à cystéine (Rinne et al. 2002). Une étude a également montré que le VNTR pouvait adopter une structure secondaire d'ADN tétraplexe, à laquelle on a associé la répression de la transcription du gène CSTB par l'altération de la structure de la chromatine (Saha et Usdin 2001). L'étude du minisatellite présent en amont du gène pour la monoamine oxydase A (MAOA) a suggéré que la variation de taille d'un minisatellite pouvait interférer avec une longueur optimale pour la région régulatrice. À cet effet, les allèles contenant 3,5 ou 4 copies du motif de 30 pb sont transcrits de 2 à 10 fois plus efficacement que ceux formés de 3 ou 5 copies (Sabol et al. 1998).

Deux minisatellites ont été identifiés dans le gène du transporteur de la sérotonine (5-HTT), un premier situé dans la région régulatrice (5-HTTLPR) et le second retrouvé dans le deuxième intron (STin2). Le polymorphisme de 5-HTTLPR implique deux allèles majeurs, S (*short*) et L (*long*), qui correspondent à la présence de 14 ou 16 répétitions d'un motif de 20 à 23 pb. Une plus faible transcription du gène 5-HTT a été montrée pour l'allèle S qui résulte en une baisse de l'expression de la protéine et de la recapture de la sérotonine dans des cellules lymphoblastiques (Lesch et al. 1996). On retrouve également deux allèles principaux pour le VNTR STin2, STin2.10 et STin2.12, constitués de 10 et 12 répétitions d'un motif de 17 pb. Des évidences indiquent que ce minisatellite agirait comme régulateur transcriptionnel, l'allèle STin2.12 possédant une activité transcriptionnelle supérieure à l'allèle STin2.10 (MacKenzie et Quinn 1999). On a d'ailleurs identifié dans ce minisatellite des sites de liaison pour les facteurs de transcription YB-1 (*Y box binding-protein 1*) et CTCF (*CCTC-binding factor*) et démontré que ce dernier interfère avec la capacité de YB-1 à induire l'expression d'un gène rapporteur (Klenova et al. 2004). Les résultats de méta-analyses sur l'association du VNTR 5-HTTLPR à des maladies sont contradictoires et ne permettent pas de tirer de conclusion sur son implication (Munafò et al. 2005; Sen et al. 2004). Une autre méta-analyse permet cependant d'associer l'allèle STin2.12 à un facteur de risque pour la susceptibilité à la schizophrénie

(Fan et Sklar 2005). Ces études, ajoutées à celles sur le récepteur de la dopamine, illustrent la difficulté d'établir une association entre la variation de taille des minisatellites et des traits comportementaux. D'un côté les séquences d'ADN répétées exercent seulement de petits effets quantitatifs sur le phénotype, alors que les traits comportementaux sont typiquement multigéniques, hétérogènes, faiblement pénétrants et influencés par l'environnement (Fondon et al. 2008).

Finalement, d'autres minisatellites ayant des effets sur la transcription des gènes ont été identifiés dans des introns et des régions 3'-UTR. Un minisatellite composé de 3 à 11 répétitions d'un motif de 40 pb est retrouvé dans la région 3'-UTR du transporteur de la dopamine (DAT1) (Vandenbergh et al. 1992). Plusieurs pathologies sont associées au minisatellite ayant 10 répétitions du motif qui a un effet "*enhancer*" dépendant de sa longueur, et qui montre un niveau d'ARNm plus élevé dans différents tissus (Mill et al. 2002). La variabilité des effets observés pour différents allèles ou génotypes sur l'expression des gènes pourrait également être expliquée par la présence d'un second élément affectant l'expression du gène. À cet effet, une étude a identifié un second minisatellite au motif de 30 pb dans l'intron 8 du gène DAT1 qui provoque des effets différentiels sur l'expression d'un gène rapporteur (Guindalini et al. 2006). Un minisatellite riche en bases purines est présent dans la région 3'-UTR du gène du récepteur de la bradykinine de type 2 (B2) et dont l'allèle le plus fréquent consiste en 43 répétitions du motif (TGGAGAGCTAGAACC)_n. Contrairement à l'exemple précédent, un effet de régulateur négatif dépendant de l'orientation et de la longueur du minisatellite lui a été associé. De plus, l'ARNm contenant le VNTR a montré une demi-vie supérieure au contrôle en transfections transitoires (Zamorano et al. 2006).

1.2.2.5. Les séquences d'ADN répétitives et l'évolution des espèces

L'idée que des séquences non-codantes puissent être impliquées dans l'évolution des espèces remonte à plus de trente ans lorsque deux groupes, sur la base d'évidences indirectes, ont posé l'hypothèse que des mutations régulatrices en *cis* (*cis-regulatory mutations*) auraient possiblement un

rôle évolutif important à jouer. Après la découverte qu'une grande proportion des génomes eucaryotes était composée de séquences répétitives, une première étude par Britten et Davidson (1971) a proposé que ces séquences pourraient réguler la transcription, et que ces mutations régulatrices auraient une implication cruciale dans l'évolution des phénotypes. Une seconde étude par King et Wilson (1975) a rapporté que les séquences codantes de gènes orthologues entre l'humain et le chimpanzé étaient identiques à plus de 99 %. Les auteurs ont argumenté que ce faible degré de divergence au niveau des séquences protéiques ne pouvait expliquer les différences phénotypiques entre les espèces et suggéré que les mutations régulatrices étaient responsables de ces différences biologiques. Aujourd'hui, des évidences s'accumulent pour appuyer l'implication des séquences régulatrices dans l'adaptation de populations naturelles et la susceptibilité à plusieurs maladies (Kleinjan et van Heyningen 2005). Longtemps considérées comme des déchets (*junk*) et encore parfois de nos jours jugées inutiles (*usefulness*) par certains auteurs (ex. Ellegren 2004), les recherches sur les séquences d'ADN répétitives montreraient de plus en plus qu'elles ont un rôle important à jouer dans l'évolution. Mis à part leurs implications dans plusieurs maladies, des études réalisées dans différents organismes rapportent des effets bénéfiques associés à ces éléments qui permettent une adaptation rapide aux conditions environnementales par des changements morphologiques et comportementaux. La prochaine section décrit quatre exemples montrant une association entre la variabilité d'une séquence répétée présente dans un gène, la fonction moléculaire de la protéine et le phénotype observé.

Le premier exemple à avoir été documenté chez un eucaryote montre des évidences d'effets phénotypiques quantitatifs associés à un minisatellite et d'une sélection naturelle agissant sur les allèles (Sawyer et al. 1997). Le gène PER chez *Drosophila melanogaster*, dont le produit est une composante essentielle de la régulation du rythme circadien, contient dans l'exon 5 une répétition d'hexanucléotides codants pour un segment de dipeptides thréonine-glycine (Thr-Gly)_n. Pour les deux allèles les plus communs, les mouches ayant l'allèle le plus long (Thr-Gly)₂₀ montrent une sensibilité réduite du rythme circadien en réponse à une variation de température par rapport à celles ayant l'allèle

plus court (Thr-Gly)₁₇. Il a été suggéré que le domaine Thr-Gly serait impliqué dans la stabilité de la protéine en permettant de maintenir une conformation particulière en fonction de la température (Peixoto et al. 1998). Dans des populations naturelles, l'allèle (Thr-Gly)₂₀ est retrouvé à une plus haute fréquence aux latitudes plus élevées. Sur la base d'une différente sensibilité qu'il confère, il a été suggéré que cet allèle était sélectionné par le climat qui montre une plus grande variation de température dans ces régions (Kyriacou et al. 2008). D'autres études supportent cette hypothèse d'une sélection naturelle, notamment les observations faites en Israël au site du "*Evolution Canyon*". On retrouve à cet endroit un contraste microclimatique important; le versant sud ensoleillé est associé à un stress de température et de sécheresse plus grand que le versant nord. L'allèle le plus long (Thr-Gly)₂₀ a été retrouvé plus fréquemment dans la zone écologique située sur le versant nord qui est plus frais (Zamorzaeva et al. 2005). De plus, un échantillonnage de mouches fait à haute altitude (> 1000 m) en Europe a montré une fréquence des allèles qui était caractéristique d'une région plus nordique (Peixoto 2002).

Le deuxième exemple chez le campagnol illustre comment la variation d'un microsatellite complexe présent dans la région 5'-UTR gène du récepteur de la vasopressine 1a (AVPR1A) est impliqué dans la régulation de l'expression du gène et la modification des comportements sociaux (Hammock et Young 2004; 2005). Les rongeurs du genre *Microtus*, montrent des différences marquées dans la structure sociale des espèces. Le campagnol des prairies (*Microtus ochrogaster*) est social et monogame alors que le campagnol des montagnes (*Microtus montanus*) est asocial et n'est pas monogame. La comparaison de ces campagnols révèle des différences dans l'expression de la protéine au cerveau et une augmentation des comportements sociaux pour le campagnol des prairies qui possède un long microsatellite. La région 5'-UTR du gène *AVPR1A* de différentes espèces est variable. Les deux espèces sociales de campagnol possèdent un microsatellite bien développé dans cette région, alors que deux espèces asociales n'ont qu'une version courte. Le microsatellite est également conservé chez l'humain et le bonobo (*Pan paniscus*) qui ont un comportement sociosexuel

semblable, contrairement au chimpanzé (*Pan troglodites*) chez qui on a noté une délétion de 360 pb incluant une partie du microsatellite. L'effet sur la transcription provoqué par la seule variation de taille du microsatellite a été démontré *in-vitro* dans différents types cellulaires avec un gène rapporteur. De plus, l'introduction transgénique du gène AVPR1A d'un campagnol des prairies dans une souris provoque une distribution de la protéine au cerveau et des comportements en réponse à l'injection de vasopressine qui sont plus semblables à ceux de l'espèce sociale qu'à ceux de la souris contrôle (Young et al. 1999). Une association a dernièrement été faite chez l'humain entre la présence d'un long microsatellite en amont du gène et un comportement altruiste, de même qu'un plus haut niveau d'ARNm dans l'hypocampe *post-mortem*, suggérant la base génétique du mécanisme fonctionnel (Knafo et al. 2007).

La variation morphologique est commune chez les espèces canines. Une étude comparative de Fondon et Garner (2004) s'est intéressée aux séquences d'ADN répétées présentes dans des gènes impliqués dans le développement chez 92 races de chiens. L'examen de 17 gènes reconnus pour influencer les traits morphologiques a révélé que ces derniers montraient un niveau très élevé de variation, avec du polymorphisme dans presque tous les gènes. De plus, la pureté des répétitions, en comparaison avec celles retrouvées chez l'humain ou à d'autres sites, indiquerait une fluctuation récente et rapide des microsatellites ne permettant pas l'accumulation de mutations ponctuelles. Même si la fonction de la plupart de ces variations au niveau des SSR reste inconnue, les auteurs ont trouvé que le ratio de deux microsatellites consécutifs dans le gène du facteur de transcription Runx-2, contenant de 18 à 20 glutamines suivies de 12 à 17 alanines, était corrélé avec des mesures faciales entre les races. La protéine homologue humaine est un facteur de transcription spécifique aux ostéoclastes (OSF2) qui influence la structure cranio-faciale. L'expansion anormale du segment de 17 à 27 alanines est retrouvée dans une famille où des membres sont atteints de dysplasie cléidocrânienne (Mundlos et al. 1997). Ils ont également observé, chez la race des chiens de montagne des Pyrénées qui est caractérisée par de la polydactylie, la présence d'un ergot supplémentaire qui était associé à une

délétion de 51 pb codant pour un segment de 17 dipeptides répétés proline-glutamine (Pro-Gln) dans le gène *Alx-4*. Cette forme de polydactylie est semblable à celle retrouvée chez la souris ayant les deux copies inactivées de ce gène qui code pour un activateur transcriptionnel (Qu et al. 1997). Ces résultats suggèrent que la sélection artificielle effectuée par les éleveurs de chiens de race pour des caractères particuliers n'est pas si différente de la sélection naturelle.

Le dernier exemple montre comment la variation de taille d'un minisatellite codant permet une diversité fonctionnelle d'un antigène à la surface cellulaire et possiblement une adaptation de l'organisme aux conditions environnementales (Verstrepen et al. 2005). Comme mentionné précédemment, plusieurs gènes dont les produits sont impliqués dans la formation de la paroi cellulaire chez *S. cerevisiae* contiennent des minisatellites codants. Parmi eux, des membres de la famille des mannoprotéines FLO affichent différents variants dont la taille du motif peut être de 30 pb (*FLO11*), 81 pb (*FLO10*) ou 135 pb (*FLO1*, *FLO5* et *FLO10*) (Richard et Dujon 2006). Ces gènes sont les orthologues d'ALS et d'EPA chez *C. albicans* et *C. glabrata* respectivement, lesquels sont impliqués dans l'adhérence cellulaire et la pathogénicité. En utilisant des variants de FLO1 ne différant que par le nombre de motifs formant le minisatellite, les auteurs ont observé une altération quantitative d'un phénotype; la longueur du minisatellite étant directement corrélée avec l'adhésion au polystyrène et aux autres cellules. Il a également été rapporté qu'une souche de *S. cerevisiae* utilisée dans la production de vin de cerises avait des mutations dans le gène FLO11 qui lui conférait l'habileté de flotter pour ainsi avoir accès à l'oxygène (Fidalgo et al. 2006). La variation de taille du minisatellite codant a été identifiée comme le déterminant majeur de cette flottabilité, un nombre accru de motifs dans le domaine central de la protéine conférant à la cellule un niveau très élevé d'hydrophobicité de surface.

1.3. Mise en contexte

La furunculose, causée par le pathogène bactérien *Aeromonas salmonicida*, occasionne des pertes économiques importantes dans les piscicultures québécoises. Exprimée à la surface des cellules présentatrices d'antigènes, la protéine du complexe majeur d'histocompatibilité de classe II (CMHII) est essentielle au système immunitaire puisqu'elle présente des peptides dérivés principalement de pathogènes extracellulaires aux récepteurs des lymphocytes T CD4⁺. Ceci permet de guider le développement et l'activation de ces derniers afin de combattre l'infection. La variabilité de certains résidus de la niche peptidique du CMHII influence directement les propriétés de liaison de différents allèles pour les peptides antigéniques. De plus, un système de contrôle strict de la transcription des gènes permet de restreindre l'expression à certains types de cellules et de varier leurs niveaux d'expression.

Plusieurs études chez l'humain, le poulet et d'autres espèces animales font la corrélation entre la présence de certains allèles de gènes du CMH et une plus grande résistance ou susceptibilité à des agents infectieux. Cependant, peu d'études ont évalué l'effet de ce polymorphisme face au pathogène bactérien *Aeromonas salmonicida*. Chez le saumon atlantique, quelques-unes ont montré une association significative entre des allèles particuliers et une résistance accrue au pathogène. Identifier des souches d'omble de fontaine possédant une résistance accrue aux infections causées par cette bactérie constituerait un atout économique intéressant.

Dans un premier temps, nous avons effectué une étude en ce sens afin de caractériser les allèles du gène MHII β présents dans une population d'omble de fontaine de souche Rupert et tester par la suite leurs performances lors d'une infection par *Aeromonas salmonicida*. Nous avons donc développé une méthode de génotypage par SSCP (Single-Strand Conformation Polymorphism) pour évaluer la variabilité et déterminer la séquence des allèles du gène MHII β présents dans la population. Cette technique s'avère très efficace pour identifier rapidement de légères différences de séquences (parfois un seul nucléotide) pour un gène particulier. Après avoir généré des familles à partir de géniteurs de la

population de départ, la descendance a été exposée au pathogène *A. salmonicida* par immersion et la mortalité analysée en fonction des allèles du gène MHII β .

Le séquençage des différents allèles MHII β chez l'omble de fontaine de souche Rupert nous a également révélé la présence d'un minisatellite polymorphique situé dans le second intron du gène. Les minisatellites sont composés de courtes séquences d'ADN répétées en tandem et ont une forte tendance à muter. Ceux-ci sont généralement considérés comme "junk DNA". Cependant, des études récentes montrent des effets moléculaires et phénotypiques associés à la variation dans la taille de minisatellites particuliers. En effet, les séquences d'ADN répétées en tandem ont le potentiel de former des structures secondaires de l'ADN, pour lesquelles des évidences suggèrent un rôle fonctionnel dans la réplication et la transcription. De plus, la variabilité dans la taille de minisatellites particuliers a été associée à la susceptibilité pour certaines maladies. Par exemple, un minisatellite présent en amont du gène de l'insuline est associé à la susceptibilité ou à la protection contre le diabète de type I, démontré par un effet régulateur différentiel sur la transcription dans le thymus et le pancréas en fonction de sa taille.

Dernièrement, l'hypothèse d'un rôle évolutif des minisatellites comme source importante de variation génétique associée à l'adaptation a été proposée. Afin de déterminer si un rôle fonctionnel est associé au minisatellite présent dans le second intron du gène MHII β , nous avons étudié dans un deuxième temps l'effet associé à la présence de minisatellites de tailles différentes sur la transcription des allèles du gène MHII β . Le niveau d'expression a été étudié par RT-PCR en temps réel à partir d'ARN total obtenu de lymphocytes ou du pronéphros d'individus gardés à basse et haute températures (6 °C et 18 °C, respectivement). Cette technique est actuellement la plus sensible pour détecter de légères variations transcriptionnelles d'un gène (ou d'un allèle). De plus, l'utilisation d'amorces spécifiques permet de comparer les niveaux d'expression des deux allèles présents dans un poisson hétérozygote, limitant ainsi la variation interindividuelle.

L'omble de fontaine étant un poisson poïkilotherme, les résultats de la présente étude pourraient permettre de poser les bases pour étudier de façon complémentaire la résistance ou

susceptibilité aux pathogènes en fonction (1) d'allèles spécifiques du CMH et (2) de leur niveau d'expression. Éventuellement, l'analyse de plusieurs souches infectées par *Aeromonas salmonicida* pourrait mener à l'identification d'acides aminés particuliers de la niche peptidique du gène MHII qui sont associés à une résistance accrue à la furonculose. Ultimement, il devrait être possible d'identifier des poissons intrinsèquement plus résistants aux infections, minimisant ainsi l'utilisation d'antibiotiques en aquaculture.

Chapitre 2

Identification of MHC class II β resistance/susceptibility alleles
to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*)

Identification of MHC class II β resistance/susceptibility alleles to
Aeromonas salmonicida in brook charr (*Salvelinus fontinalis*)

Sébastien Croisetière^{a,1}, Philippe D. Tarte^a, Louis Bernatchez^b, Pierre Belhumeur^a.

^aDépartement de Microbiologie et Immunologie, Université de Montréal, Montréal, Québec, H3C 3J7, Canada. ^bDépartement de Biologie, Université Laval, Québec, Québec, G1V 0A6, Canada.

¹ Corresponding author. Tel.: 1-514-343-6111 #1432; fax: 1-514-343-5701.

Cet article a été publié dans : Molecular Immunology 45(11):3107-3116. Juin 2008.

CONTRIBUTIONS DU CANDIDAT

En tant que premier auteur, j'ai contribué de façon majoritaire à la rédaction de l'article, à l'analyse des résultats, à la réalisation des travaux de laboratoire ainsi qu'à l'élaboration des travaux de recherche. Philippe D. Tarte a collaboré aux travaux de laboratoire menant à la publication de cet article. Louis Bernatchez et Pierre Belhumeur ont participé à la rédaction de l'article (commentaires, corrections) ainsi qu'à l'élaboration du projet de recherche.

2.1. Abstract

Pathogen-driven selection is believed to be important in the evolution and maintenance of the polymorphism of the major histocompatibility complex (MHC) genes in vertebrates. Yet, it has been tested empirically for only a handful of species thus far. In this study, we first investigate by SSCP (Single Strand Conformational Polymorphism) the diversity found at the MHC class II β gene in a population of brook charr (*Salvelinus fontinalis*) from the Rupert River (Québec, Canada). Secondly, to explore the survival performances conferred by specific alleles and genotypes, individuals from 23 half- and full-sibling families were infected with *Aeromonas salmonicida*, the causative agent of furunculosis. From the initial brook charr population, a total of six MHC class II β alleles were identified; four complete and two partial coding sequences that include the complete polymorphic β 1 domain. We found that one allele, *Safo-DAB*0101*, was significantly associated with resistance against *Aeromonas salmonicida*. In addition to homozygotes for this allele, its resistance effect was also detected in heterozygote individuals for two specific genotypes. In contrast, other allelic combinations, namely heterozygous genotypes *Safo-DAB*0201/*0301* and *Safo-DAB*0301/*0401* were significantly associated with increased susceptibility to furunculosis. As previously found in other salmonids, the intronic sequences obtained from the MHC class II β gene in brook charr indicated that a single locus is expressed in this population. Given that its frequency was relatively low (0.0873), the negative frequency dependent selection hypothesis could explain the advantage associated with the allele *Safo-DAB*0101* over the other alleles and highlight the importance of this mechanism to sustain variation at the MHC in brook charr.

Keywords: Histocompatibility antigens class II; alleles; bacterial infections; disease resistance; *Salvelinus fontinalis*; *Aeromonas salmonicida*; furunculosis; selection

2.2. Introduction

The major histocompatibility complex (MHC) class I and class II genes encode cell-surface proteins specialised in the presentation of self and non-self antigen peptides to T-lymphocytes in the adaptive immune system. These represent the most polymorphic genes known to date, with multiple loci and high allelic diversity at each of these loci (Trowsdale and Parham, 2004). Unlike the situation in other vertebrates, the two classical MHC regions, class I and class II, are not found in a complex in bony fishes (Bingulac-Popovic et al., 1997; Sato et al., 2000). For this reason, the expression “major histocompatibility” (MH) genes is more appropriate in teleosts (Dixon and Stet, 2001). Studies have shown that rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) express a single “minimal, essential” gene of the MH class I (Aoyagi et al., 2002; Grimholt et al., 2002) and class II (Glamann, 1995; Langefors et al., 2000). Also, MH class II α and class II β loci are genetically linked in Atlantic salmon (Stet et al., 2002).

The evolution and maintenance of the polymorphism of the MHC genes can be attributed to two major types of mechanisms: the pathogen-driven and reproductive mechanisms (Aguilar and Garza, 2007; Bernatchez and Landry, 2003). Two main hypotheses have been proposed to explain the pathogen-driven selection: the overdominance and the negative frequency dependent selection hypothesis. In the overdominance model (or heterozygous advantage), the heterozygous individuals are assumed to present a broader range of pathogen-derived antigens due to a larger number of different MHC molecules and have increased fitness relative to homozygous (Hughes and Nei, 1989). The second model of negative frequency dependent selection (or rare-allele advantage) argues that individuals bearing low-frequency alleles have an advantage because of the limited co-evolution of the

pathogens facing these MHC alleles (Takahata and Nei, 1990; Slade and McCallum, 1992). A third mechanism implicates selection that fluctuates in time and/or space whereby the presence or absence of pathogens over time in an indigenous bacterial or viral population can influence the polymorphism at MHC loci (Hedrick, 2002).

The implications of different MHC class I and class II alleles in disease resistance or susceptibility have been documented previously. Pioneer studies on congenic chicken lines have clearly demonstrated the B haplotype effect on Marek's disease resistance (Bacon, 1987) and thereafter MHC associations toward resistance to other viral or bacterial pathogens in this species (Zekarias et al., 2002). Since then, strong associations have been reported between the human MHC (HLA) and autoimmune diseases (Gebe et al., 2002; Singer et al., 1997) and resistance to diseases (Carrington and O'Brien, 2003; Shiina et al., 2004). In fish, the functional implications of MH polymorphism on the resistance against or susceptibility to infectious diseases have essentially all been investigated in Atlantic salmon (*Salmo salar*). Langefors et al. (2001) were the first to establish a correlation between survival probability and three different MH class II β alleles following *Aeromonas salmonicida* exposure. Their observations were confirmed by Lohm et al. (2002) in an experiment where the effect of non-MH background genes could be controlled. Another study identified significant associations between resistance towards infectious diseases caused by both bacterial or viral pathogens and MH class I and class II alleles (Grimholt et al., 2003). It is noteworthy that different alleles were identified as associated with resistance to each pathogen. More recently, a combination of particular MH class I and class II alleles has been associated with resistance or susceptibility to infectious salmon anaemia virus (ISAV) (Kjoglum et al., 2006). Finally,

results that could be interpreted as the manifestation of a disease-mediated natural selection have been observed for the MH class II locus, but not for MH class I (de Eyto et al., 2007).

Aeromonas salmonicida, a gram-negative, facultatively anaerobic, rod-like bacterium, is an invasive pathogen capable of surviving and replicating within intraperitoneal macrophages and non-phagocytic cells (Daly et al., 1996; Garduno et al., 2000). This economically important pathogen is the causative agent of furunculosis that has a strong impact on the survival of salmonids. Previous studies have demonstrated an important genetic component associated with resistance to this pathogen (Perry et al., 2004; Gjedrem, 2000). Brook charr (*Salvelinus fontinalis*) is an important endemic salmonid fish from eastern North America and the major breeding species of the Québec aquaculture industry, but there has been no research regarding the MH genes and their performance against *A. salmonicida*, although a heritable genetic basis for resistance to furunculosis has been evidenced in this species (Perry et al., 2004). In this study, our aims were (i) to document the diversity of alleles at the MH class II β gene of brook charr and (ii) for the first time, to document survival conferred by specific alleles and/or genotypes following exposure to furunculosis in a fish other than Atlantic salmon.

2.3. Materials and methods

2.3.1. Fish and challenge test

Wild brook charr (*Salvelinus fontinalis*) originating from the Rupert River and associated waterways were maintained at the LARSA (LABoratoire Régional des Sciences Aquatiques; Université Laval). This strain is of particular interest for aquaculture because it demonstrates rapid growth and large adult size that may exceed 5 kg (Sutton et al., 2002). Fish kept at LARSA were used to generate 23 half- and full-sibling families as previously described (Perry et al., 2004). This allowed us to take into account the effect of genetic background in interpreting the putative role of MH alleles in resistance to furunculosis. To identify individuals of interest for breeding, adult brook charr were first genotyped at the MH class II β locus by SSCP (Single Strand Conformational Polymorphism). Fish were exposed to a virulent strain of *Aeromonas salmonicida* by a bath immersion method (Dautremepuits et al., 2006; Perry et al., 2004; Lutwyche et al., 1995) at 1×10^6 bacteria/ml previously determined to be the LD₅₀ at 72 hours. Individual survival was recorded every 4 hours for the following 90 hour-period. No mortality was observed in control fish. Dead and surviving individuals were collected and SSCP genotyped at the MH class II β locus. The final experimental dataset contained the surviving time of 860 genotyped individuals for statistical analyses.

2.3.2. MH class II β genotyping of brook charr

Genomic DNA from 63 adult brook charr was extracted from adipose fin for genetic analysis. The different MH class II β alleles were first identified using SSCP. A 315-318 bp fragment containing the complete MH class II β exon 2, coding for the polymorphic β 1 domain of the protein, was amplified by PCR (polymerase chain reaction) using intronic sense primer

SP4501 (5'-CCTGTATTTATGTTCTCCTTTC-3') and antisense primer SP4502 (5'-TAAGTGTTGCTACGGAGCC-3') in a thermocycler (Biometra). PCR was carried out in a total volume of 50 μ l containing 20-40 ng of genomic DNA, 0.2 mM dNTPs, 8 μ l of 10X PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 2 mM MgCl₂, 5 pmol each of SP4501 and SP4502 primers, 1 unit of *Taq* DNA polymerase. The following PCR conditions were used: denaturation for 3 min at 94°C followed by 35 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 47°C and extension for 1 min at 72°C. The final extension was for 10 min at 72°C. Three μ l of formamide were added to 9 μ l of the PCR products which were then loaded onto a non-denaturing acrylamide gel (10% 49:1 acrylamide:bis-acrylamide, 5% glycerol and 1X TBE) for a 2.5 hours and 200 V migration in a Mini-protean III tank (Bio-Rad) in an ice-cold water bath. The procedure resolves alleles into discrete bands that were finally revealed on the gels by silver staining (Budowle et al., 1991).

2.3.3. RNA isolation from peripheral blood leukocytes (PBL) of brook charr

RNA from PBL of brook charr was extracted to assure the identification of expressed MH class II β alleles. Fish were anaesthetised with 0.1% of tricaine methanesulfonate (MS-222) (Sigma). The blood was collected from the caudal vein using a heparinised 3 ml syringe with a 21G1 needle (Becton Dickson) and added to 1 volume of ice-cold PBS containing 10 U/ml heparin sodium salt (Sigma). The cellular suspension (1 ml) was then spun at 500 x g for 5 min at 4 °C and the buffy layer and plasma transferred to a new Eppendorf. The PBL were then pelleted and suspended in Trizol (Invitrogen) for total RNA extraction according to the manufacturer's protocol.

2.3.4. Reverse Transcriptase (RT)-PCR from PBL

Total RNA (500 ng) from fish of different genotypes was reverse transcribed using SuperScript II reverse transcriptase (Invitrogen), oligo(dT)₁₂₋₁₈ primer (Invitrogen) and RNaseOUT (Invitrogen) according to the manufacturer's protocol. The complete open reading frame of MH class II β was obtained by PCR using primers in the 5'-UTR (P472; 5'-CAGCAGAGGAACATGTCGATG-3') and the 3'-UTR (P473; 5'-TTTCTGCTGCAGATTCAGCA-3') designed from sequences at positions conserved between Atlantic salmon and rainbow trout (Syed et al., 2003). The reactions were carried out on homozygous individuals for product homogeneity. PCR was carried out in a total volume of 50 μ l containing 0.25 μ l of cDNA from a 1/20 dilution of the RT reaction, 0.2 mM dNTPs, 5 μ l of 10X Expand High Fidelity buffer (Roche), 2 mM MgCl₂, 5 pmol each of P472 and P473 primers, 2.6 units of Expand High Fidelity enzyme mix (Roche). The following PCR conditions were used: denaturation for 3 min at 94°C followed by 40 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 45°C and extension for 1 min at 72°C. The final extension was for 10 min at 72°C. Sequencing of the PCR products was realized on a 3730 DNA analyser (Applied Biosystems).

2.3.5. Data analysis

Alignment of the deduced amino acid sequence of MH class II β peptide was performed using ClustalX (Thompson et al., 1994). Sequences used for comparison and their GenBank™ accession numbers were as follows: *Oncorhynchus mykiss* : U20943 (Glamann, 1995), *Salmo salar* : X70166 (Hordvik et al., 1993), *Salmo trutta* : AF296398 (Shum et al., 2001), *Cyprinus carpio* : Z49064 (Ono et al., 1993), *Brachydanio rerio* : L04805 (Ono et al., 1992) and human : P04229; UniProt (Bell et al., 1985).

Nucleotide sequences were analysed with the Syn-SCAN program (Gonzales et al., 2002) to calculate synonymous and nonsynonymous nucleotide substitution rates according to the method of Nei and Gojobori (1986) using a Jukes-Cantor correction.

Phylogenetic tree was created using the Neighbor-Joining method (Saitou and Nei, 1987). The percentages of replicate trees in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

We calculated the proportion of mortality for each genotype with 95% exact confidence intervals obtained by using the F distribution method given in Collet (1991) and also described by Leemis and Trivedi (1996). To compare survival times between the genotypes for MH class II proteins, we conducted a log-rank test with the procedure LIFETEST. Then the Cox proportional hazards model (Cox, 1972) was fitted by using the procedure TPHREG to compare survival times between genotypes with adjustment for covariates (weight and infection tank) selected by the stepwise method. In the Cox model, each individual has its own hazard function expressed as:

$$\lambda_i(t) = \lambda(t; \mathbf{Z}_i) = \lambda_0(t) \exp(\mathbf{Z}_i' \boldsymbol{\beta})$$

where $\lambda_i(t)$ is the risk function for the individual i at time t , \mathbf{Z}_i is the vector of covariates for individual i , $\lambda_0(t)$ is the common part of the model and $\exp(\mathbf{Z}_i' \boldsymbol{\beta})$ is the individual specific part of the model. Results are presented as a hazard ratio (HR). To compare allele and genotype effect on survival time, we computed contrasts on HR with 95% confidence intervals.

2.4. Results

2.4.1. Sequences of brook charr MH class II β gene

Six MH class II β alleles were identified from the 63 brook charr individuals by means of SSCP. The cDNA coding region (741-738 bp) of the four most frequent alleles were obtained from PBL (Peripheral Blood Leukocytes) total RNA by using 5'-UTR and 3'-UTR primers. According to accepted nomenclature rules (Klein et al., 1990), these alleles were designated *Safo-DAB*0101* to *Safo-DAB*0401*. Partial coding sequences of two alleles, denoted *Safo-DAB*0501* and *Safo-DAB*0601*, were obtained from amplified exon 2 by PCR on genomic DNA using primers in intron 1 and intron 2 because of the impossibility to proceed blood extraction. The sequences of these two far less frequent alleles cover the complete polymorphic β 1 domain. The nucleotide alignment of the six coding sequences of brook charr MH class II β alleles are shown in Figure 2.1 (GenBank accession nos. [EU478851](#), [EU478852](#), [EU478853](#), [EU478854](#), [EU478855](#) and [EU478856](#)). The frequency of each allele in the initial population was: *DAB*0101* – 8.73%, *DAB*0201* – 39.7%, *DAB*0301* – 17.5%, *DAB*0401* – 30.2%, *DAB*0501* – 3.17% and *DAB*0601* – 0.79%. The open reading frame of *Safo-DAB* alleles encodes a 245-246 residues long polypeptide chain. Alignment of the sequences with teleosts and mammal representative sequences revealed a characteristic structure typical of the MHC class II β chain (Fig. 2.2). The molecule is composed of a hydrophobic signal peptide, a polymorphic β 1 domain, a conserved β 2 domain, a hydrophobic transmembrane region, and a short cytoplasmic tail. The four conserved cysteine residues forming the disulfide bridges in the β 1 and β 2 domain are preserved. A potential N-linked glycosylation site (N-X-S/T) is observed in the β 1 domain. From a crystallographic study of the mouse homolog, Wang and co-workers (2001) identified twelve conserved CD4-binding residues (STLNGDWTFQLV) of

the MHC class II β chain. Seven of these residues (STNGDWQ) are also found in the β 2 domain of the brook charr.

2.4.2. Polymorphism of MH class II β

The pairwise nucleotide diversity of the sequences obtained from the four brook charr complete cDNA clones ranged from 3.8% to 5.4% (28 to 40 nucleotides) with a mean of 4.6% (34 nucleotides). Derived amino acid sequences divergence ranged from 7.3% to 11% (18 to 26 residues) with a mean of 8.6% (21 residues) (Table 2.1). The level of diversity was higher in the β 1 domain coding sequences compared to the non- β 1 domain.

Within the polymorphic β 1 domain of the six reported sequences from brook charr, the mean pairwise nucleotide divergence was 9.5% (range: 3.0% to 12%), while it was 18% (7.6% to 22%) for the amino acids. Among the teleost β 1 domain amino acid sequences, 25 positions were conserved whereas a deletion of one residue was shared by the *Safo-DAB* alleles (Fig. 2.2). A deletion of two residues was shared in the alleles *DAB*0401* and *DAB*0601* that also showed the greatest similarity between the alleles with only 7.6% divergence at amino acid variation. Sequence comparison showed that 30 of the 91 residues were polymorphic, each involving replacement of two to five different amino acids. The polymorphic positions found in *Safo-DAB* alleles corresponded well with the PBR (Peptide Binding Region) residues identified in HLA-DRB (Brown et al., 1993), particularly in the N-terminal region of the domain.

Non- β 1 domain coding sequences were much less variable, with pairwise divergence values ranging from 0.85% to 2.8% (mean of 1.7%) for nucleotides and from 1.3% to 3.9% (mean of 2.37%) for amino acid positions. The 20 residues long signal peptide of *DAB*0101*

allele differed from the three other alleles at four positions, with the exception of *DAB*0301* which had an additional mutation. The $\beta 2$ domain, connecting peptide, transmembrane region, and the cytoplasmic tail were all highly conserved (Fig. 2.2).

The average rates of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitution and their ratio (d_N/d_S) for the brook charr MH class II β sequences are presented in Table 2.1. For the complete cDNA comparisons, d_S and d_N were similar and the resulting ratio d_N/d_S was 1.1 (± 0.13). In contrast, for the 91 codons of the polymorphic $\beta 1$ domain, the rate of nonsynonymous nucleotide substitutions was 2.2 (± 1.1) times higher than the synonymous nucleotide substitutions rates, providing evidence for the role of positive selection in maintaining diversity in this domain, as generally reported. In contrast, the corresponding ratio for the non- $\beta 1$ domain was 0.31 (± 0.082), being more indicative of purifying selection.

A phylogenetic tree constructed with the exon 2 from the six MH class II β alleles of brook charr indicated that the sequences clustered together and separately from those observed in Atlantic salmon (Hordvik et al., 1993) and Pacific salmon (*Oncorhynchus*) species (Miller and Withler, 1996). The two sequences characterized by an additional deletion, *DAB*0401* and *DAB*0601*, formed a separate outgroup in the brook charr cluster (Fig. 2.3).

2.4.3. Single expressed locus

We sequenced the intron 2 of the six MH class II β alleles identified in brook charr. We identified in all the alleles a minisatellite composed of a 32 nucleotide motif: ACATTACAGTATGACTAGTCAGCTATGTAGTG, which has already been shown to be present in *Salmo salar* (Stet et al., 2002) and as a single motif in Pacific salmon (*Oncorhynchus*) species (Miller and Withler, 1996). With the exception of the minisatellite varying in size, the

identity of intronic sequences ranged from 84% to 99% (mean of 91%). From each fish, only one or two different PCR products covering the intron 2 was obtained, for which a particular length always corresponded to the same allele. In addition to SSCP and sequencing analysis that indicated no more than two alleles per fish, these results strongly suggest that the identified alleles corresponded to a single locus. As for other salmonids (Glamann, 1995; Langefors et al., 2000), it thus appears that a single locus expresses the MH class II β gene in the brook charr.

2.4.4. Analysis of MH class II β in resistance to furunculosis

The survival time of 860 *Aeromonas salmonicida* infected brook charrs was recorded at 4 hour intervals for a total of 90 hours following infection (Fig. 2.4). The mortality started 52 hours after the challenge, peaked at 68 hours and reached a cumulative mortality of 94.4% at the end of the experiment.

Fish harbouring genotypes *DAB*0101/*0101* or *DAB*0101/*0201* showed retarded mortality and superior survival at the end of the experiment compared to the other individuals. In single allele analysis, individuals possessing the MH class II β *DAB*0101* or *DAB*0301* alleles were the most resistant, while those with *DAB*0201* or *DAB*0401* were the most susceptible to furunculosis (Table 2.2). However, fish carrying the *DAB*0101* allele were the only individuals to show a statistically significant increase in resistance to *Aeromonas salmonicida*. Thus, observed mortality for fish carrying the *DAB*0101* allele was 84.03%, which was approximately 10% lower than that observed for fish not possessing this allele.

In the combined (genotype) analysis, two allelic combinations returned statistically significant results. Namely, the genotypes *DAB*0201/*0301* and *DAB*0301/*0401* were associated with increased susceptibility to furunculosis (Table 2.3). The allele *DAB*0101* was

mostly associated with genotypes showing increased resistance, and homozygote genotype *DAB*0101/*0101* was nearly significant ($p = 0.0657$). Thus, survival of individuals with the *DAB*0101/*0101* and *DAB*0101/*0201* genotypes was clearly higher than the other combination of alleles, their respective mortality being 84.62% and 79.71%, or 8.5% to 13.4% lower than for the other allelic combinations on average. Finally, fish with *DAB*0201/*0201* genotypes showed a tendency towards susceptibility, albeit not significant.

