

UNIVERSITÉ DE MONTRÉAL

**VARIABILITÉ DANS LES PROMOTEURS DES GÈNES  
DU MÉTABOLISME DU GLUTATHION**

Par  
Catherine Labbé

Faculté des études supérieures

Mémoire présenté à la Faculté des études supérieures  
En vue de l'obtention du grade de  
Maître ès sciences (M. Sc.)  
En biologie moléculaire

Août 2005

©, Catherine Labbé, 2005





## AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

## NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

UNIVERSITE DE MONTREAL  
Faculté des études supérieures

Ce mémoire intitulé :  
VARIABILITÉ DANS LES PROMOTEURS DES GÈNES  
DU MÉTABOLISME DU GLUTATHION

Présenté par :  
Catherine Labbé

A été évalué par un jury composé des personnes suivantes :

Dr Maja Krajinovic

.....  
président-rapporteur

Dr Damian Labuda

.....  
directeur de recherche

Dr Francine M. Mayer

.....  
membre du jury

## RÉSUMÉ

Les polymorphismes des régions régulatrices avec un impact fonctionnel pourraient expliquer la susceptibilité à certaines maladies et/ou la réponse différentielle aux médicaments. Les enzymes du métabolisme du glutathion sont particulièrement intéressants à cause de leur implication dans la détoxification des cellules et la protection contre les oxydants. Pour quantifier la diversité génétique dans la région promotrice des gènes correspondants à ces enzymes, nous avons caractérisé la variabilité dans des segments de 2kb en amont des sites d'initiation de la transcription. Les polymorphismes ont été identifiés dans un échantillon mondial de 40 individus et étudiés dans un échantillon élargi provenant de cinq populations continentales (n=80) et de 8 sous-populations (n=142). Sept gènes ont été étudiés : GSS, GSTT1, GSTP1, GSTM3, GSTM4, GPX2, GPX3. La diversité nucléotidique varie de 0,061% (GSTM3 avec 4 sites polymorphiques and 5 haplotypes) à 0,205% (GSTP1 avec 13 sites polymorphiques et 13 haplotypes). L'« haplotype » de GSTT1 le plus fréquent (44%) est une délétion complète du gène. Les haplotypes de GSS et GSTT1 peuvent être expliqués par une suite de mutations alors que les réseaux haplotypiques des autres gènes suggèrent des épisodes de recombinaisons. Les analyses MDS basées sur les Fst par paire montrent des différences marquées entre l'Afrique et le reste du monde. En général, les haplotypes africains majeurs ne sont pas les plus importants dans les autres populations et il y a plusieurs haplotypes spécifiques à l'Afrique. Les populations européenne, amérindienne et moyenne-orientale semblent peu différenciées. La population asiatique présente des différences majeures avec l'Afrique et le bloc Europe/Amérique/Moyen-Orient. Les populations islandaise et aborigène de Taiwan sont génétiquement rapprochées de l'Europe et l'Asie respectivement, quoique plus homogènes. Les sous-groupes québécois sont généralement peu différenciés de la population européenne, mais il y a quelques singularités dues entre autres à l'effet fondateur. Les résultats de ce projet pourront être utilisés comme outils génétiques pour étudier la relation entre la susceptibilité aux maladies ou la réponse thérapeutique et les variants dans l'ADN humain.

Mots-clés : polymorphismes, régions régulatrices, métabolisme du glutathion, haplotypes, génétique des populations

## SUMMARY

Regulatory polymorphisms that modify genetic expression may have an impact on the susceptibility to diseases and/or on the differential response to treatment. Enzymes of glutathione metabolism are of special interest because of their involvement in detoxification of the cell and its protection against reactive oxygen species. To learn about genetic diversity in the bona fide promoter regions of the corresponding genes, we characterized variability of 2 kb segments upstream of the transcription start site. Polymorphisms were ascertained in a world sample of 40 individuals and studied further in an extended sample of five continental populations ( $n = 80$ ) and 8 subpopulations ( $n=142$ ). Seven genes have been studied GSS, GSTT1, GSTP1, GSTM3, GSTM4, GPX2, GPX3. Nucleotide diversities range from 0.061% (GSTM3 with 4 segregating sites and 5 haplotypes) to 0.205% (GSTP1 with 13 segregating sites and 13 haplotypes). The most frequent “haplotype” of GSTT1 (44%) is a deletion of the entire gene. While GSTT1 and GSS haplotypes can be explained entirely through mutations, the haplotype networks of the other genes indicate recombination episodes. MDS analyses based on pairwise  $F_{st}$  in continental populations demonstrate striking differences between Africa and the rest of the world for these genes. In general, the major haplotypes in Africa are not the most important in other populations and there are many African specific haplotypes. European, Amerindian and Middle-eastern population appear very similar. Asia is different from both Africa and the Europe/America/Middle East block. Populations from Iceland and Taiwan are close to European and Asian populations respectively, although they are more homogenous. Quebec subpopulations are generally similar to the European population but there are some singularities due to founder effect. These results could be used as genetic tools to study relationships between disease susceptibility or therapeutic responses and human DNA variants.

Key words : polymorphisms, regulatory regions, glutathion metabolism genes, haplotypes, population genetics

<b>TABLE DES MATIÈRES</b>	
<b>RÉSUMÉ</b>	<b>III</b>
<b>SUMMARY</b>	<b>IV</b>
<b>TABLE DES MATIÈRES</b>	<b>V</b>
<b>LISTE DES TABLEAUX</b>	<b>VII</b>
<b>LISTE DES FIGURES</b>	<b>IX</b>
<b>LISTE DES ABRÉVIATIONS</b>	<b>XI</b>
<b>REMERCIEMENTS</b>	<b>XIII</b>
<b>CHAPITRE 1 : INTRODUCTION</b>	<b>1</b>
<b>1 LE MÉTABOLISME DES SUBSTANCES XÉNOBIOTIQUES</b>	<b>2</b>
1.1 LES OXYDANTS	2
1.2 LE GLUTATHION	4
<b>2 LES GÈNES IMPLIQUÉS DANS LE CHEMINEMENT DU GSH</b>	<b>7</b>
2.1 GSS	7
2.2 LES TRANSFERASES	7
2.2.1 <i>La famille M – GSTM3 et GSTM4</i>	11
2.2.2 <i>La famille T – GSTT1</i>	12
2.2.3 <i>La famille Pi – GSTP1</i>	13
2.3 LES PEROXYDASES	13
2.3.1 <i>GPX2</i>	14
2.3.2 <i>GPX3</i>	14
<b>3 LA VARIABILITÉ</b>	<b>15</b>
<b>4 LA RÉGULATION</b>	<b>17</b>
<b>5 RELATIONS ENTRE LE MÉTABOLISME DES XÉNOBIOTIQUES, LA MALADIE ET LA RÉPONSE THÉRAPEUTHIQUE</b>	<b>20</b>
<b>6 LES PARAMÈTRES DE LA DIVERSITÉ</b>	<b>23</b>
<b>7 OBJECTIFS DU PROJET</b>	<b>25</b>
<b>CHAPITRE II : MATERIEL ET METHODES</b>	<b>26</b>
<b>1 ÉCHANTILLONS</b>	<b>27</b>
<b>2 SÉQUENCES</b>	<b>29</b>
<b>3 AMPLIFICATION PCR</b>	<b>30</b>
<b>4 HYBRIDATION ASO</b>	<b>33</b>
<b>5 GÉNOTYPER LA DÉLÉTION GSTT1</b>	<b>35</b>
<b>6 BASE DE DONNÉES ET LOGICIELS</b>	<b>36</b>

<b>CHAPITRE III : RÉSULTATS</b>	<b>37</b>
<b>1 DIVERSITÉ GÉNÉTIQUE</b>	<b>38</b>
GSS	42
GSTT1	45
GSTP1	50
GSTM3	53
GSTM4	56
GPX2	59
GPX3	63
<b>2 APPORT DES MUTATIONS ET DES RECOMBINAISONS</b>	<b>66</b>
<b>3 VARIATIONS À TRAVERS LES POPULATIONS</b>	<b>73</b>
<b>4 LA DIVERGENCE</b>	<b>78</b>
<b>5 VARIABILITÉ DANS LES PARTIES INTRONIQUES ET EXONIQUES</b>	<b>80</b>
<b>CHAPITRE IV : DISCUSSION</b>	<b>82</b>
DIVERSITÉ DANS LES GÈNES DE LA VOIE MÉTABOLIQUE DU GLUTATHION	84
LA DIVERSITÉ QUÉBÉCOISE	89
ABORIGÈNES DE TAIWAN ET ISLANDAIS : DES POPULATIONS TYPES	92
GSTs, GPXS ET VARIABILITÉ	94
<b>CONCLUSION</b>	<b>96</b>
<b>BIBLIOGRAPHIE</b>	<b>97</b>
<b>ANNEXES</b>	<b>XIV</b>



## **LISTE DES TABLEAUX**

Tableau I. Les GSTs et leurs substrats	10
Tableau II. Localisation tissulaire des GSTs les plus étudiés	11
Tableau III. Origine des échantillons d'ADN	28
Tableau IV. Longueur et position des séquences étudiées	29
Tableau V. Protocole de PCR utilisés	31
Tableau VI. Composition de la réaction de PCR multiplex pour la délétion de GSTT1	35
Tableau VII. Amorces pour le PCR multiplex de la délétion de GSTT1	35
Tableau VIII. Statistiques descriptives pour GSS	43
Tableau IX. Haplotypes pour GSS et fréquences dans les populations continentales	43
Tableau X. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GSS	44
Tableau XI. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et des Islandais pour GSS	44
Tableau XII. Statistiques descriptives pour GSTT1	47
Tableau XIII. Haplotypes pour GSTT1 et fréquences dans les populations continentales	47
Tableau XIV. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GSTT1	48
Tableau XV. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GSTT1	49
Tableau XVI. Statistiques descriptives pour GSTP1	51
Tableau XVII. Haplotypes pour GSTP1 et fréquences dans les populations continentales	51
Tableau XVIII. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GSTP1	52
Tableau XIX. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GSTP1	52
Tableau XX. Statistiques descriptives pour GSTM3	54
Tableau XXI. Haplotypes pour GSTM3 et fréquences dans les populations continentales	54
Tableau XXII. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GSTM3	55
Tableau XXIII. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GSTM3	55
Tableau XXIV. Statistiques descriptives pour GSTM4	57

Tableau XXV. Haplotypes pour GSTM4 et fréquences dans les populations continentales	57
Tableau XXVI. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GSTM4	58
Tableau XXVII. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GSTM4	58
Tableau XXVIII. Statistiques descriptives pour GPX2	60
Tableau XXIX. Haplotypes pour GPX2 et fréquences dans les populations continentales	61
Tableau XXX. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GPX2	62
Tableau XXXI. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GPX2	62
Tableau XXXII. Statistiques descriptives pour GPX3	64
Tableau XXXIII. Haplotypes pour GPX3 et fréquences dans les populations continentales	64
Tableau XXXIV. Statistiques descriptives (A) et fréquences haplotypiques (B) des groupes québécois pour GPX3	65
Tableau XXXV. Statistiques descriptives (A) et fréquences haplotypiques (B) des Aborigènes de Taiwan et les Islandais pour GPX3	65
Tableau XXXVI. Taux de recombinaison	67
Tableau XXXVII. Taux de divergence des séquences de chimpanzé	78
Tableau XXXVIII. Statistiques pour GPX2, variabilité dans les promoteurs	80
Tableau XXXIX. Statistiques comparatives pour GPX3, variabilité dans les promoteurs et dans la partie intronique/exonique	81
Tableau XXXX. Origine des échantillons supplémentaires utilisés	xvii
Tableau XXXXI. Liste des polymorphismes étudiés	xviii
Tableau XXXXII. Liste des amorces de PCR utilisées	xix
Tableau XXXXIII. Liste des sondes utilisées pour l'ASO	xxii

## LISTE DES FIGURES

Figure 1. Structure du GSH	4
Figure 2. Survol des activités métaboliques du GSH	5
Figure 3. Réactions catalysées par les GSTs.	8
Figure 4. Le Tandem GSTM et la délétion de GSTM1	12
Figure 5. La région du gène GSTT1.	13
Figure 6. La variabilité due aux polymorphismes	15
Figure 7. Complexe d'initiation de la transcription.	17
Figure 8. Éléments de séquence du cœur d'un promoteur	19
Figure 9. Fragments amplifiés ensemble et orientation des gènes.	32
Figure 10. Polymorphismes dans les promoteurs.	39
Figure 11. Hybridation ASO pour G564C GSTP1	40
Figure 12. Résultats de génotypage de la délétion de GSTT1.	45
Figure 13. Pourcentage de délétion par continent.	48
Figure 14. Réseau haplotypique pour GSS.	69
Figure 15. Réseau haplotypique pour GSTT1.	69
Figure 16. Réseau haplotypique pour GSTP1.	70
Figure 17. Réseau haplotypique pour GSTM3.	70
Figure 18. Réseau haplotypique pour GSTM4.	71
Figure 19. Réseau haplotypique pour GPX2.	71
Figure 20. Réseau haplotypique pour GPX3.	72
Figure 21. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $\pi$ des échantillons continentaux pour tous les gènes étudiés	75
Figure 22. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $G$ des échantillons continentaux pour tous les gènes étudiés	75
Figure 23. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $\pi$ des groupes québécois pour tous les gènes étudiés.	76
Figure 24. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $G$ des groupes québécois pour tous les gènes étudiés	76
Figure 25. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $G$ des groupes québécois pour GPX3.	77
Figure 26. Analyse MDS des $F_{st}$ calculés à partir des valeurs de $G$ de différentes populations pour GSTM3.	77
Figure 27. Nombre de polymorphismes en fonction de la divergence.	79

Figure 28. Diversité nucléotidique en fonction de la divergence.	79
Figure 29. Carte du monde	83
Figure 30. Carte du Québec.	88

## LISTE DES ABRÉVIATIONS

°C: degré Celcius.

Θ : un estimateur de la diversité, le produit du taux de mutation et de l'effectif efficace

π : diversité nucléotidique

ADN: acide désoxyribonucléique

ANC : ancestral

ARNm: acide ribonucléique messenger

ASO : allelic specific oligonucleotide hybridization

ATP: adenosine tri-phosphate

cM: centimorgan

dHPLC: chromatographie dénaturante à haute pression

EDTA : ethylenediamine-tetraacetic acid

G : diversité haplotypique

GPX2: glutathione peroxydase 2

GPX3: glutathione peroxydase 3

GSH: glutathion

GSS: synthétase du glutathion

GST: *glutathione S-transferase*

GSTM3: *glutathione S-transferase M3*

GSTM4: *glutathione S-transferase M4*

GSTP1: *glutathione S-transferase P1*

GSTT1: *glutathione S-transferase T1*

H<sub>S</sub>: hétérozygotie moyenne des sous-populations

H<sub>T</sub>: hétérozygotie totale sans division en sous-population

k : nombre d'haplotypes

kb : kilobase

Mb: Mégabase

MDS : multidimensional scaling

N: effectif efficace

pb : paire de base

PCR: polymerase chain reaction

pol II : polymérase II

RFLP: *restricted fragment length polymorphism*

ROS: *reactive oxygen species*

S: nombre de polymorphismes

SDS : sodium dodecyl sulfate, detergent

SNP: *single nucleotide polymorphism*

SSPE : saline-sodium phosphate-EDTA, solution tampon

T<sub>m</sub> : *melting temperature*

UCSC : Université de Californie, Santa Cruz

UV : ultraviolet

VIH: virus d'immunodéficience humaine

## REMERCIEMENTS

J'ai plusieurs remerciements à offrir à des personnes qui au cours de l'aventure de ma maîtrise m'ont apporté aide et soutien :

Génome Québec pour le soutien financier.

Mon directeur de recherche Dr Damian Labuda qui m'a accueillie dans son labo et initiée à ce fascinant domaine qu'est la génétique des populations.

Vania, un des meilleurs professeurs que j'ai eus.

Tous les membres de mon labo et des labos des Drs Sinnett et Krajinovic. Vous m'avez montré ce qu'est la coopération en recherche et parallèlement, vous êtes devenus mes amis. La barre est haute pour mes futurs collègues.

Caroline pour notre collaboration qui mènera à un article, j'espère !

Claudia, Stéphane, Sylvie, Hélène et (même !) Jean-François pour leur temps et leur patience.

Mes parents et ma sœur qui ont supporté mon humeur durant la rédaction et qui soutiennent toujours mes choix.

Anna et Jas. Les filles, ça aurait été pas mal moins drôle sans vous !

**Chapitre I**

**INTRODUCTION**



# 1 LE MÉTABOLISME DES SUBSTANCES XÉNOBIOTIQUES

## 1.1 Les oxydants

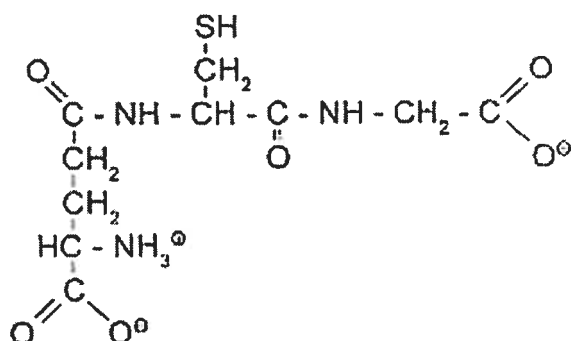
L'avènement de l'oxygène a rendu possible la vie des organismes supérieurs. Paradoxalement, il s'agit d'une substance dangereuse qui tue à petit feu. Le «côté obscur» de l'oxygène provient du fait que l'atome a un électron libre et que l'oxygène moléculaire O<sub>2</sub> en a deux, c'est-à-dire qu'ils sont des radicaux libres. Le processus de respiration cellulaire lors duquel l'oxygène est réduit jusqu'à générer une molécule d'eau est considéré relativement sûr. Cependant suite à cette chaîne de réactions, des produits toxiques subsistent, les *reactive oxygen species* (ROS) qui possèdent des radicaux libres. Parmi les ROS, le radical anion superoxyde, le peroxyde d'hydrogène et le radical hydroxyle sont constamment générés dans les cellules à la suite de réactions métaboliques normales. Ces produits attaquent la structure et la messagerie cellulaire. Au cours de l'évolution, la cellule a développé des enzymes et des molécules antioxydantes afin de contrer les dommages causés par les ROS. Mais malgré tout, la dégradation causée par les oxydants semble inhérente à la vie aérobie. En effet, le vieillissement et les maladies dégénératives comme le cancer, le Parkinson, l'Alzheimer et plusieurs autres sont directement causés par ces derniers (Ames, Shigenaga et al. 1993; Davies 1995).

Les oxydants agissent directement sur l'organisme principalement à travers l'alimentation. Étant donné la diversité des territoires sur la planète et les différentes habitudes alimentaires qui en découlent, on peut penser que les mécanismes de défense contre les oxydants ont évolué pour s'adapter à ces différences. Les gènes impliqués dans le métabolisme des substances étrangères ou xénobiotiques doivent donc porter des variations suite à des épisodes de sélection en faveur des allèles les plus adaptés à la survie. Ces variations peuvent être responsables de changements dans la structure des protéines et/ou dans le niveau d'expression des gènes et par la même occasion, entraîner des réponses différentielles à l'environnement. C'est cette prémisse qui a été à l'origine du projet *Regulatory Genetics* de Génome Québec. Dans le cadre ce projet, on tente d'identifier des polymorphismes régulateurs, dans la région promotrice définie

arbitrairement à 2 kb en amont du site d'initiation de la transcription, associés à certaines maladies et réponses thérapeutiques. Dans ce but, quelques centaines de gènes ont été analysés par chromatographie dénaturante à haute pression (dHPLC) puis par séquençage pour cibler des variations. Ce projet de maîtrise fait partie du volet génétique des populations du projet *Regulatory Genetics*. Il concerne les gènes de la famille du glutathion qui, à cause de leur implication directe dans la protection contre les oxydants et dans l'alimentation, sont de bons gènes candidats pour étudier la susceptibilité à certaines maladies et la réponse thérapeutique. Nous nous proposons de recenser la distribution des variants, individuellement dans sept gènes impliqués dans le métabolisme des xénobiotiques (GSTT1, GSS, GSTP1, GSTM3, GSTM4, GPX2 et GPX3) mais surtout dans la voie métabolique en général afin d'avoir un aperçu des facteurs responsables du profil génétique actuel des populations incluses dans notre analyse.

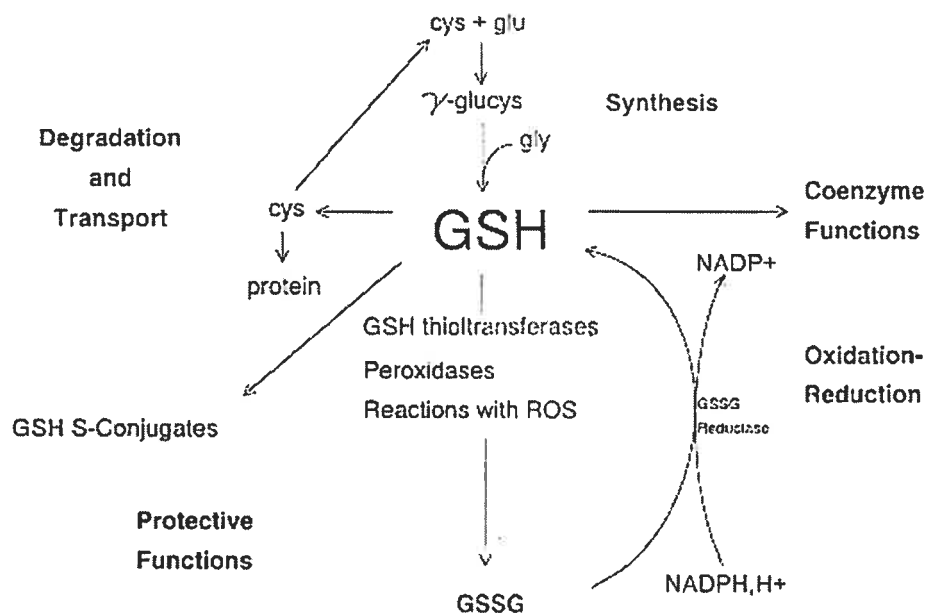
## 1.2 Le glutathion

Le glutathion (figure 1 page 4) et sa voie métabolique (figure 2 page 5) ont été choisis parce qu'ils se trouvent au cœur des processus de protections contre les oxydants. Le glutathion est une molécule composée de trois peptides glutamate, cystéine, glycine qui possède un lien peptidique  $\delta$  inhabituel qui l'empêche d'être hydrolysée par la plupart des peptidases (Anderson 1998).



**Figure 1. Structure du GSH : L-g-glutamyl-L-cysteinylglycine.**(Anderson 1998)

Le GSH est présent dans la majorité des cellules mammifères et dans plusieurs cellules procaryotes. Il remplit une foule de fonctions essentielles au métabolisme cellulaire. C'est le thiol (molécule possédant un groupement SH) intracellulaire le plus abondant (0,2 à 10 mM ). La figure 2 présente un survol de ses activités.



**Figure 2. Survol des activités métaboliques du GSH (Anderson 1998)**

Le GSH est souvent impliqué comme coenzyme dans diverses réactions. De plus, il contribue au transport des acides aminés. Il est impliqué dans le métabolisme et le maintien des groupements thiol des protéines et des molécules de faible poids moléculaire comme la coenzyme A. Il agit également de façon à maintenir l'acide ascorbique dans sa forme réduite et participe à la formation des désoxyribonucléotides (Anderson 1998). Il réagit rapidement de façon non enzymatique avec les radicaux hydroxyles formés lors de la réaction de Fenton et avec le  $N_2O_3$  produit par la réaction de l'oxyde nitrique avec l' $O_2$  et le superoxyde pour les détoxifier (Kalyanaraman, Karoui et al. 1996; Luperchio, Tamir et al. 1996; Briviba, Klotz et al. 1999). Dans les réactions catalysées par les GSH peroxydases, le GSH participe à la réduction et la détoxification du peroxyde d'hydrogène et des peroxydes lipides (Kehrer and Lund 1994; Ursini, Maiorino et al. 1995). Toutes ces réactions aboutissent directement ou indirectement à la formation de disulfide de glutathione (GSSG) par l'enzyme *glutathione disulfide reductase* dans une réaction NADPH – dépendante (Kehrer and Lund 1994).

Même, si le GSH réagit spontanément avec plusieurs électrophiles, la majorité des réactions nécessite une famille d'enzymes catalyseurs connus sous le nom de GSH S-transférases (GST) (Armstrong 1997). Le glutathion est synthétisé dans la cellule sous l'action de deux enzymes. La ligase glutamate cystéine catalyse la première étape qui est la réaction limitative dans la synthèse du glutathion (Deneke and Fanburg 1989). La GSH synthétase (GSS) complète la synthèse à travers l'addition de glycine à  $\delta$ -glutamylcystéine (Tanaka, Yamaguchi et al. 1993).

## 2 LES GÈNES IMPLIQUÉS DANS LE CHEMINEMENT DU GSH

Les 7 gènes étudiés ont été sélectionnés en fonction de leur disponibilité dans la base de données du projet *Regulatory Genetics* de Génome Québec et de leur implication dans les activités métaboliques du GSH.

### 2.1 GSS

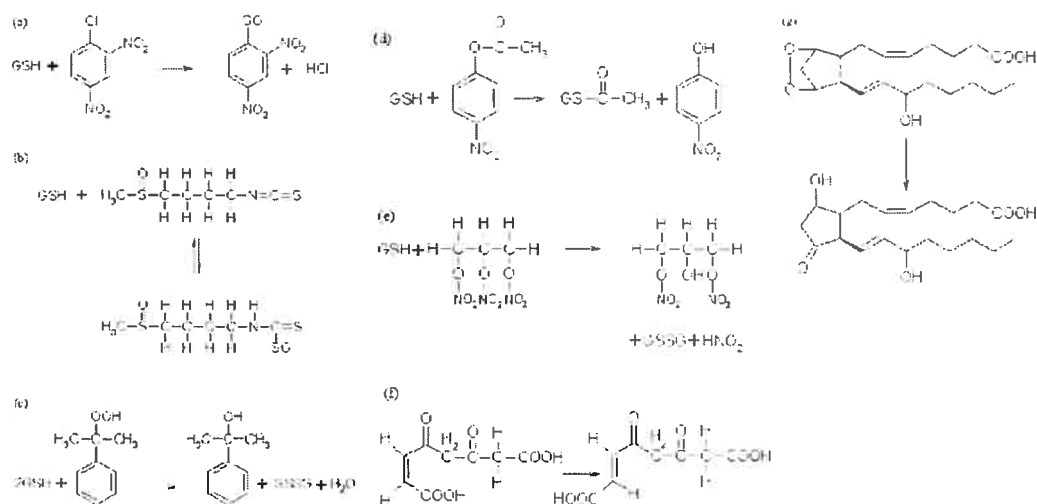
L'enzyme GSS humaine est un homodimère composé de 2 sous-unités ( $M_r$  : 52 344 ) contenant environ 2 % de glucides ( sucres neutres, sucres aminés et acides sialiques ) et dont le gène se situe au locus 20q11.2. (Shi, Habib et al. 1996). GSS est une des deux enzymes qui catalysent la synthèse du glutathion. Des études (Oppenheimer, Wellner et al. 1979) de spécificité ont prouvé que cet enzyme est hautement spécifique à la glycine et au groupement cysteinyl du  $\delta$ -glutamylcystéine. La cinétique et les mécanismes chimiques du GSS sont apparemment très semblables à ceux de la ligase glutamate cystéine (l'autre enzyme responsable de la synthèse du glutathion). Le  $\delta$ -glutamylcystéine et l'ATP réagissent et forment le L- $\delta$ -glutamyl-L-cysteinylphosphate et cet intermédiaire réagit avec la glycine pour former le GSH selon l'équation suivante (Nishimura, Dodd et al. 1964) :



### 2.2 Les transférases

Les transférases de glutathion furent historiquement appelées *glutathione S-transferase* et c'est de cette appellation que découle l'abréviation populaire GST. Ces enzymes catalysent l'attaque nucléophile du GSH réduit sur des composés non-polaires qui contiennent un carbone électrophile, un atome d'azote ou de soufre (Hayes, Flanagan et al. 2004) (Figure 3 page 8). Dans la nature, il existe trois grandes classes de protéines qui présentent des activités de transférase. Les GST cytosoliques et les mitochondriales sont plus ou moins apparentées. La troisième classe est composée des GST microsomiales qu'on appelle également MAPEG pour *membrane associated proteins in eicosanoid and*

*glutathione metabolism* (Jakobsson, Morgenstern et al. 1999). Les GST cytosoliques représentent la plus grande classe de transférases et réagissent d'une manière caractéristique. Elles catalysent la thiolase de l'acétate 4-nitrophenyl; présentent une activité de transférase de groupement thiol; réduisent la trinitroglycérine, les acides déhydroascorbique et monométhylarsonique; et catalyse l'isomérisation du maleylacetoacetate et  $\Delta^5$ -3-kétosteroides (Singhal, Piper et al. 1996).



**Figure 3. Réactions catalysées par les GSTs.** Exemples de conjugaison, réduction, thiolase, et isomérisation catalysées par les GSTs. Les substrats suivants sont présentés. a) CDNB, b) sulphoraphane, c) CuOOH, d) 4-nitrophenyl acetate, e) trinitroglycérine, f) maleylacetoacetate, et g) PGH<sub>2</sub> (Hayes, Flanagan et al. 2004).

Les GSTs cytosoliques représentent un champ de recherche en développement pour les pharmacologues et les toxicologues parce qu'elles fournissent des cibles pour la prévention et le traitement de l'asthme et des cancers. De plus, les GSTs métabolisent les agents chimiothérapeutiques, les insecticides, les herbicides et les produits générés par les réactions d'oxydoréduction (Hayes, Flanagan et al. 2004). Dans le cadre de la présente étude, nous étudions quatre de ces GSTs cytosoliques. Les GSTs cytosoliques humaines sont des dimères composés de sous-unités de 199 à 244 acides aminés. Sept familles de gènes de GST ont été décrites et désignées :  $\alpha$ ,  $\mu$ ,  $\pi$ ,  $\sigma$ ,  $\theta$ ,  $\omega$  et  $\zeta$  ou A, M, P, S, T, O et Z (Hayes and Pulford 1995). Les transférases portent donc des noms tels

GSTM3 ( pour la 3<sup>e</sup> transférase de la famille M ) et GSTP1 ( pour une transférase de la famille P ). Au moins, 17 GSTs cytosoliques existent chez l'humain, le tableau I (page 10) présente les GSTs découvertes à ce jour et leurs substrats.



Tableau I. Les GSTs et leurs substrats

Famille	Classe, enzyme	Substrats ou réactions
Cytosolique	Alpha, A1	$\Delta 5$ -ADD, BCDE, BPDE, Busulfan, Chlorambucil, DBADE, DBPDE, BPhDE, N-a-PhIP
	Alpha, A2	CuOOH, DBPDE, 7-chloro-4-nitrobenz-2-oxa-1,3-diazole
	Alpha, A3	$\Delta 5$ -ADD, $\Delta 5$ -pregnene-3,20-dione, DBPDE
	Alpha, A4	COMC-6, EA, 4-hydroxynonanal, 4-hydroxydecenal
	Alpha, A5	inconnu
Cytosolique	Mu, M1	<i>trans</i> -4-phenyl-3-buten-2-one, BPDE, CDE, DBADE, <i>trans</i> -stilbene oxide, styrene-7,8-oxide
	Mu, M2	COMC-6, 1,2-dichloro-4-nitrobenzene, aminochrome, dopa <i>O</i> -quinone, PGH <sub>2</sub> →PGE <sub>2</sub>
	Mu, M3	BCNU, PGH <sub>2</sub> →PGE <sub>2</sub>
	Mu, M4	CDNB
	Mu, M5	Faible réaction avec CDNB
Cytosolique	Pi, P1	acroléine, base propenals, BPDE, CDE, Chlorambucile, COMC-6, EA, Thiotepe
Cytosolique	Sigma, S1	PGH <sub>2</sub> →PGD <sub>2</sub>
Cytosolique	Theta, T1	BCNU, époxyde de butadiène, CH <sub>2</sub> Cl <sub>2</sub> , EPNP, oxyde d'éthylène
	Theta, T2	CuOOH, sulfate ménaphthyl
Cytosolique	Zeta, Z1	dichloroacétate, fluoroacétate, 2-chloropropionate, malelyacétoacétate
Cytosolique	Omega, O1	monomethylarsonic acid, dehydroascorbic acid
	Omega, O2	dehydroascorbic acid
Mitochondriale	Kappa, K1	CDNB, CuOOH, ( <i>S</i> )-15-hydroperoxy-5,8,11,13-eicosatetraenoic acid
MAPEG	gp I, MGST2	CDNB, LTA <sub>4</sub> →LTC <sub>4</sub> , ( <i>S</i> )-5-hydroperoxy-8,11,14- <i>cis</i> -6- <i>trans</i> -eicosatetraenoic acid
	gp I, FLAP	Liaison non-enzymatique de l'acide arachidonique
	gp I, LTC4S	LTA <sub>4</sub> →LTC <sub>4</sub>
MAPEG	gp II, MGST3	CDNB, LTA <sub>4</sub> →LTC <sub>4</sub> , acide ( <i>S</i> )-5-hydroperoxy-8,11,14- <i>cis</i> -6- <i>trans</i> -eicosatetraenoic
MAPEG	gp IV, MGST1	CDNB, CuOOH, hexachlorobuta-1,3-diene
	gp IV, PGES1	PGH <sub>2</sub> →PGE <sub>2</sub>

Abréviations: 5-ADD,  $\Delta 5$ -androstene-3,17-dione; BCDE, benzo[*g*]chrysene diol epoxide; BCNU, 1,3-*bis*(2-chloroethyl)-1-nitrosourea; BPDE, benzo[*a*]pyrene diol epoxide; BPhDE, benzo[*c*]phenanthrene diol epoxide; CDE, chrysene-1,2-diol-3,4-epoxide; COMC-6, crotonyloxymethyl-2-cyclohexenone; DBADE, dibenz[*a,h*]anthracene-diol-epoxide; DBPDE, dibenzo[*a,l*]pyrene diol epoxide; EA, ethacrynic acid; EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; N-a-PhIP, N-acetoxy-2-amino-1-methyl-6-phenylimidazo [4,5-*b*]pyridine. (Hayes, Flanagan et al. 2004)

**Tableau II. Localisation tissulaire des GSTs les plus étudiés**

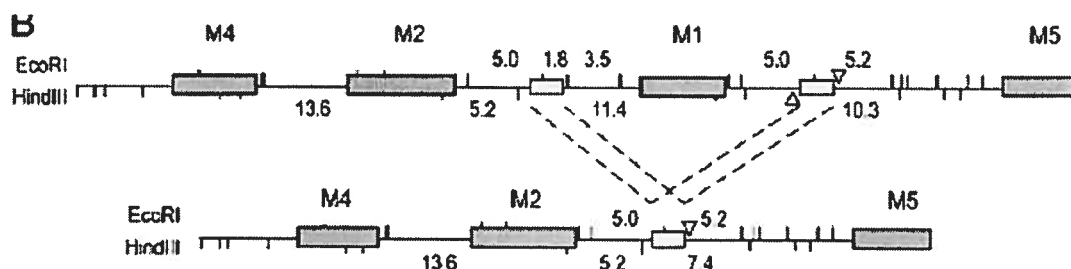
Enzymes GST	Alpha		Mu			Pi	Thêta		
	A1:2	M1-1	M2-2	M3-3	M4-4	M5-5	P1-1	T1-1	T2-2
Erythrocyte	-	-	-	-			±	+	
Cerveau	±	+	+	++	+	+	++	-	
Vessie	+	+					++		
Peau	++	±	-	-			++	-	
Coeur	±	+	(+)	(+)	(+)	(+)	++	-	
Testicule	++	+	+	++	++	+	+		
Foie	++	+	(+)	(±)	+	-	(+)	+	+
Poumon	+	+	(+)	(+)	(+)	+	++	-	
Lymphocytes		+							
Estomac/intestin	+	±					+		
Muscle	-	+	++	-	+	-	++	-	
Rein	+	+	(+)	-	+	-	+	-	
Ovaire	+				+		+		
Pancréas	+	(+)			+	-	+		
Placenta	-	-	-	-	+	-	++	-	
Prostate	+						++		
Utérus	+				+		+		

++ : expression forte ; + : expression moyenne ; (+) : expression faible ; - : pas d'expression.

(Habdous, Siest et al. 2004)

### 2.2.1 La famille M – GSTM3 et GSTM4

Les cinq gènes de la famille M sont situés en tandem dans un cluster de 20kb sur le chromosome 1p13.3 (Xu, Wang et al. 1998). GSTM4-GSTM2-GSTM1-GSTM5 sont très rapprochés, alors que GSTM3 se trouve plus loin sur le chromosome 1. Nous étudions deux de ces gènes soit M3 et M4. La séquence de ces derniers est relativement courte, 3798 pb pour GSTM3 et 5584 pb pour GSTM4. Quarante pour cent des chromosomes possèdent une délétion de 15kb qui occasionne une perte du gène GSTM1 en entier. De part et d'autre de cette délétion, il y a des séquences partageant plus de 98% d'identité sur 4 kb ( figure 4 page 12 ), ce qui fait penser qu'une recombinaison en est à l'origine. Cette délétion, lorsqu'elle est présente sur les deux chromosomes, engendre le génotype GSTM1 nul qui a été associé à plusieurs maladies (Strange, Spiteri et al. 2001; Krajinovic, Labuda et al. 2002).

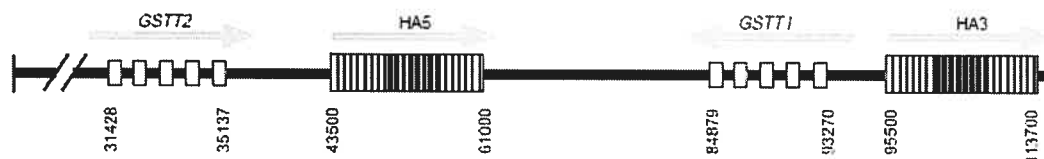


**Figure 4. Le Tandem GSTM et la délétion de GSTM1 :** Les rectangles gris pâle représentent les séquences avec un pourcentage d'identité très élevée. Le chromosome du haut a le gène GSTM1 alors que celui du bas a subi une délétion suite à une recombinaison (Xu, Wang et al. 1998).

Les protéines GSTM3 et GSTM4 agissent sous la forme d'homodimères. L'expression de GSTM4 est presque ubiquitaire alors que celle de GSTM3 se limite au cerveau et aux testicules (tableau II page 11) ce qui laisse croire que GSTM3 pourrait agir au niveau des barrières hémato-cérébrale et hémato-testiculaire.

### 2.2.2 La famille T – GSTT1

Dans la famille T, nous étudions GSTT1 qui a la particularité d'être délété chez certains individus. La classe thêta comprend deux protéines, GSTT1 et GSTT2 qui partagent 55% d'identité de séquence d'acides aminés. Elles ont toutes les deux été impliquées dans la carcinogénèse. Le gène GSTT1 est situé au locus 22q11.2 sur 8129 pb à approximativement 50 kb de GSTT2 (Strange, Spiteri et al. 2001). Les protéines traduites agissent sous forme d'homodimères. GSTT1 est exprimé dans les érythrocytes, faiblement dans les cellules hépatiques et dans les poumons, au niveau des cellules ciliées à la jonction broncho-alvéolaire (Strange, Spiteri et al. 2001). Le gène GSTT1 et son activité sont absents de 20 % de la population. Le phénotype GSTT1+ (présence du gène) peut catalyser la conjugaison de GSH au dichlorométhane. La délétion est probablement due à un événement de recombinaison puisque de part et d'autre de GSTT1, il y a des régions avec plus de 90% d'identité de séquence. La taille de cette délétion est d'environ 54kb (Sprenger, Schlagenhauer et al. 2000).



**Figure 5. La région du gène GSTT1.** Deux séquences hautement spécifiques HA5 et HA3 ( avec plus de 90% d'identité de séquence sur 18 kb ) se situent de part et d'autre de GSTT1. GSTT2 est adjacent à l'une de ces séquences. HA3 et HA5 sont des répétitions directes (Sprenger, Schlagenhauser et al. 2000).

### 2.2.3 La famille Pi – GSTP1

Dans la famille P, nous étudions la transférase la plus ubiquitaire soit GSTP1. La protéine GSTP1 est exprimée dans plusieurs tissus humains incluant la moelle osseuse et le sang. Il s'agit de la GST la plus exprimée dans les tissus fœtaux (McCarver and Hines 2002) et adultes (tableau II page 11). L'enzyme GSTP1 peut métaboliser plusieurs produits cancérigènes incluant le *benzopyrenediol epoxide* et l'*acroleine* (Hayes and Pulford 1995). Le gène de 2839 pb est localisé au locus 11q13.2.

## 2.3 Les peroxydases

La famille des peroxydases de glutathion dépendantes au sélénium (GPX) comprend quatre membres connus qui partagent certaines caractéristiques. Les GPXs catalysent la réduction du peroxyde d'hydrogène, des hydroperoxydes organiques et des peroxydes de lipides par la GSH et contribuent à la protection des cellules contre la toxicité des oxydants. Les GPXs dépendent du sélénium pour leur activité enzymatique (Rotruck, Pope et al. 1973). Elles possèdent un acide aminé sélénocystéine inséré lors de la traduction et qui confère l'activité enzymatique (Forstrom, Zakowski et al. 1978). Elles participent à une réaction lors de laquelle le sélénium dans la sélénocystéine réduit les peroxydes, s'en suit l'oxydation du sélénium suivit par sa réduction par le GSH (Awasthi, Beutler et al. 1975).

C'est en 1957 (Mills 1957) que la première activité GPX fut décrite et on lui attribua la fonction de protection des érythrocytes contre l'hémolyse par oxydation. Cette enzyme fut originalement appelée *glutathione peroxydase* puis GPX classique, mais on la connaît aujourd'hui sous le nom de GPX1. GPX1 peut métaboliser le peroxyde d'hydrogène et une panoplie de peroxydes organiques incluant le cholestérol et les peroxydes acides gras à longue chaîne (Arthur 2000).

### 2.3.1 GPX2

La seconde forme de GPX est localisée dans le cytosol des cellules hépatiques et gastro-intestinales d'où sont nom original *Gastrointestinal glutathione peroxidase*. Elle est connue aujourd'hui sous le sigle GPX2. L'enzyme agit sous la forme d'un homotétramère. Elle possède 65% d'identité de séquence d'acide aminé et 60% d'identité de séquences nucléotidiques avec GPX1 (Chu, Doroshov et al. 1993). Il s'agit de l'enzyme responsable de la majorité des activités de réduction du peroxyde d'hydrogène GSH-dépendante. Elle pourrait jouer un rôle prépondérant dans la protection des mammifères contre la toxicité des hydroperoxydes organiques dans l'alimentation. Sa séquence de 3606 pb a été localisée à la position 14q23.3 (Chu, Rohan de Silva et al. 1996).

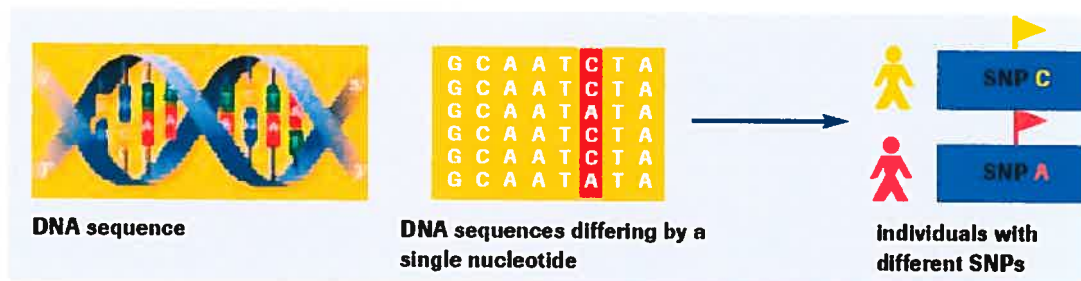
### 2.3.2 GPX3

Le gène GPX3 couvre 8360 kb dans la région q32 du chromosome 5 (Chu, Esworthy et al. 1992). GPX3 est l'unique isoforme extracellulaire de GPX. On l'appelle d'ailleurs plasma GPX. Son activité a également été identifiée dans le fluide pulmonaire et dans le lait maternel (Avissar, Slemmon et al. 1991; Avissar, Finkelstein et al. 1996). La protéine est un tétramère glycosylé possédant des sous-unités de 23 kb qui catalyse la réduction du peroxyde d'hydrogène et des peroxydes lipides par le glutathion (Takebe, Yarimizu et al. 2002). GPX3 est l'un des épurateurs majeurs des ROS. La protéine GPX3 est excrétée basolatéralement à partir des cellules tubulaires proximales du rein et des cellules pariétales épithéliales des capsules de Bowman (Whitin, Bhamre et al. 2002).

### 3 LA VARIABILITÉ

#### Les variants

Le plus grand des enjeux de la génétique moderne est d'associer des variations dans les séquences d'ADN avec des phénotypes précis et héréditaires. Ces variations incluent les RFLPs, les minisatellites, les microsatellites, les *single nucleotide polymorphisms* (SNPs) ainsi que les délétions ou les insertions de une ou plusieurs paires de bases. Ces variations sont transmises de générations en générations et représentent une grande proportion des différences qui existent entre deux personnes (Chakravarti 2001; Stoneking 2001) (figure 6 page 15). Les estimations actuelles montrent que les SNPs avec une fréquence d'allèle mineur d'au moins 1% se trouvent à un taux de un toutes les 200-300 pb dans le génome. L'extrapolation au génome entier suggère que 15 millions de SNPs ou plus seraient disponibles pour des études. Actuellement, quelques millions de polymorphismes sont déjà décrits dans des bases de données publiques (Salisbury, Pungliya et al. 2003).



**Figure 6. La variabilité due aux polymorphismes**

([www.roche.com/pages/facets/22/snps\\_e.pdf](http://www.roche.com/pages/facets/22/snps_e.pdf))

Dans le cadre de ce projet, la région étudiée est courte ( 2 kb ) et les polymorphismes étudiés sont surtout les SNPs et les insertions/délétions. En fonction de la région d'occurrence des polymorphismes, leurs effets diffèrent. Moins de 10% du génome humain est constitué de gènes, la majorité de ces derniers est donc située dans la portion non-codante et donne naissance à des variations neutres.

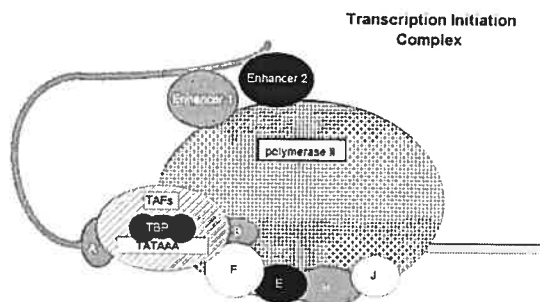
Les polymorphismes des régions codantes sont des variations qui ont lieu directement dans les exons d'un gène. Parmi ceux-ci un tiers n'ont pas d'effet puisqu'il donne lieu à des changements synonymes, c'est-à-dire à la transcription d'un même acide aminé. Le reste des polymorphismes de région codante peuvent avoir un impact majeur sur la fonction du gène s'ils occasionnent un changement d'acide aminé dans la séquence protéique. Ils peuvent alors altérer le repliement de la protéine, affecter un site de modifications post-traductionnelles, changer les sites de liaison ou le site catalytique, etc.

Les polymorphismes des régions régulatrices se trouvent dans les régions responsables de la régulation de l'expression génique, par exemple le promoteur. Ces polymorphismes peuvent modifier la liaison de l'ADN polymérase, faire disparaître ou apparaître des sites de transcription. En somme, ils peuvent altérer la concentration ou la présence d'ARNm et donc de protéines. C'est à ces polymorphismes en particulier que nous nous intéressons dans le cadre du projet. Ils ont été peu étudiés jusqu'à maintenant à cause de la difficulté à identifier leur impact précis puisqu'ils ne provoquent pas de changement de la structure de la protéine. De plus, les polymorphismes des régions régulatrices comme ceux des régions codantes paraissent les plus importants au point de vue médical. Ils forment la base de la pharmacogénétique qui s'occupe d'étudier l'influence des variations génétiques sur la tolérance et l'efficacité des médicaments.

## 4 LA RÉGULATION

Dans le cadre de ce projet, nous étudions la variation au niveau des polymorphismes des régions régulatrices. Les variants se situent donc au niveau de la régulation des gènes et non dans la structure de la protéine. La région spécifiquement étudiée correspond au promoteur du gène. Afin d'uniformiser l'étude, le « promoteur » a été défini comme la séquence allant de 2 kb en amont jusqu'au site d'initiation de la transcription.

Les promoteurs peuvent être considérés comme les interrupteurs de la transcription. La cascade d'événements qui précède l'activation du gène doit mener la machinerie transcriptionnelle au cœur du promoteur (*core promotor*). Ce dernier est donc la cible ultime de tous les facteurs impliqués dans la régulation de la transcription par la polymérase II (Butler and Kadonaga 2002). Le cœur du promoteur est la longueur minimale de séquence continue d'ADN qui est suffisante pour diriger une initiation exacte de la transcription par la machinerie de l'ARN pol II (Smale, Jain et al. 1998). Habituellement, le cœur du promoteur comprend le site de l'initiation de la transcription et s'étend en amont ou en aval sur 35 nucléotides. La plupart du temps, le cœur du promoteur ne mesure que 40 nucléotides. Le cœur du promoteur est le site d'action de l'ARN polymérase II. Cette dernière est une enzyme composée de plusieurs sous-unités qui catalyse la synthèse de l'ARNm à partir de l'ADN matrice. Pour être exacte et efficace, la transcription à partir du cœur du promoteur nécessite la liaison de la polymérase II en plus des facteurs auxiliaires qu'on appelle facteurs de transcription généraux (TF) et qui incluent TFIIA, B, D, E, F et H (Figure 7 page 17).



**Figure 7. Complexe d'initiation de la transcription.**

<http://www.dur.ac.uk/biological.sciences/Staff/Croy/GENNET4.HTML>



Le cœur du promoteur possède plusieurs motifs ( figure 8 page 19 ). Chacun de ces éléments se trouve dans certains mais pas tous les promoteurs (Butler and Kadonaga 2002).

**Motif TATA :** Il s'agit du premier élément du cœur du promoteur à être identifié (Breathnach and Chambon 1981). Chez les métazoaires, le motif TATA est situé 25 à 30 nucléotides en amont du site du début de la transcription. Chez les humains, on a démontré que 32 % de 1031 séquences promotrices potentielles contenaient un motif TATA (Suzuki, Tsunoda et al. 2001). TBP ( pour *TATA box Binding protein* ) la sous unité de TFIIB est la protéine prédominante se liant au motif TATA.

**Initiateur (Inr) :** Le Inr comprend le site du début de la transcription. Il a été décrit fonctionnellement dans le cadre d'une analyse à grande échelle du cœur du promoteur du gène de la déoxynucléotidyltransférase terminale (TdT) chez la souris (Smale, Jain et al. 1998). La séquence consensus pour Inr dans les cellules mammifères est Py-Py-A+-N-T/A-Py-Py (Corden, Wasylyk et al. 1980). Plusieurs facteurs se lient au Inr comme par exemple TFIID et ses sous-unités TAF<sub>II</sub>150 et TAF<sub>II</sub>250 (Kaufmann, Ahrens et al. 1998).

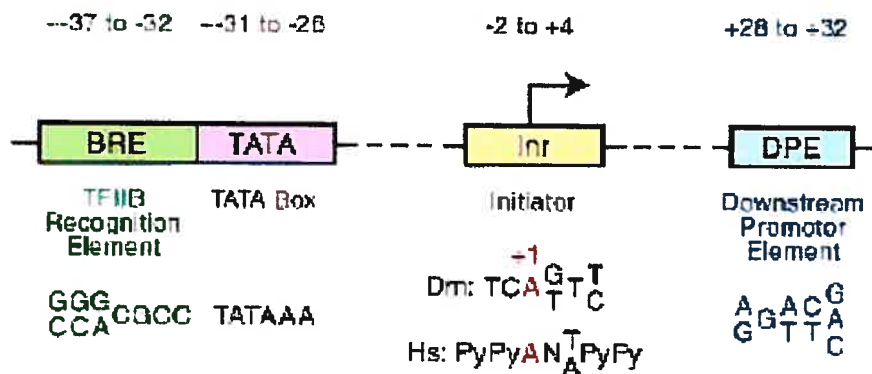
**Downstream core promoter element (DPE) :** Le DPE a été identifié comme un site de liaison pour la protéine TFIID de la drosophile. TFIID se lie à Inr et au DPE et une mutation chez l'un ou l'autre empêche la liaison de TFIID au promoteur (Burke and Kadonaga 1996). Le DPE se retrouve le plus souvent dans les promoteurs qui n'ont pas de motif TATA.

**L'élément de reconnaissance TFIIB (BRE) :** BRE est un site de liaison de TFIIB situé en amont de certains motifs TATA (Lagrange, Kapanidis et al. 1998).

**Les îlots CpG :** Ce sont des séquences riches en GC qui sont généralement non-méthylées et qui mesurent de 0,5 à 2 kb. La plupart du temps, les îlots CpG ne possèdent pas de motifs TATA ou de DPE mais ils peuvent être liés par différents facteurs de transcription (Macleod, Charlton et al. 1994).

De plus, il existe d'autres séquences d'ADN en cis qui régulent la transcription par l'ARN polymérase II : le promoteur proximal, les *enhancers*, les *silencers*, et les

*boundary/insulator elements* (Blackwood and Kadonaga 1998; West, Gaszner et al. 2002). Le promoteur proximal est la région directement autour du site du début de la transcription de  $-250$  à  $+250$  environ. Une partie de sa séquence est donc incluse dans la région étudiée (2 kb en amont du site du début de la transcription). Un seul variant dans cette séquence ou dans celles mentionnées plus haut pourrait potentiellement affecter la transcription du gène et donc sa fonction.



**Figure 8. Éléments de séquence du cœur d'un promoteur**

## **5 RELATIONS ENTRE LE MÉTABOLISME DES XÉNOBIOTIQUES, LA MALADIE ET LA RÉPONSE THÉRAPEUTHIQUE**

Les mécanismes de protection contre les xénobiotiques sont apparus avec les premières formes de vie terrestre. Au cours de l'évolution humaine ces mécanismes ont évolué en des systèmes de réactions à l'environnement, entre autres en réponse aux substances « étrangères » contenues dans la nourriture. L'exposition à des régimes alimentaires différents a favorisé la variabilité dans les enzymes responsables de la protection contre ces substances « étrangères ». Ces variations peuvent se traduire par le blocage de l'effet d'un gène ou son augmentation ou par des changements dans la structure des enzymes qui affectent leur efficacité ou l'identité de leurs substrats. Les variants génétiques peuvent affecter la susceptibilité individuelle aux infections, la réponse aux traitements, à la consommation d'alcool, à la vitesse à laquelle sont métabolisées certaines substances, etc.

Le glutathion, par son implication contre les oxydants contenus en grande quantité dans la nourriture et générés par les processus cellulaires normaux, est important pour le maintien de l'équilibre métabolique des organismes. Lorsque la biosynthèse des enzymes qui catalysent la formation du GSH (GSS,  $\delta$ -glutamylcystéine synthase) ne se fait pas correctement, cela conduit à une déficience métabolique grave qui se traduit le plus souvent par l'anémie hémolytique et des troubles neurologiques (Wellner, Sekura et al. 1974; Dahl, Pigg et al. 1997). De plus, des taux bas de GSH ont été associés avec plusieurs maladies comme le VIH, l'hépatite C, le diabète de type II, les colites ulcéreuses, les fibroses pulmonaires, le syndrome de détresse respiratoire chez l'adulte (ARDS) et les cataractes (Meister 1991; White A.C. 1994).

De toutes les associations entre GSTs et maladie, ce sont les génotypes nuls de GSTM1 et GSTT1 qui ont été le plus étudiées. De façon générale, ces génotypes sont considérés comme attribuant un risque élevé pour plusieurs cancers et maladies inflammatoires. Une étude de Moisió et al. (Moisió, Sistonen et al. 1998) a décrit une association entre eux et le cancer du colon chez des patients finlandais. Les génotypes nuls de GSTM1 et GSTT1

ont également été associés avec le cancer du poumon (Goto, Yoneda et al. 1996), le cancer des ovaires (Howells, Redman et al. 1998), l'arthrite rhumatoïde (Mattey, Hassell et al. 1999) et la leucémie lymphoblastique aiguë (Krajinovic, Labuda et al. 2002). Une autre étude démontre que le génotype nul de GSTM1 chez des enfants souffrant de leucémie lymphoblastique aiguë est associé avec une meilleure chance de rémission, possiblement à cause d'un métabolisme plus lent des médicaments chimiothérapeutiques (Hall, Autzen et al. 1994). Les autres GSTs sont moins bien étudiées, mais on a quand même pu associer des polymorphismes chez ces dernières avec l'augmentation de la susceptibilité au cancer et à certaines maladies inflammatoires comme l'asthme, les allergies, l'athérosclérose, l'arthrite rhumatoïde et la sclérose systémique (Hayes, Flanagan et al. 2004). Les peroxydases sont aussi les sujets d'étude d'association. On a relié une déficience de l'enzyme GPX3 avec des maladies coronariennes et avec les crises cardiaques (Bierl, Voetsch et al. 2004).