2.5. Discussion

2.5.1. Identification of brook charr MH class II β alleles

The goals of this study were (i) to document the diversity of alleles at the MH class II β gene of brook charr and (ii) for the first time, to document survival conferred by specific alleles and/or genotypes following exposure to furunculosis in a fish other than Atlantic salmon. Overall, we observed a relatively low level of allelic diversity in the studied population with a total of six alleles of the MH class II β gene identified in a group of 63 adult brook charr from the Rupert River. The similarities in the amino acid composition between brook charr and other vertebrates β 2 domains, especially salmonids, confirm the four identified alleles as MH class II β genes. The high ratio of d_N/d_S observed for the β 1 domain compared to the remaining of the protein indicates that a positive selection pressure is effective to select polymorphic mutations in that region. Furthermore, their identification from PBL mRNA indicates that alleles *Safo-DAB*0101* to *Safo-DAB*0401* are expressed and ultimately represent the surface molecule in charge of antigen presentation. We also found that one allele, *Safo-DAB*0101*, was significantly associated with resistance against *Aeromonas salmonicida*. In addition to homozygotes for this allele, its resistance effect was also detected in heterozygote individuals for two specific genotypes. In contrast, other allelic combinations, namely heterozygous genotypes *Safo-DAB*0201/*0301* and *Safo-DAB*0301/*0401* were significantly associated with increased susceptibility to furunculosis.

2.5.2. Variability of brook charr MH class II β alleles

The variability of certain amino acids of the peptide binding region of the MHC class II alleles directly influences the binding properties for antigenic peptides as well as

susceptibility to diseases (Jones et al., 2006). It has been shown that a single residue mutated in the $\beta 1$ domain N-terminal region could be correlated with increased resistance (Langefors et al., 2001). We compared the sequences of the $\beta 1$ domain between the *Safo-DAB*0101* allele and the other three alleles present in the challenge test to identify exclusive residues, possibly responsible of a functional effect. Based on *HLA-DRB*0101* numbering, we identified 6 exclusive residues: $\beta 9Y$, $\beta 12L$, $\beta 13A$, $\beta 53Y$, $\beta 69G$ and $\beta 70P$, with clustering of three of these residues between positions 9 and 13. The $\beta 9$ position is particularly interesting for two reasons. First, observed substitution changes the chemical characteristics of the residue at this position since it is changed from a positively ($\beta 9H$) or negatively ($\beta 9E$) charged one to an aromatic polar uncharged one ($\beta 9Y$). Second, this position is just beside the $\beta 10H$ residue that has been identified to be exclusively responsible for the emergence of resistance to furunculosis in European populations of Atlantic salmon (Langefors et al., 2001). It is interesting to note that $\beta 10H$ is also found in *Safo-DAB*0101*. Based on human MHC class II three-dimensional structure (Brown et al., 1993), this specific single changed residue identified by Langefors et al. (2001) corresponds to the $\beta 10Q$ position and is situated in the pocket 6 (P6) of the molecule (Jones et al., 2006). The presence of highly polymorphic charged residues is frequently observed in salmonids at corresponding positions between $\beta 9W$ and $\beta 13F$ of the *HLA-DRB1*0101* (Fig. 2.2). This sequence of residues is located in the bottom of the groove between pocket 6 (P6) and pocket 4 (P4), and the presence of particular residues could be critical for the efficient presentation of *Aeromonas salmonicida* antigens. From these studies, it appears that the polymorphic residues of the $\beta 1$ domain present in the pockets P4 and P6 could be important for the binding of antigenic peptide(s) from *Aeromonas salmonicida*. Namely, residue $\beta 53Y$ is located in the middle of an interface of dimerisation

between two MH class II molecules and could possibly affect the stability or the structure of the dimer (Brown et al., 1993). Residues $\beta 69G$ and $\beta 70P$ are both located at corresponding positions, in *HLA-DR1* and *I-Ak*, that have been implicated in the binding to the TCR (Hennecke et al., 2000). The presence of a proline instead of the usually conserved glutamine at the position $\beta 70$ of MH class II could have an important impact on immune system stimulation since this residue has the highest average number of contacts with the TCR (Rudolph et al., 2006). From the six exclusive residues identified for the *Safo-DAB*0101* allele, one or more could be responsible for the increase resistance of fish bearing it, considering their position in one of the three functional region identified.

2.5.3. Polymorphism of MH class II β leader sequences

As integral membrane protein, MHC class II β nascent polypeptide chain contains an amino-terminal signal sequence (signal peptide) that is bound by the SRP (Signal Recognition Particle) for protein targeting (Doudna and Batey, 2004). It has been demonstrated that reduction in either the net positive charge or the hydrophobicity of the signal peptide decreases the effectiveness of SRP recognition (Peterson et al., 2003). Comparison of the signal sequence of the *DAB*0101* allele with the three other available sequences indicates two exclusive mutations. The replacement of the aspartic acid by an alanine (D6A) modifies the charged residue in amino-terminal of the signal sequence, while the phenylalanine by an isoleucine mutation (F14I) modestly alters the hydrophobicity. Compared to the other alleles, the differences in the signal sequence of the *DAB*0101* allele could affect both SRP binding and membrane integration leading to differences in cell surface expression and presentation. Experiments are currently in progress to evaluate the surface expression levels of these different alleles.

2.5.4. Minisatellite in MH class II β intron 2

It is noteworthy that the minisatellite (motif: ACATTACAGTATGACT-AGTCAGCTATGTAGTG) present in brook charr intron 2 also occurred in Atlantic salmon (Stet et al., 2002), but only as a single motif in Pacific salmon (*Oncorhynchus*) species (Miller and Withler, 1996). This result correlates and reinforces the phylogenetic tree constructed from $\beta 1$ exonic sequences which establish a closer relationship between brook charr and Atlantic salmon alleles relative to Pacific salmon (Fig. 2.3). However, the identity of the intron 2 nucleotide sequences between the three species is highly conserved, suggesting a single and same active locus for the MH class II β genes. The single minisatellite motif found in Pacific salmon (*Oncorhynchus*) species is located between a partial *Hpa* 1 SINE (Short Interspersed Nucleotide Elements) and a complete *Hpa* 1 SINE (Miller and Withler, 1996), both absent from brook charr and Atlantic salmon. Minisatellites and SINEs are known to be implicated in recombination and gene conversion events (for reviews see: Jeffreys et al., 2004; Kazazian, 2004), raising the hypothesis that different molecular mechanisms may have been selected as a means to maintain high levels of MH class II β gene polymorphism among salmonid species.

2.5.5. Association of MH class II β alleles with resistance or susceptibility

In order to identify MH class II alleles associated with resistance or susceptibility to *Aeromonas salmonicida*, we performed a challenge experiment on 23 half- and full-sibling families. The observed resistance or susceptibility effect can falsely be attributed to a particular MH allele if no precautions are undertaken to control the genetic background effect. Here we could not integrate the genetic background effect in our statistical model because of insufficient degrees of freedom. However, the correlation of the allele *DAB*0101* with resistance to furunculosis resulted from the analysis of individuals issued from seven families

for which two different sires possessing the allele, and/or three different dams, have been utilized in the mating procedure. Thus, we believe that the use of families in this experiment minimized the effect of an eventual variable co-segregating gene while permitting variation at other unlinked genes.

From the four alleles that have been tested in the challenge experiment against *Aeromonas salmonicida*, the *DAB*0101* allele was significantly correlated with disease resistance in single analysis (Table 2.2). This result is strengthened by the combined analysis where fish homozygous for this allele, *DAB*0101/*0101*, also tended to be resistant to furunculosis (Table 2.3), which translated into a lower mortality in fish carrying allele *DAB*0101* in general and those homozygote at this allele in particular. Conversely, the results of the allele *DAB*0201* in single analysis and genotype *DAB*0201/*0201* in combined analysis are both suggestive of disease susceptibility (Table 2.2 and 2.3). Interestingly, the relatively high resistance observed for the *DAB*0101/*0201* genotype also suggested that the dominant effect on disease resistance associated with *DAB*0101* was stronger than the dominant susceptibility conferred by *DAB*0201*. This type of effect has been demonstrated previously in Atlantic salmon (Lohm et al., 2002) where a co-dominant pattern of disease resistance/susceptibility was found, indicative of qualitative difference in the immune response between individuals carrying the high- and low-resistance alleles. Overall, the resistance or susceptibility to furunculosis infection seemed to be associated more to specific MH class II β alleles rather than MH heterozygosity. Admittedly however, not enough individuals of the *DAB*0101/*0401* genotype were available to evaluate the dominance of *DAB*0101* over *DAB*0401*.

The result obtained for fish harbouring genotype *DAB*0301/*0301*, although not statistically significant, was suggestive of increased resistance, whereas the high ratio values obtained for *DAB*0201/*0301* and *DAB*0301/*0401* were more suggestive of increased susceptibility. On the other hand, when looking at the single analysis for the *DAB*0301* allele, the intermediate hazard ratio was indicative of a null or confounded effect. Thus, it is possible that the apparent increased susceptibility in heterozygotes carrying this allele could have been caused by the alternate alleles to which it was associated in given heterozygous genotype such as allele *DAB*0201* and *DAB*0401* which gave high hazard ratios, albeit without statistically significant effect. Overall, these results are suggestive of a dominant effect of *DAB*0201* and *DAB*0401* alleles over *DAB*0301*.

2.5.6. Mechanism for pathogen-driven balancing selection

The negative frequency dependent selection hypothesis may potentially explain the advantage of the allele *DAB*0101* over the other more common alleles, and underlines the importance of this mechanism in sustaining variation at the MH in brook charr. Thus, the *DAB*0101* allele was present in only 10 of the 63 brook charr (16%) originally used to generate the families, whereas the allele *DAB*0201* (76%) that was widely distributed in this population tended to be associated with disease susceptibility. It should be noted that there has been no report of a bacterial outbreak in the Rupert River population that could have led to a pre-selection of the more performing MH class II alleles. It is then reasonable to hypothesize that these allele distributions within the wild brook charr population from that system could reflect the adaptation of these fish to the dominant pathogen present in that river (Hedrick, 2002). Since we know that bacterial diversity represents an important factor to explain the MHC amino acid variability in wild salmonid populations (Dionne et al., 2007), it should be

very useful for future studies of brook charr to combine the fish sampling with an analysis of the bacterial communities present in the water as well.

In population infection studies, an important factor influencing which MHC polymorphism mechanism of evolution will be observed is the type of infection the individual is facing. Single-pathogen infection studies failed to support the heterozygote advantage (Ilmonen et al., 2007), while others, in support to the frequency-dependent selection, demonstrated a selective benefit for individuals expressing rare alleles (Langefors et al., 2001; Lohm et al., 2002; Trachtenberg et al., 2003). Regarding the heterozygous advantage hypothesis, a recent study examined multiple pathogen infections and found a pattern of MHC superiority and reduced pathogen load in heterozygotes over homozygotes (McClelland et al., 2003). Penn et al. (2002) also showed that MHC heterozygotes were more resistant to infection and had higher fitness than homozygotes in multiple-strain infections.

Overall, the results of this study did not support a heterozygous advantage against a single pathogen. For example, the majority of the heterozygous individuals have hazard ratios indicative of susceptibility to furunculosis whereas three out of four homozygous genotypes of the combined analysis presented significant or a tendency towards resistance hazard ratios (Table 2.3). Homozygosity at the MH class II locus has already been suggested to be an advantage against furunculosis in Atlantic salmon (Grimholt et al., 2003). However, as for the brook charr in the present study, these fish were facing a single pathogen and the infection with multiple pathogens could have potentially revealed heterozygous advantage, assuming they are able to present a wider array of antigen peptides to T-cells to initiate the specific immune response (Hughes and Nei, 1992).

The selected infection mode by bath immersion (Lutwyche et al., 1995) represents a valuable standardized technique, where all fish used in this experiment were exposed to almost the same bacterial charge at the same time and were eventually presenting symptoms characteristic of furunculosis as: darkening, tachybranchia, lethargy and haemorrhages (Bernoth et al., 1997). The synchronicity of the infection could explain the period of high mortality rate that we observed (around 40h) compared to published results of 6, 10 or 12 days with co-habitation infection method (Lohm et al., 2002; Grimholt et al., 2003; Kjøglum et al., 2006). Questions were previously raised regarding the rapid mortality occurring in these types of 'artificial' infections, especially pertaining to the sole contribution of the adaptive immunity (Grimholt et al., 2003). Because of the correlation of MH class II β genotype with survival in this model with a high mortality rate (< 4 days), another avenue of investigation could be the effect of MH genotypes in the initiation of the immune response. Future studies could address the evaluation of early immune functions following infection in function of MH class II genotype. Those studies could include assay for the activation and proliferation of lymphocytes, as well as phagocytic activity and cytokine quantification (Kollner et al., 2002).

To conclude, this study brings further support to the hypothesis that specific MH class II β alleles can be associated to disease resistance or susceptibility in the brook charr, as reported previously for Atlantic salmon (Langefors et al., 2001; Lohm et al., 2002; Grimholt et al., 2003; Kjøglum et al., 2006). The results also demonstrated that the level of resistance or susceptibility is affected by specific allele' combinations, which suggests different levels of dominance between them. However, the results presented in this study were obtained from a single pathogen infection, and it is likely that the fitness of different alleles would differ for different pathogens and shifts over time (Penn and Potts, 1999). Consequently, future studies

on the brook charr should compare different pathogens and combined infections. Even if precautions have been taken to control it, we cannot totally exclude a genetic background effect of co-varying genes closely linked to the MH class II β genes in the observed results. Finally, it should be noted that the resistance or susceptibility effect could also be imputed to particular residues on the alpha chain of the MH class II. Knowing that they are co-segregated (Stet et al., 2002), it should be interesting to sequence both chains of the MH class II and eventually modelize the different proteins. In combination, the identification of the major antigenic peptides derived from *Aeromonas salmonicida* could also be of importance for our understanding of the virulence of this pathogen and to refine the presentation complex model. The identification of a theoretical consensus resistance allele from sequences of different salmonids could also be relevant for the prediction of performance of new alleles to be discovered.

2.6. Acknowledgements

We wish to dedicate this study to the late Rene Stet, a pioneer in the study of adaptive significance of MHC diversity in fishes. This work was supported by Valorisation Recherche Québec (VRQ) through the auspices of the Réseau Aquaculture Québec (RAQ). As members of the former AquaNet Network, we would like to thank AquaNet and the Networks of Centres of Excellence Programs and Councils for providing funds. S.C. wishes to thank the Faculty of Graduate Studies of the Université de Montréal for partial financial support. Finally, we would also like to thank Martin Clément and Julie Shareck for helpful discussions, and Patrick C. Hallenbeck for critical reading of the manuscript.

2.7. References

- Aguilar, A., Garza, J.C., 2007. Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. *J. Mol. Evol.* 65, 34-43.
- Aoyagi, K., Dijkstra, J.M., Xia, C., Denda, I., Ototake, M., Hashimoto, K., Nakanishi, T., 2002. Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). *J. Immunol.* 168, 260-273.
- Bacon, L.D., 1987. Influence of the major histocompatibility complex on disease resistance and productivity. *Poult. Sci.* 66, 802-811.
- Bell, J.I., Estess, P., St John, T., Saiki, R., Watling, D.L., Erlich, H.A., McDevitt, H.O., 1985. DNA sequence and characterization of human class II major histocompatibility complex beta chains from the DR1 haplotype. *Proc. Natl. Acad. Sci. U.S.A.* 82, 3405-3409.
- Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* 16, 363-377.
- Bernoth, E.M., Ellis, A.E., Midtlyng, P.J., Olivier, G., Smith, P. (Eds.), 1997. *Furunculosis: multidisciplinary fish disease research*. Academic Press, San Diego, CA.
- Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W.S., Johnson, S.L., Gates, M., Postlethwait, J.H., Klein, J., 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46, 129-134.
- Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364, 33-39.

- Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J., Allen, R.C., 1991. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* 48, 137-144.
- Carrington, M., O'Brien, S.J., 2003. The influence of HLA genotype on AIDS. *Annu. Rev. Med.* 54, 535-551.
- Collett, D., 1991. *Modelling Binary Data*. Chapman et Hall, London.
- Cox, D.R., 1972. Regression Models and Life Tables. *J. R. Stat. Soc. [Ser B]*. 20, 187-220.
- Daly, J.G., Kew, A.K., Moore, A.R., Olivier, G., 1996. The cell surface of *Aeromonas salmonicida* determines in vitro survival in cultured brook trout (*Salvelinus fontinalis*) peritoneal macrophages. *Microb. Pathog.* 21, 447-461.
- Dautremepuits, C., Fortier, M., Croisetièrre, S., Belhumeur, P., Fournier, M., 2006. Modulation of juvenile brook trout (*Salvelinus fontinalis*) cellular immune system after *Aeromonas salmonicida* challenge. *Vet. Immunol. Immunopathol.* 110, 27-36.
- de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H.J., Jordan, W., Cross, T., Stet, R.J., 2007. Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. *Proc. Biol. Sci.* 274, 861-869.
- Dionne, M., Miller, K.M., Dodson, J.J., Caron, F., Bernatchez, L., 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution* 61, 2154-2164.
- Dixon, B., Stet, R.J., 2001. The relationship between major histocompatibility receptors and innate immunity in teleost fish. *Dev. Comp. Immunol.* 25, 683-699.

- Doudna, J.A., Batey, R.T., 2004. Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* 73, 539-557.
- Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783-791.
- Garduno, R.A., Moore, A.R., Olivier, G., Lizama, A.L., Garduno, E., Kay, W.W., 2000. Host cell invasion and intracellular residence by *Aeromonas salmonicida*: role of the S-layer. *Can. J. Microbiol.* 46, 660-668.
- Gebe, J.A., Swanson, E., Kwok, W.W., 2002. HLA class II peptide-binding and autoimmunity. *Tissue Antigens* 59, 78-87.
- Gjedrem, T., 2000. Genetic improvement of cold water fish species. *Aquacult. Res.* 31, 25-33.
- Glamann, J., 1995. Complete coding sequence of rainbow trout Mhc II beta chain. *Scand. J. Immunol.* 41, 365-372.
- Gonzales, M.J., Dugan, J.M., Shafer, R.W., 2002. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* 18, 886-887.
- Grimholt, U., Drablos, F., Jorgensen, S.M., Hoyheim, B., Stet, R.J., 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* 54, 570-581.
- Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., Stet, R.J., 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55, 210-219.

- Hedrick, P.W., 2002. Pathogen resistance and genetic variation at MHC loci. *Int. J. Org. Evolution* 56, 1902-1908.
- Hennecke, J., Carfi, A., Wiley, D.C., 2000. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J.* 19, 5611-5624.
- Hordvik, I., Grimholt, U., Fosse, V.M., Lie, O., Endresen, C., 1993. Cloning and sequence analysis of cDNAs encoding the MHC class II beta chain in Atlantic salmon (*Salmo salar*). *Immunogenetics* 37, 437-441.
- Hughes, A.L., Nei, M., 1992. Models of host-parasite interaction and MHC polymorphism. *Genetics* 132, 863-864.
- Hughes, A.L., Nei, M., 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U.S.A.* 86, 958-962.
- Ilmonen, P., Penn, D.J., Damjanovich, K., Morrison, L., Ghotbi, L., Potts, W.K., 2007. Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics* 176, 2501-2508.
- Jeffreys, A.J., Holloway, J.K., Kauppi, L., May, C.A., Neumann, R., Slingsby, M.T., Webb, A.J., 2004. Meiotic recombination hot spots and human DNA diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 141-152.
- Jones, E.Y., Fugger, L., Strominger, J.L., Siebold, C., 2006. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* 6, 271-282.
- Kazazian, H.H. Jr., 2004. Mobile elements: drivers of genome evolution. *Science* 303, 1626-1632.

- Kjoglum, S., Larsen, S., Bakke, H.G., Grimholt, U., 2006. How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol.* 21, 431-441.
- Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K., Watkins, D.I., 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* 31, 217.
- Kollner, B., Wasserrab, B., Kotterba, G., Fischer, U., 2002. Evaluation of immune functions of rainbow trout (*Oncorhynchus mykiss*) - how can environmental influences be detected? *Toxicol. Lett.* 131, 83-95.
- Langefors, A., Lohm, J., Von Schantz, T., Grahn, M., 2000. Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. *Mol. Ecol.* 9, 215-219.
- Langefors, A., Lohm, J., Grahn, M., Andersen, O., von Schantz, T., 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. Biol. Sci.* 268, 479-485.
- Leemis, L.M., Trivedi, K.S., 1996. A comparison of approximate interval estimators for the Bernoulli parameter. *The American Statistician* 50, 63-68.
- Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., von Schantz, T., 2002. Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc. Biol. Sci.* 269, 2029-2033.

- Lutwyche, P., Exner, M.M., Hancock, R.E., Trust, T.J., 1995. A conserved *Aeromonas salmonicida* porin provides protective immunity to rainbow trout. *Infect. Immun.* 63, 3137–3142.
- McClelland, E.E., Penn, D.J., Potts, W.K., 2003. Major histocompatibility complex heterozygote superiority during coinfection. *Infect. Immun.* 71, 2079-2086.
- Miller, K.M., Withler, R.E., 1996. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43, 337-351.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418-426.
- Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H., Klein, J., 1992. Major histocompatibility complex class II genes of zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11886-11890.
- Ono, H., O'hUigin, C., Vincek, V., Stet, R.J., Figueroa, F., Klein, J., 1993. New beta chain-encoding Mhc class II genes in the carp. *Immunogenetics* 38, 146-149.
- Penn, D.J., Damjanovich, K., Potts, W.K., 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11260-11264.
- Penn, D.J., Potts, W.K., 1999. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* 153, 145–164.
- Perry, G.M.L., Tarte, P., Croisetiere, S., Belhumeur, P., Bernachez, L., 2004. Genetic variance and covariance for 0+ weight and resistance to *Aeromonas salmonicida* in the brook charr (*Salvelinus fontinalis*). *Aquaculture* 235, 263-271.

Peterson, J.H., Woolhead, C.A., Bernstein, H.D., 2003. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J. Biol. Chem.* 278, 46155-46162.

Rudolph, M.G., Stanfield, R.L., Wilson, I.A., 2006. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* 24, 419-466.

Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4, 406-425.

Sato, A., Figueroa, F., Murray, B.W., Malaga-Trillo, E., Zaleska-Rutczynska, Z., Sultmann, H., Toyosawa, S., Wedekind, C., Steck, N., Klein, J., 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51, 108-116.

Shiina, T., Inoko, H., Kulski, J.K., 2004. An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 64, 631-649.

Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M.A., Hedrick, R.P., Nehring, R.B., Stet, R.J., Secombes, C., Parham, P., 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *J. Immunol.* 166, 3297-3308.

Singer, D.S., Mozes, E., Kirshner, S., Kohn, L.D., 1997. Role of MHC class I molecules in autoimmune disease. *Crit. Rev. Immunol.* 17, 463-468.

Slade, R.W., McCallum H.I., 1992. Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* 132, 861-864.

Stet, R.J., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B.P., Grimholt, U., 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II

B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* 54, 320-331.

Sutton, T.M., Pangle, K.L., Greil, R.W., 2002. Hatchery performance of Nipigon, Assinica and Iron River strains of age-0 brook trout. *N. Am. J. Aquac.* 64, 188–194.

Syed, M., Vestrheim, O., Mikkelsen, B., Lundin, M., 2003. Isolation of the promoters of Atlantic salmon MHCII genes. *Mar. Biotechnol. (NY)* 5, 253-260.

Takahata, N., Nei, M., 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124, 967–978.

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

Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030-11035.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., Kunstman, K., Wu, S., Phair, J., Erlich, H., Wolinsky, S., 2003. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* 9, 928-935.

Trowsdale, J., Parham, P., 2004. Mini-review: defense strategies and immunity-related genes. *Eur. J. Immunol.* 34, 7-17.

Wang, J.H., Meijers, R., Xiong, Y., Liu, J.H., Sakihama, T., Zhang, R., Joachimiak, A., Reinherz, E.L., 2001. Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10799-10804.

Zekarias, B., Ter Huurne, A.A., Landman, W.J., Rebel, J.M., Pol, J.M., Gruys, E., 2002. Immunological basis of differences in disease resistance in the chicken. *Vet. Res.* 33, 109-125.

2.8. Tables and Figures

Table 2.1

Mean diversity in the nucleotides (nt.) and amino acid (AA) sequences, mean proportion of observed synonymous (d_S) and nonsynonymous substitutions (d_N) and the ratio between d_N and d_S of brook charr MH class II β gene

MH class II β sequence	N	Diversity (%)		d_S	d_N	d_N/d_S
		nt.	AA			
Complete cDNA	4	4.6 [3.8-5.4]	8.6 [7.3-11]	0.046 (0.007)	0.048 (0.007)	1.1 (0.13)
β 1 domain	6	9.5 [3.0-12]	18 [7.6-22]	0.061 (0.027)	0.11 (0.029)	2.2 (1.1)
Non- β 1 domain	4	1.7 [0.85-2.8]	2.4 [1.3-3.9]	0.036 (0.014)	0.012 (0.007)	0.31 (0.082)

N gives the number of alleles included in the analysis. Standard errors in parentheses and ranges are in brackets.

Table 2.2

Hazard ratio in the single allele groups with at least one of the given class II β (DAB) alleles present

Allele	Hazard ratio	Confidence (Pr > ChiSq)	No. of fish
<i>DAB*0101</i>	0.7187 (0.1201)	0.0480	119
<i>DAB*0301</i>	0.9896 (0.1454)	0.9434	388
<i>DAB*0401</i>	1.1133 (0.1493)	0.4235	412
<i>DAB*0201</i>	1.2255 (0.1918)	0.1938	453

Standard errors are in parentheses. Significant Hazard ratio ($p < 0.05$) is indicated in bold.

Table 2.3

Hazard ratio and mortality in groups with class II β (DAB) genotypes

Genotype	Hazard ratio (S.E.)	Confidence (Pr > ChiSq)	Mortality (%) (C.I.)	No. of fish
<i>DAB*0101/*0101</i>	<i>0.6283</i> (0.1586)	0.0657	84.62 (65.13-95.64)	26
<i>DAB*0301/*0301</i>	0.7790 (0.1382)	0.1593	93.10 (83.27-98.09)	58
<i>DAB*0101/*0201</i>	0.7944 (0.1427)	0.2002	79.71 (68.31-88.44)	69
<i>DAB*0401/*0401</i>	0.8662 (0.1131)	0.2715	95.71 (90.91-98.41)	140
<i>DAB*0101/*0301</i>	1.0090 (0.3028)	0.9763	94.44 (72.71-99.86)	18
<i>DAB*0201/*0401</i>	1.0256 (0.1417)	0.8547	97.20 (92.02-99.42)	107
<i>DAB*0201/*0201</i>	<i>1.2886</i> (0.1695)	0.0538	96.77 (91.95-99.11)	124
<i>DAB*0301/*0401</i>	1.2942 (0.1605)	0.0376	95.60 (91.14-98.21)	159
<i>DAB*0101/*0401</i>	1.3027 (0.6178)	0.5771	100.0 (54.07-100.0)	6
<i>DAB*0201/*0301</i>	1.3207 (0.1642)	0.0253	96.73 (92.54-98.93)	153

Standard errors (S.E.) and confidence intervals (C.I.) are in parentheses. Significant Hazard ratios ($p < 0,05$) are presented in bold whereas tendencies ($0,05 < p < 0,075$) are shown in italic.

```

M S M P I A F Y I C L T L I L S I F Y G I D G Y F Y H R L A
Safo-DAB*0101 ATGTCGATGCCAATTGCCTTCTATATTTGCCTGACCTTGATTTTGTCCATATTCATGGAATAGATGGATATTTTATCATAGGTTGGCA 90
Safo-DAB*0201 .....C...A.....C.....TC..T.....T.....C.....G.....G.A....CG...T. 90
Safo-DAB*0301 .....T..C...A.....C.....TC..T.....T.....C.....C.....C.....G...G...G. 90
Safo-DAB*0401 .....C...A.....C.....TC..T.....C.....C.....G.A..GGTTG..AG. 90
Safo-DAB*0501 -----C.....G.T.G..AA. 27
Safo-DAB*0601 -----G.A..GGTTG..AG. 27

Q C R Y S S K D L H G I E F I D S Y Y F N Q A E Y V R F N S
Safo-DAB*0101 CAGTGCCGATACTCCTCAAAGGACCTGCATGGTATAGAGTTTATAGACTCTTACTACTTCAATCAGGCTGAATATGTCAGATTCAACAGC 180
Safo-DAB*0201 .....GC.....T...G.A..... 180
Safo-DAB*0301 .....TGTT.....A.....A..... 180
Safo-DAB*0401 .....TGTT.....T..... 180
Safo-DAB*0501 .....GC.....TGTT.....T...C.AA..... 117
Safo-DAB*0601 .....TGTT.....A..... 117

T V G K Y V G Y T E Y G V K N A E A W N K G S E L G P E L G
Safo-DAB*0101 ACTGTGGGGAAGTATGTTGGATACACTGAGTATGGTGTGAAGAATGCAGAAGCATGGAACAAAGGTTCTGAGCTGGGTCCAGAGCTAGGG 270
Safo-DAB*0201 .....CTG...T.....G.....C..A..... 270
Safo-DAB*0301 .....CTG...T.....A.....T.....C..A..... 270
Safo-DAB*0401 .....C.....A.....T.....G.A..***.C...C..A..... 267
Safo-DAB*0501 .....CTG...T..... 207
Safo-DAB*0601 .....T.....G.A..***.C...C..A..... 204

E L E R V C K R N A A I Y Y G A V L D K T V E P H V R L S S
Safo-DAB*0101 GAGCTGGAGCGTGTCTGCAAGCGTAACGCTGCTATCTACTACGGCGCGTACTGGATAAGACAGTTGAGCCCCATGTCAGACTGAGCTCA 360
Safo-DAB*0201 .....A..... 360
Safo-DAB*0301 .....A..TA.....A.....A..AA..C..... 360
Safo-DAB*0401 .....A.....A.....A.C..C...A.....A..... 357
Safo-DAB*0501 .....T...C.....A.....A.....C..... 270
Safo-DAB*0601 .....T.....TA.....A.C..C.....A.....C..... 267

V T P P S G R H P A M L M C S A Y D F Y P K P I R V T W L R
Safo-DAB*0101 GTGACTCCCCCTAGTGGCAGACACCCCTGCCATGCTGATGTGCAGCGCCTATGACTTCTACCCCAACCAATCAGAGTGACCTGGCTGAGG 450
Safo-DAB*0201 ..... 450
Safo-DAB*0301 ..... 450
Safo-DAB*0401 .....T..... 447
Safo-DAB*0501 ----- 270
Safo-DAB*0601 ----- 267

D G H E V K S D V T S T E E L A N G D W Y Y Q I H S H L E Y
Safo-DAB*0101 GACGGACATGAGGTGAAGTCTGATGTGACCTCCACTGAGGAGCTGGCTAACGGGACTGGTACTACCAGATCCACTCCCACCTGGAGTAC 540
Safo-DAB*0201 ..... 540
Safo-DAB*0301 ..... 540
Safo-DAB*0401 ..... 537
Safo-DAB*0501 ..... 270
Safo-DAB*0601 ----- 267

T P K S G E K I S C M V E H I S L T E P M M Y H W D P S L P
Safo-DAB*0101 ACACCCAAGTCTGGAGAGAAGATCTCCTGTATGTTGGAGCACATCAGCCTGACTGAGCCCATGATGTATCACTGGGACCCGTCCTGCCT 630
Safo-DAB*0201 ..... 630
Safo-DAB*0301 ..... 630
Safo-DAB*0401 .....A..... 627
Safo-DAB*0501 ..... 270
Safo-DAB*0601 ----- 267

E A E R N K I A I G A S G L V L G T V L A L A G L I Y Y K K
Safo-DAB*0101 GAGGCTGAGAGGAATAAGATCGCGATCGGGCGTCTGGTCTGGTTCTGGGAACCGTCTGGCATTAGCAGGACTGATCTACTACAAGAAG 720
Safo-DAB*0201 ..... 720
Safo-DAB*0301 .....G.....G.....A..... 720
Safo-DAB*0401 .....G.....G..... 717
Safo-DAB*0501 ----- 270
Safo-DAB*0601 ----- 267

K S S G V L *
Safo-DAB*0101 AAGTCTTCTGGGGTGCCTAG 741
Safo-DAB*0201 ..... 741
Safo-DAB*0301 ..... 741
Safo-DAB*0401 ..... 738
Safo-DAB*0501 ----- 270
Safo-DAB*0601 ----- 267

```

Figure 2.1. Nucleotide sequences of the MH class II β cDNA and partial gDNA of six different alleles from brook charr. On top is the amino acid sequence translated from *Safo-DAB*0101*.

Figure 2.2. Alignment of brook charr MH class II β amino acid sequences with MH class II β sequences from other vertebrates. Protein domains are indicated. Dots indicate identity with *Safo-DAB*0101*; asterisks, gaps introduced to maximize the alignment; dashes, unavailability of sequence information; cysteine residues are marked with light gray boxes; putative N-linked glycosylation sites are underlined; conserved CD4-binding residues are double underlined. Numbering below and residues involved in antigen binding denoted by dots above the β 1-domain sequence refer to HLA-DRB*0101 (Brown et al., 1993) CP: connecting peptide. TM: transmembrane. In addition to brook charr (*Salvelinus fontinalis*: Safo) sequences, *Oncorhynchus mykiss* (Onmy), *Salmo salar* (Sasa), *Salmo trutta* (Satr), *Cyprinus carpio* (Cyca), *Brachydanio rerio* (Brre) and human (HLA) sequences are aligned.

	Signal peptide	β 1-domain		
Safo-DAB*0101	-----MSMPIAFYICLTLLLSIFY*G	IDGYFYHRLAQCRYSSKDLHGIEFIDS		
Safo-DAB*0201	-----LD.....F....S..	T....E.SVV.....A....		
Safo-DAB*0301	-----I.LD.....F....S..	T....HQ.VG.....		
Safo-DAB*0401	-----LD.....F....S..	T....EQVVR.....		
Safo-DAB*0501	-----	-...HQMVK.....A....		
Safo-DAB*0601	-----	-...EQVVR.....		
Onmy-DAB*0101	-----LW...S..	T....EQVVR.....		
Sasa-DAB*0101	-----MSI.CVS...V...S..	T....EQVVR.....Q....		
Satr-DAB*0101	-----	----EQ.VSE.....Q.A....		
Cyca-DAB*0101	-----L.KVLCFH.I.M.A.TGA	AN..YHSWWTK.IH..R.FSDMVY..N		
Brre-DAB*0101	-----VK.YLLILFLAI.M..T.TGT	A...YQYTMLE.I..TS.YSDMVLL.		
HLA-DRB1*0101	MVCLKLPGGSC.TALTVTLMV.SSP.ALAGDT	RPRFLWQLKFE.HFFN*GTERVRLLER		
		1 10 20		
Safo-DAB*0101	Y Y F N Q A E Y V R F N S T V G K Y V G Y T E Y G V K N A B A W N K G * S E L G P E L G E L E R V C K R N A A I Y Y G A V			
Safo-DAB*0201	...V.DI.....L.L.....AQ.....H.....			
Safo-DAB*0301	.V..K...I.....L.L...T...V.AQ.....SNA			
Safo-DAB*0401	.V...V.....H.....V..R*.DY.AQ.....H..DLH.S.I			
Safo-DAB*0501	.V...V.QI.....L.L.....P..D....I			
Safo-DAB*0601	.V..K.....V..R*.DY.AQ.....F..Y..DLH...I			
Onmy-DAB*0101	.V..K..H.....R.....H.L.....SDAGI..Q.QA...Y..PS...D.S.I			
Sasa-DAB*0101	.V..K...I.....F.....L.....SDAAV.AV.R...Y..H..DLH.STI			
Satr-DAB*0101	...V.HI.....H..Y.....SDAGI.AQ.R...Y..QH.P.A.S.I			
Cyca-DAB*0101	.I..KDV.IQ.....E.....AL..Y..RF..DPNI.QQ.RAQV..Y..H..EL.QS.I			
Brre-DAB*0101	GS..KVV.D.QY.....Q..IF.RNF..NQAY.QQRKA.V.SF..RH..Q.SDS..			
HLA-DRB1*0101	CIY..E.S...D.D..E.RAV..L.RPD..Y..SQKDL.EQRRAAVDTY..RH.YGVGESFT			
	30 40 50 60 70 80 90			
		β 2-domain		
Safo-DAB*0101	LDKT VEPHVRLSSVTPPSGRHPAMLMCSAYDFYKPIRVTWLRDGEVKSVDVTS	TEELAN		
Safo-DAB*0201			
Safo-DAB*0301			
Safo-DAB*0401F.....			
Safo-DAB*0501	-----			
Safo-DAB*0601	-----			
Onmy-DAB*0101Q.....R.....			
Sasa-DAB*0101A.....R.....			
Satr-DAB*0101A..I.....G.....			
Cyca-DAB*0101	R...P.K.K...K.AG...V...R..HG.K.S.M.N.EV..T...MP.			
Brre-DAB*0101	R..A..K.K.TIQ..MQAE.K...L.D.E...K.KMS...DKV.T...I.M..			
HLA-DRB1*0101	VQRR...K.TVYPSKTQPLQ.HNL.V...VSG...GS.E.R.F.N.Q.E.AG.V..GLIQ.			
	94			
Safo-DAB*0101	<u>GDWYYQIHS</u> HSHLEYTPKSGEKISCMVEHISLTEPMMYHW	CP	TM region	
Safo-DAB*0201	DPSLPEAERNK	IAIGASGLV	
Safo-DAB*0301			
Safo-DAB*0401			
Safo-DAB*0501	-----			
Safo-DAB*0601	-----			
Onmy-DAB*0101			
Sasa-DAB*0101R.....V.....			
Satr-DAB*0101R.....V.....			
Cyca-DAB*0101E.....A.D.AG..K.IIID.S.....		
Brre-DAB*0101	.N.....Q.V...A.S.Q.ITKE.	N.HIS.SD...F.....		
HLA-DRB1*0101	..TF.TLVM..TV.R...VYT.Q...P.V.S.LTVE.	R*ARS.SAQS.	MLS.VG.F.	
		Cytoplasmic tail		
Safo-DAB*0101	LGTVLALAGLIY	YKKKSSGVL-----		
Safo-DAB*0201	-----		
Safo-DAB*0301	..A.....	-----		
Safo-DAB*0401	..A.....	-----		
Safo-DAB*0501	-----	-----		
Safo-DAB*0601	-----	-----		
Onmy-DAB*0101	..I.....	-----		
Sasa-DAB*0101	..AI.....	-----		
Satr-DAB*0101	..AI.....	-----		
Cyca-DAB*0101	..IIISA.....RILVPT----		
Brre-DAB*0101	..III.I.....T.RILVPN----		
HLA-DRB1*0101	..LLFLG...FI	.FRNQK.HSGLQPTGFLS		

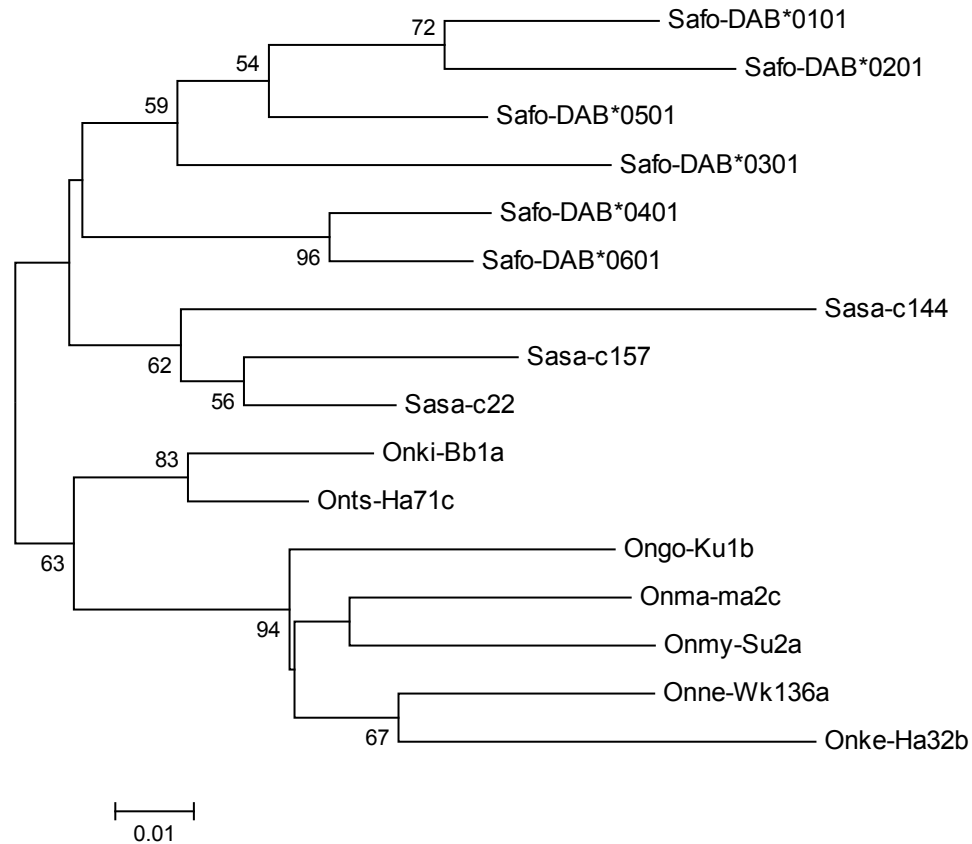


Figure 2.3. Neighbor-joining tree of the second exon of the MH class II β gene in brook charr and other salmonids. Six sequences from brook charr (*Salvelinus fontinalis*: Safo-DAB*0101 to Safo-DAB*0601; this study), three from Atlantic salmon (*Sasa-c22*, *Sasa-c144*, *Sasa-c157*; Hordvik et al. 1993) and seven from Pacific salmon (*Onts-Ha71c*, *Onki-Bb1a*, *Ongo-Ku1b*, *Onma-ma2c*, *Onmy-Su2a*, *Onke-Ha32b*, *Onne-Wk136a*; Miller and Withler 1996) are included. Tree was constructed using the method of Saitou and Nei, based on Maximum Composite Likelihood method in Mega software. Numbers on node indicate bootstrap confidence levels of 1000 bootstrap replications. Bootstrap values higher than 50% are shown.