Un polymorphisme fonctionnel affectant la régulation d'un gène du métabolisme du glutathion a donc de grandes chances d'avoir un impact sur le développement de certaines maladies et sur le métabolisme des drogues et médicaments. La caractérisation de la distribution des polymorphismes dans les gènes du métabolisme des xénobiotiques permettra ultérieurement des études d'association et l'identification de variants qui influencent le développement et influenceront potentiellement le traitement des maladies. Le supplément de novembre 2004 de *Nature Genetics* intitulé *Genetics for the Human Race* discute de ce virage de la médecine. Présentement, l'ascendance ethnique est déjà un facteur déterminant dans la santé. Par exemple, les Afro-Américains ont une incidence deux fois plus grande de mourir d'une crise cardiaque et 10% plus de chance de mourir du cancer que les Européens, les Américains et les Asiatiques du sud-est. Par la même occasion, ils sont moins susceptibles de recevoir des services de santé préventifs et sont davantage susceptibles de souffrir d'un cancer avancé lors du diagnostic (White A.C. 1994; Tate and Goldstein 2004). À ce jour, vingt-neuf médicaments ou combinaisons de médicaments ont été cités dans des articles comme ayant des différences dans l'efficacité ou la présence d'effets secondaires chez divers groupes ethniques. De 1995 à 1998, les étiquettes de 8 % des nouveaux médicaments

aux États-Unis contenaient une mention par rapport à l'ascendance ou l'ethnie (Tate and Goldstein 2004). Étant donné que le glutathion joue un rôle importants dans le métabolisme des xénobiotiques, les enzymes qui contrôle son propre métabolisme peuvent être impliqués dans l'étiologie de la maladie et/ou de la réponse variée au cours d'un traitement. Les variants génétiques des gènes sous-jacents influencent l'activité de ces enzymes et sont distribués différemment parmi les populations. Étudier la distribution des variants dans les gènes du glutathion pour comprendre leurs impacts s'inscrit donc dans un courant très actuel.

## 6 LES PARAMÈTRES DE LA DIVERSITÉ

Le profil génétique actuel d'une population est la conséquence de son passé génétique et démographique. La diversité génétique prend son origine à travers les mutations et les recombinaisons. Plusieurs événements sont ensuite responsables de maintenir, de modifier ou de faire disparaître cette diversité. Ces événements laissent des traces génétiques. Certains paramètres permettent de quantifier la diversité génétique et d'identifier les événements évolutifs dont elle provient. Ces paramètres sont appelés statistiques descriptives. Le nombre de sites polymorphiques (S) et le nombre d'haplotypes (k) sont des paramètres directement identifiables. Le paramètre  $\pi$  est la diversité nucléotidique ou l'hétérozygotie moyenne par nucléotide. Il est calculé à partir des fréquences des polymorphismes. Le paramètre G est la diversité haplotypique ou l'hétérozygotie et se calcule à partir de k.

Le paramètre  $F_{ST}$ , aussi appelé indice de fixation, mesure la réduction d'hétérozygotie liée aux différences de fréquences alléliques entre populations. C'est un indice de la diversification et de subdivision des populations. Si les populations sont génétiquement indifférenciées,  $F_{ST}$  est nul.

$F_{ST} = (H_T - H_S) / H_T$  où  $H_T$  est l'hétérozygotie totale de la population sans division en sous-populations et  $H_S$  est l'hétérozygotie moyenne des sous-populations étudiées.

$F_{ST}$  est donc toujours positif et d'une valeur comprise entre 0 et 1, puisque  $H_S$  ne peut être plus grande que  $H_T$ . L'hétérozygotie peut être représentée par  $\pi$  ou G.

La subdivision des populations peut s'effectuer par dérive génétique c'est-à-dire des fluctuations dues au hasard causées par des variations dans la contribution de chaque individu à la génération suivante. Elle peut aussi être causée par des forces sélectives sur un groupe d'individus. Les goulots d'étranglement – des réductions dans la diversité génétique qui proviennent d'une réduction majeure dans la taille de la population – et l'effet fondateur – la réduction dans la diversité génétique dans une population fondée par un petit nombre d'individus – peuvent également modifier la diversité dans les

populations. La croissance démographique et le métissage sont deux autres facteurs qui contribuent aux variations dans le profil génétique des populations (Jobling MA 2004).

La sélection naturelle est un facteur déterminant dans l'évolution. Elle se définit par les contributions différentielles des individus à la génération suivante sur la base de leur aptitude à survivre. Le lien entre les fréquences élevées d'anémie falciforme en Afrique et le paludisme (Livingstone 1976) sont un des exemples les plus connus de sélection pour le maintien d'un allèle. La sélection naturelle agit sur la quantité et la forme de variabilité maintenue dans la population. Étant donné la quantité de facteurs qui affectent les séquences nucléotidiques, il est difficile de quantifier le rôle de la sélection.

Les tests de neutralité peuvent indiquer un écart à la neutralité symptomatique de la présence de forces sélectives. Ces tests se basent sur le modèle de l'évolution neutre. Ce dernier stipule qu'en l'absence de sélection, différents estimateurs de la diversité devraient être égaux. Le paramètre  $\Theta$  est un estimateur de la diversité et le produit du taux de mutation et de l'effectif efficace. Le test de Tajima D (Tajima 1989) compare deux estimateurs  $\Theta$ ,  $\Theta_S$  calculé à partir du nombre de sites polymorphiques et  $\Theta_\pi$  calculé à partir de la diversité nucléotidique. Si aucune force de sélection ne vient influencer les fréquences alléliques, on s'attend à ce que Tajima D soit 0. Un autre test de neutralité a été développé par Fu et Li (Fu and Li 1993). Ce test compare l'estimateur  $\Theta_\pi$  avec  $k$ , le nombre de polymorphisme et un résultat nul indiquerait également une absence de sélection. Le test de Fu est plus puissant que celui de Tajima parce que les paramètres comparés le rendent plus sensible à la déviation de la neutralité (Fu and Li 1993).

Tous ces paramètres et tests qui permettent de quantifier et de comparer la diversité dans les populations ont été calculés dans le cadre de ce projet. Nous avons également identifié des traces de phénomènes évolutifs à l'origine de la diversité tout cela, afin retracer l'histoire démogénétique des populations.

## 7 OBJECTIFS DU PROJET

Les objectifs généraux qui touchent le projet *Regulatory Genetics* de Génome Québec sont de :

- Caractériser la variabilité dans des régions promotrices
  - définies arbitrairement à 2 kb en amont du début du site de la transcription
- Trouver des polymorphismes qui auraient un impact sur la régulation des gènes.

Les objectifs particuliers de mon projet de maîtrise sont de :

1. Caractériser la variabilité des gènes du métabolisme du glutathion car ce sont des bons gènes candidats pour étudier la susceptibilité génétique aux maladies et à la réponse thérapeutique.
  
2. Étudier la diversité des régions promotrices au niveau de différentes populations mondiales.
  1. En sélectionnant des échantillons appropriés
  2. En effectuant le génotypage
  3. En extrapolant les statistiques descriptives
  
3. Analyser les mécanismes sous-jacents afin de mieux comprendre l'histoire démographique des populations choisies.



## **Chapitre II**

### **MATERIEL ET METHODES**

## 1 ÉCHANTILLONS

Pour ce projet, deux cent vingt-deux échantillons humains ont été utilisés. Dans un premier temps, le génotypage a été effectué sur un groupe de quatre-vingts individus provenant des cinq continents, soit des Européens, des Asiatiques, des Africains, des Amérindiens et des Moyen-Orientaux. Ces échantillons proviennent en partie de la collection Coriell ( Corriell Institute for Medical Research ), les autres proviennent d'individus qui ont accepté de donner un échantillon de sang et certaines données démographiques (sexe, origine ethnique, etc) sur une base anonyme. Les origines de ces échantillons ont été établies en fonction de l'ascendance déclarée par chacun des individus participants. Deux populations particulières, des Islandais (  $n = 8$  ) et des Aborigènes de Taiwan (  $n = 8$  ), ainsi que des échantillons provenant de Québécois aux origines diverses ont été utilisés. L'étude comprend plusieurs groupes de Gaspésiens: des Canadiens-Français (  $n = 18$  ), des Loyalistes (  $n = 18$  ), des Acadiens (  $n = 18$  ), des descendants d'immigrants des îles anglo-normandes de Jersey et de Guernesey (  $n = 18$  ). Des échantillons de Montréalais de souche canadienne-française (  $n = 36$  ) et d'habitants de Charlevoix et du Saguenay-Lac Saint-Jean (  $n = 18$  ) font aussi partie du projet. Le tableau III (page 28) montre la liste et l'origine des échantillons utilisés. Ce projet a été accepté par le comité d'éthique institutionnel et un formulaire de consentement éclairé a été signé par chacun des participants. Quatre échantillons d'ADN de grands singes, un gibbon, un orang-outan, un gorille et un chimpanzé, ont également été génotypés.

Dans des buts de collaboration, nous avons utilisé une cohorte spéciale pour le génotypage de GPX3 et GPX2. Nous avons utilisé en plus de tous ceux mentionnés précédemment 88 échantillons supplémentaires qui avaient déjà fait l'objet d'une étude sur la variabilité dans les exons et les introns. Le tableau XXXX en annexe B (page xvii) présente la liste et l'origine de ces échantillons.

Tableau III. Origine des échantillons d'ADN.

Origine		Nombre	Provenance Commerciale
<b>Afrique sub-saharienne</b>	Pygmées Biaka	3	Coriell
	Pygmées Mbuti	1	Coriell
	Afro-Américains	4	CoriellHD50
	Afrique sub-saharienne	8	Coriell HD12
<b>Amérique</b>	Andes et Amérique du sud	8	CoriellHD17
	Amérique du sud, différentes tribus	8	CoriellHD18
<b>Asiatique du Sud-est</b>	Cambodge	3	
	Vietnam	2	
	Chine	3	
	Chine	8	Coriell HD32
<b>Europe</b>	Bulgarie	1	
	Pologne	1	
	Belgrade	1	
	Allemagne	1	
	Québec	4	
	Hongrie	4	Coriell HD26
	Russie	4	Coriell HD23
<b>Proche-Orient</b>	Juifs Ashkénazes	2	
	Juifs Sépharades	2	
	Arabes d'Égypte	2	
	Arabes de Syrie	1	
	Arabes du Liban	1	
	Indo-Pakistanaï	8	Coriell HD03
<b>Taiwan</b>	Aborigènes tribu Ami	8	Coriell HD25
<b>Islande</b>	Islande	8	Coriell HD30
<b>QUÉBEC</b>	Gaspésie : Acadiens	18	
	Canadiens-Français	18	
	Loyalistes	18	
	Jersey et Guernesey	18	
	Charlevoix et Saguenay-Lac St-Jean	18	
	Montréal ( Canadiens-Français )	36	
<b>GRANDS SINGES</b>		4	

Les 40 échantillons en gras ont été utilisés pour détecter les polymorphismes par dHPLC. La liste de chacun des échantillons utilisés se trouve en annexe A (page xv).

## 2 SÉQUENCES

Les séquences ont été choisies arbitrairement à deux mille paires de bases en amont du site d'initiation de la transcription. Le tableau IV (page 29) montre la longueur et la position des séquences étudiées.

**Tableau IV. Longueur et position des séquences étudiées**

Gène	Longueur du fragment (pb)	Chromosome	Position selon UCSC Blat	Sens
GSTT1	2116	22	22708820 - 22710935	F
GSS	2013	20	34259196 - 34261288	R
GSTP1	2098	11	67124435 - 67126682	F
GSTM3	1989	1	109581589 - 109583000	R
GSTM4	1824	1	109495652 - 109497000	R
GPX2	2080	14	63399464 - 63401621	F
GPX3	2228	5	150426201 - 150428000	F

Le sens correspond à l'orientation du gène sur le chromosome. F = avant (*forward*), R=inverse (*reverse*)

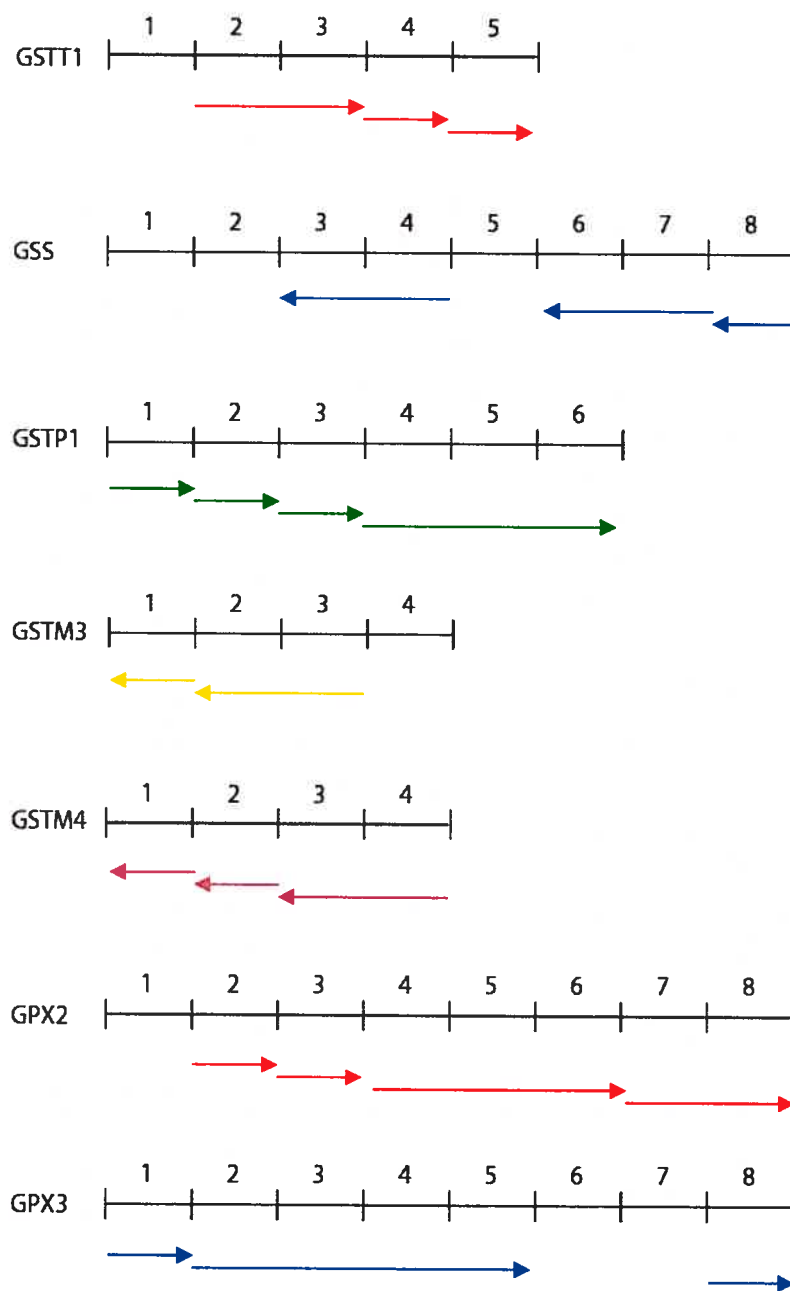
### 3 AMPLIFICATION PCR

Les amplifications ont été effectuées en plaque de 96 puits dans un volume de 30  $\mu$ l de réaction contenant de 7,5 ng d'ADN, 0,1 mM de dNTP (chaque), 1,2 mM de  $MgCl_2$ , 0,42  $\mu$ M de chaque amorce, 3  $\mu$ l de tampon ( 20mM Tris-HCl ( pH 8,4 ) et 50 mM KCl ) et 0,18  $\mu$ l ( ou 0,9 U ) d'ADN polymérase (Platinum Taq Invitrogen utilisée selon les normes du fabricant ). Pour le fragment GSTP1-1, l'amplification a eu également lieu dans un volume de 30  $\mu$ l de réaction avec les réactifs suivant : 10 ng d'ADN, 0,23 mM de dNTP, 0,42  $\mu$ M de chaque amorces, 3,3  $\mu$ l de tampon, 6,7  $\mu$ l de solution riche en GC et 0,27  $\mu$ l d'ADN polymérase (FastStart Taq Roche, utilisée selon les normes du fabricant). La dénaturation initiale de l'ADN se fait à 95°C pendant 5 minutes. Elle est suivie de cycles de dénaturation à 95°C, d'appariement et d'extension à températures et temps variant selon les amorces utilisées, et d'une extension finale de 10 minutes à 72°C, le tout dans un appareil GeneAmp PCR System 9700 (Applied Biosystems) .

Pour faciliter la détection des polymorphismes, la séquence de deux kb a été séparée en fragments de plus ou moins 300 pb. Dans ce projet, certains de ces fragments ont été combinés pour l'amplification (figure 9 page 32). Les séquences d'ADN formée de plus d'un fragment, ont été amplifiées en utilisant le protocole long, pour les autres séquences nous avons utilisé le protocole court (Tableau V page 31). La liste de toutes les amorces de PCR se trouve en annexe D (page xix). Le succès des amplifications est vérifié par électrophorèse sur gel d'agarose à 1%.

**Tableau V. Protocole de PCR utilisés**

	<b>Début</b>	<b>Touchdown</b>	<b>Amplification Principale</b>	<b>Finale</b>
<b>Protocole court</b>				
Dénaturation	95°C, 5 min	95°C, 30s	95°C, 30s	
Appariement		62°C-0,5°C par cycle jusqu'à 55°C, 30s	55°C, 30s	
Élongation		72°C, 30s	72°C, 30s	72°C, 7 min
Nb. de Cycles	1	14	30	1
<b>Protocole long</b>				
Dénaturation	95°C, 5 min	95°C, 30s	95°C, 30s	
Appariement		62°C-0,5°C par cycle jusqu'à 55°C, 1min	55°C, 1min	72°C, 7 min
Élongation		72°C, 1m15	72°C, 1m15	1
Nb. de Cycles	1	14	30	



**Figure 9. Fragments amplifiés ensemble et orientation des gènes.** Les chiffres représentent les fragments tels que définis dans la base de données THOR. Les flèches représentent les séquences amplifiées. Les fragments non-amplifiés ne sont pas polymorphiques.

## 4 HYBRIDATION ASO

Le génotypage a été effectué par hybridation ASO, cette technique requiert plusieurs étapes (Bourgeois and Labuda 2004).

### a. Transfert sur membranes

Les séquences d'ADN génomique amplifiées par PCR sont d'abord dénaturées. 180 µl de solution 0.4 N NaOH, 20mM EDTA, 0,02% orange II sont ajoutés dans chaque puit. Des aliquots de 100 µl sont par la suite transférés en duplicata sur des membranes de nylon (Hybond N+, Amersham) par aspiration (Minifold 1, Schleicher et Schuell). Les deux membranes sont placées dans un four à UV (Stratagene UV stratalinker 1800) et exposées à 120mJ de rayonnement UV afin de lier l'ADN à la membrane de façon covalente. Elles sont ensuite rincées.

### b. Marquage des sondes

La liste complète des sondes d'ASO utilisées lors de ce projet est placée en annexe E (page xxii). Elles mesurent typiquement 15 nucléotides et ont été synthétisées chez Invitrogen. Chaque sonde est marquée dans un tube séparé contenant 9 µl d'eau, 3 µl de tampon (*Forward 5x*), 1 µl de T4 kinase (Invitrogen), 1 µl de sonde à 25 µM et 1 µl de [ $\gamma^{32}$ ]ATP (10 µCi/ µl, Amersham). Le tube est incubé à 37°C de 30 minutes à 1 heure. L'ajout de 15 µl d'EDTA à 10mM arrête la réaction. Les sondes marquées sont ensuite conservées à -20°C jusqu'à utilisation.

### c. Hybridation

#### Préhybridation

Chacun des deux exemplaires de membrane est placé dans un tube à hybridation différent. Un volume adéquat de solution d'hybridation est ensuite ajouté – environ 30 ml par tube pour 3 ou 4 paires de membrane. La solution d'hybridation est composée de SSPE 1X (150 mM NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM EDTA, pH 7.4), 70mM Tris, SDS 1% et 100mg d'héparine. Par la suite, les tubes sont placés dans un four rotatif (Fisher *Isotemp dual – chamber incubator*) qui a été préchauffé une température environ 10°C supérieure à la température d'hybridation ( $T_m$ ) des sondes. Puisque que les  $T_m$ s sont variables en fonction des sondes et des conditions d'expérimentations et pour permettre



de faire plusieurs hybridation à la fois, nous faisons un changement de température appelé *touchdown*, c'est-à-dire que l'hybridation ne se fait pas à une température précise mais dans un intervalle d'environ 10°C de plus que la  $T_m$  à 10°C de moins que la  $T_m$ . Toutes les  $T_m$ s sont indiquées en annexe E. La préhybridation est de plus ou moins 20 minutes.

### Hybridation

Lors de l'hybridation, le mélange sondes/compétiteurs est ajouté. Ce mélange est composé de sondes marquées et de sondes compétitrices non-marquées dans une proportion de 1 :15 soit 8 µl de sondes marquées à 0.83 µM et 4 µl de compétiteur à 25 µM. La température du four est maintenue 10 minutes, puis le four est réglé à la température inférieure du *touchdown*. Lorsque cette température est atteinte, les tubes sont laissés dans le four pendant 30 minutes pour compléter l'hybridation des sondes.

### Lavage

À la suite de l'hybridation, la solution d'hybridation contenue dans les tubes est jetée dans des contenants prévus pour recueillir le matériel radioactif. La même quantité de solution de lavage (SSPE 1%, SDS 0,1%) est ajoutée pour rincer les membranes et éventuellement défaire les liaisons non-spécifiques. Les membranes sont rincées une secondes fois, cette fois-ci dans le four à la température inférieure du *touchdown* de 10 à 15 minutes.

### Exposition

Les membranes sont ensuite placées dans un sac de plastique scellé puis dans une cassette d'exposition avec un film (Fuji Medical X-Ray 100NIF) en présence d'écrans amplificateurs. L'exposition se fait à -80°C pendant une nuit. Par la suite, les résultats sont prêts à être relevés.

### Décrochage

Avant d'utiliser à nouveau les membranes, les sondes doivent être décrochées dans une solution bouillante de SDS 1% pendant 5 minutes.

## 5 GÉNOTYPER LA DÉLÉTION GSTT1

Pour génotyper la délétion de GSTT1 la méthode de Sprenger et al. (Sprenger, Schlagenhauer et al. 2000) a été utilisée. Cette méthode consiste en une amplification PCR multiplex suivi d'une électrophorèse sur gel d'agarose à 1%. La réaction de PCR s'est fait dans un volume total de 10 µl avec 12,5 ng d'ADN. Les détails des autres réactifs se trouvent dans le tableau VI (page 35). Le tableau VII (page 35) indique la séquence des amorces pour la réaction ainsi que la longueur des fragments.

**Tableau VI. Composition de la réaction de PCR multiplex pour la délétion de GSTT1**

Produits	Quantité (ul)
eau	2,78
10X PCR Buffer (incluant 15mM MgCl <sub>2</sub> )	1
dNTP mix (10mM de chaque )	0,8
amorce 1 (25uM)	0,08
amorce 2 (25uM)	0,08
amorce 3 (25uM)	0,08
amorce 4 (25uM)	0,08
Hotstar Taq DNA polymerase (Qiagen)	0,1
ADN génomique ( 2,5 ng/ul)	5
<b>Volume total</b>	<b>10</b>

**Tableau VII. Amorces pour le PCR multiplex de la délétion de GSTT1**

Séquence	Position*	Longueur (pb)
CAGTTGTGAGCCACCGTACCC	52069 - 52089	1460
CGATAGTTGCTGGCCCCCTC	107779 - 107760	1460
CCAGCTCACCGGATCATGGCCAG	85457 - 85479	466
CCTTCCTTACTGGTCCTCACATCTC	85922 - 85898	466

\* selon GeneBank : AF240786. (Sprenger, Schlagenhauer et al. 2000)

La dénaturation préliminaire est de 15 minutes à 95°C suivi de 40 cycles composés de dénaturation ( 94°C, 30 s ), appariement (70°C, 30 s) et élongation (72°C, 60 s), et d'une élongation finale de 7 minutes à 72°C, le tout dans un appareil GeneAmp PCR System 9700 ( Applied Biosystems).

## 6 BASE DE DONNÉES ET LOGICIELS

Les résultats de génotypage sont compilés dans la base de données du projet *Regulatory Genetics*, THOR, qui permet à tous les membres du projet d'y avoir accès.

La base de données permet de générer un fichier entrant pour le logiciel Phase 2.0. (Stephens, Smith et al. 2001). Ce dernier permet de reconstruire les haplotypes.

Les statistiques descriptives ont été effectuées à l'aide du logiciel Arlequin 2.0 de Laurent Excoffier. Ce logiciel est disponible à l'adresse suivante : <http://lgb.unige.ch/arlequin/>.

Le logiciel Network 4.1.0.9 (<http://www.fluxus-engineering.com/sharenet.htm>) a servi à générer les réseaux haplotypiques. Le logiciel Statistica 6.0 a été utilisé pour les analyses de MDS ( *multidimensional scaling* ).

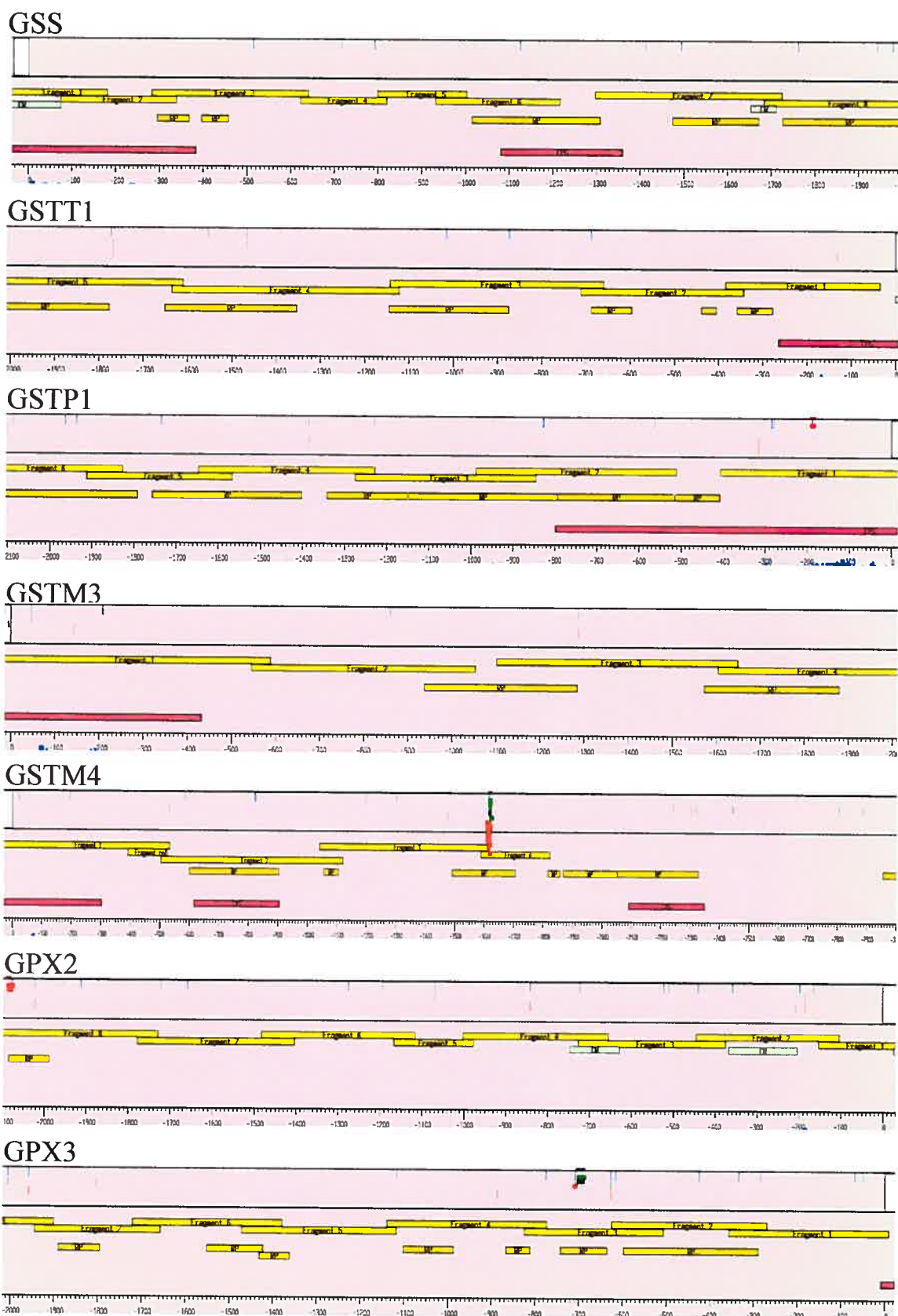
Les logiciels MaxHap et MaxDip (<http://genapps.uchicago.edu/maxdip/index.html>) ont été utilisés pour estimer les taux de recombinaisons.