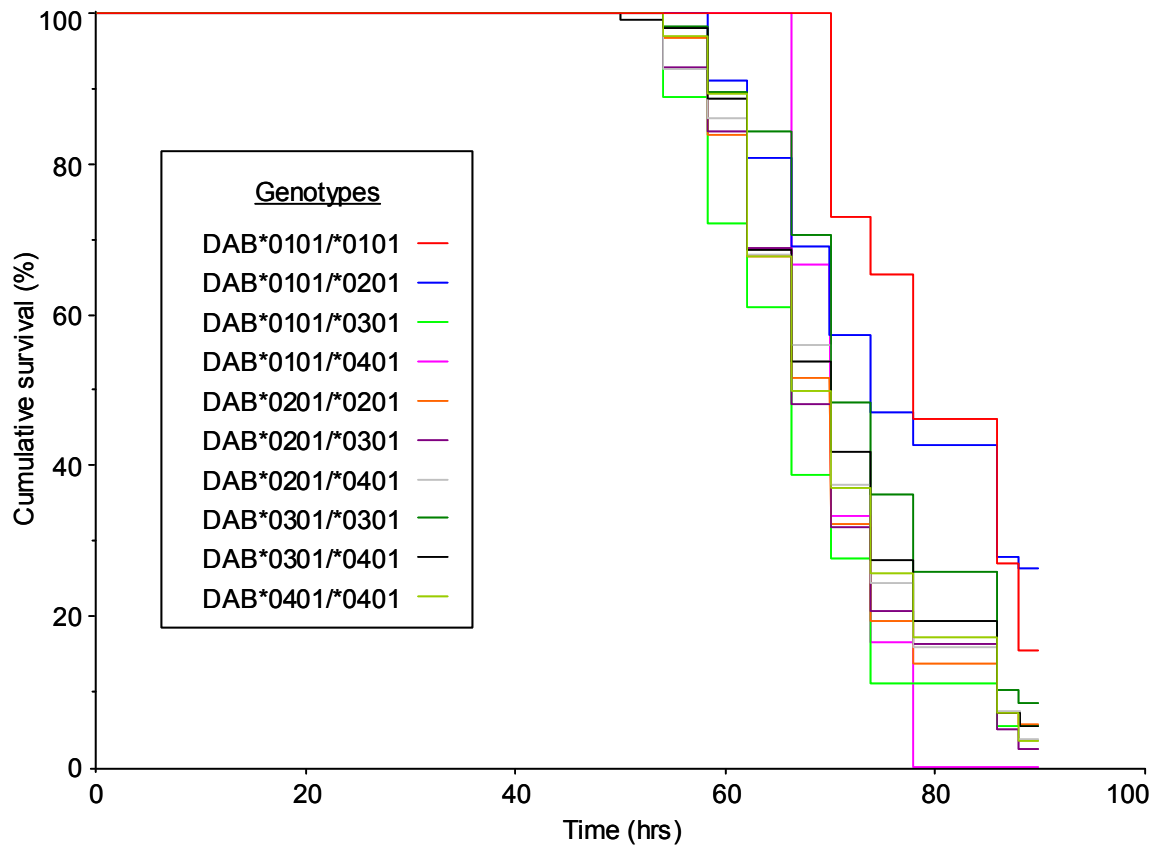


Figure 2.4. Kaplan-Meier plot showing the survival of different MH class II β genotypes in brook charr exposed to the challenge experiment with *Aeromonas salmonicida*. Cumulative survival of fish harbouring one of the ten genotypes found in this study was calculated following harvest every four hours during the infection experiment. Survival was reflected by remaining individuals at 90 hours.

Chapitre 3

Temperature and length-dependent modulation
of the MH class II β gene expression
in brook charr (*Salvelinus fontinalis*) by a *cis*-acting minisatellite.

Temperature and length-dependent modulation of the MH class II β gene expression in brook charr (*Salvelinus fontinalis*) by a *cis*-acting minisatellite.

Research article

Sébastien Croisetière^{1,a}, Louis Bernatchez² and Pierre Belhumeur¹.

¹ Département de Microbiologie et Immunologie, Université de Montréal, Montréal, Québec, H3C 3J7, Canada. ² Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, G1V 0A6, Canada.

^a Corresponding author. Phone: 1-514-343-6111 #1432, Fax: 1-514-343-5701.

CONTRIBUTIONS DU CANDIDAT

En tant que premier auteur, j'ai contribué de façon majoritaire à la rédaction de l'article, à l'analyse des résultats, ainsi qu'à l'élaboration des travaux de recherche. J'ai également réalisé la totalité des travaux de laboratoire. Louis Bernatchez et Pierre Belhumeur ont participé à la rédaction de l'article (commentaires, corrections) ainsi qu'à l'élaboration du projet de recherche.

3.1. Abstract

It is widely recognized that the variation in gene regulation is an important factor from which evolutionary changes in diverse aspects of phenotype can be observed in all organisms. Distinctive elements with functional roles on gene regulation have been identified within the non-coding part of the genome, including repeated elements. Major histocompatibility complex (MHC) genes have been the subject of an abundant literature which made them unique candidates for studies of adaptation in natural populations. Yet, the vast majority of studies on MHC genes have dealt with patterns of polymorphism in sequence variation while very few paid attention to the possible implication of differential expression in adaptive responses. In this paper, we report the identification of a polymorphic minisatellite formed of a 32 nucleotides motif (38% G+C) involved in regulation of the major histocompatibility class II β gene (MHII β) of brook charr (*Salvelinus fontinalis*). Our main objectives were: to analyze the variability of this minisatellite found in the second intron of the MHII β gene and to document its effect on the variation of expression level of this gene under different environmental conditions. Distinctive numbers of the minisatellite repeats were associated with each different MHII β alleles identified from exon 2 sequences. Relative expression levels of specific alleles in heterozygous individuals were determined from fish lymphocytes in different genotypes. We found that alleles carrying the longest minisatellite showed a significant 1.67 to 2.56-fold reduction in the transcript expression relatively to the shortest one. Results obtained in three different genotypes also indicated that the repressive activity associated to the longest minisatellite was more effective at 18 °C compared to 6 °C. In contrast, no significant difference was observed in transcript levels between alleles with comparable minisatellite length at both temperatures. We also depicted a significant up-

regulation of the total MIII β transcript at 6 °C relative to 18 °C. These results reveal for the first time that a temperature-sensitive minisatellite could potentially play an important role in the gene regulation of the adaptive immune response in fish.

Keywords: Major histocompatibility complex, gene expression regulation, repetitive sequences, selection, evolution, salmonids.

3.2. Introduction

More than three decades ago, King and Wilson (1975) proposed that the phenotypic divergence between human and chimpanzees was mostly attributable to changes in gene regulation rather than changes in the protein-coding sequences of the gene. It is now recognized that the variation in gene regulation is an important factor from which evolutionary changes in diverse aspects of phenotype can be observed in all organisms (Wray et al. 2003, Bird et al. 2007). Indeed, numerous studies demonstrated the correlation between variation in cis-regulatory sequences and significant modification in morphology, physiology or behaviour (Wray 2007). In the last few years, several groups used the approach of genome-wide association analysis to identify the genetic basis of the variation in gene expression (Brem and Kruglyak 2005, Cheung et al. 2005, Stranger et al. 2005). Being advantageous at multiple levels, these studies have the inconvenient of solely identifying the single nucleotide polymorphisms (SNPs) or marker in linkage disequilibrium with a determinant responsible for the variation of the expression level of the gene. Moreover, the correlation between the marker and the gene expression level requires a direct allele-specific expression measurement to confirm a cis-acting effect and further investigations to identify the mechanism responsible of the effect (Pastinen and Hudson 2004).

Non-coding DNA other than the one found in the transcriptional regulatory region (i.e. promoter, enhancer and suppressor), considered for a long time as “junk” DNA, recently received increasing interest (Dermitzakis et al. 2002, Mouse Genome Sequencing Consortium et al. 2002, Thomas et al. 2003, Siepel et al. 2005). Distinctive elements with functional roles in the regulation of gene expression have been identified within the non-coding part of the genome. These non-coding functional elements include introns, satellite DNAs and

transposons, as well as pseudogenes and non-coding RNA genes (see Eddy 2001, Balakirev and Ayala 2003, Beauregard et al. 2008, Usdin 2008 for reviews).

The different roles associated to introns include those in alternative splicing, transcriptional regulation, nuclear export, translation and chromatin structure. Moreover, the evolution of intron size in eukaryotes can be imputable to different models. The “selection for economy” model argues that natural selection favours short introns in highly expressed genes to minimize the cost of transcription and other molecular processes (Castillo-Davis et al. 2002, Eisenberg and Levanon 2003). On the other hand, the “genomic design” model hypothesizes that the greater length of introns found in tissue-specific genes, possibly resulting from the accumulation of multispecies conserved sequences (MCSs) (Pozzoli et al. 2007), is an indication of their functional and regulatory complexity (Vinogradov 2004, 2006). Furthermore, repeated sequences are frequently found in introns of higher eukaryotes. These repeated sequences are classified in two large families, called “tandem repeats” and “dispersed repeats” (Richard et al. 2008). The first family includes satellite DNAs, where the subfamilies of microsatellites and minisatellites are respectively tandemly repeated sequences of short (<10 bp) or longer motives (>10 and <100 bp) (Charlesworth et al. 1994). The second family contains transposons from which three subdivisions of retrotransposons can be identified; long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and long terminal repeats (LTRs) (Wicker et al. 2007). Both families of repeated sequences have been shown to be implicated in regulation of gene expression (Ugarkovic 2005, Feschotte 2008, Tomilin 2008).

Located in coding or non-coding regions of genomes, microsatellites (simple sequence repeats, SSRs) expansion or contraction can produce various effects on chromatin organization, regulation of gene activity, recombination, DNA replication, cell cycle,

mismatch repair system, etc. (Li et al. 2002, 2004). SSRs variation is also well known for being responsible of human diseases such as neurological disorders and cancers (Atkin 2001, Orr and Zoghbi 2007), and also for contributing to normal variation in brain and behavioral traits (Fondon et al. 2008). These expanded repeats have the potential to adopt altered DNA secondary structures that confer genetic instability, leading eventually to phenotypic changes (Wells et al. 2005, Wang and Vasquez 2006). On the other hand, functions associated to minisatellites (variable number of tandem repeats, VNTRs) are far less well characterized with the exception of their implication in recombination events and gene expression modulation (Bois and Jeffreys 1999, Richard and Pâques 2000). Probably the best characterized minisatellite regulating gene expression is the one found at the insulin (*INS*) gene locus (Bennett et al. 1995), for which the implication in susceptibility to human type 1 diabetes has been known for more than two decades (Bell et al. 1984). Other regulatory VNTRs were identified in additional genes as well: cystatin B (*CSTB*) (Laloti et al. 1997), serotonin transporter (*5-HTT*) (MacKenzie and Quinn 1999), dopamine transporter (*DAT1*) (Mill et al. 2002) and type 2 bradykinin receptor (*B₂*) (Zamorano et al. 2006).

At least two mechanisms have been proposed to be responsible for the differences in transcriptional activity conferred by VNTRs. The first one, found in *INS* and *5-HTT* (Stin2 VNTR) genes, implicates the presence of high-affinity binding sites for the transcription factors Pur-1 (Kennedy et al. 1995) and YB-1 (Y box binding protein 1) / CTCF (CCTC-binding factor) (Klenova et al. 2004) respectively, to modulate the VNTR function as a transcriptional regulatory domain. The other mechanism contributes to the biological effect of the VNTR by the adoption of a non-B DNA structure that could interfere with the transcriptional activity. Precisely, a four stranded G-quartet structure has been associated to *INS*-VNTR (Catasti et al. 1996) and *CSTB*-VNTR (Saha and Usdin 2001).

Classical major histocompatibility complex class II (MHCII) molecules are cell-surface glycoproteins specialized in the presentation of peptides, mainly derived from extracellular proteins, to the antigen receptor of CD4 T cells in the adaptive immune system. With the MHC class I molecules, they both represent the most polymorphic genes known to date, generally with multiple loci and high allelic diversity at each of these loci (Trowsdale and Parham 2004). However, since genes coding for these two classical proteins are not found in a complex in bony fishes (Bingulac-Popovic et al. 1997, Sato et al. 2000), the designation “major histocompatibility” (MH) genes is more appropriate than MHC in teleost (Dixon and Stet 2001). Classical major histocompatibility genes have been the subject of an abundant literature over the last 20 years or so. In particular, these genes offer several assets that made them unique candidates for studies of adaptation in natural populations (Bernatchez and Landry 2003). Yet, the vast majority of studies on MHC genes deal with patterns of polymorphism in sequence variation while very few paid attention to the possible implication of differential expression in adaptive responses. This is surprising since it has been known for many years that the modulation of the immune response is not only determined by the polymorphism of MHCII genes, but also by the cell surface expression correlation with the transcriptional level of the gene (Glimcher and Kara 1992).

In a previous study, we identified a 32 bp motif minisatellite located in the second intron of the MH class II β (MHII β) gene in brook charr, *Salvelinus fontinalis*, a salmonid species with a broad native geographic distribution in northeastern North America and which is highly valued both for sport fisheries and aquaculture (Croisetièrre et al. 2008). Here, our main objective is to investigate the possible implication of this minisatellite in the MHII β gene expression. More specifically, we (i) analyzed the variability of this minisatellite found in the

second intron of the MHII β gene in brook charr and (ii) documented its effect on the variation of the expression level of this gene under different environmental conditions. Precise expression levels were determined directly from fish lymphocytes. In addition to the identification of TNF- α and LPS as regulator of the MHII expression in rainbow trout macrophages (Knight et al. 1998), the promoter characterisation of the MHII genes in Atlantic salmon (Syed et al. 2003) and the modulation of the transcriptional activity of immune system genes by low temperature (Rodrigues et al. 1998; Nath et al. 2006), results presented here improve our understanding of the factors contributing to the modulation of the expression of the MHII genes in teleosts.

3.3. Materials and Methods

3.3.1. Fish

Individuals of approximately 50 g from a domestic strain of brook charr from a local hatchery (Chesterville, Québec, Canada) were used to evaluate the total expression level (cumulative of both alleles) of the MHIIB gene in a natural environment. Our previous work confirmed the existence of a single copy of this gene in the brook charr (Croisetière et al. 2008). A sample of 96 individuals from this population was previously genotyped at MHIIB gene which revealed eight different alleles (exon 2 coding sequence) with a distinct minisatellite size associated to each one (data not shown). Directly at the hatchery, head kidney tissue samples were taken for total RNA isolation from 23 random individuals in August when average water temperature was 18 °C and from 24 other individuals in November when average temperature was 6 °C.

A population of 93 brook charr originating from the Rupert River (Rupert strain) in northwestern Québec, Canada, and associated waterways were maintained in a controlled environment at the LARSA (Laboratoire de Recherches des Sciences Aquatiques; Université Laval). These fish of approximately 1 kg were tagged and genotyped at the MHIIB locus (*DAB*) by SSCP (Single Strand Conformational Polymorphism) as previously described (Croisetière et al. 2008). Knowing their respective genotype, fish could be specifically selected for the length of the minisatellites they were carrying at the MHIIB gene and used in real-time RT-PCR with specific primers to determine allelic expression. Before proceeding to blood sampling, fish were acclimatized 5-7 days at specified temperature (6 °C and 18 °C) to attenuate the effect of stress and assure a steady state expression of the MHIIB gene.

3.3.2. RNA isolation from peripheral blood leukocytes (PBL) and head kidneys

Head kidneys from domestic fish (Chesterville) of each group (6 °C and 18 °C) were isolated to evaluate the total expression of MHII β genes. Fish were sacrificed with an overdose of tricaine methanesulfonate (MS-222) (Sigma) before removing the head kidney that was mashed into a disposable dish using a syringe. An equivalent volume of Trizol (Invitrogen) was added before transferring the tissue sample to an Eppendorf for total RNA extraction according to the manufacturer's protocol. For each fish, a fin sample was taken for genomic DNA isolation.

RNA from PBL of brook charr (Rupert strain) was also extracted to evaluate the relative expression levels of MHII β alleles at 6 °C and 18 °C. Since it is non-lethal, RNA extraction from PBL compared to head kidneys allowed us to test gene expression of the same fish population at both experimental temperatures. Furthermore, this allowed the evaluation of the expression level of the MHII β gene for different alleles in a second strain and type of tissue. Fish were first anesthetised with 0.1 % of tricaine methanesulfonate (MS-222) (Sigma). The blood was collected from the caudal vein using a heparinised 3 ml syringe with a 21G1 needle (Becton Dickson) and added to 1 volume of ice-cold PBS containing 10 U/ml heparin sodium salt (Sigma). A fraction of this cellular suspension (1 ml) was then spun at 500 x g for 5 min at 4 °C and the buffy coat and plasma transferred to a new Eppendorf. The PBL were then pelleted and resuspended in Trizol (Invitrogen) for total RNA extraction according to the manufacturer's protocol. Dimethyl sulfoxide (DMSO; 7 % final conc) was added to the remaining blood sample in PBS and frozen for subsequent genomic DNA isolation.

3.3.3. Polymerase chain reaction (PCR), cloning and sequencing

PCR was performed on genomic DNA from domestic strain fish to evaluate the size(s) of the minisatellite present in the second intron of the MHII β alleles. The PCR fragment containing the minisatellite was obtained with primers located in conserved regions of the second intron, after analysis of thirteen alleles in brook charr (data not shown). The sequence of the sense primer P399, located 69 to 91 base pair (bp) before the first minisatellite motif, was 5'-TTCTTGAATRGT TTTATAGCTGCAAC-3' and the antisense primer P400, located 37 or 81 bp after the last motif, was 5'-CATCGCTGTASTTTATAAATTACACTTC-3'. PCR was carried out in a total volume of 50 μ l containing 20-40 ng of genomic DNA, 0.2 mM dNTPs, 7 μ l of 10X PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 2 mM MgCl₂, 5 pmol each of P399 and P400 primers, 1 unit of *Taq* DNA polymerase. The following PCR conditions were used: denaturation for 3 min at 94 °C followed by 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 45 °C and extension for 2 min at 72 °C. The final extension was for 10 min at 72 °C.

For each allele of the Rupert strain previously identified (Croiseti re et al. 2008), a MHII β gene fragment containing the complete exon coding for the β 1 domain (second exon) and the subsequent intron (second intron) was obtained by PCR on genomic DNA using primers in conserved regions for cloning and sequencing. The sequence of the sense primer SP4503, located at the end of the first intron of the gene, was 5'-CCTGTATTTATGTTCTCCTTTCAGATGG-3' and the antisense primer P232, located at the N-terminal region of the β 2 domain (exon 3), was 5'-CTGAGCTCAGTCTGACATGG-3'. The reactions were carried out in a total volume of 50 μ l containing 20-40 ng of genomic DNA, 0.2 mM dNTPs, 7 μ l of 10X Expand high fidelity buffer (Roche), 2 mM MgCl₂, 5 pmol

each of SP4503 and P232 primers, 2.6 units of Expand high fidelity enzyme mix (Roche) in a thermocycler (Biometra). The following PCR conditions were used: denaturation for 3 min at 94 °C followed by 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 52 °C and extension for 3 min at 72 °C. The final extension was for 10 min at 72 °C.

PCR products, from 1350 to 2700 bp depending on MHII β Rupert alleles, were subsequently purified and extracted from agarose gel using the Illustra GFX PCR DNA and gel band purification kit (GE Healthcare) and subcloned in both orientations in pT7Blue-2 vector (Novagen) in accordance with the recommended protocol from the manufacturer.

Sequencing of both strands was performed on a 3730 DNA analyser (Applied Biosystems) using plasmid primers P235-Blue2S·Tag (5'-ACCGCTGCTGCTAAATTCGAA-3') and P247 (5'-CGGAATTCTATAGTGTCACC-3') as well as intron 2 internal sense primer P399 (5'-TTCTTGAATRGT TTTATAGCTGCAAC-3') and antisense primer P400 (5'-CATCGCTGTASTTTATAAATTACACTTC-3'). Since the length of one particular allele (*Safo-DAB*0101*) was too long to reach the nucleotide sequences from both sides (see Results), an Exonuclease III (New England Biolabs) digestion of the vector containing the allele was realized to allow the sequencing through the minisatellite.

3.3.4. Reverse transcriptase (RT)

Quantification of extracted RNA from fish of both the domestic and Rupert strain was done on a ND-1000 Spectrophotometer (NanoDrop) after treating with the DNA-*free* kit (Ambion) and before cDNA synthesis using SuperScript II reverse transcriptase (Invitrogen). Briefly, 500 ng of total RNA was mixed with oligo(dT)₁₂₋₁₈ primer (Invitrogen), dNTPs (GE Healthcare) in a volume of 12 μ l and denatured at 65 °C for 5 min. This was followed by addition of 5X FS buffer, DTT, RNaseOUT (Invitrogen) and 200 units of SuperScript RT in a

20 μ l final reaction volume for incubation at 42 °C for 50 min, as recommended by the manufacturer.

3.3.5. Multiplex PCR

A multiplex (semi-quantitative) PCR was realized on domestic strain to evaluate the total expression level of MHII β genes against the housekeeping gene GAPDH. The generic primer sequences used in this experiment are listed in Table 3.1 (multiplex). PCR was carried out in a total volume of 50 μ l containing 1 μ l of a cDNA dilutions (1/10, 1/25, 1/50, 1/100 and 1/200), 0.2 mM dNTPs, 7 μ l of 10X PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 2 mM MgCl₂, 5 pmol each of P383, P410, P399 and P400 primers, 1 unit of *Taq* DNA polymerase. The following PCR conditions were used: denaturation for 3 min at 94 °C followed by 40 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 50 °C and extension for 1 min at 72 °C. The final extension was for 10 min at 72 °C. PCR products length were 294 bp for the P384/P410 (MHII β) primer pair and 595 bp for P339/P340 (GAPDH). Quantification of the PCR product by spot densitometry was done on a ChemiImager 5500 Imaging System (Alpha Innotech Corporation). Intensities of the PCR products were linear between dilutions 1/25 and 1/200. The ratio MHII β /GAPDH for each sample was calculated with the linear regression equation at the median dilution point.

3.3.6. Quantitative PCR

Quantitative PCR was performed using the FastStart SYBR Green Master (Roche Diagnostics) on a Rotor-Gene 3000 (Corbet Research). The primers used in these experiments are detailed in Table 3.1 (qPCR). All new primers were designed to flank a region containing an intron, to assure that amplification of contaminating DNA was controlled, which has not been detected in melting curve analysis. Amplification mixture (25 μ l) contained template

cDNA (2 μ l of 1/20 dilution), 2X FastStart SYBR Green Master (12.5 μ l) and 300 nM forward and reverse primers. The cycling conditions comprised 15 min polymerase activation and denaturation step at 95 °C, followed by 40 cycles at 95 °C for 15 sec, 60 °C for 45 sec and 72 °C for 30 sec. All samples were run in triplicate with non-template controls (NTC). Relative quantification of gene expression levels was done by comparing difference in C_T values by the $\Delta\Delta C_T$ method (Critical factors for successful real-time PCR, Qiagen), since the PCR efficiency was nearly equivalent for pairs of primers. For the domestic strain, EFIA_B, β 2m, GAPDH and TCR α were evaluated as reference genes to normalize the results before calculating the relative expression levels. The stability of the reference genes was evaluated with the Bestkeeper tool (Pfaffl et al. 2004). We kept the elongation factor I alpha, paralog B, (EFIA_B) as reference gene, as it showed the lowest variability in our samples and also because it has been found suitable for gene expression assay in different studies (Olsvik et al. 2005, Jorgensen et al. 2006, Ingerslev et al. 2006). Within the population of 93 brook charr (Rupert strain) kept at the LARSA, we identified four genotypes from which the allelic expression levels could be analyzed at temperatures of 18 °C and 6 °C: *Safo-DAB*0201/*0301*, *Safo-DAB*0401/*0301*, *Safo-DAB*0101/*0301* and *Safo-DAB*0101/*0201*. The sizes of the minisatellites present in the MHII β gene from these individuals were respectively: 27 and 20, 40 and 20, 69 and 20, 69 and 27 repetitions of the minisatellite motif (Table 3.2). The relative allelic expression levels were calculated by comparing differences in C_T values in the same sample with allele specific primers. Samples were included in different qPCR runs to control the reproductivity of the results. Differences for MHII β expression levels as a function of minisatellite length or temperature were compared using the *t*-test with bilateral distribution for samples with homogeneous variance.

3.3.7. Sequence analysis

Nucleotide sequences from the MHII β second intron alleles of the Rupert strain were submitted to GenBank (accession nos. GQ253476, GQ253477, GQ253478, GQ253479, GQ253480 and GQ253481). The coding sequences of the six MHII β alleles can also be found (GenBank accession nos. EU478851 to EU478856). Alignment of the nucleotide sequences was performed using ClustalX (Thompson et al. 1994) and MEGA 4 (Tamura et al. 2007). Identity of the intronic sequences was determined by pairwise-comparison of nucleotide sequences with the Syn-SCAN program (Gonzales et al. 2002). GeneBank published sequences from Atlantic salmon, *Salmo salar* (*Sasa-DAB*0201*; AJ439067 and *Sasa-DAB*0301*; AJ439069) and *Oncorhynchus* species (*Onmy-Su2a*; U34715, *Onne-Wk136e*; U34713, *Onne-Wk136a*; U34712 and *Onke-An13c*; U34702) (Miller and Withler 1996) were included in the alignment. All sequence alignments are available in the Supplementary Material online.

3.4. Results

3.4.1. Nucleotide sequences and minisatellite from the second intron of MHII β gene.

The first objective of this study was to document the variability of a minisatellite found in the second intron of the MHII β gene in brook charr. As a first step, we sequenced the second intron of the six brook charr MHII β alleles of the Rupert strain identified in a previous study (Croisetière et al. 2008). Characteristics of the second intron of these alleles and of other salmonids MHII β gene are given in Table 3.2. The intron length differed among all six exon alleles identified in brook charr of the Rupert strain and ranged from 1023 to 2410 nucleotides. The alignments of the second intron nucleotide sequences flanking a minisatellite in 5' (supplementary Figure S3.1, Supplementary Material online) and 3' (supplementary Figure S3.2, Supplementary Material online) are presented for selected salmonid species to illustrate the diversity found at this locus. A schematic representation of this diversity in the second intron of the MHII β gene of salmonids is also shown in Figure 3.1.

Four regions of the intron indicated as Box I to Box IV showed variable level of sequence divergence in pairwise comparisons (Figure 3.1 and supplementary Figure S3.1, S3.2). The first 261 nucleotides of the second intron (Box I) showed pairwise divergence values ranging from 1.20 % to 7.02 % (mean of 5.52 %) among brook charr alleles to a maximum of 12.9 % (mean of 8.46 %) when compared to other salmonid species. Box II corresponds to the first minisatellite motif from brook charr and Atlantic salmon, and the sole motif reported in Pacific salmon species. This 32 nucleotide sequence diverged by 0 to 4 nucleotides (mean of 1.33 nt.; 4.17 %) within brook charr while divergence for up to 8 nucleotides were found (mean of 3.05 nt.; 9.52 %) between species. Box III comprises 36 nucleotides with divergence ranging from 0 to 5 nucleotides (mean of 2.27; 9.45 %) among

brook charr alleles and up to 9 substitutions (mean of 4.22 nt.; 13.4 %) among species. Box IV is composed of the last 58 nucleotides of the intron with divergence values ranging from 0.00 % to 5.00 % (mean of 2.33 %) within brook charr alleles and up to 17.9 % (mean of 7.07 %) among species.

The minisatellite identified within the second intron has a consensus motif of 32 nucleotides (type I in Table 3.2): ACATTACAGTATGACTAGTCAGCTATGTAGTG; which is 38 % G+C rich. Search for homologue sequences did not reveal its occurrence in salmonid species other than Atlantic salmon. PCR amplification of a fragment containing this minisatellite with the primer pairs P399/P400 revealed that each allele contains a different number of the minisatellite repeats with 69, 27, 20, 40, 19 and 25 repetitions found in the *Safo-DAB*0101* to *Safo-DAB*0601* allele, respectively (Table 3.2). An example of the minisatellite found in allele *Safo-DAB*0301* is shown (Figure 3.2). With the exception of the *Safo-DAB*0101* allele, sequencing also revealed that the purity of the repeated motif was very high, ranging from 95.6 % to 99.7 %. In contrast, allele *Safo-DAB*0101* showed a high level of mutation over the whole minisatellite length; with 86 % (59 out of 69) of the motifs being mutated in comparison with the consensus sequence. Moreover, these mutated motifs included ten additional ones that showed a deletion of about 50 % of the nucleotides relative to the consensus motif (Figure 3.1).

3.4.2. MHIIB gene expression level is inversely correlated to minisatellite length.

A multiplex PCR experiment was realized on head kidney RNA of the domestic strain to evaluate the possible effect of length of minisatellite on gene expression level. This strain was selected because the fish were exposed to natural fish farm environment and also because of availability of specimens. However, since the genotypes of the individuals at the fish farm

are variable in function of new incorporated stocks, the coding sequence of the MHII β gene was not considered and the focus was oriented on the minisatellite length found in the second intron. To estimate the size of the minisatellite present in both MHII β genes, a PCR with the P399/P400 conserved primer pair was realized on genomic DNA from 23 individuals for which the head kidney was also sampled at 18 °C directly at the fish farm (in August). Depending on MHII β allele, PCR products of approximately 750, 925, 975, 1600 and 2000 bp, corresponding to approximately 19, 25, 27, 46 and 59 motif repetitions respectively, were generated (Figure 3.3). We identified 12 homozygotes and 11 heterozygotes individuals in the 23 fish sampled at 18 °C, with the former ones clearly showing increased intensity of the PCR product (e.g. #4, #8). Fish carrying both short minisatellites, heterozygote (#3 and #15; 19 and 25 repeats) or homozygote (#14; 25 repeats) were selected for the multiplex PCR experiment. Also, since only two individuals (#1 and #7; 46 repeats) were found to possess long minisatellites, a second cDNA was prepared from fish number 1. A representative result from triplicate multiplex RT-PCR is shown in Figure 3.3 for each category of minisatellite (short and long). Quantification of the total MHII β gene expression, normalized for GAPDH, revealed that fish having long minisatellites showed a significantly (*t*-test, $p < 0.05$) lower relative expression level of $2.14 \pm \text{S.E. } 0.06$ in comparison to $3.78 \pm \text{S.E. } 0.52$ for short minisatellites carriers, which represents a 1.77-fold reduction in the total MHII β gene expression in fish possessing long minisatellites (Figure 3.3).

3.4.3. MHII β gene expression level is up-regulated at low temperature.

It has been previously demonstrated that the expression of MHII genes was influenced by temperature in rainbow trout (*Oncorhynchus mykiss*) (Nath et al. 2006). To examine if temperature could also modulate the expression level in brook charr, a second sampling of 24

individuals was realized at 6 °C (in November) as previously described and the minisatellite length of MHII β gene was estimated by PCR amplification with the P399/P400 primer pair, thus revealing the presence of 11 homozygotes and 13 heterozygotes (Figure 3.4). Fish were selected at 6 °C for short minisatellites, heterozygote (#10 and #22; 19 and 25 repeats) or homozygote (#8 and #15; 25 repeats), or long minisatellites (#12; 46 and 59 repeats). Since the difference in the expression level observed previously at 18 °C was small (1.77-fold reduction), a quantitative real-time RT-PCR experiment was designed to evaluate the expression levels of MHII β at both temperatures and to confirm the previous results. In this experiment, a fourth sample (#4; 25 repeat) was added to the group of short minisatellite while only two fish samples of the long minisatellite group (#1 and #7) were kept for the samples at 18 °C. Figure 3.4 shows the relative total MHII β gene expression values (generic P456/P457 primer pair) in head kidney for short or long minisatellites present in the gene, at temperatures of 18 °C and 6 °C with EFIA_B as reference genes. The results are the cumulative of two RT-PCR experiments where each sample is tested each time in triplicate. At 18 °C, the MHII β genes having a short minisatellite (19 and 26 repeats) showed a relative expression of 1.12 ± 0.12 while it was 0.75 ± 0.17 for the genes with a long minisatellite (46 repeats). This real-time PCR results confirmed with a second housekeeping gene the multiplex PCR data obtained at 18 °C by demonstrating a significant 1.49-fold down-regulation (*t*-test, $p < 0.01$, $n=8$ and 4) of the MHII β expression for gene carrying a long minisatellite. At 6 °C, the relative expression was 2.31 ± 0.27 for the genes with a short minisatellite and 1.56 ± 0.05 for the genes with a long one, representing a significant 1.48-fold down-regulation (*t*-test, $p < 0.001$, $n=8$ and $n=2$) for the MHII β gene carrying a long minisatellite. These results further indicated that the lower amount of expression associated to the long minisatellite is observed at both

temperatures. We finally compared the total expression level of MHIIB gene from fish having a short minisatellite at both temperatures and observed a significant 2.08-fold up-regulation of the MHIIB transcript at 6 °C relative to 18 °C (*t*-test, $p < 0.001$, $n=8$).

3.4.4. The minisatellite from MHIIB gene is a functional cis-acting polymorphism.

Since the variation in MHIIB expression level seems to be modulated by the size of the minisatellite present in the gene, we developed a real-time RT-PCR assay to quantify the relative expression level of both alleles in peripheral blood leukocytes (PBL) from the same heterozygous fish. Values indicate the relative allelic expression levels of MHIIB for the target allele in function of the second one used as reference. The results from eight individuals of *Safo-DAB*0201/*0301* genotype (27 and 20 repeats) indicated an average MHIIB relative allelic expression of 0.97 ± 0.06 at 18 °C, and 1.00 ± 0.02 for seven individuals at 6 °C (primers P537-P538/P457) (Figure 3.5). This observation is not primer specific given that it was observed with a second pair of primers: P539-P540/P232 (data not shown). In fish of this genotype, the analysis of the relative allelic expression revealed no significant difference in the transcript levels between these alleles at both temperatures.

We also analyzed individuals with the *Safo-DAB*0401/*0301* genotype (40 and 20 repeats), for which the difference in minisatellite length is more important. In those fish, we detected an allelic imbalance in the expression profile between both alleles. The relative expression level of the *Safo-DAB*0401* allele in function of *Safo-DAB*0301* allele, as studied in five individuals, was 0.51 ± 0.06 and 0.59 ± 0.06 at 18 °C and 6 °C, respectively (primers P452-P478 /P457) (Figure 3.5). This represents a 1.96-fold and 1.69-fold reduction of the expression for the allele carrying the longer minisatellite. Moreover, these values indicate a significant difference in the expression levels between 18 °C and 6 °C (*t*-test: $p < 0.05$, $n=9$).

The last two genotypes for which we analyzed the relative expression levels were *Safo-DAB*0101/*0301* (69 and 20 repeats) and *Safo-DAB*0101/*0201* (69 and 27 repeats), with the *Safo-DAB*0101* allele carrying the longest minisatellite found in this population present in both ones. In the former genotype, found in three individuals (2 analyzed at 6 °C because of one mortality between sampling), the relative allelic expression was 0.45 ± 0.05 at 18 °C and 0.60 ± 0.07 at 6 °C (primers P451-P452/P457). This corresponded to a significant 2.22-fold and 1.67-fold reduction in the expression of the *Safo-DAB*0101* allele compared to *Safo-DAB*0301* at both temperatures (*t*-test: $p < 0.001$, $n=9$ (18 °C), $n=6$ (6 °C)) (Figure 3.5). This result was confirmed with the pair of primer P539-P540/P232 (data not shown). Found in 11 individuals, the second genotype in which *Safo-DAB*0101* was present also showed a reduction in the transcript expression of the allele carrying the longest minisatellite and a significant difference in the relative expression levels between temperatures (*t*-test: $p < 0.05$, $n=10$). A relative allelic expression of 0.39 ± 0.06 (2.56-fold reduction) and 0.51 ± 0.03 (1.96-fold reduction) at 18 °C and 6 °C respectively was found for the *Safo-DAB*0101/*0201* genotype (Figure 3.5). The results presented for the last three genotypes (*Safo-DAB*0401/*0301*, *Safo-DAB*0101/*0301* and *Safo-DAB*0101/*0201*) further indicates that the repressive activity associated to the longest minisatellite was more effective at 18 °C compared to 6 °C and therefore highlighted a functional cis-acting polymorphism for the minisatellite present in the second intron of the MIII β gene.

3.5. Discussion

3.5.1. Instability of the minisatellite found in the second intron of MHIIB gene.

In this study, a minisatellite composed of a 32 nucleotides motif was identified in the second intron of the MHIIB gene in brook charr. This minisatellite has already been revealed in the same gene in Atlantic salmon (Stet et al. 2002) and as a single 32 nucleotides motif in Pacific salmon (*Oncorhynchus*) species (Miller and Withler 1996). Search for homologue sequences of the minisatellite did not reveal its occurrence in other species. The length of the intron found in brook charr is comparable to that found in Atlantic salmon, but longer than for Pacific salmon. The presence of the motif as a single or as a minisatellite sequence in all salmonids species, combined with the low sequence divergence of Box I and Box IV support the hypothesis of a single and same ancestral expressed MHIIB locus in salmonids (Glamann 1995, Langefors et al. 2000).

Atlantic salmon minisatellites (type I) are showing an increase accumulation of mutations with purity of 92.0% and 95.4% over the entire length. For microsatellites, it has been demonstrated that expansion and contraction mutations on the tandem repeat tend to remove the imperfections found in the motifs by the mechanism of polymerase slippage, while in the absence of this purifying activity, the motifs accumulate punctual mutations (Ellegren 2000). Thus, if some extrapolation can be done for minisatellite, this would suggest that the high minisatellite purity found in the brook charr MHIIB gene, with the exception of the *Safo-DAB*0101* and Atlantic salmon alleles, are indicative of a purifying activity, presumably a consequence of a selection force preserving further possibility of expansion or contraction of the minisatellite.

Analysis of the motif sequences found in *Safo-DAB*0101* allele revealed the perfect repetition of a DNA fragment within the minisatellite composed of 22 motifs representing 704 nucleotides identified by asterisks in Figure 3.1. Since the minisatellite found in this allele is composed of motifs showing a complex and high level of polymorphism, this repetition seems to reflect a recombination event that occurred previously and support this mechanism to explain the expansion of this minisatellite sequence as previously suggested (Debrauwere et al. 1997).

3.5.2. Minisatellite modulation of the MHIIB gene expression level.

We also described a functional effect associated to the presence of the polymorphic minisatellite on the transcriptional activity of the MHIIB gene. Regulation of the transcriptional activity by minisatellites has already been demonstrated in few studies (e.g. Pugliese et al. 1997, MacKenzie and Quinn 1999, Zamorano et al. 2006). This genetic element is added to a different minisatellite (type II 40 nucleotides motif, Figure 3.1 and Table 3.2 and supplementary Figure S3.2), a partial and a complete *Hpa* I SINE that were identified previously at the same location in salmonids (Figure 3.1 and Table 3.2 and supplementary Figure S3.1, S3.2) (Miller and Withler 1996, Stet et al. 2002). The possibility of MHIIB gene expression modulation offered by these types of genetic elements could constitute a fine tuning mechanism to ensure an appropriate immune response to pathogens, while limiting deleterious effects to host tissues.

In fact, two strains of brook charr demonstrated a reduction in MHIIB gene expression for different alleles possessing the longest minisatellite. The effect was also seen at the total RNA expression level in head kidney tissue (Figure 3.3 and Figure 3.4) or by comparison of the allele's relative expression level in lymphocytes of heterozygous fish (Figure 3.5). This

analysis of the transcriptional allelic variation allows the detection of small differences observed between alleles and also eliminates the inter-individual variation that could come from different genetic background as well as environmental or physiological conditions rather than direct genetic factors. These results were obtained using different techniques including multiplex RT-PCR (Figure 3.3) and real-time RT-PCR (Figure 3.4 and Figure 3.5), the later data subsequently confirmed with different probes for the same allele (Table 3.1). With the exception of the few last 3' nucleotides that were allele specific, the primers for the quantification between these alleles were localized at the same position and produced same size amplicons. The selection of housekeeping gene is also known to be a critical factor in gene expression experiments. We then used two different genes as reference, GAPDH (Figure 3.3) and EFIA_B (Figure 3.4), where the second one revealed the most stable transcription level in a selection of genes in Atlantic salmon (Olsvik et al. 2005). These two reference genes showed a stable level of expression at both temperatures in the tissues tested for the quantification experiments (Croisetière et al. in preparation). Moreover, we observed this effect in multiple heterozygous individuals presenting three different genotypes where the MHII β alleles carrying the longest minisatellite (*Safo-DAB*0101* and *Safo-DAB*0401*) showed reduce transcript expression when compared to the allele with the shortest one (Figure 3.5). It is noteworthy that the difference between the expression level of the allele *Safo-DAB*0101* relatively to *Safo-DAB*0301* was not greater than *Safo-DAB*0401* relatively to *Safo-DAB*0301*, even if the number of motif found in *Safo-DAB*0101* is superior (69 versus 40). One possible explanation for this could be the presence of multiple mutations in this allele that decrease its impact.

Finally, our results also confirmed that reduced expression level associated to alleles carrying a long minisatellite is modulated by the temperature. The temperature effect on teleost MH gene expression has been addressed for more than a decade now (Rodriguez et al. 1998). Here, results clearly showed a more pronounced difference of MHII β gene expression level at 18 °C between alleles of different minisatellite length (Figure 3.5). This could possibly reflect a greater thermodynamical energy for the adoption of a DNA secondary structure as functional mechanism.

3.5.3. Reduced immune gene expression at high temperature in brook charr?

When we compared at both temperatures the total expression level of the MHII β genes carrying short minisatellites, we found that the expression was 2.08-fold higher at 6 °C than at 18 °C (Figure 3.4). The difference was approximately of the same order for long minisatellites, however with fewer samples tested. This increased expression at low temperature is in accordance with the data of Raida and Buchmann (2007) for constitutive expression of MHII β gene in rainbow trouts' head kidney. In their paper, Table 4 indicates a difference of 0.8 raw C_T between MHCII expression at 5 °C and 15 °C, while the expression of the reference gene EF1- α was stable. Since MHCII and EF1- α show almost the same PCR efficiency (101.1 % vs 100.0 %), this represents a 1.74-fold higher expression of MHII β at 5 °C, a result in accordance with ours. On the other hand, Nath et al. (2006) observed a downregulation of MHII β gene expression at mRNA and protein levels in rainbow trout at 2 °C. However, this conflicting result seems to be the consequence of the exposition to temperature below the immunologically nonpermissive temperature of \sim 4 °C (Fryer et al. 1976, Bly and Clem 1991) and near the lower lethal temperature of -0.7 °C for salmonids (Saunders 1995). Furthermore, since the gene coding for the ribosomal protein S11 also shows a decrease in transcript at 2 °C,

the effect reported by Nath et al. (2006) seems to be the result of a more general and severe repression of the transcription.

From the evaluation of the four candidate housekeeping genes that was realized, the $\beta 2m$ gene was surprisingly ranked as the worst gene in the samples tested. A close examination of the C_T obtained with the $\beta 2m$ gene revealed an interesting fact. The mean of the C_T at 18 °C for the $\beta 2m$ expression was $18.18 \pm \text{S.E. } 0.44$ while it was of $16.89 \pm \text{S.E. } 0.40$ at 6 °C. This difference was significant (t -test: $p < 0.001$, $n=6$) and represented a 2.78-fold higher expression of $\beta 2m$ at 6 °C once corrected with the reference gene $EF1A_B$ (data not shown). This result concords with studies performed in Atlantic salmon and rainbow trout that show no decrease of the $\beta 2m$ transcript after 10 days at temperature of 2 °C (Kales et al. 2006), unlike to a reduction of the expression at 6 °C reported in common carp (Rodrigues et al. 1998). The reasons why Kales et al. (2006) did not detect a significant variation in the $\beta 2m$ gene expression at low temperature could be explained by the low sensitivity of their technique or the differences in the temperatures to which the fish have been exposed. The conflictual data described in this report regarding the expression level of the $MHII\beta$ and the $\beta 2m$ as a function of temperature clearly highlights the need of further investigations on immune gene expression level at boundary or extreme habitat temperatures.

One can question what could be the advantages for fish to reduce their $MHII\beta$ expression level at higher temperature. It is well known that polymorphism within MHCII molecules influences the protein folding and binding of antigenic peptides, but also the intracellular trafficking and signaling functions. It has also been demonstrated in human and mice that polymorphism of the promoter or upstream regulatory regions in diverse isotypes or alleles affect the expression level of the gene (Beaty et al. 1999). Moreover, this difference

observed at the mRNA level subsequently result in dissimilar protein densities at cell surface. Within the four subsets of activated CD4 T cells, the Th1 and Th2 cells differ from each other in their pattern of cytokines production and implication in immune responses against intracellular pathogens or extracellular parasites, respectively (Zhu and Paul 2008). The magnitude and orientation of the CD4 T cell response are directly influenced by the concentration and strength of the antigenic peptides presented by the antigen presenting cells (APC). APC exposed to high concentration or to a potent antigen favored a stronger T cell proliferative response and a Th1 cytokine expression pattern (Valitutti et al. 1996, Crowe et al. 1998). INF- γ produced by Th1 T cells is an important activator of macrophage to increase their microbicidal activity (Suzuki et al. 1988).

Evidences suggested that the density of MHCII molecules at cell surface of APC influences T cell activity in the same way that concentration and potency of the antigen do. It has been observed that macrophages treated with IL-10 have a reduced capacity to present antigens due to a downregulation of MHCII expression, reflected in a lower proliferative response of Th1 and Th2 T cells (de Waal Malefyt et al. 1991). Transgenic mice with a reduced expression level of MHCII molecules in B cells and dendritic cells have been used to study the influence of the MHCII density at the cell surface on CD4 T cell differentiation (DiMolfetto et al. 1998). While the T cell priming and proliferation was unaffected in these mice, T cell responses were dominated by the Th1-associated cytokine INF- γ , with reduced levels of the Th2 cytokine IL-4 compared to control. Another study with transgenic mice showed that a line expressing a low level of MHCII molecules allowed the selection of the CD4 T cell in the thymus, whereas they were eliminated in a line with a high expression level

(Fukui et al. 1997). These results suggested an implication of the MHCII expression level on the formation of the T cell repertoire.

A lower expression level observed for the MHII β alleles carrying a long minisatellite in brook charr could possibly be implicated in the orientation of the immune response toward the Th1 subset of CD4 T cells and the formation of the T cell repertoire. Fish immune system has not been studied as human and mouse systems regarding the CD4 T cell subsets. Since comparative studies between these species didn't detect major differences between their immune systems, a such specialisation for T cells during their maturation could be possible for teleost fish, and surely constitute an interesting avenue of research. It is interesting to note that *Aeromonas salmonicida*, an important salmonid pathogen, can survive and replicate within brook charr macrophages (Daly et al. 1996; Garduño et al. 2000). The presence and amplification of the minisatellite found in the second intron of the MHII β gene could constitute a molecular mechanism allowing the fish to adapt the expression level in an environment rich in intracellular pathogens.

By reducing the MHII β gene expression level at higher temperature (and possibly other immune relevant genes), the advantage could also be a diminution of the energy cost to mount an efficient immune response. This hypothesis is based on the fact that temperature is important in governing the rate at which body processes are carried out. This might be the consequences of an increased cellular metabolism and/or functional activity. It has been suggested that membrane fluidity could be involved in the activation of T cells in function of temperature (Vallejo et al. 1992). As previously mentioned, other studies also demonstrated that some fish MH genes have a temperature dependent expression level (Rodrigues et al. 1998, Kales et al. 2006, Nath et al. 2006). Another possible advantage on the field of energy

economy implicates a fine tuning by the minisatellite of the MHII β gene expression level. The closely linkage of the minisatellite with the exon 2 coding for the protein β 1 domain and responsible of antigen presentation suggests a co-evolution of both elements. Since it is known that specific alleles of MHII β gene correlates with increase resistance to pathogen (Langefors et al. 2001, Grimholt et al. 2003, Croisetièrè et al. 2008), the selection for an allele with a particular minisatellite length from another one coding for an identical protein, but with a different minisatellite length, is clearly possible. The basis for this selection would be an optimal expression level of the allele, involving a certain energy cost, in function of the efficiency to present antigens and induce an appropriate immune response. Here, it would be very interesting to develop transgenic fish with the same MHII β coding sequence, but differing at minisatellite length, and to test them in challenge experiment against bacteria to evaluate the contribution of gene expression level to the resistance and susceptibility to pathogens.

3.5.4. Other elements contributing to the polymorphism of the MHII β gene in salmonids.

Two additional elements could be distinguished from the analysis of the second intron nucleotide sequences of the MHII β gene in salmonids. First, a second minisatellite composed of a 40 nucleotides motif (type II) was found exclusively in Atlantic salmon (Figure 3.1 and Table 3.2 and supplementary Figure S3.2, small white and grey rectangles). This minisatellite directly follows the first one (type I) and the 39 nucleotides making the transition are 90 % conserved (35 of 39 nucleotides) in respect to the consensus sequence motif of type I. This second minisatellite has so far been found in only one individual since the other Atlantic salmon allele sequence available, *Sasa-DAB*0301*, is incomplete. The accumulation of genetic element in the second intron of the salmonids MHII β gene with its expression

restricted to specialized tissues (Ting and Trowsdale 2002) support the hypothesis of ‘genomic design’ model for the evolution of this intron (Pozzoli et al. 2007).

Box III nucleotide sequence identified in supplementary Figure S3.2 is composed of 36 of the 40 nucleotides forming the motif sequence for the second minisatellite found in Atlantic salmon. This minisatellite appears to evolve from amplification of a nucleotide sequence already identified in brook charr and Pacific salmon species. Then, it seems that a particular nucleotide sequence present as a single copy, probably meeting specific criteria, could evolve in a minisatellite element in response to some circumstances. This type of element occurred twice in the second intron of *MHII β* gene of salmonids. The first amplification of the 32 nucleotide motif is found in brook charr and Atlantic salmon, in opposition to the single motif found in Pacific salmon species. This amplification could have occurred before the separation of the two former salmonid species. On the other hand, the amplification of the 40 nucleotides motif found only in Atlantic salmon may have appeared following the split of Atlantic salmon from other salmonids. This amplification mechanism has not been observed thus far in Pacific salmon species.

The other element that was in the second intron of the *MHII β* gene in salmonids is a *Hpa* I SINE (Kido et al. 1991) (Figure 3.1 and Table 3.2 and supplementary Figure S3.1, S3.2). As described by Miller and Withler (1996), two copies (one partial and one complete) of the element are inserted in opposite orientations on both sides of the single minisatellite motif (type I) found in Pacific salmon species. Analysis of the sequences that constitute Box III and Box IV, in Pacific salmon species that possess the complete *Hpa* I SINE (Onne-Wk136a and Onne-An13c), revealed that the insertion of this element occurred in a region where the sequences are still joined in brook charr and Atlantic salmon. This indicates that this

event took place after the separation of the Pacific salmon species from Atlantic salmon's and brook charr's ancestor.

SINEs contain an internal RNA polymerase III promoter that consists of an A box downstream from the transcription start site, and a B box located approximately 50 bp downstream from the A box (Roy et al. 2000). Moreover, it contains several potential transcription factor binding sites. Wang et al. (2004) reported that SINE retroposition does not appear randomly in the direction of insertion, with all SINEs within known host gene in rainbow trout inserted in the opposite orientation to the host gene transcription. As they suggested, insertion in both orientations could interfere with gene expression, the insertion in the same orientation showing the greatest possibility of effects. The fact that the complete *Hpa* I SINE found in brook charr MHII β gene is in the same orientation and shows almost intact A box and B box thus suggests that this element might present a regulatory role modulating the expression of the gene during stress period or pathogen infection. Furthermore, the closely insertion of two *Hpa* I SINEs in the same gene is unexpected since SINEs insertion is believed to be random (Kido et al. 1995). The dual insertion events detected in brook charr MHII β gene is puzzling and calls for further studies on the mechanisms of SINEs retroposition and regulatory functions.

The presence of multiple elements for which the implication in recombination events has been demonstrated previously probably highlights the importance of this type of mechanism in generating polymorphism of the MHII β gene. The high concentration and variability of elements found in this intron, just downstream of the exon 2 coding for the domain implicated in peptide binding and presentation could reflect the selective pressure to assure polymorphism of the MHII β gene. A higher incidence of these elements in the MHII β

gene of salmonids may be the consequence of the fact that it is the only MHII locus identified thus far (Glamann 1995, Langefors et al. 2000). Allele and gene conversion were identified as a mechanism to generate polymorphism of MHC in multiple organisms (Pease et al. 1983, Zangenberg et al. 1995). Also, it was demonstrated that the Eb gene of the mouse contains a recombinational hotspot located within the second intron. Interestingly, a segment of the intron showed strong similarity to retroposon long terminal repeat (LTR), env, and pol genes indicating that this segment of the second intron may have evolved through retroposon insertion (Zimmerer and Passmore 1991).

3.5.5. Identification of a new regulatory element submitted to natural selection?

MHII β gene exhibit polymorphism that is undoubtedly relevant for resistance or susceptibility toward infectious diseases (Trowsdale and Parham 2004, Dionne et al. 2009), where different mechanisms are responsible for the evolution and maintenance of this polymorphism (Bernatchez and Landry 2003, Aguilar and Garza 2007). Since teleosts express a single “minimal, essential” locus of both α and β chain forming the MHII molecule (Glamann 1995, Langefors et al. 2000) and also MHII α and MHII β loci are genetically linked in Atlantic salmon (Stet et al. 2002), this animal model was used in association studies to identify resistance and susceptibility alleles, with the focus on the coding sequences (Langefors et al. 2001, Grimholt et al. 2003, Croisetière et al. 2008). Other studies demonstrated the implication of non-coding region polymorphism in the modulation of the MHCII β transcriptional activity. The expression level of MHCII molecules was shown to be affected by mutations in the promoter, some of them revealing binding sites for transcription factors (Shewey et al. 1992, Baumgart et al. 1998). The results presented in this paper clearly demonstrate the implication of a new mechanism to fine tune the expression of the MHII β

gene. Some reports are emerging to implicate microsatellite in morphological and behavioral evolution in vertebrates (Fondon and Garner 2004, Hammock and Young 2005) indicating a more important contribution of VNTR regulating phenotypic expression than previously presumed. Given the amply documented implication of the MHCII β gene in the immune response and fitness, this report is the first, to our knowledge, to present a minisatellite that could potentially be submitted to selective forces and therefore play an important role in the immune response of fish populations in the wild.

3.6. Supplementary Material

Supplementary Figures S3.1 and S3.2 are available at Molecular Biology and Evolution online (<http://www.mbe.oxfordjournals.org/>). Sequence data from this article have been submitted to GenBank under accession nos. GQ253476 to GQ253481.

3.7. Acknowledgements

This work was supported by Valorisation Recherche Québec (VRQ) through the auspices of the Réseau Aquaculture Québec (RAQ). As members of the former AquaNet Network, we would like to thank AquaNet and the Networks of Centres of Excellence Programs and Councils for providing funds. S.C. wishes to thank the Faculty of Graduate Studies of the Université de Montréal for partial financial support. Finally, we would also like to thank Robert Gélinas for his collaboration and also Martin Clément and Julie Shareck for helpful discussions.

3.8. Literature Cited

- Aguilar A, Garza JC. 2007. Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. *J. Mol. Evol.* 65:34-43.
- Atkin NB. 2001. Microsatellite instability. *Cytogenet. Cell. Genet.* 92:177-181.
- Balakirev ES, Ayala FJ. 2003. Pseudogenes: are they "junk" or functional DNA? *Annu. Rev. Genet.* 37:123-151.
- Baumgart M, Moos V, Schuhbauer D, Müller B. 1998. Differential expression of major histocompatibility complex class II genes on murine macrophages associated with T cell cytokine profile and protective/suppressive effects. *Proc. Natl. Acad. Sci. U.S.A.* 95:6936-6940.
- Beaty JS, Sukiennicki TL, Nepom GT. 1999. Allelic variation in transcription modulates MHC class II expression and function. *Microbes Infect.* 1:919-927.
- Beauregard A, Curcio MJ, Belfort M. 2008. The take and give between retrotransposable elements and their hosts. *Annu. Rev. Genet.* 42:587-617.
- Bell GI, Horita S, Karam JH. 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33:176-183.
- Bennett ST, Lucassen AM, Gough SC, et al. (17 co-authors). 1995. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.* 9:284-292.
- Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* 16:363-377.

- Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J. 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics*. 46:129-134.
- Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, Miller W, Hurles ME, Dermitzakis ET. 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol*. 8:R118.
- Bly JE, Clem LW. 1991. Temperature-mediated processes in teleost immunity: in vitro immunosuppression induced by in vivo low temperature in channel catfish. *Vet. Immunol. Immunopathol*. 28:365-377.
- Bois P, Jeffreys AJ. 1999. Minisatellite instability and germline mutation. *Cell. Mol. Life Sci*. 55:1636-1648.
- Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102:1572-1577.
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. 2002. Selection for short introns in highly expressed genes. *Nat. Genet*. 31:415-418.
- Catasti P, Chen X, Moyzis RK, Bradbury EM, Gupta G. 1996. Structure-function correlations of the insulin-linked polymorphic region. *J. Mol. Biol*. 264:534-545.
- Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215-220.
- Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. 2005. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437:1365-1369.

- Croisetière S, Tarte PD, Bernatchez L, Belhumeur P. 2008. Identification of MHC class IIbeta resistance/susceptibility alleles to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*). *Mol. Immunol.* 45:3107-3116.
- Crowe PD, Boehme SA, Wong T, Gaur A, Sidney J, Sette A, Conlon PJ. 1998. Differential signaling and hierarchical response thresholds induced by an immunodominant peptide of myelin basic protein and an altered peptide ligand in human T cells. *Hum. Immunol.* 59:679-689.
- Daly JG, Kew AK, Moore AR, Olivier G. 1996. The cell surface of *Aeromonas salmonicida* determines in vitro survival in cultured brook trout (*Salvelinus fontinalis*) peritoneal macrophages. *Microb. Pathog.* 21:447-461.
- Debrauwere H, Gendrel CG, Lechat S, Dutreix M. 1997. Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. *Biochimie* 79:577-586.
- Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson BJ, Flegel V, Bucher P, Jongeneel CV, Antonarakis SE. 2002. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420:578-582.
- de Waal Malefyt R, Haanen J, Spits H, Roncarolo MG, te Velde A, Figdor C, Johnson K, Kastelein R, Yssel H, de Vries JE. 1991. Interleukin 10 (IL-10) and viral IL-10 strongly reduce antigen-specific human T cell proliferation by diminishing the antigen-presenting capacity of monocytes via downregulation of class II major histocompatibility complex expression. *J. Exp. Med.* 174:915-924.

- DiMolfetto L, Neal HA, Wu A, Reilly C, Lo D. 1998. The density of the class II MHC T cell receptor ligand influences IFN-gamma/IL-4 ratios in immune responses in vivo. *Cell. Immunol.* 183:70-79.
- Dionne M, Miller KM, Dodson JJ, Bernatchez L. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society of London* 364:1555-1565.
- Dixon B, Stet RJ. 2001. The relationship between major histocompatibility receptors and innate immunity in teleost fish. *Dev. Comp. Immunol.* 25:683-699.
- Eddy SR. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2:919-929.
- Eisenberg E, Levanon EY. 2003. Human housekeeping genes are compact. *Trends Genet.* 19:362-365.
- Ellegren H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16:551-558.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9:397-405.
- Fondon JW 3rd, Garner HR. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101:18058-18063.
- Fondon JW 3rd, Hammock EA, Hannan AJ, King DG. 2008. Simple sequence repeats: genetic modulators of brain function and behavior. *Trends Neurosci.* 31:328-334.
- Fryer JL, Pilcher KS, Sanders JE, Rohovec JS, Zinn JL, Groberg WJ, McCoy RH. 1976. Temperature, infectious disease, and the immune response in salmonid fish. *NTIS* 191:72.

- Fukui Y, Ishimoto T, Utsuyama M, Gytoku T, Koga T, Nakao K, Hirokawa K, Katsuki M, Sasazuki T. 1997. Positive and negative CD4+ thymocyte selection by a single MHC class II/peptide ligand affected by its expression level in the thymus. *Immunity* 6:401-410.
- Garduño RA, Moore AR, Olivier G, Lizama AL, Garduño E, Kay WW. 2000. Host cell invasion and intracellular residence by *Aeromonas salmonicida*: role of the S-layer. *Can. J. Microbiol.* 46:660-668.
- Glamann J. 1995. Complete coding sequence of rainbow trout Mhc II beta chain. *Scand. J. Immunol.* 41:365-372.
- Glimcher LH, Kara CJ. 1992. Sequences and factors: a guide to MHC class-II transcription. *Annu. Rev. Immunol.* 10:13-49.
- Gonzales MJ, Dugan JM, Shafer RW. 2002. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* 18:886-887.
- Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A, Saebo S, Stet RJ. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55:210-219.
- Hammock EA, Young LJ. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* 308:1630-1634.
- Ingerslev HC, Pettersen EF, Jakobsen RA, Petersen CB, Wergeland HI. 2006. Expression profiling and validation of reference gene candidates in immune relevant tissues and cells from Atlantic salmon (*Salmo salar* L.). *Mol. Immunol.* 43:1194-1201.

- Jorgensen SM, Kleveland EJ, Grimholt U, Gjoen T. 2006. Validation of reference genes for real-time polymerase chain reaction studies in Atlantic salmon. *Mar. Biotechnol. (NY)*. 8:398-408.
- Kales S, Parks-Dely J, Schulte P, Dixon B. 2006. Beta-2-microglobulin gene expression is maintained in rainbow trout and Atlantic salmon kept at low temperatures. *Fish Shellfish Immunol.* 21:176-186.
- Kennedy GC, German MS, Rutter WJ. 1995. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet.* 9:293-298.
- Kido Y, Aono M, Yamaki T, Matsumoto K, Murata S, Saneyoshi M, Okada N. 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. *Proc. Natl. Acad. Sci. U.S.A.* 88:2326-2330.
- Kido Y, Saitoh M, Murata S, Okada N. 1995. Evolution of the active sequences of the HpaI short interspersed elements. *J. Mol. Evol.* 41:986-995.
- King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107-116.
- Klenova E, Scott AC, Roberts J, Shamsuddin S, Lovejoy EA, Bergmann S, Bubb VJ, Royer HD, Quinn JP. 2004. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J. Neurosci.* 24:5966-5973.
- Knight J, Stet RJM, Secombes CJ. 1998. Modulation of MHC class II expression in rainbow trout *Oncorhynchus mykiss* macrophages by TNF α and LPS. *Fish Shellfish Immunol.* 8:545-553.

- Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE. 1997. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386:847– 851.
- Langefors A, Lohm J, Von Schantz T, Grahn M. 2000. Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. *Mol. Ecol.* 9:215-219.
- Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. Biol. Sci.* 268:479-485.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11:2453-2465.
- Li YC, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21:991-1007.
- MacKenzie A, Quinn J. 1999. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proc. Natl. Acad. Sci. U.S.A.* 96:15251-15255.
- Mill J, Asherson P, Browes C, D'Souza U, Craig I. 2002. Expression of the dopamine transporter gene is regulated by the 3' UTR VNTR: Evidence from brain and lymphocytes using quantitative RT-PCR. *Am. J. Med. Genet.* 114:975-979.
- Miller KM, Withler RE. 1996. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43:337-351.

- Moore LJ, Somamoto T, Lie KK, Dijkstra JM, Hordvik I. 2005. Characterisation of salmon and trout CD8alpha and CD8beta. *Mol. Immunol.* 42:1225-1234.
- Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, et al. (223 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562.
- Nath S, Kales S, Fujiki K, Dixon B. 2006. Major histocompatibility class II genes in rainbow trout (*Oncorhynchus mykiss*) exhibit temperature dependent downregulation. *Immunogenetics* 58:443-453.
- Olsvik PA, Lie KK, Jordal AE, Nilsen TO, Hordvik I. 2005. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. *BMC Mol. Biol.* 6:21.
- Orr HT, Zoghbi HY. 2007. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30:575-621.
- Pastinen T, Hudson TJ. 2004. Cis-acting regulatory variation in the human genome. *Science* 306:647-650.
- Pease LR, Schulze DH, Pfaffenbach GM, Nathenson SG. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. U.S.A.* 80:242-246.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* 26:509-515.
- Pozzoli U, Menozzi G, Comi GP, Cagliani R, Bresolin N, Sironi M. 2007. Intron size in mammals: complexity comes to terms with economy. *Trends Genet.* 23:20-24.

Pugliese A, Zeller M, Fernandez A Jr, Zalberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD. 1997. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDD3 susceptibility locus for type 1 diabetes. *Nat. Genet.* 15:293-297.

Raida MK, Buchmann K. 2007. Temperature-dependent expression of immune-relevant genes in rainbow trout following *Yersinia ruckeri* vaccination. *Dis. Aquat. Organ.* 77:41-52.

Richard GF, Pâques F. 2000. Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* 1:122-126.

Richard GF, Kerrest A, Dujon B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72:686-727.

Rodrigues PN, Dixon B, Roelofs J, Rombout JH, Egberts E, Pohajdak B, Stet RJ. 1998. Expression and temperature-dependent regulation of the beta2-microglobulin (Cyca-B2m) gene in a cold-blooded vertebrate, the common carp (*Cyprinus carpio* L.). *Dev. Immunol.* 5:263-275.

Roy AM, West NC, Rao A, Adhikari P, Alemán C, Barnes AP, Deininger PL. 2000. Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* 302:17-25.

Saha T, Usdin K. 2001. Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases. *FEBS Lett.* 491:184-187.

Sato A, Figueroa F, Murray BW, Malaga-Trillo E, Zaleska-Rutczynska Z, Sultmann H, Toyosawa S, Wedekind C, Steck N, Klein J. 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51:108-116.

- Saunders R. 1995. Salmon aquaculture: present status and prospectus for the future. In: A Boghen (ed) Cold-water aquaculture in Atlantic Canada. The Canadian Institute for Research On Regional Development, Moncton, New Brunswick, Canada. p. 35–82.
- Shewey LM, Beaty JS, Andersen LC, Nepom GT. 1992. Differential expression of related HLA class II DQ beta genes caused by nucleotide variation in transcriptional regulatory elements. *J. Immunol.* 148:1265-1273.
- Siepel A, Bejerano G, Pedersen JS, et al. (16 co-authors). 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034-1050.
- Stet RJ, de Vries B, Mudde K, Hermsen T, van Heerwaarden J, Shum BP, Grimholt U. 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* 54:320-331.
- Stranger BE, Forrest MS, Clark AG, et al. (12 co-authors). 2005. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 1:e78.
- Suzuki Y, Orellana MA, Schreiber RD, Remington JS. 1988. Interferon-gamma: the major mediator of resistance against *Toxoplasma gondii*. *Science* 240:516-518.
- Syed M, Vestrheim O, Mikkelsen B, Lundin M. 2003. Isolation of the promoters of Atlantic salmon MHCII genes. *Mar. Biotechnol.* 5:253-60.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.
- Thomas JW, Touchman JW, Blakesley RW, et al. (71 co-authors). 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424:788-793.

- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Ting JP, Trowsdale J. 2002. Genetic control of MHC class II expression. *Cell* 109 Suppl:S21-33.
- Trowsdale J, Parham P. 2004. Mini-review: defense strategies and immunity-related genes. *Eur. J. Immunol.* 34:7-17.
- Tomilin NV. 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* 30:338-348.
- Ugarkovic D. 2005. Functional elements residing within satellite DNAs. *EMBO Rep.* 6:1035-1039.
- Usdin K. 2008. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res.* 18:1011-1019.
- Valitutti S, Müller S, Dessing M, Lanzavecchia A. 1996. Different responses are elicited in cytotoxic T lymphocytes by different levels of T cell receptor occupancy. *J. Exp. Med.* 183:1917-1921.
- Vallejo AN, Miller NW, Clem LW. 1992. Cellular pathway(s) of antigen processing in fish APC: effect of varying in vitro temperatures on antigen catabolism. *Dev. Comp. Immunol.* 16:367-381.
- Vinogradov AE. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* 20:248-253.

- Vinogradov AE. 2006. "Genome design" model: evidence from conserved intronic sequence in human-mouse comparison. *Genome Res.* 16:347-354.
- Wang G, Vasquez KM. 2006. Non-B DNA structure-induced genetic instability. *Mutat. Res.* 598:103-119.
- Wang T, Johnson N, Zou J, Bols N, Secombes CJ. 2004. Sequencing and expression of the second allele of the interleukin-1beta1 gene in rainbow trout (*Oncorhynchus mykiss*): identification of a novel SINE in the third intron. *Fish Shellfish Immunol.* 16:335-358.
- Wells RD, Dere R, Hebert ML, Napierala M, Son LS. 2005. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* 33:3785-3798.
- Wicker T, Sabot F, Hua-Van A, et al. (13 co-authors). 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973-982.
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20:1377-1419.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8:206-216.
- Zamorano R, Suchindran S, Gainer JV. 2006. 3'-Untranslated region of the type 2 bradykinin receptor is a potent regulator of gene expression. *Am. J. Physiol. Renal Physiol.* 290:F456-64.
- Zangenberg G, Huang MM, Arnheim N, Erlich H. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nat. Genet.* 10:407-414.
- Zhu J, Paul WE. 2008. CD4 T cells: fates, functions, and faults. *Blood* 112:1557-1569.

Zimmerer EJ, Passmore HC. 1991. Structural and genetic properties of the Eb recombinational hotspot in the mouse. *Immunogenetics* 33:132-140.

3.9. Tables and Figures

Table 3.1 Primer sequences for multiplex and quantitative RT-PCR in brook charr

Gene	Allele	Name	Sequence (5'→3')
MHII β	N/A	P384-Fwd	CTCCTCAAAGGACCTGCATGG
(multiplex)	N/A	P410-Rev	AGCATGGCAGGGTGTCTGC
GAPDH	N/A	P339-Fwd	AACGGATTTGGCCGTATTGGCC ^b
(multiplex)	N/A	P340-Rev	GAGGCAGGGATGATGTTCTGG ^b
MHII β ^a	*0101	P451-FwdGGAAGTATGTTGGATACACTGAGTAT
(qPCR)	*0201 and *0301	P452-FwdGTATGTTGGATACACTGAGCTG
	*0101 to *0401 (generic)	P456-Fwd	GTGGGGAAAGTATGTTGGATACACTGAG
	*0401	P478-FwdGGAAGTATGTTGGATACACTGAGCAT
	*0101, *0201 and *0401	P537-Fwd	. . .CTAGGGGAGCTGGAGCGTGT
	*0301	P538-Fwd	GAGCTAGGGGAGCTGGAGAGTTA
	*0101 and *0201	P539-Fwd	. . .GCATGGTGCAGAGTTTATAGACTCTTACTAC
	*0301 and *0401	P540-Fwd	CCTGCATGGTATAGAGTTTATAGACTCTTATGTT
	*0101 to *0401	P232-Rev	CTGAGCTCAGTCTGACATGG
	*0101 to *0401 (generic)	P457-Rev	ATGGCAGGGTGTCTGCCACTA
β 2m	N/A	P454-Fwd	AGAGTGTGGATTTCACACCAGCCA ^c
(qPCR)	N/A	P455-Rev	TCCTTACATATCTGCCCTCCAGGT ^c
EFIA _B	N/A	P460-Fwd	TGCCCCCTCCAGGATGTCTAC ^d
(qPCR)	N/A	P461-Rev	CACGGCCACAGGTACTG ^d
GAPDH	N/A	P462-Fwd	AGATTGTGGCCATCAACGACCC ^e
(qPCR)	N/A	P463-Rev	AAACGTCCGTGGGTGGAGTCATA ^e
TCR α	N/A	P464-Fwd	GACAGCTACTACAGCCAGGTT ^f
(qPCR)	N/A	P465-Rev	CAGAAATGGTCAGGGATAGGAAGTT ^f

Abbreviations: MHII β , major histocompatibility class II gene beta; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; β 2m, beta-2 microglobulin; EFIA_B, elongation factor I alpha, paralog B; TCR α , T cell receptor alpha; Fwd, forward; Rev, reverse; N/A, not applicable. Dots (.) are introduced to aligned primers.

^aPrimer sequences are from Croisetière et al. 2008.

^b*Oncorhynchus mykiss* gene sequence. GeneBank reference: AF027130.

^cfrom conserved region of *Oncorhynchus mykiss* gene sequences.

^dReference: Olsvik et al. 2005.

^e*Salmo salar* gene sequence. GeneBank accession no. AM230811.

^fReference: Moore et al. 2005.

Table 3.2 Characteristics of the second intron in MHII β alleles from brook charr (*Safo-DAB**) and other salmonid fishes

Allele	Intron length (nt.)	Minisatellite			<i>Hpa</i> I SINE	
		Motif ^b	No. repeats	Purity (%)	Partial	Complete
<i>Safo-DAB*0101</i>	2410	I	69 ^c	88.1	no	no
<i>Safo-DAB*0201</i>	1382	I	27	99.7	no	no
<i>Safo-DAB*0301</i>	1095	I	20	99.4	no	no
<i>Safo-DAB*0401</i>	1696	I	40	99.1	no	no
<i>Safo-DAB*0501</i>	1023	I	19	95.6	no	no
<i>Safo-DAB*0601</i>	1255	I	25	99.6	no	no
<i>Sasa-DAB*0201</i>	1334	I	25 ^d	95.4	no	no
		II	6 ^d	98.7		
<i>Sasa-DAB*0301</i> ^a	478	I	≥7	92.0	no	N/A
<i>Onmy-Su2a</i>	563	I	1	88	yes	no
<i>Onne-Wk136e</i>	547	I	1	88	yes	no
<i>Onne-Wk136a</i>	731	I	1	88	yes	yes
<i>Onke-An13c</i>	773	I	1	78	yes	yes

^a incomplete intronic sequence.

^b consensus sequence, type I (32 nt.):
consensus sequence, type II (40 nt.):

ACATTACAGTATGACTAGTCAGCTATGTAGTG.

TATGATAAACATCCAATAATATATGGTGAACATGGAATA.

^c 10 partial motifs.

^d 1 partial motif.

N/A, not available.

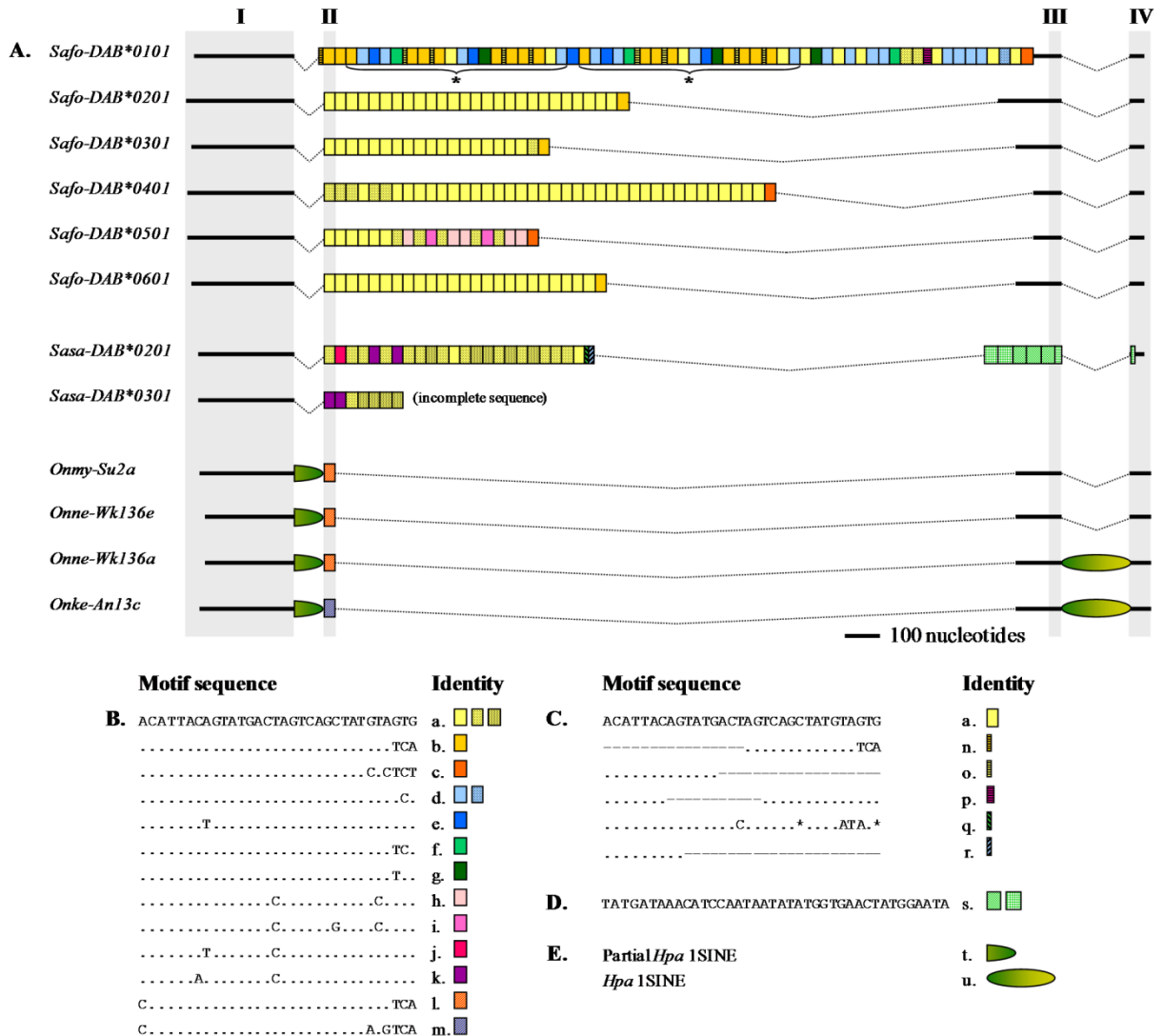


Figure 3.1. Properties of the second intron of the MHII β gene in salmonids. (A) Schematic representation of the allelic diversity found in the MHII β gene of various salmonid species. Identity of the motifs and elements correspond to the letters and color shapes indicated in (B), (C), (D) and (E). The sequence of 22 motifs repeated twice in *Safo-DAB*0101* is marked by an asterisks (*) (B) Complete motif sequences with corresponding identity indicated below by a letter and color boxes. Motifs with one or two mutations relative to the original are represented by boxes with few or many shading respectively. Dots (.) indicate identity with the consensus motif. (C) Partial motif sequences with the element identity indicated below. Asterisks (*) indicate gaps introduced to maximize the alignment; dashes (-), sequence not related to the motif. (D) Motif sequence of the second minisatellite (type II) present in the second intron of MHII β gene from Atlantic salmon. (E) Identity of the partial and complete *Hpa* I SINES.