**Chapitre III**

**RÉSULTATS**

## 1 Diversité génétique

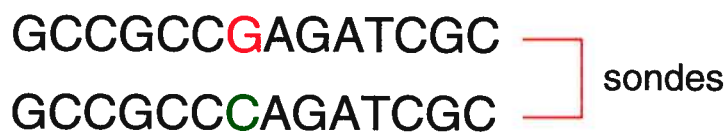
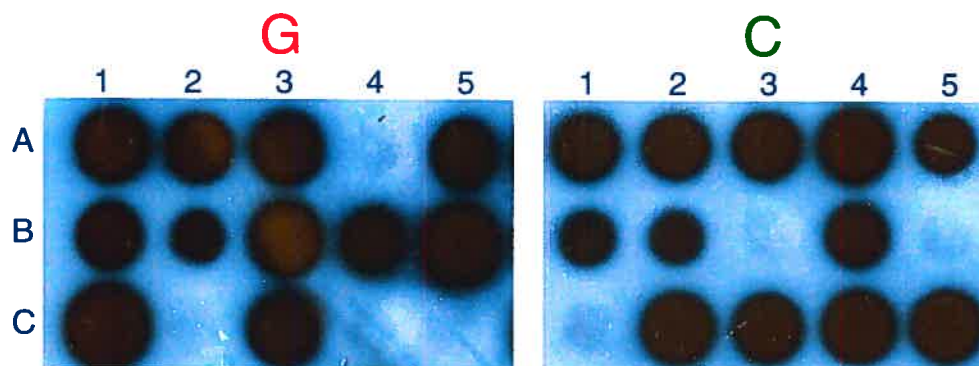
Les polymorphismes ont été détectés utilisant la technique de chromatographie dénaturante à haute pression (dHPLC) et séquençage sur le groupe de quarante individus mentionné dans le chapitre Matériel et Méthodes. On a détecté soixante-neuf polymorphismes dans les sept gènes. La figure 10 (page 39) représente les gènes et les positions des polymorphismes identifiés. Nous avons analysé la distribution géographique des 69 polymorphismes trouvés dans les segments d'ADN de 2 kb en amont des sites d'initiation de la transcription des sept gènes impliqués dans le métabolisme du glutathion qui font l'objet de cette étude. Ces 69 sites polymorphiques ont été trouvés par criblage de 40 individus par dHPLC suivi de séquençage en prenant le 2 kb comme « promoteur » ( voir titre de la figure 10 ).



**Figure 10. Polymorphisms dans les promoteurs.**

Traits bleu verticaux représentent les polymorphismes identifiés et étudié dans le cadre du projet. Traits verts et rouges verticaux identifient des polymorphismes déjà publiés. Rectangles rouges indiquent les régions riches en GC. Rectangles oranges correspondent à des régions répétées, Rectangles jaunes sont les fragments analysés.

Pour recenser les fréquences alléliques distribuées dans chaque population, nous avons effectué des hybridations ASO. La figure 11 (page 40) présente un exemple de résultats typiques d'ASO.



**Figure 11. Hybridation ASO pour G564C GSTP1**

Chaque point noir et point blanc représente un échantillon d'ADN. Pour l'échantillon A1, il y a un point noir sur les deux membranes, on en conclura que c'est un hétérozygote GC. Les échantillons A4 et C1 qui n'ont qu'un cercle noir sont des homozygotes CC et GG, respectivement.

Les résultats évalués par statistiques sont présentés pour chaque population. Ces statistiques permettent d'élaborer le profil de la variabilité pour chacun des gènes et pour chacune des populations analysées, pour pouvoir par la suite comparer les différents paramètres de variation. Le nombre de polymorphismes par segments de 2 kb est représenté par la lettre S. Notez que le nombre S a été déterminé à partir d'un échantillon de 40 individus, tandis que le nombre des haplotypes (k) peut différer entre l'échantillon initial et celui de 80 individus et plus ( qui inclut d'autres populations étudiées ) si des recombinaisons ont eu lieu. L'indice de diversité nucléotidique ( $\pi$ ) qui est calculé à partir des fréquences des polymorphismes est un autre paramètre analysé. On considère également le nombre d'haplotypes et l'indice de diversité haplotypique calculé à partir des fréquences haplotypiques. Le génotypage permet de déterminer le  $\pi$ , le k et donc d'effectuer les tests de neutralité qui dépendront de  $S/\pi/k$  comme Tajima ( dépend de S et  $\pi$  ) et  $F_u$  ( dépend de  $\pi$  et k ). Les tests de neutralité ( réalisés sur l'échantillon de 40 individus passé au dHPLC ) sont effectués afin de vérifier s'il y a des signes de sélection. L'expression démographique d'un goulot d'étranglement peut aussi affecter les résultats de ces tests. L'indice  $F_{st}$  total ( effectué sur l'échantillon de 80 individus ) montre la proportion de variabilité due aux différences entre les populations par rapport à la variabilité due aux différences dans les populations.

Les haplotypes ancestraux ont été obtenus à partir des résultats de génotypage de quatre primates non humains ( gibbon, orang-outan, gorille, chimpanzé ). Dans certains cas, les séquences de primates étaient trop divergentes et il n'y a pas eu d'amplification à partir des amorces humaines de PCR utilisées. Dans ces cas, les haplotypes ancestraux proviennent de la base de données de l'Université de Californie, Santa Cruz (UCSC) Pour trois gènes ( GSS, GSTM3, GSTM4 ), ces haplotypes ont également été trouvés dans l'échantillon humain.



## GSS

On a géotypé neuf sites polymorphismes dans le promoteur de la synthétase du glutathion (GSS) dans l'échantillon mondial. Neuf haplotypes ont été retrouvés dans l'ensemble des populations continentales. La valeur de  $\pi$  (0,079%) pour le Moyen-Orient est plus élevée que pour les autres populations et témoigne de la répartition des fréquences haplotypiques. Celle de l'Amérique est plus basse et reflète des valeurs de  $k$  et  $S$  plus basse que la moyenne mondiale. L'échantillon africain a plus d'haplotypes et son indice de diversité haplotypique est plus élevé. L'haplotype ancestral a été retrouvé dans les populations africaine et moyen-orientale. Les quatre haplotypes les plus fréquents sont partagés par toutes les populations continentales. Les tests de neutralité ne détectent pas de départ du modèle standard neutre et donc ne montrent pas de signes de sélection. Le  $F_{st}$  total qui indique des différenciations entre les populations n'est pas significativement différent de 0.

Les haplotypes sont distribués presque également dans les groupes québécois, c'est-à-dire qu'il n'y a pas d'haplotype majeur clairement défini. D'autre part, les Canadiens-Français de Gaspésie et le groupe anglo-normand présentent un haplotype asiatique rare avec des fréquences non négligeables (respectivement 27,7% et 36,1% pour S11). Cet haplotype n'est pas retrouvé dans la population amérindienne qui ne comprend cependant que des Amérindiens du sud. On ne le retrouve pas dans l'échantillon européen analysé, en revanche il peut s'agir d'un haplotype européen rare avec une fréquence assez importante dans les îles anglo-normandes. Il se sera retrouvé dans la population canadienne-française à la suite d'unions avec des immigrants normands. Il peut également s'agir d'un haplotype présent chez les Autochtones d'Amérique du Nord qui se sera introduit dans la population gaspésienne.

Les échantillons aborigènes de Taiwan et islandais présentent moins de polymorphismes que les populations continentales et que les groupes québécois. L'indice de diversité haplotypique du groupe de Taiwan (0,560) est dans le même ordre que celui de l'Amérique (0,528) alors que celui de l'Islande (0,675) se rapproche des G de l'Europe (0,671) et de l'Asie (0,661).

Tableau VIII. Statistiques descriptives pour GSS

	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
<b>S</b>	7	5	6	5	7	9
<b><math>\pi</math> (%)</b>	0,079	0,056	0,073	0,073	0,091	0,076
<b>k</b>	7	4	6	4	6	9
<b>G</b>	0,825	0,528	0,661	0,671	0,776	0,714
<b>Tajima's D</b>	-0,29946	0,11264	0,25845	1,3229	-0,20481	-0,32292
<b>Fu's Fs</b>	-1,21332	0,77161	0,87579	1,64142	0,98984	-1,33392
<b>Fst Total</b>	<b>0,028</b>					

Tableau IX. Haplotypes pour GSS et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient

Noms	Séquences	Fréquences					Total
		Af	Am	As	Eu	MO	
<b>ANCestral</b>	<b>G G G G A C G C G</b>	12,5				3,1	3,1
S01	- A - - - - - C	34,4	65,6	53,1	50,0	28,1	46,3
S02	A A - - - - -	12,5	21,9	25,0	21,9	28,1	21,9
S03	- A - A - - - -	6,3	9,4	6,3	21,9	28,1	14,4
S04	- A - - T - - G -	6,3	3,1	9,4	6,3	9,4	6,9
S05	- A - - - - del - -	18,8					3,8
S07	- A - - - - - - -	9,4		3,1			2,5
S08	- A - - - - - G -						
S09	- A - - T T - G -					3,1	0,6
S10	A A - A - - - - -						
S11	A A A - - - - - -			3,1			0,6

**Tableau X. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GSS. Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.**

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	5	5	5	5	5	5	5
$\pi$ (%)	0,078	0,092	0,087	0,08	0,079	0,081	0,083
<b>k</b>	4	6	4	4	4	4	6
<b>G</b>	0,703	0,749	0,754	0,706	0,713	0,705	0,722

B

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S01</b>	36,1	38,9	27,8	38,9	41,7	44,3	38,8
<b>S02</b>	38,9	25,0	13,9	11,1	27,8	22,9	23,2
<b>S03</b>	16,7	8,3			22,2	20,0	12,4
<b>S04</b>	8,3	22,2			8,3	12,9	9,2
<b>S08</b>		2,8					0,4
<b>S10</b>		2,8	30,6	13,9			6,8
<b>S11</b>			27,8	36,1			9,2

**Tableau XI. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et des Islandais pour GSS.**

A

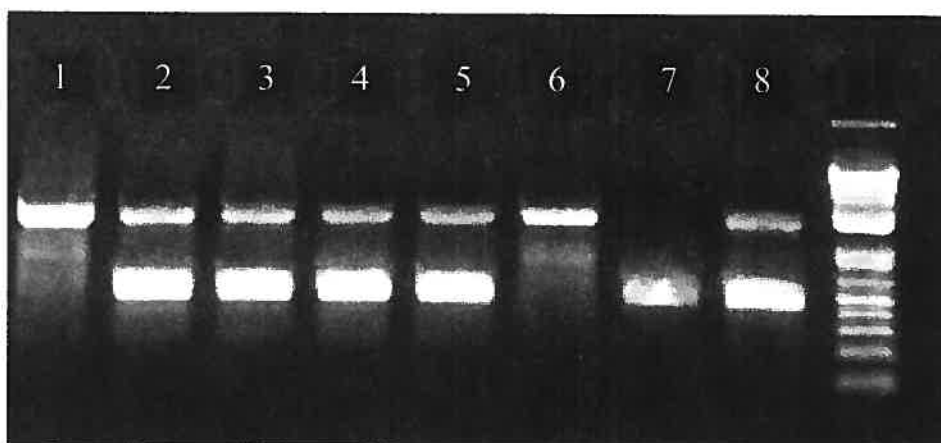
	Taiwan	Islande
<b>S</b>	4	3
$\pi$ (%)	0,069	0,067
<b>k</b>	3	3
<b>G</b>	0,560	0,675

B

	Taiwan	Islande
<b>S01</b>	64,3	43,8
<b>S02</b>	14,3	
<b>S10</b>	21,4	37,5
<b>S11</b>		18,8

## GSTT1

La délétion de GSTT1 chez une grande proportion de la population mondiale singularise ce gène par rapport aux autres gènes étudiés. La technique de l'hybridation ASO ne permet pas de distinguer les homozygotes (pas de délétion ou deux copies du gène) des hémizygotes (une délétion et une copie du gène). Pour génotyper la délétion de GSTT1, la méthode de PCR multiplex de Sprenger et al. (Sprenger, Schlagenhauser et al. 2000) a été utilisée. La figure 12 ( page 45) présente un exemple de résultats de cette méthode de génotypage.



**Figure 12. Résultats de génotypage de la délétion de GSTT1.** La bande supérieure montre la présence de la délétion et la bande inférieure celle du gène. Ainsi, l'individu 7 possède deux copies de GSTT1, l'individu 2 une seule copie et l'individu 1 n'en possède pas. Le marqueur de poids moléculaire est du 1KB+ de Invitrogen.

À la suite du génotypage, six haplotypes ont été identifiés ( outre la délétion ). L'Afrique présente cinq polymorphismes et quatre haplotypes alors que les autres populations en ont trois et deux, respectivement. Les valeurs de ces paramètres de diversité expliquent les indices plus élevés de diversité haplotypique et nucléotidique pour l'Afrique. « L'haplotype » le plus fréquent pour GSTT1 est la délétion complète du gène. Quarante-quatre pour cent des chromosomes mondiaux portent cette délétion (figure 13 page 48). La proportion est légèrement inférieure en Afrique et en Amérique avec 37,5% et 33,3% respectivement. En outre, chez les Asiatiques 63,3% des chromosomes ont la délétion du gène GSTT1.

Les fréquences de délétion des groupes québécois suivent celle des Européens sauf pour le groupe loyaliste, 33,3%, et le groupe anglo-normand, 29,4%. Les autres paramètres de diversité des groupes québécois s'apparentent à ceux des Européens.

Les aborigènes de Taïwan et les Islandais présentent respectivement 25% et 43% de délétion. Même si la fréquence de délétion de l'échantillon islandais s'apparente à celle de l'Europe, pour le reste des haplotypes, l'Islande se distingue. En effet, l'haplotype T1 est fixé pour l'échantillon islandais.

**Tableau XII. Statistiques descriptives pour GSTT1**

	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
<b>S</b>	5	3	3	3	3	6
<b><math>\pi</math> (%)</b>	0,086	0,050	0,072	0,060	0,031	0,079
<b>k</b>	4	2	2	2	2	5
<b>G*</b>	0,744	0,616	0,545	0,639	0,577	0,667
<b>Tajima's D</b>	-0,14254	-1,65231	0	-1,44751	0	0,25403
<b>Fu's Fs</b>	0,14326	0,97596	1,60944	1,4149	0,90502	
<b>FstTotal</b>	<b>0,065*</b>					

\* ces valeurs tiennent compte de la délétion.

**Tableau XIII. Haplotypes pour GSTT1 et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient**

Noms	séquences	Fréquences					Total
		Af	Am	As	Eu	MO	
<b>ANCestral</b>	ACCAA C C G T G						
<b>DEL</b>	del	37,5	34,5	63,3	43,8	43,3	44,4
<b>T1</b>	TCAT - T - C -	12,5	51,7	23,3	40,6	50,0	35,3
<b>T2</b>	- del - - C -		13,8	13,3	15,6	6,7	9,8
<b>T3</b>	- - - A - -	31,3					6,5
<b>T4</b>	- - - - C -	15,6					3,3
<b>T5</b>	- - - A C -						
<b>T6</b>	- - - - C C	3,1					0,7

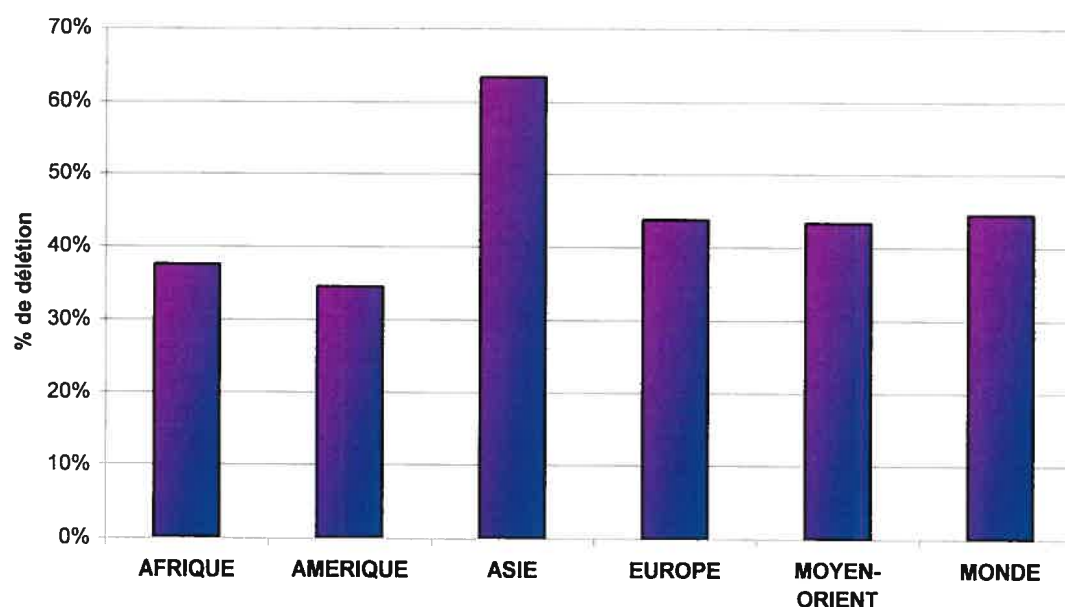


Figure 13. Pourcentage de délétion par continent.

Tableau XIV. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GSTT1. Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
S	3	3	3	3	3	4	4
$\pi$ (%)	0,04	0,02	0,04	0,012	0,012	0,018	0,022
k	2	2	2	2	2	3	3
G	0,594	0,557	0,589	0,469	0,494	0,538	0,539

B

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
DEL	44,1	46,2	47,2	29,4	33,3	41,5	40,2
T1	47,1	50,0	44,4	67,6	63,9	54,7	54,8
T2	8,8	3,8	8,3	2,9	2,8	1,9	4,6
T5						1,9	0,5

**Tableau XV. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GSTT1.**

A

	Taiwan	Islande
<b>S</b>	3	0
$\pi$ (%)	0,055	
<b>k</b>	2	1
<b>G</b>	0,621	0,527

B

	Taiwan	Islande
<b>DEL</b>	25,0	42,9
<b>T1</b>	58,3	57,1
<b>T2</b>	16,7	



## GSTP1

Le génotypage de l'ensemble des cinq populations continentales a révélé treize haplotypes pour GSTP1 dont deux haplotypes majeurs partagés par toutes les populations et huit haplotypes spécifiques à une seule population. Sur les treize haplotypes, onze ont des fréquences inférieures à 11%. L'Afrique et l'Amérique se distinguent toutes deux par des paramètres élevés de diversité comparée aux autres populations. L'Afrique possède les valeurs de  $S$  ( 13 ) et  $k$  ( 9 ) les plus élevées. Cependant, l'Amérique possède des indices de diversité haplotypique (0,786) et nucléotidique ( 0,230%) tout aussi élevés que l'Afrique (  $\pi$  = 0,231% ) et  $G$  = (0,786). L'indice  $F_{st}$  indique une différenciation modérée entre les populations.

Nous n'avons identifié que quatre haplotypes chez les Européens alors que les Québécois en ont dix. L'haplotype P03 présent à 8,7% dans l'échantillon québécois n'est pas retrouvé dans la cohorte européenne mais il est présent à 16,6% chez les Amérindiens. Il pourrait s'agir d'un haplotype amérindien introduit dans la population québécoise à la suite de métissage.

Les fréquences haplotypiques de Taiwan se rapprochent de celles de l'Asie alors que celles de l'Islande sont semblables à celles de l'Europe. D'autre part, les statistiques pour Taiwan montrent une population diversifiée. La diversité nucléotidique (0,222%) pour cette population se rapproche de la valeur africaine (0,231%).

Tableau XVI. Statistiques descriptives pour GSTP1.

GSTP1	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
S	13	12	12	12	10	13
$\pi$ (%)	0,231	0,230	0,151	0,172	0,190	0,205
k	9	6	5	4	4	13
G	0,786	0,786	0,644	0,504	0,595	0,713
Tajima's D	1,07355	1,33828	-0,75672	0,54	0,93189	1,57771
Fu's Fs	-0,84102	3,08696	1,32809	4,90108	5,30861	0,73963
<b>Fst Total</b>			<b>0,059*</b>			

\*valeur statistiquement significative.

Tableau XVII. Haplotypes pour GSTP1 et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient, Ins= Insertion de AGTCTAGTCTA.

Noms	Séquences												Fréquences (%)							
													Af	Am	As	Eu	MO	Total		
ANCeutral	Ins	C	G	T	A	C	G	C	G	DEL	A	T	C							
P01	-	-	-	A	-	-	-	-	C	T	C	C	-	9,4	28,1	56,7	65,6	56,3	43,0	
P02	-	T	A	-	G	-	-	-	-	-	-	-	-	43,8	31,3	13,3	28,1	31,3	29,7	
P03	-	-	-	A	-	-	C	T	-	T	C	C	-	12,5	15,6	16,7		9,4	10,8	
P04	del	-	-	A	-	-	-	-	-	T	C	C	G	3,1	18,8	10,0	3,1		7,0	
P05	-	-	-	A	G	-	-	-	C	T	C	C	-							
P06	-	-	-	A	-	-	-	-	-	T	C	C	-	9,4						1,9
P07	-	-	A	A	-	-	-	-	C	T	C	C	-			3,3		3,1		1,3
P08	-	-	-	A	-	-	-	-	C	-	-	C	-							
P09	-	-	-	A	-	-	-	-	C	-	-	-	-		3,1					0,6
P10	-	-	-	A	-	-	C	T	C	T	C	C	-				3,1			0,6
P11	-	T	A	-	G	-	-	-	C	-	-	-	-	6,3						1,3
P12	del	-	-	A	-	-	-	-	-	T	C	C	-	6,3						1,3
P13	del	-	-	A	-	T	-	-	-	T	C	C	-	6,3						1,3
P14	-	-	-	A	-	-	-	-	-	T	C	C	G		3,1					0,6
P15	-	-	-	A	-	T	-	-	-	T	C	C	-	3,1						0,6
P16	-	T	A	-	G	-	-	-	-	T	C	-	-							

**Tableau XVIII. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GSTP1. Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.**

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	10	10	10	10	10	10	10
$\pi$ (%)	0,181	0,206	0,197	0,173	0,214	0,205	0,198
<b>k</b>	5	4	5	6	5	5	10
<b>G</b>	0,563	0,565	0,676	0,667	0,694	0,616	0,63

B

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>P01</b>	58,8	41,2	47,1	52,9	38,9	47,1	47,5
<b>P02</b>	32,4	52,9	32,4	23,5	38,9	40,0	37,2
<b>P03</b>	2,9	2,9	11,8	8,8	13,9	10,0	8,7
<b>P05</b>			5,9	8,8	5,6	1,4	3,3
<b>P06</b>			2,9	2,9			0,8
<b>P07</b>	2,9						0,4
<b>P08</b>	2,9					1,4	0,8
<b>P09</b>					2,8		0,4
<b>P10</b>		2,9					0,4
<b>P16</b>				2,9			0,4

**Tableau XIX. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GSTP1.**

A

	Taiwan	Islande
<b>TP01</b>	42,9	21,4
<b>TP02</b>	28,6	57,1
<b>TP03</b>	14,3	
<b>TP04</b>	14,3	
<b>TP05</b>		21,4

B

	Taiwan	Islande
<b>S</b>	12	8
$\pi$ (%)	0,222	0,193
<b>k</b>	4	3
<b>G</b>	0,747	0,626

### GSTM3

Le promoteur de GSTM3 est celui, parmi les gènes étudiés, pour lequel on a trouvé le moins de polymorphismes (quatre). Cinq haplotypes ont été identifiés. Trois haplotypes peuvent être qualifiés de majeurs et il n'y a que deux haplotypes mineurs. De plus, l'haplotype ancestral ne se retrouve pas dans la population africaine alors qu'il est présent dans les autres populations continentales. La population africaine ne présente d'ailleurs que très peu de diversité pour ce gène ( $\pi = 0,006\%$  et  $G=0,123$ ) comparé aux autres populations. L'haplotype MIII2 est presque fixé pour cette population (93,8%). Les paramètres de diversité suggèrent que la dérive génétique a contribué à réduire la diversité dans la population africaine. À cause de la différence africaine, l'indice de Fst montre une très grande différenciation génétique entre les populations. En effet, 19,2% de la variabilité sont dus aux différences entre les populations.

Les groupes québécois se distinguent des Européens en présentant un seul haplotype majeur, l'ancestral. Seul le groupe de Jersey et Guernesey ne possède pas d'haplotype majeur clairement défini. C'est-à-dire que les haplotypes ANC, MIII02 et MIII03 ont des fréquences respectives de 33,3%, 36,1% et 22,2%. Ces fréquences ressemblent davantage à celle des Européens qui possèdent également trois haplotypes majeurs.

L'Islande, tout comme la majorité des groupes québécois, ne possède qu'un haplotype majeur, l'ancestral. L'haplotype le plus important pour l'échantillon de Taiwan est MIII-3. Il s'agit également de l'haplotype majeur de l'Asie et l'Amérique.

**Tableau XX. Statistiques descriptives pour GSTM3.**

<b>GSTM3</b>	<b>Afrique</b>	<b>Amérique</b>	<b>Asie</b>	<b>Europe</b>	<b>Moyen-Orient</b>	<b>Monde</b>
<b>S</b>	2	3	3	3	3	4
<b><math>\pi</math> (%)</b>	0,006	0,065	0,047	0,072	0,075	0,061
<b>k</b>	3	3	3	4	3	5
<b>G</b>	0,123	0,627	0,548	0,704	0,663	0,666
<b>Tajima's D</b>	-1,16221	1,91076	0,20063	1,99092	1,75043	1,33603
<b>Fu's Fs</b>	-0,7001	1,97983	0,92083	0,71844	1,88952	1,1901
<b>Fst Total</b>	<b>0,192*</b>					

\*valeur statistiquement significative.

**Tableau XXI. Haplotypes pour GSTM3 et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient.**

<b>Noms</b>	<b>Séquences</b>	<b>Fréquences (%)</b>					<b>Total</b>
		<b>Af</b>	<b>Am</b>	<b>As</b>	<b>Eu</b>	<b>MO</b>	
<b>ANCestral</b>	<b>C C C G</b>		25,0	12,5	37,5	43,8	23,8
<b>MIII-2</b>	<b>- T - T</b>	93,8	21,9	25,0	31,3	21,9	38,8
<b>MIII-3</b>	<b>G T - T</b>	3,1	53,1	62,5	28,1	34,4	36,3
<b>MIII-5</b>	<b>- T - -</b>						
<b>MIII-7</b>	<b>- T G T</b>	3,1					0,6
<b>MIII-8</b>	<b>G - - -</b>				3,1		0,6

**Tableau XXII . Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GSTM3.** Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G =d'origine anglo-normande.

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	3	3	3	3	3	3	3
<b><math>\pi</math> (%)</b>	0,073	0,070	0,068	0,068	0,070	0,066	0,068
<b>k</b>	3	4	3	3	3	3	4
<b>G</b>	0,640	0,656	0,624	0,684	0,641	0,648	0,645

B

	Acadiens	Char-SLSJ	CF Gasp.	Jer-Guer	Loyalistes	Montréal	Total
<b>ANC</b>	50,0	50,0	52,8	33,3	50,0	42,9	46,0
<b>MIII2</b>	20,6	25,0	27,8	36,1	27,8	37,1	30,2
<b>MIII3</b>	29,4	22,2	19,4	30,6	22,2	20,0	23,4
<b>MIII5</b>		2,8					

**Tableau XXIII. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GSTM3.**

A

	Taiwan	Islande
<b>ANC</b>	7,1	62,5
<b>MIII-2</b>	14,3	25,0
<b>MIII-3</b>	78,6	12,5
<b>MIII-4</b>	7,1	

B

	Taiwan	Islande
<b>S</b>	3	3
<b><math>\pi</math> (%)</b>	0,041	0,062
<b>k</b>	4	3
<b>G</b>	0,517	0,567

#### **GSTM4**

Pour le dernier gène de GST analysé, dix polymorphismes ont été identifiés. Dix haplotypes ont pu être reconstruits dont deux majeurs partagés par toutes les populations et cinq haplotypes spécifiques à une seule population. Les haplotypes MIV3 et MIV4 présents chez les Amérindiens mais pas les Asiatiques, pourraient être le témoignage du métissage de la population amérindienne avec les populations européenne et africaine. Cependant, l'échantillon asiatique analysé ici comprend uniquement des individus d'Asie du Sud-est et la colonisation des Amériques pourrait avoir été majoritairement faite par les populations du nord de l'Asie. Les indices de diversité nucléotidique sont uniformes (0,107% à 0,135%) à travers les populations continentales. La diversité haplotypique est sensiblement inférieure pour l'Asie (0,546) qui ne possède que deux haplotypes importants et un singleton.

Le groupe anglo-normand se distingue parmi les groupes québécois, puisque l'haplotype majeur de ce groupe est MIV2, 41,7%. Il s'agit également de l'haplotype majeur des Européens, 43,8%. Pour tous les autres groupes québécois, l'haplotype majeur de GSTM4 est MIV1.

Les paramètres de diversité de l'échantillon de Taiwan et de celui d'Irlande se rapprochent respectivement des échantillons asiatique et européen.