*Safo-DAB*0301* intron 2

EXON 2 ◀ GGTGAGCAGGGCTCCGTAGCAACACTTACAGGCCCTC
 TGACTGATCAAGTCAAGACCATGACTTTATTTTATTATGGATAAAA
 TGTATGTAGAGATCTGCTGATGAAGATAATAGTAGAGATTCAG
 GGATTAATAAAGATATCTCGACTAAAGGTAACCTAATTAATAA
 CAGGATTGTTATCTAACTTGGATGGATGAAGTGTTCCTTGAA
 TGGTTTTATAGCTGCAACTATATTTAAACACTGAGTTAAATCTGA
 AAACAAACAGAATAATAATAGACA
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 1
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 2
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 3
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 4
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 5
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 6
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 7
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 8
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 9
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 10
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 11
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 12
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 13
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 14
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 15
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 16
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 17
 ACATTACAGTATGACTAGTCAGCTATGTAGTG >Repeat 18
 ACATTACAGTAGGACTAGTCAGCTATGTAGTG >Repeat 19
 ACATTACAGTATGACTAGTCAGCTATGTATCA >Repeat 20
 ACAGTTCAGTCCATAATCATCTCCTGTATGTACATCCAGCTCATC
 TCCTAATGAAGTGAATTTATAAAGTACAGCGATGGTTC AATGA
 TATCAGAACTATGGACTATATGTTATTCATCCATCATACATAGT
 GAACTACGGTGACAAGTGTTCCTCCATTA ▶ **EXON 3**

Figure 3.2. Example of a minisatellite sequence within in the second intron of brook charr *MHIIB* gene, allele *Safo-DAB*0301*. The intron of this allele contains 20 perfect repeats of the 32 nucleotide motif. The purity and numbers of motifs found in brook charr and salmonids alleles are presented in Table 3.2.

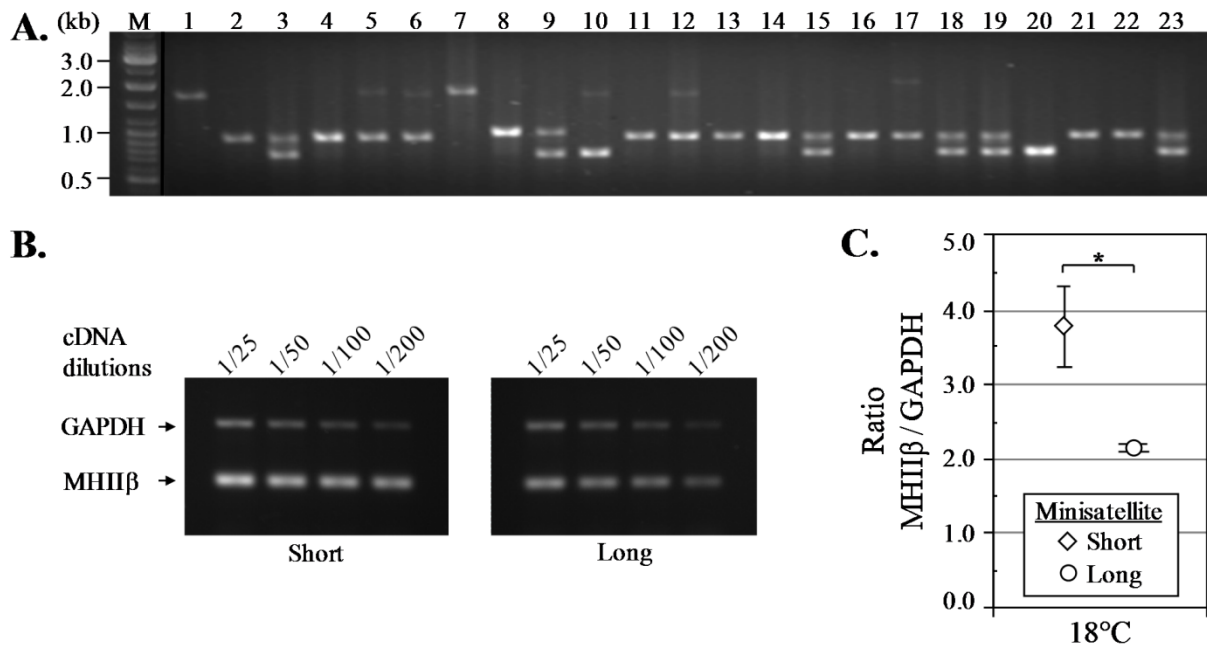


Figure 3.3. MHIIB expression level in brook charr head kidney as a function of minisatellite number of repeats. (A) Genomic PCR showing the length of the minisatellite present in the second intron of MHIIB alleles (domestic strain) amplified with the P399/P400 primers. 23 fish were sampled at 18 °C. Short and long minisatellites have 19 or 25 and 46 repeats of the 32 nucleotide motif respectively. **(B)** Agarose gel showing intensity of amplified bands for GAPDH (upper band) and total MHIIB (lower band) genes by multiplex RT-PCR. Representative results from triplicates are shown for each category (short and long minisatellite). **(C)** Quantified ratio MHIIB / GAPDH (average +/- S.D.) of the gene expression at 18 °C in fish with short (diamond) or long (circle) minisatellite. Comparisons of the distributions using the *t*-test returned significant results: * = $p < 0.05$.

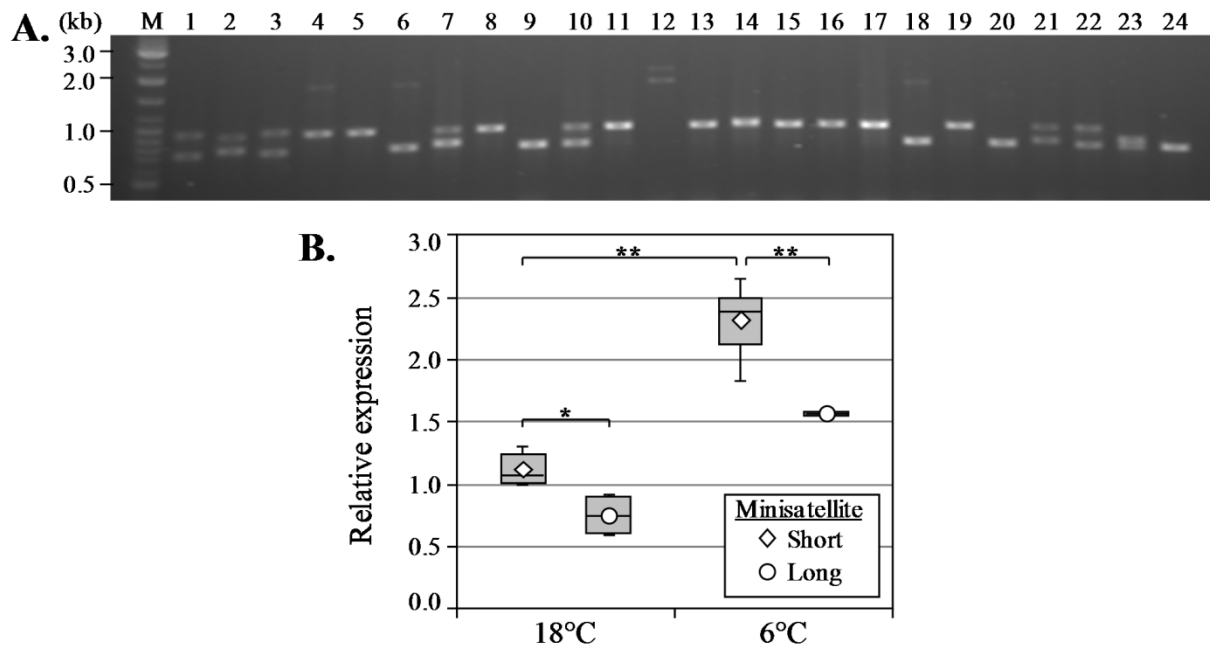


Figure 3.4. MHIIB expression level in brook charr head kidney as a function of temperature. (A) Genomic PCR showing the length of the minisatellite present in the second intron of MHIIB alleles (domestic strain) amplified with the P399/P400 primers. A total of 24 fish were sampled at 6 °C. Short and long minisatellites have 19 or 25 and 59 repeats of the 32 nucleotide motif respectively. (B) Box plots giving the relative expression distributions, evaluated by real-time RT-PCR, of total MHIIB mRNA from fish with variable minisatellite length. Symbol legend: diamond, short minisatellite; circle, long minisatellite. Results from samples at 18 °C and 6 °C are normalized for EFIA_B. The top and bottom of each rectangle represent the 75th and 25th percentiles, respectively; the bar and symbol in the rectangle indicate the median and average expression value respectively; the top and bottom whiskers show the maximal and minimal allelic expression value. Results are the cumulative data of two independent real-time RT-PCR experiments where samples are analyzed in triplicate. Comparisons of the distributions using the *t*-test returned significant results: * = $p < 0.01$ and ** = $p < 0.001$.

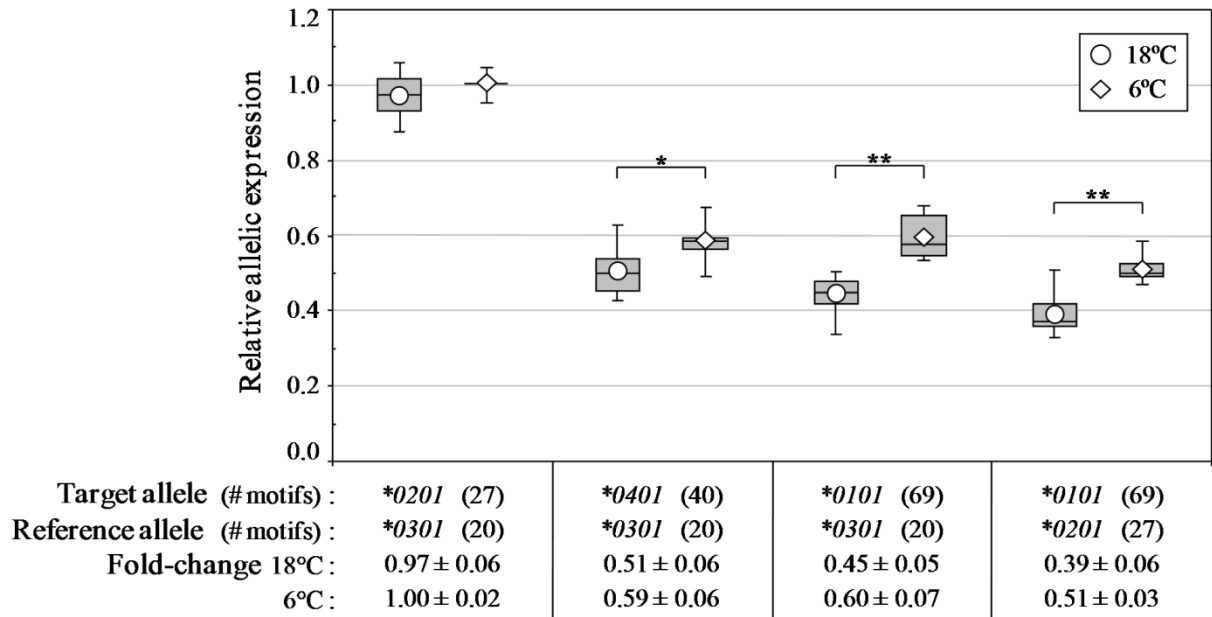


Figure 3.5. Expression level of MHIIB alleles as a function of both minisatellite length and temperature. Box plots showing the relative allelic expression distributions of MHIIB mRNA in brook charrs' lymphocytes (Rupert strain) at 18 °C and 6 °C. The allelic expression levels are expressed in function of the other one in parenthesis. The top and bottom of each rectangle represent the 75th and 25th percentiles, respectively; the bar and symbol in the rectangle indicate the median and average allelic expression value respectively; the top and bottom whiskers show the maximal and minimal allelic expression value. Symbol legend: circle, value at 18 °C; diamond, value at 6 °C. Values presented are from a single (*Safo-DAB*0201/*0301* and *Safo-DAB*0101/*0201*), two (*Safo-DAB*0401/*0301*) or three (*Safo-DAB*0101/*0301*) independent real time RT-PCR experiment, where all samples are analyzed in triplicate. Comparisons of the distributions using the *t*-test returned significant results: * = $p < 0.05$ and ** = $p < 0.001$.

Supplementary Figure S3.1. Alignment of the second intron nucleotide sequences preceding the minisatellite found in the MHI β gene from brook charr and other salmonids. The first minisatellite motif sequence of each allele is in bold. Dots (.) indicate identity with *Safo-DAB*0101*; dashes (-), gaps introduced to maximize the alignment. Portions of the intronic sequences showing high conservation are identified with boxes (I - II). The partial *Hpa* I SINE sequence is underlined. GeneBank accession numbers for Atlantic salmon (*Salmo salar*; Sasa) and Pacific salmon (*Oncorhynchus* species; Onmy, Onne and Onke) sequences are AJ439067 , AJ439069 and U34715, U34713, U34712, U34702 respectively.

	1		31		61	
<i>Safo-DAB*0101</i>	GGTGAGCAGG	GCTCCGTAGC	AACACTTACA	GGCCTCTGAC	TGAGCAAGTC	AAGACCATGA
<i>Safo-DAB*0201</i>T.....
<i>Safo-DAB*0301</i>T.....
<i>Safo-DAB*0401</i>TG.....
<i>Safo-DAB*0501</i>TG.....CA.....
<i>Safo-DAB*0601</i>CA.....
<i>Sasa-DAB*0201</i>T..C..C.....	GT.....	..G.....	..AT..T.....	..C..CA..A.....
<i>Sasa-DAB*0301</i>T..C..C.....	GT.....	..G.....	..AT..T.....	..C..CA..A.....
<i>Onmy-Su2a</i>T..A..C.....	G.....CA.....
<i>Onne-Wk136e</i>T..A..C.....	G..C.....	..T.....CA.....
<i>Onne-Wk136a</i>T..A..C.....	G..C..C.....	..T.....CA.....
<i>Onke-An13c</i>T..A..C.....	G..C.....	..G.....CA.....
	91		121		151	
<i>Safo-DAB*0101</i>	GAGATCTGTC	TGATGAAGAT	AATGGTAGAG	ATGCAGGGAT	TAATAAAGAT	AT---CTAG
<i>Safo-DAB*0201</i>G.....	..T.....C.....
<i>Safo-DAB*0301</i>A.....	..T.....C.....
<i>Safo-DAB*0401</i>G.....	..G.....	..T.....G.....
<i>Safo-DAB*0501</i>A..G.....	..T.....C.....
<i>Safo-DAB*0601</i>A.....	..T.....C.....
<i>Sasa-DAB*0201</i>G.....	..T..A.....	..T..TT.....	..G.....
<i>Sasa-DAB*0301</i>G.....	..T..A.....	..T..TT.....	..C.....
<i>Onmy-Su2a</i>	G.....	..G.....	..TG.....	..ATAT..C.....
<i>Onne-Wk136e</i>	G.....	..G.....	..T.....	..ATAT..C.....
<i>Onne-Wk136a</i>	G.....	..G.....	..T.....	..ATAT..C.....
<i>Onke-An13c</i>	G.....	..G.....	..T.....	..ATAT..C.....
	181		211		241	
<i>Safo-DAB*0101</i>	TTTTFATCTA	AATTGGATGA	ATAAC-----	ATGTTTTTTC	TTGAATAGTT	TTATAGCTGC
<i>Safo-DAB*0201</i>	..G.....	..C.....TCG.....
<i>Safo-DAB*0301</i>	..G.....	..C.....	..G..A.....TATT.....
<i>Safo-DAB*0401</i>	..G.....	..C.....	..AG..G..A.....TATT.....
<i>Safo-DAB*0501</i>	..G.....	..C..A.....	..G.....TATT.....
<i>Safo-DAB*0601</i>	..G.....	..C.....	..G..A.....TATT.....
<i>Sasa-DAB*0201</i>G..ATAACCC.....	..TATT.....
<i>Sasa-DAB*0301</i>G..ATAACCC.....	..TATT.....
<i>Onmy-Su2a</i>	..G.....	..C.....	..G..ATAAC	..A..G.....	..A.....	..GTATT..G.....
<i>Onne-Wk136e</i>	..G.....	..C.....	..G..ATAAC	..A..G.....GTATT..G.....
<i>Onne-Wk136a</i>	..G.....	..C.....	..G..ATAAC	..A..G.....GTATT..G.....
<i>Onke-An13c</i>	..G.....	..C.....	..G..ATAAC	..A..G.....GTATT..G.....
	271		301		331	
<i>Safo-DAB*0101</i>	AAA TCTGAAACAA	AGTCAGCTAT	GTAT---CA	-----	-----	-----
<i>Safo-DAB*0201</i>	CTGAGTT...	..AC..ACAGAA...	TA..AGA..	-----	-----	-----
<i>Safo-DAB*0301</i>AC..ACAGAA...	TA..AGA..	-----	-----	-----
<i>Safo-DAB*0401</i>	CTGAGTT...	..AC..GAA...	TA..AGA..	-----	-----	-----
<i>Safo-DAB*0501</i>	CTGAGTT...	C.....AC..ACAGAA...	TA..AGA..	-----	-----	-----
<i>Safo-DAB*0601</i>AC..ACAGAA...	TA..AGA..	-----	-----	-----
<i>Sasa-DAB*0201</i>T.....AC..CA..A...	TA..CAGATA-	-----	-----	-----
<i>Sasa-DAB*0301</i>T.....AC..CA..A...	TA..CAGATA-	-----	-----	-----
<i>Onmy-Su2a</i>A	CCTTGCAGC	TCGGGGATT	GAACTTGCAA
<i>Onne-Wk136e</i>A	CCTTGCAGC	TCGGGGATT	GAACTTGCAA
<i>Onne-Wk136a</i>A	CCTTGCAGC	TCGGGGATT	GAACTTGCAA
<i>Onke-An13c</i>A	C---TCAGC	TCGGGGATT	GAACTTGCAA
	361		391		421	
<i>Safo-DAB*0101</i>	A CATTACAGTA	TGACTAGTCA	GCTATGTATC	A
<i>Safo-DAB*0201</i>GT	G
<i>Safo-DAB*0301</i>GT	G
<i>Safo-DAB*0401</i>C	GT
<i>Safo-DAB*0501</i>GT	G
<i>Safo-DAB*0601</i>GT	G
<i>Sasa-DAB*0201</i>C.....	..GT	G
<i>Sasa-DAB*0301</i>A.....	..C.....	..GT
<i>Onmy-Su2a</i>	CACTCTAGCC	ACTAGGCTAC	CCTGCCGCC
<i>Onne-Wk136e</i>	CACTCTAACC	ACTAGGCCAC	C-TGCCACCC
<i>Onne-Wk136a</i>	CACTCTAACC	ACTAGGCTAC	C-TGCCACCC
<i>Onke-An13c</i>	CACTCTAACC	ACTAGGCTAC	CCTGCCGCCA..G..

Box I

Box II

Supplementary Figure S3.2. Alignment of the second intron nucleotide sequences following the minisatellite found in the MHII β gene from brook charr and other salmonids. The last minisatellite motif sequence of each allele is in bold. Dots (.) indicate identity with *Safo-DAB*0101*; dashes (-), gaps introduced to maximize the alignment. Portions of the intronic sequences showing high conservation are identified with boxes (III - IV). The motifs (type II) forming the second minisatellite in Atlantic salmon are alternatively indicated by a grey and white rectangle. The complete *Hpa* I SINE sequence is underlined. GeneBank accession numbers for Atlantic salmon (*Salmo salar*; Sasa) and Pacific salmon (*Oncorhynchus* species; Onmy, Onne and Onke) sequences are AJ439067 , AJ439069 and U34715, U34713, U34712, U34702 respectively.

Chapitre 4

Discussion

Dans ce chapitre, nous allons discuter des résultats obtenus au sujet du polymorphisme du gène MHII β , tant au niveau de la séquence codante que de la variabilité d'un minisatellite identifié dans le second intron. Ensuite, les implications fonctionnelles de ces variations seront analysées en regard de leurs effets sur la résistance ou susceptibilité au pathogène *Aeromonas salmonicida* lors d'une infection ou sur la régulation de l'expression du gène, respectivement. Le polymorphisme sera également considéré en fonction des modèles évolutifs existants : la sélection par les pathogènes pour le maintien du polymorphisme au CMH et le modèle de "genomic design" pour les introns de gènes exprimés dans un type cellulaire spécifique.

4.1 Le gène MHII β chez l'omble de fontaine (*Salvelinus fontinalis*)

Nous avons entrepris la première étude (Chapitre 2) avec comme objectif initial de documenter, pour la première fois, la diversité allélique du gène MHII β dans une souche sauvage d'omble de fontaine. L'analyse des séquences obtenues a également révélé la présence d'un minisatellite de taille variable dans le second intron du gène. Ceci posait les bases de la seconde étude (Chapitre 3) qui avait comme premier objectif d'analyser la variabilité de ce VNTR. Cette section discute des résultats sur le polymorphisme des séquences codantes et non-codantes (intron 2) du gène MHII β .

4.1.1. Identification des allèles du gène MHII β

Nous avons identifié un total de 6 allèles du gène MHII β à partir d'une population initiale de 63 individus; quatre séquences codantes complètes et deux partielles incluant le domaine polymorphique β 1. La similarité entre les séquences en acides aminés du domaine β 2 de l'omble de fontaine et celles correspondantes chez d'autres poissons téléostéens et l'humain ne montre pas de différences majeures, particulièrement chez les salmonidés, confirmant que les quatre séquences complètes identifiées sont des allèles du gène MHII β (Figure 2.2). L'identification de ces séquences à

partir d'ARN messagers isolés de lymphocytes du sang périphérique (PBL) indique que ces allèles sont exprimés et les molécules ultimement retrouvées à la surface cellulaire pour présenter des antigènes. De plus, la présence des cystéines impliquées dans la formation des deux ponts disulfures, du site de glycosylation N-X-S/T, et de plusieurs acides aminés qui seraient impliqués dans l'interaction avec le corécepteur CD4⁺ des lymphocytes T suggère fortement la conservation de la structure et de la fonction de cette molécule chez l'omble de fontaine. Cependant, on observe une délétion spécifique d'un acide aminé pour tous les allèles à la position #65 et une seconde à la position #64 pour deux d'entre eux. L'analyse des séquences humaines des molécules du CMHIIβ indique que cette dernière position (#64) ainsi que les #66 et #67 seraient potentiellement impliquées dans des contacts avec la chaîne Vβ du TCR, suggérant ici une légère modification structurale.

4.1.1.1. Polymorphisme des allèles du gène MHIIB

Le polymorphisme important identifié pour les allèles du gène MHIIB au niveau du domaine β1 n'est pas surprenant compte tenu de son implication dans la présentation antigénique. L'excès de substitutions non-synonymes (d_N) par rapport aux substitutions synonymes (d_S) observé entre les allèles au domaine β1 (Tableau 2.1) a déjà été rapporté pour d'autres salmonidés, chez le saumon atlantique et plusieurs espèces de saumons du Pacifique (Miller et Withler 1996; Langefors et al. 2001). Ce résultat indique l'action d'une pression sélective positive qui opère à cet endroit pour maintenir le polymorphisme. Au contraire, le ratio de 0,31 ($\pm 0,082$) calculé pour la molécule complète en excluant ce domaine est plutôt indicateur d'une sélection stabilisatrice. La sélection par les pathogènes peut varier en intensité dans l'espace et le temps (Hedrick 2002). Chez le saumon atlantique, il a été montré que selon la latitude, la diversité au niveau du CMH augmente en fonction de la diversité bactérienne et de la température (Dionne et al. 2007). L'obtention d'un faible ratio d_N/d_S ($2,2 \pm 1,1$), par rapport à ceux rapportés pour les salmonidés ci-haut mentionnés, indique la présence

d'une pression sélective inférieure qui pourrait être associée aux caractéristiques de la population de pathogènes présente dans l'environnement de la rivière Rupert située au nord du Québec.

Les résultats indiquant une pression sélective inférieure nous révèlent cependant que son effet au niveau du nombre de positions polymorphiques reste inchangé. Les positions polymorphiques identifiées sur les allèles de l'omble de fontaine correspondent bien avec celles présumément impliquées dans la liaison au peptide pour l'allèle *HLA-DRB1*0101* (Brown et al. 1993), particulièrement dans la région N-terminale du domaine qui montre également une variabilité plus importante (Figure 2.2). Une variabilité plus prononcée dans cette région du domaine a également été rapportée pour le saumon atlantique et la truite arc-en-ciel (Glamann 1995; Langefors et al. 2001). De plus, la comparaison des séquences montre que 30 des 91 positions du domaine $\beta 1$ sont polymorphiques entre les allèles, chacune montrant une variabilité de deux à cinq acides aminés différents. Un nombre équivalent de positions polymorphiques et une telle variabilité en acides aminés ont été trouvés dans plusieurs études sur les salmonidés (Glamann 1995; Ristow et al. 1999; Langefors et al. 2001).

4.1.1.2. Polymorphisme de la séquence signal

Les résultats de séquences indiquent que certaines positions de la séquence signal des molécules MHII β sont variables (Figure 2.2). Cette séquence signal est nécessaire au ciblage de la molécule au réticulum endoplasmique par la protéine SRP (*signal recognition particle*) (Doudna et Batey 2004). Il a été démontré qu'une réduction de la charge nette positive ou de l'hydrophobicité de la séquence signal diminue l'efficacité de reconnaissance par la protéine SRP (Peterson et al. 2003). La comparaison de la séquence signal de l'allèle *DAB*0101* par rapport aux trois autres séquences disponibles indique deux mutations exclusives. Le remplacement de l'acide aspartique par une alanine (D6A) modifie la charge à l'extrémité N-terminale de la séquence signal, alors que la substitution d'une phénylalanine par une isoleucine (F14I) altère légèrement son hydrophobicité.

Comparativement aux autres allèles, les mutations observées dans la séquence signal de l'allèle *DAB*0101* pourraient affecter sa liaison à la protéine SRP et son intégration à la membrane du réticulum endoplasmique, menant ultimement à une variation de l'expression et de la présentation antigénique par la molécule du CMH de classe II à la surface cellulaire.

4.1.2. Variabilité du second intron du gène *MHII β*

Nous avons rapporté la présence d'un minisatellite polymorphique au motif de 32 nucléotides (type I) dans le second intron du gène *MHII β* chez l'omble de fontaine (Figures 3.1, 3.2, Tableau 3.2). Ce minisatellite avait déjà été identifié dans le gène orthologue chez le saumon atlantique (Stet et al. 2002) et sous la forme d'un seul motif chez plusieurs espèces de saumon du Pacifique (Miller et Withler 1996). La recherche de séquences homologues à celle du minisatellite n'a pas révélé sa présence chez d'autres espèces. Chez le saumon atlantique exclusivement, un second minisatellite au motif de 40 nucléotides et de séquence différente (type II) est situé en aval du premier au motif de 32 nucléotides. La longueur du second intron chez l'omble de fontaine est comparable à celle du saumon atlantique, mais supérieure aux saumons du Pacifique; la différence étant principalement associée à la présence de ces minisatellites. D'autres éléments sont cependant présents et contribuent à la variabilité de l'intron chez les espèces de saumon du Pacifique. Un élément *Hpa* I SINE partiel est situé en amont du seul motif de 32 nucléotides chez tous les individus, alors qu'un second *Hpa* I SINE complet est retrouvé en aval chez certaines espèces seulement (Miller et Withler 1996).

Un arbre phylogénétique, construit à partir des séquences exoniques du domaine $\beta 1$ des 6 allèles du gène identifiés chez l'omble de fontaine, indique qu'ils forment un groupe différent et séparé des allèles identifiés chez le saumon atlantique et chez sept espèces de saumon du Pacifique (Figure 2.3). La distribution des éléments retrouvés dans le second intron du gène *MHII β* corréle et renforce cet arbre phylogénétique qui établit un lien plus proche entre l'omble de fontaine et le saumon atlantique, par rapport aux espèces de saumon du Pacifique. Puisque l'implication des minisatellites et

des éléments SINE dans des événements de recombinaisons et de conversions géniques est bien connu (Jeffreys et al. 2004; Kazazian 2004), la présence de tous ces éléments dans le second intron soulève l'hypothèse que plusieurs mécanismes moléculaires auraient été sélectionnés afin de maintenir un niveau élevé de polymorphisme au gène *MHII β* chez les salmonidés.

4.1.2.1. Instabilité du minisatellite chez l'omble de fontaine

À l'exception du motif de l'allèle *Safo-DAB*0101* qui affiche un taux élevé de mutations sur toute sa longueur, la pureté du motif présent dans les autres allèles est très élevée, variant entre 95.6 % et 99.7 % (Figure 3.1, Tableau 3.2). Chez le saumon atlantique, le minisatellite de type I montre également une accumulation de mutations ponctuelles reflétant une pureté entre 92.0 % et 95.4 % sur toute sa longueur, alors que le minisatellite de type II a une pureté supérieure à 98.7 %. Pour les microsatellites, il a été démontré que l'expansion et la contraction des séquences d'ADN répétées en tandem tend à supprimer les imperfections retrouvées sur les motifs par le mécanisme de glissement de la polymérase, alors qu'en l'absence de cette activité purificatrice, les motifs accumulent des mutations ponctuelles (Ellegren 2000). Si on peut faire une extrapolation pour les minisatellites, ceci suggérerait que la grande pureté retrouvée au niveau du minisatellite de type I des allèles de l'omble de fontaine, à l'exception de l'allèle *Safo-DAB*0101* et de ceux du saumon atlantique, indique une activité purificatrice, présumément une conséquence de la force de sélection préservant la capacité d'expansion ou de contraction du minisatellite. Dans le même ordre d'idée, le minisatellite de type II chez le saumon atlantique aurait également conservé cette capacité d'expansion et contraction, et il est possible que ce deuxième minisatellite assure maintenant une activité auparavant associée au minisatellite de type I chez l'ancêtre de cette espèce.

L'analyse de la séquence du minisatellite retrouvé dans l'allèle *Safo-DAB*0101* révèle la présence d'une répétition parfaite d'un fragment d'ADN composé de 22 motifs représentant 704 nucléotides (Figure 3.1). Puisque le minisatellite retrouvé dans cet allèle est composé de motifs

complexes affichant un degré élevé de polymorphisme, cette répétition semble refléter le résultat d'un évènement de recombinaison et supporte ce mécanisme pour expliquer l'instabilité de ce minisatellite, comme suggéré pour ce type d'élément (Debrauwere et al. 1997).

4.1.2.2. L'origine du minisatellite chez les salmonidés

La comparaison du second intron du gène *MHIIβ* entre les différents salmonidés révèle qu'un seul motif de 32 nucléotides est présent chez les espèces de saumon du Pacifique, alors qu'il forme un minisatellite chez les autres salmonidés. Une telle situation a été rapportée dans la littérature. Chez le chimpanzé, un seul motif de 30 nucléotides est retrouvé dans l'intron 8 du gène du transporteur de la dopamine (*DAT1*), alors que sa répétition forme un minisatellite composé de 5 ou 6 copies du motif dans le gène orthologue humain (Guindalini et al. 2006). De plus, cette séquence n'est pas conservée chez le rat et la souris qui affichent plutôt à cet endroit un long microsatellite composé du dinucléotide $(GA)_n$ et $(CA)_n$, respectivement. Peu d'études ont étudié la variabilité d'un minisatellite particulier entre différentes espèces, mais celles effectuées sur les microsatellites indiquent qu'un SSR peut être parfois long et fortement polymorphique, ou court (2 à 4 répétitions) et presque invariable entre des espèces apparentés (Zardoya et al. 1996).

Un mécanisme, basé sur la présence de deux courtes séquences voisines et identiques, a été proposé pour expliquer la formation des minisatellites chez la levure (Haber et Louis 1998). Il a été suggéré que l'évènement initial de duplication d'une séquence quelconque située entre ces deux séquences identiques résultait d'un glissement de la polymérase lors de la réplication de l'ADN suivi d'un ré-appariement hors-cadre entre ces séquences. Il s'en suivrait une seconde synthèse de la séquence formant le motif qui est composé des nucléotides situés entre les deux séquences identiques. La première de celles-ci est donc ensuite retrouvée tout juste avant le premier nucléotide du minisatellite, alors que la seconde est répétée à la fin de chaque motif. L'étude de Richard et Dujon (2006) appuie ce modèle avec l'identification de ces séquences dans 49 minisatellites chez *S.*

cerevisiae, alors que ces dernières n'ont pu être identifiées dans 6 autres VNTR. Ce modèle ne semble cependant pas pouvoir expliquer l'amplification du minisatellite au motif de 32 nucléotides (type I) puisque de telles séquences ne sont pas retrouvées aux endroits attendus.

Les données actuelles proposent deux autres hypothèses pour expliquer la genèse des microsatellites (Buschiazzo et Gemmell 2006). La première suggère qu'un microsatellite *de novo* est créé spontanément à partir d'une séquence unique (Messier et al. 1996). Dans ce modèle, le microsatellite est généré via l'amplification d'un intermédiaire "proto-microsatellite" composé de 3 ou 4 motifs, et qui résulte d'évènements mutationnels. Étant donné la complexité de la séquence de 32 nucléotides, la formation d'un proto-minisatellite est improbable. De plus, puisque nous retrouvons la séquence sous la forme d'un seul motif, ce modèle ne semble pas pouvoir expliquer la création du minisatellite chez les salmonidés. La deuxième hypothèse suggère que les microsatellites sont adoptés d'une autre région du génome via des éléments transposables, notamment des SINE et des LINE qui contiennent un proto-microsatellite (Wilder et Hollocher 2001). Cette hypothèse est intéressante puisque nous avons identifié un *Hpa* I SINE partiel tout juste en amont du seul motif de 32 nucléotides retrouvé chez les espèces de saumon du Pacifique (Figure 3.1, Figure Supplémentaire S3.1). Cependant, elle expliquerait difficilement la présence du minisatellite chez l'omble de fontaine et le saumon atlantique qui sont d'origine plus ancienne, il faudrait alors pouvoir expliquer la délétion franche du *Hpa* I SINE dans ces espèces. Pour l'adaptation du modèle aux microsatellites, un seul motif aurait été apporté par le SINE plutôt qu'un proto-minisatellite. Cependant, le modèle n'explique pas la transition entre l'état d'un seul motif et un minisatellite et la recherche de séquences homologues au motif est restée infructueuse. Comme nous allons le suggérer dans une section suivante, les propriétés intrinsèques de la séquence du motif de 32 nucléotides pourraient possiblement expliquer l'évènement initial de blocage de la polymérase lors de la réplication de l'ADN, une étape qui semble primordiale à la formation d'un minisatellite.

4.1.2.3. Un seul locus MHII β exprimé chez l'omble de fontaine

Quatre régions (Box I à Box IV) du second intron du gène MHII β affichent une divergence qui est en moyenne sous les 10 % entre les différentes espèces de salmonidés (Figure 3.1, Figures Supplémentaires S3.1, S3.2). Les régions les mieux conservées, Box 1 et Box 4, sont plus longues et situées aux extrémités de l'intron, alors que les régions Box 2 et Box 3 sont plus courtes et incluent les nucléotides faisant partie des motifs des deux minisatellites identifiés. Cette faible divergence entre les séquences introniques combinée à la présence des motifs sous forme de séquences uniques ou de minisatellites supportent l'hypothèse d'un seul et même locus exprimé chez l'omble de fontaine et les autres salmonidés (Glamann 1995; Langefors et al. 2000). Cette hypothèse est également appuyée par le fait que pour chaque individu, seulement un ou deux produits PCR couvrant le second intron ont été obtenus, avec une longueur particulière correspondant toujours au même allèle. De plus, les analyses SSCP et le séquençage de l'exon 2 n'ont jamais permis d'identifier plus de deux allèles par poisson.

4.2. Régulation de l'activité transcriptionnelle du gène MHII β

La plupart des études sur les gènes du CMH s'intéressent au polymorphisme de la séquence codante alors que peu d'entre elles ont porté attention à l'implication du niveau d'expression des gènes dans la réponse adaptative. Chez les salmonidés, il a été rapporté que le TNF- α et le LPS agissent comme régulateur de l'expression des gènes MH de classe II dans les macrophages (Knight et al. 1998). On a caractérisé leur promoteur chez le saumon atlantique (Syed et al. 2003) et montré la modulation de l'activité transcriptionnelle de gènes du système immunitaire par la température (Rodrigues et al. 1998; Nath et al. 2006). Cette section discutera des résultats présentés dans le Chapitre 3 concernant l'effet de la variation de taille du minisatellite sur le niveau d'expression du gène MHII β dans différentes conditions environnementales qui constituait d'ailleurs le second objectif de l'étude de cette étude.

4.2.1. Modulation de l'expression du gène MHII β par le minisatellite

Nous avons décrit un effet fonctionnel sur l'activité transcriptionnelle du gène MHII β qui est associé à la présence d'un minisatellite polymorphique dans le second intron. La modulation de l'activité transcriptionnelle par des minisatellites a été démontrée pour plusieurs gènes tels que celui de l'insuline (INS) (Bennett et al. 1995), de la cystatine B (CSTB) (Lalioti et al. 1997), de la monoamine oxydase A (MAOA) (Sabol et al. 1998), du transporteur de la sérotonine (5-HTT) (Lesch et al. 1996; MacKenzie and Quinn 1999), du transporteur de la dopamine (DAT1) (Mill et al. 2002) et du récepteur de la bradykinine de type 2 (B2) (Zamorano et al. 2006). Les résultats présentés au Chapitre 3 indiquent que les allèles du gène MHII β qui portent un long minisatellite montrent une réduction de l'expression du gène d'un facteur 1,67 à 2,56 par rapport aux allèles qui en contiennent un court (Figures 3.3, 3.4, 3.5). Au contraire, des allèles qui possèdent des minisatellites de tailles similaires n'affichent pas de différence significative au niveau de l'abondance des transcrits (Figure 3.5), confirmant l'association entre la taille du minisatellite et l'effet sur la transcription. La réduction de l'expression est de la même ampleur de ce qui est typiquement observé pour une variation allélique, soit de l'ordre de 1,5 à 3,0 fois (Cowles et al. 2002; Yan et al. 2002).

4.2.1.1. Identification de l'effet du minisatellite dans différents tissus

Les résultats de notre étude ont été confirmés dans une variété de tissus. Premièrement, deux souches différentes d'omble de fontaine ont permis de démontrer la réduction de l'expression du gène MHII β associé à un long minisatellite. Cet effet semble indépendant du degré de maturité et de la taille du poisson puisque les individus de la souche domestique (environ 50 g) ont été produits à l'automne précédant le premier échantillonnage et n'avaient pas encore un an, alors que ceux de la souche Rupert (environ 1 kg) proviennent de familles créées en décembre 2001 et avaient donc plus de cinq ans (Perry et al. 2004). Ensuite, l'action du minisatellite sur l'activité transcriptionnelle du gène a été démontrée dans deux types de tissus différents; au niveau de l'ARN total extrait du pronéphros (*head kidney*) par

rapport à un gène de référence (Figures 3.3, 3.4), ou par comparaison de l'expression relative des allèles dans des lymphocytes de poisson hétérozygotes pour le MHII β (Figure 3.5). Des études indiquent que ces deux tissus expriment modérément le gène MHII β (Juul-Madsen et al. 1992; Koppang et al. 1998; Nath et al. 2006). L'analyse de la variation allélique de l'activité transcriptionnelle dans un même échantillon a plusieurs avantages (Yan et al. 2002). Elle permet la détection de petites différences dans les niveaux d'expression entre les allèles tout en éliminant la variation interindividuelle qui pourrait provenir d'un bruit de fond ou d'un effet de famille. La caractérisation de l'effet génétique direct est aussi favorisée par rapport aux conditions environnementales ou physiologiques qui pourraient masquer ou fausser sa valeur.

4.2.1.2. Identification de l'effet du minisatellite dans plusieurs conditions expérimentales

Plusieurs conditions expérimentales ont permis de confirmer l'effet du minisatellite sur la transcription du gène MHII β . Les résultats ont été obtenus par les techniques de RT-PCR multiplex (Figure 3.3) et de PCR quantitatif en temps réel (Figures 3.4, 3.5). Les données obtenues par cette dernière technique ont par la suite été confirmées par l'utilisation d'amorces différentes pour un même allèle (Tableau 3.1). À l'exception des derniers nucléotides de l'extrémité 3' qui étaient allèles spécifiques, les amorces utilisées pour la quantification entre les allèles étaient localisées à la même position sur le gène et produisaient des amplicons de même taille afin d'éviter un effet indirect produit par la réaction d'amplification. La sélection du gène de référence est également un facteur critique dans les expériences sur l'expression génique. À cet effet, nous avons utilisé deux gènes de références différents le GAPDH (Figure 3.3) et le EFIA_B (Figure 3.4), le second affichant le niveau d'expression le plus stable parmi une sélection de gènes chez le saumon atlantique (Olsvik et al. 2005). Ces deux gènes de référence ont montré un niveau transcriptionnel stable dans les tissus étudiés aux deux températures de 6 °C et 18 °C. Les effets associés à la présence d'un minisatellite particulier sont fréquemment issus d'études comparatives de seulement deux allèles majoritairement retrouvés dans

une population. Il en est ainsi pour deux études qui rapportent l'effet de minisatellites introniques retrouvés dans les gènes 5-HTT et DAT1 humains (MacKenzie et Quinn 1999; Guindalini et al. 2006). Nos résultats ont été observés dans plusieurs individus hétérozygotes et impliquent quatre allèles dans trois génotypes différents. L'allèle possédant le plus long minisatellite (*Safo-DAB*0101* et *Safo-DAB*0401*) a toujours montré un niveau d'expression inférieur par rapport à celui qui en possédait un plus court (Figure 3.5). De plus, les différences de taille des minisatellites retrouvés sur chaque allèle sont fréquemment mineures; 10 ou 12 répétitions d'un motif de 17 pb pour celui du gène 5-HTT, 5 ou 6 copies d'un motif de 30 pb dans le gène DAT1 (MacKenzie et Quinn 1999; Guindalini et al. 2006). La variation de taille des différents minisatellites identifiés dans le gène MHII β de l'omble de fontaine est plus importante (20, 27, 40, 69 répétitions), se comparant davantage aux variations de tailles observées pour les microsatellites de gènes impliqués dans le développement de pathologies et pour lesquels l'effet fonctionnel du microsatellite a été bien établi (Usdin 2008). Il est intéressant de constater que la différence entre le niveau d'expression de l'allèle *Safo-DAB*0101* par rapport à *Safo-DAB*0301* n'est pas plus grande que celle entre *Safo-DAB*0401* et *Safo-DAB*0301*, malgré un nombre supérieur de motifs retrouvés dans *Safo-DAB*0101* (69 par rapport à 40). Une explication possible pourrait être la présence de plusieurs mutations dans cet allèle qui réduisent son activité.

4.2.2. Modulation de l'expression du gène MHII β par la température

La comparaison de poissons ayant des courts minisatellites nous a permis d'observer une augmentation significative de l'expression totale du gène MHII β d'un facteur 2,08 à la basse température de 6 °C par rapport à 18 °C (Figure 3.4). La différence est sensiblement de la même ampleur pour les poissons possédant des longs minisatellites, un résultat toutefois démontré avec moins d'individus. L'augmentation du niveau de l'expression du gène à basse température est en accord avec les résultats de Raida et Buchmann (2007) sur l'expression constitutive du gène MHII β dans le pronéphros chez la truite arc-en-ciel. Ces auteurs indiquent les valeurs des cycle-seuils (*threshold cycle*,

CT) à différentes températures au sujet du niveau d'expression de gènes impliqués dans la réponse immunitaire. Pour l'expression du gène MHII β , on remarque une différence de 0,8 CT brut entre l'expression à 5 °C et 15 °C, alors que l'expression du gène de référence EF1- α est stable. Puisque l'efficacité d'amplification des deux gènes est presque identique, 101,1 % et 100,0 % respectivement, ceci représente une augmentation de l'expression d'un facteur 1,74 pour le gène MHII β à la température de 5 °C, un résultat semblable à ce que nous avons obtenu. D'un autre côté, Nath et collaborateurs (2006) ont observé une baisse du niveau d'expression du gène MHII β au niveau de l'ARNm et de la protéine à la basse température de 2 °C. Cependant, ce résultat conflictuel semble refléter la conséquence de l'exposition des individus sous la température non permissive pour le système immunitaire d'environ 4 °C (Fryer et al. 1976; Bly and Clem 1991), et près de la température létale de -0,7 °C pour les salmonidés (Saunders 1995). En plus, l'expression du gène encodant la protéine ribosomale S11 montre également une réduction du niveau du transcrit à 2°C, suggérant que l'effet rapporté serait plutôt le résultat d'une répression générale et sévère de l'activité transcriptionnelle. Il faut cependant considérer l'étude de Rodrigues et collaborateurs (1998) chez la carpe qui a montré une baisse de l'expression de la β_2 -microglobuline à la basse température de 6 °C, alors qu'un niveau stable a été rapporté chez la truite arc-en ciel et le saumon atlantique (Kales et al. 2006). Les résultats de ces deux études suggèrent que des moyens alternatifs de contrôle pourraient opérer afin de réguler l'activité transcriptionnelle de gènes impliqués dans l'activité du système immunitaire à basse température chez différentes espèces.

4.2.2.1 Effet de la température sur l'activité du minisatellite

Les résultats obtenus de poissons de souche Rupert hétérozygotes pour le gène MHII β dans trois génotypes différents, indiquent qu'à la température élevée de 18 °C par rapport à 6 °C, l'effet répressif associé au long minisatellite sur l'activité transcriptionnelle du gène est amplifié (Figure 3.5). Plusieurs études ont rapporté une modulation par la température de l'expression de gènes impliqués

dans l'activité du système immunitaire (Rodriguez et al. 1998; Nath et al. 2006; Kales et al. 2006; Raida et Buchmann, 2007), mais aucune ne s'est intéressée au mécanisme moléculaire qui pouvait en être responsable. Un microsatellite codant identifié dans le gène PER aurait un rôle à jouer dans la régulation du rythme circadien face aux fluctuations de température chez la drosophile (Sawyer et al. 1997). Ce microsatellite serait impliqué au niveau structurel dans le maintien d'une conformation optimale de la protéine en fonction des conditions environnementales (Peixoto et al. 1998). Pour les minisatellites situés dans la région régulatrice des gènes INS et CSTB, deux études ont proposé l'adoption de structures non-B de l'ADN comme mécanisme permettant d'affecter l'activité transcriptionnelle (Catasti et al. 1996; Saha et Usdin 2001). L'effet supplémentaire qui est observé à haute température pour les allèles qui possèdent un long minisatellite pourrait constituer un moyen additionnel permettant de réguler finement leur niveau d'expression en réponse aux conditions environnementales. Ceci serait reflété par une plus grande énergie thermodynamique à haute température pour favoriser l'adoption d'une structure secondaire de l'ADN comme mécanisme fonctionnel.

4.2.2.2. Modulation de l'expression de certains gènes du système immunitaire par la température

Lors de l'évaluation des gènes de référence potentiels pour l'expérience de PCR quantitatif en temps réel, celui de la β_2m s'est étonnamment classé comme le plus variable. L'examen des résultats obtenus pour ce gène a révélé que le CT moyen à 18 °C était de $18,18 \pm 0,44$ et de $16,89 \pm 0,40$ à 6 °C, ce qui représente une augmentation significative (*t*-test: $p < 0.001$, $n=6$) par un facteur 2,78 de l'expression à 6 °C une fois les données corrigées par rapport au gène de référence EFIA_B (voir Annexe 7.1). Ce résultat concorde avec les études réalisées chez le saumon atlantique et la truite arc-en-ciel qui ne montrent pas de diminution du transcrit de la β_2m après 10 jours à 2 °C (Kales et al. 2006), contrairement à la réduction de l'expression rapportée à 6 °C chez la carpe (Rodrigues et al. 1998). La raison pour laquelle Kales et collaborateurs (2006) n'ont pas détecté de variation significative du

niveau d'expression du gène de la β_2m à basse température pourrait être la faible sensibilité de la technique utilisée ou les différentes températures auxquelles les poissons ont été exposés. Les résultats conflictuels rapportés au sujet des niveaux d'expression des gènes MHII β et β_2m en fonction de la température illustrent le besoin d'approfondir l'étude de l'expression des gènes impliqués dans l'action du système immunitaire aux températures extrêmes retrouvées dans les habitats, en prenant également soin d'inclure des conditions intermédiaires.

Quels seraient les avantages pour le poisson d'avoir un niveau d'expression du MHII β réduit à haute température sachant qu'une maladie humaine, le syndrome des lymphocytes dénudés (*Bare lymphocytes syndrome*, BLS), est caractérisée par une absence des molécules du CMH de classe II à la surface cellulaire et l'incapacité de monter une réponse immunitaire lymphocytes T-dépendent (Reith and Mach 2001)? En réduisant le niveau d'expression du gène MHII β à haute température (et possiblement d'autres gènes d'intérêt immunologique), l'avantage pourrait être une diminution du coût en énergie nécessaire pour monter une réponse immunitaire efficace. Cette hypothèse est basée sur le fait que la température s'avère être un facteur important pour gouverner la vitesse des processus cellulaires, que ce soit le métabolisme et/ou les activités fonctionnelles. Comme mentionné précédemment, des études ont démontré que quelques gènes MH ont un niveau d'expression dépendant de la température (Rodrigues et al. 1998; Kales et al. 2006; Nath et al. 2006; Raida et Buchmann 2007). Il a été suggéré que la fluidité des membranes pourrait être impliquée dans l'activation des lymphocytes T en fonction de la température (Vallejo et al. 1992). Une plus faible expression du gène MHII β pourrait être compensée par une activité ou efficacité supérieure des molécules à la membrane pour l'activation des lymphocytes T CD4⁺ à haute température.

Un autre avantage au niveau de l'économie d'énergie implique l'ajustement fin par le minisatellite du niveau d'expression du gène MHII β . La proximité physique du minisatellite retrouvé dans le second intron du gène et de la séquence codante pour le domaine $\beta 1$ (exon 2) responsable de la présentation antigénique suggère une coévolution des deux séquences. Il est connu que des allèles

spécifiques du gène MHII β sont associés à une résistance accrue aux infections (Langefors et al. 2001; Lohm et al. 2002; Grimholt et al. 2003; Kjølglum et al. 2006; 2008; de Eyto et al. 2007; Croisetièrre et al. 2008). Une sélection est envisageable entre deux allèles qui possèdent une séquence codante identique et qui ne diffèrent que par la longueur du minisatellite. La base de cette sélection pourrait reposer sur un niveau d'expression optimal pour l'allèle, qui implique un certain coût en énergie, en fonction de son efficacité à présenter les antigènes et à induire une réponse immunitaire appropriée. Dans cette optique, il serait intéressant de développer différents poissons ou organismes transgéniques avec une même séquence codante pour la molécule du CMH de classe II, mais ne différant que par la taille du minisatellite, et de les tester ensuite dans une expérience d'infection afin d'évaluer la contribution du niveau d'expression du gène à la résistance ou susceptibilité face aux pathogènes.

4.2.3. Analyse du motif du minisatellite et mécanisme fonctionnel

Deux mécanismes par lesquels les minisatellites peuvent affecter l'activité transcriptionnelle d'un gène ont jusqu'à maintenant été documentés. Un premier implique la présence de sites de liaison pour des facteurs de transcription à l'intérieur du motif, alors que l'autre implique la formation d'une structure secondaire de l'ADN de type non-B qui perturbe les activités métaboliques (Kennedy et al. 1995; Catasti et al. 1996; Saha et Usdin 2001; Klenova et al. 2004). La séquence consensus du motif qui forme le minisatellite que nous avons identifié est : ACATTACAGTATGACT-AGTCAGCTATGTAGTG (Figures 3.1, 3.2). Dans un premier temps nous avons tenté, mais sans succès, d'identifier sur le minisatellite des sites de liaison connus pour différentes protéines en utilisant le logiciel de reconnaissance de motifs consensus *Cluster-Buster* (Frith et al, 2003). Ensuite nous nous sommes intéressés à la possibilité qu'aurait le minisatellite d'adopter une structure secondaire de l'ADN. Les séquences d'ADN répétées en tandem retrouvées dans une orientation directe peuvent adopter une structure glissantes/épingle à cheveux (Bacolla et Wells 2004). De plus, celles répétées dans une orientation inversée ou qui contiennent un centre de symétrie dans le motif peuvent générer

une structure cruciforme. La Figure 4.1 illustre une structure secondaire de type épingle à cheveux proposée à partir de la séquence de trois motifs d'ADN simple-brin. L'appariement des nucléotides est élevé et implique 26 des 32 nucléotides du motif formant cette structure. De plus, on y retrouve un centre de symétrie qui permettrait possiblement la formation d'une structure à partir d'un seul motif (Figure 4.1, 2^e motif en rouge). Cette structure est hypothétique mais illustre tout de même les caractéristiques particulières et intéressantes de la séquence du motif en vue d'investiguer plus à fond l'hypothèse de la formation d'une structure secondaire de l'ADN comme mécanisme fonctionnel du minisatellite pour réguler l'activité transcriptionnelle du gène MHIIB.

4.2.4. Implications fonctionnelles de la variation de l'expression du MHIIB

Le polymorphisme au niveau de la séquence codante des molécules présentatrices d'antigènes peut influencer le ciblage et la signalisation intracellulaire ainsi que la liaison du peptide antigénique. Il a été démontré chez l'humain et la souris que le polymorphisme observé dans le promoteur et la région régulatrice des gènes du CMH de classe II, pour différents isotypes ou allèles, est responsable de la variation du niveau d'expression du gène (Beaty et al. 1999). De plus, les différences alléliques observées au niveau de l'ARNm se traduisent par une variation de la densité des molécules à la surface en fonction du type cellulaire.

Parmi les 4 sous-types fonctionnels de lymphocytes T CD4⁺ identifiés jusqu'à maintenant, les Th1 et Th2 se distinguent au niveau de la sécrétion des cytokines et de leur implication dans la réponse immunitaire face aux pathogènes intracellulaires ou aux parasites extracellulaires, respectivement (Zhu et Paul 2008). L'amplitude et le caractère de la réponse immunitaire initiale assurée par les lymphocytes T CD4⁺ sont directement influencés par l'abondance et la nature du peptide présenté par les molécules du CMH de classe II. Les cellules présentatrices d'antigènes (APC) exposées à une forte concentration d'antigènes ou présentant un puissant antigène tendent à favoriser l'activation du sous-



Figure 4.1. Structure secondaire hypothétique pour le minisatellite de type I. La structure secondaire de type épingle à cheveux est illustrée à partir de la séquence de trois motifs de type I d'ADN simple-brin. L'appariement implique 26 des 32 nucléotides du motif et le second motif (rouge) montre l'adoption d'une structure intra-motif. Le centre de symétrie est indiqué par la ligne pointillée.

type Th1 (Valitutti et al. 1996; Crowe et al. 1998). L'INF- γ produit par ces lymphocytes T CD4⁺ Th1 est un activateur important de l'activité microbicide des macrophages (Suzuki et al. 1988).

Des évidences suggèrent que la densité des molécules du CMH de classe II à la surface cellulaire influence l'activité des lymphocytes CD4⁺ de la même manière que la concentration ou la nature des antigènes. Il a été observé que des macrophages traités à l'IL-10 ont une capacité réduite à présenter des antigènes, qui résulte d'une baisse de l'expression des molécules du CMH de classe II. Ceci se traduit par une réponse proliférative plus faible des lymphocytes T CD4⁺ Th1 et Th2 en réponse à la présentation d'un antigène par des macrophages exposés à différentes doses (de Waal Malefyt et al. 1991). Des souris transgéniques exprimant un niveau plus faible de la molécule du CMH de classe II dans les lymphocytes B et les cellules dendritiques ont été utilisées pour étudier l'influence de la densité de la molécule à la surface sur la différenciation des lymphocytes T CD4⁺ (DiMolfetto et al. 1998). Alors que l'activation et l'expansion des lymphocytes T étaient normales chez ces souris, la réponse immunitaire était dominée par la sécrétion de INF- γ associé au sous-type Th1 et une diminution du niveau d'IL-4 qui est associé au sous-type Th2. Une autre étude faite avec des souris transgéniques montre que celles qui expriment un niveau plus faible de la molécule du CMH de classe II permettent la sélection des lymphocytes T CD4⁺ dans le thymus, alors que ces derniers sont éliminés dans les souris qui ont un niveau plus élevé, démontrant également l'implication du niveau d'expression de la molécule du CMH de classe II sur la formation du répertoire de lymphocytes T (Fukui et al. 1997).

Un plus faible niveau d'expression du gène MHIIB identifié pour les allèles portant un long minisatellite chez l'omble de fontaine pourrait possiblement avoir des implications sur l'orientation de la réponse immunitaire vers les lymphocytes T du sous-type Th1 et sur la formation du répertoire. Le système immunitaire des poissons téléostéens n'a cependant pas été étudié avec autant de profondeur que ceux de la souris et de l'homme, plus particulièrement à savoir si le développement des lymphocytes T en sous-type Th1 et Th2 est également retrouvé. Néanmoins, puisque jusqu'à

maintenant les études comparatives de ces espèces n'ont pas permis de détecter de différences majeures entre les systèmes immunitaires, une telle spécialisation des lymphocytes T lors de leur développement est envisageable chez les poissons téléostéens, et constitue certainement une avenue intéressante de recherche. Il est intéressant de constater qu'un des pathogènes importants des salmonidés (*Aeromonas salmonicida*) est capable de survivre et de se répliquer à l'intérieur des macrophages chez l'omble de fontaine (Daly et al. 1996; Garduño et al. 2000). La variation de taille du minisatellite présent dans le second intron du gène pourrait constituer un mécanisme moléculaire permettant aux poissons d'adapter le niveau d'expression de la molécule MHIIB dans un environnement riche en pathogènes intracellulaires. Il faudrait cependant confirmer dans une future étude que la variation de l'expression de l'ARNm se traduit également au niveau de la densité de la molécule à la surface cellulaire.

4.3 Le gène MHIIB et la résistance ou susceptibilité à une infection

Des associations entre différents allèles des gènes du CMH de classe I et de classe II et plusieurs maladies ont été trouvées dans une variété d'espèces (Paterson et al. 1998; Kaufman 2000; Shiina et al. 2009). Les salmonidés n'expriment qu'un seul gène MHIIB (Glamann 1995; Langefors et al. 2000) et les loci de classe II α et de classe II β sont génétiquement liés chez le saumon atlantique (Stet et al. 2002). Ce modèle animal a été utilisé dans plusieurs études d'association pour identifier des allèles de résistance ou de susceptibilité face à différents pathogènes (Langefors et al. 2001; Grimholt et al. 2003; Dionne et al. 2009). Suite aux travaux de Perry et collaborateurs (2004) qui établissaient une base génétique héritable pour la résistance de l'omble de fontaine à la furunculose, nous avons entrepris l'étude présentée au Chapitre 2 dont le second objectif était de documenter pour la première fois les performances de survie conférées par des allèles ou génotypes spécifiques, suite à l'exposition d'un poisson autre que le saumon atlantique au pathogène *Aeromonas salmonicida*. Les résultats obtenus sont discutés dans cette section.

4.3.1. Performances de survie associées aux allèles et aux génotypes

Des quatre allèles présents dans les poissons testés lors de l'expérience d'infection avec le pathogène *Aeromonas salmonicida*, l'allèle *Safo-DAB*0101* a été significativement associé à une résistance à la furonculose dans l'analyse allélique (analyse simple) (Tableau 2.2). Ce résultat est appuyé par l'analyse génotypique (analyse combinée) (Tableau 2.3) dans laquelle les poissons homozygotes pour cet allèle, *DAB*0101/*0101*, ont montré une tendance à être résistants à la furonculose, et aussi par les résultats de mortalité qui indiquent généralement une plus faible mortalité pour les poissons qui possèdent au moins une copie de l'allèle *DAB*0101*, particulièrement les homozygotes. À l'opposé, les résultats de l'analyse simple pour l'allèle *DAB*0201* et de l'analyse combinée pour le génotype *DAB*0201/*0201* suggèrent tous deux une susceptibilité à la maladie. De plus, il est intéressant de constater que la résistance relativement élevée observée pour le génotype *DAB*0101/*0201* suggère que l'effet dominant associé à la résistance pour l'allèle *DAB*0101* est plus fort que l'effet dominant de susceptibilité conféré par l'allèle *DAB*0201*. Un tel effet a auparavant été démontré chez le saumon atlantique (Lohm et al. 2002), où un patron de codominance pour la résistance ou la susceptibilité a été trouvé, reflétant des différences qualitatives de la réponse immunitaire entre des individus possédant des allèles de faible ou de forte résistance. Somme toute, les résultats d'analyse indiquent que la résistance ou la susceptibilité à la furonculose semblent être associées à des allèles MHI β spécifiques plutôt qu'à l'hétérozygotie des génotypes. Il faut cependant prendre en considération le fait que trop peu d'individus de génotype *DAB*0101/*0401* étaient disponibles pour évaluer la dominance de *DAB*0101* par rapport à *DAB*0401*.

Les résultats obtenus pour les poissons de génotype *DAB*0301/*0301*, même s'ils ne sont pas significatifs, suggèrent une résistance accrue, alors que les facteurs de risque élevés pour les génotypes *DAB*0201/*0301* et *DAB*0301/*0401* suggèrent plutôt une association à la susceptibilité. D'un autre côté, lorsqu'on regarde les résultats de l'analyse simple pour l'allèle *DAB*0301*, une valeur intermédiaire du facteur de risque indique un effet nul ou confondu. Il est donc possible que

l'augmentation de la susceptibilité apparente des individus hétérozygotes possédant cet allèle ait été causée par le second allèle formant le génotype, notamment les allèles *DAB*0201* et *DAB*0401* qui sont associés à un facteur de risque élevé (non significatif) dans l'analyse simple. Ainsi, ces résultats suggèrent un effet dominant des allèles *DAB*0201* et *DAB*0401* associés à la susceptibilité par rapport à l'allèle *DAB*0301*.

4.3.2. Conditions expérimentales de l'infection avec *Aeromonas salmonicida*

Dans le but de caractériser les performances de survie des allèles et des génotypes face à *Aeromonas salmonicida*, l'expérience d'infection a été réalisée avec un nombre équivalent de poissons provenant de 23 familles d'omble de fontaine distinctes (Perry et al. 2004). L'association qui a été faite entre l'allèle *DAB*0101* et la résistance à la furunculose provient de l'analyse des performances de survie d'individus provenant de sept familles qui ont été conçues à partir de 3 femelles et/ou 2 mâles différents possédant au moins une copie de l'allèle. Des précautions doivent ainsi être prises pour contrôler le bruit de fond ou l'effet de famille afin d'éviter une association qui serait plutôt causée par la coségrégation d'un locus avec celui du gène *MHIIβ*, en permettant tout de même la variation aux autres loci.

Le mode d'infection par immersion qui a été choisi (Lutwyche et al. 1995) représente une technique rapide et standardisée où tous les poissons utilisés dans l'expérience sont exposés à environ la même charge bactérienne simultanément. Elle permet également le développement des symptômes caractéristiques de la furunculose tels que le noircissement des poissons, la léthargie et l'apparition d'hémorragies (Bernoth et al. 1997). Le synchronisme de l'infection pourrait expliquer le taux de mortalité élevé qui est rapidement observé vers la 40e heure en comparaison avec les résultats publiés montrant un tel taux vers le 6^e, 10^e ou 12^e jour lorsque la méthode d'infection par cohabitation est utilisée (Lohm et al. 2002; Grimholt et al. 2003; Kjøglum et al. 2006). Des questions ont également été soulevées en regard de la mortalité rapide qui survient dans ces types d'infections "artificielles", plus

spécifiquement au sujet de la contribution unique du système immunitaire adaptatif (Grimholt et al. 2003). Puisque nous avons observé une association entre des génotypes et des performances de survie dans ce modèle d'infection produisant rapidement une mortalité élevée, une autre avenue de recherche pourrait être l'effet qu'ont des allèles du MHII β sur l'initiation de la réponse immunitaire. Ces études pourraient s'intéresser à l'évaluation des fonctions immunitaires primaires suite à l'infection d'individus de différents génotypes. Elles pourraient inclure par exemple des essais d'activation et de prolifération de lymphocytes, et d'autres pour caractériser l'activité phagocytaire, la présentation antigénique et la quantification des cytokines sécrétées (Kollner et al. 2002).

4.3.3. Analyse des positions polymorphiques de l'allèle *Safo-DAB*0101*

La variabilité de certains acides aminés de la région de liaison au peptide des molécules du CMHII influence directement les propriétés de liaison du peptide antigénique ainsi que la susceptibilité aux maladies (Jones et al. 2006). Il a été rapporté que la mutation d'un seul acide aminé dans la région N-terminale du domaine β 1 d'un allèle est suffisante pour lui permettre de conférer une résistance accrue (Langefors et al. 2001). Nous avons comparé les séquences des domaines β 1 entre l'allèle *Safo-DAB*0101* et les trois autres allèles à l'étude dans l'expérience d'infection pour identifier les positions où l'identité de l'acide aminé lui était exclusive, et possiblement responsable d'un effet fonctionnel.

Sur la base de la numérotation de l'allèle *HLA-DRB*0101*, nous avons identifié 6 acides aminés exclusifs à l'allèle *DAB*0101* dans le domaine β 1: β 9Y, β 12L, β 13A, β 53Y, β 69G et β 70P, trois d'entre eux étant concentrés entre les positions 9 et 13 (Figure 2.2). La position β 9 est particulièrement intéressante pour deux raisons. Premièrement, la substitution observée modifie la polarité de l'acide aminé à cette position, changeant une chaîne latérale chargée positivement (β 9H) ou négativement (β 9E) pour une autre non-chargée, polaire et aromatique (β 9Y). Deuxièmement, cette position est située juste à côté de l'histidine (β 10H) qui a été identifiée comme responsable de l'émergence d'un allèle associé à la résistance contre la furonculose chez le saumon atlantique

(Langefors et al. 2001). Il est intéressant de constater qu'une histidine est également présente à la position $\beta 10$ dans l'allèle *Safo-DAB*0101*. Sur la base de la structure tridimensionnelle de la molécule du CMHII (Brown et al. 1993), cet acide aminé chargé identifié par Langefors et collaborateurs (2001) correspond à $\beta 10Q$, une position sur l'allèle *HLA-DRB*0101* qui est située au fond de la poche P6 de la molécule (Jones et al. 2006). Chez les salmonidés, la présence d'acides aminés chargés est fréquemment observée entre les positions $\beta 9$ et $\beta 13$ qui sont d'ailleurs fortement polymorphiques (Figure 2.2). Cette séquence d'acides aminés est située au fond de la niche peptidique entre les poches P4 et P6, et la présence d'acides aminés particuliers pourrait être critique pour une présentation efficace des antigènes. De ces études, il apparaît que les positions polymorphiques impliquées dans la formation des poches P4 et P6 du domaine $\beta 1$ pourraient être importantes pour la liaison de peptides antigéniques dérivés d'*Aeromonas salmonicida*. La tyrosine située à la position $\beta 53$ ($\beta 53Y$) est un autre acide aminé exclusivement retrouvé dans l'allèle *DAB*0101*. Elle est située au milieu de l'interface de dimérisation entre deux molécules du CMHII et pourrait possiblement affecter la stabilité ou la structure du dimère (Brown et al. 1993). Les acides aminés $\beta 69G$ and $\beta 70P$ sont tous deux localisés à des positions qui ont été impliquées dans l'interaction entre les molécules HLA-DR1 ou I-Ak avec le TCR des lymphocytes T $CD4^+$ (Hennecke et al. 2000). La présence d'une proline à la place d'une glutamine habituellement conservée à la position $\beta 70$ pourrait avoir un impact important sur la stimulation de la réponse immunitaire puisque l'acide aminé à cette position possède le plus grand nombre moyen de contacts avec le TCR (Rudolph et al. 2006). Considérant la localisation des six acides aminés exclusifs identifiés pour l'allèle *Safo-DAB*0101* dans une des trois régions fonctionnelles identifiées, un ou plusieurs pourrait (ent) être responsable (s) de la résistance accrue associée aux poissons possédant cet allèle. Il faut noter que la résistance ou la susceptibilité peuvent également être imputables à des acides aminés retrouvés sur la chaîne α de la molécule MHII. Sachant que les gènes qui codent pour les deux chaînes formant la molécule sont coségrégés (Stet et al. 2002), il serait intéressant de séquencer les

deux composantes et éventuellement de modéliser les différentes molécules. En combinaison, l'identification des peptides antigéniques majeurs dérivés d'*Aeromonas salmonicida* pourrait également s'avérer importante pour notre compréhension de la virulence de ce pathogène, en plus de raffiner une éventuelle structure de la présentation antigénique par le complexe.

4.4 Pressions évolutives sur les gènes du CMH

L'évolution et la maintenance du polymorphisme des gènes du CMH peuvent être attribuables à deux types de mécanismes : la sélection par les pathogènes et les mécanismes reproductifs (Aguilar et Garza 2007; Bernatchez et Landry 2003). Les trois modèles proposés pour expliquer la sélection par les pathogènes incluent : l'avantage hétérozygote (*overdominant selection*), la sélection dépendante de la fréquence (*negative frequency dependent selection*) et la sélection fluctuante dans l'espace et le temps (*fluctuating selection*) (Hughes et Nei 1988; 1989; Takahata et Nei 1990; Slade et McCallum 1992; Hedrick 2002). La variation de la taille des introns est également soumise à des pressions évolutives selon la fonction du gène : le modèle de la sélection pour l'économie (*selection for economy*) pour les gènes fortement exprimés et celui du motif génomique (*genomic design*) pour les gènes ayant une régulation et une fonction complexe (Castillo-Davis et al. 2002; Eisenberg et Levanon 2003; Vinogradov 2004; 2006; Pozzoli et al. 2007). Dans ce second modèle sur l'évolution des introns, différents éléments non-codants peuvent être retrouvés et affecter la régulation de l'expression du gène (Ugarkovic 2005; Feschotte 2008; Tomilin 2008). Cette section discutera des résultats en regard de ces différents modèles évolutifs.

4.4.1. Modèles de sélection par les pathogènes

Nous avons trouvé une association significative entre la présence de l'allèle *Safo-DAB*0101* dans un poisson et une résistance accrue à la furonculose (Tableaux 2.2, 2.3). Le modèle de la sélection dépendante de la fréquence peut potentiellement expliquer l'avantage de cet allèle par rapport

aux autres qui sont retrouvés à une plus grande fréquence dans la population. Ainsi, l'allèle *DAB*0101* a été retrouvé dans seulement 10 des 63 poissons (16 %) utilisés pour la création des familles, alors que l'allèle *DAB*0201* était largement distribué dans cette population (76 %) et a montré une tendance vers l'association à la susceptibilité pour la furonculose. Ceci souligne l'importance de ce mécanisme pour maintenir la variation des gènes MH chez l'omble de fontaine. Nous n'avons pas trouvé d'étude identifiant *Aeromonas salmonicida* comme un pathogène important retrouvé dans la rivière Rupert, ce qui aurait pu mener à une présélection pour des allèles MHII plus performants lors d'une infection. Il est ainsi raisonnable de supposer que la distribution des allèles retrouvés dans cette population d'omble de fontaine reflète l'adaptation de ces poissons aux pathogènes dominants qui sont présents dans cette rivière (Hedrick 2002). En effet, puisqu'il est connu que la diversité bactérienne représente un facteur important pour expliquer la variabilité au niveau des gènes MH dans les populations sauvages de salmonidés, les études à venir devraient combiner l'analyse de la communauté bactérienne présente dans l'environnement à l'échantillonnage des poissons, comme certains l'ont déjà fait (Dionne et al. 2007; Dionne et al. 2009).

Dans ce genre d'études sur une population, le type d'infection utilisée est un facteur important qui influence lequel des mécanismes de sélection sera observé. Des études utilisant un seul pathogène n'ont pu démontrer l'avantage hétérozygote, alors que d'autres appuient la sélection dépendante de la fréquence en rapportant un avantage pour les individus possédant un allèle rare (Langefors et al. 2001; Lohm et al. 2002; Trachtenberg et al. 2003; Ilmonen et al. 2007). Au sujet de l'avantage hétérozygote, une étude a examiné l'effet d'infections multiples chez la souris et révélé la supériorité des individus hétérozygotes par rapport aux homozygotes, en plus d'une réduction de la charge pathogénique (McClelland et al. 2003). Penn et collaborateurs (2002) ont également montré que des souris hétérozygotes étaient plus résistantes et plus en forme que les homozygotes lorsqu'elles étaient exposées à plusieurs souches pathogéniques différentes.

En général, les résultats de notre étude ne supportent pas l'avantage hétérozygote suite à l'infection avec un seul pathogène. Par exemple, la majorité des individus hétérozygotes affichent un facteur de risque élevé associé à une susceptibilité face à la furonculose, alors que trois des quatre génotypes homozygotes présentent un facteur de risque significatif ou montrant une tendance vers une résistance (Tableau 2.3). L'homozygotie des gènes MHII a déjà été suggérée comme un avantage contre la furonculose chez le saumon atlantique (Grimholt et al. 2003). Comme pour l'omble de fontaine dans cette étude, les poissons avaient été exposés à un seul pathogène alors qu'une infection multiple aurait pu révéler l'avantage hétérozygote, en assumant qu'ils peuvent présenter une plus grande variété de peptides antigéniques aux lymphocytes T CD4⁺ afin d'initier la réponse immunitaire spécifique (Hughes et Nei 1992). En conséquence, de futures études devraient inclure différents pathogènes et/ou souches dans des infections combinées.

4.4.2. Modèle du “*genomic design*” pour l'évolution du second intron

En plus du minisatellite au motif de 32 pb (type I) retrouvé dans le second intron du gène MHII β chez l'omble de fontaine et le saumon atlantique, des éléments additionnels ont été identifiés chez d'autres salmonidés par l'analyse des séquences du gène. Le premier élément est un second minisatellite composé d'un motif de 40 nucléotides (type II) qui est retrouvé exclusivement chez le saumon atlantique (Figure 3.1, Tableau 3.2, Figure Supplémentaire S3.2, rectangles gris et blanc). Ce minisatellite suit immédiatement le premier de type I et les 39 nucléotides qui assurent la transition entre les deux sont conservés à 90 % (35 des 39 nucléotides) par rapport à la séquence consensus du motif de type I. Ce minisatellite a été identifié chez un seul individu puisque l'autre séquence allélique du saumon atlantique est incomplète (*Sasa-DAB*0301*). L'accumulation d'éléments génétiques dans le second intron du gène MHII β des salmonidés, avec son expression qui est restreinte à des tissus ou cellules spécialisés (Ting et Trowsdale 2002) supportent l'hypothèse du modèle du “*genomic design*” pour l'évolution de cet intron (Pozzoli et al. 2007).

La séquence identifiée par la Box III dans la Figure Supplémentaire S3.2 est composée de 36 des 40 nucléotides du motif consensus du minisatellite de type II identifié chez le saumon atlantique. Ce minisatellite semble avoir évolué par l'amplification de cette séquence qui est également retrouvée chez l'omble de fontaine et les espèces de saumon du Pacifique. L'analyse des séquences introniques indiquerait qu'une séquence particulière présente en une seule copie, et rencontrant probablement des critères spécifiques, peut évoluer en un minisatellite dans certaines circonstances. Ce genre d'évènement se serait produit à deux reprises dans le second intron du gène MHII β des salmonidés. La première amplification du motif de 32 pb a probablement dû se produire avant la séparation de l'omble de fontaine et du saumon atlantique, en opposition aux espèces de saumon du Pacifique où on ne retrouve qu'une seule séquence. D'un autre côté, l'amplification du motif de 40 pb retrouvée seulement chez le saumon atlantique semble être apparue à la suite de la séparation du saumon atlantique d'un ancêtre commun avec l'omble de fontaine. L'amplification de ces séquences n'a pas été notée chez les espèces de saumon de Pacifique jusqu'à maintenant.

L'autre élément identifié dans le second intron du gène MHII β chez les salmonidés est un élément *Hpa* I SINE (Short Interspaced Nucleotide Element) (Kido et al. 1991). Comme décrit par Miller et Withler (1996), deux copies (une partielle et une complète) de l'élément sont insérées dans une orientation opposée de chaque côté du seul motif de 32 pb (type I) retrouvé chez les espèces de saumon du Pacifique. L'analyse des séquences en nucléotides qui constituent les Box III et Box IV, chez les espèces de saumon du Pacifique possédant l'élément *Hpa* I SINE complet (Onne-Wk136a et Onne-An13c; Figure Supplémentaire S3.2), révèle que l'insertion de l'élément a eu lieu dans une région où les séquences sont continues chez l'omble de fontaine et le saumon du Pacifique. Ceci indique que cet évènement s'est produit après la séparation des espèces de saumon du Pacifique de l'ancêtre commun avec saumon atlantique et l'omble de fontaine.