Tableau XXIV. Statistiques descriptives pour GSTM4

GSTM4	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
S	9	5	4	4	6	10
$\pi$ (%)	0,135	0,112	0,113	0,107	0,113	0,117
k	8	5	3	3	5	10
G	0,835	0,665	0,546	0,663	0,762	0,710
Tajima's D	0,1955	1,26085	2,34321	2,19182	1,02392	0,15618
Fu's Fs	-0,04383	0,41914	2,99297	2,90129	1,38458	-0,19043
<b>Fst Total</b>			<b>0,004</b>			

Tableau XXV. Haplotypes pour GSTM4 et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient

Noms	Séquences	Fréquences (%)					Total
		Af	Am	As	Eu	MO	
ANCestral	A G C C T T C T C G	6,3	3,1				1,9
MIV-1	G A - - - - - - C	15,6	46,9	46,9	34,4	34,4	35,6
MIV-2	- - T - - - - - -	31,3	34,4	50,0	43,8	28,1	37,5
MIV-3	- A - - - - - - -	21,9	12,5		21,9	21,9	15,6
MIV-4	- A - - - G - - - -	6,3	3,1			12,5	4,4
MIV-6	G A - - C - - - - C	9,4					1,9
MIV-7	- A - G - - - C G -	6,3					1,3
MIV-8	G A - - - - - - -			3,1			0,6
MIV-9	- A - - - - T - - -					3,1	0,6
MIV-10	- A - - - G - C - -	3,1					0,6



**Tableau XXVI. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GSTM4.** Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	5	5	5	5	5	5	5
$\pi$ (%)	0,113	0,114	0,107	0,109	0,114	0,107	0,110
<b>k</b>	4	4	4	4	4	4	4
<b>G</b>	0,583	0,679	0,646	0,713	0,633	0,659	0,653

B

	Acadiens	Char-SLSJ	CF Gasp.	Jer-Guer	Loyalistes	Montréal	Total
<b>MIV1</b>	52,8	41,2	50,0	27,8	41,7	48,6	44,4
<b>MIV2</b>	38,9	38,2	30,6	41,7	44,4	28,6	35,9
<b>MIV3</b>	5,6	14,7	16,7	22,2	11,1	17,1	14,9
<b>MIV4</b>	2,8	5,9	2,8	8,3	2,8	5,7	4,8

**Tableau XXVII. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GSTM4.**

A

	Taiwan	Islande
<b>MIV-1</b>	56,3	50,0
<b>MIV-2</b>	43,8	43,8
<b>MIV-3</b>		6,3

B

	Taiwan	Islande
<b>S</b>	4	4
$\pi$ (%)	0,115	0,116
<b>k</b>	2	3
<b>G</b>	0,525	0,592

## GPX2

Pour GPX2, on a identifié seize polymorphismes dont trois mutations spécifiques à l'Afrique. L'Afrique montre une diversité haplotypique presque deux fois supérieure à celle des autres populations. Le nombre d'haplotype ( $k$ ) ne devrait pas dépasser  $S+1$  en absence de recombinaison. Or, dix-huit haplotypes ont été reconstruits, ce qui est supérieur au nombre attendu de dix-sept haplotypes. Ce nombre élevé d'haplotypes suggère la présence de recombinaisons et/ou de conversions géniques. Il n'y a qu'un haplotype majeur et une grande quantité (seize) d'haplotypes mineurs avec une fréquence inférieure à 6,3%. Le test de neutralité de Fu témoigne de cet excès d'haplotypes rares. Le test est significatif pour l'Afrique ( $p < 0,003$ ) et le monde ( $p < 0,014$ ) et suggère une sélection positive en faveur d'un haplotype. Ce pourrait également être l'effet des recombinaisons. Cependant, l'indice de  $F_{st}$  total indique une grande différenciation entre les populations. D'ailleurs, les fréquences des haplotypes importants varient beaucoup à travers les populations continentales de même que celles des polymorphismes. Le 12<sup>e</sup> polymorphisme identifié, G512T, est presque exclusivement asiatique. Pour l'Amérique, l'Europe et le Moyen-Orient, l'haplotype XII01 est majoritaire (respectivement 65,6%, 73,3% et 71,9%). Pour l'Asie, l'haplotype majeur est XII02 (53,1%; qui comprend 512T). Les fréquences de l'Afrique sont partagées plus ou moins également entre treize haplotypes.

Chez les Québécois, l'haplotype XII01 est presque fixé surtout pour les Canadiens-Français de Gaspésie (83,3%). L'indice de diversité haplotypique de ce groupe est sensiblement inférieur (0,294) à la moyenne québécoise (0,446).

Le groupe de Taiwan possède deux haplotypes majeurs XII01 (41,7%) et XII02 (50%). Il possède donc les haplotypes majeurs de l'Asie et du groupe Amérique, Europe et Moyen-Orient. Dans l'échantillon islandais, c'est XII01 qui est majoritaire (87,5%).

Tableau XXVIII. Statistiques descriptives pour GPX2.

GPX2	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
<b>S</b>	11	9	9	7	6	16
<b><math>\pi</math> (%)</b>	0,115	0,079	0,097	0,072	0,061	0,098
<b>k</b>	13	8	6	6	6	18
<b>G</b>	0,905	0,567	0,661	0,460	0,480	0,697
<b>Tajima's D</b>	-0,37802	-0,22818	-0,35003	-1,4195	-0,34679	-0,85172
<b>Fu's Fs</b>	<b>-5,26782</b>	-0,36248	0,18164	-0,1331	-0,21284	<b>-6,3668</b>
<b>Fst Total</b>	<b>0,154*</b>					

\*valeur statistiquement significative.

**Tableau XXIX. Haplotypes pour GPX2 et fréquences dans les populations continentales.** Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient.

Noms	Séquences	Fréquences (%)					Total
		Af	Am	As	Eu	MO	
<b>ANCestral</b>	<b>T A C G A T C A T A A G T A C A</b>						
XII01	- - - - - C - G - - - - - - - -	25,0	65,6	25,0	73,3	71,9	51,9
XII02	- - - - - C - - - del - T - - - - -	12,5	3,1	53,1	3,3	9,4	16,5
XII03	- - - A - C - - - del G - - - - -		6,3	9,4	10,0	6,3	6,3
XII04	- C - - - C - - C del - - - - - - -	9,4	3,1	6,3	6,7		5,1
XII05	- - - - - C - G - del - - - - - - -	6,3	3,1		3,3		2,5
XII06	C C - - - C - G - - - - - - - G	6,3	6,3				2,5
XII07	- - - - - C - - - - - T - - - - -		9,4	3,1	3,3		3,2
XII08	- - - - - C - - - del - - - - - - -	12,5	3,1				3,2
XII09	- - - - - C - - - - - - - - - - -	6,3				3,1	1,9
XII10	- - - A - C - - - - - G - - - - -					3,1	0,6
XII11	- - - - - C - - - del - - - - - G	6,3					1,3
XII12	- C - - - C - G C del - - - - - - -						
XII13	- - - - - C - - - - - - - - - G						
XII14	- - - - - C - G - - - - - G - - -				6,3		1,3
XII15	- - - - - C - - - del - T C - - - -						
XII16	- - - - - C - - - del G - - - - - -						
XII17	- - - - - C - - - - - - - - - G -	3,1					0,6
XII18	- - - - - C - G - del - - - - - G						
XII19	- - - - - C - G - del G - - - - - -						
XII20	- - - - - C - G - - - - - - - G	3,1					0,6
XII21	- - - - - C T G - del - - - - - -	3,1					0,6
XII22	- - - - - - - - del - - - - - - -	3,1					0,6
XII23	- - - - - - - G - - - - - - - - -						
XII24	- - - - G - - - - del - - - - - -	3,1					0,6
XII25	- - - A - C - G - del G - - - - -						
XII26	- - - A - C - G - - - - - - - - -						
XII27	- - T - - C - - - del - T C - - - -			3,1			0,6
XII28	- C - - - C - G - - - - - - - - -						
XII29	- C - - - C - G - - - - - - - - G						

**Tableau XXX. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GPX2.** Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	7	7	6	7	7	9	9
$\pi$ (%)	0,066	0,068	0,056	0,098	0,08	0,075	0,073
<b>k</b>	6	4	3	6	5	9	12
<b>G</b>	0,427	0,405	0,294	0,597	0,465	0,478	0,446

B

	Acadiens	Char-SLSJ	CF Gasp.	Jer-Guer	Loyalistes	Montréal	Total
<b>XII01</b>	75,0	76,5	83,3	61,8	72,2	71,4	73,2
<b>XII02</b>	2,8	11,8		8,8	2,8	2,9	4,5
<b>XII03</b>	13,9	8,8	13,9	14,7	11,1	11,4	12,2
<b>XII04</b>	2,8	2,9	2,8	8,8	11,1	5,7	5,7
<b>XII05</b>	2,8				2,8	2,9	1,6
<b>XII06</b>						1,4	0,4
<b>XII09</b>				2,9			0,4
<b>XII10</b>				2,9			0,4
<b>XII16</b>						1,4	0,4
<b>XII19</b>						1,4	0,4
<b>XII26</b>	2,8						0,4
<b>XII28</b>						1,4	0,4

**Tableau XXXI. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GPX2.**

A

	Taiwan	Islande
<b>XII01</b>	41,7	87,5
<b>XII02</b>	50,0	
<b>XII03</b>		6,3
<b>XII04</b>	8,3	
<b>XII25</b>		6,3

B

	Taiwan	Islande
<b>S</b>	5	4
$\pi$ (%)	0,093	0,040
<b>k</b>	3	3
<b>G</b>	0,621	0,242

### GPX3

Les résultats du génotypage de GPX3 ont permis de reconstruire onze haplotypes. Dans la population mondiale, un seul haplotype est majeur avec une fréquence de 61,8%. Il y a sept haplotypes avec une fréquence inférieure ou égale à 2,6%. Cependant, la population africaine présente cinq haplotypes importants : XII01 20%, XII02 30%, XII03 20%, XII04 10% et XII06 13,3%. Ce partage des fréquences haplotypiques est la cause d'un indice de diversité haplotypique africain (0,828) très supérieur à la moyenne mondiale (0,587). De plus, l'indice  $F_{st}$  témoigne d'une diversité modérée entre les populations continentales.

Pour GPX3, il y a un seul haplotype majeur, XIII01, pour tous les groupes québécois sauf les gens de Charlevoix et du Saguenay-Lac St-Jean. Chez ce groupe, il y a trois haplotypes importants et trois haplotypes mineurs, une distribution qui se démarque de celle des autres groupes et de la population européenne. Les paramètres de diversité de ce groupe sont d'ailleurs beaucoup plus élevés ( $\pi = 0,149\%$ ,  $G = 0,754$ ) que ceux des autres groupes québécois ( Québec :  $\pi = 0,091\%$  ,  $G = 0,549$ ) ou que l'échantillon européen ( $\pi = 0,065$ ,  $G = 384$ ). L'indice  $F_{st}$  de l'échantillon québécois, 0,29, témoigne de cette différenciation génétique du groupe Charlevoix et Saguenay-Lac St-Jean.

Les échantillons de Taiwan et d'Islande ne se démarquent pas de la moyenne mondiale.

Tableau XXXII. Statistiques descriptives pour GPX3

GPX3	Afrique	Amérique	Asie	Europe	Moyen-Orient	Monde
S	10	8	9	9	10	11
$\pi$ (%)	0,156	0,092	0,095	0,065	0,078	0,104
k	7	5	4	5	6	11
G	0,828	0,434	0,405	0,384	0,583	0,587
Tajima's D	1,15454	0,66903	-0,42287	-1,41244	-1,01323	0,43768
Fu's Fs	0,92859	1,28585	1,942	1,89995	1,03954	-0,15903
<b>Fst Total</b>	<b>0,074*</b>					

\*valeur statistiquement significative.

Tableau XXXIII. Haplotypes pour GPX3 et fréquences dans les populations continentales. Af= Afrique, Am= Amérique, As= Asie, Eu= Europe, MO= Moyen-Orient

Noms	Séquences	Fréquences (%)					Total
		Af	Am	As	Eu	MO	
ANCestral	A G C A C A C C C T T						
XIII01	G A - - T - - T T A -	20,0	75,0	76,7	78,6	59,4	61,8
XIII02	- - - - T - - T T A -	30,0	6,3		7,1	28,1	14,5
XIII03	G - - C - T - - - -	20,0	9,4	10,0	7,1	3,1	9,9
XIII04	- - - - T - - T T A A	10,0		10,0	3,6	3,1	5,3
XIII05	G - - - T - - T T A -	3,3	3,1		3,6		2,0
XIII06	G - - - T - - T T A A	13,3					2,6
XIII07	- - - - T - - - T A -						
XIII09	G - - C - - - - - -		6,3				1,3
XIII10	- A - - T - - T T A -						
XIII11	G - - - - - - T - - -						
XIII13	G - - C T - - T - - -						
XIII14	G - - C - - - T - - -						
XIII16	G - - C - T - T - - -			3,3			0,7
XIII17	G - - C - T - - - A -					3,1	0,7
XIII18	G - - C - T T - - - -	3,3					0,7
XIII19	G - T - T - - T T A -					3,1	0,7

**Tableau XXXIV. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des groupes québécois pour GPX3.** Char-SLSJ = Charlevoix et Saguenay Lac St-Jean, CF Gasp. = Canadiens-Français de Gaspésie, J-G = d'origine anglo-normande.

A

	Acadiens	Char-SLSJ	CF Gasp.	J-G	Loyalistes	Montréal	Total
<b>S</b>	7	7	6	7	7	9	9
$\pi$ (%)	0,066	0,068	0,056	0,098	0,08	0,075	0,073
<b>k</b>	6	4	3	6	5	9	12
<b>G</b>	0,427	0,405	0,294	0,597	0,465	0,478	0,446

B

	Acadiens	Char-SLSJ	CF Gasp.	Jer-Guer	Loyalistes	Montréal	Total
<b>XIII01</b>	73,5	34,4	75,0	68,8	71,9	57,6	62,9
<b>XIII02</b>	14,7	31,3	13,9	21,9	25,0	25,8	22,4
<b>XIII03</b>	2,9	21,9	8,3	9,4	3,1	10,6	9,5
<b>XIII04</b>	5,9	6,3	2,8			1,5	2,6
<b>XIII07</b>	2,9					1,5	0,9
<b>XIII10</b>		3,1					0,4
<b>XIII11</b>						1,5	0,4
<b>XIII13</b>		3,1					0,4
<b>XIII14</b>						1,5	0,4

**Tableau XXXV. Statistiques descriptives ( A ) et fréquences haplotypiques ( B ) des Aborigènes de Taiwan et les Islandais pour GPX3.**

A

	Taiwan	Islande
<b>S</b>	9	8
$\pi$ (%)	0,135	0,094
<b>k</b>	4	3
<b>G</b>	0,592	0,385

B

	Taiwan	Islande
<b>XIII01</b>	58,8	91,7
<b>XIII02</b>	5,9	8,3
<b>XIII03</b>	17,6	16,7
<b>XIII04</b>	11,8	



## 2 Apport des mutations et des recombinaisons

La variabilité génétique provient des mutations et des recombinaisons. Pour déterminer les taux de recombinaisons des gènes des régions étudiées, des programmes d'estimation (MaxHap et Maxdip (Hudson 2001)) et une méthode de calcul à partir de taux de recombinaison régionaux au niveau chromosomique provenant de données familiales ont été utilisés (Tableau XXXVI page 67). Les programmes estiment le taux de recombinaison à travers le paramètre rho qui est égal à  $4Nr$  où  $N$  est l'effectif efficace et  $r$  est le taux de recombinaison. MaxDip calcule rho à partir de données polymorphiques diploïdes, MaxHap le calcul à partir des données haplotypiques. Nous avons d'abord calculé  $N$  à partir de la diversité nucléotidique  $\pi$  et du taux de mutation  $\mu$  par génération (20 ans) obtenu en calculant le taux de divergence avec le chimpanzé sur 10kb ( section 4 du présent chapitre ) selon l'équation  $N = \pi / 4\mu$ . Nous avons ensuite calculé  $r$  par kb à partir de la définition de rho. Selon MaxHap et MaxDip GPX2 présente un taux de recombinaison élevé ( $7,2 \times 10^{-5}$  et  $5,3 \times 10^{-5}$  par kb), ce qui suggère une région riche en recombinaisons. La 3<sup>e</sup> méthode consiste en un calcul de la pente d'un graphique du taux de recombinaisons en fonction de la distance physique sur un Mb en utilisant des marqueurs microsatellites. Ces taux proviennent d'une étude de cartographie du génome humain (Kong, Murphy et al. 2004). Selon cette méthode, les taux de recombinaisons sont généralement plus élevés. Cependant, GPX2 et GPX3 ont des valeurs inférieures à la moyenne,  $2 \times 10^{-6}$  par kb et  $1 \times 10^{-6}$  par kb respectivement. Les marqueurs de distance physique dans ces régions étaient plutôt rares, il y a d'ailleurs une brèche dans les données du graphique. Ce manque de données pourrait être la cause de la divergence entre les estimations des programmes et les calculs sur données familiales pour GPX2. Par ailleurs, les taux de GSTP1, GSTM3 et GSTM4 par données familiales sont très élevés comparé à ceux obtenus par les deux autres méthodes ce qui pourrait suggérer que les taux varient de kb en kb dans ces régions.

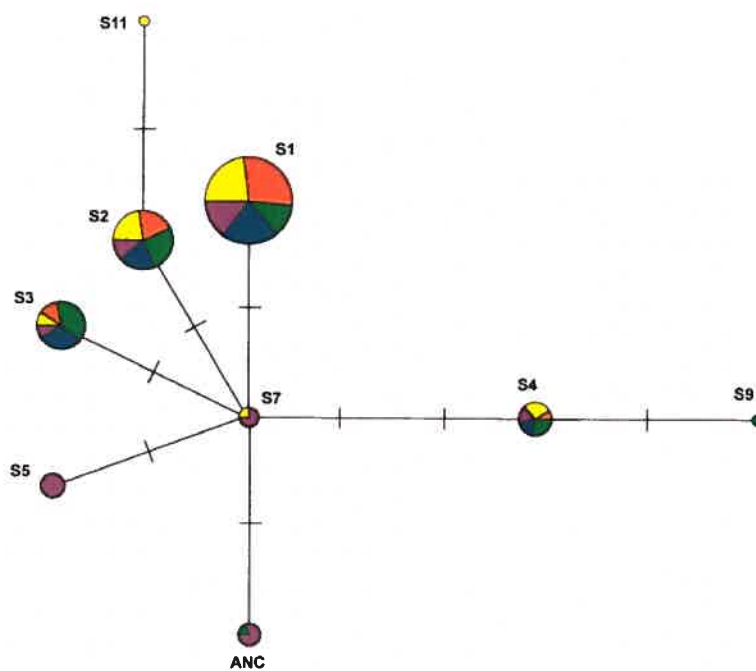
**Tableau XXXVI. Taux de recombinaison. D.familiales = données familiales**

Gènes	N (années)	Recombinaisons/kb x 10 <sup>5</sup>		
		MaxDip	MaxHap	D. familiales
<b>GSS</b>	4897	0,8	0,3	7,0
<b>GSTT1</b>	6718	2,5	1,2	3,0
<b>GSTP1</b>	18174	0,8	0,7	100,0
<b>GSTM3</b>	5332	3,8	2,8	80,0
<b>GSTM4</b>	9435	0,4	0,4	80,0
<b>GPX2</b>	13130	7,2	5,3	0,2
<b>GPX3</b>	10569	1,7	1,7	0,1

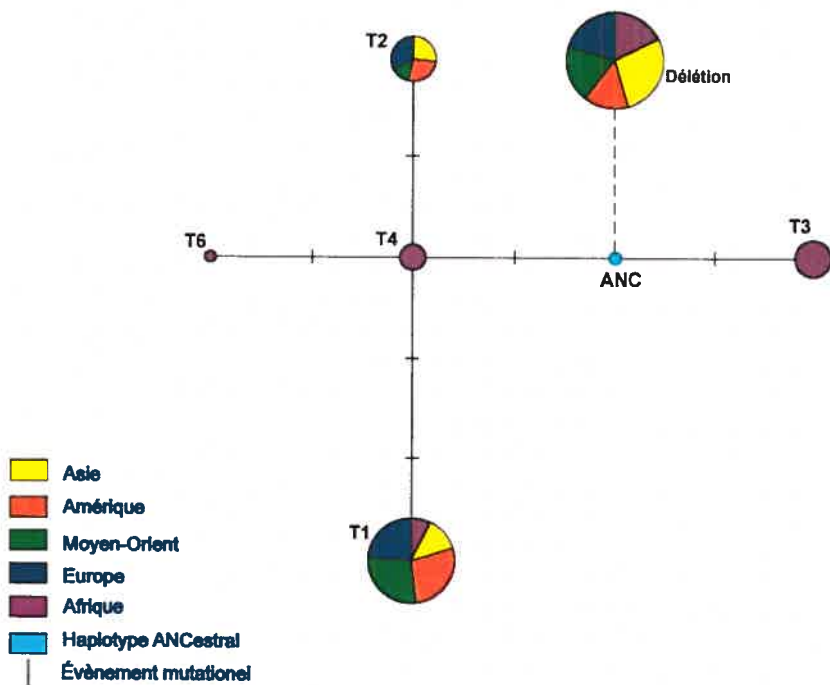
Les cladogrammes ou réseaux haplotypiques ( figures 14 à 20, pages 69 à 72) permettent de déterminer les liens génétiques entre les divers haplotypes. Ils ont été générés par Network 4.1.0.9 en utilisant un algorithme de médianes réduites (Bandelt, Forster et al. 1999). Lorsqu'il y a eu recombinaisons, le programme traduit par des mutations récurrentes. Comme ces mutations sont très peu probables et donc négligeables, les réseaux recombinants (GSTP1, GSTM3, GSTM4, GPX2, GPX3) ont été redessinés manuellement. Les réseaux sont basés sur les données haplotypiques des 80 individus des populations continentales. Les couleurs représentent chacune de ces populations (voir légende figure 15 page 69) et le carré du rayon des cercles est proportionnel à la fréquence des haplotypes. Un trait transversal représente une mutation. Certains réseaux peuvent être expliqués uniquement par des événements mutationnels (GSTT1, GSS). Dans d'autres cas, des lignes pointillées font référence à un événement de recombinaison entre deux haplotypes.

Pour GSTM3 et GSTM4, il y a un seul événement de recombinaison. Pour GSTP1, GPX2 et GPX3, plusieurs recombinaisons ont eu lieu. Le schéma global est compliqué et deux haplotypes n'ont pu être intégrés aux réseaux. XIII16 (figure 20 page 72), un haplotype asiatique rare, pourrait être issu d'une suite de deux recombinaisons entre XIII3 et XIII1, XIII2, XIII4, XIII5 ou XIII6 et ce nouvel haplotype avec l'ancestral. Mais ce scénario n'est pas la seule option possible, il pourrait s'agir d'une conversion génique. Dans un même ordre d'idées, l'haplotype XII10 (moyen-oriental, figure 19 page 71) pourrait être le résultat d'une recombinaison entre XII3 et l'ancestral suivi d'une autre recombinaison entre ce nouvel haplotype et l'ancestral.

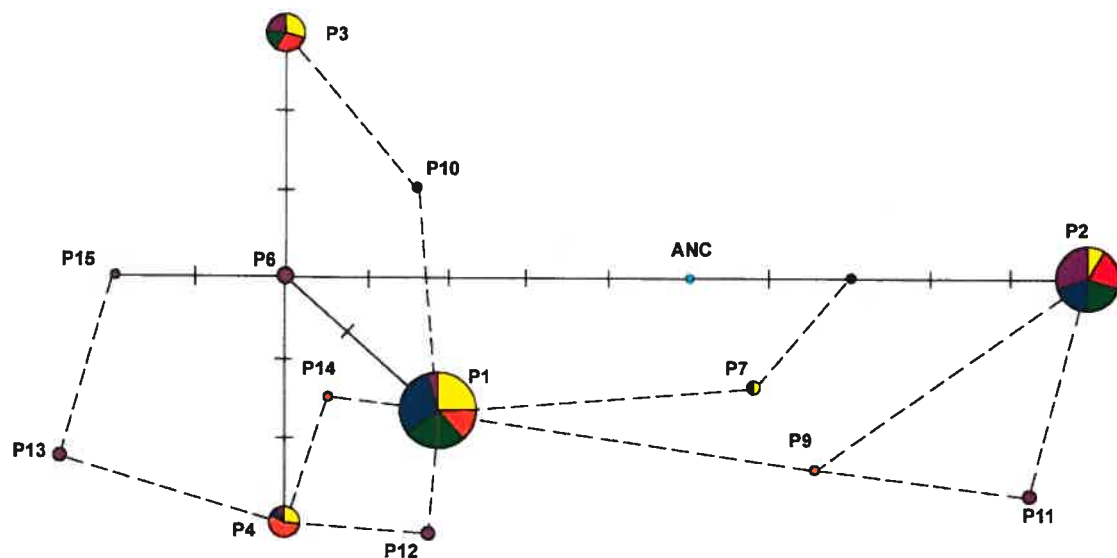
Dans l'ensemble, les réseaux les plus compliqués sont ceux avec le plus d'haplotypes (GSTP1, GPX2, GPX3). Ils sont le résultat d'une suite de mutations et de recombinaisons qui ont façonné l'histoire haplotypique de chacun des gènes.



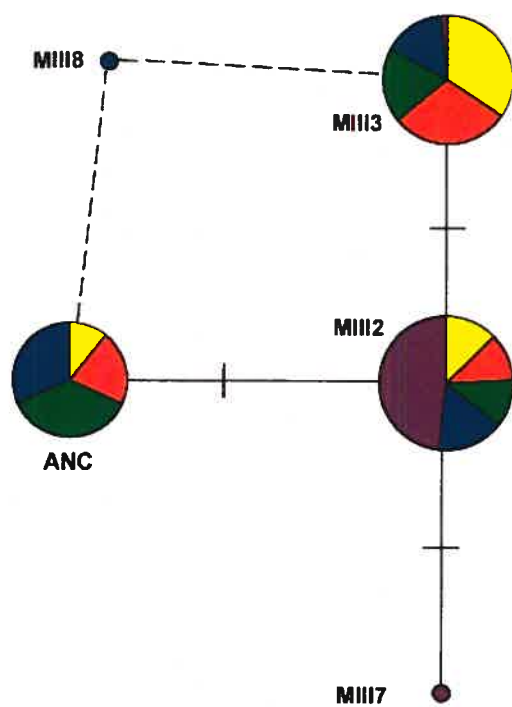
**Figure 14. Réseau haplotypique pour GSS.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.



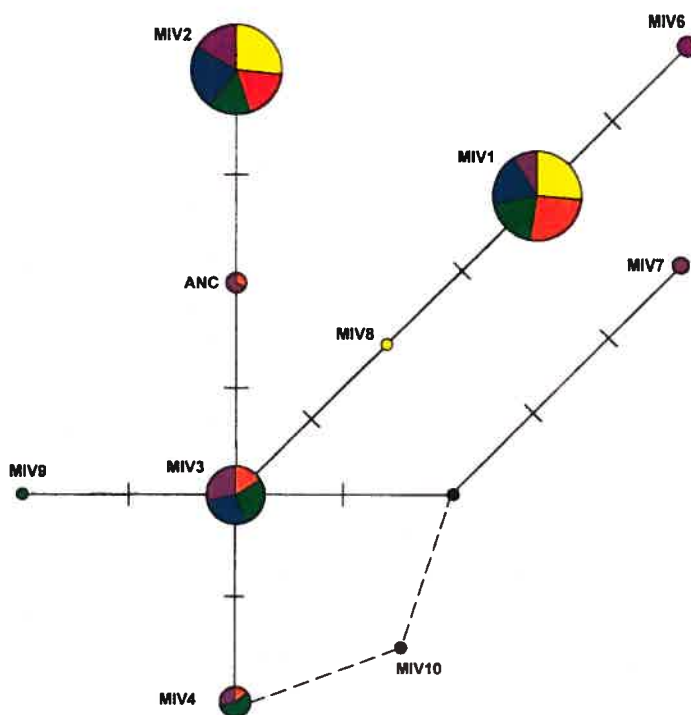
**Figure 15. Réseau haplotypique pour GSTT1.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique.



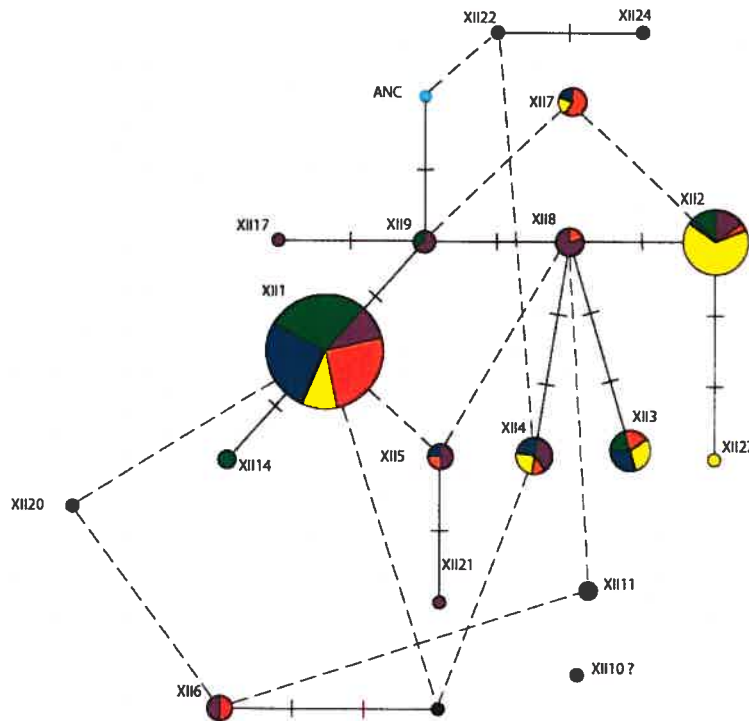
**Figure 16. Réseau haplotypique pour GSTP1.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.



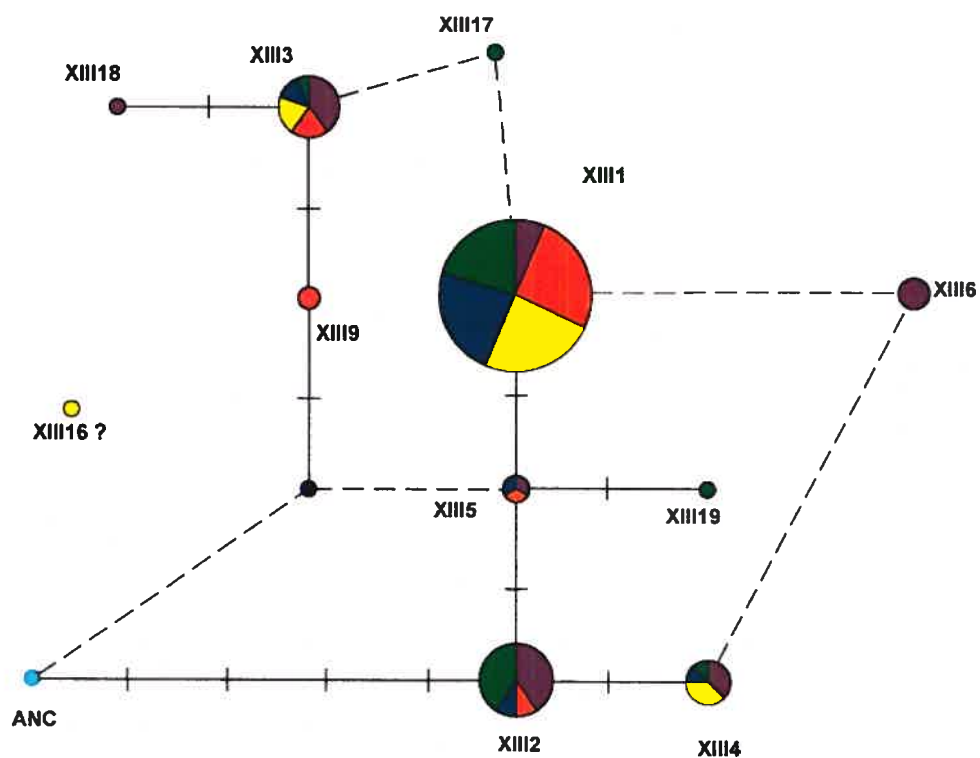
**Figure 17. Réseau haplotypique pour GSTM3.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.



**Figure 18. Réseau haplotypique pour GSTM4.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.



**Figure 19. Réseau haplotypique pour GPX2.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.



**Figure 20. Réseau haplotypique pour GPX3.** Le carré du rayon des cercles est proportionnel à la fréquence haplotypique. La légende se trouve à la figure 15.

### 3 Variations à travers les populations

Pour déterminer la proximité génétique des différentes populations, des analyses MDS (*multidimensional scaling*) à partir des moyennes de  $F_{st}$  par paire de tous les gènes étudiés (sauf GSTT1) ont été faites. Les  $F_{st}$  par paire permettent d'établir la distance génétique entre deux populations. Le but de l'analyse MDS est de représenter géométriquement sur deux dimensions la variations due à un nombre plus ou moins grand de facteurs, qui sont ici les valeurs de  $F_{st}$ . Les graphiques montrent les distances, entre chacune des données, analysées dans une matrice de covariance. Ils donnent un aperçu de la variation à travers différentes populations mondiales.

Les  $F_{st}$  ont été calculés de deux façons, à partir de la diversité nucléotidique ( $\pi$ ) et à partir de la diversité haplotypique ( $G$ ). La diversité nucléotidique est calculée à partir des fréquences des polymorphismes. Ce paramètre est donc approprié pour illustrer les différences entre des populations qui ont divergé il y a longtemps comme les populations continentales. Le paramètre de diversité haplotypique est plus adéquat pour les populations dont l'histoire est récente comme les groupes québécois. Il n'y a pas eu de nouvelle mutation à travers ces groupes, le paramètre  $\pi$  manque donc de sensibilité pour les différencier.

Les figures 21 et 22 (page 75) montrent les distances entre les populations continentales. L'analyse effectuée à partir du paramètre  $G$  montre que les distances génétiques entre les populations sont toutes plus ou moins égales. Celle effectuée à partir de  $\pi$  montrent des différences considérables. En considérant que la dimension 1 du graphique représente davantage de variabilité que la dimension 2, la population africaine est radicalement différente des autres populations. L'Amérique, l'Europe et le Moyen-Orient forment un bloc quasi indifférencié. L'Asie se distingue de ce bloc et de l'Afrique.

Les figures 23 et 24 (page 76) montrent les distances entre les groupes québécois. L'analyse effectuée à partir de  $\pi$  ne montrent pas de différences significatives dans la diversité nucléotidique de ces groupes. Celle effectuée à partir de  $G$ , montre que Montréal, Jersey-Guernesey et Charlevoix-Saguenay-Lac St-Jean ne présentent pas de



différences dans les fréquences haplotypiques. Par contre, les Canadiens-Français de Gaspésie, les loyalistes et les Acadiens se distinguent entre eux et du reste de l'échantillon québécois. À cause de la différence marquée entre les paramètres de diversité des groupes québécois pour le gène GPX3, une analyse MDS (à partir de G) a été effectuée (figure 25 page 77). Il y a une différence significative entre le groupe Charlevoix-Saguenay-Lac St-Jean et le reste des groupes québécois.

Les analyses MDS des  $F_{st}$  des populations aborigène de Taiwan et islandaise ont été effectuées même si la taille de ces échantillons rendait plus ou moins significative la comparaison avec les échantillons continentaux. Le schéma global est illustré à la figure 26 (page 77) qui représente l'analyse pour GSTM3. Taiwan est près de l'Asie, alors que l'Islande se rapproche de l'Europe et du Moyen-Orient. Taiwan et Islande sont toutes les deux éloignées génétiquement de l'Afrique.

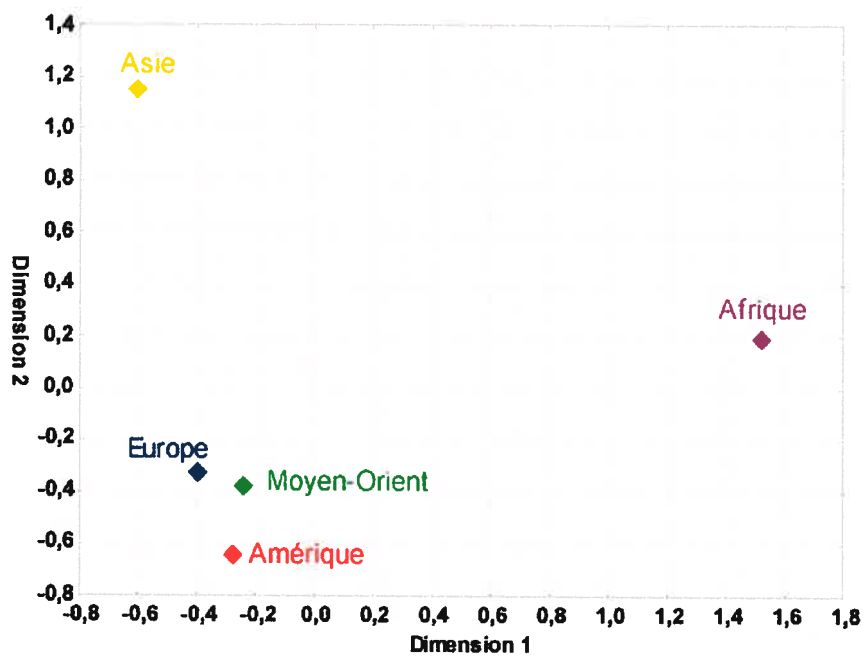


Figure 21. Analyse MDS des  $F_{st}$  calculés à partir des valeurs de  $\pi$  des échantillons continentaux pour tous les gènes étudiés.

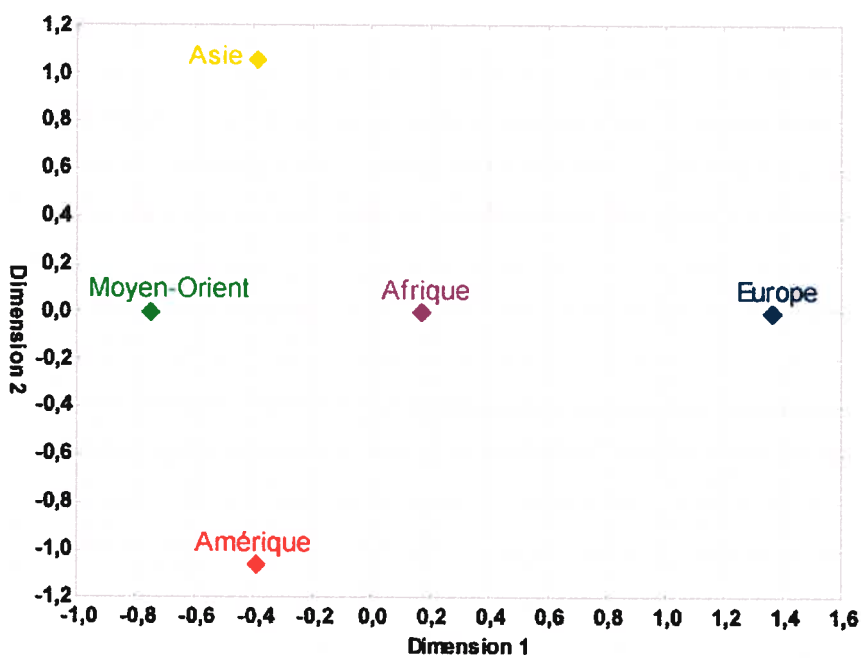


Figure 22. Analyse MDS des  $F_{st}$  calculés à partir des valeurs de G des échantillons continentaux pour tous les gènes étudiés.

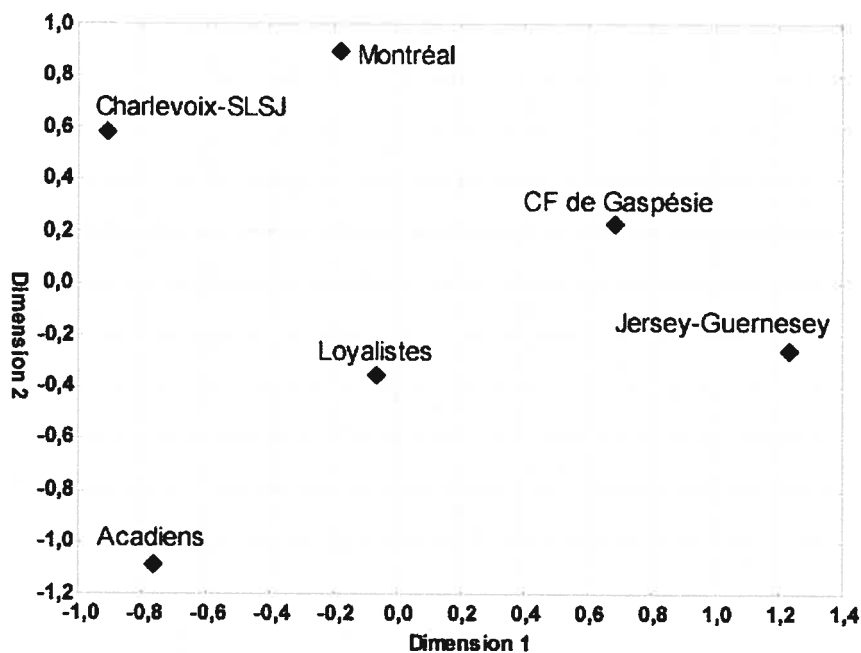


Figure 23. Analyse MDS des  $F_{st}$  calculés à partir des valeurs de  $\pi$  des groupes québécois pour tous les gènes étudiés. CF = Canadiens-Français, SLSJ=Saguenay- Lac St-Jean.

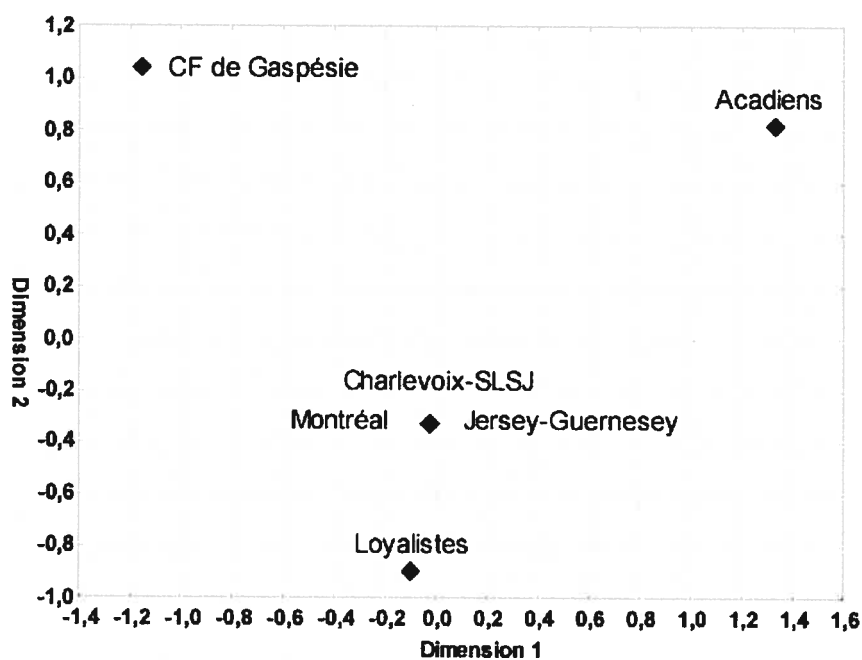
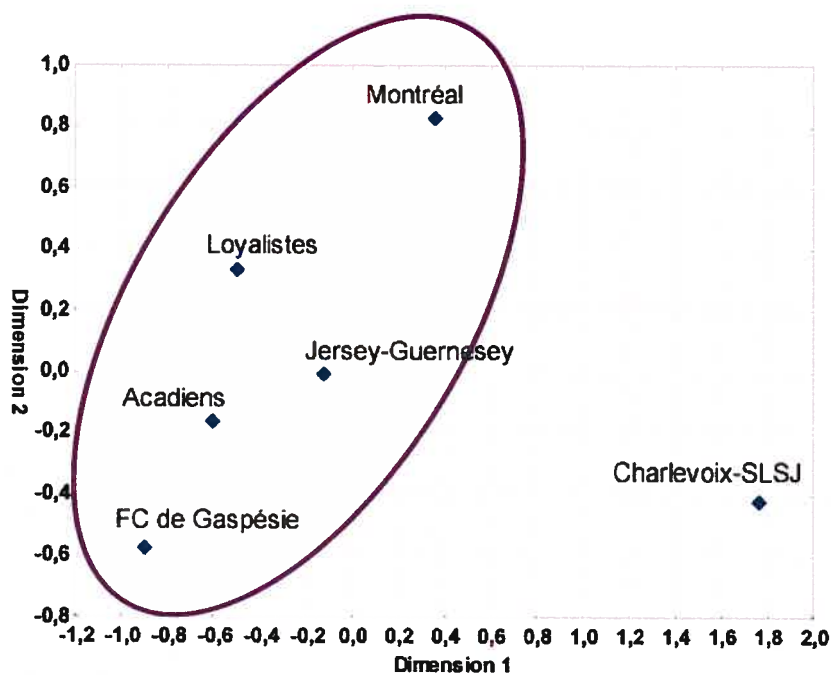
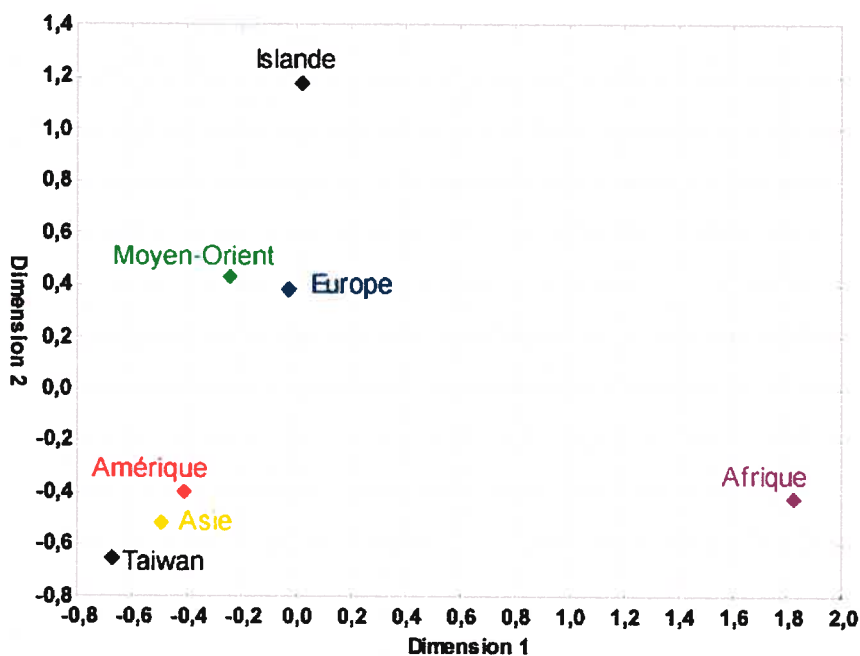


Figure 24. Analyse MDS des  $F_{st}$  calculés à partir des valeurs de  $G$  des groupes québécois pour tous les gènes étudiés. CF = Canadiens-Français, SLSJ=Saguenay- Lac St-Jean.



**Figure 25.** Analyse MDS des Fst calculés à partir des valeurs de G des groupes québécois pour GPX3. Les groupes non-différenciés sont encadrés. FC = Canadiens-Français, SLSJ=Saguenay- Lac St-Jean.



**Figure 26.** Analyse MDS des Fst calculés à partir des valeurs de G de différentes populations pour GSTM3.

#### 4 La divergence

Pour déterminer le taux de divergence, les séquences de 2 kb ont été comparées aux séquences de chimpanzé de la base de donnée UCSC. Le taux de divergence est calculé en fonction de la distance évolutive avec le chimpanzé soit dix millions d'années ( cinq millions d'années d'évolution indépendante pour l'homme et pour le chimpanzé ). Le taux de divergence moyen est de  $1,11 (x 10^{-9})$  par nucléotide par année (tableau XXXVIIA page 78). Il est légèrement supérieur pour GSTT1 (1,45) et inférieur pour GSTM4 (0,76) et GPX2 (0,89). Les taux de divergence étant très variables sur 2kb, des séquences de dix kb ont été comparées, soit cinq de chaque côté de la séquence. Le taux moyen soit  $1,42 (x 10^{-9})$  par nucléotide par année est supérieur au taux sur 2kb. Le taux de divergence de GPX2, (0,93) est toujours inférieur à la moyenne (tableau XXXIIB page 78). Il ne semble pas y avoir de rapport proportionnel entre le nombre de polymorphismes ou la diversité nucléotidique et la divergence ( figures 27 et 28 page 79).

**Tableau XXXVII: Taux de divergence des séquences de chimpanzé**

Gène	Longueur (nucléotides)	Nombre de mutations	Divergence	Taux de divergence /nt/an	ET Taux divergence	Taux de divergence /nt/an $\times 10^9$
<b>GSS</b>	<b>1584</b>	19	0,011995	1,20E-09	1,38E-10	1,20
<b>GSTT1</b>	<b>2005</b>	29	0,014464	1,45E-09	1,70E-10	1,45
<b>GSTP1</b>	<b>1644</b>	20	0,012165	1,22E-09	1,41E-10	1,22
<b>GSTM3</b>	<b>1988</b>	22	0,011066	1,11E-09	1,48E-10	1,11
<b>GSTM4</b>	<b>1833</b>	14	0,007638	7,64E-10	1,18E-10	0,76
<b>GPX2</b>	<b>2128</b>	19	0,008929	8,93E-10	1,38E-10	0,89
<b>GPX3</b>	<b>2196</b>	25	0,011384	1,14E-09	1,58E-10	1,14
<b>B</b>						
<b>GSS</b>	<b>11523</b>	223	0,019353	1,94E-09	4,72E-10	1,94
<b>GSTT1</b>	<b>11953</b>	176	0,014724	1,47E-09	4,20E-10	1,47
<b>GSTP1</b>	<b>10747</b>	151	0,014050	1,41E-09	3,89E-10	1,41
<b>GSTM3</b>	<b>5788</b>	83	0,014340	1,43E-09	2,88E-10	1,43
<b>GSTM4</b>	<b>11968</b>	186	0,015541	1,55E-09	4,31E-10	1,55
<b>GPX2</b>	<b>12007</b>	112	0,009328	9,33E-10	3,35E-10	0,93
<b>GPX3</b>	<b>12016</b>	148	0,012317	1,23E-09	3,85E-10	1,23

A. par rapport à la séquence étudiée (2Kb)

B. sur une séquence d'environ 10kb dans la région étudiée

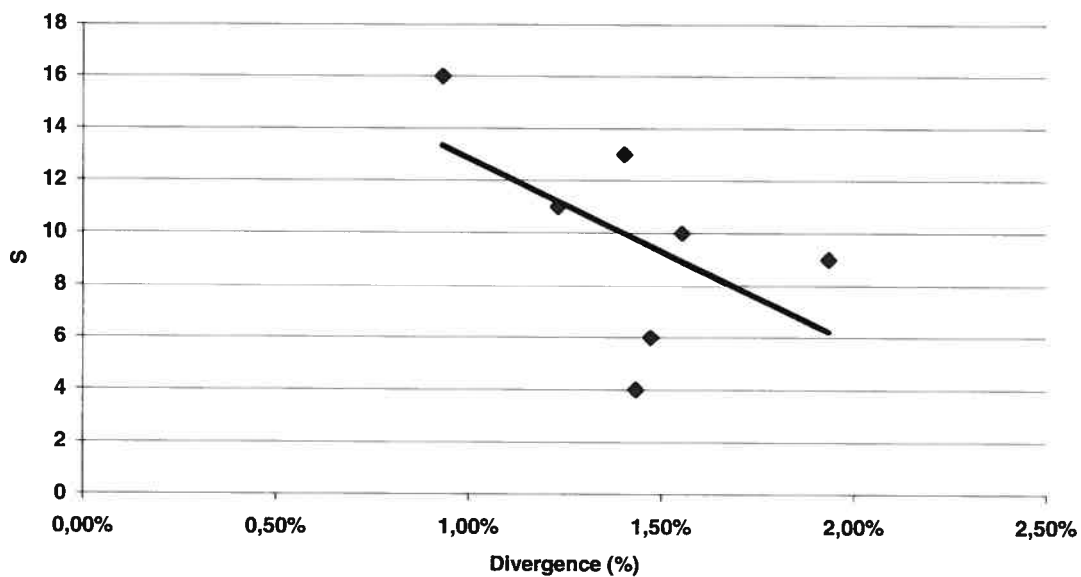


Figure 27. Nombre de polymorphismes en fonction de la divergence.

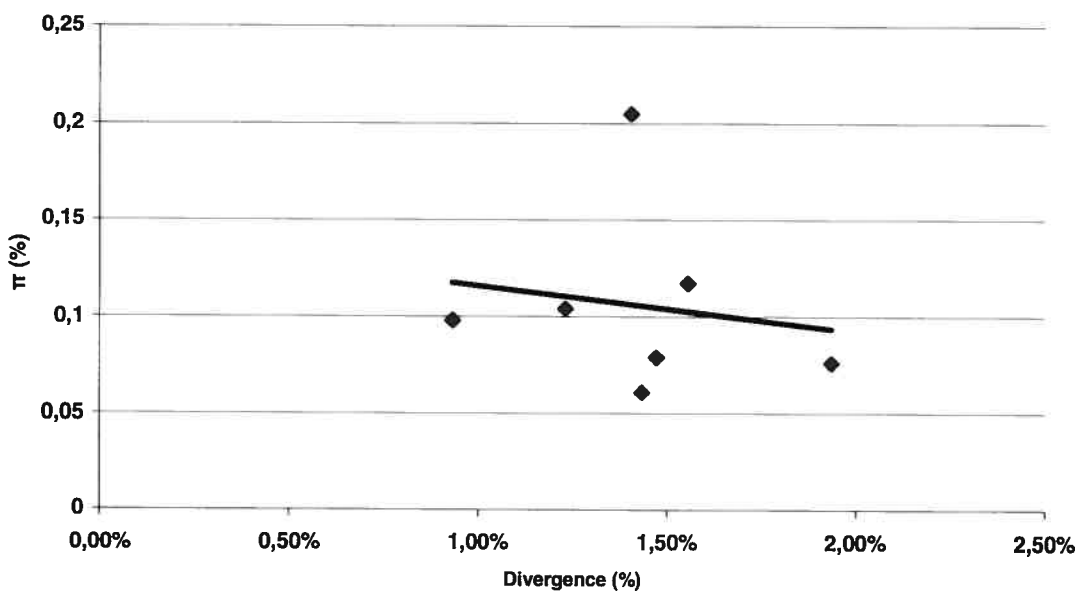


Figure 28. Diversité nucléotidique en fonction de la divergence.

## 5 Variabilité dans les parties introniques et exoniques

Afin de comparer la variabilité de la région promotrice avec les régions exoniques et introniques (REI), une cohorte spéciale qui a servi aux études d'une collègue de laboratoire (Caroline Léger), a été utilisée. Pour le gène GPX2, aucune variabilité n'a été recensée dans les sections étudiées. Par ailleurs, les données compilées dans le tableau XXXVIII (page 80) montrent une diversité non négligeable dans le promoteur. Le nombre de polymorphismes (13) est élevé et le Fst indique une subdivision modérée entre les populations. Pour GPX3, les statistiques sont présentées dans le tableau XXXIX (page 81). La région promotrice et les REI montrent des nombres de polymorphismes et d'haplotypes comparables. Il y a un polymorphisme de plus pour le promoteur et cinq haplotypes de plus pour les REI. Cependant, la diversité nucléotidique du promoteur est supérieure, 0,122 % au niveau mondial contre 0,076% pour les REI. Le Fst total est également légèrement supérieur pour le promoteur, 0,104, contre 0,086 pour les REI. Tous deux indiquent une différenciation génétique modérée entre les populations.

**Tableau XXXVIII. Statistiques pour GPX2, variabilité dans les promoteurs**

	S	$\pi$ (%)	K	G
<b>Afrique</b>	8	0,112	13	0,862
<b>Amérique</b>	7	0,065	6	0,454
<b>Asie</b>	9	0,106	8	0,740
<b>Europe</b>	8	0,092	7	0,589
<b>Moyen-Orient</b>	6	0,050	5	0,328
<b>Monde</b>	13	0,097	19	0,681
Fst Total	0,130*			

\* = valeur de Fst significative. Les données sont tirées de la cohorte spéciale.

**Tableau XXXIX. Statistiques comparatives pour GPX3, variabilité dans les promoteurs et dans la partie intronique/exonique**

	Promoteur				Exon et Intron			
	S	$\pi$ (%)	K	G	S	$\pi$ (%)	K	G
<b>Afrique</b>	9	0,128	8	0,733	8	0,089	10	0,800
<b>Amérique</b>	8	0,093	3	0,381	5	0,062	7	0,703
<b>Asie</b>	7	0,111	4	0,413	7	0,072	7	0,716
<b>Europe</b>	8	0,055	4	0,518	8	0,071	8	0,762
<b>Moyen-Orient</b>	9	0,118	5	0,630	7	0,063	5	0,615
<b>Monde</b>	10	0,112	10	0,633	9	0,076	15	0,755
Fst total	0,104*				0,086*			

\* = valeur de Fst significative. Les données sont tirées de la cohorte spéciale.



**Chapitre IV**

**DISCUSSION**

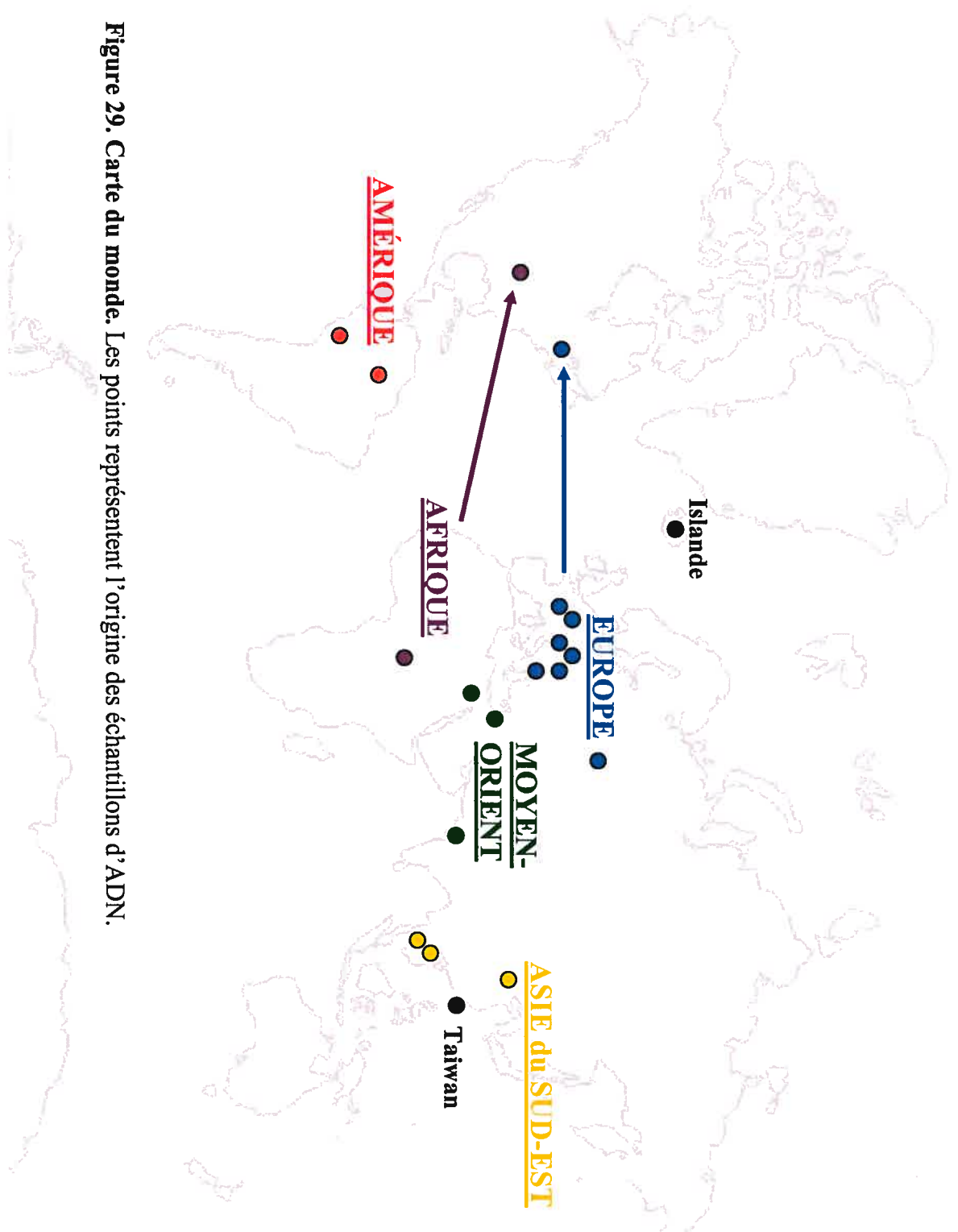


Figure 29. Carte du monde. Les points représentent l'origine des échantillons d'ADN.