Les éléments SINE contiennent un promoteur interne pour l'ARN polymérase de type III et plusieurs sites potentiels pour la liaison de facteurs de transcription (Roy et al. 2000). Wang et

collaborateurs (2004) ont rapporté que l'évènement de rétrotransposition de l'élément SINE ne semble pas aléatoire au niveau de son orientation, avec tous les SINE connus à l'intérieur de gènes chez la truite arc-en-ciel qui sont retrouvés dans l'orientation opposée à celle de la transcription du gène. Comme ils l'ont suggéré, l'insertion dans chacune des orientations peut interférer avec l'expression du gène, l'insertion dans la même orientation que celle de la transcription produisant l'effet le plus prononcé. Puisque l'élément *Hpa* I SINE complet retrouvé dans le second intron du gène *MHII β* de l'omble de fontaine est dans la même orientation que celle de la transcription, ceci suggère que cet élément pourrait avoir une fonction régulatrice en modulant l'expression du gène au cours de périodes de stress ou lors d'infections. Il est intéressant de noter que l'effet de cet élément serait répressif, tout comme celui associé aux longs minisatellites présents chez l'omble de fontaine. De plus, l'insertion de deux *Hpa* I SINE dans le même gène est inattendu puisque les évènements d'insertion devraient être aléatoires (Kido et al. 1995). Leur identification dans le gène *MHII β* est intrigante et des études sur les mécanismes de rétrotransposition et leurs fonctions régulatrices devraient être tenues.

La présence de plusieurs éléments pour lesquels l'implication dans des évènements de recombinaison a été démontrée souligne probablement l'importance de ce type de mécanisme pour générer le polymorphisme du gène *MHII β* . La forte concentration et la variabilité des éléments génétiques retrouvés dans cet intron, tout juste en aval de l'exon 2 codant pour le domaine β 1 impliqué dans la liaison et la présentation des peptides antigéniques, pourraient refléter la pression sélective pour assurer le polymorphisme du gène *MHII β* . La forte incidence de ces éléments dans le second intron du gène chez les salmonidés pourrait être la conséquence du fait qu'il s'agit du seul locus MH de classe II identifié jusqu'à maintenant (Glamann, 1995, Langefors et al. 2000). La conversion génique a été identifiée comme un mécanisme permettant de générer le polymorphisme des gènes du CMH chez plusieurs espèces (Pease et al. 1983; Zangenberg et al. 1995). Il a également été démontré que le gène *Eb* chez la souris contient un site chaud de recombinaison (*hot spot*) situé dans le second intron. De plus, il est intéressant de constater qu'un segment de cet intron montre une grande similarité avec les

séquences LTR (*long terminal repeat*) et les gènes ENV et POL des rétrotransposons, suggérant qu'il aurait évolué par l'insertion de ce type d'élément (Zimmerer et Passmore 1991).

Chapitre 5

Conclusion

Depuis l'identification du premier gène majeur d'histocompatibilité (MH) chez un poisson téléostéen par Hashimoto et collaborateurs en 1990, d'autres ont été identifiés dans une variété d'espèces, dont plusieurs salmonidés, où un degré élevé de polymorphisme a été confirmé pour les loci de classe I et de classe II (Hordvik et al. 1993; Grimholt et al. 1993; 2000; Glamann, 1995; Miller et Withler, 1996; 1997). La variabilité des gènes du CMHII chez l'humain et la souris n'est pas retrouvée qu'au niveau de la séquence codante mais également au promoteur (Guardiola et al. 1996; Cowell et al. 1998), indiquant que la régulation de l'expression pourrait être aussi un facteur clé de l'adaptation, comme suggéré il y a plus de trente ans (King et Wilson, 1975). Plusieurs études ont également associé des allèles de gènes MH à la résistance ou la susceptibilité face à différents pathogènes chez le saumon atlantique (Langefors et al. 2001; Grimholt et al. 2003; de Eyto et al. 2007). Les travaux présentés dans le cadre de ce projet de recherche avaient pour objectifs de caractériser le polymorphisme du gène MHII β chez l'omble de fontaine (*Salvelinus fontinalis*) et de documenter ses effets au niveau de la survie conférée par des allèles et/ou génotypes particuliers lors d'une infection, ainsi que sur la variation du niveau d'expression du gène dans différentes conditions.

Dans la première étude, nous avons identifié 6 allèles du gène MHII β . L'analyse des séquences du domaine β 1 a permis de reconnaître l'effet d'une pression sélective positive pour maintenir le polymorphisme dans cette région de la molécule. Les résultats de l'expérience d'infection appuient l'hypothèse voulant que des allèles spécifiques du gène MHII β soient associés à la résistance ou la susceptibilité aux maladies. De plus, les résultats laissent entrevoir que ce niveau de résistance ou susceptibilité est affecté par des combinaisons particulières d'allèles, ce qui impliquerait également différents niveaux de dominance entre eux. Étant donné la faible fréquence à laquelle l'allèle *DAB*0101* a été trouvé dans la population, le modèle de la sélection dépendante de la fréquence peut potentiellement expliquer l'avantage qu'il procure et souligne l'importance de ce mécanisme pour le maintien du polymorphisme du gène MHII β chez l'omble de fontaine.

Dans la seconde étude, nous avons rapporté la présence d'un minisatellite polymorphique dans le second intron du gène *MHIIβ*. Nous avons démontré qu'un allèle du gène ayant un long minisatellite montre une réduction de l'expression par rapport à l'allèle qui en contient un court, suggérant l'implication d'un nouveau mécanisme pour réguler finement l'expression du gène *MHIIβ*, en plus du polymorphisme pouvant être retrouvé au niveau du promoteur. De plus, cet effet répressif associé au long minisatellite est amplifié à la température élevée de 18 °C par rapport à 6 °C, ce qui pourrait constituer un moyen supplémentaire permettant de réguler le niveau d'expression en réponse aux conditions environnementales. Ces résultats confirment que les séquences non-codantes participent à la régulation de l'activité transcriptionnelle des gènes. Ils suggèrent également qu'un minisatellite sensible aux différences de températures pourrait être soumis aux forces sélectives et jouer un rôle au niveau de l'expression de gènes et de l'évolution des organismes poïkilothermes.

Chapitre 6

Bibliographie

- Abendroth, A., Slobedman, B., Lee, E., Mellins, E., Wallace, M., Arvin, A.M. 2000. Modulation of major histocompatibility class II protein expression by varicella-zoster virus. *J. Virol.* 74:1900-7.
- Aguilar, A., Garza, J.C. 2007. Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. *J. Mol. Evol.* 65:34-43.
- Amarger, V., Gauguier, D., Yerle, M., Apiou, F., Pinton, P., Giraudeau, F., Monfouilloux, S., Lathrop, M., Dutrillaux, B., Buard, J., Vergnaud, G. 1998. Analysis of distribution in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* 52:62-71.
- Androlewicz, M.J., Anderson, K.S., Cresswell, P. 1993. Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. *Proc. Natl. Acad. Sci. U S A* 90:9130-4.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., *et al.* 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301-10.
- Appelgren, H., Cederberg, H., Rannug, U. 1997. Mutations at the human minisatellite MS32 integrated in yeast occur with high frequency in meiosis and involve complex recombination events. *Mol. Gen. Genet.* 256:7-17.
- Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D., Hedrick, P.W. 2002. Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Can. J. Fish. Aquat. Sci.* 59:966-975.
- Asghari, V., Sanyal, S., Buchwaldt, S., Paterson, A., Jovanovic, V., Van Tol, H.H. 1995. Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *J. Neurochem.* 65:1157-65.
- Bacolla, A., Wells, R.D. 2004. Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* 279:47411-4.
- Bacon, L.D., Hunt, H.D., Cheng, H.H. 2000. A review of the development of chicken lines to resolve genes determining resistance to diseases. *Poult. Sci.* 79:1082-93.
- Bailey, J.R., Williams, T.M., Siliciano, R.F., Blankson, J.N. 2006. Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. *J. Exp. Med.* 203:1357-69.

- Baran, N., Lapidot, A., Manor, H. 1991. Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)_n and d(GA)_n tracts. *Proc. Natl. Acad. Sci. U S A* 88:507-11.
- Batalia, M.A., Collins, E.J. 1997. Peptide binding by class I and class II MHC molecules. *Biopolymers* 43:281-302.
- Batzer, M.A., Deininger, P.L. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3:370-9.
- Bartl, S., Weissman, I.L. 1994. Isolation and characterization of major histocompatibility complex class IIB genes from the nurse shark. *Proc. Natl. Acad. Sci. U S A* 91:262-6.
- Bartl, S., Baish, M.A., Flajnik, M.F., Ohta, Y. 1997. Identification of class I genes in cartilaginous fish, the most ancient group of vertebrates displaying an adaptive immune response. *J. Immunol.* 159:6097-104.
- Beaty, J.S., Sukiennicki, T.L., Nepom, G.T. 1999. Allelic variation in transcription modulates MHC class II expression and function. *Microbes Infect.* 1:919-27.
- Beauregard, A., Curcio, M.J., Belfort, M. 2008. The take and give between retrotransposable elements and their hosts. *Annu. Rev. Genet.* 42:587-617.
- Belancio, V.P., Hedges, D.J., Deininger, P. 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 18:343-58.
- Bennett, S.T., Lucassen, A.M., Gough, S.C., Powell, E.E., Undlien, D.E., Pritchard, L.E., Merriman, M.E., Kawaguchi, Y., Dronsfield, M.J., Pociot, F., *et al.* 1995. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.* 9:284-292.
- Bennett, S.T., Wilson, A.J., Esposito, L., Bouzekri, N., Undlien, D.E., Cucca, F., Nisticò, L., Buzzetti, R., Bosi, E., Pociot, F., *et al.* 1997. Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. *Nat. Genet.* 17:350-2.
- Bernatchez, L., Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* 16:363-77.
- Bernoth, E.M., Ellis, A.E., Midtlyng, P.J., Olivier, G., Smith, P. (Editors) 1997. *Furonculosis: multidisciplinary fish disease research*. San Diego, CA, USA: Academic Press.
- Bishop, A.J., Schiestl, R.H. 2000. Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum. Mol. Genet.* 9:2427-334.

- Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W.S., Johnson, S.L., Gates, M., Postlethwait, J.H., Klein, J. 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46:129-34.
- Blackwell, J.M., Jamieson, S.E., Burgner, D. 2009. HLA and infectious diseases. *Clin. Microbiol. Rev.* 22:370-85.
- Bly, J.E., Clem, L.W. 1991. Temperature-mediated processes in teleost immunity: in vitro immunosuppression induced by in vivo low temperature in channel catfish. *Vet. Immunol. Immunopathol.* 28:365-377.
- Bois, P., Jeffreys, A.J. 1999. Minisatellite instability and germline mutation. *Cell Mol Life Sci.* 55:1636-48.
- Bois, P.R. 2003. Hypermutable minisatellites, a human affair? *Genomics* 81:349-55.
- Borstnik, B., Pumpernik, D. 2002. Tandem repeats in protein coding regions of primate genes. *Genome Res.* 12:909-15.
- Britten, R.J., Kohne, D.E. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161:529-40.
- Britten, R.J., Davidson, E.H. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46:111-38.
- Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238:115-34.
- Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33-9.
- Brown, D.R., Qin, K., Herms, J.W., Madlung, A., Manson, J., Strome, R., Fraser, P.E., Kruck, T., von Bohlen, A., Schulz-Schaeffer, W., *et al.* 1997. The cellular prion protein binds copper in vivo. *Nature* 390:684-7.
- Buard, J., Vergnaud, G. 1994. Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* 13:3203-10.
- Burwinkel, B., Kilimann, M.W. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol.* 277:513-7.

- Buschiazzo, E., Gemmell, N.J. 2006. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* 28:1040-50.
- Buzdin, A.A. 2004. Retroelements and formation of chimeric retrogenes. *Cell. Mol. Life Sci.* 61:2046-59.
- Caillat-Zucman, S. 2009. Molecular mechanisms of HLA association with autoimmune diseases. *Tissue Antigens* 73:1-8.
- Canaday, D.H., Ziebold, C., Noss, E.H., Chervenak, K.A., Harding, C.V., Boom, W.H. 1999. Activation of human CD8+ alpha beta TCR+ cells by Mycobacterium tuberculosis via an alternate class I MHC antigen-processing pathway. *J. Immunol.* 162:372-9.
- Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, F.A. 2002. Selection for short introns in highly expressed genes. *Nat. Genet.* 31:415-418.
- Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K., O'Brien, S.J. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283:1748-1752.
- Catasti, P., Chen, X., Moyzis, R.K., Bradbury, E.M., Gupta, G. 1996. Structure-function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.* 264:534-45.
- Cederberg, H., Agurell, E., Hedenskog, M., Rannug, U. 1993. Amplification and loss of repeat units of the human minisatellite MS1 integrated in chromosome III of a haploid yeast strain. *Mol. Gen. Genet.* 238:38-42.
- Cereghetti, G.M., Schweiger, A., Glockshuber, R., Van Doorslaer, S. 2003. Stability and Cu(II) binding of prion protein variants related to inherited human prion diseases. *Biophys. J.* 84:1985-97.
- Charbonnel, N., Pemberton, J. 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity* 95:377-88.
- Charlesworth, B., Sniegowski, P., Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215-20.
- Clarke, B., Kirby, D.R. 1966. Maintenance of histocompatibility polymorphisms. *Nature* 211:999-1000.
- Cochran, E.J., Bennett, D.A., Cervenáková, L., Kenney, K., Bernard, B., Foster, N.L., Benson, D.F., Goldfarb, L.G., Brown, P. 1996. Familial Creutzfeldt-Jakob disease with a five-repeat octapeptide insert mutation. *Neurology* 47:727-33.

- Collins, M., Rubin, G.M. 1984. Structure of chromosomal rearrangements induced by the FB transposable element in *Drosophila*. *Nature* 308:323-7.
- Cowell, L.G., Kepler, T.B., Janitz, M., Lauster, R., Mitchison, N.A. 1998. The distribution of variation in regulatory gene segments, as present in MHC class II promoters. *Genome Res.* Feb:124-34.
- Cowles, C.R., Hirschhorn, J.N., Altshuler, D., Lander, E.S. 2002. Detection of regulatory variation in mouse genes. *Nat. Genet.* 32:432-7.
- Crowe, P.D., Boehme, S.A., Wong, T., Gaur, A., Sidney, J., Sette, A., Conlon, P.J. 1998. Differential signaling and hierarchical response thresholds induced by an immunodominant peptide of myelin basic protein and an altered peptide ligand in human T cells. *Hum. Immunol.* 59:679-89.
- Cullell-Young, M., Barrachina, M., López-López, C., Goñalons, E., Lloberas, J., Soler, C., Celada, A. 2001. From transcription to cell surface expression, the induction of MHC class II I-A alpha by interferon-gamma in macrophages is regulated at different levels. *Immunogenetics* 53:136-44.
- Daly, J.G., Kew, A.K., Moore, A.R., Olivier, G. 1996. The cell surface of *Aeromonas salmonicida* determines in vitro survival in cultured brook trout (*Salvelinus fontinalis*) peritoneal macrophages. *Microb. Pathog.* 21:447-61.
- Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., Darnell, R.B. 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 107:489-99.
- Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., Housman, D.E. 1997. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. U S A* 94:7388-93.
- Debrauwere, H., Gendrel, C.G., Lechat, S., Dutreix, M. 1997. Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. *Biochimie* 79:577-86.
- Debrauwère, H., Buard, J., Tessier, J., Aubert, D., Vergnaud, G., Nicolas, A. 1999. Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat. Genet.* 23:367-71.
- de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H.J., Jordan, W., Cross, T., Stet, R.J. 2007. Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. *Proc. Biol. Sci.* 274:861-9.
- de Waal Malefyt, R., Haanen, J., Spits, H., Roncarolo, M.G., te Velde, A., Figdor, C., Johnson, K., Kastelein, R., Yssel, H., de Vries, J.E. 1991. Interleukin 10 (IL-10) and viral IL-10 strongly reduce antigen-specific human T cell proliferation by diminishing the antigen-presenting capacity of

- monocytes via downregulation of class II major histocompatibility complex expression. *J. Exp. Med.* 174:915-24.
- Delamarre, L., Holcombe, H., Mellman, I. 2003. Presentation of exogenous antigens on major histocompatibility complex (MHC) class I and MHC class II molecules is differentially regulated during dendritic cell maturation. *J. Exp. Med.* 198:111-22.
- Denoeud, F., Vergnaud, G., Benson, G. 2003. Predicting human minisatellite polymorphism. *Genome Res.* 13:856-67.
- Dere, R., Napierala, M., Ranum, L.P., Wells, R.D. 2004. Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. *J. Biol. Chem.* 279:41715-26.
- Dessen, A., Lawrence, C.M., Cupo, S., Zaller, D.M., Wiley, D.C. 1997. X-ray crystal structure of HLA-DR4 (DRA*0101, DRB1*0401) complexed with a peptide from human collagen II. *Immunity* 7:473-81.
- Desseyn, J.L., Tetaert, D., Gouyer, V. 2008. Architecture of the large membrane-bound mucins. *Gene* 410:215-22.
- Dewannieux, M., Esnault, C., Heidmann, T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35:41-8.
- DiMolfetto, L., Neal, H.A., Wu, A., Reilly, C., Lo, D. 1998. The density of the class II MHC T cell receptor ligand influences IFN-gamma/IL-4 ratios in immune responses in vivo. *Cell. Immunol.* 18370-9.
- Ding, Y.C., Chi, H.C., Grady, D.L., Morishima, A., Kidd, J.R., Kidd, K.K., Flodman, P., Spence, M.A., Schuck, S., Swanson, J.M., *et al.* 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc. Natl. Acad. Sci. U S A* 99:309-14.
- Dionne, M., Miller, K.M., Dodson, J.J., Caron, F., Bernatchez, L. 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution* 61:2154-64.
- Dionne, M., Miller, K.M., Dodson, J.J., Bernatchez, L. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:1555-65.
- Dixon, B., van Erp, S.H., Rodrigues, P.N., Egberts, E., Stet, R.J. 1995. Fish major histocompatibility complex genes: an expansion. *Dev. Comp. Immunol.* 19:109-33.

- Dixon, B., Stet, R.J. 2001. The relationship between major histocompatibility receptors and innate immunity in teleost fish. *Dev. Comp. Immunol.* 25:683-99.
- Doherty, P.C., Zinkernagel, R.M. 1975 Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256:50-2.
- Dombrowski, C., Lévesque, S., Morel, M.L., Rouillard, P., Morgan, K., Rousseau, F. 2002. Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* 11:371-8.
- Dongre, A.R., Kovats, S., deRoos, P., McCormack, A.L., Nakagawa, T., Paharkova-Vatchkova, V., Eng, J., Caldwell, H., Yates, J.R. 3rd, Rudensky, A.Y. 2001. In vivo MHC class II presentation of cytosolic proteins revealed by rapid automated tandem mass spectrometry and functional analyses. *Eur. J. Immunol.* 31:1485-94.
- Doudna, J.A., Batey, R.T. 2004. Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* 73:539-557.
- Drake, J.W., Charlesworth, B., Charlesworth, D., Crow, J.F. 1998. Rates of spontaneous mutation. *Genetics* 148:1667-86.
- D'Souza, U.M., Russ, C., Tahir, E., Mill, J., McGuffin, P., Asherson, P.J., Craig, I.W. 2004. Functional effects of a tandem duplication polymorphism in the 5'flanking region of the DRD4 gene. *Biol. Psychiatry* 56:691-7.
- Dubrova, Y.E., Nesterov, V.N., Krouchinsky, N.G., Ostapenko, V.A., Neumann, R., Neil, D.L., Jeffreys, A.J. 1996. Human minisatellite mutation rate after the Chernobyl accident. *Nature* 380:683-6.
- Ecker, M., Mrsa, V., Hagen, I., Deutzmann, R., Strahl, S., Tanner, W. 2003. O-mannosylation precedes and potentially controls the N-glycosylation of a yeast cell wall glycoprotein. *EMBO Rep.* 4:628-32.
- Edwards, S.V., Grahn, M., Potts, W.K. 1995. Dynamics of Mhc evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Mol. Ecol.* 4:719-29.
- Egid, K., Brown, J. L. 1989. The major histocompatibility complex and female mating preferences in mice. *Animal Behaviour* 38:548-550.
- Eisenberg, E., Levanon, E.Y. 2003. Human housekeeping genes are compact. *Trends Genet.* 19:362-365.

- Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16:551-8.
- Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5:435-45.
- Fabre, E., Dujon, B., Richard, G.F. 2002. Transcription and nuclear transport of CAG/CTG trinucleotide repeats in yeast. *Nucleic Acids Res.* 30:3540-7.
- Fan, J.B., Sklar, P. 2005. Meta-analysis reveals association between serotonin transporter gene STin2 VNTR polymorphism and schizophrenia. *Mol. Psychiatry* 10:928-38, 891.
- Ferbeyre, G., Smith, J.M., Cedergren, R. 1998. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* 18:3880-8.
- Feschotte, C., Pritham, E.J. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41:331-68.
- Fidalgo, M., Barrales, R.R., Ibeas, J.I., Jimenez, J. 2006. Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U S A* 103:11228-33.
- Fitzgerald, D.J., Dryden, G.L., Bronson, E.C., Williams, J.S., Anderson, J.N. 1994 Conserved patterns of bending in satellite and nucleosome positioning DNA. *J. Biol. Chem.* 269:21303-14.
- Flajnik, M.F., Ohta, Y., Namikawa-Yamada, C., Nonaka, M. 1999 Insight into the primordial MHC from studies in ectothermic vertebrates. *Immunol. Rev.* 167:59-67.
- Flajnik, M.F., Kasahara, M. 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 15:351-62.
- Fojtík, P., Vorlícková, M. 2001. The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res.* 29:4684-90.
- Fondon, J.W. 3rd, Garner, H.R. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U S A* 101:18058-63.
- Fondon, J.W. 3rd, Hammock, E.A., Hannan, A.J., King, D.G. 2008. Simple sequence repeats: genetic modulators of brain function and behavior. *Trends Neurosci.* 31:328-34.
- Freeman-Gallant, C.R., Johnson, E.M., Saponara, F., Stanger, M. 2002. Variation at the major histocompatibility complex in Savannah sparrows. *Mol. Ecol.* 11:1125-30.
- Freudenreich, C.H., Kantrow, S.M., Zakian, V.A. 1998. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279:853-6.

- Freund, A.M., Bichara, M., Fuchs, R.P. 1989. Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc. Natl. Acad. Sci. U S A* 86:7465-9.
- Frith, M.C., Li, M.C., Weng, Z. 2003. Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* 31:3666-8. Site internet : <http://zlab.bu.edu/cluster-buster/>
- Fryer, J.L., Pilcher, K.S., Sanders, J.E., Rohovec, J.S., Zinn, J.L., Groberg, W.J., McCoy, R.H. 1976. Temperature, infectious disease, and the immune response in salmonid fish. *NTIS* 191:72.
- Fukui, Y., Ishimoto, T., Utsuyama, M., Gytoku, T., Koga, T., Nakao, K., Hirokawa, K., Katsuki, M., Sasazuki, T. 1997. Positive and negative CD4+ thymocyte selection by a single MHC class II/peptide ligand affected by its expression level in the thymus. *Immunity* 6:401-10.
- Galvão, R., Mendes-Soares, L., Câmara, J., Jaco, I., Carmo-Fonseca, M. 2001. Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis. *Brain. Res. Bull.* 56:191-201.
- Garduño, R.A., Moore, A.R., Olivier, G., Lizama, A.L., Garduño, E., Kay, W.W. 2000, Host cell invasion and intracellular residence by *Aeromonas salmonicida*: role of the S-layer. *Can. J. Microbiol.* 46:660-8.
- Génin, E., Ober, C., Weitkamp, L., Thomson, G. 2000. A robust test for assortative mating. *Eur. J. Hum. Genet.* 8:119-24.
- Glamann, J. 1995. Complete coding sequence of rainbow trout Mhc II beta chain. *Scand. J. Immunol.* 41:365-72.
- Glickman, M.H., Ciechanover, A. 2002. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* 82:373-428.
- Glimcher, L.H., Kara, C.J. 1992. Sequences and factors: a guide to MHC class-II transcription. *Annu. Rev. Immunol.* 10:13-49.
- Goellner, G.M., Tester, D., Thibodeau, S., Almqvist, E., Goldberg, Y.P., Hayden, M.R., McMurray, C.T. 1997. Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *Am. J. Hum. Genet.* 60:879-90.
- Goldfarb, L.G., Brown, P., McCombie, W.R., Goldgaber, D., Swergold, G.D., Wills, P.R., Cervenakova, L., Baron, H., Gibbs, C.J. Jr, Gajdusek, D.C. 1991. Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. *Proc. Natl. Acad. Sci. U S A* 88:10926-30.

- Gorer, P.A. 1937. The genetic and antigenetic basis of tumor transplantation. *J. Pathol. Bacteriol.* 44:691-697.
- Grabczyk, E., Usdin, K. 2000. Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. *Nucleic Acids Res.* 28:4930-7.
- Greer, S.F., Zika, E., Conti, B., Zhu, X.S., Ting, J.P. 2003. Enhancement of CIITA transcriptional function by ubiquitin. *Nat. Immunol.* 4:1074-82.
- Greer, S.F., Harton, J.A., Linhoff, M.W., Janczak, C.A., Ting, J.P., Cressman, D.E. 2004. Serine residues 286, 288, and 293 within the CIITA: a mechanism for down-regulating CIITA activity through phosphorylation. *J. Immunol.* 173:376-83.
- Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C., Lie, O. 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics* 37:469-73.
- Grimholt, U. 1997. Transport-associated proteins in Atlantic salmon (*Salmo salar*). *Immunogenetics* 46:213-21.
- Grimholt, U., Getahun, A., Hermsen, T., Stet, R.J. 2000. The major histocompatibility class II alpha chain in salmonid fishes. *Dev. Comp. Immunol.* 24:751-63.
- Grimholt, U., Drabløs, F., Jørgensen, S.M., Høyheim, B., Stet, R.J. 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* 54:570-81.
- Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., Stet, R.J. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55:210-9.
- Grossberger, D., Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics* 36:166-74.
- Groothuis, T.A., Griekspoor, A.C., Neijssen, J.J., Herberts, C.A., Neeffjes, J.J. 2005. MHC class I alleles and their exploration of the antigen-processing machinery. *Immunol. Rev.* 207:60-76.
- Guardiola, J., Maffei, A., Lauster, R., Mitchison, N.A., Accolla, R.S., Sartoris, S. 1996. Functional significance of polymorphism among MHC class II gene promoters. *Tissue Antigens.* 48:615-25.

- Guindalini, C., Howard, M., Haddley, K., Laranjeira, R., Collier, D., Ammar, N., Craig, I., O'Gara, C., Bubb, V.J., Greenwood, T., *et al.* 2006. A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. *Proc. Natl. Acad. Sci. U S A* 103:4552-7.
- Haber, J.E., Louis, E.J. 1998. Minisatellite origins in yeast and humans. *Genomics* 48:132-5.
- Hashimoto, K., Nakanishi, T., Kurosawa, Y. 1990. Isolation of carp genes encoding major histocompatibility complex antigens. *Proc. Natl. Acad. Sci. U. S. A.* 87:6863-7.
- Hammock, E.A., Young, L.J. 2004. Functional microsatellite polymorphism associated with divergent social structure in vole species. *Mol. Biol. Evol.* 21:1057-63.
- Hammock, E.A., Young, L.J. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* 308:1630-4.
- Hansen, J.D., Strassburger, P., Du Pasquier, L. 1996. Conservation of an alpha 2 domain within the teleostean world, MHC class I from the rainbow trout *Oncorhynchus mykiss*. *Dev. Comp. Immunol.* 20:417-25.
- Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P., Du Pasquier, L. 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. *J. Immunol.* 163:774-86.
- Havecker, E.R., Gao, X., Voytas, D.F. 2004. The diversity of LTR retrotransposons. *Genome Biol.* 5:225.
- Hedrick, P.W., Thompson, G. 1987. Evolutionary genetics and HLA: another classical example. *Biol. J. Linn. Soc.* 31:311-331.
- Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* 56:1902-8.
- Henikoff, S., Ahmad, K., Malik, H.S. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293:1098-102.
- Hennecke, J., Carfi, A., Wiley, D.C. 2000. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J.* 19:5611-5624.
- Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J., Greenwood, B.M. 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352:595-600.

- Hordvik, I., Grimholt, U., Fosse, V.M., Lie, O., Endresen, C. 1993. Cloning and sequence analysis of cDNAs encoding the MHC class II beta chain in Atlantic salmon (*Salmo salar*). *Immunogenetics* 37:437-41.
- Howell, R.M., Woodford, K.J., Weitzmann, M.N., Usdin, K. 1996. The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *J. Biol. Chem.* 271:5208-14.
- Hughes, A.L., Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167-70.
- Hughes, A.L., Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U S A* 86:958-62.
- Hughes, A.L., Nei, M. 1992. Models of host-parasite interaction and MHC polymorphism. *Genetics* 132, 863–864.
- IHGSC; International human genome sequencing consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- Ilmonen, P., Penn, D.J., Damjanovich, K., Morrison, L., Ghotbi, L., Potts, W.K. 2007. Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics* 176:2501-2508.
- Itano, A.A., McSorley, S.J., Reinhardt, R.L., Ehst, B.D., Ingulli, E., Rudensky, A.Y., Jenkins, M.K. 2003. Distinct dendritic cell populations sequentially present antigen to CD4 T cells and stimulate different aspects of cell-mediated immunity. *Immunity* 19:47-57.
- Itano, A.A., Jenkins, M.K. 2003. Antigen presentation to naive CD4 T cells in the lymph node. *Nat. Immunol.* 4:733-9.
- Janeway, C.A.Jr., Travers, P., Walport, M. and Shlomchik, M.J. 2001. *Immunobiology : the immune system in health and disease, 5th ed.*, Garland Publishing, pp. 167-181. New York.
- Jankowski, C., Nasar, F., Nag, D.K. 2000. Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc. Natl. Acad. Sci. U S A* 97:2134-9.
- Jankowski, C., Nag, D.K. 2002. Most meiotic CAG repeat tract-length alterations in yeast are SPO11 dependent. *Mol. Genet. Genomics* 267:64-70.
- Jauert, P.A., Edmiston, S.N., Conway, K., Kirkpatrick, D.T. 2002. RAD1 controls the meiotic expansion of the human HRAS1 minisatellite in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 22:953-64.

- Jeffreys, A.J., Wilson, V., Thein, S.L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-73.
- Jeffreys, A.J., Royle, N.J., Wilson, V., Wong, Z. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332:278-81.
- Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., Armour, J.A. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* 6:136-45.
- Jeffreys, A.J., Neumann, R. 1997. Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.* 6:129-32; 134-6.
- Jeffreys, A.J., Murray, J., Neumann, R. 1998. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell.* 2:267-73.
- Jeffreys, A.J., Holloway, J.K., Kauppi, L., May, C.A., Neumann, R., Slingsby, M.T., Webb, A.J. 2004. Meiotic recombination hot spots and human DNA diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359:141-152.
- Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., Wessler, S.R. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431:569-73
- Jin, P., Warren, S.T. 2000. Understanding the molecular basis of fragile X syndrome. *Hum. Mol. Genet.* 9:901-8.
- Jones, E.Y., Fugger, L., Strominger, J.L., Siebold, C. 2006. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* 6:271-82.
- Jordan, W.C., Bruford, M.W. 1998. New perspectives on mate choice and the MHC. *Heredity* 81:127-33.
- Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. U S A* 94:1872-7.
- Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V. 2007. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8:241-59.
- Juul-Madsen, H.R., Glamann, J., Madsen, H.O., Simonsen, M. 1992. MHC class II beta-chain expression in the rainbow trout. *Scand. J. Immunol.* 35:687-94.
- Kales, S., Parks-Dely, J., Schulte, P., Dixon, B. 2006. Beta-2-microglobulin gene expression is maintained in rainbow trout and Atlantic salmon kept at low temperatures. *Fish Shellfish Immunol.* 21:176-186.

- Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M., Fry, M. 2001. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)_n trinucleotide repeat sequence. *J. Biol. Chem.* 276:16439-46.
- Kandil, E., Namikawa, C., Nonaka, M., Greenberg, A.S., Flajnik, M.F., Ishibashi, T., Kasahara, M. 1996. Isolation of low molecular mass polypeptide complementary DNA clones from primitive vertebrates. Implications for the origin of MHC class I-restricted antigen presentation. *J. Immunol.* 156:4245-53.
- Kang, S., Jaworski, A., Ohshima, K., Wells, R.D. 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat. Genet.* 10:213-8.
- Kapitonov, V.V., Jurka, J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* 3:e181.
- Kasahara, M., Vazquez, M., Sato, K., McKinney, E.C., Flajnik, M.F. 1992. Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the cartilaginous fish. *Proc. Natl. Acad. Sci. U S A* 89:6688-92.
- Kasahara, M., McKinney, E.C., Flajnik, M.F., Ishibashi, T. 1993. The evolutionary origin of the major histocompatibility complex: polymorphism of class II alpha chain genes in the cartilaginous fish. *Eur. J. Immunol.* 23:2160-5.
- Kaufman, J., Völk, H., Wallny, H.J. 1995. A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol. Rev.* 143:63-88.
- Kaufman, J., Milne, S., Göbel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., Beck, S. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401:923-5.
- Kaufman, J. 2000. The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355:1077-84.
- Kaufmann, S.H., Schaible, U.E. 2005. Antigen presentation and recognition in bacterial infections. *Curr. Opin. Immunol.* 17:79-87.
- Kazazian, H.H. Jr. 2000. Genetics. L1 retrotransposons shape the mammalian genome. *Science* 289:1152-3.
- Kazazian, H.H. Jr. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626-1632.
- Kennedy, G.C., German, M.S., Rutter, W.J. 1995. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet.* 9:293-8.

- Kenneson, A., Zhang, F., Hagedorn, C.H., Warren, S.T. 2001. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.* 10:1449-54.
- Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., Okada, N. 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. *Proc. Natl. Acad. Sci. U. S. A.* 88:2326-2330.
- Kido, Y., Saitoh, M., Murata, S., Okada, N. 1995. Evolution of the active sequences of the HpaI short interspersed elements. *J. Mol. Evol.* 41:986-995.
- Kielar, M.L., Sicher, S.C., Penfield, J.G., Jeyarajah, D.R., Lu, C.Y. 2000. Nitric oxide inhibits INFgamma-induced increases in CIITA mRNA abundance and activation of CIITA dependent genes--class II MHC, Ii and H-2M. Class II TransActivator. *Inflammation* 24:431-45.
- King, M.C., Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107-116.
- Kjøglum, S., Larsen, S., Bakke, H.G., Grimholt, U. 2006. How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol.* 21:431-41.
- Kjøglum, S., Larsen, S., Bakke, H.G., Grimholt, U. 2008. The effect of specific MHC class I and class II combinations on resistance to furunculosis in Atlantic salmon (*Salmo salar*). *Scand. J. Immunol.* 67:160-8.
- Klein, J. *The natural history of the major histocompatibility complex*. New-York : John Wiley; 1986.
- Klein, J., Sato, A., Nagl, S., O'hUigin, C. 1998. Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* 29:1-21.
- Klein, J., O'hUigin, C. 1994. The conundrum of nonclassical major histocompatibility complex genes. *Proc. Natl. Acad. Sci. U S A* 91:6251-2.
- Klein, J., Sato, A. 2000. The HLA system. First of two parts. *N. Engl. J. Med.* 343:702-9.
- Kleinjan, D.A., van Heyningen, V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* 76:8-32.
- Klenova, E., Scott, A.C., Roberts, J., Shamsuddin, S., Lovejoy, E.A., Bergmann, S., Bubb, V.J., Royer, H.D., Quinn, J.P. 2004. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J. Neurosci.* 24:5966-73.

- Knafo, A., Israel, S., Darvasi, A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., Feldman, E., Lerer, E., Laiba, E., Raz, Y., *et al.* 2007. Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. *Genes Brain Behav.* 7:266-275
- Knight, J., Stet, R.J.M., Secombes, C. 1998, Modulation of MHC class II expression in rainbow trout *Oncorhynchus mykiss* macrophages by TNF α and LPS. *Fish Shellfish Immunol.* 8:545-553.
- Kollner, B., Wasserrab, B., Kotterba, G., Fischer, U. 2002. Evaluation of immune functions of rainbow trout (*Oncorhynchus mykiss*) - how can environmental influences be detected? *Toxicol. Lett.* 131:83-95.
- Koppang, E.O., Lundin, M., McL. Press, C., Ronningen, K., Lie, Ø. 1998. Differing levels of Mhc class II β chain expression in a range of tissues from vaccinated and non-vaccinated Atlantic salmon (*Salmo salar* L.). *Fish Shellfish Immunol.* 8:183-196.
- Kokoska, R.J., Stefanovic, L., Tran, H.T., Resnick, M.A., Gordenin, D.A., Petes, T.D. 1998. Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). *Mol. Cell. Biol.* 18:2779-88.
- Kurahashi, H., Emanuel, B.S. 2001. Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum. Mol. Genet.* 10:2605-17.
- Kurahashi, H., Inagaki, H., Yamada, K., Ohye, T., Taniguchi, M., Emanuel, B.S., Toda, T. 2004. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. *J. Biol. Chem.* 279:35377-83.
- Kwak, B., Mulhaupt, F., Myit, S., Mach, F. 2000. Statins as a newly recognized type of immunomodulator. *Nat. Med.* 6:1399-402.
- Kyriacou, C.P., Peixoto, A.A., Sandrelli, F., Costa, R., Tauber, E. 2008. Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends Genet.* 24:124-32.
- Lafrenière RG, Rochefort DL, Chrétien N, Rommens JM, Cochiu JI, Kälviäinen R, Nousiainen U, Patry G, Farrell K, Söderfeldt B, *et al.* 1997. Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat. Genet.* 15:298-302.

- Lalioi, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A., Antonarakis, S.E. 1997. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386:847-851.
- Landmann, S., Mühlethaler-Mottet, A., Bernasconi, L., Suter, T., Waldburger, J.M., Masternak, K., Arrighi, J.F., Hauser, C., Fontana, A., Reith, W. 2001. Maturation of dendritic cells is accompanied by rapid transcriptional silencing of class II transactivator (CIITA) expression. *J. Exp. Med.* 194:379-91.
- Landry, C., Bernatchez, L. 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Mol. Ecol.* 10:2525-39.
- Landry, C., Garant, D., Duchesne, P., Bernatchez, L. 2001. 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc. Biol. Sci.* 268:1279-85.
- Langefors, A., Lohm, J., Von Schantz, T., Grahn, M. 2000. Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. *Mol. Ecol.* 9:215-219.
- Langefors, A., Lohm, J., Grahn, M., Andersen, O., von Schantz, T. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. Biol. Sci.* 268:479-85.
- Laplanche, J.L., Hachimi, K.H., Durieux, I., Thuillet, P., Defebvre, L., Delasnerie-Lauprêtre, N., Peoc'h, K., Foncin, J.F., Destée, A. 1999. Prominent psychiatric features and early onset in an inherited prion disease with a new insertional mutation in the prion protein gene. *Brain* 122:2375-86.
- Lee, Y.J., Han, Y., Lu, H.T., Nguyen, V., Qin, H., Howe, P.H., Hocevar, B.A., Boss, J.M., Ransohoff, R.M., Benveniste, E.N. 1997. TGF-beta suppresses IFN-gamma induction of class II MHC gene expression by inhibiting class II transactivator messenger RNA expression. *J. Immunol.* 158:2065-75.
- LeibundGut-Landmann, S., Waldburger, J.M., Krawczyk, M., Otten, L.A., Suter, T., Fontana, A., Acha-Orbea, H., Reith, W. 2004. Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. *Eur. J. Immunol.* 34:1513-25.

- Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., Benjamin, J., Müller, C.R., Hamer, D.H., Murphy, D.L. 1996. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274:1527-31.
- Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5:e254.
- Li, Y.C., Korol, A.B., Fahima, T., Nevo, E. 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21:991-1007.
- Li, Y.X., Kirby, M.L. 2003. Coordinated and conserved expression of aliphoid repeat and aliphoid repeat-tagged coding sequences. *Dev. Dyn.* 228:72-81.
- Lindsey, J.C., Leach, D.R. 1989. Slow replication of palindrome-containing DNA. *J. Mol. Biol.* 206:779-82.
- Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., Ranum, L.P. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293:864-7.
- Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P.O., Zhao, K. 2001. Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* 106:309-18.
- Liu, Y., Kasahara, M., Rumfelt, L.L., Flajnik, M.F. 2002. Xenopus class II A genes: studies of genetics, polymorphism, and expression. *Dev. Comp. Immunol.* 26:735-50.
- Lobachev, K.S., Stenger, J.E., Kozyreva, O.G., Jurka, J., Gordenin, D.A., Resnick, M.A. 2000. Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* 19:3822-30.
- Lohm, J., Grahn, M., Langefors, A., Andersen, Ø., Storset, A., von Schantz, T. 2002. Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc. Biol. Sci.* 269:2029-33.
- Lopes, J., Debrauwère, H., Buard, J., Nicolas, A. 2002. Instability of the human minisatellite CEB1 in rad27Delta and dna2-1 replication-deficient yeast cells. *EMBO J.* 21:3201-11.
- Louis, E.J., Naumova, E.S., Lee, A., Naumov, G., Haber, J.E. 1994. The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* 136:789-802.
- Lutwyche, P., Exner, M.M., Hancock, R.E., Trust, T.J. 1995. A conserved *Aeromonas salmonicida* porin provides protective immunity to rainbow trout. *Infect. Immun.* 63:3137-3142.

- MacKenzie, A., Quinn, J. 1999. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proc. Natl. Acad. Sci. U S A* 96:15251-5.
- Maffei, A., Perfetto, C., Ombra, N., Del Pozzo, G., Guardiola, J. 1989. Transcriptional and post-transcriptional regulation of human MHC class II genes require the synthesis of short-lived proteins. *J. Immunol.* 142:3657-61.
- Maleki, S., Cederberg, H., Rannug, U. 2002. The human minisatellites MS1, MS32, MS205 and CEB1 integrated into the yeast genome exhibit different degrees of mitotic instability but are all stabilised by RAD27. *Curr. Genet.* 41:333-41.
- Mar Albà, M., Santibáñez-Koref, M.F., Hancock, J.M. 1999. Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *J. Mol. Evol.* 49:789-97.
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., Okazaki, T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell. Biol.* 109:1963-73.
- Maxam, A.M., Gilbert, W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U S A* 74:560-4.
- McClelland, E.E., Penn, D.J., Potts, W.K. 2003. Major histocompatibility complex heterozygote superiority during coinfection. *Infect. Immun.* 71:2079-2086.
- Mellman, I., Turley, S.J., Steinman, R.M. 1998. Antigen processing for amateurs and professionals. *Trends Cell. Biol.* 8:231-7.
- Meloni, R., Albanèse, V., Ravassard, P., Treilhou, F., Mallet, J. 1998. A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. *Hum. Mol. Genet.* 7:423-8.
- Messier, W., Li, S.H., Stewart, C.B. 1996. The birth of microsatellites. *Nature* 381:483.
- Meyer, E., Wiegand, P., Rand, S.P., Kuhlmann, D., Brack, M., Brinkmann, B. 1995. Microsatellite polymorphisms reveal phylogenetic relationships in primates. *J. Mol. Evol.* 41:10-4.
- Meyer, D., Thomson, G. 2001. How selection shapes variation of the human major histocompatibility complex: a review. *Ann. Hum. Genet.* 65:1-26.

- MHC Sequencing Consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex (MHC). *Nature* 401:921–23.
- Mill, J., Asherson, P., Browes, C., D'Souza, U., Craig, I. 2002. Expression of the dopamine transporter gene is regulated by the 3' UTR VNTR: Evidence from brain and lymphocytes using quantitative RT-PCR. *Am. J. Med. Genet.* 114:975-9.
- Miller, J.W., Urbinati, C.R., Teng-Ummuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., Swanson, M.S. 2000. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* 19:4439-48.
- Miller, K.M., Withler, R.E. 1996. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43:337-51.
- Miller, K.M., Withler, R.E. 1997. Mhc diversity in Pacific salmon: population structure and trans-species allelism. *Hereditas* 127:83-95.
- Miller, K.M., Kaukinen, K.H., Beacham, T.D., Withler, R.E. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* 111:237-57.
- Miller, H.C., Belov, K., Daugherty, C.H. 2005. Characterization of MHC class II genes from an ancient reptile lineage, *Sphenodon* (tuatara). *Immunogenetics* 57:883-91.
- Mirkin, S.M. 2007. Expandable DNA repeats and human disease. *Nature* 447:932-40.
- Miskey, C., Izsvák, Z., Kawakami, K., Ivics, Z. 2005. DNA transposons in vertebrate functional genomics. *Cell. Mol. Life Sci.* 62:629-41.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-62.
- Muhlethaler-Mottet, A., Otten, L.A., Steimle, V., Mach, B. 1997. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. *EMBO J.* 16:2851-60.
- Munafò, M.R., Clark, T., Flint, J. 2005. Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol. Psychiatry* 10:415-9.
- Mundlos, S., Otto, F., Mundlos, C., Mulliken, J.B., Aylsworth AS, Albright S, Lindhout D, Cole WG, Henn W, Knoll JH, *et al.* 1997. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* 89:773-9.

- Mungall, A.J., Palmer, S.A., Sims, S.K., Edwards, C.A., Ashurst, J.L., Wilming, L., Jones, M.C., Horton, R., Hunt, S.E., Scott, C.E. et al. 2003. The DNA sequence and analysis of human chromosome 6. *Nature* 425:805-11.
- Muntasell, A., Carrascal, M., Alvarez, I., Serradell, L., van Veelen, P., Verreck, F.A., Koning, F., Abian, J., Jaraquemada, D. 2004. Dissection of the HLA-DR4 peptide repertoire in endocrine epithelial cells: strong influence of invariant chain and HLA-DM expression on the nature of ligands. *J. Immunol.* 173:1085-93.
- Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., Matsuo, M. 1993. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* 91:1862-7.
- Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., et al. 2000. A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* 154:1773-84.
- Naruse, K., Shima, A. Nonaka, M. 2000. MHC organisation in the bony fish, medaka. In : Kasahara M, editor. *Major histocompatibility complex : evolution, structure, and function*. Tokyo : Springer-Verlag.
- Nasar, F., Jankowski, C., Nag, D.K. 2000. Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol. Cell. Biol.* 20:3449-58.
- Nath, S., Kales, S., Fujiki, K., Dixon, B. 2006. Major histocompatibility class II genes in rainbow trout (*Oncorhynchus mykiss*) exhibit temperature dependent downregulation. *Immunogenetics* 58:443-53.
- Nickerson, K., Sisk, T.J., Inohara, N., Yee, C.S., Kennell, J., Cho, M.C., Yannie, P.J. 2nd, Nunez, G., Chang, C.H. 2001. Dendritic cell-specific MHC class II transactivator contains a caspase recruitment domain that confers potent transactivation activity. *J. Biol. Chem.* 276:19089-93.
- Nigumann, P., Redik, K., Mätlik, K., Speek, M. 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 79:628-34.
- Nonaka, M., Namikawa, C., Kato, Y., Sasaki, M., Salter-Cid, L., Flajnik, M.F. 1997. Major histocompatibility complex gene mapping in the amphibian *Xenopus* implies a primordial organization. *Proc. Natl. Acad. Sci. USA* 94:5789-91.

- Nucifora, F.C. Jr, Sasaki, M., Peters, M.F., Huang, H., Cooper, J.K., Yamada, M., Takahashi H, Tsuji S, Troncoso J, Dawson VL, *et al.* 2001. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* 291:2423-8.
- Ober, C., Weitkamp, L.R., Cox, N., Dytch, H., Kostyu, D., Elias, S. 1997. HLA and mate choice in humans. *Am. J. Hum. Genet.* 61:497-504.
- Ober, C., Hyslop, T., Elias, S., Weitkamp, L.R., Hauck, W.W. 1998. Human leukocyte antigen matching and fetal loss: results of a 10 year prospective study. *Hum. Reprod.* 13:33-8.
- Ohta, Y., Okamura, K., McKinney, E.C., Bartl, S., Hashimoto, K., Flajnik, M.F. 2000. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *Proc. Natl. Acad. Sci. U S A* 97:4712-7.
- O'Keefe, G.M., Nguyen, V.T., Benveniste, E.N. 1999. Class II transactivator and class II MHC gene expression in microglia: modulation by the cytokines TGF-beta, IL-4, IL-13 and IL-10. *Eur. J. Immunol.* 29:1275-85.
- O'Keefe, G.M., Nguyen, V.T., Ping Tang, L.L., Benveniste, E.N. 2001. IFN-gamma regulation of class II transactivator promoter IV in macrophages and microglia: involvement of the suppressors of cytokine signaling-1 protein. *J. Immunol.* 166:2260-9.
- Olsvik, P.A., Lie, K.K., Jordal, A.E., Nilsen, T.O., Hordvik, I. 2005. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. *BMC Mol. Biol.* 6:21.
- Orr, H.T., Zoghbi, H.Y. 2007. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30:575-621.
- Ostertag, E.M., Kazazian, H.H. Jr. 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35:501-38.
- Pai, R.K., Convery, M., Hamilton, T.A., Boom, W.H., Harding, C.V. 2003. Inhibition of IFN-gamma-induced class II transactivator expression by a 19-kDa lipoprotein from *Mycobacterium tuberculosis*: a potential mechanism for immune evasion. *J. Immunol.* 171:175-84.
- Pace, J.K. 2nd, Feschotte, C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome. Res.* 17:422-32.
- Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D., Baralle, F.E. 2000. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J. Biol. Chem.* 275:21041-7.

- Palm, W., de Lange, T. 2008. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* 42:301-34.
- Pamer, E, Cresswell, P. 1998. Mechanisms of MHC class I - restricted antigen processing. *Annu. Rev. Immunol.* 16:323-58.
- Parham, P. 1988. Function and polymorphism of human leukocyte antigen-A,B,C molecules. *Am. J. Med.* 85:2-5.
- Paterson, S., Wilson, K., Pemberton, J.M. 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. *Proc. Natl. Acad. Sci U S A* 95:3714-9.
- Pati, U., Pati, N. 2000. Lipoprotein(a), atherosclerosis, and apolipoprotein(a) gene polymorphism. *Mol. Genet. Metab.* 71:87-92.
- Pearson, C.E., Eichler, E.E., Lorenzetti, D., Kramer, S.F., Zoghbi, H.Y., Nelson, D.L., Sinden, R.R. 1998. Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* 37:2701-8.
- Pease, L.R., Schulze, D.H., Pfaffenbach, G.M., Nathenson, S.G. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. U. S. A.* 80:242-246.
- Peixoto, A.A., Hennessy, J.M., Townson, I., Hasan, G., Rosbash, M., Costa, R., Kyriacou, C.P. 1998. Molecular coevolution within a Drosophila clock gene. *Proc. Natl. Acad. Sci. U S A* 95:4475-80.
- Peixoto, A.A. 2002. Evolutionary behavioral genetics in Drosophila. *Adv. Genet.* 47:117-50.
- Penn, D.J., Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* 153:145-164.
- Penn, D.J., Damjanovich, K. Potts, W.K. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Natl. Acad. Sci. USA* 99:11260-11264.
- Perry, G.M.L., Tarte, P., Croiseterie, S., Belhumeur, P., Bernachez, L. 2004. Genetic variance and covariance for 0+ weight and resistance to *Aeromonas salmonicida* in the brook charr (*Salvelinus fontinalis*). *Aquaculture* 235:263-271.
- Peterson, J.H., Woolhead, C.A., Bernstein, H.D. 2003. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J. Biol. Chem.* 278:46155-46162.

- Pfeifer, J.D., Wick, M.J., Roberts, R.L., Findlay, K., Normark, S.J., Harding, C.V. 1993. Phagocytic processing of bacterial antigens for class I MHC presentation to T cells. *Nature* 361:359-62.
- Philips, A.V., Timchenko, L.T., Cooper, T.A. 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* 280:737-41.
- Piskurich, J.F., Lin, K.I., Lin, Y., Wang, Y., Ting, J.P., Calame, K. 2000. BLIMP-1 mediates extinction of major histocompatibility class II transactivator expression in plasma cells. *Nat. Immunol.* 1:526-32.
- Plohl, M., Luchetti, A., Mestrovic, N., Mantovani, B. 2008. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409:72-82.
- Potts, W.K., Manning, C.J., Wakeland, E.K. 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* 352:619-21.
- Potts, W.K., Wakeland, E.K. 1990. Evolution of diversity at the major histocompatibility complex. *Trends Ecol. Evol.* 5:181-186.
- Powis, S.J., Young, L.L., Joly, E., Barker, P.J., Richardson, L., Brandt, R.P., Melief, C.J., Howard, J.C., Butcher, G.W. 1996. The rat cim effect: TAP allele-dependent changes in a class I MHC anchor motif and evidence against C-terminal trimming of peptides in the ER. *Immunity* 4:159-65.
- Pozzoli, U., Menozzi, G., Comi, G.P., Cagliani, R., Bresolin, N., Sironi, M. 2007. Intron size in mammals: complexity comes to terms with economy. *Trends Genet.* 23:20-24.
- Pugliese, A., Zeller, M., Fernandez, A. Jr, Zalberg, L.J., Bartlett, R.J., Ricordi, C., Pietropaolo, M., Eisenbarth, G.S., Bennett, S.T., Patel, D.D. 1997. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat. Genet.* 15:293-7.
- Qu, S., Niswender, K.D., Ji, Q., van der Meer, R., Keeney, D., Magnuson, M.A., Wisdom, R. 1997. Polydactyly and ectopic ZPA formation in Alx-4 mutant mice. *Development* 124:3999-4008.
- Raca, G., Siyanova, E.Y., McMurray, C.T., Mirkin, S.M. 2000. Expansion of the (CTG)(n) repeat in the 5'-UTR of a reporter gene impedes translation. *Nucleic Acids Res.* 28:3943-9.
- Raida, M.K., Buchmann, K. 2007. Temperature-dependent expression of immune-relevant genes in rainbow trout following *Yersinia ruckeri* vaccination. *Dis. Aquat. Organ.* 77:41-52.

- Raida, M.K., Buchmann, K. 2008. Bath vaccination of rainbow trout (*Oncorhynchus mykiss* Walbaum) against *Yersinia ruckeri*: effects of temperature on protection and gene expression. *Vaccine* 26:1050-62.
- Rammensee, H.G., Friede, T., Stevanović, S. 1995. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41:178-228.
- Reche, P.A., Reinherz, E.L. 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J. Mol. Biol.* 331:623-41.
- Reith, W., Mach, B. 2001. The bare lymphocyte syndrome and the regulation of MHC expression. *Annu. Rev. Immunol.* 19:331-373.
- Reith, W., LeibundGut-Landmann, S., Waldburger, J.M. 2005. Regulation of MHC class II gene expression by the class II transactivator. *Nat. Rev. Immunol.* 5:793-806.
- Rich, A., Zhang, S. 2003. Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet.* 4:566-72.
- Richard, G.F., Pâques, F. 2000. Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* 1:122-6.
- Richard, G.F., Cyncynatus, C., Dujon, B. 2003. Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. *J. Mol. Biol.* 326:769-82.
- Richard, G.F., Dujon, B. 2006. Molecular evolution of minisatellites in hemiascomycetous yeasts. *Mol. Biol. Evol.* 23:189-202.
- Richard, G.F., Kerrest, A., Dujon, B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72:686-727.
- Rinne, R., Saukko, P., Järvinen, M., Lehesjoki, A.E. 2002. Reduced cystatin B activity correlates with enhanced cathepsin activity in progressive myoclonus epilepsy. *Ann. Med.* 34:380-5.
- Ristow, S.S., Grabowski, L.D., Thompson, S.M., Warr, G.W., Kaattari, S.L., de Avila, J.M., Thorgaard, G.H. 1999. Coding sequences of the MHC II beta chain of homozygous rainbow trout (*Oncorhynchus mykiss*). *Dev. Comp. Immunol.* 23:51-60.

- Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P., Marsh, S.G.E. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Research* 31:311-314
- Rodrigues, P.N., Dixon, B., Roelofs, J., Rombout, J.H., Egberts, E., Pohajdak, B., Stet, R.J. 1998. Expression and temperature-dependent regulation of the beta2-microglobulin (Cyca-B2m) gene in a cold-blooded vertebrate, the common carp (*Cyprinus carpio L.*). *Dev. Immunol.* 5:263-75.
- Rohn, W., Tang, L.P., Dong, Y., Benveniste, E.N. 1999. IL-1 beta inhibits IFN-gamma-induced class II MHC expression by suppressing transcription of the class II transactivator gene. *J. Immunol.* 162:886-96.
- Rosa, F.M., Fellous, M. 1988. Regulation of HLA-DR gene by IFN-gamma. Transcriptional and post-transcriptional control. *J. Immunol.* 140:1660-4.
- Roy, A.M., West, N.C., Rao, A., Adhikari, P., Alemán, C., Barnes, A.P., Deininger, P.L. 2000. Upstream flanking sequences and transcription of SINES. *J. Mol. Biol.* 302:17-25.
- Rudensky, A.Yu, Preston-Hurlburt, P., Hong, S.C., Barlow, A., Janeway, C.A. Jr. 1991. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 353:622-7.
- Rudolph, M.G., Stanfield, R.L., Wilson, I.A. 2006. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* 24:419-466.
- Sabol, S.Z., Hu, S., Hamer, D. 1998 A functional polymorphism in the monoamine oxidase A gene promoter. *Hum. Genet.* 103:273-279.
- Saha, T., Usdin, K. 2001. Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases. *FEBS Lett.* 491:184-7.
- Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M., Wells, R.D. 2001. Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J. Biol. Chem.* 276:27171-7.
- Sambrook, J.G., Russell, R., Umrana, Y., Edwards, Y.J., Campbell, R.D., Elgar, G., Clark, M.S. 2002. Fugu orthologues of human major histocompatibility complex genes: a genome survey. *Immunogenetics* 54:367-80.
- Sammut, B., Du Pasquier, L., Ducoroy, P., Laurens, V., Marcuz, A., Tournefier, A. 1999. Axolotl MHC architecture and polymorphism. *Eur. J. Immunol.* 29:2897-907.

- Sammut, B., Marcuz, A., Pasquier, L.D. 2002. The fate of duplicated major histocompatibility complex class Ia genes in a dodecaploid amphibian, *Xenopus ruwenzoriensis*. *Eur. J. Immunol.* 32:2698-709.
- Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U S A* 74:5463-7.
- Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., Kimura-Yoshida, C., Matsuo I, Sumiyama K, *et al.* 2008. Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. U S A* 105:4220-5.
- Sato, K., Flajnik, M.F., Du Pasquier, L., Katagiri, M., Kasahara, M. 1993. Evolution of the MHC: isolation of class II beta-chain cDNA clones from the amphibian *Xenopus laevis*. *J. Immunol.* 150:2831-43.
- Saunders, R. 1995 Salmon aquaculture: present status and prospectus for the future. In: *A Boghen (ed) Cold-water aquaculture in Atlantic Canada*. The Canadian Institute for Research On Regional Development, Moncton, New Brunswick, Canada, pp 35–82.
- Sawyer, L.A., Hennessy, J.M., Peixoto, A.A., Rosato, E., Parkinson, H., Costa, R., Kyriacou, C.P. 1997. Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* 278:2117-20.
- Schaible, U.E., Winau, F., Sieling, P.A., Fischer, K., Collins, H.L., Hagens, K., Modlin, R.L., Brinkmann, V., Kaufmann, S.H. 2003. Apoptosis facilitates antigen presentation to T lymphocytes through MHC-I and CD1 in tuberculosis. *Nat. Med.* 9:1039-46.
- Schneider, S.C., Sercarz, E.E. 1997. Antigen processing differences among APC. *Hum. Immunol.* 54:148-58.
- Schweitzer, J.K., Reinke, S.S., Livingston, D.M. 2001. Meiotic alterations in CAG repeat tracts. *Genetics* 159:1861-5.
- Schroth, G.P., Chou, P.J., Ho, P.S. 1992. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.* 267:11846-55.
- Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., Willard, H.F. 2001. Genomic and genetic definition of a functional human centromere. *Science* 294:109-15.

- Sen, S., Burmeister, M., Ghosh, D. 2004. Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 127B:85-9.
- Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T., Kikkawa, E., Iwata, K., Tomizawa, M., Okuaki, N., Kuwano, Y., *et al.* 1999. Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. *Proc. Natl. Acad. Sci. U S A* 96:13282-7.
- Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J.K., Inoko, H. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J. Immunol.* 172:6751-63.
- Shiina, T., Hosomichi, K., Inoko, H., Kulski, J.K. 2009. The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* 54:15-39.
- Shum, B.P., Avila, D., Du Pasquier, L., Kasahara, M., Flajnik, M.F. 1993. Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the *Xenopus* MHC. *J. Immunol.* 151:5376-86.
- Skowronski, J., Fanning, T.G., Singer, M.F. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* 8:1385-97.
- Slade, R.W., McCallum, H.I. 1992. Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* 132:861-4.
- Slotkin, R.K., Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8:272-85.
- Smith, K.J., Pyrdol, J., Gauthier, L., Wiley, D.C., Wucherpfennig, K.W. 1998. Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. *J. Exp. Med.* 188:1511-20.
- Spradling, A.C., Stern, D.M., Kiss, I., Roote, J., Laverly, T., Rubin, G.M. 1995. Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. U S A* 92:10824-30.
- Steinman, R.M., Mellman, I.S., Muller, W.A., Cohn, Z.A. 1983. Endocytosis and the recycling of plasma membrane. *J. Cell. Biol.* 96:1-27.
- Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., Wiley, D.C. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215-21.

- Stern, L.J., Potoicchio, I., Santambrogio, L. 2006. MHC class II compartment subtypes: structure and function. *Curr. Opin. Immunol.* 18:64-9.
- Stet, R.J., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B.P., Grimholt, U. 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* 54:320-31.
- Stratikos, E., Wiley, D.C., Stern, L.J. 2004. Enhanced catalytic action of HLA-DM on the exchange of peptides lacking backbone hydrogen bonds between their N-terminal region and the MHC class II alpha-chain. *J. Immunol.* 172:1109-17.
- Subramanian, S., Mishra, R.K., Singh, L. 2003a. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.* 4:R13.
- Subramanian, S., Madgula, V.M., George, R., Mishra, R.K., Pandit, M.W., Kumar, C.S., Singh, L. 2003b. Triplet repeats in human genome: distribution and their association with genes and other genomic regions. *Bioinformatics* 19:549-52.
- Suzuki, Y., Orellana, M.A., Schreiber, R.D., Remington, J.S. 1988. Interferon-gamma: the major mediator of resistance against *Toxoplasma gondii*. *Science* 240:516-8.
- Syed, M., Vestrheim, O., Mikkelsen, B., Lundin, M. 2003. Isolation of the promoters of Atlantic salmon MHCII genes. *Mar. Biotechnol. (NY)* 5:253-60.
- Tam, M., Erin Montgomery, S., Kekis, M., Stollar, B.D., Price, G.B., Pearson, C.E. 2003. Slipped (CTG)(CAG) repeats of the myotonic dystrophy locus: surface probing with anti-DNA antibodies. *J. Mol. Biol.* 332:585-600.
- Takahata, N., Nei, M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124:967-78.
- Takami, K., Zaleska-Rutczynska, Z., Figueroa, F., Klein, J. 1997. Linkage of LMP, TAP, and RING3 with Mhc class I rather than class II genes in the zebrafish. *J. Immunol.* 159:6052-60.
- Terado, T., Okamura, K., Ohta, Y., Shin, D.H., Smith, S.L., Hashimoto, K., Takemoto, T., Nonaka, M.I., Kimura, H., Flajnik, M.F., Nonaka, M. 2003. Molecular cloning of C4 gene and identification of the class III complement region in the shark MHC. *J. Immunol.* 171:2461-6.
- Thierry, A., Bouchier, C., Dujon, B., Richard, G.F. 2008. Megasatellites: a peculiar class of giant minisatellites in genes involved in cell adhesion and pathogenicity in *Candida glabrata*. *Nucleic Acids Res.* 36:5970-82.

- Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C., Moxley, R.T. 3rd. 1997. Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet.* 16:407-9.
- Thursz, M.R., Thomas, H.C., Greenwood, B.M., Hill, A.V. 1997. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat. Genet.* 17:11-12.
- Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B., Thornton, C.A. 2000. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA* 6:79-87.
- Tidow, N., Boecker, A., Schmidt, H., Agelopoulos, K., Boecker, W., Buerger, H., Brandt, B. 2003. Distinct amplification of an untranslated regulatory sequence in the egfr gene contributes to early steps in breast cancer development. *Cancer Res.* 63:1172-8.
- Ting, J.P., Trowsdale, J. 2002. Genetic control of MHC class II expression. *Cell* 109 Suppl:S21-33.
- Tomilin, N.V. 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* 30:338-48.
- Tomiya, H., Miwa, K., Shiga, H., Moore, Y.I., Oka, S., Iwamoto, A., Kaneko, Y., Takiguchi, M. 1997. Evidence of presentation of multiple HIV-1 cytotoxic T lymphocyte epitopes by HLA-B*3501 molecules that are associated with the accelerated progression of AIDS. *J. Immunol.* 158:5026-34.
- Tosi, G., Jabrane-Ferrat, N., Peterlin, B.M. 2002. Phosphorylation of CIITA directs its oligomerization, accumulation and increased activity on MHCII promoters. *EMBO J.* 21:5467-76.
- Tóth, G., Gáspári, Z., Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10:967-81.
- Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., *et al.* 2003. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* 9:928-935.
- Trowsdale, J. 1995. "Both man et bird et beast": comparative organization of MHC genes. *Immunogenetics* 41:1-17.
- Trowsdale, J., Parham, P. 2004. Mini-review: defense strategies and immunity-related genes. *Eur. J. Immunol.* 34:7-17.

- Ugarković, D., Plohl, M. 2002. Variation in satellite DNA profiles--causes and effects. *EMBO J.* 21:5955-9.
- Ugarkovic, D. 2005. Functional elements residing within satellite DNAs. *EMBO Rep.* 6:1035-9.
- Umezū, K., Hiraoka, M., Mori, M., Maki, H. 2002. Structural analysis of aberrant chromosomes that occur spontaneously in diploid *Saccharomyces cerevisiae*: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. *Genetics* 160:97-110.
- Usdin, K., Woodford, K.J. 1995. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res.* 23:4202-9.
- Usdin, K. 2008. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res.* 18:1011-9.
- Vafiadis, P., Bennett, S.T., Colle, E., Grabs, R., Goodyer, C.G., Polychronakos, C. 1996. Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes. *J. Autoimmun.* 9:397-403.
- Valgardsdottir, R., Chiodi, I., Giordano, M., Cobianchi, F., Riva, S., Biamonti, G. 2005. Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. *Mol. Biol. Cell.* 16:2597-604.
- Valitutti, S., Müller, S., Dessing, M., Lanzavecchia, A. 1996. Different responses are elicited in cytotoxic T lymphocytes by different levels of T cell receptor occupancy. *J. Exp. Med.* 183:1917-21.
- Vallejo, A.N., Miller, N.W., Clem, L.W. 1992. Cellular pathway(s) of antigen processing in fish APC: effect of varying in vitro temperatures on antigen catabolism. *Dev. Comp. Immunol.* 16:367-381.
- Vandenbergh, D.J., Persico, A.M., Hawkins, A.L., Griffin, C.A., Li, X., Jabs, E.W., Uhl, G.R. 1992. Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics* 14:1104-6.
- van den Elsen, P.J., van der Stoep, N., Viëtor, H.E., Wilson, L., van Zutphen, M., Gobin, S.J. 2000. Lack of CIITA expression is central to the absence of antigen presentation functions of trophoblast cells and is caused by methylation of the IFN-gamma inducible promoter (PIV) of CIITA. *Hum. Immunol.* 61:850-62.
- van den Elsen, P.J., Holling, T.M., van der Stoep, N., Boss, J.M. 2003. DNA methylation and expression of major histocompatibility complex class I and class II transactivator genes in human developmental tumor cells and in T cell malignancies. *Clin. Immunol.* 109:46-52.

- Van de Water, N., Williams, R., Ockelford, P., Browett, P. 1998. A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. *Thromb. Haemost.* 79:938-42.
- Van Tol, H.H., Wu, C.M., Guan, H.C., Ohara, K., Bunzow, J.R., Civelli, O., Kennedy, J., Seeman, P., Niznik, H.B., Jovanovic, V. 1992. Multiple dopamine D4 receptor variants in the human population. *Nature* 358:149-52.
- Verstrepen, K.J., Jansen, A., Lewitter, F., Fink, G.R. 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37:986-90.
- Villadangos, J.A., Bryant, R.A., Deussing, J., Driessen, C., Lennon-Duménil, A.M., Riese, R.J., Roth, W., Saftig, P., Shi, G.P., Chapman, H.A., *et al.* 1999. Proteases involved in MHC class II antigen presentation. *Immunol. Rev.* 172:109-20.
- Vinogradov, A.E. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* 20:248-253.
- Vinogradov, A.E. 2006. "Genome design" model: evidence from conserved intronic sequence in human-mouse comparison. *Genome Res.* 16:347-354.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833-7.
- Wahls, W.P., Wallace, L.J., Moore, P.D. 1990. The Z-DNA motif d(TG)₃₀ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol. Cell. Biol.* 10:785-93.
- Wang, T., Johnson, N., Zou, J., Bols, N., Secombes, C.J. 2004. Sequencing and expression of the second allele of the interleukin-1beta1 gene in rainbow trout (*Oncorhynchus mykiss*): identification of a novel SINE in the third intron. *Fish Shellfish Immunol.* 16:335-358.
- Waring, M., Britten, R.J. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* 154:791-4.
- Weber, J.L., Wong, C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2:1123-8.
- Wedekind, C., Fürti, S. 1997. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. Biol. Sci.* 264:1471-9.
- Weller, P., Jeffreys, A.J., Wilson, V., Blanchetot, A. 1984. Organization of the human myoglobin gene. *EMBO J.* 3:439-46.

- Wells, R.D., Dere, R., Hebert, M.L., Napierala, M., Son, L.S. 2005. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* 33:3785-98.
- Wessler, S.R. 1988. Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science* 242:399-405.
- West, M.A., Lucocq, J.M., Watts, C. 1994. Antigen processing and class II MHC peptide-loading compartments in human B-lymphoblastoid cells. *Nature* 369:147-51.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973-82.
- Wiegertjes, G.F., Egberts, E., Stet, R.J.M. 1997. Segregation of MHC class I and class II genes in the common carp (*Cyprinus carpio L.*). *Dev. Comp. Immunol.* 21:105.
- Wierdl, M., Dominska, M., Petes, T.D. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146:769-79.
- Wilder, J., Hollocher, H. 2001. Mobile elements and the genesis of microsatellites in dipterans. *Mol. Biol. Evol.* 18:384-92.
- Wilhelm, M., Wilhelm, F.X. 2001. Reverse transcription of retroviruses and LTR retrotransposons. *Cell. Mol. Life Sci.* 58:1246-62.
- Wyman, A.R., White, R. 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. U S A* 77:6754-8.
- Yamazaki, K., Boyse, E.A., Miké, V., Thaler, H.T., Mathieson, B.J., Abbott, J., Boyse, J., Zayas, Z.A., Thomas, L. 1976. Control of mating preferences in mice by genes in the major histocompatibility complex. *J. Exp. Med.* 144:1324-35.
- Yamazaki, K., Beauchamp, G.K., Egorov, I.K., Bard, J., Thomas, L., Boyse, E.A. 1983. Sensory distinction between H-2b and H-2bm1 mutant mice. *Proc. Natl. Acad. Sci. U S A* 80:5685-8.
- Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., Kinzler, K.W. 2002. Allelic variation in human gene expression. *Science* 297:1143.
- Young, L.J., Nilsen, R., Waymire, K.G., MacGregor, G.R., Insel, T.R. 1999. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* 400:766-8.

- Yrlid, U., Wick, M.J. 2000. Salmonella-induced apoptosis of infected macrophages results in presentation of a bacteria-encoded antigen after uptake by bystander dendritic cells. *J. Exp. Med.* 191:613-24.
- Zamorano, R., Suchindran, S., Gainer, J.V. 2006. 3'-Untranslated region of the type 2 bradykinin receptor is a potent regulator of gene expression. *Am. J. Physiol. Renal Physiol.* 290:F456-64.
- Zamorzaeva, I., Rashkovetsky, E., Nevo, E., Korol, A. 2005. Sequence polymorphism of candidate behavioural genes in *Drosophila melanogaster* flies from 'Evolution canyon'. *Mol. Ecol.* 14:3235-45.
- Zangenberg, G., Huang, M.M., Arnheim, N., Erlich, H. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nat. Genet.* 10:407-414.
- Zardoya, R., Vollmer, D.M., Craddock, C., Streelman, J.T., Karl, S., Meyer, A. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc. Biol. Sci.* 263:1589-98.
- Zhu, J., Paul, W.E. CD4 2008. T cells: fates, functions, and faults. *Blood* 112:1557-69.
- Zhu, Y., Queller, D.C., Strassmann, J.E. 2000. A phylogenetic perspective on sequence evolution in microsatellite loci. *J. Mol. Evol.* 50:324-38.
- Zimmerer, E.J., Passmore, H.C. 1991. Structural and genetic properties of the Eb recombinational hotspot in the mouse. *Immunogenetics* 33:132-140.

Chapitre 7

Annexes

7.1. Figure supplémentaire

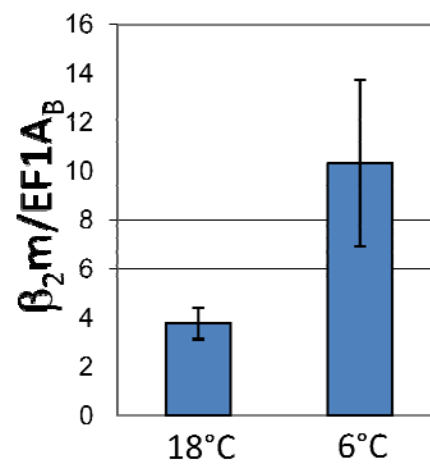


Figure 7.1 Niveaux d'expression de la β_2 -microglobuline en fonction de la température. Quantification des ratios moyens de l'expression de $\beta_2m / EF1A_B$ (+/- écart type) dans le pronéphros de l'omble de fontaine à 18 °C et 6 °C. La comparaison des distributions par le test-*t* de Student renvoie un résultat significatif $p < 0,001$, $n = 6$.

7.2. Curriculum vitae

Formation académique

2007-		UNIVERSITÉ DE MONTRÉAL
	Cours complétés	Microprogramme en enseignement supérieur PLU6035 - Pratique de l'enseignement supérieur PPA6015 - Méthodes d'enseignements et TICS
2005-		UNIVERSITÉ DE MONTRÉAL
	Diplôme à venir:	Doctorat en Microbiologie et Immunologie Laboratoire du Dr. Pierre Belhumeur
	Titre de la thèse:	Étude du polymorphisme du gène majeur d'histocompatibilité de classe II bêta (MHII β) chez l'omble de fontaine (<i>Salvelinus fontinalis</i>).
1997-2000		UNIVERSITÉ DE MONTRÉAL
	Diplôme:	Maîtrise en Biochimie Laboratoire du Dre. Sylvie Mader
	Titre du mémoire:	Étude des voies de signalisation activées par l'acide ricinoléique, le composant majeur de l'huile de ricin.
1994-1997		UNIVERSITÉ DE MONTRÉAL
	Diplôme:	Baccalauréat en Biochimie
1992-1994		COLLÈGE DE SHAWINIGAN
	Diplôme:	D.E.C. en Sciences pures

Expériences de travail

Depuis septembre 2005		UNIVERSITÉ DE MONTRÉAL
	Poste:	Auxiliaire d'enseignement
	Cours :	MCB2986 Immunologie générale MCB1979 Microbiologie générale MCB2979 Bactériologie générale MCB2990 Microbiologie et pathogènes
Août 2001 - Août 2005		UNIVERSITÉ DE MONTRÉAL
		Microbiologie/Immunologie Laboratoire du Dr. Pierre Belhumeur
	Poste:	Agent de recherche
	Projet:	Résistance de l'omble de fontaine au pathogène <i>Aeromonas salmonicida</i> .
Juin 2000 - Août 2001		UNIVERSITÉ DE MONTRÉAL
		Biochimie Laboratoire du Dr. Sylvie Mader
	Poste:	Agent de recherche
	Projet:	Rôle de la signalisation des récepteurs à tyrosine kinase, ErbB, dans le pronostic de cancer du sein.

Publications:

Croisetière, S., Bernatchez, L. and Belhumeur, P. Temperature and length-dependent modulation of the MH class II β gene expression in brook charr (*Salvelinus fontinalis*) by a *cis*-acting minisatellite. (soumis).

Croisetière, S., Tarte, P.D., Bernatchez, L. and Belhumeur, P. (2008) Identification of MHC class IIbeta resistance/susceptibility alleles to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*). Mol Immunol. 45(11):3107-3116.

Dautremepuits, C., Fortier, M., **Croisetière, S.**, Belhumeur, P. and Fournier, M. (2006) Modulation of juvenile brook trout (*Salvelinus fontinalis*) cellular immune system after *Aeromonas salmonicida* challenge. Vet Immunol Immunopathol. 110(1-2):27-36.

Dayan, G., Lupien, M., Auger, A., Anghel, S.I., Rocha, W., **Croisetière, S.**, Katzenellenbogen, J.A. and Mader, S. (2006) Tamoxifen and raloxifene differ in their functional interactions with aspartate 351 of estrogen receptor alpha. Mol Pharmacol. 70(2):579-588.

Perry, G.M.L., Tarte, P.D., **Croisetière, S.**, Belhumeur, P. and Bernatchez, L. (2004) Genetic variance and covariance for 0+ brook charr (*Salvelinus fontinalis*) weight and survival time of furunculosis (*Aeromonas salmonicida*) exposure. Aquaculture 235: 263-271.

Communications orales:

Croisetière, S., Tarte, P.D., Bernatchez, L. et Belhumeur, P. Identification of MHC class II β resistance/susceptibility alleles to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*). International Symposium on *Aeromonas* and *Plesiomonas* (ISAP2008), 10-12 septembre 2008, Vila Real (Portugal).

Croisetière, S., Tarte, P.D., Bernatchez, L. et Belhumeur, P. Corrélation entre des allèles des gènes du complexe majeur d'histocompatibilité (CMH) et une résistance à la furonculose chez l'omble de fontaine. Réunion annuelle du Réseau Aquaculture Québec (RAQ), 6 novembre 2006, Rimouski.

Croisetière, S., Alaoui-Jamali, M. et Mader, S. Le cremophor EL et l'acide ricinoléique activent AP-1 et les MAPKs par plusieurs mécanismes indépendants. Rencontre scientifique Louis-Philippe Bouthillier, mai 1999, Montréal.

Croisetière, S., Alaoui-Jamali, M. et Mader, S. Importance des acides gras poly-insaturés en C18 comme molécules signalisatrices. XVIIIe congrès de l'AEGSFM, janvier 1999, Montréal.

Communication par affichage:

Croisetière, S., Bernatchez, L. et Belhumeur, P. Caractérisation d'un minisatellite retrouvé dans le gène du complexe majeur d'histocompatibilité (CMH) de classe II β chez l'omble de fontaine (*Salvelinus fontinalis*). Réunion annuelle du Réseau Aquaculture Québec (RAQ), 6 novembre 2006, Rimouski.

Croisetière, S., Nassif, R., Tarte, P.D., Bernatchez, L. et Belhumeur, P. Vers l'identification d'omble de fontaine plus résistantes à la furonculose. Rencontre technologique organisée conjointement par le CQVB et le RAQ sur le thème : « Recherche en Aquaculture : Défis et perspectives », septembre 2005, Québec.