Dans ce projet, nous avons voulu analyser les mécanismes génétiques et évolutifs à l'origine de la diversité dans les promoteurs des gènes impliqués dans le métabolisme du glutathion. Ces gènes ont été choisis à cause de leurs rôles dans la protection contre les oxydants et dans la détoxification qui les placent au premier plan dans le métabolisme des xénobiotiques tels la nourriture et les médicaments. Étant donné la diversité dans les diètes entre les populations mondiales, les enzymes du métabolisme glutathion se sont vraisemblablement adaptés à ces différences. Par conséquent, ils sont variables dans leur efficacité comme dans leurs substrats. Ils réagissent différemment aux oxydants et ne métabolisent pas les médicaments de la même façon. Ce projet de caractérisation de la diversité des promoteurs dans les gènes du métabolisme du glutathion vise à identifier des variants qui pourraient faire l'objet d'étude d'association avec certaines maladies ou réponses thérapeutiques. Nous croyons qu'il serait souhaitable de connaître l'impact réel de la diversité génétique sur ces processus.

### **Diversité dans les gènes de la voie métabolique du glutathion**

La distribution géographique des populations qui font l'objet de cette étude est illustrée dans les figures 29 (page 83) et 30 (page 88). La figure 29 montre la carte du monde et la « définition » des populations continentales. Ainsi, l'échantillon africain inclut des Afro-américains et l'échantillon européen, des Québécois.

Les polymorphismes géotypés ont été identifiés à partir d'un échantillon de 80 chromosomes appartenant aux cinq populations continentales. Statistiquement, la taille de cet échantillon de départ permet d'identifier des polymorphismes présents une fois sur 80 (1,25%) ou plus dans la population mondiale. Les polymorphismes dans les populations sont identifiés s'ils sont présents une fois sur 16 (6,25%) ou plus. Soixante-neuf polymorphismes ont été identifiés sur la séquence promotrice de sept gènes. Ces polymorphismes ont été utilisés pour géotyper les 80 chromosomes initiaux, 80 supplémentaires provenant des mêmes populations continentales et 284 autres provenant d'autres populations.

En général, les résultats de diversité présentés ici sont en accord avec plusieurs études qui observent des taux de diversité plus élevés dans la population africaine que dans les non-africains ((Jorde, Rogers et al. 1997; Jorde, Watkins et al. 2000; Garte, Gaspari et al. 2001; Romualdi, Balding et al. 2002). La grande diversité africaine comparée aux autres populations continentales est compatible avec l'hypothèse que l'homme moderne est apparu en Afrique et a ensuite colonisé les autres continents (Stoneking 1994). Les résultats tendent également à montrer que l'Amérique, l'Europe et le Moyen-Orient ont des populations qui partagent beaucoup de leur diversité. Selon la théorie de l'origine africaine récente, les populations européenne, moyenne-orientale et asiatique seraient issues d'une vague de migrations hors de l'Afrique datant d'environ 100 000 ans (Nei and Roychoudhury 1993). La population amérindienne origine de migrations asiatiques il y a dix à trente mille ans (Bianchi, Catanesi et al. 1998; Ruiz-Linares, Ortiz-Barrientos et al. 1999; Uinuk-Ool, Takezaki et al. 2002).

Seuls, les paramètres de diversité de GSTM3 montrent une diversité réduite pour l'Afrique. Les tests de neutralité ne montrent pas de sélection et cette réduction de la diversité semble provenir de la dérive génétique. La singularité africaine se présente également dans les résultats pour GPX2. Nous avons trouvé 17 haplotypes dont 16 haplotypes mineurs. Les résultats du test  $F_s$  de  $F_u$  pour ce gène suggèrent de la sélection positive. Cependant, les fréquences haplotypiques des polymorphismes de GPX2 ne violent pas la loi de Hardy-Weinberg. De plus, le taux de divergence de ce gène est très bas. Le gène est donc très conservé. Les polymorphismes sont majoritairement apparus dans la population africaine. Le résultat du test  $F_s$  de  $F_u$  montre aussi que la population est en expansion d'où l'excès d'haplotypes rares.

Il y a plusieurs événements évolutifs qui peuvent expliquer la divergence des Asiatiques par rapport au bloc Europe, Moyen-Orient et Amérique pour les gènes étudiés. Premièrement, la variabilité divergente de l'Asie est due à la dérive génétique ou la sélection et l'un ou l'autre de ces événements a eu lieu après les migrations vers l'Amérique. Les tests de neutralité ne sont pas significativement différents de zéro

mais les indices de diversités nucléotidiques et haplotypiques généralement plus bas pour l'Asie que pour les autres populations continentales suggèrent des signes de dérive.

Deuxièmement, la population amérindienne a subi une réduction de la variabilité asiatique ancestrale et porte des haplotypes provenant d'autres populations à la suite de métissage. Plusieurs études font état de la réduction de la variabilité de la population amérindienne suite aux sanglantes guerres de colonisation (Yunis, Yunis et al. 2001; Cifuentes, Morales et al. 2004; Araujo, Novaes et al. 2005). Ce goulot d'étranglement est perceptible dans le nombre inférieur de polymorphismes en Amérique par rapport à l'Asie notamment pour les gènes GSS et GPX3. Ces valeurs peuvent aussi être interprétées comme un effet fondateur au sein de la population amérindienne. Un petit nombre de gens ont peuplé l'Amérique apportant avec eux une fraction de la variabilité asiatique. Cependant, pour plusieurs gènes, ( GSTP1, GSTM4, GPX2 et GPX3) l'Amérique possède plus d'haplotypes que l'Asie. Plusieurs de ces haplotypes non asiatiques sont partagés par la population amérindienne et les Européens ou les Africains (GSTM4 : ancestral, MIV3, MIV4 ; GPX3 : XIII5, XIII2 ; GPX2 : XII5, XII6 et XII8). Ils indiquent un métissage des Amérindiens avec les Européens et subséquemment, les Africains à la suite de la colonisation et de l'introduction des esclaves noirs en Amérique.

Troisièmement, les résultats de cette étude semblent indiquer qu'il y a peu de liens génétiques entre les Asiatiques du sud-est ( la population « asiatique » étudiée ici ) et les Amérindiens. En fait, l'Asie du sud-est semblent être isolée génétiquement par rapport au bloc Europe/Moyen-Orient/Amérique. Cette population pourrait vraisemblablement provenir d'une vague indépendante de migration hors d'Afrique. Les populations à l'origine de la colonisation du continent Américain seraient originaire de Sibérie et du nord de l'Asie. La proximité de l'accès au détroit de Béring semble appuyer cette hypothèse.

En somme, la population asiatique présente des caractéristiques de variabilité qui la distingue des autres populations continentales. Ces caractéristiques peuvent être dues à la dérive génétique. L'Amérique, le Moyen-Orient et l'Europe présentent une variabilité très semblable. Cependant, les échantillons continentaux représentent plusieurs pays en peu d'individus, ce qui donne un aperçu plutôt général de la distribution de la variabilité. Nous avons voulu préciser notre étude en analysant des groupes représentant une population plus restreinte. Nous avons choisi une sous-population européenne : les Québécois.



**Figure 30. Carte du Québec.**

Les points noirs représentent les régions étudiées.

## La diversité québécoise

La carte du Québec montre les trois régions étudiées soit Montréal, Charlevoix et le Saguenay-Lac St-Jean et la Gaspésie. La variabilité a été analysée dans un échantillon québécois de 126 individus provenant de différentes régions du Québec et se réclamant de diverses ascendances. Les paramètres de diversité sont ici beaucoup plus uniformes à cause de l'histoire récente de la colonisation. En effet, les premiers colons français sont arrivés en Amérique, il y a presque quatre cents ans ou une douzaine de générations. Ils ont formé des colonies le long du Saint-Laurent, entre autres à Québec et à Montréal. La région de Charlevoix fait partie de celles qui ont été colonisées au 17<sup>e</sup> siècle. Au 19<sup>e</sup> siècle, suite à une augmentation de la population, certaines régions, comme le Saguenay-Lac St-Jean, ont été ouvertes à la colonisation. La majorité des premiers colons de cette région provenait de la région voisine, Charlevoix (Heyer 1995). Au 17<sup>e</sup>, d'autres colons français, appelés plus tard Acadiens, se sont établis dans la Baie de Fundy et au Cap-Breton. En 1755, suite à la déportation, certains se sont retrouvés jusqu'en Louisiane. L'immigration des loyalistes a eu lieu vers la fin du 18<sup>e</sup> siècle lors de la guerre d'indépendance américaine. L'immigration anglo-normande s'est produite vers la fin du 18<sup>e</sup> et au début du 19<sup>e</sup> siècle (Ommer 1991). Ce court laps de temps n'a pas permis la propagation de nouvelles mutations québécoises et les paramètres de diversité des échantillons québécois ressemblent beaucoup à ceux des Européens. D'ailleurs le nombre de polymorphismes identifié chez les québécois dans cette étude est toujours égale ou plus petit que celui identifié chez les européens et ce même si sept fois plus de québécois des trois régions ciblées ont été analysés. La diversité génétique entre les sous-groupes québécois est plus faible à cause du petit nombre de fondateurs à l'origine de la colonie (Scriver 2001). L'indice de diversité nucléotidique n'est donc pas un paramètre très évocateur pour les groupes québécois. La diversité haplotypique est plus sensible et traduit les différences de fréquences haplotypiques entre les groupes.



En général, les valeurs des paramètres de diversité des groupes québécois suivent celles des Européens. Les groupes québécois partagent également beaucoup de diversité, ce qui leur confère une certaine homogénéité. D'autre part, la taille de l'échantillon québécois analysé (126 individus) permet de faire ressortir certains haplotypes rares qu'on a pas vus dans le groupe européen (16 individus dont 8 Canadiens-Français), par exemple l'haplotype S11 de GSS retrouvé dans les groupes de Canadiens-Français de Gaspésie et Jersey-Guernesey et dans la population asiatique. On ne le retrouve pas dans l'échantillon européen ou amérindien. Il peut s'agir d'un haplotype rare que la taille réduite de ces échantillons empêche de trouver. L'haplotype P03 de GSTP1 présent à 8,7% dans l'échantillon québécois n'est pas retrouvé dans la cohorte européenne. En revanche, on le retrouve chez 16,6% des chromosomes amérindiens. P03 pourrait être un haplotype amérindien introduit dans la population québécoise à la suite de métissage. Il pourrait s'agir du même scénario pour MIV4 (GSTM4).

Pour le gène GPX3, XIII01 est l'unique haplotype majeur chez les Européens (78,6%) et chez les groupes québécois sauf le groupe Charlevoix-Saguenay Lac St-Jean. Pour ce dernier, il y a trois haplotypes importants, une distribution qui se démarque de celle des autres groupes québécois en présentant plus de variabilité. Les paramètres de diversité de Charlevoix-Saguenay-Lac St-Jean sont d'ailleurs beaucoup plus élevés ( $\pi = 0,149\%$ ,  $G = 0,754$ ) que ceux des autres groupes québécois et de la population européenne. L'indice  $F_{st}$  indique que 2,9% de la variation génétique sont dus aux différences entre les groupes, ce qui est non négligeable vu l'homogénéité génétique de la population québécoise. L'origine de cette diversité charlevoisienne ne trouve pas d'explication dans la littérature. Le hasard des fluctuations alléliques suite à l'effet fondateur semblent en être la cause.

La variabilité dans les promoteurs des gènes du glutathion est, en général, homogène dans les groupes québécois. Cela témoigne du petit nombre de fondateurs à l'origine de la population. L'analyse MDS à partir des paramètres de diversité haplotypique (G) de tous les gènes étudiés montre que Montréal, les descendants anglo-normands

de Gaspésie et Charlevoix – Saguenay Lac Saint-Jean ont des populations très semblables. Les loyalistes se distinguent un peu de ces groupes. Toutefois, ce sont les populations canadienne-française de Gaspésie et acadienne qui montrent le plus de caractéristiques distinctives.

### **Aborigènes de Taiwan et Islandais : des populations types**

Les aborigènes de Taiwan et les Islandais sont deux populations insulaires qui ont subi une isolation géographique et culturelle qui transparait dans leur diversité génétique. Les aborigènes de Taiwan constituent une population qui a été isolée génétiquement pendant plusieurs milliers d'années. La population islandaise présentent un exemple d'effet fondateur bien connu dans la littérature.

Des données paléogéographiques suggèrent que Taiwan était connecté au continent asiatique durant la dernière ère glaciaire (Voris 2000). Les premiers humains modernes se sont probablement installés, il y a 20 000 à 30 000 ans et se sont retrouvés isolés par la montée des eaux vers 11 000 ans avant notre ère. Taiwan compte aujourd'hui 22 millions d'habitants dont 1,5% d'Aborigènes (Tajima, Sun et al. 2003). L'Islande a été colonisée il y a 1100 ans par environ 8000 à 16 000 fondateurs. La population islandaise compte aujourd'hui 280 000 personnes (Helgason, Nicholson et al. 2003). Les résultats de la présente étude suggèrent que les populations aborigènes de Taiwan et les Islandais forment des groupes beaucoup plus homogènes que les autres populations mondiales ou les groupes québécois. Nous avons trouvé nettement moins de polymorphismes et d'haplotypes chez ces deux groupes que chez les autres populations analysées. Particulièrement pour les gènes GSS et GPX2. De plus, à part pour la présence de la délétion, l'haplotype T1 est fixé dans la population islandaise. Les analyses de fréquences haplotypiques montrent que les Aborigènes de Taiwan sont génétiquement plus diversifiés que les Islandais et plus près des Asiatiques que des autres populations. Cette observation suit le modèle de l'histoire de la colonisation de Taiwan par les Asiatiques il y a 20 000 ans. Elle pourrait traduire également un certain métissage avec la population d'immigrants chinois récents. D'autres études sur les liens génétiques entre différentes tribus aborigènes de Taiwan et des populations asiatiques montrent des résultats semblables (Wei, Chiang et al. 1999; Tajima, Sun et al. 2003). Les résultats pour l'échantillon islandais décrivent une population très homogène avec peu de polymorphismes et d'allèles différents. Cette population est plus homogène que l'échantillon européen.

Elle se rapproche davantage de l'échantillon québécois, notamment dans ses fréquences haplotypiques. Cette homogénéité de la population islandaise semble être due à la dérive génétique et à l'effet fondateur. Des études de plus grande envergure sur les Islandais ont montré les traces génétiques de la présence de ces mécanismes évolutifs (Helgason, Sigureth ardottir et al. 2000; Helgason, Nicholson et al. 2003).

### **GSTs, GPXs et variabilité**

En comparant les paramètres de diversité des gènes étudiés dans ce projet, on remarque que les GSTs ont des degrés de variabilité très différents. Même les GSTs d'une même famille tel que GSTM3 et GSTM4 ont des différences marquantes. En effet, si le promoteur de GSTM3 ne contient que quatre polymorphismes, celui de GSTM4 en contient dix et son histoire haplotypique semble beaucoup plus complexe.

Par ailleurs, la délétion de GSTT1 le distingue des autres GSTs. Un autre gène codant pour une GST, GSTM1, est également fréquemment délété. Il est également absent de 70% des chromosomes mondiaux (Xu, Wang et al. 1998). Dans le cas de GSTT1, les résultats de la présente étude montrent qu'il y a 44% des chromosomes mondiaux qui portent une délétion de ce gène. 24,7% des individus n'ont aucune copie de ce gène et 40,3% n'en ont qu'une seule. Cette distribution ne dévie pas de l'équilibre Hardy-Weinberg et la fréquence élevée de cette délétion dans la population asiatique (63,3%) semble être due à la dérive génétique. Ces résultats concordent avec ceux déjà publiés par Sprenger et al. (20% génotype nul et 44% une seule copie du gène) (Sprenger, Schlagenhauser et al. 2000). Ces fréquences élevées pourraient indiquer que le gène n'accomplit pas une fonction essentielle et qu'une autre transférase peut remplacer GSTT1. Cependant, plusieurs études semblent indiquer une association entre le génotype nul de GSTT1 avec certaines formes de cancers et d'autres maladies dégénératives (Goto, Yoneda et al. 1996; Howells, Redman et al. 1998; Matthey, Hassell et al. 1999). Par contre, très peu d'études d'associations distinguent les effets sur les porteurs d'une seule copie du gène par rapport à ceux sur les porteurs de deux copies de GSTT1. Cette distinction pourrait approfondir notre compréhension des impacts de la délétion de GSTT1. De plus, il n'existe aucune étude sur l'histoire de la délétion de GSTT1. On ne sait toujours pas s'il s'agit d'un événement qui a eu lieu une seule fois ou plusieurs fois à travers les populations. La région importante (18 Kb) d'identité de séquence semble favoriser la deuxième hypothèse mais une étude des haplotypes de part et d'autre de cette délétion pourrait être nécessaire afin d'obtenir la réponse définitive.

Bien que très différentes dans leurs paramètres de diversité, les GSTs présentent des taux de divergences très semblables. Cela suggère une évolution parallèle au sein de la voie métabolique du glutathion.

Les GPXs se distinguent des GSTs par leur nombre élevé de polymorphismes (GPX2 16 et GPX3 11) et d'haplotypes ( GPX2 18 et GPX3 11). Les taux de recombinaisons calculés à partir de marqueurs microsatellites (Kong, Murphy et al. 2004) indiquent que ces deux gènes seraient moins souvent recombinaisonnés que les autres gènes étudiés. Cependant, les programmes d'estimation MaxHap et Max Dip (Hudson 2001) indiquent que GPX3 a un taux de recombinaison moyen et que GPX2 serait une région hautement recombinaisonnée. Les données microsatellites pour GPX2 sont incomplètes et pourraient affecter le calcul du taux de recombinaison mais cet écart entre les deux estimations est tel qu'il suggère un autre évènement génétique, la conversion génique. Des études indiquent d'ailleurs que ce phénomène se produit plus fréquemment qu'on serait porté à le croire (Frisse, Hudson et al. 2001). La conversion génique expliquerait la complexité des réseaux haplotypiques pour GPX2 et GPX3 et l'impossibilité de représenter certains haplotypes dans les cladogrammes.

Dans un autre ordre d'idées, nous avons analysé la variabilité exonique et intronique de GPX2 et GPX3. Dans les deux cas, la diversité nucléotidique est supérieure dans les promoteurs. Pour GPX3, elle est de 0,122% contre 0,076% dans la région exonique et intronique et pour GPX2 est de 0,097% dans le promoteur et de 0% dans le gène. Ces valeurs montrent l'importance des régions régulatrices dans l'évolution. La diversité de GPX2 est d'ailleurs entièrement dans le promoteur. Le taux de divergence de GPX2 ( $0,93 \times 10^{-9}$  par nucléotide par année ) indique que ce gène est très conservé et confirme les valeurs de diversité.

## Conclusion

Une recherche du terme *glutathione* sur Pubmed génère plus de 66 000 entrées. Ce nombre illustre l'importance de l'implication de la voie métabolique du glutathion dans une panoplie de maladies. Plusieurs polymorphismes ont été identifiés dans les promoteurs des gènes impliqués dans cette voie. Ils pourront éventuellement être utilisés dans des études d'association. Ces polymorphismes montrent des distributions ethniques caractéristiques qu'il est important de connaître pour faciliter le choix des cohortes lors d'études d'association. Les Africains présentent plus de polymorphismes et d'haplotypes que les autres populations continentales. Ce phénomène reflète l'antériorité de l'origine de cette population par rapport aux autres. D'autres parts, les populations moyen-orientale, européenne et amérindienne montrent des caractéristiques de diversité communes. Ce partage de diversité témoigne à la fois d'une origine commune et de rencontres entre ces populations suivi de métissages successifs. Les caractéristiques de la population asiatique indiquent que la dérive génétique a différencié cette population des autres. La population islandaise est un sous-groupe européen homogène. Les caractéristiques de diversité des aborigènes de Taiwan s'apparentent à celles de l'Asie. La population québécoise présente les mêmes polymorphismes que la population européenne et est très homogène. Cependant quelques singularités, comme les fréquences haplotypiques de GPX3 pour Charlevoix et Saguenay-Lac Saint-Jean, mériteraient d'être investiguées dans le cadre d'études plus approfondies. Avec l'avènement du projet HapMap, de plus en plus de chercheurs se rendent compte de l'importance des haplotypes par rapport aux polymorphismes simples dans les études d'association et de linkage. Dans le cadre de mon projet, j'ai identifié plusieurs de ces haplotypes. Ils pourront être utilisés en épidémiologie ou pharmacogénétique pour étudier la relation entre la susceptibilité aux maladies ou la réponse thérapeutique et les variants dans l'ADN humain. Ces disciplines impliquent de comprendre les effets du profil génétique d'un individu. Le défi est de traiter avec des différences biologiques significatives à travers les populations mondiales. La génétique des populations devient alors un outil essentiel pour découvrir et étudier la distribution de cette variation.

## **BIBLIOGRAPHIE**



- Ames, B. N., M. K. Shigenaga, et al. (1993). "Oxidants, antioxidants, and the degenerative diseases of aging." Proc Natl Acad Sci U S A 90(17): 7915-22.
- Anderson, M. E. (1998). "Glutathione: an overview of biosynthesis and modulation." Chem Biol Interact 111-112: 1-14.
- Araujo, F. G., F. C. Novaes, et al. (2005). "Prevalence of deltaF508, G551D, G542X, and R553X mutations among cystic fibrosis patients in the North of Brazil." Braz J Med Biol Res 38(1): 11-5.
- Armstrong, R. N. (1997). "Structure, catalytic mechanism, and evolution of the glutathione transferases." Chem Res Toxicol 10(1): 2-18.
- Arthur, J. R. (2000). "The glutathione peroxidases." Cell Mol Life Sci 57(13-14): 1825-35.
- Avissar, N., J. N. Finkelstein, et al. (1996). "Extracellular glutathione peroxidase in human lung epithelial lining fluid and in lung cells." Am J Physiol 270(2 Pt 1): L173-82.
- Avissar, N., J. R. Slemmon, et al. (1991). "Partial sequence of human plasma glutathione peroxidase and immunologic identification of milk glutathione peroxidase as the plasma enzyme." J Nutr 121(8): 1243-9.
- Awasthi, Y. C., E. Beutler, et al. (1975). "Purification and properties of human erythrocyte glutathione peroxidase." J Biol Chem 250(13): 5144-9.
- Bandelt, H. J., P. Forster, et al. (1999). "Median-joining networks for inferring intraspecific phylogenies." Mol Biol Evol 16(1): 37-48.
- Bianchi, N. O., C. I. Catanesi, et al. (1998). "Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations." Am J Hum Genet 63(6): 1862-71.
- Bierl, C., B. Voetsch, et al. (2004). "Determinants of human plasma glutathione peroxidase (GPx-3) expression." J Biol Chem 279(26): 26839-45.
- Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science 281(5373): 61-3.
- Bourgeois, S. and D. Labuda (2004). "Dynamic allele-specific oligonucleotide hybridization on solid support." Anal Biochem 324(2): 309-11.
- Breathnach, R. and P. Chambon (1981). "Organization and expression of eucaryotic split genes coding for proteins." Annu Rev Biochem 50: 349-83.
- Briviba, K., L. O. Klotz, et al. (1999). "Defenses against peroxynitrite." Methods Enzymol 301: 301-11.
- Burke, T. W. and J. T. Kadonaga (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." Genes Dev 10(6): 711-24.
- Butler, J. E. and J. T. Kadonaga (2002). "The RNA polymerase II core promoter: a key component in the regulation of gene expression." Genes Dev 16(20): 2583-92.
- Chakravarti, A. (2001). "To a future of genetic medicine." Nature 409(6822): 822-3.
- Chu, F. F., J. H. Doroshov, et al. (1993). "Expression, characterization, and tissue distribution of a new cellular selenium-dependent glutathione peroxidase, GSHPx-GI." J Biol Chem 268(4): 2571-6.

- Chu, F. F., R. S. Esworthy, et al. (1992). "Expression of plasma glutathione peroxidase in human liver in addition to kidney, heart, lung, and breast in humans and rodents." Blood 79(12): 3233-8.
- Chu, F. F., H. A. Rohan de Silva, et al. (1996). "Polymorphism and chromosomal localization of the GI-form of human glutathione peroxidase (GPX2) on 14q24.1 by in situ hybridization." Genomics 32(2): 272-6.
- Cifuentes, L., R. Morales, et al. (2004). "DYS19 and DYS199 loci in a Chilean population of mixed ancestry." Am J Phys Anthropol 125(1): 85-9.
- Corden, J., B. Wasyluk, et al. (1980). "Promoter sequences of eukaryotic protein-coding genes." Science 209(4463): 1406-14.
- Dahl, N., M. Pigg, et al. (1997). "Missense mutations in the human glutathione synthetase gene result in severe metabolic acidosis, 5-oxoprolinuria, hemolytic anemia and neurological dysfunction." Hum Mol Genet 6(7): 1147-52.
- Davies, K. J. (1995). "Oxidative stress: the paradox of aerobic life." Biochem Soc Symp 61: 1-31.
- Deneke, S. M. and B. L. Fanburg (1989). "Regulation of cellular glutathione." Am J Physiol 257(4 Pt 1): L163-73.
- Forstrom, J. W., J. J. Zakowski, et al. (1978). "Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine." Biochemistry 17(13): 2639-44.
- Frisse, L., R. R. Hudson, et al. (2001). "Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels." Am J Hum Genet 69(4): 831-43.
- Fu, Y. X. and W. H. Li (1993). "Statistical tests of neutrality of mutations." Genetics 133(3): 693-709.
- Garte, S., L. Gaspari, et al. (2001). "Metabolic gene polymorphism frequencies in control populations." Cancer Epidemiol Biomarkers Prev 10(12): 1239-48.
- Goto, I., S. Yoneda, et al. (1996). "Prognostic significance of germ line polymorphisms of the CYP1A1 and glutathione S-transferase genes in patients with non-small cell lung cancer." Cancer Res 56(16): 3725-30.
- Habdous, M., G. Siest, et al. (2004). "[Glutathione S-transferases genetic polymorphisms and human diseases: overview of epidemiological studies]." Ann Biol Clin (Paris) 62(1): 15-24.
- Hall, A. G., P. Autzen, et al. (1994). "Expression of mu class glutathione S-transferase correlates with event-free survival in childhood acute lymphoblastic leukemia." Cancer Res 54(20): 5251-4.
- Hayes, J. D., J. U. Flanagan, et al. (2004). "Glutathione Transferases." Annu Rev Pharmacol Toxicol.
- Hayes, J. D. and D. J. Pulford (1995). "The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance." Crit Rev Biochem Mol Biol 30(6): 445-600.
- Helgason, A., G. Nicholson, et al. (2003). "A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift." Ann Hum Genet 67(Pt 4): 281-97.

- Helgason, A., S. Sigurethardottir, et al. (2000). "mtDNA and the origin of the Icelanders: deciphering signals of recent population history." Am J Hum Genet 66(3): 999-1016.
- Heyer, E. (1995). "Genetic consequences of differential demographic behaviour in the Saguenay region, Quebec." Am J Phys Anthropol 98(1): 1-11.
- Howells, R. E., C. W. Redman, et al. (1998). "Association of glutathione S-transferase GSTM1 and GSTT1 null genotypes with clinical outcome in epithelial ovarian cancer." Clin Cancer Res 4(10): 2439-45.
- Hudson, R. R. (2001). "Two-locus sampling distributions and their application." Genetics 159(4): 1805-17.
- Jakobsson, P. J., R. Morgenstern, et al. (1999). "Common structural features of MAPEG -- a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism." Protein Sci 8(3): 689-92.
- Jobling MA, Hurles M., Tyler-Smith C (2004). Human Evolutionary Genetics. New York, Garland Science.
- Jorde, L. B., A. R. Rogers, et al. (1997). "Microsatellite diversity and the demographic history of modern humans." Proc Natl Acad Sci U S A 94(7): 3100-3.
- Jorde, L. B., W. S. Watkins, et al. (2000). "The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data." Am J Hum Genet 66(3): 979-88.
- Kalyanaraman, B., H. Karoui, et al. (1996). "Detection of thiyl radical adducts formed during hydroxyl radical- and peroxynitrite-mediated oxidation of thiols--a high resolution ESR spin-trapping study at Q-band (35 GHz)." Anal Biochem 241(1): 75-81.
- Kaufmann, J., K. Ahrens, et al. (1998). "CIF150, a human cofactor for transcription factor IID-dependent initiator function." Mol Cell Biol 18(1): 233-9.
- Kehrer, J. P. and L. G. Lund (1994). "Cellular reducing equivalents and oxidative stress." Free Radic Biol Med 17(1): 65-75.
- Kong, X., K. Murphy, et al. (2004). "A combined linkage-physical map of the human genome." Am J Hum Genet 75(6): 1143-8.
- Krajinovic, M., D. Labuda, et al. (2002). "Glutathione S-transferase P1 genetic polymorphisms and susceptibility to childhood acute lymphoblastic leukaemia." Pharmacogenetics 12(8): 655-8.
- Lagrange, T., A. N. Kapanidis, et al. (1998). "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB." Genes Dev 12(1): 34-44.
- Livingstone, F. B. (1976). "Hemoglobin history in West Africa." Hum Biol 48(3): 487-500.
- Luperchio, S., S. Tamir, et al. (1996). "No-induced oxidative stress and glutathione metabolism in rodent and human cells." Free Radic Biol Med 21(4): 513-9.
- Macleod, D., J. Charlton, et al. (1994). "Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island." Genes Dev 8(19): 2282-92.

- Mattey, D. L., A. B. Hassell, et al. (1999). "Association of polymorphism in glutathione S-transferase loci with susceptibility and outcome in rheumatoid arthritis: comparison with the shared epitope." Ann Rheum Dis 58(3): 164-8.
- McCarver, D. G. and R. N. Hines (2002). "The ontogeny of human drug-metabolizing enzymes: phase II conjugation enzymes and regulatory mechanisms." J Pharmacol Exp Ther 300(2): 361-6.
- Meister, A. (1991). "Glutathione deficiency produced by inhibition of its synthesis, and its reversal; applications in research and therapy." Pharmacol Ther 51(2): 155-94.
- Mills, G. C. (1957). "Hemoglobin catabolism. I. Glutathione peroxidase, an erythrocyte enzyme which protects hemoglobin from oxidative breakdown." J Biol Chem 229(1): 189-97.
- Moisio, A. L., P. Sistonen, et al. (1998). "Genetic polymorphisms in carcinogen metabolism and their association to hereditary nonpolyposis colon cancer." Gastroenterology 115(6): 1387-94.
- Nei, M. and A. K. Roychoudhury (1993). "Evolutionary relationships of human populations on a global scale." Mol Biol Evol 10(5): 927-43.
- Nishimura, J. S., E. A. Dodd, et al. (1964). "Intermediate Formation of Dipeptide-Phosphate Anhydride in Enzymatic Tripeptide Synthesis." J Biol Chem 239: 2553-8.
- Ommer, R. (1991). From outpost to outport: a structural analysis of the Jersey-Gaspé cod fishery, 1767-1886. Montréal.
- Oppenheimer, L., V. P. Wellner, et al. (1979). "Glutathione synthetase. Purification from rat kidney and mapping of the substrate binding sites." J Biol Chem 254(12): 5184-90.
- Romualdi, C., D. Balding, et al. (2002). "Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms." Genome Res 12(4): 602-12.
- Rotruck, J. T., A. L. Pope, et al. (1973). "Selenium: biochemical role as a component of glutathione peroxidase." Science 179(73): 588-90.
- Ruiz-Linares, A., D. Ortiz-Barrientos, et al. (1999). "Microsatellites provide evidence for Y chromosome diversity among the founders of the New World." Proc Natl Acad Sci U S A 96(11): 6312-7.
- Salisbury, B. A., M. Pungliya, et al. (2003). "SNP and haplotype variation in the human genome." Mutat Res 526(1-2): 53-61.
- Scriver, C. R. (2001). "Human genetics: lessons from Quebec populations." Annu Rev Genomics Hum Genet 2: 69-101.
- Shi, Z. Z., G. M. Habib, et al. (1996). "Mutations in the glutathione synthetase gene cause 5-oxoprolinuria." Nat Genet 14(3): 361-5.
- Singhal, S. S., J. T. Piper, et al. (1996). "Rabbit aorta glutathione S-transferases and their role in bioactivation of trinitroglycerin." Toxicol Appl Pharmacol 140(2): 378-86.
- Smale, S. T., A. Jain, et al. (1998). "The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes." Cold Spring Harb Symp Quant Biol 63: 21-31.

- Sprenger, R., R. Schlagenhauser, et al. (2000). "Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation." Pharmacogenetics 10(6): 557-65.
- Stephens, M., N. J. Smith, et al. (2001). "A new statistical method for haplotype reconstruction from population data." Am J Hum Genet 68(4): 978-89.
- Stoneking, M. (1994). "Mitochondrial DNA and human evolution." J Bioenerg Biomembr 26(3): 251-9.
- Stoneking, M. (2001). "Single nucleotide polymorphisms. From the evolutionary past." Nature 409(6822): 821-2.
- Strange, R. C., M. A. Spiteri, et al. (2001). "Glutathione-S-transferase family of enzymes." Mutat Res 482(1-2): 21-6.
- Suzuki, Y., T. Tsunoda, et al. (2001). "Identification and characterization of the potential promoter regions of 1031 kinds of human genes." Genome Res 11(5): 677-84.
- Tajima, A., C. S. Sun, et al. (2003). "Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese." Hum Genet 113(1): 24-33.
- Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics 123(3): 585-95.
- Takebe, G., J. Yarimizu, et al. (2002). "A comparative study on the hydroperoxide and thiol specificity of the glutathione peroxidase family and selenoprotein P." J Biol Chem 277(43): 41254-8.
- Tanaka, T., H. Yamaguchi, et al. (1993). "Flexibility impaired by mutations revealed the multifunctional roles of the loop in glutathione synthetase." Biochemistry 32(46): 12398-404.
- Tate, S. K. and D. B. Goldstein (2004). "Will tomorrow's medicines work for everyone?" Nat Genet 36(11 Suppl): S34-42.
- Uinuk-Ool, T. S., N. Takezaki, et al. (2002). "Origin and affinities of indigenous Siberian populations as revealed by HLA class II gene frequencies." Hum Genet 110(3): 209-26.
- Ursini, F., M. Maiorino, et al. (1995). "Diversity of glutathione peroxidases." Methods Enzymol 252: 38-53.
- Voris, H. K. (2000). "Maps of pleistocene sea levels in southeast Asia: shorelines, river systems and time durations." J Biogeography 27: 1153-1167.
- Wei, C. C., F. T. Chiang, et al. (1999). "The spectrum of microsatellite loci on chromosomes 7 and 8 in Taiwan aboriginal populations: a comparative population genetic study." Hum Genet 104(4): 333-40.
- Wellner, V. P., R. Sekura, et al. (1974). "Glutathione synthetase deficiency, an inborn error of metabolism involving the gamma-glutamyl cycle in patients with 5-oxoprolinuria (pyroglutamic aciduria)." Proc Natl Acad Sci U S A 71(6): 2505-9.
- West, A. G., M. Gaszner, et al. (2002). "Insulators: many functions, many mechanisms." Genes Dev 16(3): 271-88.
- White A.C., Thannickal V. J., Fanburg B.L. (1994). "Glutathione deficiency in human disease." J. Nutr. Biochem. 5: 218-226.

- Whitin, J. C., S. Bhamre, et al. (2002). "Extracellular glutathione peroxidase is secreted basolaterally by human renal proximal tubule cells." Am J Physiol Renal Physiol 283(1): F20-8.**
- Xu, S., Y. Wang, et al. (1998). "Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion." J Biol Chem 273(6): 3517-27.**
- Yunis, J. J., E. J. Yunis, et al. (2001). "Genetic relationship of the Guambino, Paez, and Ingano Amerindians of southwest Colombia using major histocompatibility complex class II haplotypes and blood groups." Hum Immunol 62(9): 970-8.**

**ANNEXES**

**ANNEXE A : Noms de chacun des échantillons d'ADN utilisés**

**Afrique** : NA10469A, NA10470, NA10471, NA10492 ; **HD50** : NA17101, NA17108, NA17109, NA17110 ; **HD12** : NA17341, NA17342, NA17343, NA17344, NA17345, NA17346, NA17347, NA17348

**Amérique** : **HD18** : NA17311, NA17312, NA17314, NA17316, NA17317, NA17318, NA17319, NA17320 ; **HD17** : NA17303, NA17304, NA17305, NA17306, NA17307, NA1730, NA17309, NA17310

**Asie** : **HD32** : NA16654, NA16688, NA16689, NA17014, NA17015, NA17016, NA17017, NA17018 ; GQ02, GQ14, GQ15, CP56, BM60, A43, BM64, BM128

**Europe** : GQ01, GQ03, GQ05, GQ07, GQ08, GQ09, GQ10, GQ11 ; **HD26** : NA15199, NA15200, NA15201, NA15202 ; **HD23** : NA13838, NA13849, NA13852, NA13876

**Moyen-Orient** : **HD03** : NA17021, NA17022, NA17023, NA17024, NA17026, NA17027, NA17028, NA17029 ; GQ04, GQ06, GQ12, GQ13, BM19, BM25, BM33, BM58

**Aborigènes de Taiwan** : **HD25** : NA13607, NA13608, NA13609, NA13610, NA13611, NA13613, NA13614, NA13615

**Islande** : **HD30** : NA15755, NA15756, NA15757, NA15758, NA15761, NA15762, NA15763, NA15764

**Acadiens** : LC3, LC4, LC5, LC7, LC11, LC16, LC17, LC20, LC26, LC27, LC28, LC29, LC30, LC32, LC35, LC36, LC37, LC38

**Canadiens-Français de Gaspésie** : JCL300, JCL301, JCL302, JCL303, JCL304, JCL305, JCL306, JCL307, JCL308, JCL309, JCL310, JCL311, JCL312, JCL313, JCL314, JCL315, JCL316, JCL318

**Charlevoix et Saguenay-Lac St-Jean** : cp2, cp42, A108, MQF55, MQF56, MQF57, MQF58, MQF59, MQF61, MQF62, MQF99, MQF135, PEDI-1, PEDI-2, II-2, III-1, III-2, CO14

**Jersey et Guernesey** : GEL202, GEL206, GEL208, GEL210, GEL214, GEL215, GEL216, GEL217, GEL218, GEL221, GEL226, GEL232, GEL237, GEL238, GEL239, GEL243, GEL248, GEL250



**Loyalistes** : DS102, DS103, DS104, DS108, DS114, DS115, DS116, DS117, DS118, DS119, DS120, DS121, DS122, DS123, DS124, DS125, DS127, DS129

**Montréal** : A131, A169, A172, A176, A180, A190, A199, A200, A265, A267, A287, A288, MQF80, MQF93, MQF96, MQF102, MQF104, MQF160, cp1, cp10, cp24, cp27, cp32, cp39, cp41, cp46, cp50, cp71, cp74, cp86, cp89, cp106, cp139, cp146, cp149, cp151

**Primates** : Chimp101, Gorille102, Orang-Outan, Gibbon

## ANNEXE B

Tableau XXXX. Origine des échantillons supplémentaires utilisés

Origine		Nombre	Provenance commerciale	
<b>Africains</b>	Congo	1	Coriell	
	Ghana	1		
	Ouganda	1		
	Éthiopie	1		
	Burkina Faso	1		
	Pygmée	7		
	Seychelles	1		
	Érythrée	1		
	Afrique Subsaharienne	9		CoriellHD12
	Afro-américain	1		CoriellHD50AA
<b>Américains</b>	Chipewyans	4	CoriellHD28 CoriellHD17	
	Cree	4		
	Mexique	5		
	Amérique du sud	3		
<b>Asiatiques</b>	Chine	4	CoriellHD13	
	Vietnam	2		
	Mongolie	5		
	Asie du Sud-Est	4		
	Cambodge	1		
<b>Européens</b>	Royaume-Uni	2	CoriellHD16	
	Espagne	1		
	Italie	3		
	Bulgarie	1		
	Québec	5		
	Grèce	3		
	France	1		
	<b>Moyen-Orientaux</b>	Égypte		4
Iran		1		
Juifs Ashkénazes		4		
Maroc		3		
Syrie		1		
Liban		2		
Arménie		1		

## ANNEXE C

Tableau XXXXI. Liste des polymorphismes étudiés

	Gène	Position	Variant		Gène	Position	Variant
1	GSS	-1979	G->A	43	GPX2	-2077	T->C
2		-1943	A->G	44		-2022	A->C
3		-1762	G->A	45		-1913	C->T
4		-1494	G->A	46		-1723	G->A
5		-1412	A->T	47		-1659	A->G
6		-1125	C->T	48		-1276	C->T
7		-792	DEL G	49		-1195	C->T
8		-716	C->G	50		-1070	G->A
9		-514	C->G	51		-843	T->C
10	GSTT1	-1773	TCAT->ACCAA	52	-724	DEL A	
11		-1554	DEL C	53	-524	A->G	
12		-1466	T->C	54	-512	G->T	
13		-1016	G->A	55	-442	T->C	
14		-874	C->T	56	-360	A->G	
15	GSTP1	-689	G->C	57	-199	C->G	
16		-2088	INS AGTCTAGTCTAGT	58	-166	A->G	
17		-1964	C->T	59	GPX3	-2005	A->G
18		-1937	G->A	60		-1958	G->A
19		-1736	A->T	61		-776	C->T
20		-1386	A->G	62		-709	A->C
21		-1231	C->T	63		-694	T->C
22		-829	C->G	64		-628	A->T
23		-828	T->C	65		-617	C->T
24		-564	G->C	66		-335	T->C
25		-286	DEL T	67		-285	T->C
26	-285	C->A	68	-69		A->T	
27	-280	C->T	69	-51		T->A	
28	GSTM3	-189	C->G				
29		-1286	C->G				
30		-859	C->T				
31	GSTM4	-208	C->G				
32		-46	G->T				
33		-1617	A->G				
34		-1614	A->G				
35		-1302	C->T				
36		-1196	C->G				
37		-823	T->C				
38		-822	T->G				
39		-821	C->T				
40		-584	T->C				
41		-195	C->G				
42		-22	G->C				

## ANNEXE D

Tableau XXXXII. Liste des amorces de PCR utilisées

Gène	Frag.	Noms	Séquence génomique de l'oligo
GPX2	1	GPX2.1F.-153	CAACCTCCTTGTTCAAACAGC
		GPX2.1R.55	AGGACTTGGCAATGAAAGC
	2	GPX2.2F.-446	TTCATCTTTTCAGATCTCCGC
		GPX2.2R.-104	TTAGCAAAAACAGGTCCCC
	3	GPX2.3F.-727	AGGAGATGGTACAGTGCAGG
		GPX2.3R.-376	GCGGGACGAATGCCTGCTGAGTGTCAAT
	4	GPX2.4F.-1002	GCCTCTTGTGAAAGCATCA
		GPX2.4R.-655	AAGTCTTTCAGTGCCTGGTG
	5	GPX2.5F.-1167	AAGCTAGGTCAAGCCCCATT
		GPX2.5R.-978	ACCCCTGATGCTTTCAACAA
	6	GPX2.6F.-1483	AGGGGACACACTTCCCTCTA
		GPX2.6R.-1117	ATGCCTTGATTTCTTCATCAGT
	7	GPX2.7F.-1778	GAAGGGTCCAAGGTCACAC
		GPX2.7R.-1404	GGCTGCTCCCTAAAAACAA
	8	GPX2.8F.-2103	GTGGCAGAATTGTGGTTGTA
		GPX2.8R.-1729	TCCTTAGGTACATGCCATCC
GPX3	1	GPX3.1F.-357	TCTACAGGCATGTCAGCTGTG
		GPX3.1R.10	TCTTTTGGCTCCAAAGTGC
	2	GPX3.2F.-625	TTCTGCTTCTAAGCCAGGTG
		GPX3.2R.-270	TCTGGGTAGGGCATTAGGAC
	3	GPX3.3F.-825	TGTGCCAGGTGGTGATTATG
		GPX3.3R.-507	GAGTGGGGAAAAAGGGAATG
	4	GPX3.4F.-1139	AGGCCAACTGCCTCACATA
		GPX3.4R.-774	ACGGAGACCAGGAGTCCTTT
	5	GPX3.5F.-1471	TTGCTCCTCATCTCTAGGCTT
		GPX3.5R.-1118	GTTATTGTGAGGCAGTTGGC
	6	GPX3.6F.-1720	CAACTCTTTAGCATCTCTCATGG
		GPX3.6R.-1379	TGGGAGCATATGAACAATGG
	7	GPX3.7F.-1944	GGTAGCATCTCAGGTCTCTGG
		GPX3.7R.-1657	TCTAGTGACCCTCAAATAGGC
	8	GPX3.8F.-2228	CACTGTCATGCTCTTTGTGC
		GPX3.8R.-1901	TCTTTAGGGGAGACAAGAAGG
GSS	1	GSS.1F.+8	CCCAACCTACTAGTTCGCC
		GSS.1R.-181	CGACGCACTCCTAATGCTAA
	2	GSS.2F.-74	GGCTAGGCTTAAAGAGCAGG
		GSS.2R.-338	CTGAGGTAGTGACTGTCCCC
	3	GSS.3F.-283	ACGACTAGTGGAAGATGGGG
		GSS.3R.-641	CCCAGATCCTGATGTAAACG

	4	GSS.4F.-622	CGTTTACATCAGGATCTGGG
		GSS.4R.-820	GGTTAATTCCCTAGAAGCTGG
	5	GSS.5F.-799	GCCCCCGCCGTCCAGCTTCTAGGGAATTAACC
		GSS.5R.-1003	CTTTTGTTCCTGCAATTTGG
	6	GSS.6F.-932	GCCCCCGCCGGTAGTGAAGGAGGTAGGGG
		GSS.6R.-1217	CTCAAGAATTCTCACGCCG
	7	GSS.7F.-1298	AAACAAAACAAACGAGCAGC
		GSS.7R.-1725	AGTCAAAAGGATCTTGGTCCC
	8	GSS.8F.-1684	GTAAATCAGTGGCCCAACC
		GSS.8R.-2051	TTATTTAAAGAGTCAAAGGGGC
GSTM3	1	GSTM3.1F.44	GAGAACCATAGACGACTCGC
		GSTM3.1R.-588	AGCAAGTATTGTTGGGTTGG
	2	GSTM3.2F.-545	CCCGTCACAACCTTCATAAGC
		GSTM3.2R.-1053	GAATGTAAACTTCCTGAGGGC
	3	GSTM3.3F-1101	GGTGGTGATACACATCCTATGG
		GSTM3.3R-1649	GCACCTCAAACCCTTAGACC
	4	GSTM3.4F-1604	CATGGGAGCACGAGTAGC
		GSTM3.4R-2036	GACTCCTTCCTCTGTCACTCC
GSTM4	1	GSTM4.1F.37	GGCTGGAAGGAGTAAGATCC
		GSTM4.1R.-533	GCACTTGTGTGTGCATGG
	2	GSTM4.2F.-503	CAAGACGATAGCACCATGC
		GSTM4.2R.-1120	GATTGTTTTATTCCCAGTGCC
	3	GSTM4.3F-1041	TCACCCATTTGGTCATTCC
		GSTM4.3R.-1619	AATTATGGAAAGGCTGAGACG
	4	GSTM4.4F.-1589	ACTTACCTTCCGTCTCAGCC
		GSTM4.4R.-1824	AAGGAAGGATAACCAGAAGC
GSTP1	1	GSTP1.1F.-407	GCAATTTCCCTTTCCTCTAAGC
		GSTP1.1R.+7	TACTCACTGGTGGCGAAGAC
	2	GSTP1.2F.-988	GCGATCAACACACACTGGT
		GSTP1.2R.-511	TTTATGACGTAGTCTCACGCTGT
	3	GSTP1.3F.-1275	GCCCCCGCCGAGTTGATGGAATAGGAGAAGC
		GSTP1.3R.-846	TTACATAGGTGTGTGCGCC
	4	GSTP1.4F.-1647	ATGCCGAGTACCCAATAGTTC
		GSTP1.4R.-1228	CCCCCGCCCGCTTGTGGCTTTAGACATGACA
	5	GSTP1.5F.-1913	GCCCCCGCCGACCCGGCTTTAAAATGTCTC
		GSTP1.5R.-1568	ACAGACACTGTGGTTGACTCC
	6	GSTP1.6F.-2208	CCCCCAGACTTGATAGGATT
		GSTP1.6R.-1828	CCCCCGCCCGTGGGTCCCTTTCCTATTTTCC
GSTT1	1	GSTT1.1F.-384	TCTCCTCCTCTAAAATGCGG
		GSTT1.1R.-35	GGAGTCAGCAAACCTCCAGTG
	2	GSTT1.2F.-711	CTCAGCAGATGAATCCAGGG

	GSTT1.2R.-343	GTGTGTCAGAATTTCCCACC
	GSTT1.3F.-1143	CCAGTTGTAGTGGCTCACATC
3	GSTT1.3R.EX1.- 660	TAGAGCCCTCCAGAGTCCTT
	GSTT1.4F.-1635	TTGGAGTTACTCTCTTGTCGC
4	GSTT1.4R.-1122	AGATGTGAGCCACTACAACCTGG
	GSTT1.5F.-2150	GCTTCCTAGGAGACTACACTGC
5	GSTT1.5R.-1610	GCTGGGCGACAAGAGAGTA

## ANNEXE E

Tableau XXXXIII. Liste des sondes utilisées pour l'ASO

	Gène	Position ( - )	Séquence	T <sub>m</sub> (°C)
1	GPX2	166	TCAGTCTGTTAATAGGA	59
			TCAGTCTGCTAATAGGA	60
2		199	CACCCCTGCTGCAACC	67
			CACCGTGCTGCAACC	67
3		360	CGTCTGTGTCAAATGAG	63
			CGTCTGTGCCAAATGAG	66
4		524	AAAGCAACAAAGAGAAG	59
			AAAGCAACGAAGAGAAG	61
5		512	GAAGAACGTGAACAG	59
			GAAGAACTTGAACAG	56
6		442	TCTGAAAGATGAAGAAG	59
			TCTGAAAGGTGAAGAAG	61
7		843	GTGGATGAACTGCACCT	66
			GTGGATGAGCTGCACCT	68
8		724	AATATAGGAGATGGTAC	59
	AATATAGGGATGGTACA		59	
9	1070	CAGAAACCGAAAAGA	56	
		CAGAAACTGAAAAGA	53	
10	1195	TTTCAATAACAAAGC	50	
		TTTCAATAATAAAGC	48	
11	1276	AGCAAGTCCGGGGAGCA	71	
		AGCAAGTCTGGGGAGCA	68	
12	1723	GGAAGAGGGAACCCCA	71	
		GGAAGAGGAAACCCCA	68	
13	1659	CAAAAGAGCAGAGAATA	59	
		CAAAAGAGCGGAGAATA	61	
14	2077	CACTGAATTGGAATC	56	
		CACCGAATTGGAATC	59	
15	2022	GAGAATTGTTTGTAGGA	59	
		GAGAATTGGTTGTAGGA	61	
16	1913	TAAGCAGCGTATGATTC	61	
		TAAGCAGCATATGATTC	59	
17	GPX3	285	GGGCATTAGGACTGT	61
			GGGCATTGGGACTGT	64
18		69	TCAGTTCTAGGGAGCTA	63
			TCAGTTCTTGGGAGCTA	63
19		51	CAGGGCAAGGGGGCTAA	71
			CAGGGCATGGGGGCTAA	71
20		335	CATAAGGATCACAGC	59
			CATAAGGGTCACAGC	61
21		776	GACAGACGGAGACCA	64
			GACAGACAGAGACCA	61

22		709	GATGGGATTGATTAGTG	61
			GATGGGATGGATTAGTG	63
23		694	CATTACATAGCTGAGG	60
			CATTACACAGCTGAGG	62
24		628	TTTATTTTCTGCTTCT	52
			TTTTTTTTCTGCTTCT	52
25		617	TGGCTTAGAAGCAGA	59
			TGGCTTAAAAGCAGA	56
26		1958	TTGCTACGAAAAGTA	56
			TTGCTACAGAAAAGTA	53
27		2005	GTCCACCTATTTAAC	56
			GTCCACCTGTTTAAAC	59
28	<b>GSS</b>	514	CTGGGATACCAGCAGGG	71
			CTGGGATAGCAGCAGGG	71
29		792	ATGGAATGTATCAGT	53
			ATGGAATTATCAGTC	53
30		1125	TAGAGACGGGGTTTC	61
			TAGAGACAGGGTTTC	59
31		1412	CAACAGGACTTAGGC	61
			CAACAGGTCTTAGGC	61
32		1494	AGGGGAAGGACATGG	64
			AGGGGAAAGACATGG	61
33		1762	AACAACAACGAAAACCTC	59
			AACAACAATGAAAACCTC	56
34		1943	CAGACCAATCTGGCCAA	66
			CAGACCAACCTGGCCAA	68
35		1979	TCCACCCGCCCCAGC	72
			TCCACCCACCCCAGC	69
36		716	AGGTGCCTTCCCTGTTT	61
			AGGTGCCTTCCGTGTTT	64
37	<b>GSTM3</b>	859	TAAATGCGTGAGCAT	56
			TAAATGCATGAGCAT	53
38		1286	CACAGCAGAGTGGCT	64
			CACAGCACAGTGGCT	64
39		46	TATACCCGACATAAG	56
			TATACCCTACATAAG	53
40		208	AGGAGTGCTGCAGGA	64
			AGGAGTGGTGCAGGA	64
41	<b>GSTM4</b>		GCCTTTCCATAATTG	56
		1614 - 1617	GCCTTTCCGTAATTG	59
42			GCCTTTCCATAGTTG	59
			GCCTTTCCGTAGTTG	61
43		22	TCCGGAAGCACTAGG	64
			TCCGGAACCACTAGG	64
44		195	AGCCGAGAGGCAGAAGG	71



		AGCCGAGACGCAGAAGG	71
45	1196	AAAATACGGGACACT	56
		AAAATAGGGGACACT	56
46	1302	TGAGGAGCCAGGAAGAC	68
		TGAGGAGCTAGGAAGAC	66
47	584	TAAGAATGTGTCATT	50
		TAAGAACGTGTCATT	53
		ACTGGTTTTCGAACTC	59
48		ACTGGTTGCGAACTC	61
		ACTGGTCTTGAACTC	59
49	821-822 -823	ACTGGTCTCGAACTC	61
		ACTGGTCGCGAACTC	64
		ACTGGTCGTGAACTC	61
50		ACTGGTTTTTGAACTC	56
		ACTGGTTGTGAACTC	59
51	<b>GSTP1</b>	CGCGCCAGTTCGCTG	72
	285 -286	CGCGCCAGTTAGCTG	69
52		GCGCGCCAGTCGCTG	69
		GCGCGCCAGTAGCTG	67
53	280	GCTGCGCACACTTCG	64
		GCTGTGCACACTTCG	61
54	189	CGGCCAGCTGCGCGG	75
		CGGCCAGGTGCGCGG	75
55		ACATGTGGAGGATGT	61
	828 - 829	ACATGTGCAGGATGT	63
56		ACATGTAGAGGATGT	59
		ACATGTACAGGATGT	57
57	564	AGCCGCCGAGATCGC	69
		AGCCGCCCAGATCGC	69
58	1386	GTTTTAAAGATACAA	48
		GTTTTAAGGATACAA	50
59	1736	CCAAAACAGAAATAA	50
		CCAAAAC T GAAATAA	50
60	1231	GTCTAAAGCCAACAAG	56
		GTCTAAAGCCAATAAG	53
61	1964	GGTTTTACGGCTCAG	61
		GGTTTTATGGCTCAG	59
62	1937	AACCTGCGGACATGT	61
		AACCTGCAGACATGT	59
63	2088	TAGTCTAGTCTAGTT	53
		ATTCTAGTTTCGCCG	59
64	<b>GSTT1</b>	CAGGGGGGTGGACTT	67
	689	CAGGGGGCTGGACTT	67
65	1016	ACAGGTGCGCACAAA	61
		ACAGGTGTGCACAAA	59

66	874	TAAAACAGTTGTTAC	50
		TAAAATAGTTGTTAC	48
67	1466	CAGAGATAGGGTTTC	59
		CAGAGATGGGGTTTC	61
68	1554	AGAATCGCTTGAACCTC	61
		AGAATCGTTGAACCTC	61
69	1773	TGTGTGGGATGATCGTTTGT	68
		TGTGTGGGTTGGTTCGTTTGT	70